CURSOS DE VERÁN

Estudios de lógica borrosa y sus aplicaciones

Edición a cargo de
ALEJANDRO SOBRINO E SENÉN BARRO

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

RSIDADE
ANTIAGO
MPOSTELA
icacións
P.

Estudios de lógica borrosa y sus aplicaciones

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CURSOS E CONGRESOS DA
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Nº 78

SERIE CURSOS DE VERÁN

Estudios de lógica borrosa y sus aplicaciones

Edición a cargo de
ALEJANDRO SOBRINO E SENÉN BARRO

1993

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ESTUDIOS de lógica borrosa y sus aplicaciones / Edición a cargo de Alejandro Sobrino e Senén Barro. – Santiago de Compostela : Universidade, Servicio de Publicacións e Intercambio Científico, 1993. – 378 p. ; 24 cm. – (Cursos e Congresos da Universidade de Santiago de Compostela ; 78). – Índice: p. 5-6. – D.L.: C. 1230-1993. – ISBN 84-8121-018-8.

I. Sobrino, Alejandro, ed. lit; II. Barro, Senén, ed. lit; III. Universidade de Santiago de Compostela. Servicio de Publicacións e Intercambio Científico, ed.

1. Lóxica simbólica e matemática; 2. Conxuntos difusos, Teoría de
510.64

(PUBLICACIÓNS EN LÓXICA ; 2)

© Universidade de
Santiago de Compostela, 1993

EDITA: UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
Servicio de Publicacións e Intercambio Científico
Campus Universitario

IMPRIME: IMPRENTA UNIVERSITARIA
Campus Universitario
Pavillón de Servicios

ISBN: 84-8121-018-8
Depósito Legal: C. 1230-1993

INDICE

PREFACIO 7

CAP. I. TEMAS DE LÓGICA BORROSA Y RAZONAMIENTO APROXIMADO: PASADO, PRESENTE Y PROYECCIONES FUTURAS
Alejandro Sobrino, José Angel Olivas y Carmen Adán 13

APLICACIONES DE LA LÓGICA BORROSA EN CONTROL

CAP. II. APLICACIÓN DE LA LÓGICA BORROSA EN CONTROL DE PROCESOS Y ROBÓTICA
Alfonso J. García Cerezo y Aníbal Ollero 51

CAP. III. CONTROL Y SUPERVISIÓN DE PROCESOS INDUSTRIALES CON LÓGICA BORROSA
Antonio Barreiro y Rafael Sanz 69

CAP. IV. FABRICACIÓN AUTOMATIZADA DE AMPLIFICADORES ELECTRÓNICOS EN TELEVÉS UTILIZANDO LÓGICA BORROSA
Ricardo García Rosa y Pedro F. Zuliani 95

CAP. V. APLICACIÓN DEL CONTROL BORROSO EN MEDICINA
Carlos Hernández y Andrés Carollo 105

OTROS ÁMBITOS DE PROYECCIÓN DE LA LÓGICA BORROSA

CAP. VI. CLASIFICACIÓN BORROSA BASADA EN CONOCIMIENTO
Ramón Ruiz 123

CAP. VII. RAZONAMIENTO TEMPORAL BORROSO
Roque Marín, Senén Barro y José Mira 163

CAP. VIII. COMPUTACIÓN NEURONAL BORROSA
José Mira, Ana E. Delgado, Roque Marín y Senén Barro 197

CAP. IX. UNCERTAINTY IN C&L MACHINES
Panos A. Ligomenides 225

CAP. X. EL MODELO DIFUSO COMO PROGRAMA DE INVESTIGACIÓN LINGÜÍSTICA
Teresa Moure 245

CAP. XI. APLICACIÓN DE LA LÓGICA BORROSA A LA SELECCIÓN DE CARTERAS
José A. Redondo y José C. de Miguel 275

HARDWARE BORROSO

CAP. XII. SOLUCIONES HARDWARE EN CONTROL BORROSO
Alberto Bugarín, Senén Barro y Ramón Ruiz 295

CAP. XIII. EL PROYECTO ORBE
Claudio Moraga 327

TRABAJOS ESPECIALIZADOS EN LÓGICA BORROSA

CAP. XIV. UNA APROXIMACIÓN T-NORMA A LA LÓGICA BUNGEANA DE MEDIOS Y FINES ¿HACIA UNA METODOLOGÍA HEURÍSTICA EN ÉTICA?
Enric Trillas 335

CAP. XV. ON SOME BISYMMETRIC FUNCTIONS ON CLOSED INTERVALS
Sergei Ovchinnikov 353

CAP. XVI. ON SOME MEASURES OF ENTROPY RELATED TO AGGREGATION OPERATORS
Ronald R. Yager 365

PREFACIO

El libro que ahora presentamos al lector tiene su origen, en su mayor parte, en un curso de verano de la Universidad de Santiago de Compostela celebrado en Julio de 1992 y titulado 'La lógica borrosa en Galicia: teoría y aplicaciones', del que los encargados de la edición de este texto fueron sus directores. El objetivo del curso era reunir en un foro de discusión común a investigadores de las Universidades Gallegas que, aún teniendo una ubicación científico-profesional diferente, participan de un nexo común: todos estudian aspectos de la lógica borrosa o utilizan esta herramienta formal para el análisis o aplicación a algún dominio conceptual propio de su profesión. Sin embargo, en el curso tuvimos también el placer de contar con profesionales de otras universidades o centros de investigación, españoles o europeos, con alguna vinculación con las Universidades Gallegas, bien por haber ejercido su labor profesional tiempo atrás en ellas, bien por su participación, colaboración o asesoramiento en proyectos de investigación. El libro se completa con tres contribuciones de profesores de universidades estadounidenses que, aunque no participaron en el curso, mantienen una estrecha relación científica con algunos de los ponentes.

La descripción de la génesis del libro puede explicar la diversidad de temas que, bajo el título de **Estudios de lógica borrosa y sus aplicaciones**, se incluyen en este texto. Difícilmente un título puede ser fiel reflejo de un contenido tan diverso como el que a continuación se presenta. Sin embargo, el significado de amplio espectro que se tiende a asociar al término 'lógica borrosa', es una ayuda en esta difícil tarea, en la medida en que éste va mucho más allá de los límites estrictamente asociados a estudios o aplicaciones 'canónicos' en lógica borrosa. Evidentemente, esta diversidad no permite un tratamiento conceptual ascendente y coherente del texto, como sí suele ser usual en un libro de un único autor y de contenido más homogéneo. En efecto, en él se incluyen desde trabajos puramente teóricos, de desarrollo conceptual y formal propio de la lógica borrosa, hasta artículos donde se describen aplicaciones muy concretas. Esta circunstancia nos ha inducido a pensar que sería aconsejable proporcionar un índice informativo al lector, en el que se han

tratado de agrupar los capítulos de contenido más afín. Además, pensando que esta diversidad favorecerá el que muchos lectores se acerquen al libro con un interés en trabajos concretos, los encargados de la edición del texto hemos pedido a los autores que sus trabajos fuesen, en la medida de lo posible, autocontenidos. Se trata, así, de facilitar una lectura aislada de los mismos, aunque en algunos trabajos altamente especializados en lógica borrosa, como los de los profesores Ovchinnikov, Trillas o Yager, se deba poseer, además, una familiaridad notable con algunos ámbitos de esta lógica.

Tres de los trabajos incluidos, los de los profesores estadounidenses Ligomenides, Ovchinnikov y Yager, no han surgido de su participación en el curso de verano que fue la génesis de este libro, sino que sus contribuciones han respondido a una petición expresa de los responsables de esta edición. Su excelente disposición para participar en este proyecto es una actitud que no podemos sino agradecer encarecidamente. El que sus trabajos no hayan sido traducidos al castellano ha sido por el deseo de respetarlos en su versión original, entendiendo que su contenido altamente especializado interesará a lectores para los que muy probablemente no será un problema mayor el que estén escritos en inglés.

Continuando con el idioma, haremos a continuación un breve comentario sobre la problemática de la traducción de algunos términos ingleses. Nos referimos, en particular, a la traducción de 'fuzzy', 'fuzzyfication' y 'defuzzyfication'. Respecto a la traducción de 'fuzzy' se ha convenido en hacerlo por 'borroso' cuando su contexto de uso lo sea en un medio formal y por 'difuso' cuando lo sea en un ámbito teórico no formal. Así, p. ej., se ha uniformizado la traducción de 'fuzzy logic' por lógica borrosa; en cambio, en el caso de desarrollos relacionados con una lingüística no discreta se usa 'difuso' en vez de 'fuzzy'. En ambos casos, se ha adoptado una convención que no vulnera el uso correcto del castellano. Más dudas ofrece la traducción de los términos 'fuzzyfication' y 'defuzzyfication', pero antes permítasenos hacer otro comentario sobre la traducción de 'fuzzy logic'.

Se traduce 'fuzzy logic' por 'lógica borrosa' y no por 'lógica fuzzy' porque pensamos que la posible confusión de atribuir borrosidad a una lógica, atribuir un carácter poco explícito o equívoco a sus fórmulas y reglas, forma parte de prejuicios ya pasados o de un desconocimiento, que la sustitución de 'borroso' por 'fuzzy' no ayudará a despejar. Es de normal entendimiento que por lógica borrosa se quiere decir lógica que estudia la relación de consecuencia en argumentos que incluyen enunciados vagos; de la misma forma que por lógica temporal se entiende lógica que estudia la relación de consecuencia en argumentos que incluyen enunciados que hacen alusión a situaciones temporales. Pero, así como sería insensato pensar que la lógica temporal caduca, como sí le pasa a un yogur, de la misma forma, nadie que se dedique a estudiar lógica puede sentirse guiado al equívoco por un rótulo

como lógica borrosa: es obvio, como indicó A. Deaño, que no se trata de 'explicar la tartamudez tartamudeando'.

Más problemático ha sido adoptar una convención para la traducción de 'fuzzyfication' y 'defuzzyfication', donde se entabla una pugna entre la corrección lingüística y el uso impuesto por la necesidad de discriminar significados. Así, una traducción acorde con el castellano aconsejaría, por ejemplo, traducir 'defuzzyfication' por 'discretear' o 'discontinuizar'. Pero los beneficios que supondría esta traducción, se verían contrarrestados por su ambigüedad en el dominio conceptual de varios de los trabajos de este libro, donde por los términos 'discretear' o 'discontinuizar' es frecuente (en un uso ya consolidado), entender cosas bien diferentes, incluso contrapuestas a lo que la traducción del término inglés 'defuzzyfication' debería señalar. Para evitar posibles equívocos, hemos adoptado la convención, a sabiendas de que con ello forzamos al castellano –como ocurre con cierta frecuencia, por otra parte, con otros términos que se acuñan en áreas de producción mayoritariamente anglosajonas– de traducir 'fuzzyfication' por 'borrosificación' y 'defuzzyfication' por 'desborrosificación'. Otra opción, empleada en algunos textos originales, era usar las palabras 'fuzzyficación' y 'defuzzyficación', traducción que no nos pareció aconsejable, ya que ni son palabras inglesas ni se asemejan a palabras castellanas.

Finalmente, a modo de breve guía introductoria para el lector, pasamos a describir la estructura del libro y a comentar los trabajos en él incluídos.

El texto se inicia con el trabajo *Temas de lógica borrosa y razonamiento aproximado*, en el que se intentan rastrear algunos antecedentes históricos al trabajo de Zadeh en la preocupación por la vaguedad lingüística y su modelización formal. Se apuntan también algunos problemas actuales de razonamiento aproximado, relacionados fundamentalmente con la caracterización de la lógica borrosa como una lógica de valores de verdad lingüísticos. Por último, se comentan brevemente algunos temas que están siendo o podrían ser objeto de investigación en los próximos años, como el fuzzy-Prolog o la cuasi-tautologicidad.

Cuatro de los capítulos del libro abordan el control borroso, lo que no hace sino reflejar el hecho de que el control de procesos es el ámbito donde se asientan la mayoría de las aplicaciones actuales de la lógica borrosa. El primer trabajo en esta línea es *Aplicación de la lógica borrosa en el control de procesos y robótica*, en el que se expone una taxonomía de los controladores borrosos según su modo de operación, subrayándose, además, el interés del razonamiento aproximado en general y el borroso en particular, en la robótica móvil y en la navegación autónoma.

El trabajo *Control y supervisión de procesos industriales con lógica borrosa* aborda la exposición de distintas estrategias para el control y la supervisión de

procesos industriales que se integran en lo que ya es usual denominar como 'controladores inteligentes'. Se contextualiza, por tanto, el papel que la lógica borrosa comparte con los sistemas expertos en tiempo real y, más recientemente, con las redes neuronales artificiales.

El éxito del control borroso se ha hecho patente, en buena medida, cuando las aplicaciones dejaron de ser simples proyectos de laboratorio y han afianzado sus posibilidades de comercialización, desde el control de procesos industriales hasta objetos de consumo tan próximos a nosotros como cámaras, aspiradoras, sistemas de aire acondicionado, etc. La casi totalidad de estos productos son japoneses, pero en otros países, entre ellos España, se están empezando a desarrollar productos basados en la lógica borrosa. Un ejemplo se refleja en el trabajo titulado *Fabricación automatizada de amplificadores electrónicos en Televés utilizando lógica borrosa*, donde se presenta un sistema que automatiza el proceso de ajuste de amplificadores electrónicos, desarrollado para la empresa Televés, con sede en Santiago de Compostela.

Siguiendo en el área del control, pero vinculado al ámbito de la medicina, se enmarca el trabajo *Aplicación del control borroso en medicina*. En este dominio las aplicaciones del control borroso son más recientes, como lo son, en general, las estrategias de control aplicadas a pacientes, aunque las aplicaciones comienzan a ser abundantes: relajación muscular en anestesia, presión arterial y dolor post-operatorio, son algunas de las que se recogen en este trabajo.

Bajo el epígrafe de «Otros ámbitos de proyección de la lógica borrosa», se incluyen un grupo de seis capítulos en los que los temas de estudio o aplicación son muy variopintos, lo que da muestra de la potencialidad de los conceptos y herramientas de la lógica borrosa para el análisis y síntesis en diversos ámbitos de conocimiento. El artículo *Clasificación borrosa basada en conocimiento*, pone de manifiesto el desarrollo de modelos borrosos como una forma natural e intuitiva de formular y resolver problemas de clasificación y reconocimiento de patrones, a la vez que permiten el desarrollo sencillo de esquemas heurísticos de análisis, compatibles con los procesos de comunicación de los expertos en tales problemas.

En el trabajo *Razonamiento temporal borroso* se revisa el conocimiento actual en área del razonamiento temporal. Se parte de perspectivas que se podrían denominar convencionales hasta llegar a analizar el interés y las posibilidades de aplicar el modelo formal de la lógica borrosa al tratamiento de la información temporal imprecisa, con frecuencia presente en el lenguaje natural cuando nos referimos al tiempo y a la ordenación temporal de hechos.

El capítulo *Computación neuronal borrosa* pretende dar una visión integral de la computación neuronal en su doble aspecto, biológico y computacional. En

ambas perspectivas se ilustra la posible utilidad de la matemática borrosa para ayudar en la representación, análisis y síntesis de problemas relacionados con las redes neuronales.

En el trabajo *Uncertainty in C & L Machines* se recoge una propuesta singular: en aplicaciones de la Inteligencia Artificial, la incertidumbre asociada a procesos de inferencia inductiva es tratada a menudo mediante métodos heurísticos. Para simular estos procesos en computadores digitales se acude a teorías de la probabilidad, de la posibilidad o de la creencia. En este trabajo se sugiere la posible conveniencia de que la incertidumbre pueda ser manipulada de modo natural e, incluso, que pueda ser utilizada de modo beneficioso en sistemas físicos dinámicos, autoorganizativos y cooperativos, los cuales pueden ser regulados para funcionar como máquinas capaces de computar y aprender.

Uno de los propósitos de la investigación lingüística tradicional era establecer una categorización de los distintos elementos que en el análisis del lenguaje podían ser distinguidos. De esta actividad resultaban unas taxonomías discretas. Esto ha facilitado la explicación de la lengua, pero ha privado a la Lingüística de aquellas manifestaciones del lenguaje –por lo demás, bastante frecuentes– que incluyen algún tipo de indeterminación. El artículo *El modelo difuso como programa de investigación lingüística* plantea algunos presupuestos sobre los que tendría que conformarse una lingüística no discreta y cómo su diseño repercutiría, en general, sobre el área de estudio de lo difuso.

El trabajo *Aplicación de la lógica borrosa a la selección de carteras* utiliza y evalúa la aplicabilidad de la lógica borrosa en decisiones financieras, más concretamente en las inversiones, describiéndose la aplicación de esta lógica a la selección de una cartera de riesgo para el caso particular de dos activos financieros bajo borrosidad.

El coste en tiempo de cálculo que imponen multitud de soluciones basadas en lógica borrosa, plantea la necesidad de desarrollar *hardware* dedicado a esta tarea. Por ejemplo, en sistemas de control, donde la necesidad de operar en tiempo real ha propiciado el diseño y desarrollo de numerosos dispositivos *hardware* capaces de ejecutar la base de conocimiento de un controlador borroso. En el trabajo *Soluciones hardware en control borroso* se realiza una revisión de los desarrollos más destacables en esta línea, aportando soluciones propias. Se refleja la gran homogeneidad que existe en la mayoría de las propuestas de implementación y realizaciones físicas existentes actualmente para la evaluación de la base de reglas en sistemas de control borroso. No obstante, a medida que surgen nuevas aplicaciones de la lógica borrosa se demandan implementaciones hardware más flexibles y potentes. También en este punto Japón marca la pauta, pero, de nuevo, tampoco España queda al margen, como se pone de manifiesto en el trabajo *El proyecto ORBE*.

Este trabajo, tal como le hemos pedido a su autor, no persigue tanto una descripción detallada de desarrollos específicos, cuanto el relato del origen, planteamiento, participantes, logros ya alcanzados y previsiones de futuro del Proyecto de Innovación Tecnológica ORBE (ORdenador Borroso Experimental). Este proyecto persigue un doble objetivo: desarrollar arquitecturas para el procesamiento de información borrosa que vayan más allá de las soluciones estrictamente dedicadas al control borroso y servir de nexo a una amplia comunidad de investigadores dedicados al estudio de la lógica borrosa.

Finalmente, los tres últimos capítulos del libro recogen algunos temas avanzados y bastante especializados relacionados con diferentes aspectos de la lógica borrosa. El artículo *On some Bisymmetric Functions on Closed Intervals* presenta algunos resultados relacionados con el uso de operaciones binarias sobre intervalos cerrados de números reales. El principial objetivo de este trabajo es desarrollar una aproximación unificada de las t-normas, t-conormas y medias.

El trabajo *Una aproximación t-norma a la lógica bungeana de medios y fines. ¿Hacia una metodología heurística en Ética?* constituye un intento de asomarse a una posible lógica deóntica borrosa, aunque restringida en este caso a las propuestas de Bentham y Bunge. Aporta interés, por tanto, como un apunte a una posible extensión de la lógica borrosa a un dominio todavía inédito: la ética. La formalización y desarrollo de conceptos frecuentes en el ámbito de una teoría de la decisión bajo incertidumbre extiende también su interés a esta área.

En el artículo *On some Measures of Entropy related to Aggregation Operators* se sugieren nuevas medidas para el cálculo de la entropía de una distribución de posibilidad que puedan servir de instrumentos para la aplicación del pricipio de máxima entropía. Se estudian tres clases de medidas de entropía, basadas, respectivamente, en el operador OWA, en la distancia de una distribución de posibilidad a partir de una distribución uniforme y, por último, en el concepto de t-conorma.

En otro orden de cosas, y para terminar con este prefacio, los responsables de la edición del texto queremos manifiestar nuestro agradecimiento a los autores de los distintos trabajos, que con su contribución generosa han hecho posible este libro. Esta gratitud se extiende al Vicerrectorado de Política Cultural de la Universidad de Santiago de Compostela, quien facilitó los recursos organizativos y financieros necesarios para la organización del curso de verano, cuyas conferencias fueron el embrión oral de estos textos escritos. Gracias también a los revisores anónimos del libro, que con sus comentarios y sugerencias han contribuido a mejorarlo. Finalmente, en la labor tediosa e inacabable, pero no por ello menos útil e imprescindible, de conformación final del texto, bibliografía y figuras, ha contribuido José Felípez, ayudado por Carlos V. Regueiro y Alberto Bugarín. Gracias también a ellos por su labor generosa y desinteresada.

Capítulo I

TEMAS DE LÓGICA BORROSA Y RAZONAMIENTO APROXIMADO: PASADO, PRESENTE Y PROYECCIONES FUTURAS

Alejandro Sobrino, José Ángel Olivas y Carmen Adán
Área de Lógica y Filosofía de la Ciencia. Universidad de Santiago de Compostela

1. JUSTIFICACIÓN

Sin duda, y por lo abarcador de su título, los objetivos de este trabajo pueden parecer pretenciosos. Dado el complejo y voluminoso desarrollo que alcanzan casi todas las áreas de la lógica borrosa, parece ya una tarea difícil hablar en profundidad de alguno de sus temas. De ahí que los artículos tengan habitualmente títulos más restringidos. Podría parecer entonces que, al rotular nuestra contribución como *Temas de lógica borrosa y razonamiento aproximado*, estamos transgrediendo normas de prudencia habitualmente respetadas.

Lejos de pretender mostrar una actitud desmesurada o imprudente, el título de este trabajo responde a dos razones: una, el carácter interdisciplinar y diversificador de la investigación que nos ha ocupado hasta el momento, pero que al fin no es otra que el análisis lógico del lenguaje ordinario y de la racionalidad común. En esta aventura nos hemos encontrado con objetos de estudio diferenciados que demandaban diferentes instrumentales teóricos y, por tanto, distintas vías de estudio. La segunda razón atiende a la conveniencia de que, dada la relativa disparidad de temas que serán presentados, pensamos que sería bueno que, en la medida de lo posible, fueran autocontenidos. Por estos dos motivos hablaremos de 'temas' y no de 'un tema' de lógica borrosa y razonamiento aproximado.

No obstante, esperamos no sea difícil entrever el hilo conductor que une a los mismos. Nuestra pretensión es que, al final del trabajo, los lectores observen que los diferentes tópicos estudiados tienen un vínculo común: indagar acerca de la naturaleza del lenguaje y las propiedades formales relativas a su carácter aproximado o vago; en opinión de Zadeh, una de las características definitorias del cerebro humano frente al 'cerebro' electrónico. El interés acerca de cómo el lenguaje vago, como característica humana especie-específica, puede ser simulado en un artefacto interactivo de la I. A., debe mostrar además nuestra preocupación porque los interfaces de comunicación persona-máquina acerquen cada vez más esos dos tipos de 'cerebros'.

Por último, indicar que el trabajo se acompaña de una serie de *[Propuestas]* de investigación (señaladas con esa tipografía) que creemos constituyen, en general, retos interesantes dentro de la teoría de lo borroso.

2. OBJETIVO

El objetivo de este trabajo es mostrar, de un modo resumido y –como ya se ha anunciado– en lo posible integrador, una parte importante de la investigación que hemos realizado sobre vaguedad lingüística (la materia prima de estudio), lógica borrosa (el instrumento para estudiar esa materia prima) y el razonamiento aproximado (una interesante forma en la que se presenta la materia prima desde el punto de vista de la lógica y, más en concreto, de las lógicas para la I.A.). La investigación se aglutina en torno a un proyecto de Investigación, financiado respectivamente por la D.G.I.C.Y.T. y la Consellería de Educación de la Xunta de Galicia y de título: "Fundamentos de sistemas expertos: formalización y gestión de la vaguedad en el razonamiento aproximado".

3. PREOCUPACIÓN INICIAL: LA VAGUEDAD LINGÜÍSTICA

Nuestro primer acercamiento al tema de la vaguedad fue a través del texto de A.Deaño [1978], *Introducción a la lógica formal*, que en su reedición de 1978 (en una parte titulada 'Más allá de este libro', por ser una incorporación póstuma), comentaba brevemente las ideas centrales de la lógica borrosa desde distintos puntos de vista: desde la propia lógica, la lingüística y la psicología. No obstante, aún prescindiendo de este capítulo, la primera edición del libro ya dejaba traslucir su interés por el uso de la lógica en el análisis formal del lenguaje ordinario. Su

texto, quizás a contracorriente en aquella época, más que hacer énfasis en una presentación *more matemático* de la lógica, acentuaba su utilidad como instrumento clarificador del lenguaje. De ahí que tratase siempre de ilustrar los conceptos lógicos con ejemplos extraídos de la literatura, forma de expresión mundana por la que tanta pasión sentía y de la que era un excelente conocedor. Fue quizás la preocupación de Deaño por el lenguaje ordinario y sus lógicas, la que decantó los primeros pasos de nuestra inmersión en la vaguedad.

La principal preocupación cuando iniciamos nuestra investigación en el área de lo borroso y en las lógicas borrosas fue acercarnos a lo que antes denominamos materia prima de estudio: la vaguedad lingüística. Consultados algunos textos propios de lingüística y de filosofía, constatamos que el tema de la vaguedad tenía una presencia escasa en sus índices, cuando no inexistente. Esto nos llevó a plantear una primera cuestión: ¿Es quizás la vaguedad un tema reducible a otros aspectos del lenguaje o, por el contrario, tiene una identidad propia?. Contestar a esta pregunta parecía interesante, aunque el caudal de trabajos que había sobre lógica borrosa cuando iniciamos está investigación hacía pensar razonablemente en que, en efecto, constituía una característica del lenguaje con entidad autónoma. No obstante, más allá de los trabajos en lógica borrosa que daban por bueno el *dictum* de Zadeh de la inevitabilidad de la vaguedad, en la bibliografía que en aquellos momentos había sobre el tema, esto no era algo obvio (hoy tampoco lo es para todo el mundo, aunque cada día parecen ser menos los escépticos).

La investigación fue, entonces, emprendida a un doble nivel:

1) Examinar si en lingüística era considerado, y de qué forma, el tema de la vaguedad.

2) Estudiar si en filosofía (y, más en concreto, en filosofía analítica, por las razones que apuntaremos a continuación) la vaguedad era objeto de debate en alguno de los círculos de pensamiento más relevantes.

3.1. La vaguedad en lingüística

Aunque en los manuales de lingüística no aparecían casi nunca capítulos dedicados al tema de la vaguedad (y sabemos, desde la obra de Kuhn [1962], *La estructura de las revoluciones científicas*, la importancia que esto tiene para señalar si un tema es generalmente aceptado por una comunidad científica, en tanto que los libros de texto reflejan los contenidos no cuestionados del paradigma o paradigmas dominantes; esto es, de lo que él ha llamado técnicamente 'período de ciencia

normal'), en el ámbito de la investigación sí había preocupación por el tema. Este interés tenía dos notas:

1) Era un interés participado, *grosso modo,* por dos escuelas importantes de Lingüística: la generativista de Chomsky y el Círculo de Praga.

2) A pesar del interés de ambas escuelas, el énfasis en el problema de la vaguedad siempre fue limitado.

En el caso de los generativistas se redujo a unos breves trabajos de Chomsky o Katz (entre los más representativos) y en el caso del Círculo de Praga a una contribución un poco más extensa, pero globalmente también escasa. Aunque cronológicamente posteriores a los trabajos citados, es necesario destacar también las aportaciones a este tema de Ross y de Lakoff. Todas estas contribuciones tuvieron eco en España en un informadísimo e interesante trabajo de Bosque [1979], al que desafortunadamente el autor no dio continuidad.

Lo indicado antes nos permite destacar lo siguiente: ninguna de las escuelas ni lingüistas mencionados logró proponer unas bases amplias para el diseño de una lingüística no discreta. En un área tan importante para nosotros como es la de la ciencia que estudia esa materia prima con la que trabajamos, el lenguaje, alguien debería hacer este trabajo. Afortunadamente, en esta tarea está empeñada T. Moure, cuya contribución a este curso es una todavía inicial pero sin duda interesante muestra.

Por último, nos parece interesante señalar que los trabajos de las escuelas generativista y de Praga datan de finales de los 50 e inicio de los 60, época en la que Zadeh publicó sus primeros estudios sobre la teoría de conjuntos borrosos. Varias preocupaciones eran comunes en la misma época, sobre un mismo tema y entre comunidades científicas que no se conocían: siguiendo otra vez a Kuhn, quizás se pudiese decir (desde la diacronía) que se estaban gestando las primeras bases para el establecimiento de un paradigma que estudiase, de alguna forma, el tema de la vaguedad.

Pero ya antes, y también en un campo diferente, se había mostrado una preocupación por la vaguedad: era el ámbito de la filosofía analítica.

3.2. La vaguedad en filosofía analítica

Es común decir que uno de los principales objetivos de la filosofía analítica es escudriñar lo que, en otras áreas de la filosofía en particular y de otras ciencias

en general, se ha escrito o dicho acerca de su objeto de su estudio. Por eso, la tarea de análisis que emprende esta filosofía no lo es tanto acerca de un tema como de un instrumento: del lenguaje en que se expresan otras ciencias y de los argumentos que se construyen con ese lenguaje. Del examen no se escapa la propia filosofía, necesitada de esta propedéutica quizás más que ninguna otra disciplina, ya que, debido a su tradición sofista, se articula con frecuencia argumentativamente (aunque no siempre correctamente), a fin de persuadir y convencer. La filosofía analítica emprende así una tarea de desinfección del lenguaje, de destape de aquellas falacias presentes en razonamientos de la filosofía, la física o la matemática.

En esta empresa no es de extrañar que, si el tema de estudio era el lenguaje argumentativo y su corrección formal, la herramienta a utilizar fuese la lógica. La lógica participaba, ya desde Frege, en el análisis de argumentos formalmente correctos bajo la siguiente máxima: para razonamientos formados por proposiciones precisas; esto es, susceptibles de ser catalogadas como verdaderas o falsas, un argumento correcto es aquel que, de asumir que las premisas son verdaderas, la regla que transporta información de premisas a conclusión no nos puede llevar a una conclusión falsa. En este contexto, parecía que la vaguedad debería estar ausente de las preocupaciones de estos filósofos. Mas adelante veremos, sin embargo, que esta predicción no se cumple.

Lo dicho en el párrafo anterior entronca con otro presupuesto, sin duda también asumido por esta escuela de pensamiento: Para ellos, el lenguaje científico es, si se me admite esta expresión radical, 'superior' al lenguaje ordinario, pues es el lenguaje de la ciencia el que mejor expresa un conocimiento más profundo del mundo, nuestras ideas más depuradas. El lenguaje ordinario, en cambio, está lleno de vaguedad, lo que propicia, hasta cierto punto, su carácter a veces contradictorio e (in)coherente. Por tanto, si bien es útil en la comunicación ordinaria, deberíamos prescindir del mismo para fines más serios, precisamente aquellos que les interesan a los analíticos y que no son otros que los propios de la actividad científica.

3.2.1. La vaguedad en Russell

No obstante, la despreocupación que cabía esperar de los analíticos por la vaguedad (una vez eliminada) no se cumplió, bien es verdad que sólo relativamente. De manera sorprendente, Russell (que como se sabe, acometió con Whitehead en la célebre obra, *Principia Mathematica,* el colosal intento de una logicización

(estándar, clásica) de la matemática -luego problematizada por las antinomias encontradas en la teoría de conjuntos y por los teoremas de incompletud de Church y Gödel-), se ocupa de la vaguedad en un pequeño pero sustancioso trabajo. El trabajo se denomina "*Vagueness*", data del año 1923, y en él deja claro varias ideas fundamentales para lo que podría ser denominado como una filosofía de la vaguedad:

(R.I). Dado que el lenguaje natural es vago, la lógica no puede tener como argumentos palabras o expresiones que refieran *"a esta vida terrenal, sino a una existencia celestial imaginaria*" [1923: 20]. Y claro está, entre las palabras o expresiones que refieren a asuntos terrenales se encuentran aquellas que son vagas, que problematizan la validez universal de las leyes lógicas más elementales. Por tanto, si se quiere evitar que la lógica clásica tenga problemas, es necesario prescindir o reducir la vaguedad.

(R.II). El uso del lenguaje científico parece un buen método para reducir la vaguedad. Por eso Russel indica que *"hay menos vaguedad bajo el microscopio"* [1923: 23].

Con la idea de que la vaguedad existe, pero debe ser eliminada, Russell no hacía más que sumarse a otro lógico relevante de finales del siglo pasado y al cual ya hemos hecho alusión: Frege, el sistematizador de la lógica moderna. Como es conocido, según Frege, la referencia de una oración se puede reducir a uno de los dos valores de verdad clásicos, 'verdadero' y 'falso' o, en representación numérica habitual, 1 y 0. Pero 'verdadero' o 'falso' sólo pueden atribuirse a oraciones precisas, que son las que en opinión del autor deben ser objeto de análisis lógico. Participa Frege, por tanto, de un cierto desdén hacia el lenguaje ordinario y, al mismo tiempo, de un indudable aprecio por el lenguaje de la ciencia; en su opinión, el que transmite nuestros pensamientos más acabados.

Por ello, quizás, Frege no dedicó nunca un trabajo completo a la vaguedad. Pero en una de las pocas referencias que dedica al tema manifiesta con claridad su opinión acerca de los términos vagos: *"It can be objected that such words are used thousands of times in the language of life. Yes, but our vernacular languages are also not made for conducting proofs"* [Frege, 1879]. Por tanto, sólo en el lenguaje científico, dada su precisión, no son contrariados principios clásicos elementales y reglas de inferencia fundamentales. Es en ese caso cuando su lógica funciona conforme a los propósitos por él diseñados: la caracterización inequívoca de la noción de consecuencia lógica.

No obstante, aunque el trabajo de Russell coincide con las ideas fregeanas acerca de la vaguedad, dice más, porque señala otras dos notas esenciales para dirimir el problema antes anunciado de si la vaguedad es un fenómeno con entidad propia, si cabe reducirla a otros fenómenos lingüísticos o, incluso, si es posible estirparla. Russell dice en el trabajo ya reseñado que,

(R.III). *"Todo el lenguaje* (natural) *es vago"* [1923: 15],
(R.IV). La vaguedad es una característica intrínseca del lenguaje, es una nota propia del mismo.

La característica (R.III) es consecuencia de una propiedad interesante de la vaguedad, anotada en muchos trabajos clásicos sobre el tema: Ante la desinfección a menudo no respetuosa con el medio lingüístico habitual de expresión propuesta por la filosofía analítica, la vaguedad se muestra, en su nicho ecológico –el lenguaje– como una propiedad extraordinariamente infecciosa. P. ej., nadie duda de que la oración '3 es un número' es precisa. Pero si se le antepone un predicado vago; p. ej., 'mucho más grande que' , entonces la oración 'mucho más grande que el número 3', se convierte en vaga.

La característica (R.IV) alude al carácter diferenciable de la vaguedad. Desde un punto de vista lingüístico, la vaguedad no parece subsumible en la ambigüedad o en la generalidad, como bien se apunta en textos como los de [Alston, 1964] o [Kempson, 1977]. Así, p. ej., se dice que 'plato' es una palabra ambigua, ya que puede significar 'pieza de bicicleta' o 'recipiente para comer', pero no que es vaga. En cada contexto por separado, 'plato' tiene un significado más o menos preciso. Del mismo modo, de 'asignatura' se dice que es una palabra genérica, en tanto al nombrarla no se especifica cuál de ellas es, pero, como en el caso anterior, una vez especificada, su significado sería más bien preciso que vago. Ambos casos no parecen ser, por tanto, ejemplares (otra vez en el sentido de Kuhn) de vaguedad.

Pero, ¿es posible mostrar de un modo más radical la irreductibilidad de la vaguedad?. Si ello fuere así, tendríamos garantizada nuestra contestación a la primera de nuestras preocupaciones, una vez que su presencia está mas que asegurada en el lenguaje por su carácter infeccioso. Pues bien, una respuesta positiva está justificada. Además de las evidencias arrojadas desde la lingüística, o desde la filosofía (sobre todo por el concepto de textura abierta, [Waismann, 1945], [Fine, 1975]), el teorema de Parikh ofrecía, desde nuestro punto de vista, un amparo muy atractivo a esta tesis, ya que proporcionaba una prueba formal de la irreductibilidad de la vaguedad en los predicados observacionales (no en vano son los más propicios a ser enseñados por ostensión). En breve haremos una descripción

informal del teorema, pero antes deseamos exponer otra apoyatura a la no subsunción de la vaguedad. Proviene del área de la filosofía de la ciencia (más en concreto, de filosofía de la física) y es debida a K. R. Popper.

De un modo informal, Popper apoya la tesis russeliana de que 'bajo el microscopio hay menos vaguedad', pero con matizaciones muy interesantes. En su autobiografía intelectual indica lo siguiente: "*No estoy sugiriendo, por supuesto, que un incremento en la precisión de, digamos, una predicción, o incluso de una formulación, no pueda ser a veces altamente deseable. Lo que sugiero es que siempre es indeseable hacer un esfuerzo para aumentar la precisión por ella misma –especialmente la precisión lingüística– puesto que ello conduce usualmente a una pérdida de claridad, y a un derroche de tiempo y esfuerzo en preliminares con frecuencia inútiles, ya que son superados por el avance real de la materia: nunca se debería intentar ser más preciso de lo que demanda la situación del problema*". [Popper, 1974: 33].

Veamos ahora la objeción formal a la tesis de la eliminabilidad de la vaguedad; es decir, el teorema de Parikh, [Parikh, 1983]. De un modo intuitivo, el teorema indica que todo predicado que es el resultado de algún tipo de observación es intrínsecamente vago, debido a la existencia de un límite observacional que, para cualquier situación, garantiza la existencia de elementos que no están localizados ni son localizables. Una descripción más exhaustiva requiere definir previamente el concepto de espacio α-conexo.

Se dice que un espacio métrico (U, d) es α-conexo, $\alpha > 0$, cuando no hay una bipartición $U = A \cup A'$, $(U \supseteq A, A \neq \varnothing)$ tal que si $a \in A$, $b \in A'$ se verifique que $d(a,b) \geq \alpha$. Pues bien, en los espacios métricos α-conexos se verifica el teorema de Parikh. En efecto, sea (U, d) un espacio métrico α-conexo y P un predicado que refiere a U y tal que $\varnothing \neq P \neq U$, entonces existen elementos $a \in P$ y $b \in P'$ tales que $d(a, b) < \alpha$. Supongamos que el espacio (U, d) presenta como tope observacional lo comprendido entre 0 y α $(0<\beta<\alpha)$ con respecto a un predicado P que cumple el teorema anterior. De ocurrir esto, no podemos distinguir puntos que disten entre sí menos de β. Pero puede suceder que, si U tiene suficientes puntos, para P los puntos a y b estén por debajo del umbral de observación; esto es que, $d(a, b) < \beta < \alpha$, con lo que no serían distinguibles.

Una consecuencia intuitiva del teorema es entonces la que sigue. Si en una medida hay un margen de error, entonces pueden existir parejas de puntos indistinguibles que por observación jamás seremos capaces de separar. Los elementos susceptibles de caer bajo la extensión de P son siempre imperfectos,

porque siempre es posible pensar (en la práctica o en teoría) en algún instrumento (microscopio, regla, ...) que precise más la distinción. De esto se sigue, como indica el profesor Trillas, que el teorema de Parikh vincula *"tanto a sastres con una cinta métrica, como a físicos con sofisticados aparatos de medida"* [Trillas, 1984: 10]. La inevitabilidad de la vaguedad queda así afirmada en un terreno que en principio no parecía el más propicio: el de los predicados observacionales y los procedimientos métricos, tan próximos a la actividad científica.

3.2.2. La vaguedad en Black

Asentada la idea russelliana de que la vaguedad es una propiedad intrínseca del lenguaje, fue Black [Black, 1937, 1963], quien avanzó tres propuestas de gran repercusión en lo que antes denominábamos como instrumental para trabajar la materia prima; esto es, tres aspectos importantes relacionados con las lógicas del lenguaje vago y, en concreto, con una de ellas: la lógica borrosa. En efecto, Black alimenta, por primera vez, la idea de que la vaguedad se puede –dicho de una forma laxa– cuantificar o, como diría Deaño, que se puede explicar de manera no vaga lo vago. Para ello hace alusión a tres temas interesantes:

(B.I). El análisis de la vaguedad debe hacerse respecto a una comunidad de hablantes.

La aproximación a la vaguedad lo ha de ser, por tanto, respecto al uso local del lenguaje. Es Black quien resalta por vez primera la importancia del dominio restringido o universo de discurso en una posible lógica de los conceptos vagos. Creemos que por esto puede ser tomado como precursor del carácter local de la lógica borrosa, según Zadeh, una de las características definitorias de la misma.

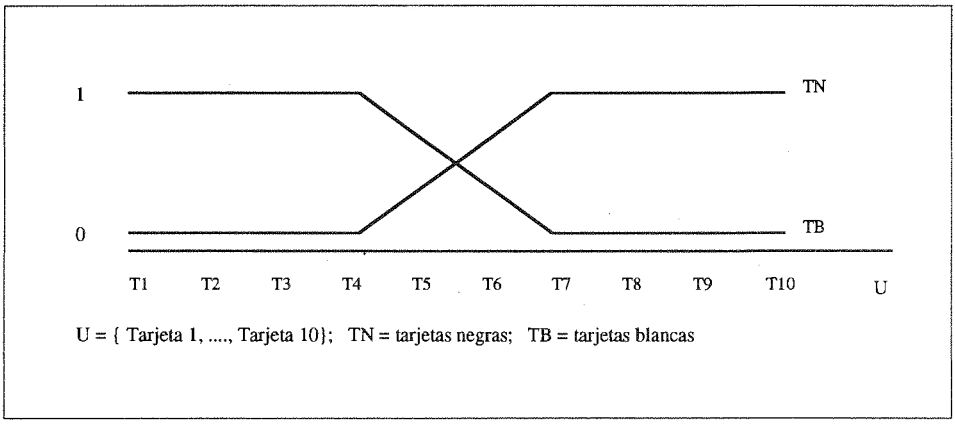
El concepto de 'comunidad de hablantes' es, del mismo modo, fundamental en la teorías formales de contextos, a las que también hemos dedicado atención [Sobrino, 1989]. Estas teorías tienen como objetivo lo que nosotros denominamos como 'la precisión periférica de la vaguedad'. El motivo de este rótulo, brevemente descrito, es que formalizar el contexto no es formalizar la propia imprecisión del predicado vago, como lo haría la lógica borrosa, sino reconvertir (o reducir) la vaguedad a precisión por homogeneización máxima de la comunidad de hablantes y de algunas de las presuposiciones subyacentes a su lenguaje (sus presupuestos (background), el foco de atención, etc.). P. ej., 'alto', en el contexto del uso lingüístico de cualquier humano, seguramente es un término muy vago, pero si restringimos la evaluación de ese predicado a la comunidad gallega, probablemente

su vaguedad disminuiría, de modo que, podemos establecer la hipótesis razonable de que, cuanto más homogénea y pequeña sea la comunidad, menos vaguedad habría en la utilización del término.

Hay ejemplos que muestran, sin embargo, que reducir completamente la vaguedad por acotación del universo de discurso no es posible, lo que confirma también la tesis de la irreductibilidad de la vaguedad. Piénsese sino en el concepto de 'número mucho más grande que el número 3' referido al universo de los mil y de los cien primeros números naturales. Obsérvese como la pertenencia de cualquier elemento a ese concepto (p. ej., 10), aún reduciendo el universo de mil a cien, sigue siendo vaga en ambos casos (aunque, naturalmente con distinta valoración).

(B.II). Dado que los predicados vagos se caracterizan por poseer una zona penumbral o franja que queda entre la extensión de un concepto vago y la de su complementario, es necesario precisar que se entiende por el concepto de 'franja'.

En opinión de Black, esta noción queda bien caracterizada si somos capaces de cuantificar la 'consistencia de aplicación' [C (x, L)], concepto que alude a una medida numérica que indica cuál es la discriminación de un objeto o serie de objetos por la comunidad lingüística que los verbaliza. La representación gráfica de C (x, L) –la medida de la consistencia de aplicación de un término de L al objeto x– arroja un perfil en forma poligonal abierta llamado 'perfil de consistencia', que señala, para el ejemplo escogido, la aplicación de términos de color (p. ej. 'blanco' y 'negro' para una serie de diez tarjetas donde el color se va difuminando progresivamente.



Como se observa, el perfil de consistencia tiene un aire de familia con la caracterización de un número borroso; quizás por esto Dubois y Prade afirman que puede ser considerado como *"the ancestors of fuzzy membership functions"* [Dubois y Prade, 1980: 4].

(B.III) Por último, Black ejemplifica también como la vaguedad puede hacer naufragar razonamientos aparentemente correctos, como ocurre en la paradoja del Sorites, bien conocida en nuestro ámbito de estudio. El examen de esta paradoja es un tema muy interesante por abordarse en la misma temas centrales de lógica como cuantificación [Schwartz, 1989] o Modus Ponens [Rolf, 1981].

A pesar de lo innovador de sus propuestas, Black mantiene no obstante un nexo de unión con Frege y Russel: aún reconociendo que la vaguedad exige un marco y un instrumental teórico diferente al de la lógica clásica, como ha quedado claro en (B.I), (B.II) y (B.III), no obstante, indica que deben seguir manteniéndose las leyes lógicas clásicas. Esta elevación de la lógica –como diría Russell– al reino de los cielos, quizás le impidió a Black recorrer más distancia en un camino que tan prometedoramente había empezado (para una posible lógica de la vaguedad) con el concepto de 'consistencia de aplicación'.

Pero en torno a los años 60, con el segundo Wittgenstein y la escuela oxoniense, se produce un vuelco en la filosofía del lenguaje. No sólo hay un interés por la estructura del lenguaje, sino, sobre todo, por sus usos concretos. Surge entonces, poco a poco, una preocupación por el lenguaje ordinario que se ve reforzado cuando la I.A. muestra la necesidad de su gestión, ya que es el lenguaje común el medio habitual para expresar nuestros conocimientos, precisamente aquellos que pretendemos simular en un Sistema Basado en el Conocimiento o Sistema Experto. Por tanto, el lenguaje debe estudiarse tal como aparece, en su inmediatez. Esta es la idea que subyace a los primeros trabajos de Zadeh, que ve las deficiencias del lenguaje preciso en el área del control, lo que le lleva a enunciar su conocido Principio de Incompatibilidad: tanto en cuanto la complejidad de un sistema se incrementa, precisión e información son propiedades que, a partir de un límite, resultan mutuamente exclusivas [Zadeh, 1984].

El acercamiento al lenguaje ordinario hace que se desarrollen las lógicas del lenguaje ordinario; entre otras, la lógica borrosa. Esta debería, dado que es una lógica orientada computacionalmente donde los aspectos semánticos preponderan sobre los sintácticos y donde las cuestiones metalógicas son periféricas más bien que centrales [Bellman y Zadeh, 1977], responder a las formas usuales de cognición y uso de los términos vagos de nuestro lenguaje. Y aquí entra en escena un área de

estudio que ha sido central en nuestra investigación: la teoría de prototipos, una rama de la psicología cognitiva que tiene como objeto de estudio el escudriñar la cognición subyacente a la categorización o estructuración del medio conceptual que nos ayuda a reconocer nuestro entorno.

4. LA TEORÍA DE PROTOTIPOS

Una hipótesis central de la teoría de prototipos es asumir que:

1) Cognitivamente hablando, hay objetos que pertenecen a una misma clase pero no son igualmente representativos de la misma. P. ej., citando la célebre experiencia de Rosch [Rosch, 1973], nadie dudaría en decir que un murciélago es más un pájaro que cualquier otra cosa. Por tanto, pertenece, por decirlo en términos conjuntistas, a la clase de los pájaros. Sin embargo, el murciélago no es tan representativo de la clase pájaro como otros pájaros (tómese como ejemplo el gorrión). Es decir (también en términos conjuntísticos), a ambos animales no se le atribuye normalmente la misma pertenencia a la clase pájaro. Por tanto, y en resumen, aún habiendo elementos que (por ahora sin más disquisiciones) pertenecen a una clase o son miembros de una categoría conceptual (terminología más acorde con la psicología cognitiva), no tienen la misma representatividad para ella. Esto nos conduce al segundo supuesto de la teoría de prototipos.

2) Las clasificaciones que establecemos los humanos son más bien graduadas que dicotómicas. Es decir, aludiendo otra vez al trabajo de Rosch, presentada una colección de pájaros que presentan diferencias entre sí (granívoro/no granívoro, vuela/no vuela,...) a un grupo de sujetos a los que se les pide que los clasifiquen según su pajaridad, estos los agruparían a todos como pájaros, pero le atribuirían distinto grado de pertenencia al conjunto de los pájaros.

El concepto de 'grado de pertenencia' posibilita algo interesante: necesario en la psicología cognitiva para explicar la gradación conceptual y consustancial a las lógicas polivalentes, como la lógica borrosa, que atribuyen grados de verdad, vincula a ambas a través de la idea de que una lógica para la I.A. debe responder a las cogniciones humanas más comunes: no en vano son éstas las que se intentan simular y, no en vano, tanto la psicología como las lógicas para la I.A. forman parte de ese ámbito general de estudio que se ha dado en llamar 'Ciencias Cognitivas'.

La gradación valorativa que adopta la lógica borrosa (como un tipo de extensión de las lógicas polivalentes) responde a la convicción expresada por Zadeh de que "*more often than not, the classes of objects encountered in the real physical world do not have precisely defined criteria of membership*". [Zadeh, 1965: 29]. Por tanto, los instrumentos que proporciona la lógica borrosa para modelizar los conceptos del lenguaje ordinario, deberían ser adecuados a la cognición a la que representan esos conceptos. Pues bien, después de los trabajos de Rosch, una pareja de psicólogos norteamericanos, [Osheron y Smith, 1981], se han mostrado muy críticos y combativos con la idea de que el instrumental que la lógica borrosa proporciona para conectar conceptos (esto es, que las definiciones que la lógica borrosa da para las conectivas), sea adecuado y refleje la forma usual con la que los humanos comprendemos lo que, en el ámbito de la psicología cognitiva, se denomina combinación conceptual.

Algunos de los ejemplos que tratan de ilustrar su crítica es dudoso que representen, siquiera, cogniciones humanas normales. El ejemplo de la manzana rayada como la conjunción del concepto 'manzana' y del concepto 'rayada' es paradigmático (comprendemos que es una manzana, pero no que es un rayado, independientemente de su soporte material o físico). El ejemplo más logrado es, sin duda, el que dedican a la disyunción conceptual, y del cual proponemos a continuación la siguiente versión.

Sea R un concepto (riqueza), que es el resultado de la disyunción (interpretada como reunión) de otros dos: A (ahorro) e I (capacidad de inversión) y sean a, b y c tres individuos. Estos individuos tienen la capacidad de inversión y el ahorro que se indica en la siguiente tabla:

Riqueza				
Ind.	Ahorro	$\mu_A \cup \mu_I$		Inversión
a	110.000 $	0.8	0.05	500 $
b	100.000 $	0.7	0.7	100.000 $
c	500 $	0.05	0.8	110.000 $

Según Osheron y Smith, el problema que se plantea es el siguiente: Dado que la lógica borrosa queda en su opinión reducida (es de suponer que por su desconocimiento de la misma) a la lógica max-min de Zadeh, la disyunción conceptual debería quedar convenientemente reflejada a través de la conectiva max; esto es, para a, b y c, [v (A$\vee$ I) = max (v (A), v (I)]. Pero aplicando esta definición para la disyunción se da la siguiente secuencia:

$$\mu_R(a) = \mu_R(c) > \mu_R(b)$$

resultado claramente antiintuitivo que contraría cualquier cognición normal del problema. De este resultado, (y de otros para las restantes conectivas, basados en ejemplos más peregrinos), Osheron y Smith concluyen que la lógica borrosa no es una lógica adecuada al análisis de los conceptos ordinarios.

Un examen más detallado del problema puede hacer ver cómo se transforma la expresión de arriba en un ordenamiento más adecuado a nuestra intuición ordinaria. P. ej., usando como definición de la conectiva '$\vee$' a la suma acotada [v (A$\vee$I) = min (1,v(A)+v(I)] o la suma probabilística, [v (A$\vee$ I) = (v (A) + v (I)) - (v (A) $\times$ v (I))], resulta la siguiente secuencia,

$$\mu_R(b) > \mu_R(a) = \mu_R(c)$$

acorde con lo que cualquier evaluación racional del problema daría como resultado. La clave fundamental para deshacer la antiintuitividad ha sido considerar que el concepto de 'riqueza' vincula al de 'ahorro' y 'capacidad inversora', mientras que la definición del valor de la disyunción como el máximo de las valoraciones no es interactiva. Por lo tanto, no podía modelar apropiadamente el significado de 'riqueza'.

No obstante, permanece aún una dificultad, ya que los valores obtenidos con las nuevas definiciones de las conectivas no expresan todavía suficientemente bien una evaluación adecuada del problema. Es decir, intuitivamente parece casi el doble de rico el individuo 'b' que el individuo 'a' o el individuo 'c'. Sin embargo, las valoraciones obtenidas, aunque más acordes (en lo que respecta a la serie que generan) con el sentido común, no indican con rotundidad esto. Y ello plantea a menudo un conflicto serio en esta lógica.

En efecto, aunque tanto los ejemplos de Osheron y Smith, como otros posteriores de Jones, 1982, Mervis y Roth, 1983 y Smith, 1985 (casos que hemos analizado en [Adán y Sobrino, 1991]) pueden ser falsables, lo que permite extender la frase de Zadeh de que *"las dificultades encontradas por Osheron y Smith no*

descalifican a la teoría de conjuntos borrosos como fundamento para una teoría de prototipos" [Zadeh, 1982: 297], no obstante, todavía es necesario trabajar en una mayor adecuación de la lógica borrosa a la teoría de prototipos. En esta tarea suele plantarse el conflicto del que hablábamos, ya que cuando se desea tener una lógica donde la capacidad calculística sea adecuada a la capacidad expresiva del lenguaje, podemos escoger dos posibilidades:

(i). Indicar qué propiedades querríamos que verificase esa lógica y, en función de ellas, p. ej., definir sus conectivas.

(ii). Pero el lenguaje vago presenta continuas dificultades, ya que cada frase es en cierto modo única, por lo que parecería deseable tener en esa lógica tantas conectivas, como sentidos de vaguedad distintos hubiese. Sin llegar a este extremo, la definición de t-normas, t-conormas y funciones de negación [Alsina y col., 1983] ha representado un avance muy importante para conciliar (i) e (ii), ya que, verificando propiedades estructurales importantes para cualquier lógica (y, en particular, para la borrosa) recogen bastantes formas particulares de combinación conceptual.

[Propuesta 1]. Sobre la base de lo ya hecho, pensamos que cualquier investigación en este aspecto sería interesante, y no sería inútil hacer un catálogo de usos y comprensión de expresiones vagas complejas que hacen los hablantes y estudiar la adecuación de las definiciones de conectivas borrosas a las mismas.

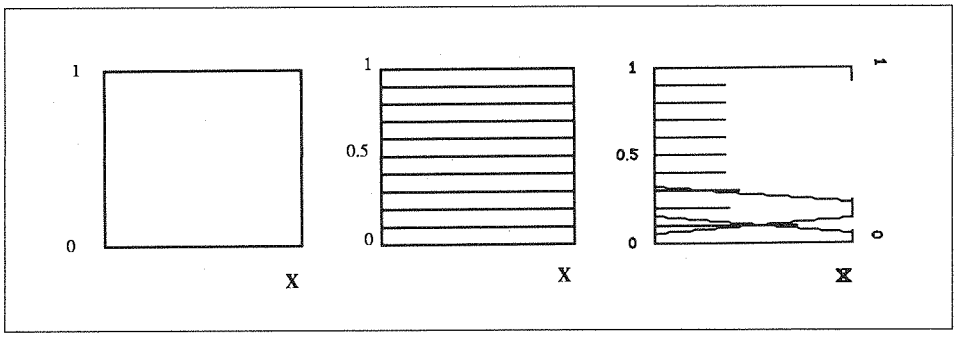
[Propuesta 2]. Consideramos también interesante investigar la noción de prototipo como un tema común a una clase de lógica borrosa, la lógica disposicional de Zadeh y un tipo de lógica no-monótona: el razonamiento por defecto, en tanto ambas tratan de modelizar expresiones del tipo 'Usualmente un pájaro vuela' o 'Normalmente (esto es, en ausencia de información en contrario), un pájaro vuela'. Obsérvese la similitud de estas expresiones con otras como: 'Típicamente, un pájaro vuela' o 'Si un pájaro es típico, vuela'. Pero mientras en la primera frase 'típicamente' puede ser visto como un cuantificador sentencial, en la segunda 'típico' es un modificador de 'pájaro'. Estudiar el distinto papel de la cuantificación en las oraciones que modelan una y otra lógica parece un tema de interés.

5. PROTOTIPOS Y APROXIMACIÓN LINGÜÍSTICA

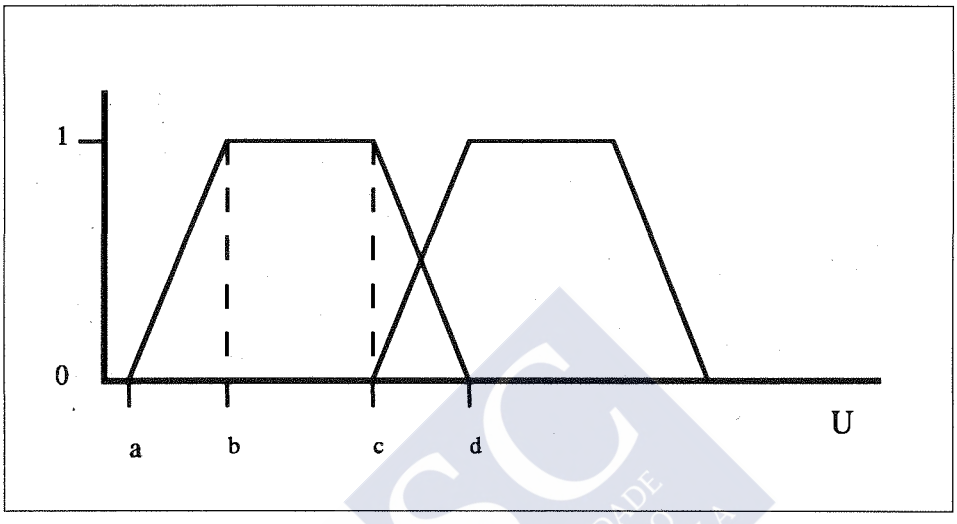
La noción de prototipo ha constituido un punto de inflexión todavía más acentuado en nuestra investigación. Aunque muy operativa en lo que hemos

denominado como 'teoría de prototipos ingenua', que responde básicamente a los estudios de Rosch antes reseñados, esa concepción de lo que es un prototipo no nos parecía plenamente satisfactoria, no era tan intuitiva como aparentaba. Por un lado, ya sabíamos que pertenencia no era lo mismo que representatividad: es dudoso que para un hombre de nuestro medio rural con escasa formación, 2 sea tan representativo del concepto 'número' como 2^{-1}. Sin embargo, ambos serían admitidos por él como perteneciendo a la clase de los números. Por otro, intuíamos que un prototipo, si responde a las cogniciones humanas, no tiene porqué coincidir con el elemento máximamente representativo de la clase. En efecto, nos parecía dudoso que si un terráqueo cualquiera le diese una descripción de sus congéneres a un extraterrestre, éste identificase mejor o más rápido como 'humano' a un/a 'modelo' (prototipo) humano, que a un individuo con características normales. Esto nos llevó a estudiar el concepto de prototipo en Zadeh, que inmediatamente se nos reveló como una noción que podía ser fructífera en su aplicación a diversos temas, incluido uno fundamental de razonamiento aproximado: el problema de la no clausura del conjunto de términos o term-set valorativo para operaciones con valores de verdad lingüísticos.

Antes de entrar en el problema debemos, no obstante, atender a la descripción de algunas ideas preliminares que introduzcan al lector en el tema. Como ya indicamos, cuando la lógica borrosa se convierte en una extensión compleja de las lógicas polivalentes, sus valores de verdad son lingüísticos. Así como una función de valoración clásica ('verdadero' o 'falso', si hablamos de funciones de verdad; 'encendido', 'apagado', si hablamos de funciones de conmutación) se resume en un número, 1 o 0; un valor de verdad lingüístico queda bien caracterizado por un número borroso. Figurativamente, mostramos en los siguientes cuadros lo que sería un valor (número) clásico, un valor (número) polivalente y un valor (número) borroso.



Los trapecios de la figura de la derecha son mostrados con más claridad a continuación:



donde el intervalo [a, b] indica la ascendencia en la pertenencia, el intervalo [b, c] la pertenencia absoluta y el intervalo [c, d] la decadencia en la pertenencia.

Veamos ahora, brevemente, cómo se genera un conjunto de valores lingüísticos o term-set. Ante todo debemos fijarnos en el problema a representar y elegir, en consonancia, la variable apropiada. Imaginemos que deseamos valorar; la variable escogida entonces debe ser 'verdad'. Elegimos, a continuación, la especificación de la variable más habitual: en el caso de 'verdad' será, sin duda, 'verdadero'. Acto seguido se fija su antónimo, 'falso', y , el resto de especificaciones se consiguen añadiéndole al término primitivo o base y al antónimo las restricciones que proporcionan las cercas semánticas, palabras como 'muy', 'ligeramente' o 'bastante' que constriñen el significado de la palabra a la que van asociadas. Un conjunto de términos valorativo usual (ya ordenado) es el siguiente: muy verdadero, verdadero, bastante verdadero, ligeramente verdadero, ni verdadero ni falso, ligeramente falso, bastante falso, falso y muy falso que, como se observa, consta de 7 elementos. Cada uno de esos elementos, caracterizado por un subconjunto borroso, debe representar el significado de cualquier variable que contenga las mismas especificaciones, p. ej., muy alto, alto, bastante alto,..., bastante bajo, bajo, muy bajo, dado que, significado (alto) $=_{def.}$ verdadero (alto) = $\mu_{alto}(x)$, $x \in U$.

Después de esta introducción a algunos conceptos básicos, pasamos a describir brevemente la teoría de prototipos de Zadeh. Sea C una categoría o un concepto (o lo que es más familiar, un predicado) vago y $e_1,\ldots, e_n$ una serie de elementos suficientemente similares entre sí y susceptibles todos ellos de caer bajo la extensión de C. En lógica borrosa la modelización formal de lo que hasta aquí hemos expresado en letra se consigue atribuyendo a cada uno de los e_i un valor en el intervalo [0, 1]; esto es, a través de un subconjunto borroso. En función del valor obtenido por los e_i, éstos pueden ser clasificados como buenos, regulares o malos representantes de C. Obtenemos así tres subconjuntos borrosos que son sometidos, de nuevo, a otro proceso valorativo: los representantes buenos respecto a su alta representatividad del concepto, los regulares respecto a su mediana representatividad y los malos respecto a su pobre representatividad. De esta forma habremos conseguido estructurar el concepto C en tres secciones, de manera que, su valoración sería la reunión de los tres subconjuntos parciales logrados por supervaloración. Esa reunión representaría al valor de C, pero no del prototipo de C (PT (C)). Veamos cómo lograrlo.

Descrito de una manera resumida, podría quedar caracterizado así: La obtención de la valoración del valor de PT (C) es el resultado de someter a los tres subconjuntos parciales que resumían el valor de C de nuevo a un proceso de supervaloración: los ejemplares buenos son valorados respecto a en qué medida son prototipos de los buenos ejemplares de C. Y, en correspondencia, indicamos en qué medida ese 'grado de prototipicidad' es alto. A continuación, el proceso se repite con los representantes regulares y malos. La suma, entendida como reunión de todos ellos, constituye el valor de PT (C). Como es fácil observar, ese valor no tiene porqué coincidir con un número de [0, 1]; normalmente –y esto ocurrirá siempre que la modelización demande casos interesantes–, la valoración será compleja y tendrá similitud con un número borroso. Lo descrito en letra queda ilustrado por las siguientes fórmulas, [Zadeh, 1982]:

$$C = \text{Alto}/C_{\text{bueno}} + \text{Regular}/C_{\text{fronterizo}} + \text{Bajo}/C_{\text{pobre}}.$$

$$PT\,(C) = \text{Alto}/PT(C_{\text{bueno}}) + \text{Regular}/PT\,(C_{\text{fronterizo}}) + \text{Bajo}/PT\,(C_{\text{pobre}}).$$

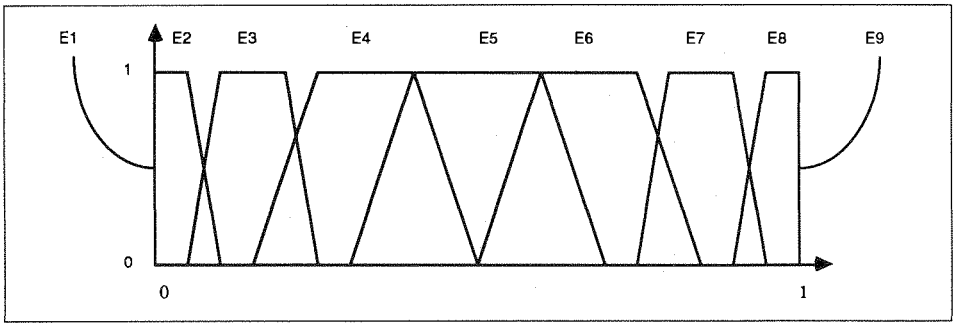
La teoría de prototipos de Zadeh, aún siendo menos intuitiva que la teoría de prototipos ingenua, representa de una manera mas precisa las intuiciones ordinarias de los hablantes.

[Propuesta 3]. Su uso todavía no ha sido suficientemente explotado, incluso dentro de la propia área de creación de teoría borrosa. Así, p. ej., la interesante noción de sinonimia, un tipo de relaciones de similaridad de Zadeh desarrollada por Trillas, Riera y López de Mántaras usa la noción ingenua de prototipo, no la borrosa. Un estudio de cómo esta noción puede ayudar a profundizar lo que es una relación de sinonimia (y, por ende, en las propias relaciones de similaridad), creemos que es un trabajo que también ofrece interés y que debería hacerse.

Como decíamos al principio, conscientes de su utilidad, nosotros hemos intentado aplicar la noción de prototipo de Zadeh al tema de la aproximación lingüística. En lo que sigue, describiremos brevemente esta aportación.

La aproximación lingüística es un tema fundamental en el razonamiento aproximado. Como es fácil comprobar, –y en esto se ha fijado la lógica borrosa– los humanos sabemos hacer inferencias utilizando valores de verdad lingüísticos. Así, si sabemos que 'si corro mucho me canso (es muy verdadero)' y que 'corro mucho (es verdadero)', inferiré que 'me canso (es muy verdadero)'. Pero un ordenador no manipula directamente el lenguaje, como nuestro cerebro, sino números que están en vez de la semántica de ese lenguaje. En la representación y gestión de inferencias con proposiciones vagas el problema de la aproximación lingüística está relacionado estrechamente con el tema de cómo mejorar los interfaces persona-máquina. Someramente descrito, el problema es el siguiente: Deseamos representar y gestionar algún tipo de conocimiento humano vago, para lo cual hemos generado un conjunto de términos valorativos. Entonces, si el interfaz es amistoso, la máquina debe permitir como input frases del lenguaje humano (normalmente dos valoraciones lingüísticas, correspondientes a las dos premisas de un modus ponens, la regla más empleada para expresar conocimientos). Estas valoraciones son traducidas a números borrosos y la máquina opera con estos números hasta que encuentra la solución al problema, que es otro número. Pero puede ocurrir que ese número no se corresponda exactamente con ninguno de los números que están en vez del conjunto de valores lingüísticos inicialmente dados para representar el problema. Esta es la dificultad que se señala en la aproximación lingüística. A continuación presentamos una posible aplicación de la teoría de prototipos de Zadeh para solucionar esta dificultad.

Sea el siguiente conjunto de términos, que pretende reflejar una descripción estándar de lo que se entiende por 'altura' (referida a personas):



Valores lingüísticos	(n_1, n_2, n_3, n_4)
E1: Bajísimo	(0, 0, 0, 0)
E2: Muy bajo	(0, 0, 0.05, 0.10)
E3: Bastante bajo	(0.05, 0.1, 0.2, 0.25)
E4: Ligeramento bajo	(0.15, 0.25, 0.40, 0.50)
E5: Normal	(0.30, 0.40, 0.60, 0.70)
E6: Ligeramento alto	(0.5, 0.6, 0.75, 0.85)
E7: Bastante alto	(0.75, 0.8, 0.9, 0.95)
E8: Muy alto	(0.9, 0.95, 1, 1)
E9: Altísimo	(1, 1, 1, 1)

Como se observa, el term-set consta de 9 etiquetas lingüísticas definidas por sus correspondientes números borrosos. Dado que la definición de prototipo de Zadeh no incluye un criterio operacional explícito para identificar de una forma precisa al conjunto PT (C), empezamos a desplegar el desarrollo embrionario de ese concepto. Agrupemos las nueve etiquetas en tres bloques (altos, mediocres y bajos) correspondientes a cómo de prototípicos son los buenos, regulares y malos elementos de C, pudiendo definir los valores que le corresponden de la siguiente forma:

$$Bajo\,(PT(C_{pobre})) = (\min(n_{1E1}, n_{1E2}, n_{1E3}), (\min(n_{2E1}, n_{2E2}, n_{2E3}), \max(n_{3E1}, n_{3E2}, n_{3E3}), \max(n_{4E1}, n_{4E2}, n_{4E3}).$$

$$Mediocre\,(PT(C_{fronterizo})) = (\min(n_{1E4}, n_{1E5}, n_{1E6}), (\min(n_{2E4}, n_{2E5}, n_{2E6}), \max(n_{3E4}, n_{3E5}, n_{3E6}), \max(n_{4E4}, n_{4E5}, n_{4E6}).$$

$Alto\,(PT(C_{bueno})) = (\min(n_{1E7}, n_{1E8}, n_{1E9}), (\min(n_{2E7}, n_{2E8}, n_{2E9}), \max(n_{3E7}, n_{3E8}, n_{3E9}), \max(n_{4E7}, n_{4E8}, n_{4E9})$.

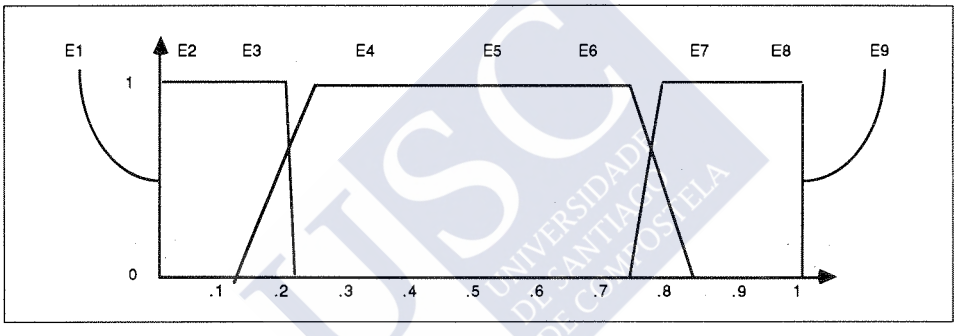
que para el conjunto de términos propuesto en la figura, tomaría las siguientes valoraciones:

$Bajo\,(PT(C_{pobre})) = (0, 0, .2, .25)$

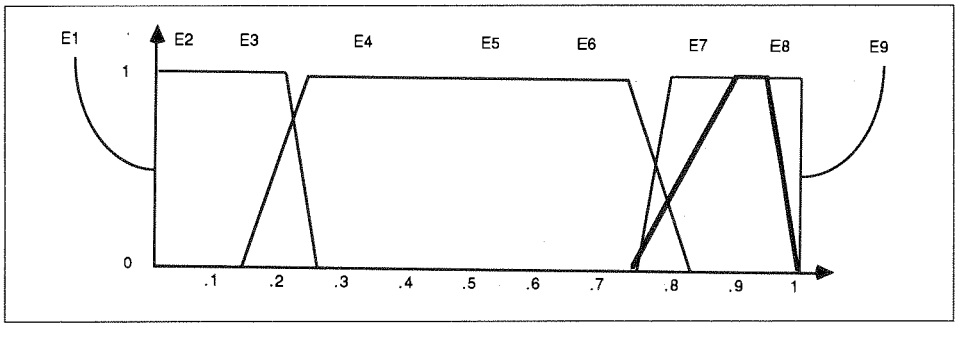
$Mediocre\,(PT(C_{fronterizo})) = (.15, .25, .75, .85)$

$Alto\,(PT(C_{bueno})) = (.75, .80, 1, 1)$

que alcanzan representatividad en el siguiente gráfico:



Imaginemos ahora que el resultado de una inferencia sea un número borroso con los siguientes valores (.75, .90, .95, 1) que, como observamos, no coincide exactamente con ninguno de los números borrosos asignados a las valoraciones lingüísticas del conjunto de términos inicial.



Como se comprueba, existe un problema importante para etiquetarlo, para darle un rótulo lingüístico, ya que está entre dos valores del conjunto valorativo. Una de las soluciones planteadas para superar esta anomalía pasa por ampliar ese conjunto utilizando etiquetas intermedias, [Godó, 1990]. En este caso, y dado que el valor conflictivo está entre E_7 y E_8, su valor de verdad lingüístico podría ser el correspondiente a la etiqueta E_{78}. Una dificultad de esta solución es que los resultados obtenidos pueden tener una traducción al lenguaje común engorrosa y poco natural.

Nosotros pensamos que la herramienta conceptual que nos proporciona la teoría de prototipos borrosa nos permite un etiquetado de los casos conflictivos bastante natural. Nuestra aportación [Olivas y Sobrino, 1991] se concreta en los siguientes pasos: 1. Constatar con qué prototipo tiene más afinidad el valor problemático. 2. Hacer una supervaloración de primer nivel respecto al prototipo más afín. 3. Hacer una supervaloración de segundo nivel que debe indicar en qué grado el caso problemático es un caso prototípico del prototipo marcado en (2) como afín. Para este caso limitamos las supervaloraciones posibles (de acuerdo a lo indicado por Zadeh) a tres casos: bueno, regular y malo.

Para el ejemplo que hemos propuesto aquí, la solución, según los pasos descritos, sería la siguiente: 1'. El prototipo más afín es Alto (PT (C_{bueno}). 2'. La primera supervaloración resultaría de observar que el caso del ejemplo constituye un 'ejemplar alto'. 3'. La sobresupervaloración (supervaloración de segundo orden) indicaría que es 'regular' dentro de los altos. El resultado definitivo sería entonces que la etiqueta que le corresponde al número borroso de arriba es 'regular como prototipo de alto'.

Como se observa, en este caso el lenguaje se muestra dócil y creemos que esta forma de valorar está dentro de la capacidades cognitivas habituales del ser humano. No obstante, como indicamos, permanecen dificultades, como las que tienen lugar cuando un número borroso está entre dos prototipos. *[Propuesta 4].* Indagar más en estas dificultades ofrece interés desde una doble perspectiva: (i). Tomando a la aproximación lingüística como el resultado de aproximar a un valor de verdad lingüístico el resultado de una inferencia borrosa. Este es un tema en el que soluciones ad hoc son previsibles y tienen su justificación en la resolución de problemas concretos para el diseño de los sistemas expertos. (ii). Cabe pensar en una orientación alternativa, formalmente más ortodoxa que la anterior: definir metalógicamente propiedades que permitan delimitar conjuntos de valores lingüísticos clausurados para las operaciones de uso habitual en razonamiento aproximado.

6. LA LÓGICA BORROSA COMO LÓGICA LINGÜÍSTICA

Dado nuestro interés por la lógica borrosa como una extensión compleja de la lógica polivalente; esto es, como una lógica de valores de verdad lingüísticos, otro objetivo fundamental de nuestra investigación fue estudiar si cabría una aproximación a la misma que permitiese exponerla, no *more matematico,* ni siquiera enfatizar su utilidad para la modelización del lenguaje común, sino *more lingua*, como una lógica lingüística, describiendo una especie de gramática formal (borrosa). Esta aproximación era posible y en ella la presentación de la lógica borrosa y de sus conceptos alcanzan un sabor inequívocamente lingüístico. Así, se definen categorías como:

6.1. Espacio nuclear o foco de interés

Como ya habíamos indicado respecto a la filosofía de Black, en lógica borrosa tiene una importancia fundamental la localidad, pues el significado de un predicado vago varía mucho según su contexto local. Pues bien, la noción de espacio nuclear hace referencia a esta característica. De manera informal, un espacio nuclear queda caracterizado por la clase de objetos más pequeña en la que nos fijamos cuando pronunciamos una oración vaga. Esas clases pueden ser concretas ('Santiago', 'estaciones temporales', etc.) o abstractas (números entre 0 y 1, triángulos equiláteros, etc.). Si L=lenguaje y N_L = núcleo (en un lenguaje), f_L es una función generadora de núcleos, que nos permite agrandar o ceñir el foco de atención tanto como queramos cuando pronunciamos una oración.

$$f_L^{\,n}\,(N) = f_L\,(f_L\,(f_L\,(\ldots(f_L\,(N))\ldots))), \quad \text{con } n = 1, 2, \ldots \quad \text{y} \quad f_L\,(N) = f_L^{\,1}\,(N).$$

Si a N la dotamos de un contenido semántico, como parece preceptivo hacer, surge entonces el concepto de

6.2. Espacio semántico

Sea N un espacio nuclear. El espacio semántico E_N generado por N se puede identificar con aquel conjunto tal que,

(i). $E_N \supseteq N^n$, $\qquad n = 1, 2,\ldots$

(ii). $E_N \supseteq f_L^{\,n}\,(N)^m$, $\qquad m, n = 1, 2, \ldots$

(iii). $E_N \supseteq A_1 \times \ldots \times A_p, A_j \begin{cases} N^n \\ f_L^{\,n}(N)^m \end{cases}$

Por tanto, E_N es el conjunto más pequeño que recubre a N^n. Además, posee la importante propiedad de estar clausurado para el producto cartesiano de los A_i generados por el operador *f*. Así pues, el espacio semántico permite escoger, entre los posibles núcleos, a aquél que es suficientemente grande, pero no demasiado, para delimitar el significado vago de la oración proferida. (i) y (ii) permiten definir otro concepto importante: el de contexto o universo de discurso.

6.3. Contexto

Es un subconjunto U de E_N ($E_N \supseteq U$), generado por N. Esta noción por sí sola resume la localidad de la lógica borrosa.

A continuación, una vez que tenemos el espacio significativo delimitado, exponemos las unidades estructurales que se consideran significativas en lógica borrosa.

6.4. Sintagma

En el contexto de la lógica borrosa se puede definir un sintagma como una palabra sencilla (un predicado vago, usualmente un adjetivo calificativo) o un conjunto de dos palabras (un nombre y un predicado vago) agrupadas. Ejemplos posibles de este concepto son los siguientes: 'carretera ancha' , 'subnormal profundo', 'número redondo', etc.

La caracterización formal de un sintagma (vago) es como sigue. Sea t un conjunto de sintagmas de un lenguaje (natural) L. Sean elementos de t , *A, B, C.* Convencionalmente asumimos en lógica borrosa que, para todo $A \in$ t y para todo x $\in$ U hay una evaluación que expresa en que grado x es compatible con el sintagma *A*.

$$\tau \times U \supseteq S.$$

Por tanto S $\langle A, x \rangle$ se puede leer como el grado en el que x es coincidente con el significado del sintagma *A*.

6.5. Significado de un sintagma

Sea A_0 un sintagma cualquiera. El significado S de A_0, $S(A_0)$ queda caracterizado por el subconjunto borroso

$$S(A_0) = \cup\, S\,\langle A_0, x\rangle / x, \text{ para todo } x \in U$$

Es decir, el significado de A_0 viene dado por la compatibilidad de ese sintagma con los distintos elementos del subconjunto del universo elegido.

6.6. La descripción de una entidad del contexto

Sea $x_0 \in U$. Su descripción $D(x_0)$ queda caracterizada por

$$D(x_0) = \cup\, S\,\langle A, x_0\rangle / x, \text{ para todo } A \in \tau$$

La descripción de una entidad del contexto o universo resulta entonces de expresar la compatibilidad del elemento con los distintos sintagmas. Tanto en este caso como en el anterior, se ha seguido la convención usual de que la compatibilidad sea expresada por un valor en el intervalo [0, 1].

6.7. Sinonimia de sintagmas

Dos sintagmas *A* y *A'* son sinónimos si, como definen los diccionarios de la lengua, 'su significado es el mismo o casi el mismo'. Nosotros diríamos que la noción de sinonimia se ajusta mejor, tal como es usada por los lingüistas, a la segunda parte de la disyunción de esta definición: Esto es, dos sintagmas son sinónimos si su significado es aproximadamente el mismo. Si fuese exactamente el mismo, entonces estaríamos ante el caso trivial de que S (*A*) = S (*A'*). Si fuese aproximadamente el mismo, entonces el grado de compatibilidad del significado de los términos bajo la relación de sinonimia queda caracterizado por un tipo especial de las relaciones de similaridad de Zadeh, las relaciones de sinonimia de [Riera y Trillas, 1981] y [López de Mántaras y Trillas, 1982], usando las cuales es posible calcular en qué grado es comparable S (*A*) y S (*A'*). [Aguilar-Martín y López de Mántaras, 1982] han usado con éxito un algoritmo de sinonimia en clasificaciones borrosas.

6.8. Antonimia de sintagmas

Dos sintagmas *A* y *A'* son antónimos, si el significado de uno es el opuesto o contrario del significado del otro. De manera que, si el significado de *A* queda –como indicamos anteriormente– caracterizado por

$$S(A_0) = \cup S \langle A_0, x \rangle / x, \text{ para todo } x \in U$$

entonces el significado del antónimo de A es $S(\text{Ant}(A_0)) = \cup S \langle \text{Ant}(A_0), x \rangle / x$, $= \cup S \langle (A_0), u_s\text{-}x \rangle / x$ para todo $x \in U$, siendo u_s el elemento superior del Universo elegido.

P. ej., sea U = N = {0,...,100} y A_1 = {número pequeño}

$S(A_1)$ = S {número pequeño, x} = {0/1, ..., 10/0.90, ..., 50/0.50, ..., 90/0.10,..., 100/0} S (Ant (A_0)) = S (número grande, x) = S (número pequeño, 100-x)= ={100/1,...,90/0.90,...,50/0.50,..., 10/0.10, ..., 0/0}.

Estas definiciones deben hacer ver cómo es posible poner en un formato 'lingüístico' conceptos importantes de la lógica borrosa. La presentación parece adecuada para enfatizar su carácter de lógica lingüística o, mejor, de lógica que opera con valores de verdad lingüísticos. Con los conceptos que hemos definido y una vez obtenidas, en este marco, formas expresivas para los modificadores o cercas, que acompañan frecuentemente a los sintagmas en las oraciones del lenguaje, así como las t-normas, t-conormas y funciones de negación adecuadas para recoger el significado de sintagmas vagos complejos (esto es, conjuntados, disyuntados o negados), estamos en condiciones de generar una gramática formal (una gramática de Chomsky) donde las reglas semánticas asociadas a la estructura sintáctica aislada según los elementos descritos arriba, permiten una presentación restringida de la lógica borrosa como una gramática lógica, donde los significados se transmiten con naturalidad y adecuación al lenguaje común y donde, y esto es también interesante, se deja ver su aplicabilidad computacional a la gestión de inferencias con proposiciones vagas. Para más detalles al respecto, véase [Novak, 1989] y [Sobrino, 1992].

7. OTROS TEMAS

Lo relatado hasta aquí, aunque ha constituido el grueso de nuestra investigación en lógica borrosa y razonamiento aproximado, no la ha agotado totalmente.

Otros temas que han suscitado nuestro interés, y sobre los que también hemos reflexionado son los siguientes:

7.1. Como ya hemos indicado, debido a las distintas clases de vaguedad y debido a la necesidad de resolver problemas de modelización muy concretos para los sistemas expertos, han surgido muchas lógicas de la vaguedad. Buscar un marco general que las unificase ha parecido siempre un objetivo interesante y los primeros intentos al respecto han partido de Gaines y su propuesta SUL (Standard Uncertainty Logic). Recientemente, el profesor Trillas [Trillas y Sobrino, 1990] ha propuesto un procedimiento para generar lógicas, en el que, definido un espacio semántico y la relación de condicional material como un preorden y anotando cómo ambas se combinan en el Modus Ponens, es posible generar distintas lógicas de la vaguedad, de la que un ejemplo fue la lógica de la contradicción. *[Propuesta 5]*. No es necesario enfatizar que el estudio de la contribución del profesor Trillas es muy interesante para el propósito indicado arriba; esto es, para chequear si y en qué medida, este marco aporta un criterio para algún tipo de clasificación o 'genealogía' de las lógicas de la vaguedad. Por ello no estaría demás hacer un acopio de estas lógicas y contrastarlas a la luz del esquema introducido.

7.2. Otro tema que ha merecido nuestra atención es el relativo a si existen o no objetos vagos. El tema nos interesó por lo atractivo de su rótulo y porque enseguida vimos que tenía relación con otros tópicos extensamente descritos en este trabajo. Después de haber destacado la importancia que en nuestra etapa inicial de contacto con la vaguedad tuvo el teorema de Parikh; el Teorema de Evans (TE), de provocador título (¿puede haber objetos vagos?), despertó nuestra atención.

El teorema dice lo siguiente:

«*It is sometimes said that the world might itself be vague. Rather than vagueness being a deficiency in our mode of describing the world, it would then be a necessary feature of any true description of it. It is also said that amongst the statements which may not have a determinate truth value as a result of their vagueness are identity statements. Combining these two views we would arrive at the idea that the world migth contain certain objects about which it is a fact that they have fuzzy boundaries. But is this idea coherent?*». [Evans,1978] y su prueba formal transcurre así:

Sean a y b términos singulares tales que la sentencia $(a = b)$ tenga un valor de verdad indeterminado, lo que puede ser indicado a través del operador ' ∇ '. Supongamos entonces que:

$$(1)\ \nabla\ (a = b)$$

lo que caracteriza a *b* de alguna forma, ya que indica su indeterminación respecto a *a*. Esto puede ser expresado, con el operador lambda, como sigue:

(2) $\lambda x [\nabla (x = a)] b$

Pero parece siempre cierto que,

(3) $\neg \nabla (a = a)$

Por tanto, (3) también caracteriza a *a*, ya que deja clara su no indeterminación. Esto puede ser expresado –al igual que arriba– con la ayuda de λ, de la siguiente forma:

(4) $\neg \lambda x [\nabla (x = a)] a$

Por último, por la Ley de Leibniz, es posible derivar a partir de (2) y (4),

(5) $\neg (a = b)$

lo cual, en opinión de Evans, contradice directamente a (1), con lo que se puede negar ese supuesto de partida.

Como puede observarse por la propia descripción de Evans, el tema que se debate no tiene que ver tanto con objetos como con elementos simbólicos que participan en algún tipo de relación; en particular, con la relación de identidad cuando los símbolos que están a ambos lados de la misma representan a designadores rígidos; esto es, términos que designan al mismo individuo en todo mundo posible. Es esto lo que le permite, en cierto modo, hablar a Evans de objetos.

Ejemplos de designadores rígidos son los nombres propios y las descripciones definidas, (p. ej., para la relación de identidad, (Manuel Fraga=Manuel Fraga), (agua = H_2O). De ser el teorema correcto habría, pues, una parte del lenguaje que no podría ser vaga: aquella a la que aluden los designadores rígidos. Pero la vaguedad no ha estado vinculada normalmente con la relación de identidad, sino con la relación de igualdad. Puede pensarse –y ejemplos hay en la bibliografía acerca de lo borroso, como p. ej., el del 'cabinet'– en un debilitamiento de la relación de identidad. Así, se alude a que si se somete a una persona en una cabina a sesiones progresivas de mutación genética, no sabríamos en qué instante la persona al salir de la cabina sería igual a si misma o un ser diferente, existiendo, por tanto, un ámbito de indeterminación en nuestro juicio. Pero este tipo de vaguedad, si bien puede ser de interés para debatir algún problema de orden teórico-

filosófico, lo cierto es que tiene poca relevancia en problemas relacionados con actividades cotidianas. Es en esta área donde muestra más importancia que la identidad, la igualdad o el debilitamiento de ésta: la similitud o la indistinguibilidad. Estas relaciones han sido estudiadas por Zadeh y Trillas respectivamente y su desarrollo formal y aplicaciones han puesto de manifiesto su importancia e interés, entre otras, en el área de la clasificación de datos vinculados a la cognición ordinaria.

[Propuesta 6]. Un estudio de los compromisos ontológicos involucrados en el TE y en los teoremas de Zadeh y Trillas sobre similaridad e indistinguibilidad puede arrojar luz acerca de distintas formas de comprender la vaguedad e incluso la propia lógica: si como un conjunto de entes ideales o platónicos, como un conjunto de instrumentos con los que se construyen teorías formales abstractas o como un lenguaje formal que pretendemos responda a cogniciones humanas.

8. PERSPECTIVAS

Estrechamente relacionadas con alguno de los temas tratados, dos son las áreas a las que este equipo le gustaría hacer alguna contribución en el futuro. La primera está relacionada con el tema de la aproximación lingüística y constituye en investigar qué es una cuasi-tautología. La segunda se relaciona con la presentación de la lógica borrosa como una lógica lingüística y tiene que ver con los lenguajes de programación específicamente borrosos como el Fuzzy-Prolog.

8.1. Como se sabe, si bien en lógica bivalorada los valores de verdad que admiten las proposiciones son dos, V y F, las implicaciones pueden tener tres caracterizaciones posibles: tautologías, contradicciones y consistencias o, por decirlo de otra forma, expresiones que uniformemente toman el valor 1 para cualquier combinación de valores de verdad de las variables, que uniformemente toman el valor 0 o que toman alternativamente valores 1 y 0. Por tanto, parece razonable pensar que una implicación que tuviese más 1 (unos) en su tabla de verdad que otra, sería más tautológica (en un sentido distinto al concepto clásico de tautología) que ésta.

La investigación sobre cuasi-tautologías tiene una tradición arraigada en lógicas polivalentes [Rescher, 1969] donde, para casos interpretativos difíciles como ocurre con la tabla de verdad del condicional, es posible tener

valoraciones múltiples para las combinaciones de valores más conflictivas (F, V) y (F, F). Así,

p	q	$p \rightarrow q$
V	V	V
V	F	F
F	V	V, F
F	F	V, F

donde el par (V, F) no debe ser interpretado como un valor de verdad nuevo, sino como que cualquiera de las entradas, V o F, son posibles, dependiendo del contexto de evaluación. El par (V, F) puede ser sustituido por el nuevo valor I (indeterminado) que resume bien en el lenguaje ordinario esa situación de conflicto. Por tanto, la tabla resultante para el condicional no es veritativo-funcional, sino cuasi veritativo-funcional.

Pues bien, a un sistema de lógica proposicional en el cual,

(i). Sus fórmulas son las mismas que las de un sistema bivalente clásico,

(ii). Dada una asignación de valores de verdad $\vee$ a las variables [p.e.], de la siguiente forma: $v(p) = (V, F)$, el $v(q) = V$ y el $v(r) = F$], el valor de cualquier fórmula se halla de la manera que se ejemplifica a continuación:

$v(\neg p) = (V, F)$, $v(p \wedge q) = (V, F)$, $v(p \wedge r) = F$, $v(p \vee q) = V$; $v(p \vee r) = (V, F)$.

(iii). La interpretación veritativo-funcional de las fórmulas en las que intervienen las demás conectivas es totalmente idéntica a las de la lógica bivalente,

se denomina sistema de lógica cuasi vertitativo-funcional (Q). En él, un esquema tautológico como el Modus Ponens alcanza la siguiente formulación:

$$\text{Si } v(p) = V \quad \text{y} \quad v(p \rightarrow q) \text{ es V; entonces } v(q) = V.$$

que responde al uso real del Modus Ponens en los motores de inferencia de los sistemas expertos, ya que se supone que ningún perito va a expresar sus conocimientos partiendo de premisas falsas, aunque ello garantizase la corrección de la inferencia.

Obsérvese como, en el sistema Q no son tautologías las paradojas de la implicación material como $[p \rightarrow (q \rightarrow p)]$ y $[p \rightarrow (\neg p \rightarrow q)]$. A cambio, se pierden muchas tautologías clásicas como $p \rightarrow p$, $p \wedge q \rightarrow q$, etc.

Pues bien, la ausencia de tautologicidad en esquemas como $A \rightarrow A$, (con apariencia tautológica) puede ser suplida a través del concepto de cuasi-tautologicidad; esto es, de fórmulas que pueden tomar valores de verdad designados para toda asignación de valores de verdad de sus variables proposicionales, aunque no como único valor al valor designado (V). P. ej., la fórmula $[p \rightarrow (q \rightarrow p)]$ es una cuasi-tautología de Q.

a	b	$[p \rightarrow (q \rightarrow p)]$
V	V	V
V	F	V, F
F	V	V, F
F	F	V, F

Es posible mostrar que todas las tautologías bivalentes son tautologías de Q. Para ello hay que verificar que todos los axiomas del cálculo bivalente son cuasi-tautologías y establecer la aplicabilidad del Modus Ponens de la siguiente forma:

$$\text{Si } p \rightarrow q \in \text{ c-T (Q) y } p \in \text{ c-T (Q), entonces } q \in \text{ c-T (Q)}$$

donde c-T (X) es el conjunto de las cuasi-tautologías del sistema X. P. ej., si p es una c-T, entonces, para cualquier asignación de valores de verdad, o bien v (p) = V o v (p) = V, F. Pero si $p \rightarrow q$ es una c-T, entonces v (q) no puede ser F para el primer caso y debería ser (V,F) en el segundo, de forma que tendría que ser una c-T.

[Propuesta 7]. Las ideas sobre cuasi-tautologicidad pueden hacer ver cómo desarrollar unas tablas de verdad que, aún siendo cuasi-veritativo funcionales, eviten el problema de la no clausura para las operaciones con valoraciones de verdad en los 'term-sets'. Puede también enseñar cómo diseñar unas tablas de verdad complejas que ayuden a proponer intentos de solución al problema de la aproximación lingüística. No obstante, la dificultad que supone la variabilidad de la vaguedad hará brotar problemas que por ahora todavía no intuimos.

8.2. El segundo tema al que tenemos pensado dedicar nuestra atención se relaciona con lenguajes de programación específicos para datos o conocimientos borrosos, en concreto, el Fuzzy-PROLOG.

Como es sabido, [Kowalski, 1979] consideró que la lógica clásica bivalente podía ser aproximada a un lenguaje de programación una vez que sus fórmulas fuesen puestas de una determinada forma: en forma clausulada. Los resultados de Kowalski se basaron en los descubrimientos de Robinson: en su teorema de resolución, y en el concepto de prueba por refutación, herramientas a partir de las cuales era posible automatizar, de una manera sencilla, las deducciones en lógica. Ambas ideas se apoyaban en un teorema bien conocido de la lógica clásica, que indica, de modo informal, que para cualquier expresión bien formada del cálculo lógico de primer orden, existe una formulación equivalente que consta únicamente de literales (afirmados o negados) unidos por disyunciones (minitérminos). De las fórmulas que constan de un producto de sumas (Prenex) se dice que están en primera forma canónica y sobre ellas pueden actuar métodos de minimización tradicionales.

Pues bien, cada minitérmino puede ser resuelto con otro minitérmino o literal empleando la regla de resolución. Y cada minitérmino puede ser puesto en forma clausulada, de manera que los literales negativos pasan al lado derecho de la implicación y los positivos al lado izquierdo. P. ej. y para el Modus Ponens, tenemos estas tres presentaciones posibles

$A \,\&\, A \rightarrow B$, luego B.	(regla de inferencia lógica)
$A \,\&\, \neg A \vee B$ resuelve B.	(resolución)
$B \leftarrow A$, A, se sigue B.	(procedimiento).

Liu y Li, [1988] y Shen y col. [1988] han mostrado cómo puede darse una generalización de las formas clausales más interesantes (las cláusulas de Hörn) para el caso 'borroso'. El examen comprende tres niveles:

(i). Sintáctico. A la implicación de una regla, formada por una condición y una conclusión, se le asocia un factor *f*, llamado 'fuerza de la implicación'. Si el valor de verdad de la condición es *t*, entonces, el valor de la conclusión será $t \times f$,

(ii). Semántico. El concepto de interpretación se obtiene introduciendo la noción de grado de verdad en el intervalo [0, 1], una vez que se considera a una interpretación borrosa como subconjunto de la base Herbrand en el sentido de [Zadeh, 1965].

(iii). Teoría de la prueba. Se generaliza la búsqueda en árboles Y/O asociándole grados a los arcos.

En el caso (i), la fuerza de la implicación no tiene por qué ser una constante o una variable, sino que puede ser una función. *[Propuesta 8]*. Investigar cómo las funciones generadoras de Modus Ponens de Trillas y Valverde pueden ayudar a caracterizar (i) quizás es una tarea interesante. Caracterizar el uso de los cuantificadores borrosos y cómo estos podrían ser eliminados de la misma forma que lo son, en el caso clásico, los cuantificadores existenciales a través de las funciones de Skolem, es otra labor a completar.

(i), (ii) e (iii) permiten hablar de un fuzzy-PROLOG, considerablemente desarrollado ya por Li y Liu para el diseño de bases de datos relacionales [Li y Liu, 1990]. Sin menoscabo de su interés, de momento resulta de una generalización de los resultados de Kowalski para la lógica de predicados como un lenguaje de programación. *[Propuesta 9]*. Pero el carácter lingüístico de la lógica borrosa quizá demande lenguajes orientados específicamente para tratar con valoraciones lingüísticas. La presentación de la lógica borrosa como una gramática formal borrosa podría ayudar en este intento.

9. RECONOCIMIENTOS

Una parte de la investigación expuesta en este trabajo ha sido sugerida y alentada por el profesor Enric Trillas, al que expresamos nuestro reconocimiento. El profesor Settimo Termini nos ha dado su sabio consejo en múltiples ocasiones, y con él estamos en deuda.

REFERENCIAS

Adan, M. y Sobrino, A., «Nota sobre la teoría de prototipos», en *Actas del Primer Congreso Español sobre Tecnologías y Lógica Fuzzy*. Universidad de Granada, (1991), 173-177.

Aguilar-Martín, J. y López de Mántaras, R., «The Process of Classification and Learning the Meaning of Linguistic Descriptions of Concepts», en Gupta, M.M. y Sánchez, E. (Eds.), *Approximate Reasoning in Decision Analisys*, North-Holland, The Netherlands, (1982), 165-177.

Alsina, C., Trillas, E. y Valverde, L., «On some logical conectives of fuzzy set theory», *Journal of Mathematical Analalysis and Applications*, Vol. 93, N. 1, (1983), 15-36.

Alston, W.P., *Filosofía del Lenguaje*, Alianza Universidad, Madrid, 1980². Ver. orig. en Prentice-Hall, Inc. Englewood Cliffs, New Yersey, USA, 1964.

Bellman, R. y Zadeh, L.A., «Local and fuzzy logics», en Dunn y Epstein (Eds.), *Modern Uses of Multiple-Valued Logic*, Reidel Pub. Co., 1977, 105-165.

Black, M., «Vagueness: An Exercice in Logical Analysis», *Philosophy of Science*, N. 4, (1937), 427-455.

Black, M., «Reasoning with Loose Concepts», en *Margins of Precision*, Cornell, U.P., 1963.

Bosque, I., «Perspectivas de una lingüística no discreta», en Abad y otros, *Metodología y gramática generativa*, Sociedad Española de Librería, S.A., Madrid, (1979), 81-111.

Deaño, A., *Introducción a la lógica formal*, Alianza Universidad, Textos, Madrid, 1978.

Dubois, D. y Prade, H., *Fuzzy Sets and Systems. Theory and Applications*. Academic Press, New York, 1980.

Evans, G., «Can there be vague objects?», *Analysis*, N. 38, (1978), 208.

Fine, K., «Vagueness, truth and logic», *Synthese*, N. 30, (1975), 265-300.

Frege, G., «Lettera del sig. G. Frege all'editore (G. Peano)», (1879), en Angelelli, I. (Ed.). *Gottlob Frege*, Kleine Schriften, Olms, Hildesheim, 1967, 235-236.

Godó, Ll., *Contribució a l'estudi dels models d'inferència en els sistemes posibilístics*, Tesis Doctoral, Universidad Politécnica de Cataluña, 1990.

Kempson, R., *Teoría semántica*, Ed. Teide, 1982, Barcelona. Vers. orig. en Cambridge U. P. Cambridge, 1977.

Kowalski, R., *Lógica, programación e I. A.*, Editorial Díaz de Santos, Madrid, 1986. Vers. orig. en Elservier Science Pub., New York, 1979.

Kuhn, T.S., *La estructura de las revoluciones científicas*, F.C.E. México. Vers. orig. en University of Chicago Press, 1962.

Li, D. y Liu, D., *A Fuzzy Prolog Database System*, John Wiley and Sons, Inc., New York, 1991.

Liu, D. y Li, D., «Fuzzy Reasoning Based on f-Hörn Clause Rules», in Grabowski, J. y col. (Eds.), *Algebraic and Logic Programming*, Springer Verlag, (1988), 214-222.

López de Mántaras, R. y Trillas, E., «Towards a measure of the degree of synonymy between concepts», en Sánchez, E. (Ed.), (1982), *Fuzzy information, knowledge representation in decision analysis*, Pergamon Press, (1984).

Novak, V., *Fuzzy Sets and their applications*, Adam Hilger, New York, 1989.

Olivas, J.A. y Sobrino, A., «Aproximación lingüística y teoría de prototipos de Zadeh», en *Actas del Primer Congreso Español sobre Tecnologías y Lógica Fuzzy*, Universidad de Granada, (1991), 179-183.

Osheron, D. y Smith, E.E., «On the Adecuacy of prototype theory as a theory of concepts», *Cognition*, N. 9, (1981), 35-58.

Parikh, R., «The Problem of Vague Predicates», en Cohen y Wartofsky (Eds.), *Language, Logic and Method*, Reidel, (1983), 241-261.

Popper, K.R., *Búsqueda sin término*, Tecnos, Madrid, 1974.

Rescher, N., *Many-valued Logics*, McGraw-Hill, 1969.

Riera, T. y Trillas, E., «Towards representation of synonyms in antonyms by fuzzy sets», *Busefal*, N. 5, (1981), 42-68.

Rolf, B., *Topics on vagueness*, Tesis Doctoral, Lunds Universitet Filosofiska Institutionen, 1981.

Rosch, E.H., «On the Internal structure of perceptual and semantic categories», en Moore, T.E. (Ed.), *Cognitive Development and the Acquisition of Language*, Academic Press, New York, (1973).

Russell, B., «Vagueness», *Australasian Journal of Philosophy*, N. 1, (1923), 84-92.

Schwartz, S.P., «Vagueness and Incoherence: A Reply to Burns», *Synthese*, N. 80, (1989), 395-406.

Shen, Z., Ding, L. y Mukaidono, M., «A theoretical framework of Fuzzy-Prolog machine», en Gupta y Yamakawa (Eds.), *Fuzzy Computing*, Elservier Science Pub., (1988), 89-100.

Sobrino, A., *La vaguedad: Hacia un estudio crítico de algunos aspectos formales y de gestión*, Tesis Doctoral, Universidad de Santiago de Compostela, 1989.

Sobrino, A., «Lógica borrosa y lingüística», en Trillas, E. y Gutiérrez Ríos, J. (Eds.), *La lógica borrosa y sus aplicaciones*, Publicaciones del CSIC, Serie Nuevas Tendencias, (1992)

Trillas, E., «El tratamiento cualitativo de la vaguedad. ¿Hacia una nueva lógica descriptiva?», en Jaimez, R.G. (Ed.), *Algunas cuestiones actuales en teoría de la probabilidad y procesos estocásticos*, Publicaciones de la Universidad de Granada, (1984), 15-32.

Trillas, E. y Sobrino, A., «Primeros conceptos lógicos (expresados de una forma bastante general)», *Agora*, N. 9, (1990), 57-65.

Waismann, F., «Verificabilidad», en Parkinson, G.H.R. (Ed.), *La teoría del significado*, F.C.E Madrid. Vers. orig. en *Proc. of the Aristotelian Society*, Suppl., N. 19, (1945), 119-150.

Yager, R.R. y col., eds., *Fuzzy Sets and Applications. Selected Papers by L. A. Zadeh*, John Wiley and Sons, New York, 1987.

Zadeh, L.A., «Fuzzy Sets», (1965), en Yager, R. y col. (Eds.), (1987), 29-44.

Zadeh, L.A., «A note on prototype theory and fuzzy sets», *Cognition*, N. 12, (1982), 291-297.

Zadeh, L.A., «Coping with the Imprecision of the Real World: An Interview with Lotfi A. Zadeh», (1984), en Yager, R. y col. (Eds.), (1987), 9-28.

APLICACIONES DE LA LÓGICA BORROSA EN CONTROL

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CAP. II	Aplicación de la lógica borrosa en control de procesos y robótica
CAP. III	Control y supervisión de procesos industriales con lógica borrosa
CAP. IV	Fabricación automatizada de amplificadores electrónicos en Televés utilizando lógica borrosa
CAP. V	Aplicación del control borroso en Medicina

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

APLICACIÓN DE LA LÓGICA BORROSA EN CONTROL DE PROCESOS Y ROBÓTICA

Alfonso J. García Cerezo y Aníbal Ollero
Departamento de Ingeniería de Sistemas y Automática. Universidad de Málaga

1. INTRODUCCIÓN

En la última década se ha constatado un crecimiento significativo de aplicaciones basadas en lógica borrosa. En este trabajo se analizan las posibilidades de aplicación de la lógica borrosa en control de procesos y en robótica, especialmente en robótica móvil y navegación autónoma.

Esencialmente, los sistemas de control borroso fueron concebidos con el espíritu de incorporar la experiencia del operador o técnico del proceso [Mamdani, 1974], expresada de forma lingüística, mediante sentencias del tipo:

«Si se aumenta la temperatura, la presión también aumenta y viceversa. De forma análoga, la velocidad puede aumentarse o disminuirse abriendo o cerrando la válvula. Sin embargo, para una determinada apertura de la válvula la velocidad aumenta si la presión aumenta. Así mismo, abriendo o cerrando la válvula se origina una disminución o aumento de la presión».

Un controlador borroso está normalmente descrito por un conjunto de reglas borrosas que constituyen el protocolo de control. En estas reglas se expresan las relaciones cruzadas que existen entre las variables de medida del proceso y las variables de control. Alternativamente, pueden emplearse otras formas para

representar el conocimiento, como tablas de decisión, o funciones de distribución de relaciones borrosas.

En los siguientes apartados se presentarán las principales características de los sistemas de control borroso, y las líneas de trabajo que actualmente se siguen en este campo.

2. ELEMENTOS DE UN CONTROLADOR BORROSO

En muchas de las aplicaciones de razonamiento aproximado, premisas y conclusiones se representan simbólicamente y, por tanto, sus correspondientes funciones de pertenencia no requieren tratamiento significativo adicional.

En el caso del control de procesos, las premisas y conclusiones que intervienen son medidas y actuaciones del proceso, por lo que su representación no es un conjunto borroso. Para el proceso de inferencia se precisa, por lo tanto, un operador de conversión escalar-borroso, es decir, un operador de borrosificación.

Mamdani [1974], introduce operadores de borrosificación abruptos, al considerar que las variables de entrada está adecuadamente medidas. En la Fig. 1 puede observarse este proceso de borrosificación. Otros autores como Graham y Newell [1988] utilizan representaciones borrosas de la medida en base a los términos lingüísticos definidos para las entradas (Fig. 2).

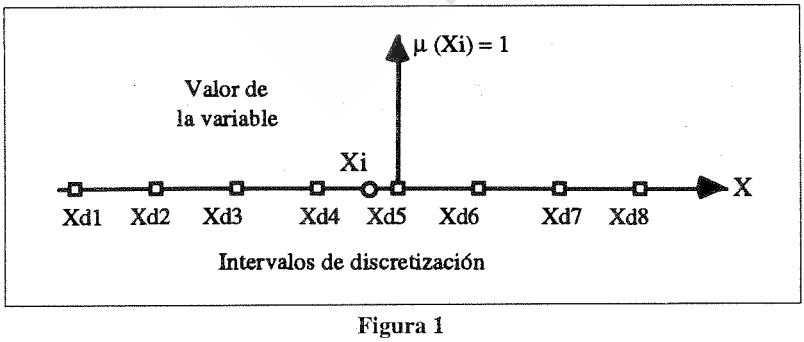


Figura 1

El resultado de la aplicación de la regla de composición produce un subconjunto borroso. La representación de la variable actuación por un subconjunto borroso no es apta para su aplicación directa como variable de control. En consecuencia, es necesario aplicar un operador de reducción a escalar, al que denominaremos operador de desborrosificación.

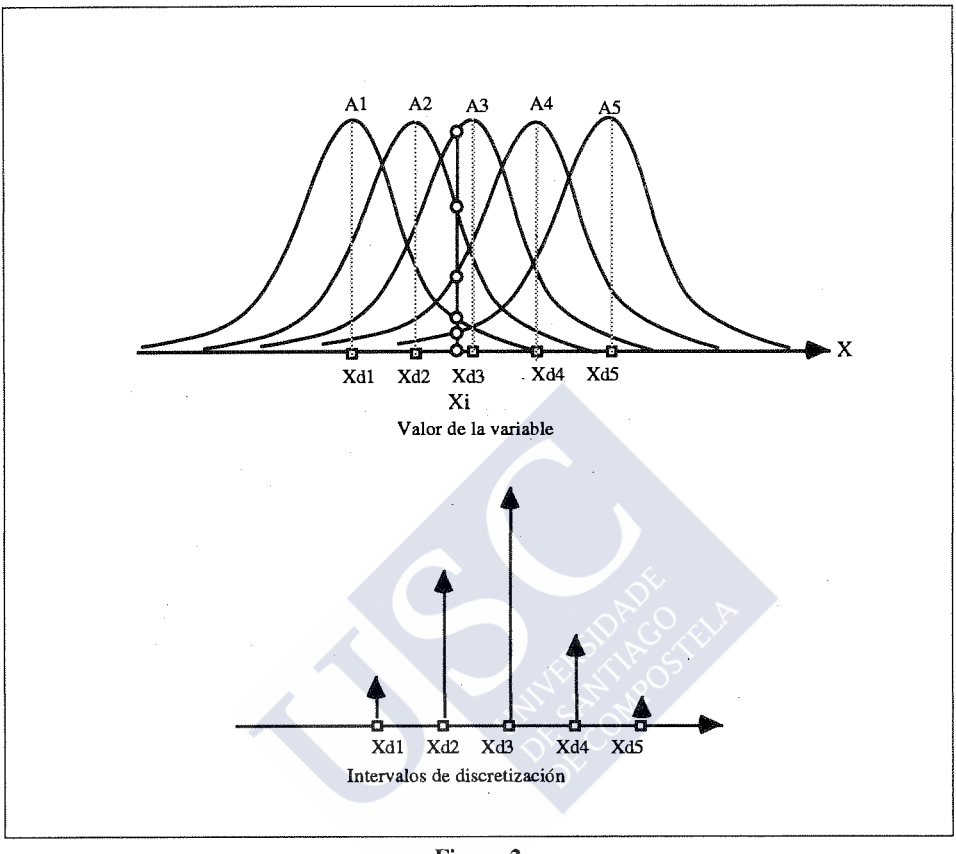


Figura 2

Entre las técnicas de reducción a escalar más utilizadas están: el argumento del máximo (AM), el argumento del área mitad (AAM), y el del Centro de gravedad (ACG). De los tres, éste último es el más utilizado.

3. MECANISMOS DE INFERENCIA

Mediante la Regla de Composición es posible obtener conclusiones a partir de las premisas y la relación establecida en el protocolo. El proceso de inferencia es, en general, lento. Por ello, desde las primeras aplicaciones se ha apreciado un interés notable de los investigadores por la búsqueda de algoritmos más eficientes para la aplicación de la regla general de composición.

Uno de los resultados más significativos lo introdujo Mamdani, debido a su simple y eficiente operador de borrosificación. En este caso, y como se demuestra en Mamdani [1974], Tong [1978], la función de pertenencia del consecuente puede calcularse mediante:

$$\mu_{C'}(u) = \mu_R(u,x) \mid x=x(t)$$

siendo R la relación que define el protocolo de control y x(t) el valor medido en el proceso.

Tong [1978], introduce una reducción importante en el tiempo de procesamiento de las inferencias, aunque de forma indirecta, al analizar el comportamiento de los controladores borrosos sobre la Matriz de Relación (MR), también denominada Matriz Global de Implicación (véase «Soluciones Hardware en Control Borroso»). Esta matriz es una relación borrosa obtenida a partir de las funciones de pertenencia discretizadas.

La introducción por Mamdani de la discretización en la función de pertenencia se debe a requerimientos en el procesado de la inferencia. De ese modo, se obtiene una representación matricial de la relación (la matriz de relación, MR), en vez de un mapa continuo (La relación borrosa, R).

La matriz de relación es, de hecho, una discretización de la relación borrosa definida por el conjunto de reglas.Sin embargo, su mayor interés radica en que representa una versión precompilada de las reglas, lo que desde un punto de vista práctico reduce considerablemente los esfuerzos de control.

Como desventaja, estas representaciones discretizadas introducen comportamientos dinámicos no deseados en el bucle de control.

En García-Cerezo [1987] y Ollero y García-Cerezo [1989], se introduce una mejora significativa en el tiempo de procesado del controlador borroso al aplicar previamente el operador de desborrosificación sobre la matriz de relación. El resultado es otra matriz, denominada matriz de inferencia (F), que unifica los pasos previos de construcción de la matriz de relación y el proceso de desborrosificación.

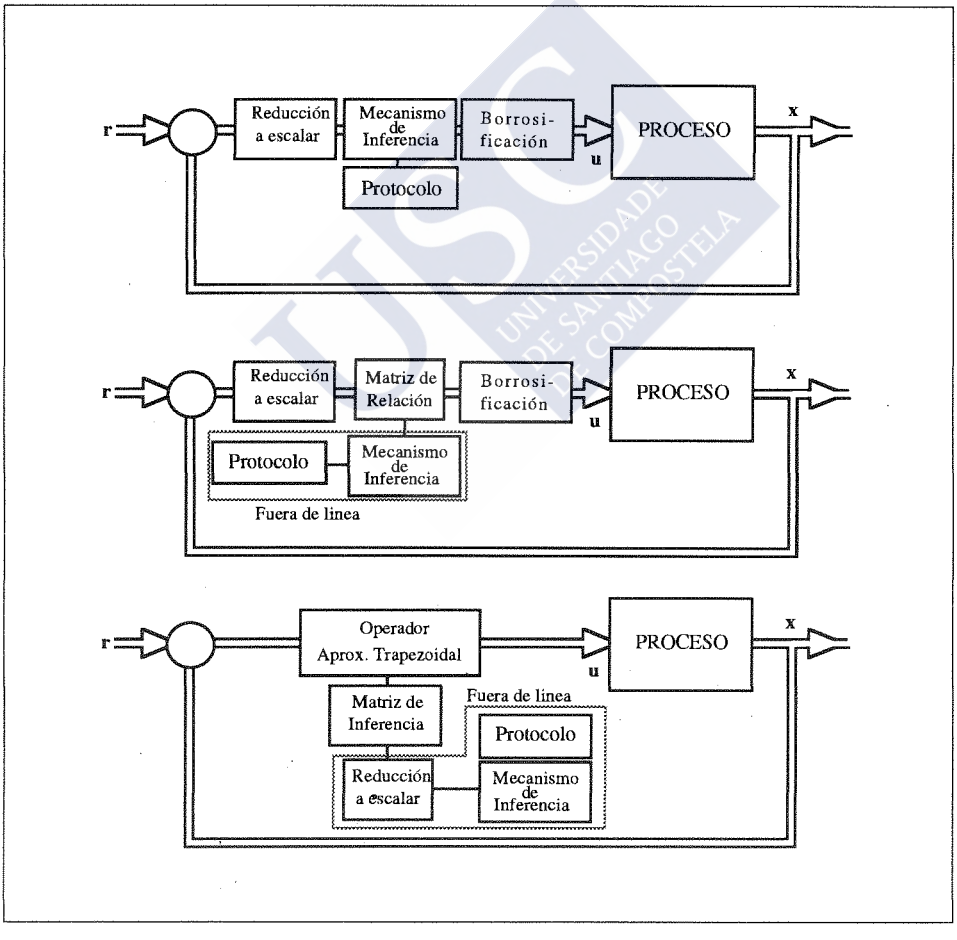
Obviamente, y como en el caso anterior, el proceso se realiza normalmente fuera de línea. La aplicación directa de la matriz de inferencia (interpolación rectangular) permite obtener tiempos de respuesta muy rápidos.

Para resolver el problema de la discretización de las matrices de relación e inferencia, en García-Cerezo [1987] se introduce la técnica de interpolación trapezoidal. Aunque precisa de un mayor tiempo de cálculo que por interpolación

rectangular, mediante ésta técnica se eliminan los efectos no deseados sobre la dinámica del sistema producidos por la discretización. De hecho, la interpolación trapezoidal permite reconstruir los resultados de la inferencia como un mapa continuo Ø, a partir de su discretizado, F.

Para otros operadores de conversión de escalar a borroso, es preciso emplear la regla de composición original. En García-Cerezo [1991] se muestran distintas versiones optimizadas de la regla de composición.

En la figuras 3, 4 y 5, pueden observarse las estructuras elementales de controladores borrosos, atendiendo a la técnica de inferencia utilizada.



Figuras 3, 4 y 5

4. ESTRUCTURAS DE CONTROL

Atendiendo al modo de operar de los controladores borrosos, pueden distinguirse los siguientes tipos:

i) Controladores borrosos directos.
ii) Controladores borrosos auto-organizados.
iii) Controladores borrosos con autoaprendizaje.
iv) Controladores basados en modelos borrosos. Control predictivo.
v) Controladores borrosos híbridos.

4.1. Control borroso directo

Por lo general, las características del controlador se establecen por la naturaleza de las variables de entrada y salida utilizadas. En la mayoría de las aplicaciones se utilizan variables como el error (e) , la derivada o el incremento del error (Δe) y la integral o suma de errores (s). Como variables de control se emplean la actuación (u) y la derivada o el incremento de la actuación. Frecuentemente, también se utilizan representaciones equivalentes a leyes de realimentación no-lineal del estado.

Los tipos más usuales de controladores pueden resumirse en:

Controlador proporcional:	$u = \emptyset(e)$
Controlador integral:	$u = \emptyset(s)$
Proporcional-derivativo:	$u = \emptyset(e, \Delta e)$
Proporcional-integral:	$u = \emptyset(e,s)$
Realimentación no-lineal:	$u = \emptyset(x)$

4.2. Controladores borrosos auto-organizados

Un control borroso auto-organizado (S.O.C.) es un sistema capaz de modificar automáticamente, y sin intervención humana, su base de conocimientos. A partir de las discrepancias con respecto a unos determinados criterios prefijados, el control auto-organizado establece las oportunas modificaciones en la base de reglas.

Introducidos por Mamdani en el 79, los S.O.C. se están mostrando de especial interés en multitud de aplicaciones. Scharf [1985] y Mandic y col. [1985],

introducen los controladores borrosos auto-organizados en el control de robots. Los enfoques seguidos para el desarrollo de los S.O.C. se han basado en criterios de evaluación local, es decir, orientados a acciones de control individuales.

Los elementos que constituyen el S.O.C. (*Self-Organizing Controller*) son los siguientes (ver Fig. 6):

1) El controlador lingüístico, formado por el conjunto de reglas de control, el mecanismo de inferencia, y los operadores de borrosificación y desborrosificación.
2) El algoritmo modificador de reglas.
3) El evaluador de características.
4) El modelo de referencia.

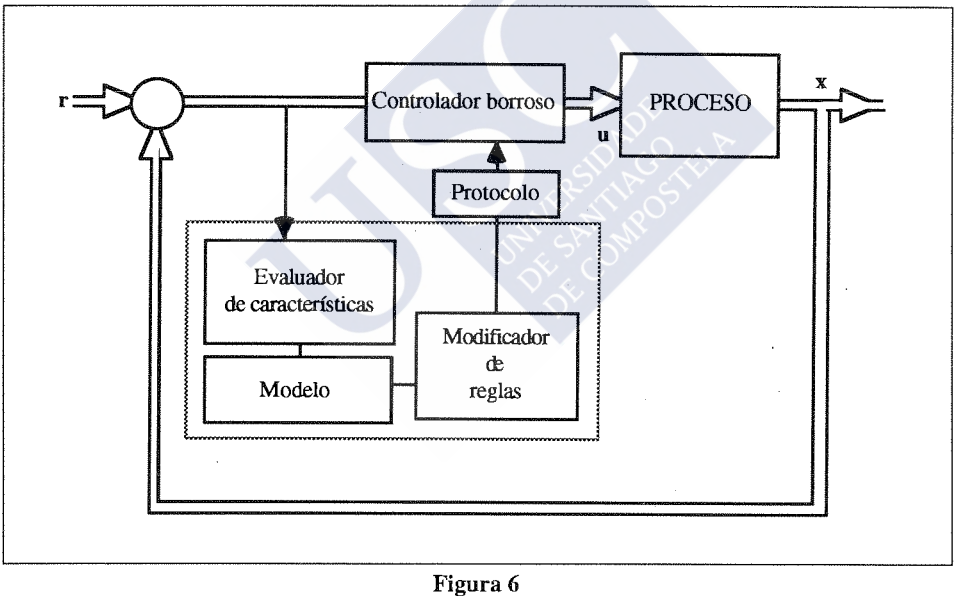


Figura 6

El evaluador de características está constituido por una tabla de decisión que, en función del error y su variación, devuelve una indicación sobre la corrección requerida respecto a una respuesta ideal. Representa de algún modo la idea del diseñador de la tolerancia mínima sobre la respuesta.

A partir del Jacobiano del sistema se construye un modelo incremental que relaciona los incrementos de la actuación con los cambios en las variables de salida.

De la repuesta del evaluador de características y el modelo, se obtiene la corrección necesaria en la actuación.

Finalmente, para producir una modificación en las reglas, se borrosifican los valores instantáneos del error y la variación del error correspondientes al punto de la trayectoria, así como el valor de actuación necesaria, introduciéndolos en el conjunto de reglas de control. Si el sistema presenta un retardo t_r, la nueva regla se introduce para aquellos valores de las variables de entrada aparecidos t_r segundos antes de producirse la evaluación.

Hay que señalar que, en realidad, este mecanismo no modifica las reglas ya adquiridas, sino que incluye nuevas reglas en el protocolo.

Los planteamientos utilizados en los S.O.C. generan una serie de problemas de cuya solución apenas si se ha tratado. Entre otros, el problema de la convergencia del algoritmo del S.O.C., la generación en exceso de nuevas reglas de control, la falta de optimización del protocolo lingüístico, etc.

Parte de los problemas originados por la continua generación de reglas (entre 60 y 100) para cada operación, pueden ser eliminados diseñando sistemas de evaluación alternativos, limitando el número de términos en los antecedentes, e incluyendo modificadores con capacidad de optimización del protocolo de control.

En García-Cerezo [1987] se introduce un controlador borroso auto-organizado basado en un evaluador global de características y en un modelo no lineal del proceso bajo control.

En este caso, la evaluación de características se realiza de forma cíclica sobre las trayectorias del sistema realimentado, mediante un algoritmo borroso. En este se utilizan cuantificadores borrosos del tipo «La mayor parte», «Todos», «Algunos», «El 90 %»,... con objeto de realizar la evaluación. Las reglas utilizadas son del tipo:
«Si la mayoría de los puntos de la trayectoria sobre la regla R son inaceptables entonces evaluar la actuación precisa...»

Aún cuando el número de desarrollos de los S.O.C. no es hoy demasiado significativo, la investigación sobre controladores borrosos auto-organizados presenta un notable interés teórico y práctico; es en aplicaciones en robótica donde más desarrollos se han realizado.

En Shao [1988] se introducen algunas mejoras en cuanto al comportamiento en aprendizaje del esquema propuesto por Mamdani, aplicándolo al control de un

servomecanismo. En Tanscheit y Scharf [1988], se presentan nuevos experimentos en control y seguimiento de trayectorias en robots. En Zeungnan y Jongcheol [1991], se utiliza un S.O.C. para el seguimiento visual en robots.

Finalmente, en Hayashi [1991] se propone una regulador PI auto-sintonizado, que incluye un paso adicional de escalado, y que mejora el número de pasos requeridos en la modificación del protocolo. Como se comprueba en Braae y Rutherford [1979] o en Aracil y col. [1989], la dinámica del sistema en bucle cerrado se ve notablemente afectada por los factores de escalado (fundamentalmente en la actuación).

Cabe señalar que la mejora del comportamiento de los S.O.C. depende en gran medida de la selección del tipo de elementos a ajustar. Por esta línea se están desarrollando nuevas alternativas de controladores borrosos.

4.3. Controladores borrosos con autoaprendizaje

Entre las técnicas de control borroso más reciente se encuentran las que podríamos denominar «controladores borrosos con autoaprendizaje». En su mayor parte responden a procedimientos y técnicas de aprendizaje del tipo gradiente descendente.

En Takagi [1990] y Horikawa y col. [1990] se utilizan las similitudes entre los controladores borrosos y autómatas con capacidad de aprendizaje en la realización de controladores con aprendizaje (ver Fig. 7). Nomura y col. [1991] proponen un método de aprendizaje por propagación hacia atrás, que opera sobre un conjunto de reglas del tipo

$$\text{Regla i: Si } x_1 \text{ es } A_1 \text{ y } \ldots x_n \text{ es } A_n \text{ entonces y es } w_i.$$

en donde A_i está representado por un conjunto borroso de forma triangular, definido por su centro, c_i ,y su ancho, a_i. Los consecuentes, w_i , son escalares.

Como puede comprobarse, en éste método se introducen dos simplificaciones importantes. La salida del controlador queda definida como una función de los parámetros a_i, c_i y w_i .

En Glorennec [1991], se emplean técnicas similares para la realización de un control borroso adaptativo. En éste caso, el controlador opera esencialmente como un modelador de la dinámica inversa del proceso.

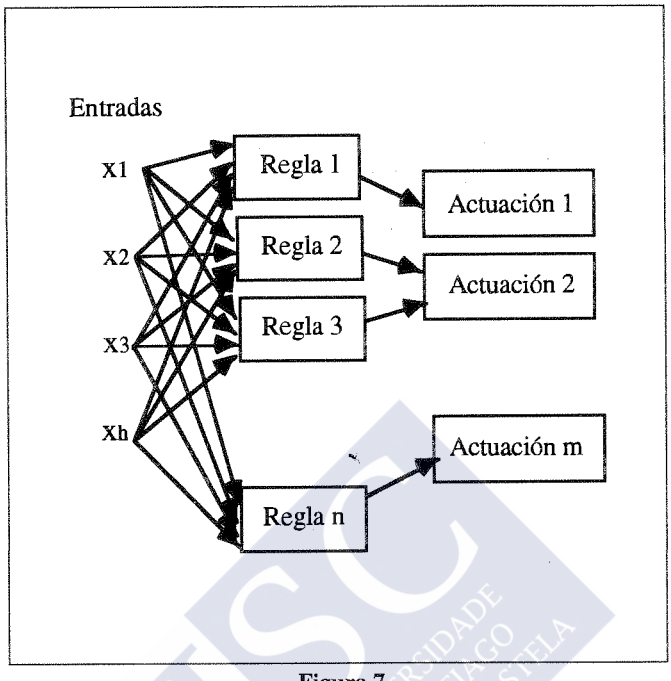


Figura 7

4.4. Controladores basados en modelado borroso

Otra alternativa de gran interés son los controladores basados en modelos borrosos del proceso a controlar. Estas técnicas tienen sus orígenes en los primeros trabajos de Takagi y Sugeno [1985], Sugeno y Nishida [1985] sobre identificación de sistemas borrosos y su aplicación al modelado de sistemas.

Como en casos anteriores, la aplicación de las técnicas de modelado precisan unas serie de simplificaciones sobre los parámetros relativos a antecedentes y consecuentes. Por ejemplo, Sugeno y Nishida [1985] emplea reglas en donde los consecuentes están descritos por combinaciones lineales de las entradas. En Ichihashi [1991], se siguen simplificaciones similares. En Graham y Newell [1988], se emplea una técnica de modelado basada en la modificación directa de la matriz de relación. En este caso, su modificación es sumamente simple, dada la técnica de borrosificación utilizada por Graham. En todo caso, y como se indica en García-Cerezo [1991], los controladores

obtenidos por aplicación de las técnicas de Graham y Mamdani son equivalentes.

Básicamente, los sistemas de control basados en modelos borrosos actúan como modeladores de la dinámica inversa del proceso. Este modelo puede ser realizado fuera de línea, como en el caso de Graham, o realizarse en tiempo real mediante un proceso de auto-aprendizaje.

En Nakamori y col. [1991], se propone un método de diseño de un controlador predictivo mediante el uso de modelos borrosos de sistemas. El diseño consta de dos fases. La primera, de modelado, permite la construcción de un modelo de la dinámica del proceso. La segunda, es una implantación de un esquema de control predictivo.

4.5. Controladores borrosos híbridos

Se denominan así aquellos sistemas de control formados por dos controladores interconectados, de los cuales uno es convencional (como los PID), y otro es borroso. El primero se encarga básicamente del control, garantizando un comportamiento estable.

El controlador borroso actúa en paralelo, introduciendo el componente heurístico en el proceso. Usualmente sus acciones se orientan a la mejora de ciertas características, como reducción de oscilaciones, mejoras del tiempo de establecimiento, etc.

Las primeras referencias a este tipo de estructuras se tienen en Willaeys [1980]. Pero su mayor presencia es, sin duda, a nivel industrial. Una buena parte de los Reguladores Borrosos Industriales han sido realizados mediante esta técnica de hibridación. Un ejemplo se tiene en el regulador de temperatura E5AF de OMROM. Es un regulador PID hibridado con un regulador borroso que funciona en paralelo. En la figura 8 puede observarse algunos de los efectos del ajuste de los parámetros del controlador borroso sobre la dinámica del proceso.

Cabe señalar que estrategias de este tipo son muy adecuadas desde una visión puramente industrial; al disponer de un control clásico, el técnico de proceso se siente apoyado por su propia experiencia con controladores convencionales, y a su vez, se adapta a los nuevos planteamientos introducidos por el controlador borroso.

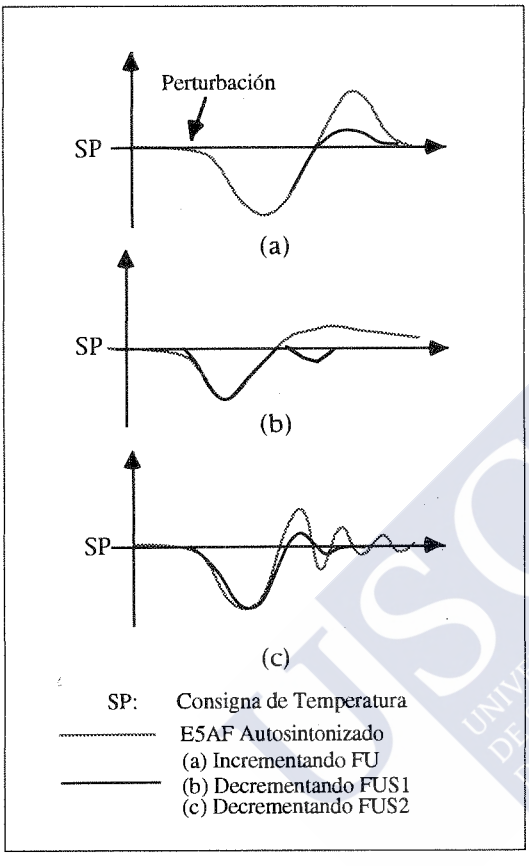


Figura 8. Respuesta del ESAF ante perturbaciones.

5. ROBOTS MÓVILES Y NAVEGACIÓN AUTÓNOMA

Los robots móviles pueden caracterizarse en términos de su movilidad, autonomía, inteligencia y conocimiento previo del entorno.

El sistema de navegación incluye tareas de Planificación, Percepción, y Control de movimientos.

La primera comprende la planificación de la misión, planificación de la ruta, planificación de la trayectoria, y las funciones específicas para evitar obstáculos imprevistos, no considerados en las planificaciones anteriores.

La percepción incluye tareas relacionadas con la detección de obstáculos, construcción de mapas, modelado del entorno y reconocimiento de objetos.

El control de movimientos convierte las órdenes de control en acciones, utilizando para ello los errores entre la posición deseada y la estimación de la posición actual del robot. Se incluyen también tareas de monitorización de la ejecución.

Existen diversas posibilidades de aplicación de métodos de control borroso. La mayoría de los trabajos existentes parten de una identificación del escenario con extracción de características del entorno y obtención de valores de variables de situación del robot (distancia a pared, distancia a entrada de estacionamiento, etc). Las reglas de control borroso relacionan valores cualitativos de estas variables con las acciones de control, típicamente el ángulo de dirección del vehículo. En algunos casos, en los consecuentes de las reglas, en vez de utilizarse directamente la variable de control, se emplea una combinación lineal de los valores de las variables de situación empleando unos coeficientes que se calculan, mediante identificación de las actuaciones de un operador humano, trabajando en simulación. En este punto conviene mencionar el interés de la aplicación de técnicas de aprendizaje, tal como se mencionaba en la sección segunda.

En cualquier caso, una vez obtenidos los valores inferidos por cada regla, la acción final de control resulta de una suma ponderada normalizada de los resultados de cada regla.

Por otra parte, cabe mencionar que el conjunto de reglas puede descomponerse en subconjuntos relativos a operaciones diferentes (conducir hacia delante, girar a la izquierda en intersección) o bien a maniobras diferentes (aparcar en paralelo, aparcar en batería).

Existen también trabajos de aplicaciones de controladores borrosos adaptativos (autoorganizados).

Las aplicaciones mencionadas en los párrafos anteriores cabe considerarlas como de control borroso directo.

Otra posibilidad consiste en aplicar las técnicas de razonamiento aproximado para seguir una trayectoria previamente definida. Considérese un sistema de navegación que realiza sucesivamente los pasos: planificación de la ruta, generación de un camino con buenas propiedades cinemáticas y, finalmente, control del seguimiento de la trayectoria. Para este último paso existen a su vez diferentes

métodos. En efecto, la generación del ángulo de dirección (curvatura) puede realizarse mediante métodos geométricos o considerando también la dinámica.

De acuerdo con el esquema del párrafo anterior, la aplicación de métodos de control borroso puede realizarse directamente (sustituyendo los métodos geométricos y dinámicos), o en combinación con ellos.

De esta forma es posible compensar errores de seguimiento mediante reglas que tienen como antecedentes la desviación de ángulo y la variación de la desviación de ángulo, y como consecuentes la modificación de la dirección para compensar el error.

Asimismo, otra estrategia muy interesante, en línea con lo propuesto en la sección 2, consiste en seleccionar mediante técnicas de razonamiento aproximado los parámetros del método de control, que típicamente se eligen mediante consideraciones heurísticas. Por ejemplo, los métodos geométricos o dinámicos de seguimiento de trayectorias tienen como parámetro importante el número de puntos o distancia sobre el camino que se pretende seguir ("*lookahead*"). Este parámetro puede elegirse mediante técnicas de razonamiento aproximado teniendo en cuenta las características del camino, la velocidad, y la posición del vehículo con respecto al camino.

Los comentarios anteriores se refieren fundamentalmente al control de la dirección. Existe también el problema del control de la velocidad del vehículo. Aunque hay un evidente acoplamiento cinemático y dinámico con el anterior, se ha demostrado que en diversas aplicaciones prácticas pueden realizarse desacoplamientos. En cualquier caso, el problema de planificación y control de velocidad es notablemente complejo y suele realizarse con criterios heurísticos. Por consiguiente, la aplicación de técnicas de razonamiento aproximado es particularmente interesante.

6. CONCLUSIONES

Las técnicas de razonamiento aproximado pueden ser aplicadas a diferentes niveles de control desde el nivel inferior de control en bucle cerrado hasta los superiores de planificación.

En control se precisa de herramientas de diseño de controladores que faciliten la adquisición de conocimiento, y el análisis del sistema de control

resultante, incluyendo el análisis de las propiedades dinámicas de estabilidad y robustez.

En caso contrario el diseño puede convertirse en un proceso muy tedioso, y no garantizarse el comportamiento del sistema de control.

Estos métodos de control pueden aplicarse en control de brazos articulados y vehículos autónomos, en los cuales los modelos matemáticos significativos son muy complejos. En muchos casos interesa combinar las propiedades de un control basado en el modelo, con el de reglas heurísticas, las cuales pueden emplearse para seleccionar o ajustar automáticamente los parámetros de un controlador basado en el modelo.

Asimismo, las técnicas de razonamiento aproximado son interesantes para los niveles superiores de control y planificación de robots cuando el entorno no es conocido en forma precisa.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

REFERENCIAS

Aracil, J., Ollero, A. y García-Cerezo, A. «Stability indices for the global analysis of expert control systems», *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 19, N. 5, Sept-Oct, (1989), 998-1007.

Braae, M. y Rutherford, D.A., «Selection of parameters for a fuzzy logic controller», *Fuzzy Sets and Systems*, (1979).

García-Cerezo, A., *Aplicaciones del razonamiento aproximado en el control y supervisión de procesos*, Tesis Doctoral, Ed. Universitaria, (Univ. de Santiago), 1987.

García-Cerezo, A., «Sistemas de Control Borroso. Panorámica actual. Nuevas aportaciones.», *Jornadas de control inteligente*, Sevilla, Enero, (1991).

Graham, B.P. y Newell, R.B., «Fuzzy identification and control of a liquid nlevel rig.», *Fuzzy Sets and Systems*, Vol. 26, (1988), 255-273.

Glorennec, P.Y., «Adaptive fuzzy control», *V IFSA Congress Vol. Engineering*, (1991), 33-36.

Hayashi, S., «Autotuning fuzzy PI controller», *V IFSA Congress Vol. Engineering*, (1991), 41-44.

Horikawa, S. y col., «A fuzzy controller using neural network and its capability to learn expert's control rules», *Proc. of IIZUKA'90*, Fukuoka, (1990).

Ichihashi, H., «Iterative fuzzy modelling and a hierachical network», *V IFSA Congress Vol. Engineering*, (1991), 49-52.

Mamdani, E.H., «Application of Fuzzy Control Algorithms for Control of simple dynamic plant», *Proc. IEEE*, Vol. 121, N. 12, (1974), 1585-1588.

Mandic, N.J., Scharf, E.M. y Mamdani, H.E., «Practical Application of a Heuristic Fuzzy Rule-Based Controller to the Dynamic Control of a Robot Arm», *IEE Proc.*, Vol. 132, N. 4, (1985), 190-203.

Nakamori, Y., Suzuki, K. y Yamanaka, T., «Model predictive control ussing fuzzy dynamic models», *V IFSA Congress Vol. Engineering*, (1991), 135-138.

Nomura, H., Hayashi, I. y Wakami, N., «A self-tuning methods of fuzzy control by descent methods», *V IFSA Congress Vol. Engineering*, (1991), 155-158.

Ollero, A. y García-Cerezo, A., «Direct Digital Control, Autotuning and Supervision ussing Fuzzy Logic», *Fuzzy Sets and Systems, an international journal*, Vol. 30, (1989), 135-153.

Scharf, E.M., «Fuzzy Logic could Redefine Robot Control», *Robotics*, Febrero, (1985), 11-13.

Shao, S., «Fuzzy self-organizing controller and its applications for dynamic processes», *Fuzzy Sets and Systems*, Vol. 26, N. 2, (1988), 151-164.

Sugeno, M. y Nishida, M., «Fuzzy Control of Model Car», *Fuzzy Sets and Systems*, Vol. 16, (1985), 103-113.

Takagi, «Fusion Technology of Fuzzy Theory and Neural Networks», *Proc. of IIZUKA'90*, Fukuoka, (1990).

Takagi, T. y Sugeno, M., «Fuzzy Identification of Systems and its Applications to Modeling and Control», *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 15, N. 1, (1985).

Tanscheit, R. y Scharf, E.M., «Experiments with the use of a rule based self-organising controller for robotics applications», *Fuzzy Sets and Systems*, Vol. 26, N. 2, (1988), 195-214.

Tong, M.R., «Analysis of Fuzzy Control Algorithms using the relation matrix», *International Journal of Man Machine studies*, Vol. 8, (1978), 679-686.

Willaeys, D., «Optimal Control of Fuzzy Systems», *Fuzzy Sets and Posibility Theory*, (1980).

Zeungnan, B. y Jongcheol, V., «Fuzzy self-organizing visual tracking controller», *V IFSA Congress Vol. Engineering*, (1991), 21-24.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CONTROL Y SUPERVISIÓN DE PROCESOS INDUSTRIALES CON LÓGICA BORROSA

Antonio Barreiro y Rafael Sanz
Depto. de Ingeniería de Sistemas y de Lenguajes y Sistemas Informáticos. Universidad de Vigo

1. INTRODUCCIÓN

La aplicación de las modernas tecnologías de control ha conducido al desarrollo de sofisticados sistemas de control en sectores tales como la industria espacial, plantas químicas, centrales nucleares y otros sectores industriales con entornos peligrosos o dañinos para la salud. El empleo de controles automáticos ha evitado la exposición y manipulación directa por los operarios de los procesos. Obviamente, el comportamiento de estos sistemas tiene que alcanzar unos niveles muy altos de fiabilidad y robustez de funcionamiento. Muchas veces también es necesario que posean un cierto grado de inteligencia que les posibilite la toma de decisiones de manera automática.

Aunque la teoría de control permite resolver muchos de los problemas que surgen en la implantación industrial de controladores robustos, en muchas ocasiones es necesario disponer de mecanismos extras de monitorización y supervisión del comportamiento de estos controladores. Estos mecanismos han sido desarrollados, en gran medida, mediante sistemas de detección de situaciones anómalas en el proceso [Isermann, 1984; Goedecke, 1985; Rault y col., 1984]. Sin embargo, la complejidad de los modernos sistemas de control aconseja, en algunas ocasiones, el empleo de técnicas de inteligencia artificial para la implantación de los mecanismos de supervisión de procesos.

En este contexto, se denominan *controladores inteligentes* aquellos controladores que son diseñados e implantados empleando alguna técnica de inteligencia artificial. Estos controladores intentan modelar el comportamiento de un operador humano que lleva a cabo tareas de control y supervisión.

Los *sistemas expertos* (SSEE) constituyen la herramienta de inteligencia artificial empleada para modelar el comportamiento del experto humano en un dominio específico. Actualmente existen SSEE que resuelven determinados problemas con éxito. Las aplicaciones a control de procesos de los SSEE también se han ido incrementando de forma paulatina. Sin embargo, la mayoría de ellos y las herramientas de desarrollo que se comercializan en la actualidad han sido diseñados para trabajar fuera de línea como equipo de diagnóstico. Son útiles, por tanto, para sistemas de ayuda al diseño, pero no se adecuan correctamente a las operaciones de control y supervisión de procesos en tiempo real. Es necesario, pues, analizar e implantar otros SSEE cuyo comportamiento permita realizar tareas de control y supervisión en línea acoplado al proceso.

2. CONTROL Y SUPERVISIÓN DE PROCESOS

La automatización de un proceso persigue la sustitución, parcial o total, de un operador humano por otro artificial.

Un *sistema de control realimentado* es aquél que tiende a mantener una relación preestablecida entre la salida del sistema y la consigna de referencia, comparando ambas y utilizando la diferencia como parámetro de control. Un sistema realimentado está compuesto de tres partes esenciales (ver figura 1): el sistema controlado, el órgano de medida y el órgano de control.

El *sistema controlado* está constituido por un conjunto de elementos unidos unos a otros de manera que forman un todo. El sistema controlado corresponde a los elementos del proceso a controlar, como por ejemplo un intercambiador de calor.

El *órgano de medida* está formado por un conjunto de captadores que permiten conocer el valor o valores de las variables a controlar. Por ejemplo, en un sistema de control de un intercambiador de calor el órgano de medida empleado sería un termopar.

El *órgano de control* es el encargado de mantener el punto de consigna de la salida; en el caso de regulación, ante posibles perturbaciones externas; y en el caso de servomecanismos, conseguir que la salida responda a los cambios de la consigna.

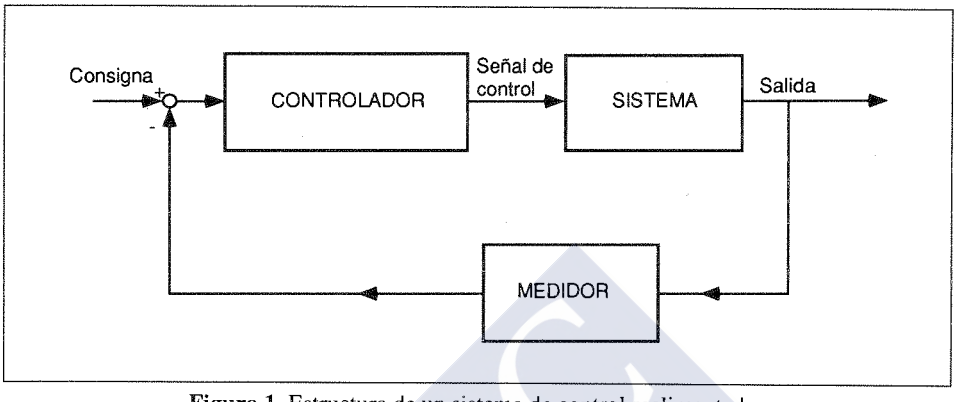


Figura 1. Estructura de un sistema de control realimentado.

La noción de *realimentación* se puede ilustrar claramente mediante multitud de casos prácticos. Por ejemplo, en la figura 2 se muestra un sistema de control de la posición de un timón de un barco. Desde el cuadro de mando se puede elegir un rumbo de consigna. La señal de salida está constituida por la dirección efectiva de la nave. Un instrumento de comparación indica, en todo momento, la diferencia entre la señal de consigna y la salida. Cuando existe una diferencia entre ambas, el sistema de control se encarga de actuar sobre el motor que gobierna el timón, con objeto de reducir esta diferencia. Las posibles modificaciones de rumbo debida a corrientes marinas y otras perturbaciones externas pueden ser corregidas de esta manera (caso *regulación*). Si se cambia el rumbo de consigna, el sistema debe igualmente responder a este cambio haciendo que la nave siga el nuevo rumbo indicado (caso *servomecanismo*).

Actualmente los computadores se emplean con éxito para el control de procesos industriales. Además permiten incorporar otras funciones como las de supervisión, captura y tratamiento de datos, presentación de sinópticos del proceso en pantalla, etc. Sin embargo, muchos procesos continuos complejos se siguen controlando de forma manual por la dificultad existente en la aplicación de la teoría de control cuando no se dispone de un modelo matemático suficientemente fiable. El modelado cualitativo del comportamiento del operador del proceso lleva a considerar técnicas de Inteligencia Artificial (IA) para la implantación de controladores inteligentes.

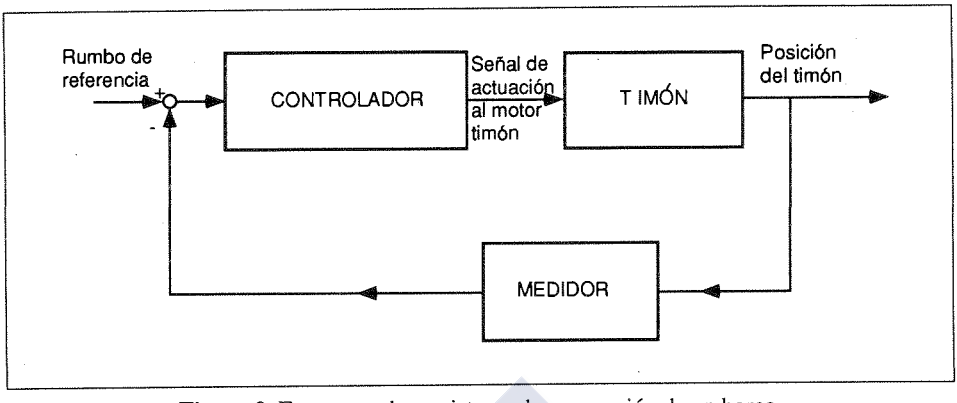


Figura 2. Estructura de un sistema de navegación de un barco.

Por otra parte, la *supervisión* de un sistema de control puede ser considerada, desde el punto de vista funcional, como un sistema que implanta las diferentes funciones de monitorización y supervisión. Estas funciones pueden corresponder a rutinas, que una vez escritas se integran como parte del programa de control [Gertler y Chang, 1986], o bien son implantadas empleando alguna técnica de inteligencia artificial.

En general, un sistema de supervisión debe llevar a cabo las siguientes operaciones:

a) *Monitorización* de las operaciones de la planta y del controlador.

b) *Detección de posibles fallos y funcionamientos incorrectos* de algunos de los componentes, en cuyo caso, si es posible, se reconfiguraría el sistema (modificando o cambiando los algoritmos de control) para mantener el comportamiento dentro de unos determinados márgenes de estabilidad.

c) *Selección de los parámetros de los algoritmos*, de manera que el comportamiento se ajuste a cada situación particular.

Teniendo en cuenta la supervisión, y desde un punto de vista jerárquico, el sistema de control puede estar compuesto por tres niveles:

•*Nivel de control*, donde se ejecutan los algoritmos de control.

- *Nivel de captación* o *de monitorización del comportamiento*, donde se realizan los cálculos con los datos del nivel de control, suministrando la información procesada al nivel superior.

- *Nivel de decisión*, donde propiamente se llevan a cabo las funciones denominadas inteligentes y la toma de decisiones: localización y clasificación de sucesos, selección de algoritmos, ajuste de parámetros, etc.

3. SISTEMAS EXPERTOS EN CONTROL Y SUPERVISIÓN

Un sistema experto (SE) se puede definir como un programa de computador que incorpora, de forma operativa, el conocimiento de expertos sobre un dominio particular, y es capaz de razonar usando técnicas de razonamiento simbólico para resolver problemas dentro de dicho dominio; también tiene que ser capaz de explicar sus conclusiones. En la figura 3 se representa la estructura de un SE de aplicación general y sus principales componentes:

- *Base de conocimiento*, que contiene el conocimiento general sobre el problema.

- *Base de datos*, que contiene información sobre el problema particular.

- *Motor de inferencia*, que está constituido por los métodos para aplicar el conocimiento general a la resolución del problema.

- *Componente explicativo*, que informa al usuario de las conclusiones obtenidas.

- *Acoplamiento con el usuario*, que presenta la información al usuario de forma inteligible.

- *Componente de adquisición del conocimiento*, que se encarga de comunicarse con el experto de conocimiento del problema y procesa esta información.

En cualquier caso, ésta no es la única estructura posible. Con frecuencia, algunos de estos componentes están íntimamente ligados y no son separables. Asimismo, se puede extender esta estructura básica a sistemas con capacidad de aprendizaje.

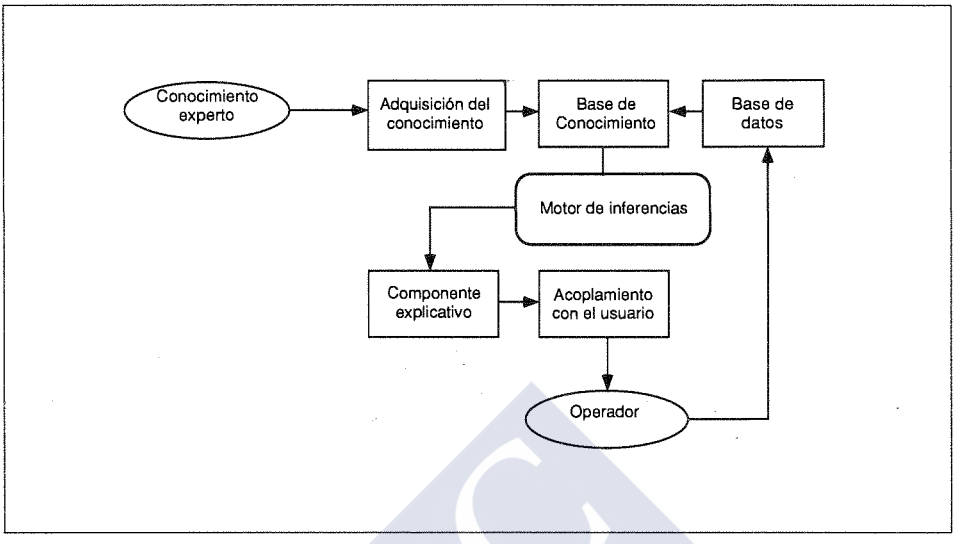


Figura 3. Estructura de un sistema experto de aplicación general.

3.1. Esquemas de sistemas expertos en control y supervisión

Se pueden considerar como esquemas básicos de SSEE en tiempo real los siguientes:

Sistemas basados en reglas, que intentan modelar la estrategia de control manual del operador de planta. Esto se lleva a cabo mediante reglas que expresan de forma cualitativa cómo se debe elegir la señal de control en diferentes situaciones. Un ejemplo de técnica que utiliza este tipo de representación son los *sistemas de control borroso* (figura 4). Estos sistemas se han empleado en procesos tales como hornos de industrias cementeras, en los cuales no existe o es difícil obtener un modelo matemático del proceso que permita diseñar el control.

Sistemas basados en conocimiento, que emplean técnicas de sistemas expertos para codificar el conocimiento general y la heurística que intervienen en cualquier proceso (figura 5). Este esquema, que inicialmente fue propuesto en Aström y Anton [1984] y en Aström y col. [1986], permite combinar de forma inteligente los algoritmos de control, identificación y monitorización de un proceso [Sanz, 1987].

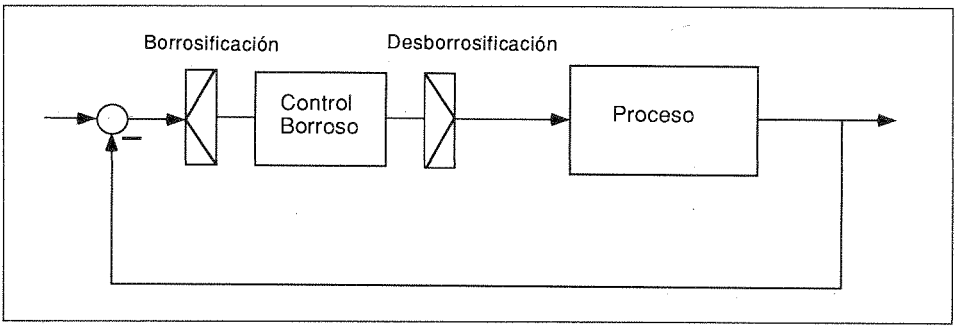


Figura 4. Sistema de control borroso.

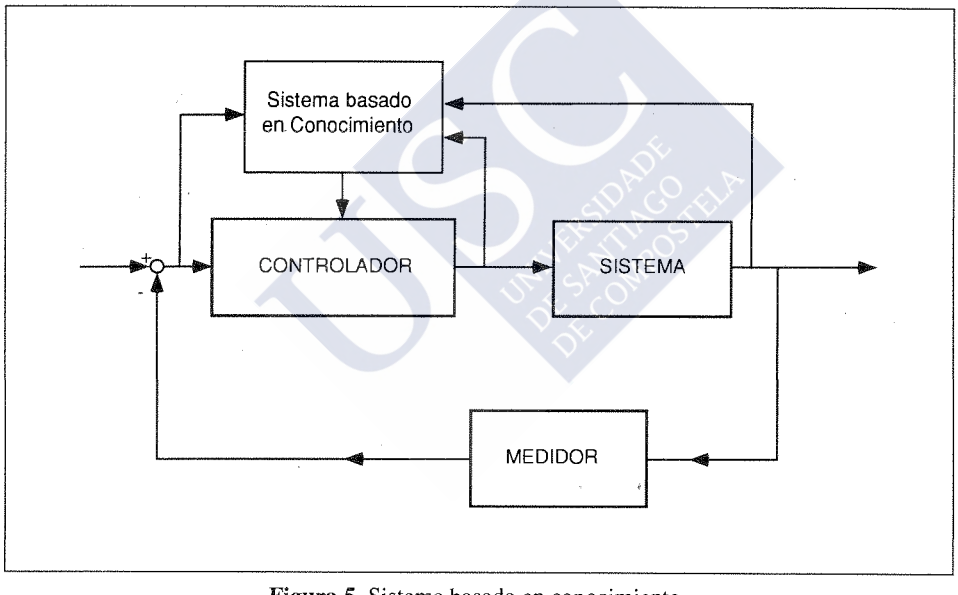


Figura 5. Sistema basado en conocimiento.

Sistemas de inteligencia distribuida, que se caracterizan por una estructura en donde los componentes del SE están fuertemente diluidos en la misma. Un ejemplo de este tipo de sistemas son los que utilizan *redes neuronales* [Widrow y Lehr, 1990] (ver figura 6). Estos sistemas poseen una cierta capacidad de adquirir información de forma redundante, lo cual implica una naturaleza fiable, robusta y tolerante a fallos.

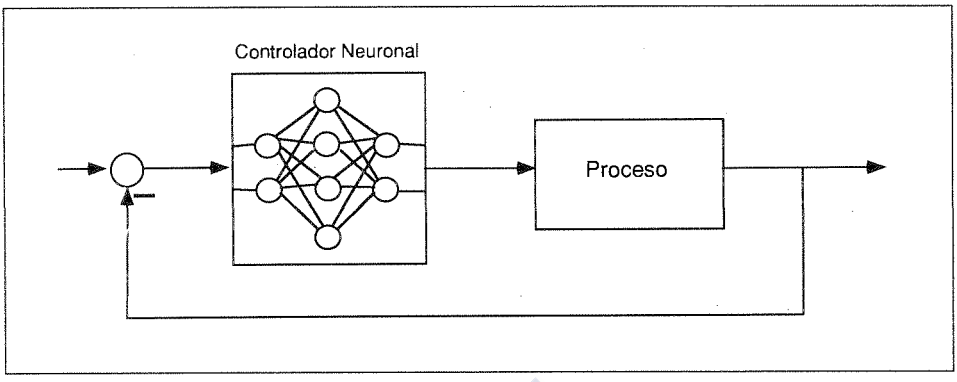


Figura 6. Utilización de redes neuronales.

4. CONTROL Y SUPERVISIÓN DE PROCESOS MEDIANTE CONJUNTOS BORROSOS

En los últimos años, el control borroso se ha convertido en una de las áreas más activas y fructíferas dentro de las investigaciones en teoría de conjuntos borrosos. Los trabajos pioneros sobre control borroso, desarrollados por Mamdami y sus colegas en los años 75-77, fueron inspirados por los trabajos previos de Zadeh (1965-1975) sobre teoría de conjuntos borrosos y razonamiento lingüístico aproximado.

Hay un gran número de aplicaciones recientes entre las cuales se encuentran las de control de calidad de agua, conducción automática de trenes, control de ascensores y grúas, aplicaciones en reactores nucleares, automóviles, máquinas lavadoras, etc. Existen también desarrollos de dispositivos de hardware borroso, memorias borrosas y computadores borrosos.

Todas estas aplicaciones han puesto de manifiesto que la utilización efectiva de un sistema borroso, en el contexto de procesos complejos mal modelados, puede ser controlada por un operador humano experimentado, sin el conocimiento de la dinámica subyacente.

La literatura sobre control borroso ha crecido rápidamente, haciendo difícil la presentación de una recopilación signicativa de las contribuciones fundamentales, que siempre será subjetiva. En la tabla I se resumen cronológicamente algunas de las primeras aportaciones.

TABLA I

Año	*Autor*	*Aportación*
1972	Zadeh	Controlador borroso
1973	Zadeh	Variables lingüísticas
1974	Mamdani y Assilian	Control máquina de vapor
1977	Ostergaard	Control intercambiador de calor y horno cementera
1980	Tong y col.	Procesador de aguas residuales
1983	Takagi y Sugeno	Derivación de reglas de control borroso

Posteriormente, el crecimiento de los trabajos fue explosivo, éxito quizá debido a que la lógica borrosa, que es la lógica sobre la que se asienta el control borroso, está mucho más próxima al pensamiento humano y al lenguaje natural que los sistemas lógicos tradicionales.

Básicamente, la lógica borrosa proporciona un medio efectivo de aprehender la naturaleza aproximada e imprecisa del mundo real. Visto desde este perspectiva, la parte esencial del control y supervisión borrosos es un conjunto de reglas lingüísticas relacionadas por los conceptos duales de implicación borrosa y de regla de inferencia composicional.

En esencia, el control y supervisión borrosos proporcionan mecanismos que permiten convertir una estrategia lingüística basada en el conocimiento experto en una estrategia automática. La experiencia muestra cómo los resultados de aplicación de control borroso resultan mejores que los obtenidos por técnicas de control y supervisión convencional.

En particular, la metodología del control borroso resulta muy útil cuando los procesos son demasiado complejos para ser analizados por técnicas cuantitativas convencionales, o cuando las fuentes disponibles de información son interpretadas de modo cualitativo, impreciso o incierto. Por tanto, el control y supervisión borrosos pueden ser vistos como un punto de aproximación o encuentro entre un control analítico, basado en un modelo matemático exacto, y una estrategia de toma de decisiones.

A continuación, y aunque por el momento no existe un procedimiento sistemático para diseñar un controlador borroso, se establecerá un resumen de la metodología y problemas que surgen en control y supervisión borrosos: las estrategias de *borrosificación* y *desborrosificación*, la derivación de la base de

datos y reglas borrosas, los mecanismos de razonamiento, etc. Nos centraremos en aquellos aspectos específicos de control y supervisión.

4.1. Estructura de un controlador borroso

La configuración básica de un controlador borroso directo se indica en la figura 7 [Lee, 1990]. Esta estructura puede tener algunas variantes, pero siempre se puede organizar en cuatro componentes principales: la *interfase de borrosificación*, la *interfase de desborrosificación*, la *base de conocimiento*, y la *lógica de toma de decisiones*.

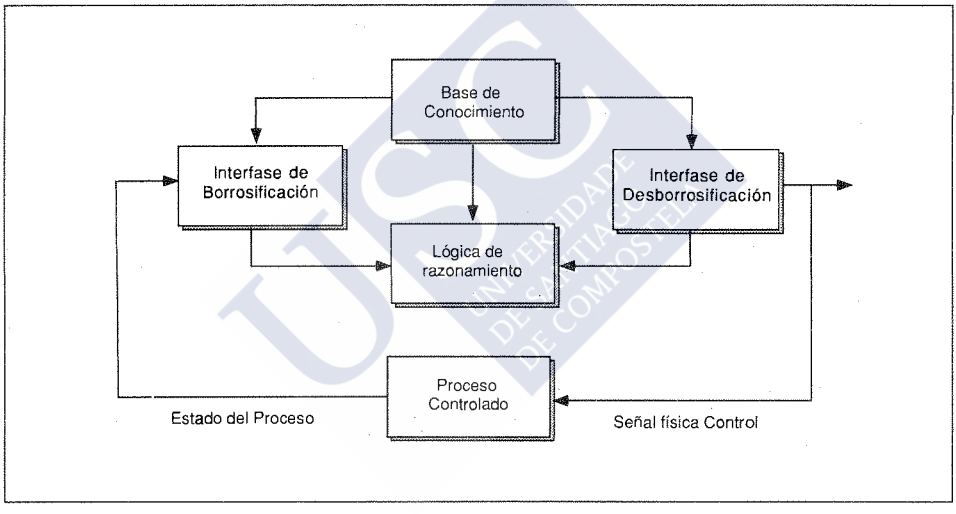


Figura 7. Estructura de un controlador borroso.

En el caso de que el sistema borroso no se dedique al control directo, sino a la supervisión, el esquema es el mismo, salvo que el bloque «proceso» ya no representa la planta aislada, sino la planta con un controlador convencional, algunos de cuyos parámetros se ajustan desde el sistema borroso, en base a las informaciones recibidas sobre el funcionamiento del proceso.

• *Interfase de borrosificación*. Debe realizar las siguientes tareas:

a) Medida de las variables del proceso.

b) Escalado de las mismas, que transfiera su rango de valores a los correspondientes universos de discurso.
c) *Borrosificación*, esto es, conversión a valores lingüísticos que pueden ser considerados como etiquetas de conjuntos borrosos.

• *Base de conocimiento*, que consta de:

d) Una base de datos con las definiciones utilizadas en las reglas lingüísticas y
e) La base de reglas lingüísticas que caracteriza los objetivos del control.

• *Lógica de toma de decisiones*, que es el núcleo del sistema borroso. Tiene la capacidad de simular el razonamiento humano, utilizando las reglas de inferencia de la lógica borrosa.

• *Interfase de desborrosificación* , que realiza las tareas:

f) *Escalado* que transforma los universos de discurso en rangos de variables de actuación sobre el proceso, y
g) *La desborrosificación*, que produce una acción de control no borrosa a partir de una inferencia borrosa.

4.2. Características de un controlador borroso

A continuación se citan algunos aspectos importantes que deben ser considerados en el diseño de un controlador borroso.

Partición borrosa de los conjuntos de entradas y salidas. Este es el problema de la decisión del número de conjuntos borrosos primarios o básicos, número que determinará la granularidad o finura del control borroso que se puede obtener. Por ejemplo, en un sistema con dos entradas (p, q) y una salida, y con Np, Nq conjuntos primarios para la p y la q, el máximo número de reglas borrosas es Np × Nq. Los valores óptimos de los tamaños de particiones Np, Nq se determinan normalmente mediante tanteos sucesivos.

Completitud. Típicamente, un controlador borroso deberá ser capaz de inferir la acción de control adecuada para cualquier estado del proceso. Esto está relacionado tanto con la base de conjuntos primarios como con la base de reglas.

El soporte de un conjunto borroso primario es el intervalo de valores para los cuales el grado de pertenencia es distinto de cero. Para asegurar la completitud es necesario que la unión de todos los soportes de los conjuntos borrosos primarios cubra todo el universo de discurso, de forma que cada punto tenga un grado de pertenencia mayor que α con respecto de algún conjunto. Normalmente, se elige α=0.5. La *completitud* se incorpora a la base de reglas a través de la experiencia del diseñador. Cuando en alguna situación no se dispara ninguna regla con un grado suficiente de pertenencia, se deben incluir reglas adicionales.

Fuentes de obtención de las reglas. Hay tres fuentes, no mutuamente exclusivas, para obtener las reglas de control borroso:

- Experiencia y conocimiento del ingeniero de control. Consiste en la verbalización de la experiencia diaria sobre la planta, expresada de forma lingüística. Queda formalizada, por ejemplo, en el manual de operación del proceso, que registra las acciones de control del operador humano.

- En base a un modelo borroso del proceso. Este método es más fiable, en el sentido de que se parte de una estructura (modelo borroso del proceso) cuyo tratamiento se puede abordar metodológicamente con más rigor. Sin embargo, este método no está suficientemente desarrollado.

- En base al aprendizaje. Aquí existen dos niveles, el de la base de reglas, y el de las «meta-reglas», que crean y modifican reglas de base, hasta conseguir un comportamiento adecuado. Un ejemplo de aplicación de este método lo constituyen los *controladores autoorganizados*.

5. SISTEMAS BASADOS EN CONOCIMIENTO

Desde el punto de vista de control de procesos, estos sistemas pueden ser considerados como un sistema experto añadido a un sistema de control convencional.

Sin embargo, hay que resaltar que estos sistemas van a diferir, en su modo de operar, de los SSEE tradicionales. En efecto, el esquema indicado en la figura 3, puede ser considerado como un sistema que trabaja en lazo cerrado a través del operador. El sistema dialoga interactivamente, con lo cual el razonamiento sólo puede ser realizado en la medida en que el operador introduce datos para realizar sucesivos razonamientos. Para construir un SE para control y supervisión de

procesos, a partir de uno convencional, es necesario, desde el punto de vista conceptual, desplazar al operador humano del bucle, construir un subsistema de adquisición de datos para preprocesar las salidas de los sensores, y sintetizar los algoritmos de control apropiados (figura 8).

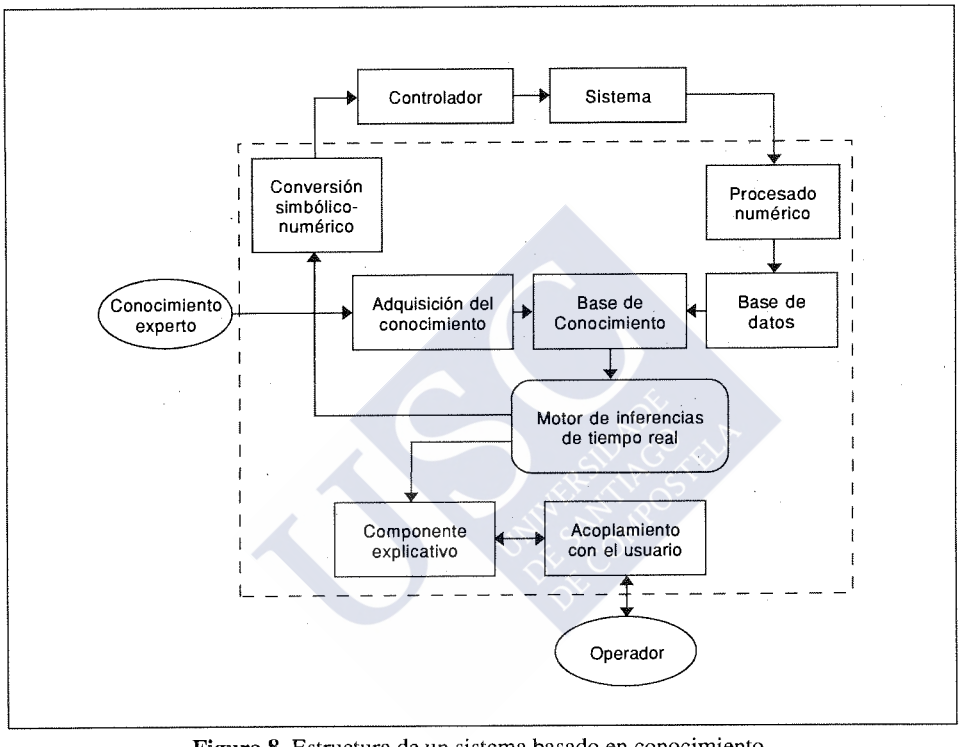


Figura 8. Estructura de un sistema basado en conocimiento.

Por otra parte, la naturaleza de tiempo real de estos sistemas obliga a realizar consideraciones específicas sobre el comportamiento de un sistema basado en conocimiento de tiempo real, ya que éstos se diferencian sustancialmente de los sistemas expertos desarrollados para aplicaciones de consulta interactiva.

Para la construcción de sistemas basados en conocimiento se pueden seguir básicamente dos alternativas. Por un lado, adaptar un SE de aplicación general, creando un módulo de procesado numérico (de las variables provenientes de los sensores y de los algoritmos de control e identificación) y un módulo de conversión simbólico-numérico que transforme las conclusiones simbólicas en valores

numéricos que son capturadas por el algoritmo de control. La otra posibilidad es el empleo de un sistema experto de tiempo real, cuya estructura está diseñada de manera que pueda ser integrado en un sistema de control convencional. Es evidente que, en la medida en que no existían productos adaptados a la manipulación directa de los valores proporcionados por los sistemas de control y a su integración en tiempo real, era necesario emplear sistemas de aplicación general. Sin embargo, existen en la actualidad productos comerciales diseñados para este cometido. Cabe mencionar entre ellos [Arzén, 1992]: G2, RTAC, COGSYS, RTworks y TDC3000 Expert.

5.1. Aspectos de tiempo real de los sistemas basados en conocimiento

Un sistema basado en conocimiento de tiempo real puede definirse como un sistema para el cual existe un tiempo fijo, al término del cual debe producir una respuesta [Laffey y col., 1988]. Las características de tiempo real se concretan en [Arzén, 1989]:

Razonamiento no monótono. El razonamiento se lleva a cabo con datos que provienen de sensores y que son variables en el tiempo, lo cual provoca un cambio en el estado del sistema y la posible invalidez de determinados sucesos. El razonamiento debe ser capaz de retraer de forma automática las situaciones o eventos inferidos con objeto de mantener una visión consistente del entorno.

Sucesos asíncronos. El sistema debe ser capaz de responder a sucesos que pueden aparecer de una forma arbitraria en el tiempo. Es de suma importancia la manipulación de interrupciones para atender a procesos prioritarios.

Razonamiento temporal. Los sistemas de tiempo real tienen que permitir representar el tiempo y razonar sobre sucesos pasados, presentes y futuros, y sobre la secuencia de sucesos.

Restricciones en el tiempo de razonamiento. Las conclusiones del razonamiento deben estar disponibles cuando la solución es necesaria, dentro de unos márgenes de tiempo fijados. Aquí se incluye también la posibilidad de estimar o medir el tiempo necesario para realizar el razonamiento.

Razonamiento en paralelo. Muchos de los problemas de razonamiento se pueden plantear como un conjunto de actividades que se pueden ejecutar en paralelo. Ello obliga a disponer de mecanismos de sincronización entre las

diferentes actividades y la posibilidad de suspender una actividad un cierto tiempo o cuando ocurre una determinada situación.

5.2. Características de un sistema basado en conocimiento de supervisión de procesos

Los sistemas basados en conocimiento de tiempo real están especialmente orientados a la monitorización y a la supervisión de procesos, ya que pueden interaccionar con algoritmos clásicos de control sin necesidad de eliminarlos. En todo caso, el sistema debe tener un comportamiento análogo al de un operador experimentado ante las diferentes situaciones de funcionamiento del sistema.

Los principales objetivos de un sistema basado en conocimiento de supervisión en control de procesos son:

- Mejora del comportamiento en bucle cerrado.
- Selección, en línea, de los algoritmos de control y, en su caso, de identificación a emplear.
- Modificación en línea de las funciones objetivo, por ejemplo, la modicación de los índices cuadráticos de un control óptimo.
- Verificación de los índices de robustez.
- Detección de situaciones anómalas o cambios bruscos de funcionamiento.
- Vigilancia de determinados comportamientos o situaciones de tipo algorítmico: «*antiwindup*», traza de la matriz de covarianza, etc.
- Aprendizaje de situaciones patrón o estereotipadas.
- Coordinación con el funcionamiento de otros posibles controladores y con los objetivos globales de funcionamiento, con un determinado horizonte de tiempo.

Es evidente que no todas estas funciones pueden ser llevadas a cabo simultáneamente. De hecho, existen tareas fáciles de supervisar y de introducir dentro del programa de control, como, por ejemplo, la toma de decisiones ante el valor de la señal de control. Sin embargo, existen otras funciones de supervisión

que, aparte de su complejidad, por restricciones temporales no pueden ser llevadas a cabo con la misma periodicidad. Esta periodicidad sugiere la división del esquema de supervisión inteligente en varios niveles. Cada uno de éstos realiza diferentes funciones de supervisión a distintas escalas de tiempo. Cabe pensar que cada nivel puede ser implantado por un SE con características propias y funcionando de forma independiente con escalas de tiempo diferentes. Estos sistemas colaboran entre sí comunicándose datos mediante un mecanismo denominado de *pizarra* («blackboard»). La pizarra es un conjunto de posiciones de memoria central del computador en el que un SE puede almacenar información, de forma que otro SE pueda emplearla si ello es necesario para alcanzar sus objetivos. Por lo tanto, la pizarra es una estructura que contiene información que puede ser vista por los SSEE cooperantes. La manipulación de esta información depende de la aplicación.

Hay que resaltar que las técnicas de supervisión inteligente introducen siempre una complejidad adicional en el diseño del sistema, por tanto, no deben ser empleadas en aquellos casos en los cuales se puede lograr un funcionamiento aceptable del sistema de control mediante otras técnicas de control y de supervisión.

6. REDES NEURONALES EN CONTROL Y SUPERVISIÓN INTELIGENTE DE PROCESOS

En los comienzos de la IA, se podían encontrar dos objetivos principales, que originalmente parecían dos aspectos del mismo problema: diseñar máquinas inteligentes y entender la inteligencia humana.

J. von Neumann buscaba una teoría que estableciese las diferencias lógicas entre autómata natural y artificial. McCulloch y Pitts mostraron cómo las neuronas formales, similares a las células nerviosas (neuronas naturales), alcanzan la misma capacidad de computación que una máquina de Turing.

La cuestión de cómo hacerlas aprender fue abordada por Rosenblatt y Minsky mediante el *perceptrón*. Aunque esas primeras redes aprendían y reconocían patrones sencillos, sus limitaciones detuvieron momentáneamente, a finales de los 60, las investigaciones en este campo.

Recientemente, ha aumentado el interés en IA y redes neuronales, debido a los avances en algoritmos de entrenamiento y a los desarrollos en hardware, que hacen posible implantar las redes mediante computadores analógicos o digitales

más rápidos y económicos, bien sea en equipos específicos o en computadores convencionales.

En general, el trabajo de investigación en redes neuronales se divide en tres áreas:

- Algoritmos para determinar conexiones y pesos;
- Realización microelectrónica de las redes;
- Investigación de las redes neuronales biológicas naturales.

Los desarrollos y aplicaciones incluyen procesado de señal, reconocimiento adaptativo de patrones, procesado de voz e imagen.

En particular, en control y supervisión se han desarrollado técnicas neuronales para identificación, control adaptativo y control de sistemas de gran dimensión, que se han aplicado en las industria aeroespacial, cementeras, robótica, etc.

6.1. Redes neuronales y control inteligente

Como indica Bavarian [1988], uno de los principales propósitos del control clásico realimentado es el de aumentar la *robustez* del proceso, es decir, incrementar el margen de buen funcionamiento aunque aparezcan perturbaciones o incertidumbre en el sistema de control. Hay que observar que el control clásico proporciona robustez, pero con un margen relativamente pequeño de incertidumbre. Para mejorar este aspecto, se han introducido diferentes técnicas de control avanzado. Entre éstas se encuentran los sistemas de control, denominados *controladores adaptativos*, que abordan los problemas de cambios en los parámetros del proceso, y de otras señales en el entorno. El control adaptativo generalmente añade sobre el lazo de realimentación un segundo lazo, cuya tarea consiste en el seguimiento de los parámetros del proceso con objeto de obtener un sistema de control óptimo desde el punto de vista del diseño del controlador. La figura 9 muestra un diagrama de bloques general para este esquema de control.

Desde el punto de vista de diseño, el objetivo de un sistema de control inteligente es similar al de control adaptativo. La diferencia es que para un sistema inteligente, generalmente se supone un rango de incertidumbre que pueda ser mucho mayor que el tolerado por un sistema de control adaptativo. Por ejemplo, el rango de incertidumbre puede ser mucho más amplio del esperado debido a la presencia de perturbaciones del proceso no cuantificables y no estructuradas.

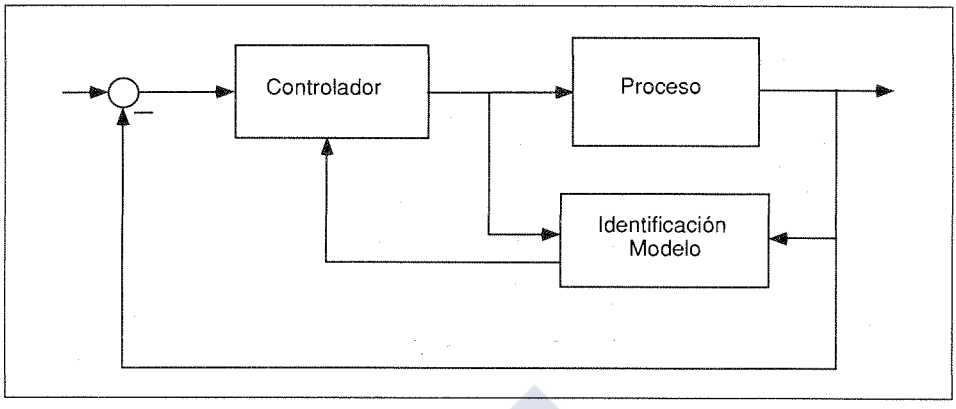


Figura 9. Estructura de un sistema de control adaptativo.

Con la utilización de redes neuronales, el problema de control puede ser asimilado al problema de *reconocimiento de patrones*. El controlador neuronal inteligente debe reconocer y aislar, de la lectura de variables, patrones de comportamiento, aprender, a partir de la experiencia, a reconocer rápidamente los cambios de situación e, incluso, trabajar con información incompleta. Estas propiedades de reconocimiento de patrones, realización de funciones no lineales, y la capacidad de autoaprendizaje rápido son las ventajas potenciales que proporcionan las redes neuronales para el diseño e implantación de controladores inteligentes.

A modo de ejemplo, en la Figura 10a se muestra un elemento individual constitutivo de la red, o neurona simple. La neurona efectúa la suma de entradas x_i, ponderada por los pesos w_i e indica en la salida si el resultado arroja un balance positivo (y=+1) o negativo (y=-1). Pueden definirse otras neuronas similares utilizando una función de activación de salida similar a la función signo.

La Figura 10b muestra una red neuronal, basada en una interconexión masiva de neuronas simples y organizada en dos capas. Las redes neuronales multicapa permiten aproximar funciones no lineales arbitrarias entre señales de entrada y salida, con tal de incluir un número suficientemente grande de neuronas simples.

De esta forma, la red puede ser entrenada, por modificación de los coeficientes o pesos, de manera que sea capaz de reconocer situaciones o patrones especiales, tales como la aparición de un fallo, de una situación de alarma, etc., y así poder generar las acciones adecuadas.

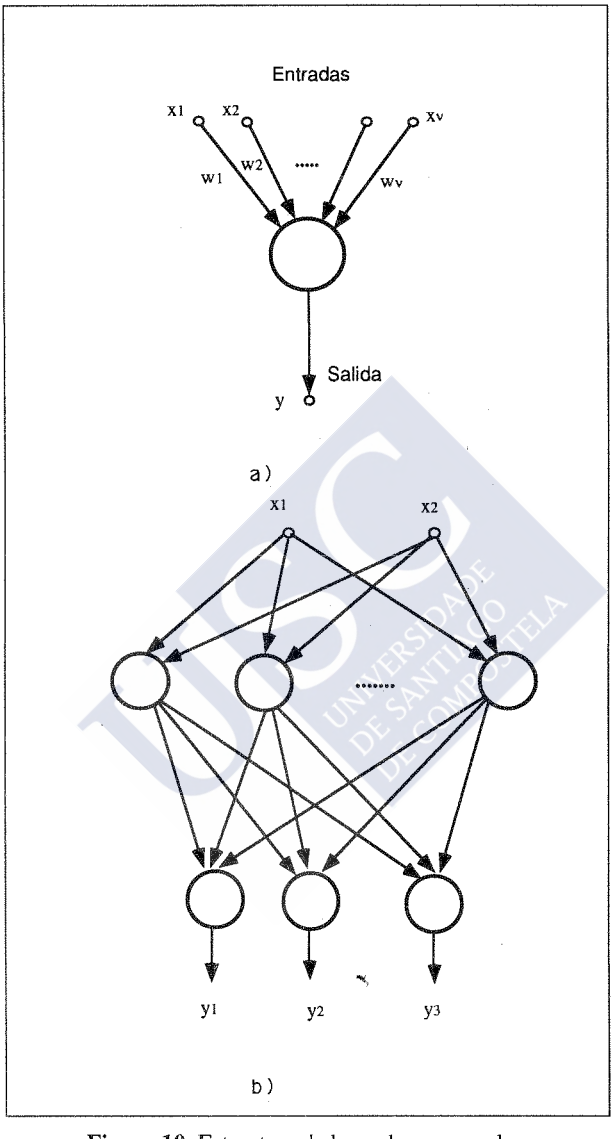


Figura 10. Estructura de las redes neuronales.

Por este motivo, la red multicapa puede sustituir a un controlador convencional, con las ventajas de la posibilidad de aprendizaje, en cualquiera de los esquemas usuales de identificación y control adaptativo (Ver figura 11). Una buena referencia técnica para estas cuestiones es el artículo de Narendra y Parthasarathy [1990].

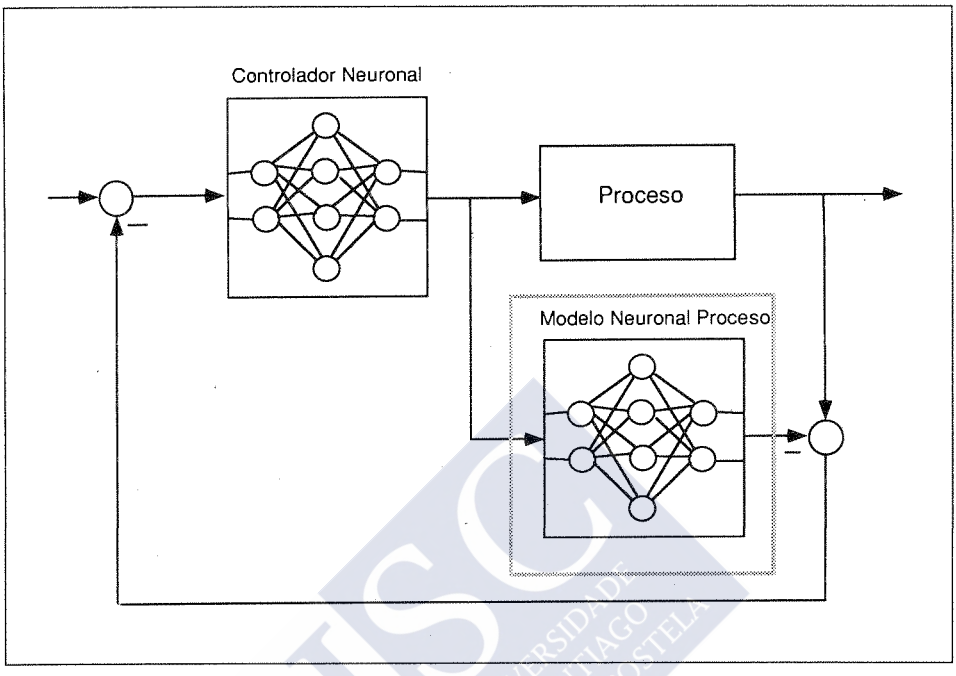


Figura 11. Estructura de un sistema de control borroso con redes multicapa.

7. APLICACIONES DE LA INTELIGENCIA ARTIFICIAL EN CONTROL Y SUPERVISIÓN DE PROCESOS

Sería imposible elaborar una recopilación más o menos exhaustiva de aplicaciones IA en control y supervisión de procesos, debido a la enorme difusión actual de estas técnicas. No obstante, en este apartado se analiza una muestra significativa de los desarrollos dentro de la comunidad científica y técnica en el área del control de procesos e inteligencia artificial. En las aplicaciones seleccionadas se discute, no sólo el interés de la aplicación en sí, sino especialmente los aspectos de IA: motivaciones en la elección de este método, ventajas alcanzadas, problemas derivados, etc. Hemos omitido deliberadamente las aplicaciones de robótica que son tratadas en otro trabajo (véase «Aplicación de la lógica borrosa en el control de procesos y robótica») y, por idéntico motivo, las aplicaciones en otros ámbitos de la ciencia, tales como las médicas (véase «Aplicación del control borroso en Medicina»).

7.1. Desarrollos recientes en control y supervisión borrosos

A continuación se muestran algunas aplicaciones industriales recientes de control y supervisión que utilizan técnicas de control borroso.

7.1.1. El coche borroso de Sugeno [Sugeno y Nishida, 1985]

Este coche sigue una trayectoria de referencia complicada y aparca por sí mismo en un recinto prefijado. El sistema borroso está formado por reglas con antecedentes borrosos y cuyos consecuentes toman la forma de una combinación lineal de las variables de interés (distancias, orientación del vehículo, etc.).

Los parámetros de las reglas se entrenan en una fase en la que un operador guía al coche en diferentes situaciones. Por tanto, el sistema posee la capacidad de aprendizaje.

7.1.2. Conducción automática de trenes [Yasunobu y Miyamoto, 1985]

Este sistema ha sido desarrollado e implantado en 1987, por Hitachi Ltd., para el tren de Sendai, en Japón. En este sistema, un evaluador borroso compara las diversas posibles acciones de control y decide cuál es la más apropiada.

Los criterios utilizados son: seguridad, comodidad, exactitud en la parada, tiempo de transporte, etc. Un subsistema realiza el control a velocidad constante y otro subsistema se encarga de las fases de arranque y parada.

7.1.3. Control borroso de grúas de contenedores [Yasunobu y Hasegawa, 1986]

Las variables de diseño son: seguridad, exactitud, y tiempo de arrastre del contenedor. Esta organizado en dos subsistemas: el movimiento y rotación de la grúa, y el ascenso y descenso del cable con la carga. Cada subsistema tiene a su vez un nivel de decisión y un nivel de activación.

La automatización ha sido probada con éxito en el puerto de Kytakyusu en Japón. El funcionamiento del sistema borroso es comparable al de un operador experimentado, y ambos superan en 30 cargas por hora al funcionamiento con un operador no experimentado.

7.2. Aplicaciones industriales de los sistemas basados en conocimiento

7.2.1. Sistema de gestión de alarmas en una planta de destilación [Jacob y Suslenchi, 1990]

Este sistema basado en reglas, denominado EXTASE, permite localizar las causas de las señales de alarma que ocurren en la superficie caliente de una columna de destilación de una refinería, sirviendo de ayuda al operador de planta.

El sistema se basa en dos importantes principios de diseño. El primero consiste en representar el conocimiento heterogéneo mediante funciones que describen causalmente el proceso. El segundo principio consiste en construir progresivamente una red de hipótesis que permiten un control de búsqueda local, almacenando la información dinámica dependiente del tiempo y formando una base para el proceso de explicación.

7.2.2. Monitorización y diagnóstico de un telescopio espacial [Perkins y Austin, 1990]

Esta aplicación, denominada *wheels*, consiste en la monitorización y diagnóstico de anomalías de los dispositivos de control de altitud primario del telescopio espacial Hubble, a partir de datos telemétricos.

El sistema inteligente está basado en el sistema LES de la Lockheed, habiendo sido modificado con objeto de incluir la posibilidad de emplear razonamiento temporal (véase «Razonamiento temporal borroso»).

7.2.3. Sistema integrado de supervisión de sistemas de control [Marcos y col., 1992]

Este sistema consiste en un conjunto de paquetes de programas desarrollados en Ada para la supervisión de sistemas de control. Este sistema realiza el control del proceso mediante controladores clásicos e inteligentes, empleando un sistema de monitorización en línea para detectar situaciones anómalas.

Este paquete de programas ha sido empleado para el desarrollo de un sistema inteligente que compensa las zonas muertas en un servomecanismo de posición. El supervisor realiza, en este caso, la vigilancia y detección de situaciones no deseadas del comportamiento del sistema en bucle cerrado, tales como: errores en régimen permanente, ciclos límites y zonas de funcionamiento no deseadas.

7.3. Aplicaciones recientes de redes neuronales en control y supervisión de procesos

7.3.1. Aplicación de redes neuronales a control óptimo de vuelo [McKelvey, 1992]

Se trata de un problema de control óptimo de vuelo de aeronaves: el problema de ascenso (o descenso) de altura, optimizando el tiempo transcurrido o el consumo de energía. El modelo simplificado de la dinámica de la aeronave no es lineal.

Es importante resaltar que no existe solución analítica al problema. La ausencia de una solución analítica directa sugiere usar una red neuronal. El controlador se realiza mediante una red neuronal de 4 entradas, 2 salidas y 2 capas, de 10 y 5 neuronas. La red se entrena por el método de «*propagación hacia atrás del gradiente*», a partir de ejemplos constituidos por leyes de control óptima generadas numéricamente por simulación. Los resultados se consideran satisfactorios.

7.3.2. Aplicación de redes neuronales a la medida no intrusiva de caudal en tuberías multifase [Shepperd y Russell, 1992]

Se trata de la medida de caudales extraídos de estaciones petrolíferas. El producto contiene una fase de gas natural, dos fases líquidas (petróleo y agua) y partículas sólidas. Existen distintos tipos de régimen de flujo (estratificado, distribuido). Los requisitos legales-fiscales exigen una precisión del 1% en caudales. Las medidas se hacen con sensores no intrusivos (conductividad, rayos X, ultrasónicos).

El procesado de señal convencional presenta problemas de dificultad de análisis de la medidas de los diferentes sensores. La utilización de un sistema experto tiene la desventaja de que las posibles reglas son difíciles de especificar.

Por estos motivos, se ha decidido la utilización de redes neuronales. Las redes son ventajosas en situaciones en las que existe una abundancia de ejemplos (para entrenamiento). Otras ventajas que presentaron fueron su capacidad de aprendizaje y generalización, su capacidad de tratar problemas no lineales en presencia de ruido, y su rapidez y eficiencia computacional.

7.3.3. Procesamiento con redes neuronales para el mantenimiento de posiciones de naves en el mar [Xiang-Gu y col., 1992]

El problema es el del mantenimiento de naves o plataformas flotantes en una posición fija en el mar, evitando las perturbaciones del viento, corrientes y oleaje. En este tipo de situaciones es aconsejable la utilización de control en prealimentación (*feedforward*), que se anticipe a la perturbación, corrigiéndola. Para ello es necesario modelar las perturbaciones. En este proyecto se usan redes neuronales para modelar el efecto no lineal del oleaje sobre la posición de la nave. La red funciona satisfactoriamente y ha sido entrenada para modelar la fuerza del oleaje.

8. CONCLUSIONES

En este trabajo se ha intentado mostrar una parte de lo que son los desarrollos actuales en los campos de IA e Ingeniería de Control. En primer lugar se han esbozado los objetivos de la automatización de procesos industriales, y se han establecido los conceptos de control y supervisión. A continuación se han considerado los esquemas generales de SSEE aplicados a control y supervisión, y se han clasificado en: Sistemas basados en reglas y lógica borrosa; Sistemas basados en conocimiento; y Sistemas de inteligencia distribuida y redes neuronales.

Para cada uno de estos tres enfoques, se ha descrito el esquema básico de funcionamiento y se han resaltado aquellos aspectos principales que deben ser tenidos en cuenta en el diseño e implantación en tiempo real, terminando con una selección de algunas de las aplicaciones industriales recientes más significativas.

Estas técnicas de control basadas en conjuntos borrosos, o en otros métodos de IA, se están expandiendo actualmente a un ritmo creciente. Como resultado, el ingeniero de control se encuentra con una variedad de productos comerciales para resolver un problema concreto de automatización. La elección del sistema (control borroso, neuronal, o basado en conocimiento) depende de los objetivos y del proceso a automatizar. En este trabajo se intenta recoger algunos aspectos clave de los distintos métodos, de modo que puedan servir de orientación en el problema de selección de la herramienta adecuada.

REFERENCIAS

Arzén, K.E., «An Architecture for Expert System Based Feedback Control», *Automatica*, Vol. 25, N. 6, (1989), 813-827.

Arzén, K.E., «A Survey of Commercial Real-time Expert System Environments», *Proc. AI in Real-time Control*, Delf, The Netherlands, (1992), 611-618.

Aström, K.J. y Anton, J.J., «Expert Control», *Proc. 9th IFAC World Congress*, Vol. VI, Budapest, Hungary, (1984), 471-486.

Aström, K.J., Anton, J.J. y Arzén, K.E., «Expert Control», *Automatica*, Vol. 22, (1986), 277-286.

Bavarian, B., «Introduction to Neural Networks for Intelligent Control», *IEEE Control Systems Magazine*, Abril, (1988), 3-7.

Gertler, J. y Chang, H.S., «An instability Indicator for Expert Control», *IEEE Control Systems Magazine*, Agosto, (1986), pág. 1417.

Goedecke, W., «Fault Detection in a Tubular Heat Exchanger Based on Modelling and Parameter Estimation», *IFAC Identification and System Parameter Estimation*, York, U.K, (1985), 1095-1100.

Isermann, R., «Process Fault Detection Based on Modeling and Estimation Methods- A Survey», *Automatica*, Vol. 20, N. 4, (1984), 387-404.

Jacob, F. y Suslenschi, P., «Situation Assessment for Process Control», *IEEE Expert*, Abril, (1990), 49-58.

Laffey, T.J., Cox, P.A., Schmidt, J.L., Kao, S.M. y Read, J.Y., «Real-Time Knowledge-Based Systems», *AI Magazine*, Vol. 9, N. 1, (1988), 27-45.

Lee, C.C., «Fuzzy Logic in Control Systems: Fuzzy Logic Controller», *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 20, N. 2, (1990), 404-432.

Marcos, M., Artaza, F. y Iriondo, N., «An Integrated System for Supervisory Process Control Using Ada», *Proc. SICICA'92*, Málaga, (1992), 801-805.

McKelvey, T., «Neural Networks Applied to Optimal Flight Control», *Proc. IFAC/IFIP/IFORS Conference on Artificial Intelligence in Real-Time Control*, Delft, Netherlands, Junio, (1992), 43-48.

Narendra, K.S. y Parthasarathy, K., «Identification and Control of Dynamical Systems using Neural Networks», *IEEE Transactions on Neural Networks*, Vol 1, N. 1, Marzo, (1990), 4-26.

Perkins, W.A. y Austin, A., «Adding Temporal Reasoning to Expert-System-Building Environments», *IEEE Expert*, Febrero, (1990), 23-30.

Rault, A., Jaume, D. y Verg, M., «Industrial Processes Fault Detection and Localization», *Proc. 9 th. IFAC World Congress*, Vol III, Budapest, Hungary, (1984), 190-195.

Sanz, R., *Diseño de controladores autoajustables: Nivel Supervisor experto*, Tesis Doctoral, Univ. Santiago, 1987.

Sheppard, C.P. y Russell, D., «The Application of Artificial Neural Networks to Non-Intrusive Multi-Phase Metering», *Proc. IFAC/IFIP/IFORS Conference on Artificial Intelligence in Real-Time Control*, Delft, Netherlands, Junio, (1992), 227-232.

Sugeno, M. y Nishida, M., «Fuzzy Control of Model Car», *Fuzzy Sets and Systems*, Vol. 19, (1985), 103-113.

Yasunobu, S. y Hasegawa, T., «Evaluation of an Automatic Container Crane Operation System based on Predictive Fuzzy Control», *Control Theory and Advanced Technology*, Vol. 2, N. 3, (1986), 419-432.

Yasunobu, S. y Miyamoto, S., «Automatic Train Operation by Predictive Fuzzy Control», *Industrial Applications of Fuzzy Control*, North Holland, Amsterdam, (1985), 1-18.

Widrow, B. y Lehr, M.A., «30 Years of Adaptive Neural Networks: Perceptron, Madaline, Backpropagation», *Proc. of the IEEE*, Vol. 78, N. 9, (1990).

Xiang-Gu, M., Yoh-Hau, P., y Yip, P.P.C., «Neural Network Computing for Maintaining Positions of Vessels at Sea», *Proc. IFAC/IFIP/IFORS Conference on Artificial Intelligence in Real-Time Control*, Delft, Netherlands, Junio, (1992), 397-402.

FABRICACIÓN AUTOMATIZADA DE AMPLIFICADORES ELECTRÓNICOS EN TELEVÉS UTILIZANDO LÓGICA BORROSA

Ricardo García Rosa y Pedro F. Zuliani
Instituto de Automática Industrial, CSIC, Madrid

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1. INTRODUCCIÓN

La lógica borrosa no nació como una disciplina teórica, sino como una herramienta práctica que intentaba describir, para posteriormente simular, el comportamiento de las personas, cuando éstas realizan el control manual de determinados procesos. Los procesos que desde un principio se han puesto como ejemplo van desde los aparentemente sencillos de conducir un automóvil o hacer aterrizar un aeroplano hasta el casi imposible, como muestra la experiencia cotidiana, de mezclar con fortuna el agua de la ducha o baño sin escaldarse.

La comunidad científica dedicada al control de procesos no acogió esta nueva herramienta con los brazos abiertos precisamente, sino que puso de manifiesto constantemente las aparentes –quizás no tan aparentes– inconsistencias en la formulación de la teoría de control de procesos por medio de métodos basados en la lógica borrosa, dándose la paradoja de que las herramientas desarrolladas para resolver problemas prácticos eran rechazadas por estar excesivamente alejadas de la realidad.

La principal objeción que se le hace es la de estar al servicio de charlatanes que se dicen capaces de resolver todos los problemas de control aunque éstos estén

mal definidos, haciendo de ella la panacea universal o remedio para todos los males. Esta pretensión por parte de los científicos –o creyentes– de la lógica borrosa de conocer procedimientos para controlar procesos de cuyo funcionamiento sólo se tiene un conocimiento aproximado choca con la necesidad de tener un buen modelo matemático de ellos por parte de los que utilizan métodos «clásicos» de control.

Creemos que los dos puntos de vista son correctos; es decir, que no puede controlarse ningún proceso del que no se posea un modelo de su comportamiento y que la teoría de control que utiliza procedimientos basados en la lógica borrosa es capaz de controlar procesos mal definidos. La explicación reside en que lo que utiliza es un modelo distinto del funcionamiento del sistema a controlar, como puede ser un modelo del comportamiento de una persona que es capaz de hacerlo manualmente.

El caso que nos ocupa es un ejemplo claro de sistema físico del que no se posee un modelo matemático, pero sí existen personas que son capaces de controlarlo; lo describiremos a continuación.

2. EL PROBLEMA DEL AJUSTE DE EQUIPOS ELECTRÓNICOS

Los amplificadores monocanales de UHF se utilizan en instalaciones domésticas para obtener buenas señales de cada una de las emisoras, separándolas de las de las demás. Estos equipos se fabrican como amplificadores de gran ancho de banda provistos de dos filtros que ajustados convenientemente se particularizan para transformarlos en amplificadores monocanales; este proceso de singularizar cada uno de los equipos lo realizan unos operadores provistos de dos destornilladores con los que modifican las posiciones de unos elementos móviles, observan las figuras que aparecen en la pantalla de un analizador de ondas y repiten su actuación hasta que las figuras obtenidas sean las correctas, evidentemente el tiempo utilizado en este proceso depende de la experiencia de los operadores.

El costo del tiempo dedicado al ajuste es una parte alta del precio del equipo fabricado y Televés[1] consideró interesante automatizar este proceso de ajuste dado que el número que se fabrica de estos equipos es elevado. Como nosotros no necesitamos aquí concretar este número, diremos de una manera difusa que es «aproximadamente bastante».

[1] Televés es una empresa de fabricación e investigación electrónica.

Las características eléctricas de los equipos fabricados dependen de un conjunto de parámetros que se corresponden con las posiciones de los elementos móviles a los que nos hemos referido anteriormente, pero esta correspondencia no está definida matemáticamente, sino que depende de la dispersión de los valores eléctricos debida al proceso de fabricación de los componentes; por esta razón no puede automatizarse el ajuste mediante un procedimiento «clásico», en su lugar se ha procedido a automatizar el procedimiento que utilizan los operadores humanos.

Podemos describir el sistema creado para automatizar el ajuste desde dos puntos de vista de distinta naturaleza: el físico y el funcional (más divertido sería llamarles material y espiritual respectivamente).

3. COMPONENTES FÍSICOS

Los componentes físicos del sistema son: un banco auxiliar de sujeción de los equipos, del que no hablaremos más aquí, dos manos robotizadas para actuar sobre los elementos de ajuste, un sistema de percepción para analizar las características de los equipos fabricados y un computador central en el que reside toda la programación.

La parte electromecánica está formada por dos «manos» con un total de ocho dedos, siendo cada uno de los dedos capaz de manejar un destornillador (aquí radica una de las ventajas del sistema creado, ya que un humano sólo puede manejar dos destornilladores simultáneamente); cada cuatro dedos están gobernados por un controlador de ejes de robot basado en un computador.

El sistema de percepción coincide con el mismo equipo que manejan los operadores, es un analizador de ondas programable que se utiliza para alimentar a los amplificadores en estudio, que recibe su salida y que presenta el resultado en forma de dos gráficas sobre una pantalla; el conjunto de puntos que constituyen estas gráficas es el que deberá ser suministrado al computador central.

El computador central está encargado de la recepción de todos los datos, de su tratamiento y de proporcionar las órdenes a todos los componentes del sistema.

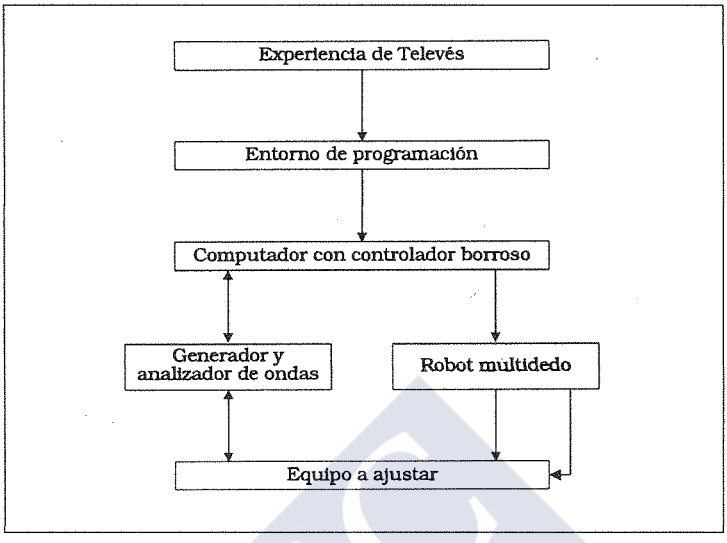


Figura 1. Diagrama de bloques.

4. DESCRIPCIÓN FUNCIONAL

En el computador central se ha realizado un programa que está encargado de indicar al analizador de ondas cuándo y cómo debe comenzar su función y, posteriormente, de recibir los datos por él proporcionados. Estos datos directos son analizados por un sistema de interpretación y transformados en un vector de características que representa el estado del equipo y que constituye la entrada a un sistema experto basado en lógica borrosa que almacena el conocimiento de los expertos humanos; la salida del sistema experto es el resultado de un proceso de concreción que proporciona valores numéricos que son los que se suministran al sistema de control de los motores de los dedos robotizados. Veamos a continuación los sistemas de interpretación y de control borroso.

5. SISTEMA DE INTERPRETACIÓN

La forma de la onda que proporciona el analizador, que representa los valores de la ganancia en función de la frecuencia, se asemeja a la de la figura 2. De esta figura destacaremos las siguientes características:

Los canales tienen un ancho de banda, su banda de paso, de 8 MHz y por razones prácticas se considera que la separación entre canales consecutivos es de 16 MHz, en lugar de 8 MHz; cada canal queda especificado por dos frecuencias que se denominan marcas inferior y superior respectivamente.

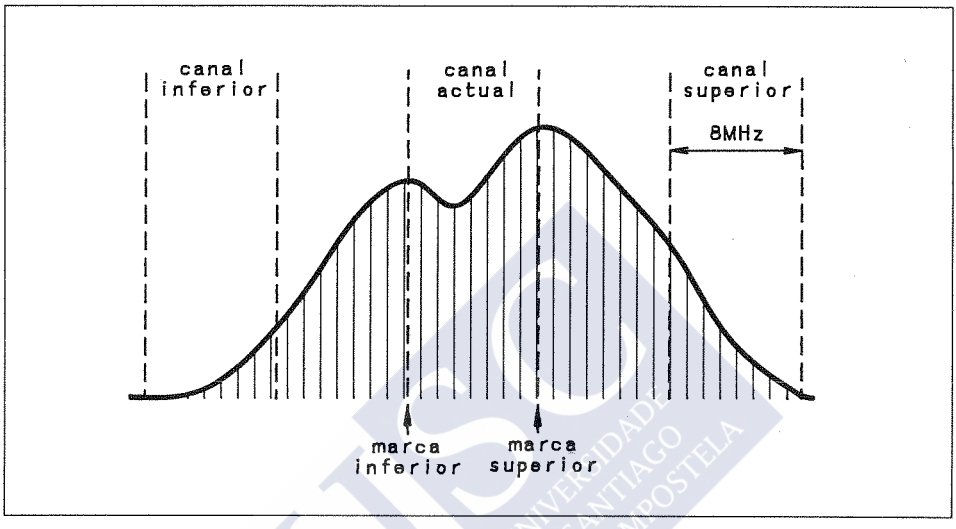


Figura 2

La forma de la onda dentro del canal que esté siendo ajustado puede definirse como formada por dos picos con un valle entre ellos; el análisis del conjunto de puntos que constituyen una curva los transforma en un vector de características cuyos componentes tienen los siguientes nombres y significados:

frecuencia: distancia del primer pico (el de la izquierda) a la marca inferior del canal actual (figura 3).

rizado: diferencia entre las máxima y mínima ganancias dentro de la banda de paso.

ganancia: valor medio de las ganancias dentro de la banda de paso (figura 4).

ancho: (o ancho de la banda) diferencia entre las frecuencias en los que se producen los dos picos.

simetría: diferencia entre las ganancias de los dos picos (figura 5).

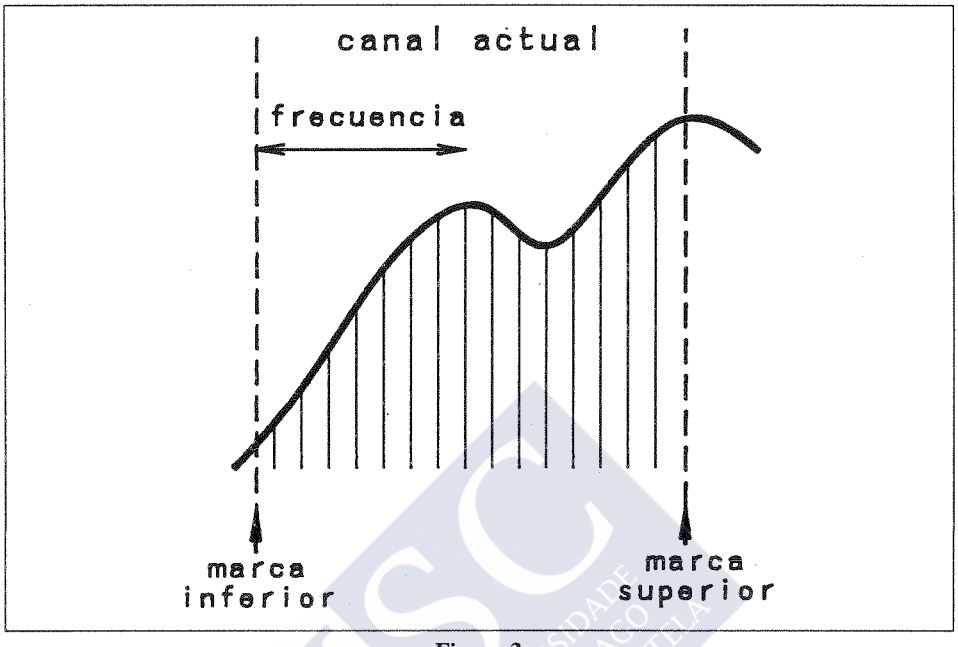


Figura 3

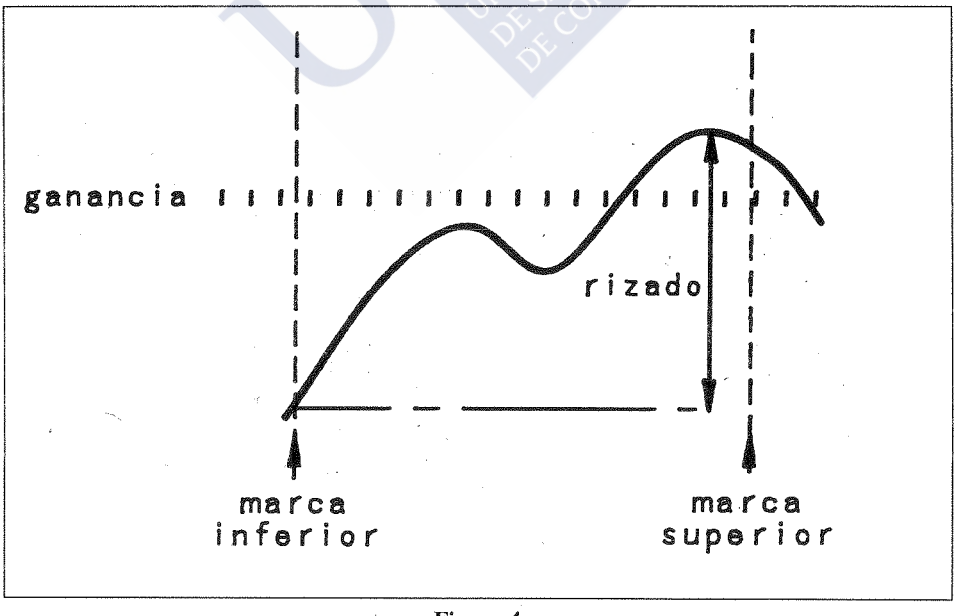


Figura 4

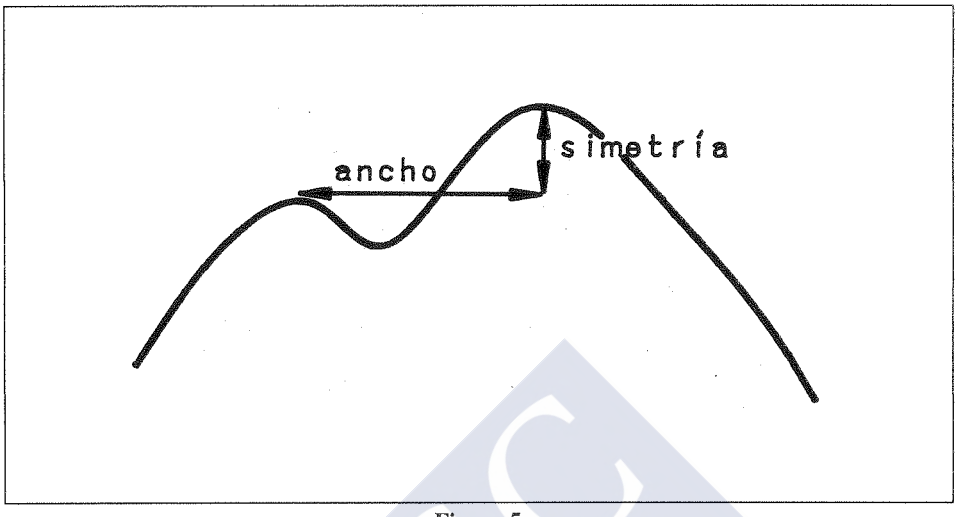


Figura 5

Estos nombres de componentes del vector de características junto con los calificadores a ellos asociados son los que utiliza el controlador lingüístico en la expresión de sus reglas.

6. LAS REGLAS DE CONTROL Y SU GENERACIÓN

Se ha construido un sistema de generación automática de controladores borrosos, tal que, suministrándole una descripción de la estrategia de control, él proporciona el algoritmo que lo realiza.

Consta de dos partes: un entorno interactivo de descripción de los calificadores lingüísticos asociados a las variables físicas y de las reglas que forman la estrategia y un controlador borroso universal que utiliza la salida del entorno para adaptarse a cada problema particular.

En el sistema de control del proceso de ajuste de Televés se han de definir los nombres de las variables de entrada, que son los indicados anteriormente: frecuencia, ganancia, rizado, ancho y simetría; los de las variables borrosas o calificadores lingüísticos asociados a cada una de ellas; y la descripción de cada una de estas variables, es decir, el tipo de función que la representa y los parámetros que la describen.

De la misma manera se han de describir las variables de salida y en este caso, en un exceso de imaginación, se les ha denominado E1, E2, E3, E4, S1, S2, S3 y S4, refiriéndose a los elementos de ajuste de los filtros de entrada y salida.

Las reglas que describen la estrategia de ajuste representan la experiencia de los operadores de Televés, relacionando una expresión borrosa de las variables de entrada con otra expresión borrosa que describe las acciones que la experiencia del operador dice que son las apropiadas. Una parte de la descripción del problema es la siguiente:

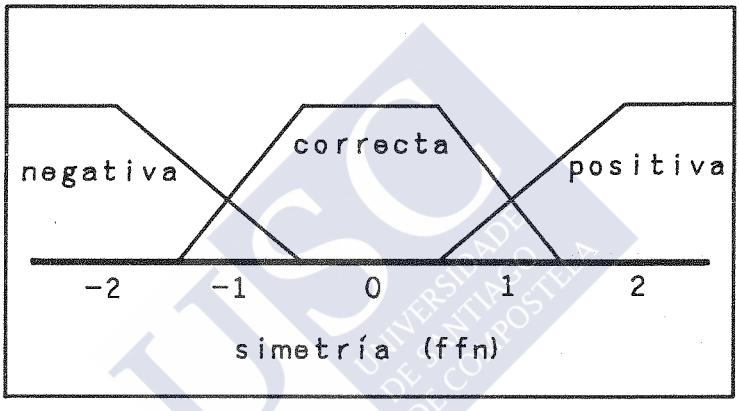


Figura 6

Variable física: simetría (ffn)				
negativa	–39	–39	–2	–0,5
correcta	–1,5	–0,3	0,3	1,5
positiva	0,5	2	39	39

Variable física: E4 (ffn)				
medio–izq	–45	–45	–15	–5
poco–izq	–15	–10	–5	0
cero	–3	0	0	3
poco–der	0	5	10	15
medio–der	5	15	45	45

Regla nº 0:

Si rizado (correcto) y frecuencia (alta) Entonces E1 (poc–der) y E2 (poco–der)

Regla nº 1

Si frecuencia (alta) y simetría (positiva) Entonces E1 (poco–der)

El resultado de una sesión de trabajo con el entorno de programación es un fichero que, de acuerdo con una gramática formal bien definida, describe la estrategia a utilizar.

El controlador borroso universal consta también de dos partes: un traductor y un motor de inferencias o proceso de concreción.

La descripción del proceso de ajuste realizada por el entorno de programación es analizada por un intérprete cuya función es traducirla a otro lenguaje cuya gramática permite describir una base de datos apropiada para el motor de inferencias.

El motor de inferencias proporciona una semántica a la base de datos que describe el ajuste y su función es mezclar la aportación de cada una de las reglas al valor de una variable de salida en un momento determinado. No se ha realizado un único método de concreción sino que el motor de inferencias utiliza, si puede, el método de la media ponderada o el del centro de gravedad, si el otro no es aplicable.

7. ESTRATEGIA GENERAL

La realización del control del robot de ajuste se ha dividido en cuatro fases, una de preparación y tres de estrategias borrosas. La fase de preparación es muy importante, ya que tiene como función la obtención de un estado en el que sea aplicable lo apuntado para las estrategias borrosas, es decir, la existencia de una función de ganancia definida por dos picos y un valle entre ellos; para conseguirlo, y dado que el estado inicial suele estar muy alejado de éste, se ha construido una estrategia basada en la derivación numérica de la función de ganancia y su posterior interpretación.

Las tres fases borrosas tienen una apariencia semejante, estando formadas por dieciséis reglas cada una de ellas, aunque la semántica asociada a los calificadores borrosos mediante las funciones de pertenencia que los describen es distinta. También se corresponden con diferentes elementos de ajuste; primeramente se realiza el ajuste del filtro de salida, a continuación el del filtro de entrada y finalmente existe una fase de ajuste global.

8. OTRAS CONSIDERACIONES

Quiere destacarse que la realización de un sistema robotizado como el aquí descrito ha exigido no sólo la construcción de la estrategia borrosa y del robot

multidedo, sino que también ha sido necesario que Televés volviese a definir los productos a ajustar, de manera que fuesen accesibles todos los tornillos al robot y también que construyese un sistema de manipulación auxiliar y de conmutación automática de señales para que pudiesen realizarse las tres fases borrosas sin intervención humana. Se destaca todo esto para mostrar como los procesos de automatización llevan asociados otros muchos que exigen modificar la tecnología.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

APLICACIÓN DEL CONTROL BORROSO EN MEDICINA

Carlos Hernández y Andrés Carollo
Departamento de Física Aplicada. Universidad de Santiago de Compostela

1. INTRODUCCIÓN

La lógica borrosa y los controladores borrosos fueron originalmente propuestos por Zadeh [1973] y Mamdani y Assilan [1975] como un medio de implementar la experiencia humana y tratar con la incertidumbre. Estos conceptos fueron inicialmente aplicados a procesos poco definidos. Tales sistemas son normalmente controlados por operadores humanos que consiguen buenos resultados basándose en información imprecisa. El origen de esta imprecisión en los sistemas de control pueden ser retardos entre la aplicación de la señal de control y la respuesta del sistema, no-linealidades en el sistema, degradación de los sensores, etc.

El control borroso fue primeramente aplicado por Mamdani y Assilan [1975]. En muchos casos el control mediante controladores borrosos fue superior al control estrictamente manual. El éxito en estas aplicaciones ha enseñado que el control borroso puede tratar con sistemas poco definidos.

Zadeh introdujo la lógica borrosa para describir sistemas que son demasiado complejos o poco definidos para admitir un análisis matemático preciso. Su principal característica es el uso de variables lingüísticas más bien que variables numéricas y la caracterización de las relaciones entre variables por sentencias

condicionales borrosas. Mamdani y col. extendieron este concepto para tratar el control de procesos industriales para los cuales no era posible conseguir medidas precisas de las variables de estado.

A diferencia de los controladores clásicos, el control borroso no utiliza modelos matemáticos del proceso a controlar. La idea básica es imitar el razonamiento y la experiencia de un operador humano al controlar un proceso. Este conocimiento se implementa en un conjunto de reglas lingüísticas que describen la estrategia de control. La principal ventaja de esta aproximación es la posibilidad de implementar estrategias «heurísticas», basadas en la experiencia e intuición del operador humano.

Los operadores humanos utilizan términos cualitativos más bien que cuantitativos, en contraposición con los controladores convencionales, para describir los estados de un proceso. Esta manera cualitativa de razonar se implementa por medio de reglas en los controladores borrosos. Las etiquetas de estas reglas son conjuntos borrosos, de esta forma se implementa la imprecisión que conlleva el asignar datos numéricos con etiquetas lingüísticas. Por ejemplo, un valor numérico de 7 puede pertenecer a la etiqueta «Alto» con un valor de 0.8 y a la etiqueta «Medio» con un valor de 0.3.

2. COMPARACIÓN DEL CONTROL BORROSO CON CONTROLADORES ANALÍTICOS

Para hacer un controlador analítico, un investigador en el diseño de modelos debe hacer un modelo matemático del proceso, de forma que una función de transferencia o un modelo en variables de estado describa el comportamiento de la planta. Con esta información y las especificaciones de comportamiento, un especialista en control diseñaría un controlador. En esta aproximación, el trabajo está claramente definido y ni el diseñador de la planta ni el especialista en control necesitan ser un experto en el campo del otro.

En el diseño de controladores borrosos la situación es bastante diferente. No se utilizan modelos matemáticos del sistema, sino que el controlador especifica la acción de control por un conjunto de reglas de forma que al ocurrir determinadas acciones se produzca una determinada salida de control. Esta aproximación implica que el diseñador conozca a fondo la operación de la planta, para después implementar la estrategia de control en forma de reglas.

La información implementada en las reglas puede ser obtenida tanto del conocimiento del funcionamiento de la planta como también del control manual del operador humano. Este último conocimiento puede ser conseguido mediante entrevistas, cuestionarios o grabaciones en tiempo real de las acciones de control del operador.

En los controladores analíticos se utiliza un criterio de comportamiento para diseñar el controlador. Sin embargo, no hay un modo directo para utilizar un criterio de comportamiento para el control borroso. El controlador borroso tiene que ser construido y probado para determinar posteriormente su comportamiento. No hay un método claro para determinar cuantas etiquetas borrosas utilizar, ni que función de pertenencia utilizar. Por otra parte, la calibración del controlador borroso tiene que ser por prueba y error; es necesario repetir pruebas en lazo cerrado y con la información resultante sintonizar las reglas de control y los grados de pertenencia de los conjuntos borrosos.

3. DESCRIPCIÓN DE UN CONTROLADOR BORROSO

Un controlador borroso consta básicamente de 4 partes: una interfase de borrosificación, una base de conocimiento, un proceso de inferencia basado en lógica borrosa, y una interfase de desborrosificación.

1) La interfase de borrosificación tiene las siguientes funciones:

a) Mide los valores de las variables de entrada.

b) Hace un escalado entre el rango de los valores de las variables de entrada y sus correspondientes universos de discurso.

c) Borrosifica los valores convirtiéndolos en etiquetas lingüísticas.

2) La base de conocimiento contiene el conocimiento del operador y los objetivos de control. Consta de 2 partes:

a) La base de datos, que contiene las definiciones de los conjuntos borrosos.

b) La base de reglas que caracteriza las acciones de control.

3) El proceso de inferencia que basado en lógica borrosa simula la toma de decisiones del operador humano.

4) La interfase desborrosificadora realiza las siguientes funciones:

a) Un escalado que pasa de los valores del universo de discurso a los valores reales.

b) Un proceso de desborrosificación para obtener un valor numérico de la acción de control a partir del conjunto borroso final.

El controlador borroso consta de un conjunto de reglas con etiquetas borrosas que describe las estrategia de control de un operador. Las reglas son de la forma:

IF (conjunto de condiciones) THEN (conjunto de consecuencias).

En el control borroso los antecedentes de las reglas suelen ser el «Error» y el «Cambio en el Error». El Error es la diferencia entre el valor actual de la variable a controlar y el punto de referencia. El Cambio en el Error es la diferencia entre el error actual y el error en el anterior instante de muestreo.

Las reglas son codificadas de la forma:

IF ERROR= E1 y CAMBIO EN EL ERROR= CE1 then SALIDA= U1
IF ERROR= E2 y CAMBIO EN EL ERROR= CE2 then SALIDA= U2
.....

IF ERROR= En y CAMBIO EN EL ERROR= CEn then SALIDA= Un
donde Ei, CEi y Ui son conjuntos borrosos.

El conocimiento se puede representar por una relación total:

$$R= R1 \cup R2 \cup ... \cup Rn = \cup (Ei \times Cei \times Ui)$$

La unión de los Ri suele ser modelada por el operador «máximo». El producto cartesiano es representado por una t-norma, que suele ser el operador mínimo.

Para cualquier entrada, la salida se obtiene utilizando la regla composicional de inferencia:

$$U=(E \times CE) \bullet R$$

El conjunto borroso resultante tiene que desborrosificarse para obtener un valor numérico de control. Usualmente, para este proceso de desborrosificación se utiliza el centro de gravedad.

Las ventajas de la utilización de controladores borrosos se puede resumir en:

1) Los controladores borrosos son bastantes insensibles al ruido de alta frecuencia debido al gran número de reglas y al efecto de promedio que realizan.
2) Son comparables en precisión a los controladores analíticos.
3) Consiguen un buen control sobre un amplio rango de condiciones iniciales.
4) Son más tolerantes a fallos en los sensores.

Las principales desventajas de los controladores borrosos son que no existe una metodología completamente desarrollada para su implementación, necesitan de expertos sobre los cuales extraer la estrategia de control, tienen un tiempo superior de respuesta al de los controladores analíticos y son más difíciles de mantener y modificar.

El control de variables fisiológicas es un campo propicio para el uso de controladores borrosos. Muchos procesos fisiológicos son poco conocidos y no se ha conseguido desarrollar modelos matemáticos precisos. Por otra parte, muchas variables son en si mismas «borrosas», como el dolor postoperatorio, el grado de inconsciencia en anestesia, etc.

Los controladores borrosos se han aplicado en muchos campos de la medicina; se han utilizado para controlar el nivel de glucosa en plasma en pacientes diabéticos, para controlar el ritmo cardíaco [Sugima y col., 1991], para controlar el nivel de relajación muscular [Linkens y Mahkouf, 1988], la presión arterial [Isaka y Sebold, 1989], así como el dolor postoperatorio [Carollo y col., 1991].

Los procedimientos para extraer el conocimiento del experto son muy diversos, y algunos de los más comunes son:

- Observaciones visuales del modo de operar del médico.
- Análisis de datos digitalizados de las variables de interés.
- Utilización de cuestionarios sobre la estrategia de actuación del médico.
- Entrevistas y discusión de los resultados anteriores con el experto.

En resumen, se trata de identificar las variables utilizadas por el médico, los valores lingüísticos empleados y las estrategias de acción que relacionan los valores lingüísticos con acciones concretas de control. Posteriormente se utiliza un método de prueba-error para ajustar los diferentes parámetros del controlador borroso.

Para dar una visión general describiremos 3 controladores borrosos: un controlador borroso «corriente», un controlador implementado en una «*shell*» FPLOPS y un controlador borroso adaptativo.

4. CONTROL DE LA RELAJACIÓN MUSCULAR EN ANESTESIA

Los relajantes musculares se utilizan en anestesia como bloqueantes de la actividad muscular. Es comprensible que en plena cirugía, movimientos o contracciones musculares podrían ser desastrosos, sobre todo en operaciones de cerebro o de ojos. Por otra parte, en operaciones abdominales los relajantes musculares permiten al cirujano acceder a estructuras profundas que de otra manera estarían ocultas debido a la tensión del paquete muscular abdominal.

El nivel de relajación muscular se mide amplificando e integrando el EMG, y según esa información se inyecta más o menos relajantes musculares tales como vecuronium o pancuronium.

Los principales problemas al controlar este proceso son la gran variabilidad entre distintos pacientes y la no-linealidad que caracteriza la respuesta muscular a la droga.

Linkens y Mahkouf [1988] han desarrollado un controlador borroso para la administración de Pancuronium. Las reglas son de la forma:

If «Error» es grande then if «Cambio en el error» es grande then «Salida del controlador» es grande.

El autor registró las acciones de control de un operador humano en base a las observaciones del «Error» y el «Cambio en el Error», con el objetivo era obtener un modelo borroso del operador humano. Con esta información y la obtenida de simulaciones de modelos matemáticos del bloqueo muscular, se obtuvieron las etiquetas lingüísticas, sus grados de pertenencia y las reglas de control. El intervalo de muestreo se eligió en un minuto y el grado de parálisis deseado en el 80%.

Se simuló el controlador borroso varias veces, y a la vista de los resultados se fue resintonizando el controlador hasta obtener un buen comportamiento. Se

tuvo especial cuidado en que el proceso no llegase a la zona de saturación que se encuentra en el 100% del bloqueo muscular.

El conjunto inicial de reglas sacadas de la experiencia usa 8 etiquetas lingüísticas para el Error y 7 etiquetas para el Cambio en el Error y se implementaron 15 reglas borrosas mostradas en la tabla I:

TABLA I

CAMBIO EN	ERROR							
EL ERROR	NG	NM	NB	NC	PC	PB	PM	PG
NG								
NM								
NB			PB	PB	PB	PG	PG	PG
C			PB	PB	PB	PM	PB	PG
PB			C		PB			
PM								
PG								PG

Después de simularlo y ver que tenía una lenta respuesta y una ligera desestabilización, se cambio la tabla de control y se obtuvo la tabla II.

TABLA II

CAMBIO EN	ERROR						
EL ERROR	NG	NM	NB	C	PB	PM	PG
NG	PG	PG		PG	PG		
NM	PG		C	PG			
NB	PG		PB	PB			
C	PG	C	C	C	PG	PG	PG
PB		PB	PB	PM		PM	PG
PM	C	C					
PG	PG	PB	PG			PG	

El comportamiento del controlador se presenta en la figura 1 [Linkens y Mahkouf, 1988].

A. Punto de referencia.

B. Salida.

C. Señal de control

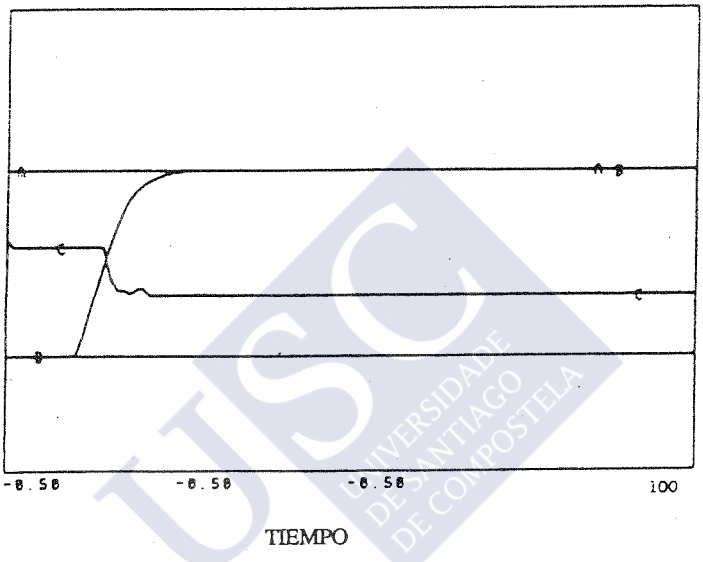


Figura 1. Resultado del control borroso. A: punto de referencia; B: salida; C: Señal de control.

Las simulaciones demuestran que el controlador es insensible a variaciones en el retardo y a variaciones en las constantes de tiempo, aunque es sensible a variaciones en la ganancia en lazo abierto del modelo simulado. En conjunto el controlador demostró un comportamiento robusto.

5. CONTROL DE LA PRESIÓN ARTERIAL

En ciertos procedimientos quirúrgicos, así como en el postoperatorio, es necesario mantener la presión arterial dentro de ciertos límites. En el quirófano, esto se realiza mediante control manual, inyectándole al paciente nitroprusiato sódico u otro vasodilatador. Sin embargo, en el periodo intraoperatorio, el anestesista tiene

que controlar la inyección de otras drogas, por lo que el control automático de la presión arterial sería de gran utilidad. El principal objetivo es evitar periodos de hipertensión o tensión alta, que puede producir complicaciones, sobre todo en pacientes cardíacos.

La estrategia a seguir es implementar mediante reglas con etiquetas borrosas la estrategia de control del anestesista. El principal problema en el control de este proceso es la variabilidad de la respuesta a la droga de distintos pacientes.

Se han propuesto diversos controladores borrosos para el control de la presión arterial, desde controladores borrosos»corrientes» (con reglas cuyos antecedentes son Error y Cambio en el Error), hasta controladores borrosos adaptativos [Isaka y Sebold, 1989] y controladores borrosos implementados en «*shell*» de sistemas expertos de lógica borrosa [Ying y col., 1988].

El esquema del controlador desarrollado por Ying y col. [1988] se muestra en la figura 2.

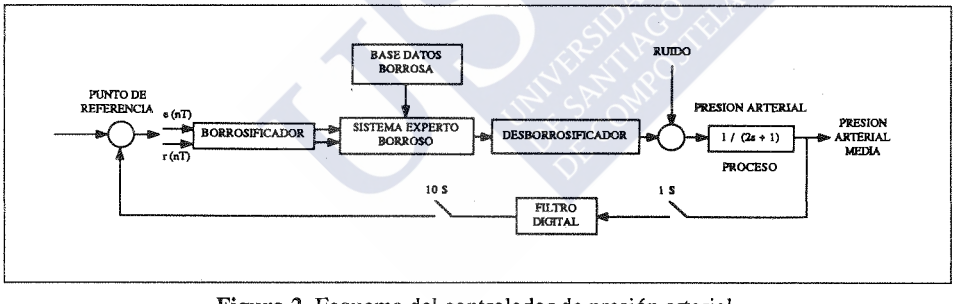


Figura 2. Esquema del controlador de presión arterial.

Este controlador fue implementado en una «*shell*» de sistemas expertos borrosos FLOPS. Este sistema utiliza 3 bases de datos. La primera base almacena las definiciones de los conjuntos borrosos «Error», «Cambio en el Error» y «Razón de cambio de la velocidad de infusión». La segunda base almacena información desborrosificada de las etiquetas borrosas «Razón de cambio de la velocidad de infusión». La tercera base almacena las reglas de control borroso en palabras directamente. En total se implementaron 121 reglas describiendo la estrategia de control. La principal ventaja de este sistema es la facilidad para cambiar las bases de datos que definen el controlador borroso y con ello poder definir con rapidez diferentes estrategias de control.

El controlador propuesto consiste en un procedimiento de borrosificación, tres bases de datos borrosas, la herramienta para sistemas expertos borrosos FLOPS, un procedimiento de desborrosificación y un filtro digital. Las funciones del controlador borroso fueron implementadas en FLOPS.

El procedimiento de borrosificación convierte los valores numéricos «Error» de la presión arterial media, e(nt), y «Cambio en el Error» de la presión arterial media, r(nt), en conjuntos borrosos. Debido a la implementación del controlador en FLOPS, los conjuntos borrosos se pueden considerar continuos y, por tanto, no hay que discretizar las variables de entrada como en otros controladores borrosos desarrollados. Las reglas de control se eligieron intentando imitar la estrategia de control de un experto. Una regla típica es de la forma:

> Si «ERROR» de la presión arterial media es POSITIVO–GRANDE y si «CAMBIO EN EL ERROR» es NEGATIVO–PEQUEÑO, entonces «CAMBIO EN RAZÓN DE INFUSIÓN» de nitroprusiato sódico es NEGATIVO–MEDIO.

Las simulaciones demuestran que este controlador funciona bien para pacientes de distintas sensibilidades a la droga.

6. CONTROL DEL DOLOR POSTOPERATORIO

Normalmente, el tratamiento del dolor postoperatorio es realizado por una enfermera que a determinados intervalos administra drogas opiáceas en la cantidad prescrita por el médico. Este tratamiento es imperfecto, debido principalmente a que el intervalo de administración es muy grande, generalmente de varias horas.

Usualmente se administran por vía intravenosa, esto da un efecto mucho más rápido y el pico de acción en el tiempo es más predecible.

Los opiáceos producen, además de la analgesia, efectos secundarios no deseados, como nauseas, vómitos y la peligrosa depresión respiratoria.

En las PCA («patient-controlled analgesia») [Tamsen y col., 1979] el paciente presiona un botón cuando siente dolor, la PCA administra un bolo de tamaño predeterminado cada vez que el botón es pulsado. Por motivos de seguridad, las PCA tienen límites de droga administrada en el tiempo y algunos modelos mejorados tienen además añadida una infusión constante predeterminada por el médico.

Las PCA han mejorado notablemente la calidad asistencial aunque, sin embargo, no son óptimas. Para la administración de droga, las PCA necesitan que el paciente tenga dolor y consecuentemente presione un botón, esto produce que el paciente oscile entre el nivel de «no dolor» y cierto grado de dolor.

Además se han hecho algunos intentos con controladores algorítmicos «Integral pulse frequency modulation/Smith delay Compensator» (IPFM/SDC) [Feng-Yu y Northrop, 1990] que ha mostrado buenos resultados en la administración de agentes hipotensores, así como controladores proporcionales, y controladores estocásticos [Jacobs y col., 1985].

La medida de la intensidad del dolor subjetivo continua siendo un tema de investigación. Se han desarrollado diversos métodos, como el McGill Pain Questionnaire [Tamsen y col., 1986], escalas analógicas, numéricas y verbales, etc.. En el dolor postoperatorio las más utilizadas por su facilidad y rapidez son las escalas visuales analógicas (VAS) y las escalas numéricas. En este controlador hemos utilizado una escala numérica. El paciente mueve el cursor de un potenciómetro lineal que tiene marcada la escala numérica de 0 a 10, indicando de ese modo su grado de dolor .

Para el dolor postoperatorio no hay un modelo matemático preciso desarrollado. Para estos tipos de sistemas poco conocidos se han desarrollado controladores basados en reglas. La idea es implementar la estrategia de control de un operador humano. Al estar basado en la experiencia del operador no necesitan un modelo matemático, lo que los hace muy apropiados para sistemas complejos, mal definidos o que estén sujetos a muchas perturbaciones.

7. CARACTERÍSTICAS DEL DOLOR POSTOPERATORIO

Discutiremos el dolor postoperatorio siguiendo un modelo aproximado desarrollado por Jacobs y col. [1985] y mostrado en la figura 3.

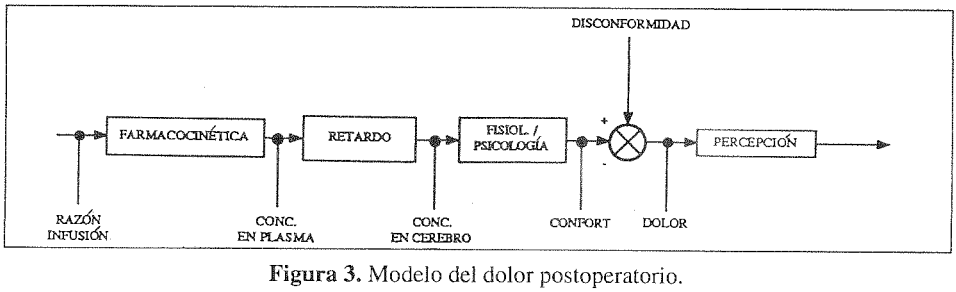


Figura 3. Modelo del dolor postoperatorio.

En la actualidad la única parte plenamente establecida y contrastada clínicamente es el modelo farmacocinético que describe la relación droga-concentración en plasma. El bloque «fisio/psicológico» incluye el efecto farmacológico de la droga que por ahora es poco conocido, pues no se conoce con detalle como interaccionan los opiáceos con los receptores cerebrales. Este bloque incluye además factores psicológicos difíciles de cuantificar y modelar. Se ha demostrado, por ejemplo, que la ansiedad y el miedo aumentan el dolor percibido, mientras que distraer al paciente poniéndole música disminuye el dolor. Esto es un aspecto de la subjetividad del dolor postoperatorio.

El bloque de percepción contiene una fuerte no-linealidad, que implica que no hay dolor con valores negativos, simplemente se queda saturado al llegar a dolor=0 o «no dolor».

El modelo farmacocinético describe la distribución cuantitativa de las concentraciones de droga en los tejidos del cuerpo.

Para una determinada droga, el cuerpo se puede considerar como compuesto de varios compartimentos, cada uno de los cuales puede ser considerado como un espacio donde la sustancia está uniformemente distribuida y tiene una cinética uniforme de distribución y transporte.

Nosotros hemos utilizado Alfentanil, un opiáceo cuyo modelo farmacocinético puede ser descrito por tres compartimentos, y representado en la figura 4.

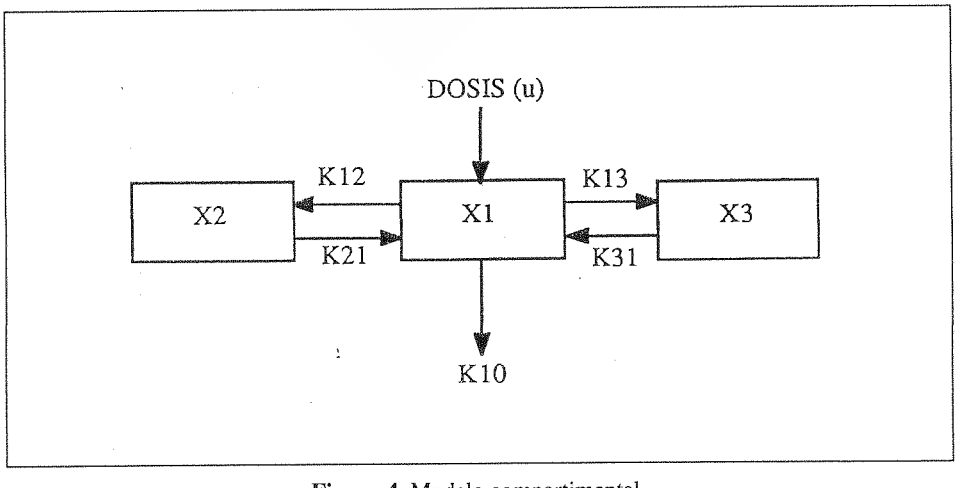


Figura 4. Modelo compartimental.

La droga se inyecta en el compartimento «central», que representa al plasma, desde el cual es eliminado (constante k10). La razón de eliminación es proporcional a la cantidad de droga en ese compartimento (Xi). Entre el compartimento central y los otros dos compartimentos se produce un intercambio de droga proporcional a la cantidad de droga en cada compartimento.

Datos de las constantes Kij para el Alfentanil han sido hallados experimentalmente. Nosotros hemos utilizado los datos hallados por Bovill y col. [1982].

Existe un retardo de transporte de la droga hasta el cerebro que hemos supuesto del orden de 1 minuto.

En cuanto al modelo farmacodinámico se han propuesto algunos modelos, sin embargo, no hay estudios clínicos contrastados para el dolor postoperatorio utilizando como medida del dolor VAS o escalas numéricas. Esto se debe principalmente a que se conoce poco el efecto farmacológico de los opiáceos y a la gran subjetividad que caracteriza al dolor postoperatorio, sobre el que influye factores psicológicos y factores ambientales difíciles de modelar.

8. ESTRUCTURA DEL CONTROLADOR

Para tratar este sistema se ha utilizado un controlador SOC [Carollo y col., 1991]. El SOC es un controlador borroso adaptativo, inicialmente desarrollado por Procyk y Mamdani [1979]. El SOC está constituido básicamente por un controlador borroso, una tabla de rendimiento y un algoritmo para modificar las reglas de control.

El SOC desarrollado tiene la estructura de la figura 5.

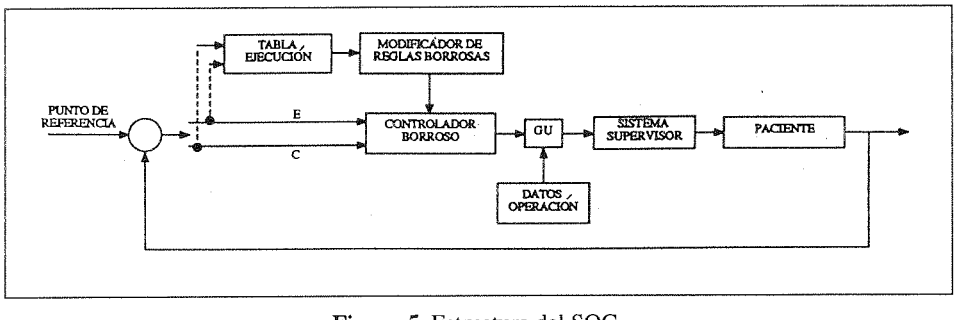


Figura 5. Estructura del SOC.

A cada instante de muestreo se chequea el «error» y el «cambio en el error», estos datos se llevan a una tabla que nos indica las condiciones de operación en ese instante. Este dato es usado por un modelo inverso incremental del sistema para obtener cuanto se tienen que incrementar o decrementar las salidas de las reglas. El modificador de reglas selecciona las reglas del controlador cuyos patrones se parecen en mayor medida al «error» y «cambio de error» muestreado y les modifica la salida.

El sistema supervisor chequea que la salida del controlador (cantidad de droga) no sobrepase ciertos limites y chequea la frecuencia respiratoria y si esta se deprime reduce paulatinamente la cantidad de droga administrada.

Las simulaciones demuestran que el SOC puede tratar con estos procesos poco definidos y mantener el dolor bajo con poca cantidad de droga.

9. CONCLUSIONES

A través de esta revisión de los controladores borrosos en medicina se ha visto que estos controladores pueden tratar con sistemas sin modelos matemáticos definidos imitando la estrategia de control de un operador humano. Esta característica es de gran utilidad en el control de los sistemas fisiológicos humanos, puesto que estos sistemas son complejos, están sujetos a muchas perturbaciones y algunos tienen componentes de subjetividad que los hace muy difíciles de modelar. Los controladores borrosos analizados demostraron un buen comportamiento frente a la variabilidad inter e intra-paciente, tanto en el estado transitorio como en el estado estacionario.

REFERENCIAS

Bovill, J.M., Sebel, P.S., Blackburn, C.L. y Heykants, J., «The pharmacokinetics of Alfentanil (R39209), a new opioid analgesic», *Anesthesiology*, N. 57, (1982), 439-443.

Carollo, A., Tobar, A. y Hernández, C., «Control Fuzzy del dolor postoperatorio», *Primer Congreso Español sobre Tecnologías y Lógica Fuzzy*, (1991), 151-154.

Feng-Yu, L. y Northrop, R.B., «A new approach to the modelling and control of postoperative pain», *IEEE Trans. Biomed. Eng.*, (1990), 1147-1158.

Isaka, S. y Sebold, A.V., «An Adaptative Fuzzy Controller for Blood Pressure Regulation», *Proc. IEEE Eng. Med. & Biol. 15th Annual International Conferences*, (1989), 1763-1764.

Jacobs, O.L.R., Bullingham, R.E.S., Linner, P., McQuay, H.J., O'Sullivan, G. y Reasbeck, M.P., «Modelling Estimation and Control in the Relief of Post-Operative Pain», *Automatica*, Vol. 4, N. 21, (1985), 349-360.

Linkens, D.A. y Mahkouf, M., «Fuzzy Log Knowledge-Based Control for Muscle Relaxant Anaesthesia», *IFAC Symposium*, (1988), 185-90.

Mamdani, E.H. y Assilan, S., «An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller», en *Int. J. Man-Machine Studies*, N. 7, (1975), 1-13.

Procyk, T.J. y Mamdani, E.H., «A Linguistic Self-Organizing Process Controller», *Automatica*, N. 15, (1979), 15-30.

Sugima, F., Mizushima, S., Kimura, M., Fukuri, Y. y Harada, Y., «A Fuzzy Approach to the Rate Control in on Artificial Cardiac Pacemaker Regalated by Respiratory Rate and Temperature: A Preliminary Report», *Journal of Medical Engineering & Technology*, Vol. 3, N. 15, (1991), 107-110.

Tamsen, A., Hartuig, P., Dahlstrom B. y Holmdahl, «Patient controlled analgesic therapy in the early postoperative period», *Acta Anaesth. Scand.*, (1979), 462-479.

Tamsen, M.P., Karoly , P. y Braver, S., «The measurement of the clinical pain intensity: a comparison of the methods», *Pain*, N. 27, (1986), 117-126.

Ying, H., Sheppard, L. y Jucker, D., «Expert-System-Based Fuzzy Control of Arterial Pressure by Drug Infusion», *Medical Progress through Technology*, N. 13, (1988), 203-215.

Zadeh, L.A., «Outline of a New Approach to the Analysis of Complex Systems and Decision Processes», *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 1, N. 3, (1973), 28-44.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

OTROS ÁMBITOS DE PROYECCIÓN DE LA LÓGICA BORROSA

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CAP. VI	Clasificación borrosa basada en conocimiento
CAP. VII	Razonamiento temporal borroso
CAP. VIII	Computación neuronal borrosa
CAP. IX	Uncertainty in C & L Machines
CAP. X	El modelo difuso como programa de investigación lingüística
CAP. XI	Aplicación de la lógica borrosa a la selección de carteras

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CLASIFICACIÓN BORROSA BASADA EN CONOCIMIENTO

Ramón Ruiz
Departamento de Informática y Automática. Universidad de Murcia

Senén Barro y Jesús R. Presedo
Departamento de Electrónica y Computación. Universidad de Santiago de Compostela

1. INTRODUCCIÓN

El problema de *clasificar* objetos en categorías puede considerarse una parte de un problema más general denominado Reconocimiento de Patrones (*Pattern Recognition*). Entender las alternativas y alcance de los algoritmos de clasificación, sean clásicos o borrosos, pasa por la consideración conjunta de las partes que intervienen en una aplicación genérica de Reconocimiento de Patrones (RP).

Vamos a dedicarnos en este trabajo a repasar los aspectos más relevantes del RP, fundamentalmente en la relación que guardan con la teoría de conjuntos borrosos y los modelos a que da lugar.

Lo primero que podemos preguntarnos es en qué consiste estrictamente el RP. Existen múltiples definiciones que intentan describir en pocas palabras lo esencial de la disciplina en sí misma. Para Duda y Hart [1973], el RP se caracteriza por ser un *«campo relacionado con el reconocimiento automático de regularidades significativas en entornos ruidosos o complejos»*. Dicho en otros términos algo más operativos, la finalidad de esta materia no es otra que la *«búsqueda de la estructura subyacente en cualquier conjunto de datos»* [Bezdek y Pal, 1992], sirviendo de

herramienta en cualquier área de conocimiento en la que se planteen problemas de análisis e interpretación de conjuntos de variables o medidas. A modo de ejemplo de la enorme variedad de áreas de aplicación de esta disciplina, podemos enumerar una serie de ellas, si bien no son las únicas posibles:

- Comunicación hombre-máquina (reconocimiento automático y comprensión del habla, reconocimiento de caracteres manuscritos, comprensión de imágenes).
- Defensa (reconocimiento automático de blancos, guía y control).
- Medicina (diagnosis médica, análisis de imágenes médicas, clasificación de enfermedades).
- Vehículos (controladores de automóviles, aviones, trenes o barcos).
- Investigación policial (identificación por voz, huellas digitales o fotografías).
- Estudios de recursos naturales (agricultura, geología, bosques y medio ambiente).
- Industria (CAD, CAM, prueba y ensamblaje de productos, control de calidad).

Si hubiera que hacer un resumen, en muy pocas palabras, de las etapas que comprende el desarrollo de un *Sistema de Reconocimiento de Patrones* (SRP) genérico, éstas podrían ser:

1.- El experto humano enumera los datos que reflejan las relaciones básicas entre las variables aparentemente importantes de un proceso.
2.- Los datos son recopilados, bien a través de sensores o introducidos vía teclado.
3.- A continuación se busca la estructura subyacente en estos datos, que proporcionará una base para hacer hipótesis sobre las relaciones y leyes que gobiernan el proceso.
4.- Las hipótesis se formalizan mediante la caracterización del proceso a través de ecuaciones, reglas o algoritmos; en resumidas cuentas, se propone un *«modelo»* del sistema.
5.- Siempre que sea posible, se analizarán aspectos teóricos del modelo (p.e. linealidad, continuidad y estabilidad) para avanzar en el conocimiento del proceso que representa.
6.- El modelo entonces se «entrena» con datos *etiquetados* (preasignados a una categoría) y se calculan los parámetros del mismo.

7.- Por último, se testea el modelo con un conjunto diferente de datos etiquetados para valorar su eficacia y, en su caso, ajustarlo. El comportamiento del modelo suele comparase con el de otros modelos alternativos del mismo proceso.

El producto final de este procedimiento de desarrollo será un conjunto hardware/software que, basado en el modelo, permitirá clasificar, predecir, estimar y/o controlar los elementos del proceso.

2. NATURALEZA Y FUENTE DE LA IMPRECISIÓN DE LOS SRP

Cuando se trabaja con clasificadores convencionales (determinísticos o probabilísticos), las características que representan los vectores de entrada son, la gran mayoría de las veces, cuantitativas en su naturaleza; los vectores con especificaciones imprecisas o incompletas usualmente son ignorados o descartados. La imprecisión en los SRP no es, sin embargo, algo infrecuente y puede presentarse en dos niveles:

- en el nivel de los vectores de características de entrada
- en el nivel de determinación de la estructura de los datos (*clustering*) o de diseño del clasificador.

En los vectores de características de entrada la imprecisión puede venir introducida por distintas causas. Los errores instrumentales, por ejemplo, o la corrupción de las medidas por ruido se presentan a menudo en ciertas aplicaciones, y conducen a que la información sea sólo parcialmente fiable; el precio a pagar para incrementar significativamente la fiabilidad o precisión de un sistema con frecuencia es demasiado alto. Piénsese, por ejemplo, en los problemas de registro asociados con ciertas señales bioeléctricas como el electrocardiograma (ECG), en donde la aparición de interferencias de red, los artefactos por problemas de contacto de electrodos o los ruidos musculares, pueden contaminar seriamente el contenido de información de la señal, introduciendo altos grados de imprecisión en las medidas numéricas puntuales. Su minimización sólo será posible a costa de añadir complejidad a la electrónica de registro y preprocesado, o bien incrementando sustancialmente la carga de computación en la extracción de características, lo que suele ser incompatible con la operación en tiempo real. En otras ocasiones, la imprecisión de los vectores de entrada puede proceder del hecho de que el problema no sea cuantitativo en esencia (problemas de índole *humanística*), lo que

condiciona una vaguedad inherente en su formulación y en las variables que lo caracterizan. Son problemas de este tipo los relacionados con: finanzas, lingüística, ciencias médicas, psicología, etc. Por los motivos apuntados, a veces resulta conveniente utilizar el formalismo de las variables lingüísticas para describir la información contenida en los vectores de características de entrada, en lugar de pretender imponer la rigidez de las representaciones numéricas exactas.

En el nivel de definición de agrupamientos de objetos y diseño del clasificador, de nuevo puede aparecer la imprecisión. Una de las razones puede ser la naturaleza solapante de las clases, en cuyo caso las técnicas convencionales de asignación no van a ser aplicables, dado que la idea básica en la que se sustentan es que una muestra sólo podrá pertenecer a una clase (*partición* en teoría clásica de conjuntos). Una segunda razón puede ser la limitación de conocimiento a priori acerca de los criterios de clasificación (medidas de similaridad entre objetos). La forma de soslayar estos problemas pasa por permitir que los objetos posean grados de pertenencia parciales a más de una categoría o, lo que es lo mismo, interpretar la estructura de clases en el marco de la teoría de conjuntos borrosos.

Para hacernos una idea algo más gráfica de cómo interviene la imprecisión en la propia estructura de las clases, consideremos el problema de determinar las fronteras de una clase a partir de sus *prototipos* (objetos preasignados). Las aproximaciones convencionales tienen como objetivo estimar de forma exacta el área de cada clase, sin solape con ninguna otra, de forma que contenga la mayoría (si no todos) de los puntos prototipo. Sin embargo, este objetivo no es deseable necesariamente, por ejemplo, para fronteras en imágenes reales; puede ser necesario, en este caso, extender las fronteras de una clase para caracterizar porciones «oscurecidas» no definidas por los puntos prototipo. Estas porciones extendidas deberán tener grados menores de pertenencia que las zonas definidas explícitamente por los puntos pertenecientes a la clase, y su tamaño se reducirá conforme se incremente el número de puntos prototipo (información disponible), al minimizarse la imprecisión.

En conclusión, y dada la naturaleza de los problemas que plantea la vaguedad en los SRP, se puede afirmar que la teoría de conjuntos borrosos constituye un marco formal adecuado para el tratamiento de esta imprecisión en cualquier nivel de su diseño. Más aún, veremos que proporciona una metodología particular para el diseño de los clasificadores que, de forma natural, incorpora el conocimiento heurístico expresado en términos lingüísticos por el experto.

3. CAMPOS CONSTITUYENTES DE UN SRP

En el procedimiento de diseño del SRP más general que podamos imaginarnos, nos encontraremos cuatro bloques o apartados principales [Bezdek y Pal, 1992], que aparecen reflejados en la fig. 1:

- Descripción del proceso (variables que intervienen en él)
- Análisis de características (exploración y transformación de los datos)
- Determinación de agrupamientos (procedimientos de *«clustering»*)
- Diseño del clasificador (criterios de asignación a clases)

Las opciones que pueden seleccionarse en cada uno de estos grandes bloques no serán independientes entre sí, de forma que la elección de una opción en uno de estos apartados va a restringir, normalmente, el espacio de opciones de los tres restantes.

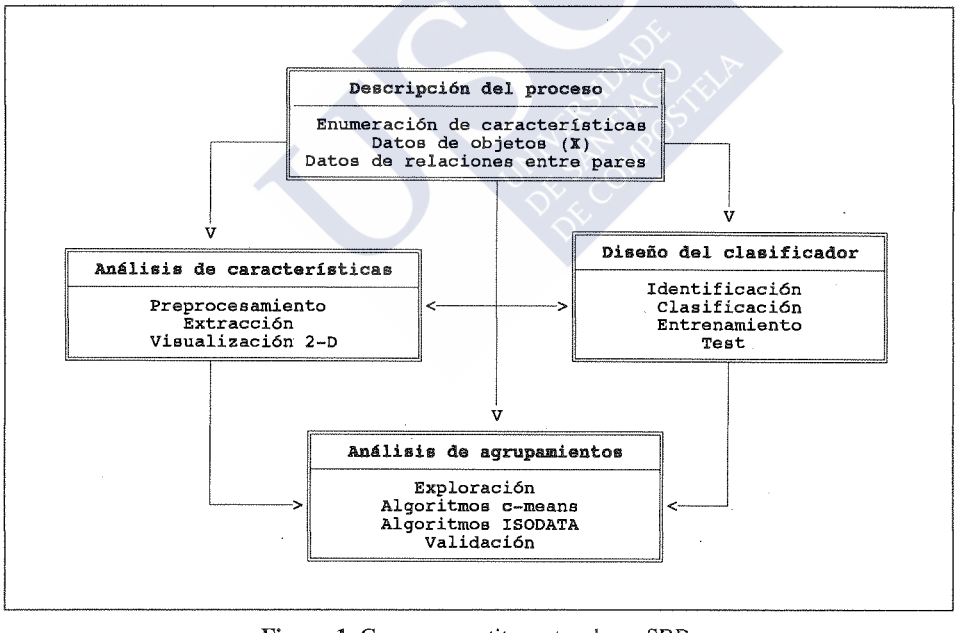


Figura 1. Campos constituyentes de un SRP.

De la estructura presentada, y dado que nuestro interés se centra en las alternativas basadas en la lógica borrosa, surge una cuestión: ¿debería existir un quinto bloque que englobará las particularidades de diseño de los SRP basados en

los modelos borrosos?. La respuesta es taxativamente no. La razón se fundamenta en que los modelos borrosos proporcionan un medio de dar solución a los problemas planteados en cada bloque de una manera particular, pero no constituyen una alternativa global diferenciada al diseño de este tipo de sistemas.

En lo que sigue, vamos a proceder a describir en cierto detalle las opciones de que dispone el diseñador de un SRP en cada uno de estos grandes apartados, con más interés en aquellas que se derivan o relacionan con los modelos borrosos.

3.1. Descripción del proceso

La primera elección que ha de hacer el diseñador de un sistema de reconocimiento de patrones tiene que ver con la forma en que ha de representarse para su estudio el proceso de interés. En reconocimiento de patrones, la situación usual es que el proceso esté gobernado (o gobierne) por objetos individuales (p.e. individuos, registros sísmicos, imágenes, etc.) o relaciones entre parejas de objetos. Si bien la elección más familiar es la de representación de estas entidades (objetos o relaciones) mediante cantidades *numéricas*, existen otras alternativas de representación como son la *sintáctica*, *contextual* y *basada en reglas*.

El reconocimiento de patrones sintáctico mantiene que los algoritmos deben ser capaces de descomponer y reconstruir los objetos (o sus relaciones) a partir de representaciones de relaciones estructurales entre sus diversas partes, de una forma semejante a como actúan, aparentemente, los seres humanos. La descripción de la estructura de los objetos se realiza en esta aproximación en el marco de las gramáticas formales, y el lector interesado puede recurrir al texto de Fu [1982], en donde encontrará una descripción detallada de los aspectos teóricos y prácticos relativos a las mismas. De otro lado, las aproximaciones contextual y basada en reglas constituyen intentos de añadir contexto, concepto y/o conocimiento experto en los niveles bajos e intermedios de procesamiento, en orden a capitalizar la experiencia previa en estas tareas de reconocimiento. Por ejemplo, el contexto de uso podría permitirnos predecir la letra que falta en la palabra «*-arco*»; será *m* si aparece en una descripción de cuadros, *b* si el contexto se refiere a la náutica, *n* si estamos hablando de productores de narcóticos, o *p* si hablamos de una persona lacónica. Los esfuerzos en esta dirección suelen incluir elementos de diseño de sistemas expertos tales como módulos de adquisición, representación y manipulación de conocimiento.

Refiriéndonos más extensamente a la alternativa más utilizada para la descripción de los objetos, la numérica, digamos que suelen utilizarse dos tipos de estructuras de datos en los SRP numéricos:

- vectores de *datos de objetos* (ya sean de características o de prototipos de clases)
- *datos relacionales* entre parejas de objetos, según una relación de similaridad o proximidad.

Los datos de objetos, esto es, los conjuntos de vectores numéricos de características, los representaremos como $X = \{x_1, x_2, x_3, ..., x_n\}$, y no son otra cosa que n vectores definidos en $\mathbb{R}^p$; un vector genérico, x_j, será la representación numérica del j-ésimo objeto observado, siendo x_{jk} la k-ésima característica de dicho objeto (k=1...p).

Puede ocurrir que, en lugar de a un conjunto X de datos de objetos, tengamos acceso a un conjunto de n^2 relaciones numéricas, $\{r_{jk}\}$, definidas entre pares de objetos. Cada coeficiente r_{jk} de esta *matriz de relaciones* es una cantidad numérica que representa la intensidad con que los objetos j y k se relacionan según una ρ definida entre:

$$\rho : O \times O \rightarrow \mathbb{R} \tag{1}$$

siendo O el conjunto de objetos, $O = \{o_1, o_2, ..., o_n\}$. A partir de la relación binaria ρ, los coeficientes de la matriz de relaciones podremos expresarlos formalmente:

$$R = [r_{jk}] = [\, \rho\,(o_j, o_k)\,] \qquad j,k = 1...n \tag{2}$$

Dentro de los *datos relacionales* pueden distinguirse, a su vez, dos clases: los *naturales* y los *inducidos*. Se habla de datos relacionales naturales cuando las definiciones de las relaciones se realizan por sí mismas, implícitamente a partir de los objetos, con independencia del conocimiento de los vectores de datos de objetos. Este es el caso, por ejemplo, de las taxonomías numéricas, donde las relaciones (entre especies, familias, etc.) son asignadas por expertos. La otra clase de datos relacionales, los inducidos, se definen a partir del conjunto de vectores de datos (no a partir de los propios objetos) por medio de alguna relación σ:

$$\sigma : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R} \tag{3}$$

siendo los coeficientes de la matriz de relaciones, en este caso, función de los vectores de características:

$$[r_{jk}] = [\ \sigma\ (x_j, x_k)\ ] \qquad (4)$$

De nuevo, la introducción de los modelos borrosos en la etapa de descripción del proceso no representa una alternativa adicional a las presentadas en este apartado, concretándose en opciones particulares en cada una de las representaciones indicadas. La aproximación sintáctica ha incorporado la lógica borrosa a dos niveles. En el primero de ellos, las *«primitivas»* (elementos) de descripción de los objetos se consideran como etiquetas representativas de subconjuntos borrosos; así, por ejemplo, en el caso de descripciones de tramos de curvas podemos encontrarnos con primitivas como *«arcos casi circulares»*, *«abrupto»* o *«suave»*. En el segundo nivel, las propias relaciones estructurales entre los elementos constituyentes de un objeto pueden definirse de forma borrosa; la gramática se «borrosifica» por medio de reglas de producción entre cláusulas borrosas. En la alternativa numérica, los conjuntos borrosos se incorporan a través de la sustitución de los componentes numéricos de los vectores o matrices por subconjuntos borrosos definidos en los espacios de datos. Aparece la *variable lingüística* en sustitución de la variable numérica de los sistemas convencionales, y de ella nos ocuparemos en extenso más adelante.

3.2. Análisis de características

En este bloque trataremos la colección de métodos y estrategias utilizables para explorar y mejorar los datos de partida, definidos y recopilados en el apartado de descripción del proceso. Los métodos incluidos dentro de este bloque corresponden a tres tipos de operaciones: *preprocesamiento, visualización 2-D* y *extracción.*

El preprocesamiento incluye operaciones de escalado, normalización, suavización y otras técnicas que tienen por finalidad la limpieza de los datos y el resalte de la información que puedan contener. La utilidad de los datos en tareas posteriores más complejas, tales como la definición de agrupamientos o el diseño del clasificador, quedará fuertemente afectada por este tipo de operaciones, lo que da una idea de su importancia en el conjunto del sistema de reconocimiento.

Para introducir las técnicas de extracción de características, es necesario comenzar definiendo qué se entiende por *«extractor»* de características. Cualquier función f_E que relacione los espacios:

$$f_E : \mathbb{R}^p \rightarrow \mathbb{R}^q \qquad \text{con } p \geq q \qquad (5)$$

será un extractor de características al aplicarse a X; las nuevas características serán la imagen de X bajo f_E. El objetivo básico de estas técnicas no es otro que comprimir el espacio de características, vía selección o transformación, eliminando características redundantes (dependientes de otras características) o no relevantes para el problema. Los métodos de extracción podrán ser de muy diversa naturaleza (analíticos, algorítmicos, lineales, etc.), siendo los más conocidos los de determinación de ejes naturales, análisis factorial y los métodos basados en redes neuronales.

La visualización 2-D, es decir, la representación gráfica de datos d-dimensionales en un plano, es una forma de explorar su estructura y sacar conclusiones a partir de la impresión visual. Los métodos de visualización 2-D pueden ser de dos clases: diagramas de dispersión (*scatterplots*) y representaciones *pictográficas*. Con referencia al primero de los tipos, en operaciones de extracción en donde $q = 2$ (que corresponden a estas visualizaciones 2-D), el conjunto de datos transformados puede ser visualizado haciendo corresponder sus valores con coordenadas cartesianas en un plano. En la figura 2 se muestra un ejemplo de uno de estos diagramas. En esta figura se representan dos características extraídas de los latidos cardíacos, muy significativas a la hora de hacer un diagnóstico de las arritmias cardíacas que puede sufrir un paciente; del contexto de puntos presentados en el gráfico se pueden extraer rápidas y valiosas conclusiones acerca de las alteraciones de ritmo cardíaco ocurridas a lo largo de un intervalo de tiempo [Ruiz y col., 1984]. La otra clase de representaciones, las pictográficas, suponen algún tipo de transformación analítica o algorítmica de los datos extraídos para generar un *pictograma*. Representaciones de este tipo son, por ejemplo, las *cuantificaciones cromáticas*, donde ciertos rangos de valores se asocian con colores del espacio cromático, o las denominadas *«caras de Chernoff»*, que asocian a la expresión de un rostro esquematizado ciertos rangos de las variables bajo representación.

De nuevo podemos cuestionarnos cómo pueden intervenir los modelos borrosos en el bloque de análisis de características. Hay que empezar constatando que no existe un interés tan intenso por los modelos borrosos en este apartado como en los que presentaremos de definición de agrupamientos o diseño del clasificador, si bien existe un cierto número de trabajos inscritos en este contexto. En las operaciones de preprocesamiento, los modelos borrosos se han venido introduciendo como alternativa en la suavización y limpieza de los datos, especialmente en áreas como la del procesamiento de imágenes, donde suelen existir problemas de bajo contraste, ruido y baja luminosidad. En las técnicas de extracción de características

y visualización 2-D la lógica borrosa se ha utilizado en apoyo de los procedimientos de reducción de dimensionalidad, por medio de la aplicación de algoritmos de *«clustering»* borrosos, así como en la selección de características a partir de la definición de funciones de posibilidad.

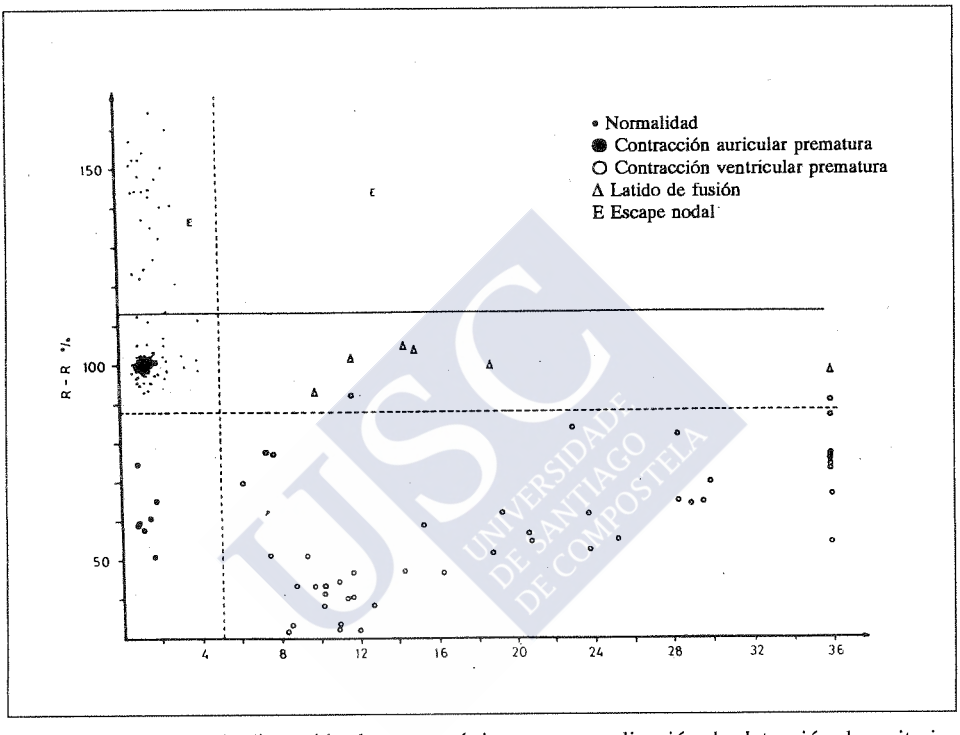


Figura 2. Diagrama de dispersión de características en una aplicación de detección de arritmias cardíacas.

3.3. Análisis de agrupamientos de objetos («clustering»)

Dado un conjunto finito de datos de objetos, X, el problema de definir *agrupamientos* en X consiste en asignar *etiquetas* de pertenencia a los elementos de dicho conjunto, que identifiquen subgrupos naturales en el mismo según un cierto criterio de similaridad. Dado que los datos no están asignados a priori, este problema suele denominarse aprendizaje no supervisado. El objetivo de este bloque es, en suma, realizar una partición de X en un número c (conocido o desconocido)

de subconjuntos naturales y homogéneos, donde los elementos de cada subconjunto deben ser tan similares entre sí como sea posible, y diferentes (según la misma relación) de los de otros subconjuntos. El realizar esta partición de forma adecuada solucionará ciertas cuestiones referentes a la descripción del proceso, lo que ayudará al correcto diseño del clasificador.

A modo de ejemplo de las técnicas de definición de agrupamientos, pensemos en el problema de segmentar una imagen; el número de regiones (c) suele ser una incógnita, y la segmentación se realiza usualmente por procedimientos multiumbral. Las subimágenes o zonas se delimitarán, por ejemplo, mediante un criterio de máximo contraste entre regiones, máxima homogeneidad dentro de cada región, o dimensiones significativas de las zonas. Evidentemente, la partición resultante, tanto en cuanto a número de clases como en cuanto a su estructura, dependerá fuertemente de los criterios utilizados en el control del algoritmo de *«clustering»*.

Existen dos tipos de técnicas de extracción de agrupamientos que se basan en ideas muy diferentes, relativas al concepto de *partición* de un conjunto de objetos: las *convencionales* y las *borrosas*. Las técnicas convencionales se sustentan en el concepto clásico de partición (*«hard»*), de forma que cada objeto pertenecerá a una y sólo una clase. Se han desarrollado multitud de algoritmos dirigidos a obtener este tipo de agrupamientos *«hard»* a partir de un conjunto dado de datos. Los más ampliamente utilizados son los algoritmos *c-means* y sus generalizaciones, los métodos de clustering *ISODATA* (*Iterative Self-Organizing Data Analysis Techniques A*) [Ball y Hall, 1967], que también presentan versiones adaptadas a la teoría de conjuntos borrosos [Bezdek, 1974]. Los algoritmos *c-means* suponen c conocido, en tanto que ese número de agrupamientos es desconocido en el caso de los algoritmos *ISODATA*, siendo ambas aproximaciones iterativas. La esencia del *c-means «hard»* es la minimización de un índice de comportamiento (la suma del cuadrado de las distancias de todos los puntos de una clase al centro de la misma), mientras que los algoritmos *ISODATA* se basan en la combinación de medidas de comportamiento heurísticas, basadas en la experiencia e intuición, con la función del criterio *c-means*. El comportamiento en ambos esquemas queda muy influido por: la elección de c, los centros iniciales elegidos para las clases, el orden en que se van introduciendo las muestras, la elección de la medida de la distancia, y las propiedades geométricas de los datos.

Las técnicas borrosas de definición de agrupamientos suponen una alternativa a las convencionales que se apoya en un concepto de pertenencia basado en la

teoría de conjuntos borrosos. En estas técnicas, el concepto básico lo constituyen los espacios de partición borrosa, que vamos a exponer brevemente a continuación. Sea X un conjunto de n vectores de características no etiquetados, definidos en $\mathbb{R}^p$, y c un entero que verifique la doble desigualdad $1 < c < n$. Decimos que c subconjuntos borrosos $\{ u_i : X \rightarrow [0,1] \}$ constituyen una *c-partición borrosa* de X, si y sólo si los $c \times n$ valores de pertenencia de los elementos de X a los c subconjuntos borrosos, definidos como:

$$\{ u_{ik} = u_i(x_k) , \ 1 \leq k \leq n \ , \ 1 \leq i \leq c \} \tag{6}$$

satisfacen las siguientes tres condiciones:

$$\begin{array}{lll} 1. & 0 \leq u_{ik} \leq 1 & \forall i, k \\ 2. & \sum_i u_{ik} = 1 & \forall k \\ 3. & 0 < \sum_k u_{ik} < n & \forall i \end{array} \tag{7}$$

Cada conjunto de valores que satisface estas tres condiciones puede organizarse en una matriz de dimensión $c \times n$, que podemos llamar U. El conjunto de todas las matrices U posibles constituye todas las *c-particiones borrosas* posibles de X, y lo llamaremos M_{fcn} [Ruspini, 1969]:

$$M_{fcn} = \{ U \text{ en } \mathbb{R}^{cn} \mid u_{ik} \text{ satisfaga las tres condiciones, } \forall i, k \} \tag{8}$$

En el caso de que todos los valores u_{ik} sean 0 ó 1, tendremos el subconjunto de *c-particiones «hard»* de X, que llamaremos M_{cn}:

$$M_{cn} = \{ U \text{ en } M_{fcn} \mid u_{ik} = 0 \text{ ó } 1, \forall i, k \} \tag{9}$$

La matriz M_{fcn} proporciona un marco descriptivo mucho más rico que M_{cn} de cara al tratamiento y representación de los datos, facilitando una herramienta valiosa para abordar las situaciones en las que se presenten objetos de adscripción dudosa. Estas situaciones son frecuentes en los casos reales, donde las fronteras y criterios de clasificación de objetos a menudo no están completamente definidas.

La idea más importante contenida en las definiciones aportadas es que todos los algoritmos de *«clustering»* del conjunto X (hard o borrosos) generan soluciones que son matrices en M_{fcn}. En otras palabras, la determinación de agrupamientos

en X es, sencillamente, la identificación de una partición óptima U de X en M_{fcn}, que agrupe en las mismas clases objetos cuyas características compartan alguna similaridad bien definida. Para hacer posible la partición de un conjunto de objetos son necesarios, por tanto, dos elementos:

1. Una medida de similaridad bien definida entre objetos y grupos de objetos.
2. Un criterio que establezca la optimalidad de soluciones alternativas en M_{fcn}.

La fig. 3 ilustra gráficamente un procedimiento de *«clustering»* de objetos en base a similaridades en su forma. La idea es simple: colocar etiquetas según el grupo natural al que pertenece el objeto. Obsérvese que en el intento de asignación de etiquetas *«hard»* (formas puras de manzana-*M*, pera-*P*, naranja-*N*) nos encontramos con el problema de una forma no estrictamente encuadrable en ninguna de las categorías definidas (*?*); este objeto representa una anomalía en los datos que no es posible tratar adecuadamente con criterios todo/nada, según se deduce de la tabla de asignación de etiquetas en la *3-partición «hard»* (tabla I.a). La alternativa borrosa a la partición permite afrontar este tipo de problemas y evitar los errores, tal como se muestra para nuestro ejemplo en la tabla I.b. La pertenencia de un objeto a más de una clase, en este caso, supondrá una señal para echar una segunda mirada a las asignaciones, a la vez que aporta un grado de detalle muy superior al de la alternativa *«hard»*.

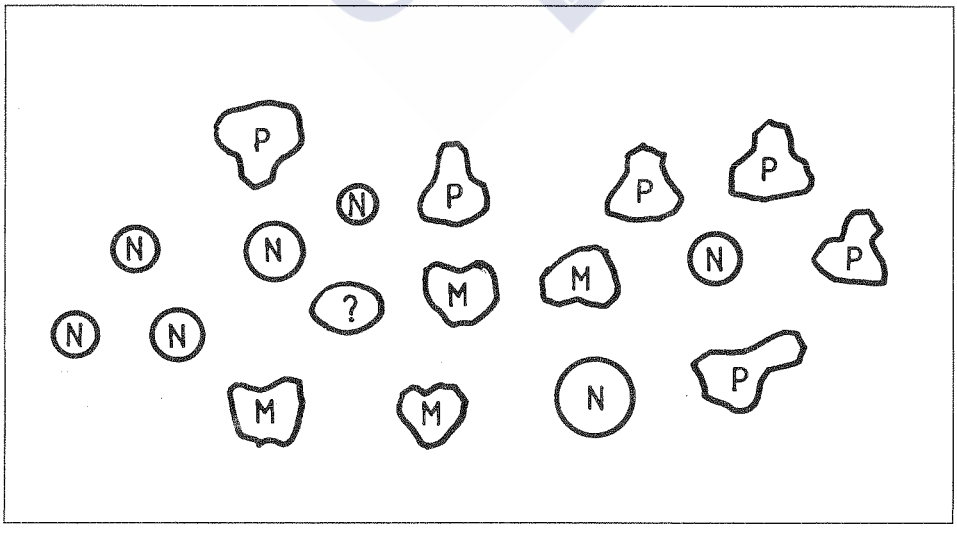


Figura 3. Asignación de etiquetas al conjunto de objetos de la figura 3.

TABLA I. Asignación de etiquetas al conjunto de objetos de la figura 3.

Asignación "hard" (en M_{cn})				
M	1 1 1 1	0 0 0 0 0 0 0	0 0 0 0 0 0	0
N	0 0 0 0	1 1 1 1 1 1 1	0 0 0 0 0 0	1
P	0 0 0 0	0 0 0 0 0 0 0	1 1 1 1 1 1	0

(a) determinística

Asignación borrosa (en M_{fcn})				
M	1 ..... 0.95	0.0 0	0 0.1 0	0.15
N	0 ..... 0.05	0.8 1	0 0.2 0	0.60
P	0 ..... 0.00	0.2 0	1 0.7 1	0.25

(b) borrosa

Podemos ilustrar los procedimientos de *«clustering»* con un ejemplo más realista, que supone una aportación al reconocimiento de muestras en electrocardiografía (ECG) desarrollado en la Universidad de Santiago [Cabello y col., 1991]. Se parte de un conjunto de registros de ECG (objetos), de los que se extraen hasta siete características o *descriptores* espectrales (datos de objetos). Haciendo un análisis de componentes principales se constata que sólo cuatro componentes (resultantes de un procedimiento de extracción) son suficientes para describir los objetos, y de estas cuatro componentes se eligen las dos más significativas para construir los gráficos de visualización 2-D. En la fig. 4 se muestra este gráfico (*«scatterplot»*) para un total de 90 registros analizados; los objetos están asignados a priori en cinco categorías (asignación no utilizada en el procedimiento de *«clustering»*), según criterios diagnósticos aplicados por un cardiólogo experto.

En el desarrollo del trabajo se eligieron aleatoriamente 50 registros para aplicar un algoritmo de definición de agrupamientos *c-means borroso*, cuyo

resultado se ilustra en la fig. 5 (líneas continuas); las líneas punteadas de esta figura reflejan los grupos que resultan de aplicar un criterio más restrictivo de asignación, en el que se exige una pertenencia mínima de 0.7 para poder realizar dicha asignación. Dado que los resultados de este algoritmo no son completamente satisfactorios, se indagó el funcionamiento de otro método de *«clustering»*, el denominado de *covarianza borrosa* [Gustafson y Kessel, 1979]. Se trata de una modificación del algoritmo *c-means borroso*, con la que puede mejorarse la respuesta en la definición de agrupamientos, si existe (y este es el caso) un conocimiento previo del aspecto de la distribución de las muestras; este conocimiento permite definir normas adecuadas y diferentes para cada porción del espacio de características. En nuestro ejemplo, la aplicación de esta variante mejora notablemente la definición de los agrupamientos, según puede apreciarse en la fig. 6.

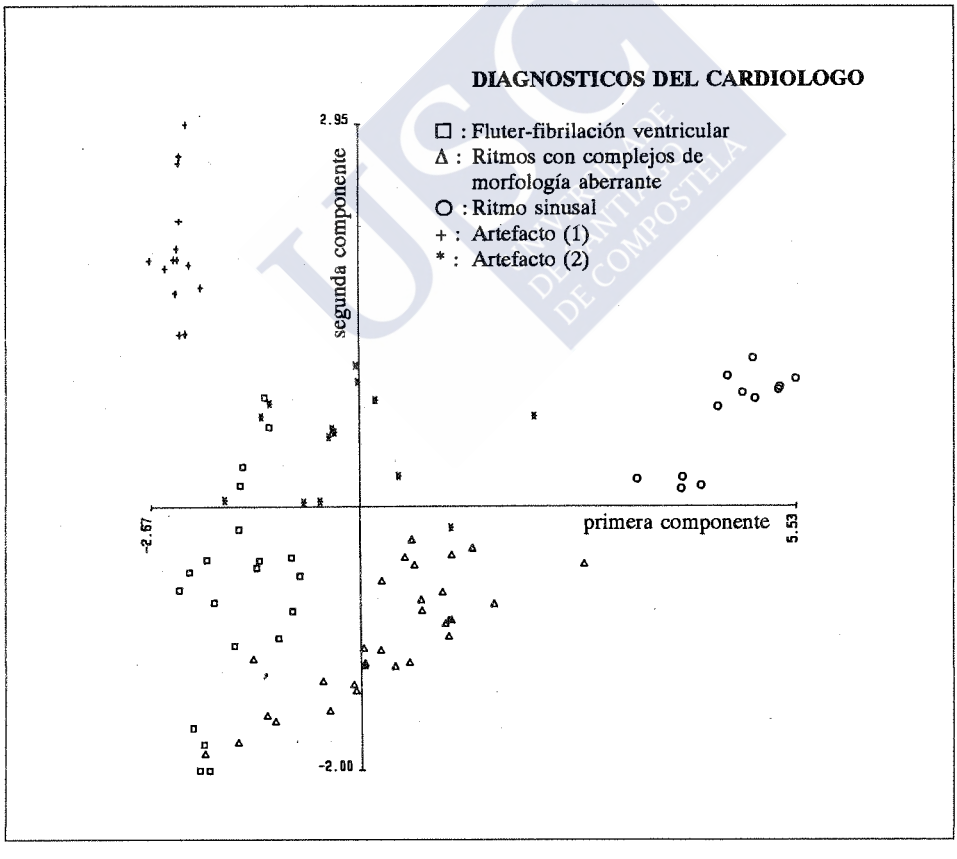


Figura 4. Dispersión de características de las dos componentes principales.

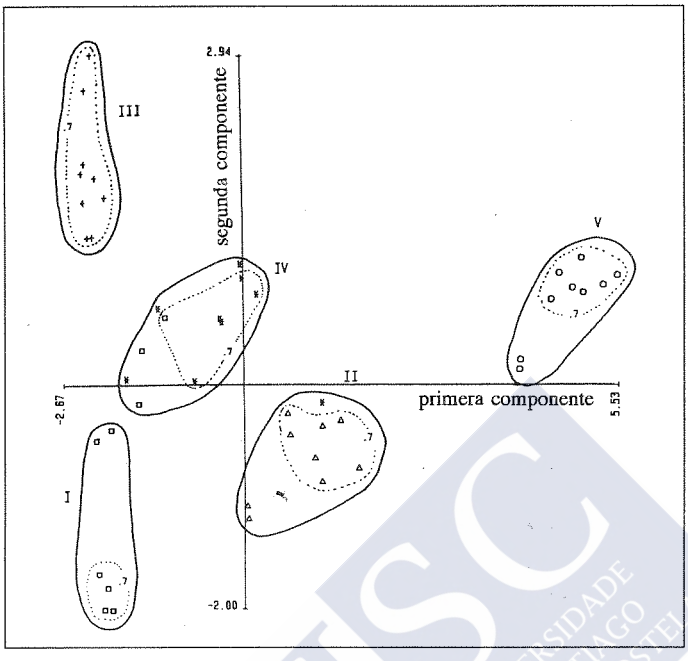


Figura 5. Resultado del *"clustering"* por el método *c-means* borroso.

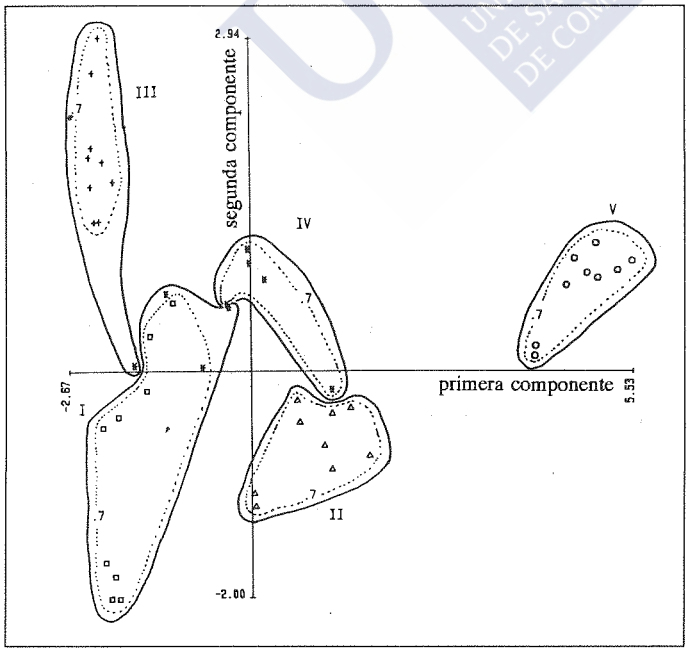


Figura 6. Resultado del *"clustering"* utilizando la técnica de covarianza borrosa.

3.4. Diseño del clasificador

El diseño del clasificador constituye un problema más ambicioso, difícil de resolver y de mayor utilidad potencial que el de determinar agrupamientos. Se refiere a encontrar una partición (*«hard»* o borrosa) de $\mathbb{R}^p$ en su totalidad. La diferencia más importante entre las fases de determinación de agrupamientos y la de clasificación reside en que en la primera se etiquetan conjuntos conocidos de datos, $X \subset \mathbb{R}^p$, mientras que un clasificador es capaz, una vez se define, de etiquetar cualquier punto en $\mathbb{R}^p$. Normalmente los clasificadores se diseñan con datos etiquetados, razón por la que a veces se denomina la fase de diseño del clasificador como de aprendizaje supervisado.

La caracterización de un clasificador pasa, en primer lugar, por la definición de un conjunto de *vectores de etiquetas* en $\mathbb{R}^c$, siendo c el número de clases. Si se trata de un clasificador *«hard»*, los vectores de etiquetas para las c clases se definen:

$$N_c = \{\ y \in \mathbb{R}^c \mid \sum y_k = 1\ \ ;\ y_k \in \{0,1\}\ \ \forall j\ \} \tag{10}$$

Si el clasificador es borroso, los vectores de etiquetas borrosos para c clases se definirán:

$$N_{fc} = \{\ y \in \mathbb{R}^c \mid \sum y_k = 1\ \ ;\ y_k \in [0,1]\ \ \forall j\ \} \tag{11}$$

N_c no es más que una base canónica usual del c-espacio euclídeo, y N_{fc} es un conjunto de puntos delimitado por N_c, que configura un «casco» convexo para éste. La fig. 7 ilustra los conjuntos de vectores de etiquetas (*«hard»* y borrosas) para $c = 3$; N_3 es el conjunto de tres vértices del triángulo mostrado en la figura (base de vectores ortonormales), mientras que N_{f3} es el propio triángulo, de forma que un vector de etiquetas borroso será, por ejemplo, el indicado en la figura, verificándose en cualquier caso el que sus componentes sean menores que la unidad y sumen 1. Puede comprobarse que las columnas de las tablas de asignación de etiquetas en particiones *«hard»* o borrosas (p.e. las indicadas en la tabla I) son vectores en N_c y N_{fc}, respectivamente.

Partiendo de N_{fc}, definimos como clasificador borroso en $\mathbb{R}^p$ a cualquier función $\underline{D}$ que tenga su imagen en N_{fc}:

$$\underline{D}:\ \mathbb{R}^p \rightarrow N_{fc} \subset \mathbb{R}^c \tag{12}$$

En otras palabras, cualquier punto de $\mathbb{R}^p$ tendrá, bajo $\underline{D}$, una imagen que será un vector de etiquetas borroso. La única restricción que introducen los

clasificadores convencionales («*hard*») respecto a los borrosos, es que la imagen se definirá, en este caso, en N_c.

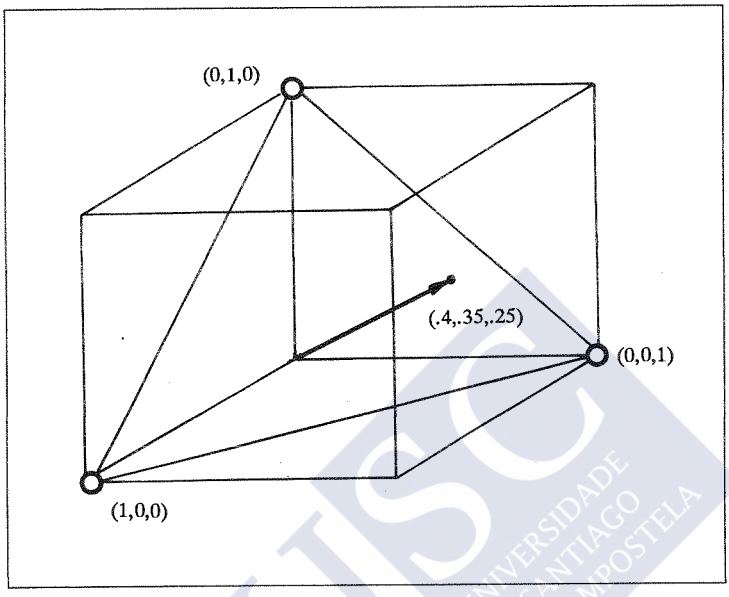


Figura 7. Vectores de etiquetas *"hard"* y borrosos (c=3).

A la hora de encarar el diseño de un clasificador, existe una cierta diversidad de alternativas o posibles opciones. A nuestro juicio, cabe distinguir como más importantes tres vías o estrategias a las que se ajustan la mayoría de los diseños de clasificadores:

a) Utilización de técnicas de «*clustering*» para producir los c vectores prototipo de cada una de las clases ($\{v_k\} \subset \mathbb{R}^p$, k=1..c), a partir de datos no etiquetados. Una vez determinados los v_k , la técnica de clasificación a utilizar será, preferentemente, la de *prototipo más próximo* (*1-NP* ó *Nearest Prototype*).

b) Partiendo de un modelo paramétrico explícito de clasificador (p.e., en términos de funciones discriminantes), en una primera fase se realiza una estimación de los parámetros por medio de un conjunto de vectores de datos etiquetados, que llamaremos de *entrenamiento*. En una segunda fase se valida la operación del clasificador por medio de un segundo conjunto de vectores de datos etiquetados, denominados de *test* o prueba.

c) Formulación del clasificador por medio de un conjunto de *reglas de producción* que recojan y estructuren el conocimiento de un experto. Para nuestros propósitos, y dada la temática del capítulo que nos ocupa, esta alternativa heurística basada en conocimiento es la que más nos interesa, y a ella nos dedicaremos más en extenso en el siguiente apartado.

Existe, por último, una cuestión que se plantea en todos los casos en los que se utilizan técnicas borrosas para resolver los aspectos de los diferentes bloques o etapas intermedias que comprende el diseño de un clasificador: ¿cómo se concilia la necesidad última de asignación de etiquetas absolutas (*«hard»*) a los objetos, con el empleo de técnicas basadas en un concepto de pertenencia parcial o borroso? La respuesta es inmediata si repasamos los conceptos vertidos en este trabajo. La búsqueda de una solución todo/nada a un problema será más fácil y correcta si se extienden los espacios de exploración en las etapas intermedias, permitiendo un número mayor de grados de libertad. En estas condiciones, la respuesta del sistema va ser, en primera instancia, un ente borroso expresado por medio de distribuciones de posibilidad. Para obtener la respuesta final absoluta, será necesario aplicar algún mecanismo de *desborrosificación*, en la línea de los métodos utilizados de manera usual en las aplicaciones de control borroso.

4. PROCEDIMIENTO DE DISEÑO DE UN CLASIFICADOR BASADO EN CONOCIMIENTO

En ciertos problemas de clasificación existe una gran cantidad de conocimiento a priori acerca del clasificador, depositado en la experiencia acumulada de humanos que conocen muy bien el problema. En tales casos, el objetivo principal a la hora del diseño del SRP consiste en extraer y estructurar dicho conocimiento, lo que da lugar a una metodología particular de diseño del clasificador que podemos denominar *«humano-dependiente»*. Dado que la comunicación y razonamiento humanos son imprecisos en su naturaleza, la representación del conocimiento en estos casos se realiza por medio de *variables lingüísticas*, basadas en la teoría de conjuntos borrosos; este tipo de variables constituyen el vehículo, tanto de la declaración del conocimiento del experto referente al problema (*reglas* o conocimiento estático), como de las componentes de los vectores de entrada (*hechos* o conocimiento dinámico), fundamentalmente aquéllas que son aportadas por el experto vía teclado.

Por la importancia que para este tipo de sistemas tienen las variables lingüísticas, es conveniente definirlas formalmente. Una variable lingüística se caracteriza por una quíntupla (H , $T(H)$, U , G , M), en donde:

- H : es el nombre de la variable.
- $T(H)$: es el conjunto de *términos* o valores lingüísticos de H; cada valor lleva asignado su *significado*, que se concreta en una distribución de posibilidad (subconjunto borroso) definida en un único universo de discurso para todos los valores.
- U : es el universo de discurso de definición de los valores lingüísticos.
- G : son las reglas sintácticas de generación de términos compuestos, a partir de los *términos atómicos* que configuran las sentencias que dan lugar a cada valor lingüístico. Estos términos atómicos pueden ser de cuatro tipos:

  - Términos *primarios* (tales como *«viejo»*, *«joven»* o *«alto»*) que son en sí mismos elementos del conjunto $T(H)$

  - Operadores *lógicos* (*NOT, AND, OR*)

  - *Modificadores* lingüísticos borrosos (tales como *MUY, LIGERAMENTE* o *MAS–O–MENOS*)

  - *Marcadores* para delimitación de significados (p.e. paréntesis)
- M : son las reglas semánticas de asignación de significados, a partir de la estructura sintáctica de cada término compuesto y del significado de los términos simples (primarios).

Para ilustrar el concepto de variable lingüística vamos a acudir a un ejemplo en donde se pretende describir la temperatura de un proceso, no por sus valores numéricos, sino por términos lingüísticos. En este caso, el nombre de la variable (H) es, obviamente, *«temperatura»*, y el conjunto de términos ($T(H)$) se diseñará tan amplio como sea necesario para mejor especificar el problema. Por ejemplo:

$$T(H) = \{ALTA\ ,\ NO\text{–}ALTA\ ,\ BAJA\ ,\ MEDIA\ ,\ MAS\text{–}O\text{–}MENOS\text{–}ALTA\ ,\ \ldots\ \} \qquad (13)$$

Con respecto al universo de discurso (U), éste comprenderá un rango de temperatura apropiado para definir los valores borrosos significativos para el

problema en toda su extensión (p.e. 0-100ºC). Lo que hemos denominado reglas sintácticas (G), de construcción de términos compuestos a partir de los atómicos, en este caso asumen como *primitivas* tres términos primarios: *ALTA*, *BAJA* y *MEDIA*; términos como *NO–MUY–BAJA* ó *NO–ALTA–O–BAJA* surgen de la *gramática* que conforma este conjunto de reglas sintácticas. Por último, las reglas semánticas de asignación de significados se apoyan en los de los términos primarios (distribuciones de posibilidad), que en nuestro ejemplo son los tres indicados anteriormente:

$$\begin{aligned} \mu_{ALTA}(Temp) &= f_1(Temp) \\ \mu_{BAJA}(Temp) &= f_2(Temp) \\ \mu_{MEDIA}(Temp) &= f_3(Temp) \end{aligned} \tag{14}$$

Cualquier término compuesto tendrá un significado que puede calcularse de estas tres funciones, sometidas a los operadores de modificación y/o composición. Así, por ejemplo, para el término *MUY–ALTA* tendremos una distribución de posibilidad que puede calcularse aplicando el modificador *MUY* (cuadrado) al significado del término *ALTA*:

$$\mu_{MUY-ALTA}(Temp) = [\mu_{ALTA}(Temp)]^2 = [f_1(Temp)]^2 \qquad \forall\, Temp \in [0,100] \tag{15}$$

En la fig. 8 se ilustran los términos primarios y algunos términos compuestos del ejemplo, descritos en términos de sus distribuciones de posibilidad.

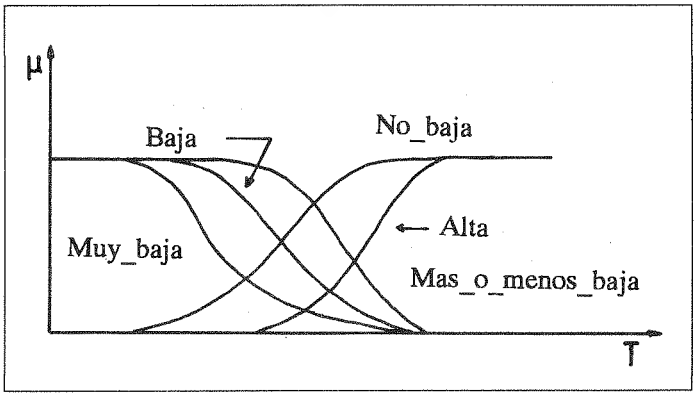


Figura 8. Algunos términos de la variable lingüística *temperatura*.

El desarrollo de un clasificador según la metodología que hemos denominado «humano-dependiente» suele dividirse en dos fases diferenciadas. En una primera etapa se realiza una enumeración de las reglas y distribuciones utilizando únicamente el conocimiento del experto, de forma que se complete un sistema de producción coherente y estructurado. Tras esta primera fase suele ser necesaria una segunda de refinamiento, que se efectúa utilizando conjuntos de muestras etiquetadas; los mecanismos de refinamiento podrán depender, de nuevo, de las habilidades del experto, o bien de técnicas automáticas de corrección de las salidas, basadas en conceptos relativos a la identificación de sistemas borrosos.

4.1. Etapa de inyección de conocimiento del experto

Para empezar, cualquier problema de clasificación en donde cada vector de entrada haya de asignarse en un conjunto de m clases, habrá de descomponerse en m problemas paralelos o *subclasificadores*, que se aplican a cada vector de entrada (x). Las funciones de pertenencia de cada uno de estos vectores de entrada a las m clases (C_i), constituyen las salidas de los subclasificadores, y las denotaremos por $MC_i(x)$.

La misión del experto en esta fase se desarrolla en una doble vertiente. Por un lado en la definición del *«diccionario»* de términos lingüísticos primarios que son relevantes para el problema de clasificación, y que han de ser expresados como distribuciones de posibilidad en los correspondientes universos de discurso. Por otro lado, en la construcción del sistema de reglas de producción borrosas que incluyan como cláusulas los subconjuntos borrosos derivados del diccionario, en orden a representar el conocimiento relativo al problema de clasificación.

Una herramienta de ayuda a la construcción del sistema de reglas es la denominada Matriz de Cuestiones que, caso de estar completamente especificada, englobaría todas las posibilidades de combinación de cláusulas. Un ejemplo muy sencillo de Matriz de Cuestiones, para un problema con dos variables de entrada y dos de salida, cada una de ellas con dos valores lingüísticos (*ALTO* y *BAJO*), se muestra en la tabla II. En los casos reales, sin embargo, la enumeración exhaustiva de todas las posibilidades de combinación de cláusulas suele ser impracticable, además de presentarse un gran número de combinaciones no relevantes o imposibles de presentarse. Por esta razón, suele construirse una matriz parcial con sólo aquellas combinaciones de cláusulas que el experto considera como relevantes, y a las que responde afirmativa (+) o negativamente (–), según la

pregunta implícitamente contenida en cada casilla (ver tabla III). La construcción de los diferentes subclasificadores se lleva a cabo utilizando las reglas asociadas a cada casilla con respuesta afirmativa (+).

TABLA II. *Matriz de cuestiones* completamente definida.

	MC_1 será		MC_2 será	
IF	BAJO	ALTO	BAJO	ALTO
H_1 es BAJO				
H_1 es ALTO				
H_2 es BAJO				
H_2 es ALTO				
H_1 es ALTO, H_2 es BAJO				
H_1 es BAJO, H_2 es ALTO				
H_1 es ALTO, H_2 es ALTO				
H_1 es BAJO, H_2 es BAJO				

TABLA III. *Matriz de cuestiones* parcial

	MC_1 será		MC_2 será	
IF	BAJO	ALTO	BAJO	ALTO
H_1 es BAJO	~	~	~	~
H_2 es ALTO	–	+	+	–
H_1 es MEDIO, H_2 es ALTO	–	+	–	+
H_1 es ALTO, H_2 es MEDIO	+	–	–	+

A efectos operativos, cada una de las reglas de producción borrosa que surgen del procedimiento anterior, llevan asociada una matriz de implicación representativa de la relación borrosa que se supone entre premisas y conclusiones. Se trata de matrices de coeficientes reales cuya dimensión es el producto de los

tamaños (número de elementos) de los universos de discurso de las cláusulas antecedentes y consecuente. Mediante estas matrices, la *evaluación* de un vector de datos de entrada de componentes borrosas, x, en el subclasificador i, se efectuará utilizando la composición asociada al *«modus ponens» generalizado*, que opera cada una de las j matrices de implicación contenidas en el subclasificador (R_{ij}) con la entrada citada:

$$MC_i = \bigcup_j (x \circ R_{ij}) \tag{16}$$

siendo MC_i la distribución de posibilidad de salida del subclasificador i , «∘» el símbolo de la composición entre matrices (normalmente la composición *«max-min»* de Zadeh), y «∪» el símbolo de la unión (máximo) del resultado de las diferentes composiciones.

Para ilustrar las operaciones que acarrea la evaluación de una entrada en cada subclasificador, consideremos la siguiente regla de producción borrosa:

$$\text{IF } [H_1 \text{ es } ALTO \text{ AND } H_2 \text{ es } MEDIO] \text{ THEN } MC_2 \text{ es } ALTO \tag{17}$$

en donde los valores lingüísticos para H_1 se definen en el universo de discurso $U_1 = \{1,2,3,4\}$ y los correspondientes a H_2 se definen en $U_2 = \{6,7,8\}$. Supongamos unas distribuciones de posibilidad expresadas como:

$$\begin{aligned} \mu_{ALTO(H_1)} &= 0.1/1 + 0.2/2 + 0.4/3 + 0.7/4 \\ \mu_{MEDIO(H_2)} &= 0.6/6 + 0.8/7 = 0.6/8 \end{aligned} \tag{18}$$

donde el símbolo «+» representa *unión*, en la terminología habitual para los subconjuntos borrosos definidos sobre universos discretos [Zadeh, 1973]. La matriz representativa de las cláusulas antecedentes de la regla indicada será la que surge de hacer el AND de cada pareja de pertenencias de los subconjuntos borrosos premisa, y el resultado se expresa en la tabla IV; las pertenencias expresadas en la tabla están definidas sobre el producto cartesiano de los universos U_1 y U_2. Supongamos ahora que el subconjunto borroso asociado a la conclusión se define sobre el universo de discurso $V = \{A,B,C,D\}$ de la forma siguiente:

$$\mu_{ALTO(MC_2)} = 0.1/A + 0.4/B + 0.6/C + 1/D \tag{19}$$

La matriz representativa de la implicación contenida en la regla será, asumiendo la interpretación de Mamdani del operador condicional (mínimo) [Lim y Takefuji,

1990], la mostrada en la tabla V; los elementos de matriz pueden contemplarse como valores de pertenencia a un subconjunto borroso definido sobre el universo producto cartesiano $U_1 \times U_2 \times V$.

TABLA IV. *ALTO* (H_1) Y *MEDIO* (H_2)

U_1/U_2	6	6	6
1	0.1	0.1	0.1
2	0.2	0.2	0.2
3	0.4	0.4	0.4
4	0.6	0.7	0.6

TABLA V. Matriz de implicación del ejemplo

$U_1 \times U_2/V$	A	B	C	D
(1,6)	0.1	0.1	0.1	0.1
(2,6)	0.1	0.2	0.2	0.2
(3,6)	0.1	0.4	0.4	0.4
(4,6)	0.1	0.4	0.6	0.6
(1,7)	0.1	0.1	0.1	0.1
(2,7)	0.1	0.2	0.2	0.2
(3,7)	0.1	0.4	0.4	0.4
(4,7)	0.1	0.4	0.6	0.7
(1,8)	0.1	0.1	0.1	0.1
(2,8)	0.1	0.2	0.2	0.2
(3,8)	0.1	0.4	0.4	0.4
(4,8)	0.1	0.4	0.6	0.6

Siempre que se necesite evaluar una entrada a través de esta regla, aplicaremos la *regla composicional de inferencia*. Por ejemplo, si dicha entrada es un valor lingüístico A de la variable H_1, definido como:

$$A = 0.05/1 + 0.1/2 + 0.2/3 + 0.4/4 \tag{20}$$

y un valor B para H_2:

$$B = 0.4/6 + 0.7/7 + 0.2/8 \tag{21}$$

el resultado de la evaluación de dicha entrada según la composición *max-min* de Zadeh será un valor de la variable borrosa MC_2, que podemos llamar D, cuyos valores de pertenencia calculados son:

$$D = 0.1/A + 0.4/B + 0.4/C + 0.4/D \tag{22}$$

Este subconjunto borroso está definido, como corresponde, en el universo de discurso V.

En este caso, como en todos aquellos en los que la metodología de diseño del clasificador se basa en la teoría de conjuntos borrosos, las salidas de los subclasificadores serán de naturaleza borrosa, por lo que será necesaria una última etapa de *«desborrosificación»* para traducir las distribuciones de posibilidad en valores numéricos significativos. Como ya hemos comentado con anterioridad, este tipo de técnicas son bien conocidas en los problemas de control borroso, que son tratados extensamente en otros capítulos. Citar tan sólo que dos de las más utilizadas son las de cálculo del *centro de gravedad* de las distribuciones, y la de localización de su *pico máximo*.

4.2. Etapa de refinamiento del clasificador

Existen métodos automáticos y semiautomáticos para corregir las desviaciones en el comportamiento deseado de este tipo de clasificadores basados en conocimiento. Operan pasando vectores de entrada etiquetados para su evaluación por los subclasificadores, corrigiendo las matrices R_{ij} en función de las desviaciones apreciadas. De entre este tipo de técnicas, son especialmente interesantes las que se basan en los métodos de identificación de los sistemas borrosos, y en este apartado nos referiremos específicamente a una propuesta de Nath y Lee [1983].

Consideremos un sistema borroso genérico gobernado por una relación de entrada/salida desconocida, que llamaremos R. El conocimiento de R significará tener completamente determinada la relación de entrada/salida del sistema borroso, de manera que toda salida (y_k) se obtendrá componiendo la entrada borrosa (x_k) con esta matriz:

$$y_k = x_k \circ R \tag{23}$$

Pues bien, se puede hacer una estimación de la matriz desconocida R a partir de secuencias apropiadas de entrada, $\{x_k\}$, y de salida, $\{y_k\}$, combinándolas por medio de *composiciones-α*.

Para introducir el concepto de composición-α es conveniente comenzar definiendo lo que se denomina *operación-α* entre los valores de pertenencia de un subconjunto borroso A, definido en un universo de discurso U:

$$\mu_A(u_1)\ \alpha\ \mu_A(u_2) = \begin{cases} 1 & si \quad \mu_A(u_1) \le \mu_A(u_2) \\ \mu_A(u_2) & si \quad \mu_A(u_1) > \mu_A(u_2) \end{cases} \qquad \forall u_1, u_2 \in U \tag{24}$$

La *composición-α* se define entre dos subconjuntos borrosos, A y B, definidos sobre los universos U y V, respectivamente. Esta composición genera un subconjunto borroso (G) que se define sobre el producto cartesiano $U \times V$ de la siguiente manera:

$$\mu_G(u, v) = \mu_A(u)\ \alpha\ \mu_B(v) \qquad \forall u \in U, v \in V \tag{25}$$

A partir de la *composición-α* de secuencias conocidas de entrada/salida de un sistema borroso, podemos hacer una estimación de la matriz R que identifica su comportamiento, según la siguiente expresión:

$$R = \bigcap_k R_k \qquad \text{siendo} \qquad R_k = x_k\ \alpha\ y_k \tag{26}$$

La *sintonía* o ajuste del clasificador se puede realizar utilizando esta técnica de identificación de sistemas, partiendo de secuencias de vectores de entrada etiquetados. Si x_k es una muestra etiquetada del conjunto de *test* perteneciente a la clase p, pero clasificada erróneamente en la clase q, el procedimiento de ajuste a seguir podemos describirlo en los dos pasos siguientes:

1. Modificación de las salidas de los subclasificadores *p* y *q*, por medio de un desplazamiento de las distribuciones de posibilidad de salida de los dos subclasificadores involucrados:

   - Subclasificador *p*: pasamos de una salida y_{pk} a y'_{pk}, desplazando los valores de pertenencia según la expresión:

   $$\mu_{y'_{pk}}(v + \delta) = \mu_{y_{pk}}(v) \qquad \forall v \in V \tag{27}$$

   siendo δ un número positivo pequeño cuyo valor dependerá del problema y de la magnitud del error de clasificación.

   - Subclasificador *q*: pasamos de una salida y_{qk} a y'_{qk} por medio del desplazamiento:

   $$\mu_{y'_{qk}}(v - \delta) = \mu_{y_{qk}}(v) \qquad \forall v \in V \tag{28}$$

   Con estos desplazamientos se pretende el corregir los centros de gravedad de los valores de salida evaluados en los subclasificadores, haciéndolos tender hacia una clasificación correcta por medio de una actuación directa sobre dichas salidas.

2. Modificación de las matrices de los subclasificadores mediante términos de *«perturbación»* resultado de las composiciones-α de las secuencias de entrada y las de salida desplazadas. La expresión de las matrices modificadas queda:

$$R'_{ij} = R_{ij} \cap (x_k \,\alpha\, y'_{ik}) \qquad \forall i \tag{29}$$

donde *i* es el subíndice representativo de cada uno de los subclasificadores. Este procedimiento se itera hasta conseguir el comportamiento deseado para el clasificador.

A modo de ejemplo de este procedimiento de ajuste, consideremos una matriz R_1 como la indicada en la tabla VI, asociada a un subclasificador dado; los elementos de los universos de discurso de entrada y salida se enumeran en dicha tabla. Consideremos una entrada, x_k, definida en el universo *U* de la siguiente manera:

$$x_k = 0.1/1 + 0.1/2 + 0.3/3 + 0.3/4 + 0.4/5 + 0.6/6 + 0.8/7 + 0.8/8 + 0.9/9 + 0.9/10 \tag{30}$$

Esta entrada, al componerse con la matriz R_1, proporciona una salida:

$$y_k = 0.1/A + 0.2/B + 0.4/C + 0.6/D + 0.8/E + 0.9/F + 0.8/G + 0.4/H + 0.3/I + 0.2/J \quad (31)$$

definida en el universo V. Si desplazamos los valores de pertenencia de la salida una cantidad negativa equivalente a una elemento del universo de discurso de salida, la nueva distribución de salida queda:

$$y'_k = 0.2A + 0.4B + 0.6C + 0.8D + 0.9E + 0.8F + 0.4G + 0.3H + 0.2I + 0/J \quad (32)$$

TABLA VI. R_1

U/V	A	B	C	D	E	F	G	H	I	J
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	0.1	0.2	1	1	1	1	1	1	1	0.2
4	0.1	0.2	1	1	1	1	1	1	1	0.2
5	0.1	0.2	1	1	1	1	1	1	0.3	0.2
6	0.1	0.2	0.4	1	1	1	1	0.4	0.3	0.2
7	0.1	0.2	0.4	0.6	1	1	1	0.4	0.3	0.2
8	0.1	0.2	0.4	0.6	1	1	1	0.4	0.3	0.2
9	0.1	0.2	0.4	0.6	0.8	1	0.8	0.4	0.3	0.2
10	0.1	0.2	0.4	0.6	0.8	1	0.8	0.4	0.3	0.2

La composición-α de esta distribución con la de entrada proporciona una matriz R_2, que se expresa en la tabla VII. Si utilizamos esta matriz para refinar la de partida R_1, el resultado lo tenemos en la tabla VIII. Podemos iterar este procedimiento, tantas veces como sea necesario, para desplazar el centro de gravedad de las distribuciones de salida en el sentido deseado. Aplicando la nueva matriz calculada, $R_1 \cap R_2$, a la entrada xk, obtendremos una salida:

$$y''_k = 0.1A + 0.2B + 0.4C + 0.6D + 0.8E + 0.8F + 0.4G + 0.3H + 0.2I + 0/J \quad (33)$$

que, desplazada de nuevo en una unidad del universo de discurso en sentido negativo, nos dará:

$$y'''_k = 0.2A + 0.4B + 0.6C + 0.8D + 0.8E + 0.4F + 0.3G + 0.2H + 0/I + 0/J \quad (34)$$

TABLA VII. R_2

U/V	A	B	C	D	E	F	G	H	I	J
1	1	1	1	1	1	1	1	1	1	0
2	1	1	1	1	1	1	1	1	1	0
3	0.2	1	1	1	1	1	1	1	0.2	0
4	0.2	1	1	1	1	1	1	1	0.2	0
5	0.2	1	1	1	1	1	1	0.3	0.2	0
6	0.2	0.4	1	1	1	1	0.4	0.3	0.2	0
7	0.2	0.4	0.6	1	1	1	0.4	0.3	0.2	0
8	0.2	0.4	0.6	1	1	1	0.4	0.3	0.2	0
9	0.2	0.4	0.6	0.8	1	0.8	0.4	0.3	0.2	0
10	0.2	0.4	0.6	0.8	1	0.8	0.4	0.3	0.2	0

TABLA VIII. $R_1 \cap R_2$

U/V	A	B	C	D	E	F	G	H	I	J
1	1	1	1	1	1	1	1	1	1	0
2	1	1	1	1	1	1	1	1	1	0
3	0.1	0.2	1	1	1	1	1	1	0.2	0
4	0.1	0.2	1	1	1	1	1	1	0.2	0
5	0.1	0.2	1	1	1	1	1	0.3	0.2	0
6	0.1	0.2	0.4	1	1	1	0.4	0.3	0.2	0
7	0.1	0.2	0.4	0.6	1	1	0.4	0.3	0.2	0
8	0.1	0.2	0.4	0.6	1	1	0.4	0.3	0.2	0
9	0.1	0.2	0.4	0.6	0.8	0.8	0.4	0.3	0.2	0
10	0.1	0.2	0.4	0.6	0.8	0.8	0.4	0.3	0.2	0

Esta distribución desplazada se compone con x_k para obtener una matriz R_3 (tabla IX), que utilizaremos de nuevo para modificar la matriz de rela-

ción obtenida en el paso anterior (R1 ∩ R2). El resultado final se muestra en la tabla X.

TABLA IX. R_3

U/V	A	B	C	D	E	F	G	H	I	J
1	1	1	1	1	1	1	1	1	0	0
2	1	1	1	1	1	1	1	1	0	0
3	0.2	1	1	1	1	1	1	0.2	0	0
4	0.2	1	1	1	1	1	1	0.2	0	0
5	0.2	1	1	1	1	1	0.3	0.2	0	0
6	0.2	0.4	1	1	1	0.4	0.3	0.2	0	0
7	0.2	0.4	0.6	1	1	0.4	0.3	0.2	0	0
8	0.2	0.4	0.6	1	1	0.4	0.3	0.2	0	0
9	0.2	0.4	0.6	0.8	0.8	0.4	0.3	0.2	0	0
10	0.2	0.4	0.6	0.8	0.8	0.4	0.3	0.2	0	0

TABLA X. $(R_1 \cap R_2) \cap R_3$

U/V	A	B	C	D	E	F	G	H	I	J
1	1	1	1	1	1	1	1	1	0	0
2	1	1	1	1	1	1	1	1	0	0
3	0.1	0.2	1	1	1	1	1	0.2	0	0
4	0.2	0.2	1	1	1	1	1	0.2	0	0
5	0.1	0.2	1	1	1	1	0.3	0.2	0	0
6	0.1	0.2	0.4	1	1	0.4	0.3	0.2	0	0
7	0.1	0.2	0.4	0.6	1	0.4	0.3	0.2	0	0
8	0.1	0.2	0.4	0.6	1	0.4	0.3	0.2	0	0
9	0.1	0.2	0.4	0.6	0.8	0.4	0.3	0.2	0	0
10	0.1	0.2	0.4	0.6	0.8	0.4	0.3	0.2	0	0

Que el procedimiento indicado produce en cada paso matrices modificadas de relación que proporcionan un desplazamiento de las distribuciones de salida, queda en evidencia en la fig. 9, donde se ilustran las salidas del sistema del ejemplo en cada una de las etapas de ajuste, ante una misma distribución de entrada.

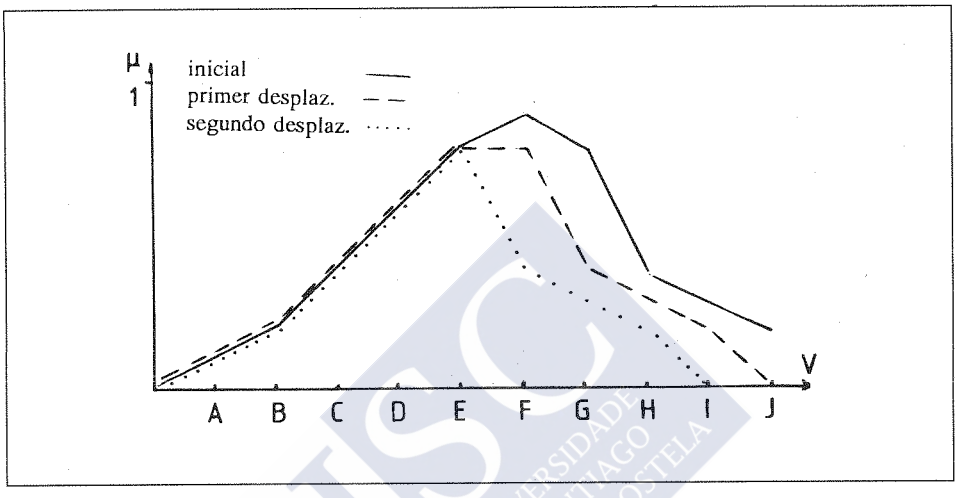


Figura 9. Distribuciones de salida en las sucesivas etapas de ajuste.

4.3. Ejemplo de diseño de un clasificador

Una de las tareas más importantes en las aplicaciones de monitorización de ritmo cardíaco es la identificación del «*tipo*» de cada latido detectado, clasificación que refleja la situación del foco del estímulo eléctrico que produce la contracción del miocardio. La fiabilidad de este procedimiento de clasificación viene exigida por el carácter pronóstico que tiene asociado esta información, que condiciona en muchos casos la terapia antiarrítmica a ser administrada.

La determinación precisa del origen de un latido cardíaco, a partir de registros electrocardiográficos superficiales, no es una tarea sencilla; factores como la posición corporal, la variabilidad entre individuos, la presencia de patologías o el ruido normalmente presente, dificultan esta determinación y aconsejan recurrir a esquemas de clasificación basados en la experiencia del cardiólogo.

En el Departamento de Electrónica y Computación de la Universidad de Santiago se ha desarrollado un procedimiento de clasificación borrosa de latidos cardíacos, basado en conocimiento, que pretende mimetizar los criterios del cardiólogo en esta tarea [Barro y col.,1990]. A efectos de clasificación se consideran en el trabajo tres clases de latidos: *sinusales* (normales), *supraventriculares* y *ventriculares*; una cuarta clase, relativa a posibles artefactos en el registro, se discrimina en el algoritmo de detección de latidos, previo al análisis de características.

La clasificación de latidos, tal como la desarrolla un cardiólogo, se realiza a partir de características de dos tipos:

- locales (ya sean de tiempo, amplitud o morfológicas), medidas sobre el propio latido y los tramos de señal contiguos al mismo
- globales, extraídas del análisis de latidos similares al de interés, que permiten aportar información de la *familia* en la que se encuadra el latido

Cuando esta sistemática de clasificación se pretende implantar en un sistema automático, es conveniente añadir un tercer tipo de variables, que podríamos denominar de usuario en las que sea el operador el que aporte información adicional al sistema. En conjunto, el esquema de bloques del sistema automático de clasificación queda de la forma que refleja el diagrama de la fig. 10.

De entre todos los atributos locales de un latido genérico, existen dos que constituyen el punto de partida en cualquier procedimiento de determinación de la posición del foco de dicho latido: su instante de ocurrencia (expresado en términos de lo que se denomina *intervalos RR*), y su aspecto morfológico, en relación con el de los latidos más frecuentes. Aparte de estos dos atributos esenciales, el sistema manipula un conjunto de características locales que pueden agruparse en cuatro bloques:

- Presencia y valores de actividad auricular
- Duración de la activación ventricular
- Presencia de las denominadas «pausas compensatorias» antes o después del latido
- Inscripción de la onda de repolarización ventricular

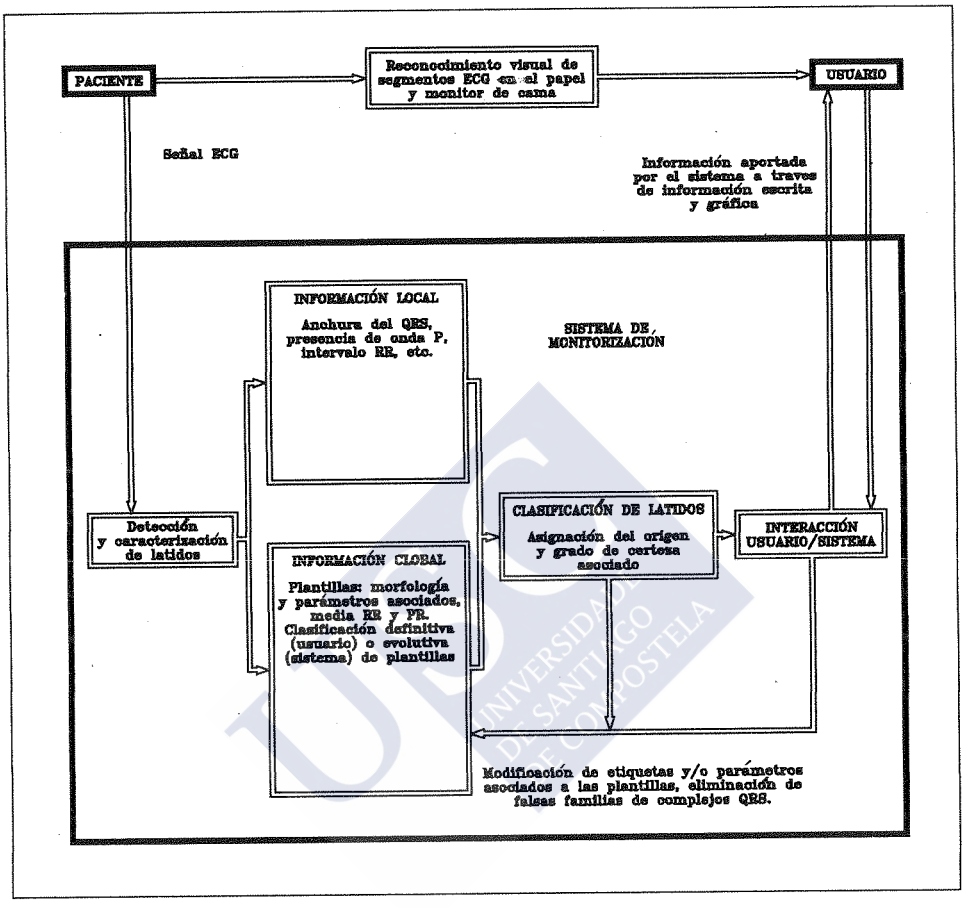


Figura 10. Diagrama de bloques del clasificador de latidos.

Lógicamente, no todos los descriptores tendrán la misma importancia en el proceso de decisión, ni tienen por qué ser utilizados en su totalidad en una clasificación dada. En la tabla XI se muestra el conjunto completo de características numéricas utilizadas, así como las variables lingüísticas que se definen a partir de ellos. Estas definiciones se realizan por medio de un tipo de funciones estándar parametrizadas, las funciones «*S*» de Zadeh [Zadeh, 1979], que se expresan en términos de tres parámetros, α , β y γ, en la forma indicada en la expresión (35). El aspecto de estas funciones estándar se muestra en la fig. 11.

$$S(u;\alpha,\beta,\gamma) = \begin{cases} 0 & u \le \alpha \\ 2((u-\alpha)/(\gamma-\alpha))^2 & \alpha \le u \le \beta \\ 1-2((u-\gamma)/(\gamma-\alpha))^2 & \beta \le u \le \gamma \\ 1 & u \ge \gamma \end{cases} \quad (35)$$

TABLA XI. Características numéricas y variables lingüísticas utilizadas en el clasificador.

Símbolo y definición	**Tipo**	**Conjunto borroso**
Existe Onda P	Boolean	————
$I_{PR} = X_{PR} / \bar{X}_{PR}$	Real	Intervalo PR Largo Δ S (I_{PR}; 1.06, 1.13, 1.20)
D_P = Índice de disimilaridad normalizado con respecto a la plantilla de onda P normal.	Real	Morfología Onda P Normal Δ 1 – S (D_P; 0.1, 0.12, 0.14)
$L = X_{RR} / \bar{X}_{RR}$	Real	Latido Prematuro Δ 1 – S (L; 0.75, 0.85, 0.95) Latido Tardío Δ S (L; 1.06, 1.15, 1.24)
W	Real	QRS Ancho Δ S (W; 0.1, 0.115, 0.13)
W_T	Real	Plantilla QRS Ancho Δ S (W_T; 0.1, 0.115, 0.13)
$C_T = (X_{RR} + X_{RR+})/2 * \bar{X}_{RR}$	Real	Pausa Compensatoria Siguiente Al Latido Δ Δ Latido Prematuro $\wedge$ (S (C_N; 0.85, 0.9, 0.95) $\wedge$ [1 – S (C_N; 1.05, 1.1, 1.15)]) Latido Interpolado Δ Δ S (C_N; 0.4, 0.42, 0.44) $\wedge$ [1 – S (C_N; 0.55, 0.57, 0.59)]
$C_P = (X_{RR} + X_{RR-})/2 * \bar{X}_{RR}$	Real	Pausa Compensatoria Previa Al Latido Δ Latido Tardío $\wedge$ (S (C_P; 0.85, 0.9, 0.95) $\wedge$ [1 – S (C_P; 1.05, 1.1, 1.15)])
$\int = \int_{QRS} / \int_T$	Real	Repolarización Ventricular Invertida Δ Δ [1 – S ($\int$; -0.9, -0.5, -0.1)] $\wedge$ S ($\int$; -3, -2, -1)
D_{QRS} = Índice de disimilaridad normalizado con respecto a la plantilla QRS normal.	Real	Morfología QRS Normal Δ 1 – S (D_{QRS}; 0.1, 0.12, 0.14) Morfología QRS Aberrante Δ S (D_{QRS}; 0.1, 0.2, 0.3)

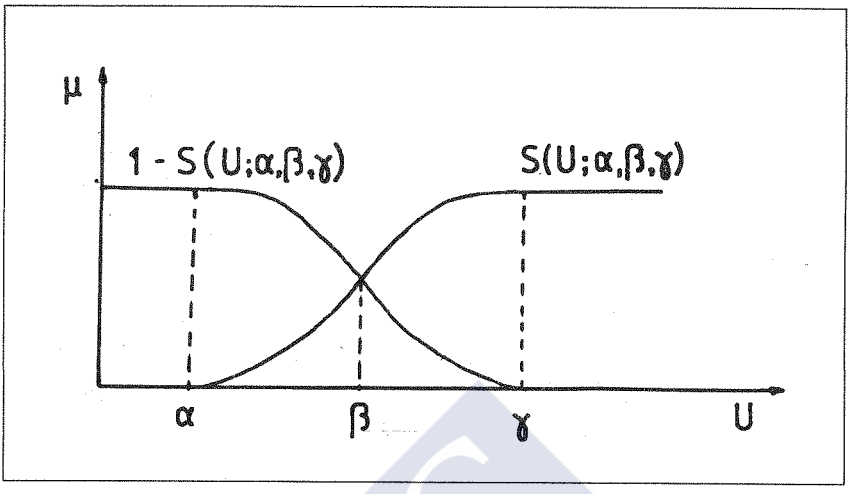


Figura 11. Funciones normalizadas "*S*".

Para ilustrar la deducción de las variables lingüísticas, en esta aplicación, a partir de los descriptores numéricos, consideremos la característica temporal de ocurrencia de un latido en relación con el precedente. El valor numérico de partida es:

$$L = X_{RR} / (X_{RR})_{av} \tag{36}$$

donde X_{RR} es el valor del intervalo temporal transcurrido desde la detección inmediatamente anterior, y $(X_{RR})_{av}$ ($\bar{X}_{RR}$ en la tabla XI) es el promedio calculado de normalidad de estos intervalos. Sobre el universo de valores posibles de este descriptor se han definido dos subconjuntos borrosos, que constituyen los significados de sendos valores lingüísticos de la variable que podríamos denominar *ocurrencia temporal*:

$$\textit{Latido–prematuro} \triangleq 1 - S(L;0.75,0.85,0.95) \tag{37}$$

$$\textit{Latido–tardío} \triangleq S(L;1.06,1.15,1.24) \tag{38}$$

A cualquiera de estos términos, a su vez, pueden aplicársele operadores de modificación para matizar su significado. En concreto, en el clasificador nos encontramos el término:

$$\textit{Latido–muy–prematuro} \triangleq [1 - S(L;0.75,0.85,0.95)]^2 \tag{39}$$

La representación de estos subconjuntos borrosos se ilustra en la fig. 12.

El clasificador borroso de nuestra aplicación adopta una estructura se sistema de reglas de producción borrosas, en donde los consecuentes no son estrictamente conjuntos borrosos, sino índices porcentuales de confirmación o exclusión de la pertenencia del latido a una categoría dada. Así, la forma de las reglas que componen el sistema queda:

$$\text{IF } \textit{antecedente (borroso)} \text{ THEN } (\mu^c,\mu^e) \qquad (40)$$

siendo μ^c y μ^e los índices de confirmación y exclusión de la clasificación, respectivamente, caso de verificarse los antecedentes de la regla. Con los índices calculados en la evaluación de cada regla, la totalización se efectúa combinando confirmaciones y exclusiones parciales para el conjunto completo de reglas; de esta manera, se calcularán un *grado global de confirmación* de la pertenencia a cada categoría, y un *grado global de exclusión*.

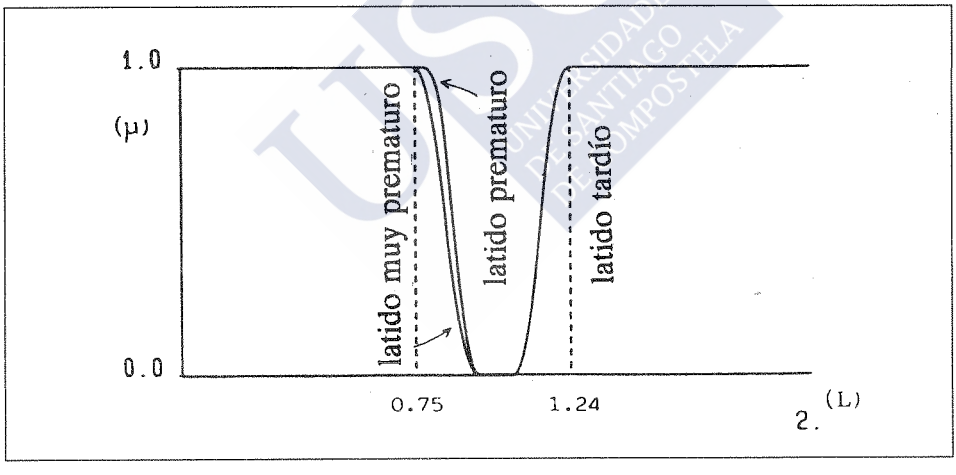


Figura 12. Términos de la variable de *ocurrencia temporal* de los latidos.

Para apreciar el comportamiento del clasificador borroso propuesto, en la fig. 13 se muestran los resultados de la clasificación de latidos para un trazado electrocardiográfico. En ella se describen los valores lingüísticos de las variables involucradas en forma de valores borrosos de verdad, en lugar de hacerlo en

términos de pertenencias numéricas. Con este esquema se pretende acercar aún más la descripción de las variables y de la propia clasificación al lenguaje utilizado por el cardiólogo.

	NUMERO DE LATIDO					
VALORES DE LATIDO	**1A**	**2A**	**3A**	**4A**	**5A**	**6A**
Existe_Onda_P	SI	SI	SI	SI	Descon.	SI
Intervalo_PR_Largo	Nulo	Nulo	Nulo	Nulo	Descon.	Nulo
Morfología_Onda_P_Normal	Absoluto	Absoluto	Muy alto	Absoluto	Descon.	Muy alto
Latido_Prematuro	Nulo	Nulo	Nulo	Nulo	Absoluto	Nulo
Latido_Tardío	Nulo	Nulo	Nulo	Nulo	Nulo	Muy Alto
Pausa_Compensatoria_Siguiente	Nulo	Nulo	Nulo	Nulo	Medio	Nulo
Pausa_Compensatoria_Previa	Nulo	Nulo	Nulo	Nulo	Nulo	Medio
QRS_Ancho	Nulo	Nulo	Nulo	Nulo	Nulo	Nulo
Repolarización_Ventricular_Invertida	Nulo	Nulo	Nulo	Nulo	Nulo	Nulo
Morfología_QRS_Normal	Absoluto	Absoluto	Absoluto	Absoluto	Absoluto	Absoluto
Morfología_QRS_Aberrante	Nulo	Nulo	Nulo	Nulo	Nulo	Nulo
CLASIFICACION DE LATIDO						
Latido_Normal	Absoluto	Absoluto	Absoluto	Absoluto	Nulo	Absoluto
Latido_Ectópico_Supraventricular	Nulo	Nulo	Nulo	Nulo	Absoluto	Nulo
Latido_Supraventricular	Nulo	Nulo	Nulo	Nulo	Nulo	Nulo

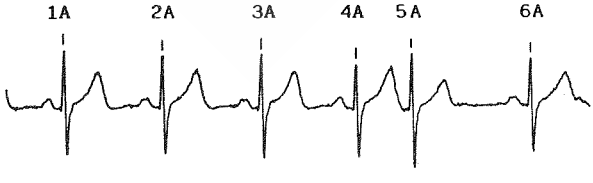


Figura 13. Resultados proporcionados por el clasificador.

5. CONCLUSIONES

Hemos pretendido en este trabajo presentar el panorama y la problemática asociada al reconocimiento de muestras, haciendo un énfasis especial en aquellos aspectos que introduce en este campo la teoría de conjuntos borrosos, como

vehículo de tratamiento de la imprecisión que con frecuencia se presenta. Siendo el diseño de los clasificadores la faceta más interesante de este tipo de sistemas, al menos desde un punto de vista académico, nos hemos extendido en este apartado, presentando los aspectos metodológicos que se aplican a un tipo de clasificadores cuyo soporte es la experiencia de un humano entrenado en la tarea que se pretende automatizar. En este sentido, hemos desarrollado un ejemplo muy en la línea de los problemas para los que esta metodología puede considerarse idónea: la identificación del foco de los latidos cardíacos a partir de su manifestación electrocardiográfica. La cantidad y el tipo de conocimiento asociado con esta tarea definen como alternativa más adecuada para abordar este problema aquélla en la que es el conocimiento del experto el que esencialmente se traduce en la automatización del procedimiento de clasificación, incluso utilizando el propio lenguaje del experto.

AGRADECIMIENTO

Los autores de este trabajo desean agradecer a la Comisión Interministerial de Ciencia y Tecnología el soporte económico aportado, a través del Programa Nacional de Tecnologías de la Información y las Comunicaciones (Ref.TIC91-0997).

REFERENCIAS

Ball, G. y Hall, D., «A clustering technique for summarizing multivariate data», *Behav. Sci.*, N. 12, (1967), 153-155.

Barro, S., Ruiz, R. y Mira, J., «Fuzzy beat labeling for intelligent arrhythmia monitoring», *Computers and Biomedical Research*, N. 23, (1990), 240-258.

Bezdek, J.C., «Cluster validity with fuzzy sets», *J. Cybernet*, N. 3, (1974), 58-73.

Bezdek, J.C. y Pal, S.K., *Fuzzy Models for Pattern recognition. Methods that search for Structures in Data*, IEEE, Inc. New York, 1992.

Cabello, D., Barro, S., Salceda, J.M., Ruiz, R. y Mira, J., «Fuzzy k-nearest neighbor classifiers for ventricular arrhythmia detection», *Int. J. Biomed. Comput.*, N. 27, (1991), 77-93.

Duda, R. y Hart, P., *Pattern classification and scene analysis*, Wiley Interscience, New York, 1973.

Fu, K.S., *Syntactic pattern recognition with applications*, Prentice Hall, Englewood Cliffs, New Jersey, 1982.

Gustafson, D.E. y Kessel, W., «Fuzzy clustering with a fuzzy covariance matrix», en Fu, K.S. (Ed.), *Proc. IEEE CDC*, Vol. 2, IEEE Press, New Jersey, (1979), 761-766.

Lim, M.H. y Takefuji, Y., «Implementing fuzzy rule-based systems on silicon chips», *IEEE Expert*, Febrero, (1990), 31-45.

Nath, A.K. y Lee, T.T., «On the design of a classifier with linguistic variables as inputs», *Fuzzy Sets and Systems*, N. 11, (1983), 265-286.

Ruiz, R., Hernández, C. y Mira, J., «Method for mapping cardiac arrhythmia in real time using microprocessor-based systems», *Med. & Biol. Eng. & Comput.*, N. 22, (1984), 160-167.

Ruspini, E., «A new approach to clustering», *Inform. Control*, N. 15, (1969), 22-32.

Zadeh, L.A., «Outline of a new approach to the analysis of complex systems and decision processes», *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 3, N. 1, (1973), 28-44.

Zadeh, L.A., «A theory of approximate reasonig», en Hayes, J.E., Michie, D. y Mikulich, L.I. (Eds.), *Machine Intelligence*, Vol.9, Elsevier, (1979), 149-194.

RAZONAMIENTO TEMPORAL BORROSO

Roque Marín y Senén Barro
Departamento de Electrónica y Computación. Universidad de Santiago de Compostela

José Mira y Ana E. Delgado
Departamento de Informática y Automática UNED. Madrid

1. INTRODUCCIÓN

La imprecisión y la incerteza aparecen frecuentemente en nuestras expresiones en lenguaje natural y, quizá con más insistencia, en aquellas que usamos para referirnos al tiempo y a la ordenación temporal de los hechos. Esta vaguedad temporal parece ser algo intrínseco a nuestra forma de tratar con la información temporal. Nuestra percepción del tiempo es indirecta. No tenemos ningún órgano que nos permita percibir directamente el tiempo, en una forma análoga a como percibimos el espacio a través de la vista. Nuestra percepción del tiempo se basa en nuestra memoria, y ésta es imperfecta. Nuestros recuerdos tienen una molesta tendencia a degradarse con el paso del tiempo. Así, a la ambigüedad característica del lenguaje natural y a la subjetividad de nuestros mecanismos de razonamiento, añadimos la imprecisión propia de nuestra percepción del tiempo [Thomson, 1985; MacKay, 1987; Ariav, 1988].

Esto obliga a que cualquier modelo computacional de tratamiento de información temporal, que pretenda aproximarse a nuestra forma de referirnos al tiempo, deba proveer mecanismos para representar y manejar la imprecisión de

nuestras expresiones. De hecho, muchos de los modelos de procesado de conocimiento temporal considerados convencionales (no borrosos) incluyen alguna forma más o menos elaborada de tratar con la imprecisión temporal. Es el caso de las aproximaciones de Kahn y Gorry [1977], Dean y McDermott [1987], Dechter y col. [1991] o incluso de la lógica de Allen [1984]. Pero ninguna de ellas alcanza la potencia de la *Lógica Borrosa* en la representación y manipulación de la imprecisión.

La Lógica Borrosa proporciona un modelo formal capaz de manejar la imprecisión y la incertidumbre propias del mundo real, a la vez que permite expresiones cercanas a las del lenguaje natural. Esto es lo que justifica su utilidad en ingeniería del conocimiento y explica el éxito de sus aplicaciones prácticas. Existe, por tanto, un gran interés en aplicar el modelo formal de la Lógica Borrosa al tratamiento de la información temporal imprecisa. Existen ya algunos resultados interesantes en esta línea. En este trabajo revisaremos el estado del conocimiento en el campo, y consideraremos las perspectivas de futuro que se ofrecen. Pero antes es conveniente estudiar el problema del razonamiento temporal desde una perspectiva convencional. Por tanto, tras hacer un planteamiento general del problema, revisaremos algunas de las aproximaciones convencionales más significativas.

2. PLANTEAMIENTO DEL PROBLEMA

Un *Sistema Basado en Conocimiento* (SBC) es un programa que incorpora, en forma computable, el conocimiento y las estrategias de razonamiento adecuadas para resolver problemas en un dominio de aplicación restringido, aparentando un comportamiento inteligente. Cuando el conocimiento en que se basa es de naturaleza empírica, y proviene de la experiencia acumulada por un experto humano, se le llama *Sistema Experto* (SE). La *Ingeniería del Conocimiento* proporciona una metodología de programación peculiar para implementar los SBC. A diferencia de un programa convencional, en un SBC existe una separación entre conocimiento y razonamiento. El primero es representado en forma declarativa en un fichero de texto, mientras el segundo es implementado por un programa de control que selecciona y aplica elementos de la base de conocimiento a los datos conocidos sobre el problema almacenados en una base de datos (fig. 1). El conocimiento puede representarse fácilmente en forma computable mediante *reglas de producción*, es decir, frases de la forma:

Si A es cierto **entonces** B es cierto

y el mecanismo de inferencias puede limitarse a una aplicación repetida de la regla de inferencia *modus ponens*:

p1: A es cierto (Hecho)

p2: Si A es cierto entonces B es cierto (Conocimiento)

.....

c: B es cierto

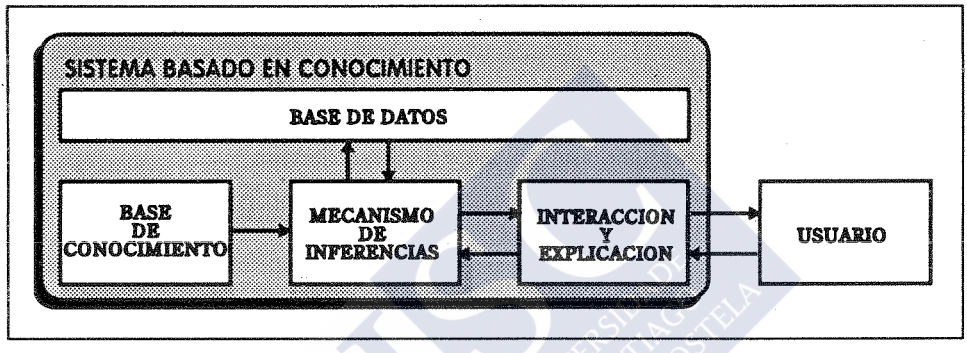


Figura 1. Esquema funcional de un Sistema Basado en Conocimiento.

Sin embargo, no siempre podemos hablar de la verdad o falsedad de un hecho en forma intemporal. Un hecho será cierto durante algún tiempo, y después puede pasar a ser falso. Esto nos va a obligar a complicar el sencillo esquema de la figura 1. Pero veamos primero porqué es necesario añadir una complejidad adicional.

El tratamiento de la información temporal es esencial en muchos dominios de aplicación de los sistemas expertos: medicina, interpretación de lenguaje natural, gestión, robótica, control de procesos, diseño y análisis de sistemas electrónicos digitales, síntesis y verificación de programas para arquitecturas paralelas, etc ... A título de ejemplo consideremos algunas de las necesidades que aparecen en el caso de la Medicina:

a) Cuando el médico razona para establecer un diagnóstico, maneja fundamentalmente información obtenida mediante exploración física en la consulta actual, pero a menudo necesita conocer también información sobre hechos pasados. Así, preguntará al paciente sobre antecedentes clínicos o sobre síntomas subjetivos previos, puede necesitar valorar

analíticas o radiografías anteriores, e incluso, en algún caso, la disposición y ordenación de los hechos sobre el eje del tiempo puede ser determinante. Las figuras 2, 3 y 4 muestran algunos ejemplos.

b) En patologías crónicas o de larga duración (oncología, por ejemplo) es preciso hacer un seguimiento de la evolución de la enfermedad y de la respuesta al tratamiento, para cambiar la terapia en caso necesario (fig. 5).

c) La existencia de un modelo temporal permite hacer análisis de coherencia sobre los datos. Por ejemplo, una diferencia de peso de 20 Kg entre la consulta de la semana pasada y la actual, indica un error en la entrada de datos.

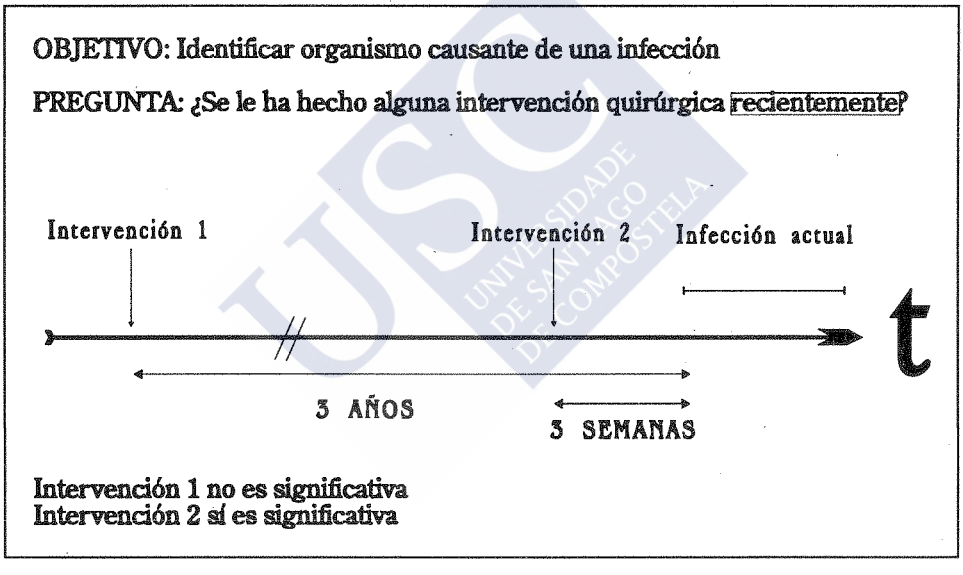


Figura 2. Ejemplo de uso de información pasada en Medicina.

Estas necesidades difícilmente pueden ser satisfechas con un sistema experto convencional que no esté provisto de la capacidad de manejar información temporal. Como paradigma de sistema experto en Medicina podemos considerar al sistema Mycin. Es un sistema de consulta única, incapaz de recordar la información pasada. Supone que el diagnóstico es una tarea relativa a un sistema fisiológico estático, cuya alteración es descrita por un conjunto de manifestaciones que no cambian mientras se mantiene el estado patológico. En los casos en que es necesario

manejar información pasada, se adoptan soluciones ad-hoc. Consideremos de nuevo el ejemplo de la figura 1. Mycin se limita a preguntar al usuario, y deja que sea él quien interprete el contexto temporal al que se hace referencia. Para solucionarlo podríamos añadirle una base de datos histórica en la que se almacenen los datos y las conclusiones de consultas previas. Pero ésto por sí sólo no es suficiente. Necesitaría además algún mecanismo capaz de encontrar relaciones de precedencia entre los hechos, identificar contextos temporales, analizar persistencias, detectar tendencias, etc ...

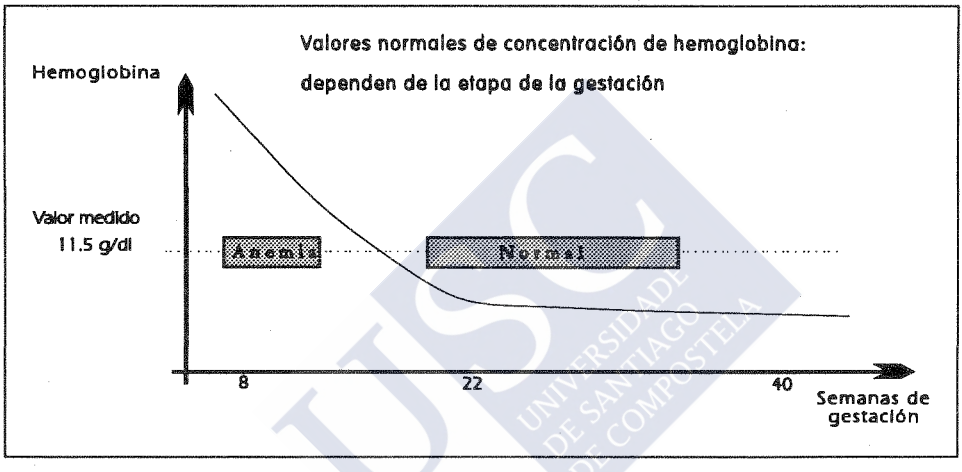


Figura 3. Ejemplo de significado dependiente del contexto en Medicina.

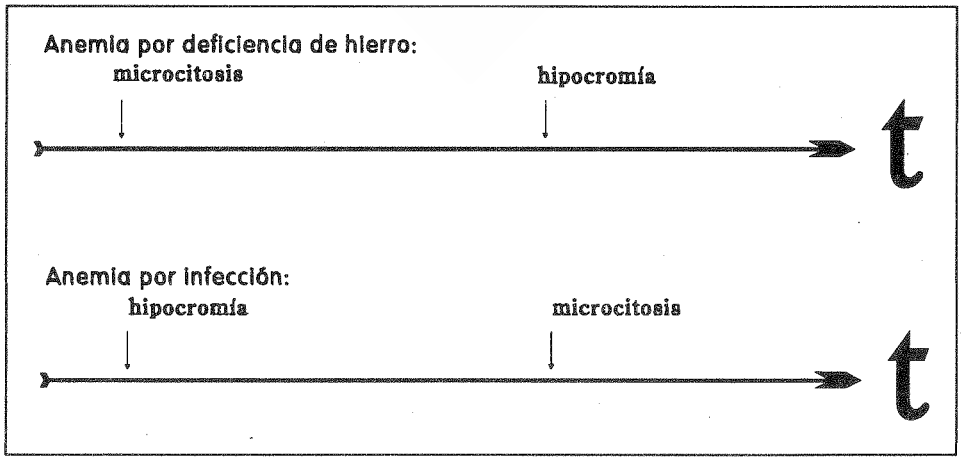


Figura 4. Ejemplo de diagnóstico dependiente de la ordenación temporal.

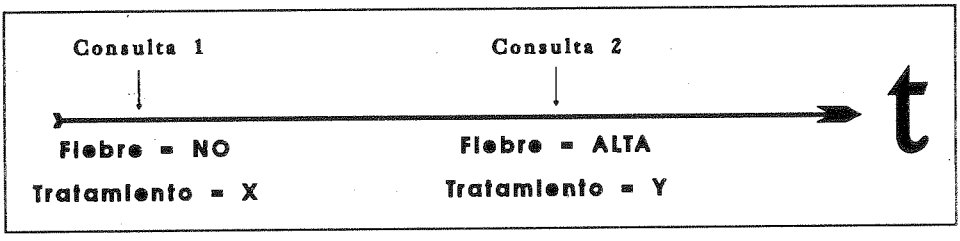


Figura 5. Adaptación del diagnóstico a la evolución del problema.

Se han propuesto numerosos formalismos para hacer computable el tratamiento del tiempo. En una revisión ortodoxa tendríamos que empezar considerando algunas de las aproximaciones clásicas basadas en *lógica de predicados de primer orden*: Lógica Situacional [McCarthy y Hayes, 1969], Especialista Temporal de Kahn y Gorry [1977], Lógica de Relaciones Temporales [Allen, 1984], Cálculo de Eventos de Kowalski [Sadri, 1987], Gestión de Mapas Temporales [Dean y McDermott, 1987]. Haríamos entonces un estudio de algunas *aproximaciones pragmáticas*, es decir, sistemas expertos prácticos que no están estrictamente basados en una aproximación formal como las anteriores, pero que funcionan bien en algún dominio de aplicación restringido: VM [Fagan, 1980]; Oncocin [Kahn y col., 1985]; Topaz [Kahn y col., 1989], o los sistemas de Perkins y Austin [1990], Cousins y col. [1989], Russ [1989] y Cohn y col. [1989]. Podríamos estudiar a continuación las limitaciones de la lógica de primer orden en razonamiento temporal (problemas de la cualificación y de la predicción extendida), consideraríamos las ventajas e inconvenientes de las lógicas temporales modales, y terminaríamos analizando la Lógica de Ignorancia Cronológica de Shoham [1988]. Sin embargo, preferimos hacer una aproximación global al tema desde un punto de vista más general. Para ello, vamos a definir un marco de referencia en el cual podamos encajar los modelos y sistemas más conocidos.

3. REPRESENTACIÓN Y RAZONAMIENTO TEMPORAL

3.1. Esquema General

Para hacer computable el tratamiento de la información dependiente del tiempo el esquema de la figura 1 debe modificarse, añadiendo algunos componentes (fig. 6):

a) Un método para representar y almacenar la ocurrencia de eventos. Necesitaremos una base de datos temporal y un lenguaje para que el usuario pueda introducir información en ella.

b) Un «especialista temporal» que obtenga inferencias de sentido común sobre el tiempo independientes del dominio. Algo tan obvio como la transitividad entre fechas debe ser indicado de forma explícita al sistema:

Si	$t_1 < t_2$
y	$t_2 < t_3$
entonces	$t_1 < t_3$

Para ello necesitaremos una base de conocimiento sobre el tiempo, con axiomas como el anterior, y un mecanismo de inferencias para actualizar la base de datos con las conclusiones obtenidas.

c) Un lenguaje para la representación del conocimiento del dominio, que permita hacer referencias a hechos que ocurren en el tiempo. En el caso más sencillo bastará con reglas de producción de la forma:

Si	A es cierto en t_1
y	B es cierto en t_2
y	.....
entonces	Z es cierto en t

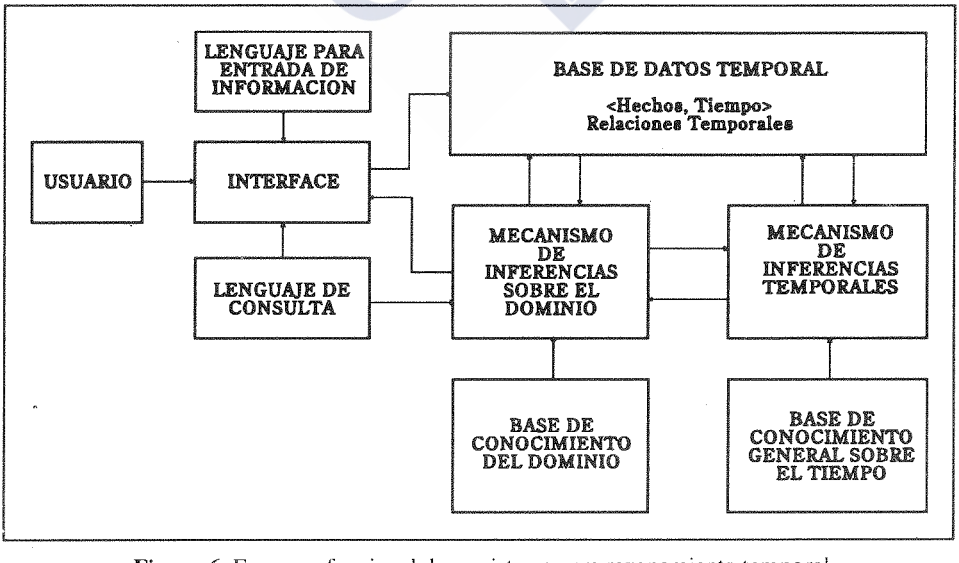


Figura 6. Esquema funcional de un sistema para razonamiento temporal.

Con algunas modificaciones menores de la sintaxis éste es el caso de muchos sistemas expertos prácticos.

Este esquema general admite numerosas variantes en cada uno de los componentes. Dado que representación y razonamiento están profundamente ligados, vamos a estudiar algunas de las posibilidades en representación, al mismo tiempo que consideramos los métodos de razonamiento asociados a cada una.

3.2. Representación Mediante Fechas Absolutas

Cada evento puede representarse en la base de datos como un hecho que tiene asociado un tiempo, expresado éste como una fecha absoluta. Por ejemplo:

s1: Tensión arterial sistólica igual a 120 mmHg el 18/9/1991 13:00

A una base de datos de este tipo, llamada *base de datos histórica*, podemos hacerle preguntas de alguno de los tipos siguientes:

p1: ¿Cual es el valor de la variable v en el tiempo t? (v, ?, t)

p2: ¿Cuando fue el valor de la variable v igual a x? (v, x, ?)

p3: ¿Es el valor de la variable v igual a x en el tiempo t? (v, x, t)

Las preguntas del tipo **p3** corresponden a *predicados temporales*,

ocurre (hecho, tiempo) = T o F

que pueden figurar en la parte de condiciones de las reglas de producción contenidas en la base de conocimientos del dominio.

Las fechas absolutas verifican una relación de orden total, es decir, se pueden representar gráficamente sobre una línea, en correspondencia biunívoca con el conjunto de los números naturales (si usamos tiempos discretos) o con el conjunto de los números reales (si suponemos un eje del tiempo denso). Debido a ello, los algoritmos para insertar y buscar eventos en la base de datos se ejecutan en tiempo lineal, ésto es, consumen un tiempo que es directamente proporcional al número de eventos en la base de datos. Además, el consumo de espacio de almacenamiento es también lineal. Es, por tanto, un problema tratable y muy sencillo.

Sin embargo, este esquema tan simple es demasiado limitado para muchos dominios prácticos, y es necesario extenderlo. Consideremos en más detalle algunos aspectos del determinismo temporal que, aunque son completamente generales, aparecen ya en su forma más simple en este tipo de representación.

3.2.1. Instantes e Intervalos

En algunos SE, todos los eventos se consideran instantáneos y tienen asociada una fecha puntual. Es propio de los sistemas que hacen descripciones basadas en un espacio de estados: a cada estado le corresponde un instante, y el estado permanece sin cambios hasta que algún nuevo evento instantáneo provoca el cambio de estado. Cada estado puede venir dado por una fecha absoluta, o puede ser representado simplemente por un símbolo tal como S_j, pero tanto fechas como estados son puntuales (no se solapan) y se recorren en una forma lineal (la secuencia de estados por la que pasa el problema está en una relación de orden total). Es el caso de la Lógica Situacional [McCarthy y Hayes, 1969], y está en la base de los sistemas expertos VM [Fagan, 1980] o Medtool [Otero, 1991].

En dominios como la Medicina puede plantear problemas: si los datos de analítica se recogen 3 días antes que los demás datos de consulta, tendrán fechas distintas y el sistema no los considerará simultáneos, a pesar de que para el médico los resultados de analítica siguen estando vigentes 3 días después. La solución es forzar al usuario a introducir todos los datos significativos para una consulta con la misma fecha que la consulta. Pero ésto es una solución artificial que impide el acceso automático a la información almacenada en bases de datos externas y, por tanto, la integración del sistema experto en una red de informática hospitalaria. Es más general considerar que algunos eventos pueden extenderse en el tiempo, y representar su intervalo de validez a partir de las fechas inicial y final del intervalo, o a partir de la fecha inicial y la duración del intervalo.

Al permitir que el usuario pueda introducir intervalos aparece un problema nuevo. Si admitimos que una variable puede tomar sólo un valor en un instante de tiempo dado (principio de exclusión), hay que vigilar que los intervalos de tiempo correspondientes a valores distintos de una misma variable no se solapen entre sí. Esto es una tarea para el módulo que hemos llamado especialista temporal, y puede ser implementada mediante un sencillo algoritmo que se ejecuta en tiempo constante (independiente del número de eventos en la base de datos): para ver si un nuevo intervalo es consistente basta ver si se superpone con los dos intervalos que comienzan inmediatamente antes e inmediatamente después de su punto de inicio.

3.2.2. Persistencia

En muchos casos los valores de las variables de interés se obtienen mediante procesos de medida. La medida puede ser ocasional (por ejemplo, una analítica en Medicina) o prefijada dentro de un plan de monitorización (por ejemplo, la monitorización de una señal ECG). En cualquier caso la medida es instantánea, pero podremos suponer que su valor se mantiene sin variación durante un cierto tiempo. Suponer persistencias es lo que hace el sistema de Perkins y Austin [1990], que diagnostica averías en el mecanismo de control de posición del telescopio espacial Hubble, a partir de los datos de telemetría provistos por 32 sensores. La figura 7 muestra los 4 tipos de persistencia que maneja. Cuando se hacen suposiciones sobre la persistencia pueden aparecer graves problemas de consistencia.

Figura 7. Tipos de persistencias en el sistema de Perkins y Austin.

Para analizar los problemas que pueden surgir vamos a considerar la cuestión desde una perspectiva más amplia. Consideremos el ejemplo de la figura 8.a: la regla de producción mostrada, que está basada en nuestro conocimiento de la física, nos permite obtener unas conclusiones para el instante t_2. Pero esta inferencia puede ser falsa si el plano está inclinado, si hay viento fuerte (fig. 8.b), si hay rozamiento, si hay un objeto interpuesto en el camino, etc... Estas situaciones pueden ser previstas en t_1, pero puede haber un número potencialmente *infinito* de tales situaciones indeseables, y todas deberían ser previstas en la parte de condiciones de la regla. Es decir, necesitaríamos una regla de tamaño infinito, lo que es obviamente intratable. Este es el conocido *problema de la cualificación*. Además, puede haber otros eventos indeseables que no son previsibles en t_1, y de los que sólo tendremos conocimiento después de t_1: la inferencia es falsa si alguien retira el objeto antes de llegar a p_2, si choca con otro objeto no previsto (fig. 8.c), si el objeto contiene un explosivo y estalla antes de tiempo, etc... Es lo que Shoham [1988] llama *problema*

de la predicción extendida. El *problema de la persistencia* (Frame Problem) es sólo un caso particular de éste, en el que se trata de ver hasta cuando se puede predecir que una situación permanece sin cambios.

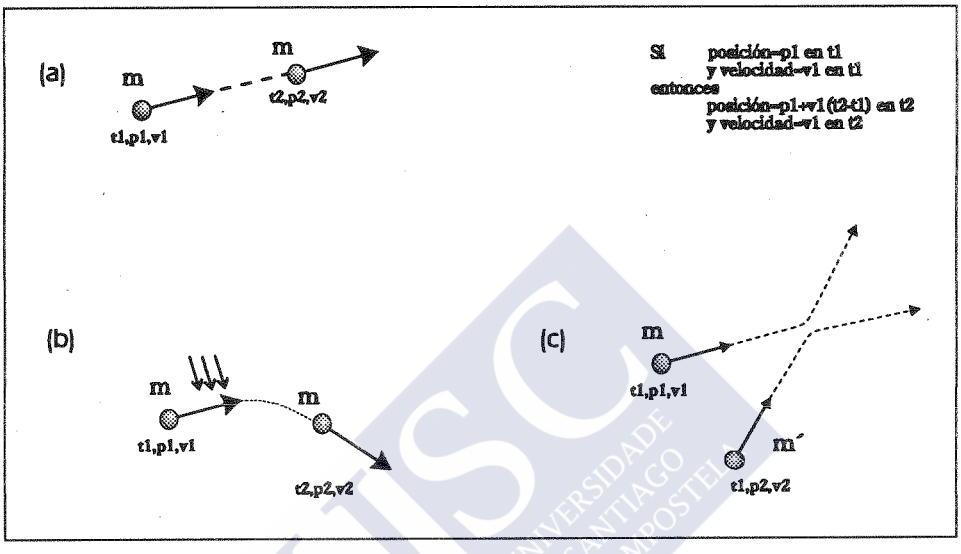


Figura 8. Problemas de la cualificación y de la predicción extendida.

El problema se puede soslayar en dominios de aplicación restringidos en los que el número de variables a considerar está muy limitado, como es el caso del sistema de control de posición del Hubble citado antes. Si el número de variables es pequeño, las reglas tendrán un tamaño manejable. No obstante, para garantizar que todas las conclusiones obtenidas sean válidas hay que imponer algunas restricciones: que las persistencias sólo se extiendan hasta el instante presente, que no se intente obtener ninguna conclusión sobre el futuro (predicciones), y que se pueda considerar que la entrada de datos es absolutamente fiable, de forma que no sea necesario corregir datos erróneos introducidos en el pasado.

En el caso general, la única solución viable consiste en arriesgarse a hacer suposiciones sobre la persistencia basadas en conocimiento incompleto, y en dotar al sistema de un mecanismo que le permita restaurar la consistencia cuando se demuestre que una suposición previa era errónea. Esto es algo que la lógica de predicados de primer orden, siendo una lógica monótona, no permite. El problema

no es trivial, porque pueden haberse obtenido numerosas falsas conclusiones a partir de la suposición errónea. Habría que reconstruir todas las cadenas de razonamiento seguidas a partir de la suposición falsa, y corregir todas las falsas conclusiones derivadas de la primera. Una solución la constituyen las lógicas modales. Un ejemplo específicamente desarrollado para resolver los problemas del razonamiento temporal es la *Lógica de Ignorancia Cronológica de Shoham*. No obstante, las lógicas modales son todavía engorrosas de manejar en la práctica. Una alternativa menos drástica consiste en añadir a la lógica de predicados un algoritmo que, basándose en una representación explícita de todas las dependencias entre variables, permita restaurar la consistencia, rehaciendo sólo las cadenas de razonamiento implicadas. Esto es lo que hacen los *Sistemas de Mantenimiento de la Certeza* (TMS), y que Russ [1989] adaptó al manejo de información relacionada con el tiempo.

3.2.3. Granularidad

Las representaciones temporales deben ser capaces de encapsular intervalos de tiempo más o menos largos, dependiendo del contexto (de picosegundos en Microelectrónica a millones de años en Paleontología), permitiendo ajustar la granularidad al nivel deseado. Mediante la granularidad variable se puede representar una forma elemental de imprecisión. Decir que una infección comenzó el mes pasado equivale a decir que comenzó en algún instante entre las 0 horas del día 1 y las 24 horas del día 31, y da cuenta de nuestro falta de conocimiento del instante preciso. Un ejemplo de sistema en el que ésto es posible es el de Sarda [1990]. No existen sistemas que hagan un tratamiento formal de la imprecisión por granularidad con el soporte de la lógica borrosa.

3.2.4. Otras escalas de tiempo

La representación de una fecha en la forma *año:mes:día:minuto:segundo* no es la única que podemos usar para fijar un instante de tiempo absoluto. De hecho, en Medicina es poco útil y es más frecuente encontrar ejemplos como:

s2: Fiebre era alta *ayer*

s3: Tensión arterial sistólica era 120 mmHg *en la 1ª consulta*

s4: Diámetro craneal era 88 mm *en la 36ª semana de gestación*

Esto equivale a una representación del tiempo en forma implícita:

s3: Tensión arterial sistólica era 120 mmHg *en el tiempo que correspondió a la 1ª consulta*

y es fácil de convertir a la escala de fechas absolutas [Marín y col., 1990]. En Oncocin [Kahn, 1985], por ejemplo, se manejan distintas escalas de tiempo que mantienen una relación jerárquica entre sí. Las relaciones jerárquicas se representan en la base de datos mediante un grafo (fig. 9). Almacenar las relaciones en la base de datos hace muy eficiente la búsqueda de equivalencias temporales, pero a costa del espacio de almacenamiento.

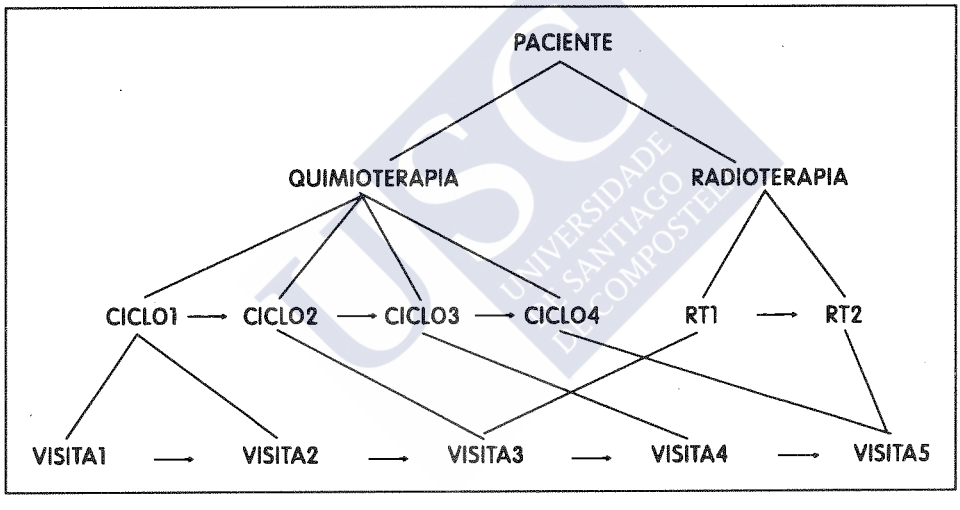


Figura 9. Ejemplo de grafo en el sistema Oncocin.

El uso de escalas relativas no supone ninguna diferencia esencial con respecto al tratamiento de fechas absolutas, siempre que se mantenga una relación de orden total entre los puntos que componen cada escala. Podemos manejar incluso *seudofechas*, asignando a cada evento un número arbitrario, de forma que la ordenación de los números refleje la ordenación temporal de los eventos. Perdemos así toda la información sobre tiempos absolutos y sobre duraciones entre eventos, pero podemos responder todavía a preguntas sobre relaciones temporales entre eventos (el evento nº 1 ocurrió antes que el evento nº 10). Las seudofechas nos permiten otra forma de tratar con la imprecisión.

3.2.5. Lenguajes de representación

Cuando se tienen en cuenta todas las consideraciones anteriores, la tarea de definir una sintaxis para el lenguaje de entrada de información empieza a adquirir entidad propia. El interface debe permitir la distinción entre instantes e intervalos, el ajuste de la granularidad, la introducción de distintos tipos de persistencias, y el uso de escalas relativas. Mayor complejidad aparece en el lenguaje de representación del conocimiento del dominio. Por ejemplo, Oncocin permite expresiones muy flexibles basadas en la sintaxis (<función> <atributo> <tiempo>). Así, para preguntar por los valores de leucocitos medidos durante todas las consultas correspondientes a los tres primeros ciclos de quimioterapia, se usa la expresión:

s5: (valor.de leucocitos (ciclo (1 3)))

Ejemplos adicionales de expresión del tiempo son:

s6: (visita (-4 -1)): cuatro últimas visitas

s7: (visita primera (ciclo (1 -2))): la primera visita en cada uno de los ciclos desde el primero al penúltimo.

Otras posibilidades de interés en Medicina [Marín y col., 1990] son la referencia a tiempos implícitos:

s8: Hemoglobina era menor que 11 g/dl *cuando Leucocitos era mayor que* 9000 μl^{-1}

y el uso de cuantificadores temporales (alguna vez, todas las veces):

s9: Linfocitos era mayor que 2500 *en todas las consultas previas.*

Un trabajo relacionado es el de Sager y col. [1987] sobre interpretación de textos médicos escritos en lenguaje natural.

3.2.6. Relaciones temporales

Hasta ahora hemos considerado sólo predicados temporales como **p3**, en los que se pregunta si un hecho ocurrió en un determinado tiempo absoluto, y reglas de producción cuyas condiciones tienen esa misma forma. El lenguaje de representación gana mucho en expresividad cuando permite manejar relaciones temporales entre instantes e intervalos. Preguntas como:

s10: ¿Pepe dejó la sección de ventas antes que Juan fuese nombrado director?

pueden ser respondidas mediante una simple comparación entre las fechas correspondientes a los dos eventos ($t_1 < t_2$). Dado que tratamos con fechas absolutas, el especialista temporal puede responder a este tipo de preguntas mediante algoritmos que se ejecutan en tiempo constante.

3.3. Representación Mediante Relaciones Temporales Cualitativas

En muchos dominios, las fechas absolutas en que ocurren los eventos son desconocidas, pero sí se conocen sus relaciones temporales, y ésto es lo único que podemos introducir en la base de datos. Existen varios modelos formales que manejan relaciones temporales, y ofrecen una solución convencional al tratamiento de la imprecisión, sin recurrir a la lógica borrosa. El ejemplo ya clásico es la Lógica de Relaciones Temporales de Allen [1984]. Se basa exclusivamente en intervalos, y no considera instantes.

Allen identifica 7 relaciones temporales (fig. 10), predicados que aplicados a 2 intervalos expresan todas las posibles relaciones temporales entre ellos. Forman un conjunto completo, de forma que cualquier otra relación temporal se puede expresar a partir de estas relaciones primitivas. Un ejemplo es la relación *EN*:

$$\text{EN}(t_1, t_2) \Leftrightarrow \text{DURANTE}(t_1, t_2) \vee \text{COMIENZA}(t_1, t_2) \vee \text{TERMINA}(t_1, t_2)$$

Junto con las relaciones inversas son excluyentes. Dos intervalos cualesquiera estarán en una y sólo una de esas relaciones, pero, como en el ejemplo anterior, usamos la disyunción lógica para expresar nuestro desconocimiento acerca de cual es precisamente la relación en que están t_1 y t_2.

Allen extiende la lógica de primer orden mediante la introducción del predicado SE-MANTIENE (HOLDS), que permite definir hechos que ocurren en el tiempo. Dada una propiedad **p** y un intervalo de tiempo **t**, el predicado SE-MANTIENE(p, t) expresa que la propiedad **p** es cierta a lo largo de todo el intervalo **t**. Es decir, el tiempo se maneja como una variable independiente más, ortogonal a las otras variables a las que se aplican los predicados. A partir de los hechos y de las relaciones almacenados en la base de datos, puede responder a preguntas tales como: ¿es cierto que el hecho $\mathbf{p}_1$ ocurrió antes que el hecho $\mathbf{p}_7$? ¿es cierto que el hecho $\mathbf{p}_1$ se solapa con el hecho $\mathbf{p}_3$?

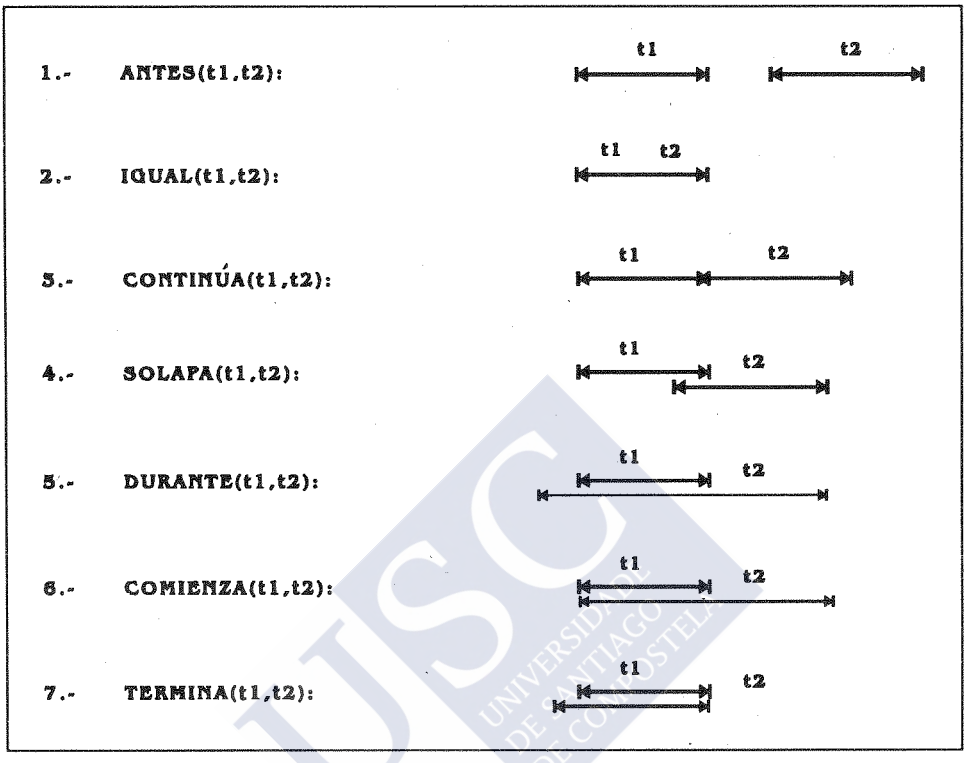


Figura 10. Relaciones temporales primitivas de Allen.

Para responder a estas preguntas el especialista temporal debe hacer inferencias relativas al tiempo, deduciendo nuevas relaciones temporales a partir de las ya conocidas. Estas inferencias se basan en axiomas como el de la transitividad:

$$\text{ANTES}(t_1, t_2) \wedge \text{ANTES}(t_2, t_3) \Rightarrow \text{ANTES}(t_1, t_3)$$

o el de convexidad de los intervalos temporales:

$$\text{SE-MANTIENE}(p,T) \Leftrightarrow \forall t_1\ [\text{EN}(t_1,T) \Rightarrow \text{SE-MANTIENE}(p,t_1)]$$

Obsérvese que todas las relaciones son cualitativas: no es necesario conocer tiempos absolutos para hacer inferencias sobre relaciones temporales.

En un sistema basado en un modelo de este tipo, la base de datos puede representarse mediante un grafo (fig. 11.a). Un grafo de relaciones permite capturar

situaciones en las que tenemos un conocimiento incompleto: en el ejemplo de la figura anterior sabemos que t_1 es anterior a t_2 y a t_3, pero no conocemos en qué relación están t_2 y t_3. Como contrapartida, hemos perdido la sencillez de manejo de las representaciones lineales del apartado anterior.

A partir del grafo, el especialista temporal puede extraer más información de la que está explícitamente representada, aplicando axiomas como el de la transitividad (fig. 11.b), que propagan las restricciones inicialmente introducidas. Es conveniente que la actualización del grafo se haga de forma incremental, es decir, que el algoritmo de propagación de restricciones se ejecute cada vez que el usuario introduce una nueva información en la base de datos. Así, se podrán detectar inmediatamente las posibles inconsistencias, como en la figura 11.c. Por tanto, es importante encontrar algoritmos de actualización incremental y detección de inconsistencias que se ejecuten lo más rápidamente posible.

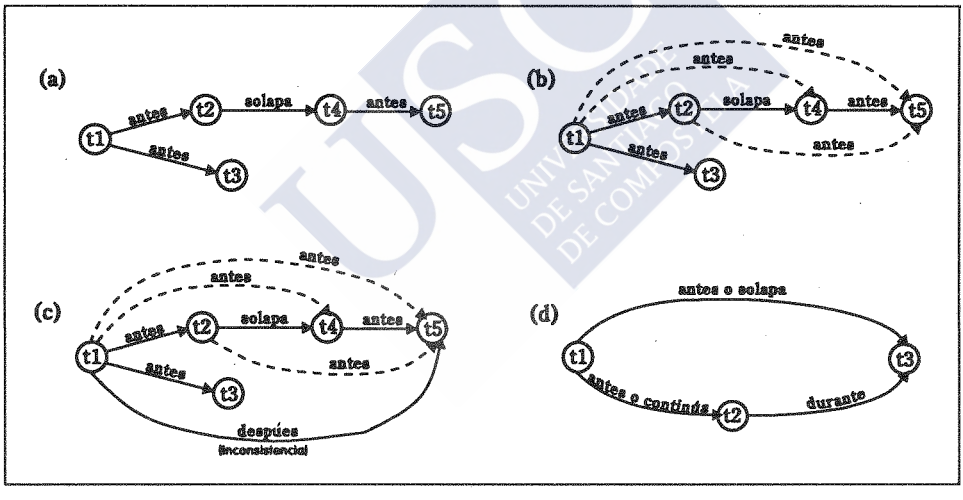


Figura 11. Grafos de relaciones temporales cualitativas.

Cuando cada arco tiene una sola etiqueta no hay demasiados problemas. Pero el modelo de Allen permite incertidumbre en las relaciones temporales, expresada mediante disyunciones entre varias relaciones primitivas (fig. 8.d), y en este caso los algoritmos son intratables (el tiempo necesario para detectar una inconsistencia crece exponencialmente con el número de nodos). Aún así, existen algoritmos de tipo heurístico que son de orden n^3, siendo n el número de nodos, y que resuelven

bien los casos más usuales, aunque no garantizan que se obtenga siempre una solución. En la actualidad se trabaja en el estudio de las restricciones que hay que imponer al caso general para obtener algoritmos que sean fiables y útiles en la práctica [Allen, 1991].

La Lógica de Relaciones Temporales es una extensión de la lógica de predicados de primer orden y, por tanto, tiene todas las limitaciones de una lógica monótona. Los problemas estudiados en el apartado anterior siguen siendo válidos aquí. Tiene también limitaciones de tipo semántico. Lo que se añade a la lógica de primer orden es: objetos temporales (intervalos), predicados que relacionan los objetos temporales entre sí (relaciones temporales) y predicados que relacionan objetos temporales con otros objetos (como el predicado SE-MANTIENE). Es decir, el tiempo se maneja como un objeto más de la lógica. Esto significa que los predicados, que corresponden a verbos en lenguaje natural, son intrínsecamente atemporales. No hay información temporal implícita en el predicado, a diferencia de lo que ocurre en nuestro lenguaje natural, en el que los verbos admiten tiempos verbales. Esta forma de tratar el tiempo, como una variable independiente a la que se aplican predicados, es característica de cualquier modelo temporal basado en la lógica de predicados de primer orden. En particular, es característica de la Física, pues su herramienta de razonamiento es el cálculo de predicados. Para poder manejar algo parecido a los tiempos verbales, hay que pasar a las lógicas modales temporales, y las existentes actualmente no alcanzan la riqueza semántica del lenguaje natural.

3.4. Representación Mediante Relaciones Temporales Cuantitativas

Las relaciones temporales cualitativas son demasiado ambiguas para muchas aplicaciones prácticas. Decir que t_1 es anterior a t_2 supone una imprecisión total en la localización de t_1 y t_2, y una imprecisión casi total en la distancia temporal entre ambos: lo único que sabemos es que la distancia entre t_1 y t_2 es positiva, pero podría ir de cero a infinito. Entre los dos tipos de representación estudiados existe un caso intermedio mucho más cercano a las situaciones reales. Es el caso de las representaciones basadas en relaciones temporales cuantitativas, a las que Dean y McDermott [1987] llamaron *Mapas Temporales*.

En un mapa temporal (fig. 12.a) los nodos representan instantes de tiempo y los arcos establecen restricciones temporales, acotando la distancia temporal entre los nodos mediante un valor máximo y un valor mínimo. En el ejemplo, t_2

ocurre entre 20 y 30 unidades de tiempo después que t_1, lo que se puede representar mediante la desigualdad:

$$20 \leq (t_2 - t_1) \leq 30$$

Estas restricciones temporales permiten representar la imprecisión, pero todavía conservan información de tipo métrico, y son más generales que las relaciones cualitativas. Por ejemplo, la relación *antes* puede representarse como la restricción $[0, \infty]$.

La gestión de este tipo de grafos es un problema tratable. El especialista temporal propaga las restricciones y busca todas las posibles soluciones, es decir, todas las posibles asignaciones de fechas a nodos que son compatibles con las restricciones. Este algoritmo, que además detecta inconsistencias, puede ejecutarse en tiempo polinomial (n^3).

Para aplicaciones como planificación, necesitamos una aproximación más general que nos permita especificar disyunciones entre restricciones. Una frase como la siguiente:

> «La duración de las obras de construcción del puente será de 8 a 12 meses si hace buen tiempo, o de 15 a 24 meses si llueve»

tiene que ser representada mediante un arco con dos etiquetas alternativas (fig. 12.b). Desgraciadamente, el problema se hace intratable, necesitándose tiempos exponenciales para resolverlo. Sin embargo, para aplicaciones prácticas en las que la topología del grafo satisface ciertas condiciones, se pueden encontrar soluciones en tiempo polinómico, como las propuestas por Dechter y col. [1991].

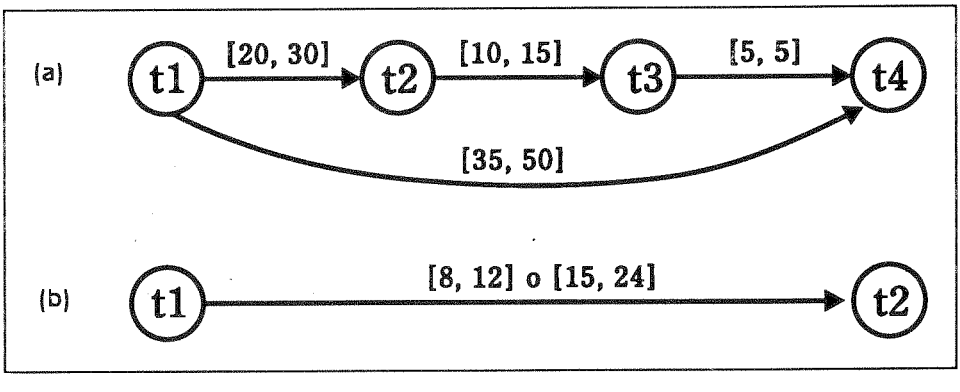


Figura 12. Grafos de relaciones temporales cuantitativas.

4. SOLUCIONES BASADAS EN LÓGICA BORROSA

4.1. Perspectiva General

Como hemos visto, las aproximaciones convencionales para representación y razonamiento temporal incluyen formas de tratar con la imprecisión. Sin embargo, están limitadas por la carencia de un modelo formal específicamente orientado al tratamiento de los distintos aspectos de la imprecisión y la incertidumbre. En el mejor de los casos, se puede representar el margen de error asociado a una estimación de distancia temporal mediante un par de desigualdades, y se pueden propagar estos errores mediante un mecanismo aditivo simple. En esencia, ésto es equivalente al tratamiento de los errores de medidas en las técnicas experimentales convencionales.

Para ir más allá de estas limitaciones, y acercarse más a los problemas reales que se presentan en dominios prácticos, es necesario recurrir a algún modelo formal capaz de representar creencias subjetivas, capaz de ofrecer un tratamiento diferenciado de la imprecisión y la incertidumbre, y capaz de manejar y combinar el soporte evidencial de las hipótesis imprecisas e inciertas. La Lógica Borrosa es uno de los modelos de razonamiento aproximado que cumple estas características, y además añade la capacidad de modelar expresiones lingüísticas llenas de vaguedad. Justamente, la representación computacional de adverbios imprecisos (cuantificadores como «muchos», «bastantes», «algunos», ... o cualificadores como «muy», «poco», «más o menos», ...) es una de sus características distintivas [Zadeh, 1984].

El primero en proponer la representación de un intervalo de tiempo como un conjunto borroso fue Vitek [1983]. Prácticamente todos los trabajos posteriores se han orientado a una generalización a relaciones borrosas de las relaciones temporales cualitativas de Allen aplicadas al tiempo. Lo único reseñable en esta línea es la aproximación de Dutta [1988a, 1988b] para tratamiento de eventos borrosos, y la de Kim y Oh [1991], que considera instantes e intervalos de tiempo borrosos. Dubois y Prade [1989] proponen un modelo de tipo más general, basado en la teoría de la posibilidad de Zadeh. Su trabajo pone los fundamentos para cualquier desarrollo posterior, y abre el camino para la representación y el manejo de relaciones temporales de tipo métrico expresadas en forma borrosa.

4.2. Lógica Temporal Borrosa de Dutta

El modelo temporal borroso de Dutta segmenta el eje del tiempo en intervalos precisos disjuntos (fig. 13.a):

$$I = \{i_1, i_2, ..., i_n\}$$

y representa un evento *e* como un subconjunto borroso del universo I:

$$e = \{<i_1, \mu_e(i_1)>; < i_2, \mu_e(i_2)>; ..., < i_n, \mu_e(i_n)>\}$$

La función de pertenencia $\mu_e(i_j)$ representa la *posibilidad* de que el evento *e* ocurra en el intervalo i_j (fig. 13.b), aunque en las aplicaciones propuestas por Dutta realmente se interpreta como intensidad de ocurrencia. Un ejemplo de aplicación al dominio médico podría ser el siguiente: un paciente relata dos síntomas, dolor desde hace unas tres semanas creciendo progresivamente en intensidad (e_1) y fiebre desde hace una semana siendo hoy más intensa (e_2). Para representarlo, se definen arbitrariamente tres intervalos precisos de tiempo:

i_1: las dos primeras semanas
i_2: los siguientes seis días
i_3: hoy

y se define la extensión temporal de los eventos sobre este conjunto de intervalos:

	i_1	i_2	i_3
e_1	0.2	0.7	0.9
e_2	0.0	0.3	0.8

A partir de este tipo de representación el modelo introduce relaciones temporales entre eventos, como las de Allen: antes, después, continúa, solapa. Las definiciones se basan en la comparación entre los segmentos de tiempo que componen el soporte de cada evento. Así, el grado $\mu_<(e_1, e_2)$ en que un evento e_1 es anterior a otro e_2, viene dado por:

$$\mu_<(e_1, e_2) = \text{máx}_{i_j \in e_1,\, i_k \in e_2,\, i_j < i_k} \text{mín}\ \{\mu_{e_1}(i_j), \mu_{e_2}(i_k)\}$$

es decir, depende de los grados de pertenencia a e_1 y e_2 de parejas de intervalos precisos que verifican la relación precisa *antes*. En el ejemplo anterior sólo tenemos las parejas:

(i_1, i_2): mín (0.2, 0.3) = 0.2
(i_1, i_3): mín (0.2, 0.8) = 0.2
(i_2, i_3): mín (0.7, 0.8) = 0.7

por tanto,

$$\mu_{<}(e_1, e_2) = \text{máx } (0.2, 0.2, 0.7) = 0.7$$

El modelo de Dutta presenta algunas limitaciones importantes. Como hemos visto, evita representaciones continuas del grado de ocurrencia de un evento mediante una discretización precisa del eje del tiempo, obedeciendo a razones de eficiencia computacional. Sin embargo, la granularidad de esta partición del tiempo es crítica y afecta considerablemente a los resultados. Además, como los segmentos son los mismos para todos los eventos, la segmentación debe elegirse globalmente, de forma que sea suficiente para capturar las variaciones en μ del evento más rápidamente cambiante. Esto supone que hay que conocer todos los hechos antes de segmentar el eje del tiempo.

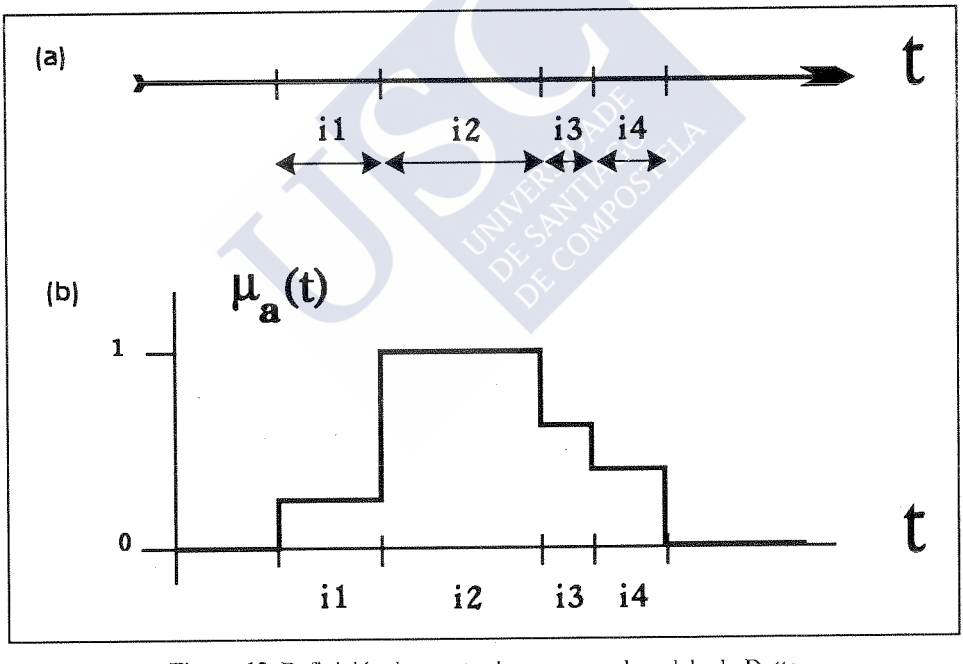


Figura 13. Definición de eventos borrosos en el modelo de Dutta.

Por otra parte, el modelo no distingue entre instantes e intervalos. En consecuencia, la duración de la ocurrencia de un evento no tiene representación borrosa. Lo que podría interpretarse como duración de *e* es la extensión del soporte temporal de *e*, dada por $(i_n - i_1)$, pero ésto es una magnitud precisa. Una limitación

adicional es que no permite ni la representación de relaciones métricas (como «e_1 ocurrió unas 3 horas antes que e_2»), ni la representación de modificadores lingüísticos de las relaciones (como «mucho antes», «poco después,...).

4.3. Modelo de Dubois y Prade

El modelo de Dubois y Prade, de tipo posibilístico, tiene una orientación más general, y pretende proporcionar una base sólida para la mayoría de las cuestiones de interés en representación y razonamiento temporal impreciso. Permite tanto el manejo de tiempos absolutos, como la representación y manipulación de relaciones temporales cualitativas.

Una fecha borrosa **a** es definida mediante un subconjunto borroso **A** del eje del tiempo **T** (fig. 14.a). De acuerdo con la interpretación estándar en Teoría de la Posibilidad, este subconjunto borroso induce una distribución de posibilidad π_a, es decir, una función que representa los valores más o menos posibles de la fecha. Dado un instante preciso t_1, la función $\pi_a(t_1)$ nos da la posibilidad entre 0 y 1 de que la fecha **a** sea justamente el instante preciso t_1, y es definida como el grado de pertenencia de t_1 a **A**:

$$\forall\, t \in \mathbf{T},\ \pi_a(t) = \mu_A(t)$$

Se consideran conjuntos borrosos normalizados y unimodales. Los elementos del conjunto **A** representan alternativas mutuamente excluyentes para **a**; es decir, se trata de un conjunto *disyuntivo*.

Un intervalo borroso [**a**, **b**] está asociado a dos fechas borrosas **a** y **b**. A partir de ellas podemos definir el conjunto de puntos que están posiblemente entre **a** y **b**, es decir [**A**, **B**], como la intersección del conjunto de puntos que están posiblemente después de **a** y el conjunto de puntos que están posiblemente antes de **b**:

$$\mu_{[\mathbf{A},\,\mathbf{B}]}(t) = \text{mín}\ \{\sup_{s_a \le t} \pi_{\mathbf{a}}(s_a),\ \sup_{s_b \ge t} \pi_{\mathbf{b}}(s_b)\}$$

Gráficamente, la función de pertenencia que define al intervalo viene dada por la envolvente de las funciones de pertenencia de las fechas **a** y **b** (fig. 14.b). De forma análoga puede definirse el conjunto de puntos]**A**, **B**[que están necesariamente entre **a** y **b**:

$$\mu_{]\mathbf{A},\,\mathbf{B}[}(t) = \text{mín}\ \{1\text{-}\sup_{s_a > t} \pi_{\mathbf{a}}(s_a),\ 1\text{-}\sup_{s_b < t} \pi_{\mathbf{b}}(s_b)\}$$

En ambos casos, no se trata de conjuntos disyuntivos, pues todos los puntos del conjunto pueden pertenecer simultáneamente al intervalo.

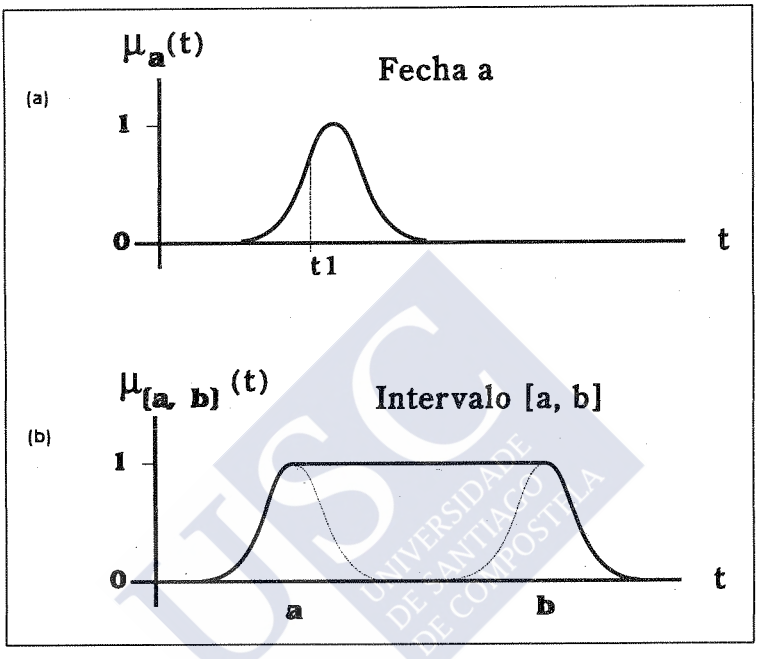


Figura 14. Fechas e intervalos borrosos en el modelo de Dubois y Prade.

Una vez introducidas las representaciones de fechas e intervalos, es fácil definir relaciones temporales entre fechas, y a partir de ellas, relaciones temporales entre intervalos. Para conocer en qué medida dos fechas **a** y **b** satisfacen una relación temporal como «antes» ($\mathbf{a} \leq \mathbf{b}$), hacen falta dos números: una medida de posibilidad Π entre 0 y 1, que estima la posibilidad de que **a** sea anterior a **b**, y una medida de necesidad N, que estima el grado de certeza. Como es norma en Teoría de la Posibilidad, para tener alguna certeza en algo ($N \neq 0$), es imprescindible que ese algo sea completamente posible ($\Pi = 1$). Para la relación temporal «antes» estas medidas se definen como:

$$\Pi (\mathbf{a} \leq \mathbf{b}) = \sup_{s \leq t} \text{mín } \{\pi_a(s), \pi_b(t)\}$$

$$N (\mathbf{a} \leq \mathbf{b}) = 1 - \Pi (\mathbf{a} > \mathbf{b}) = 1 - \sup_{s > t} \text{mín } \{\pi_a(s), \pi_b(t)\}$$

La figura 15 muestra algunos ejemplos.

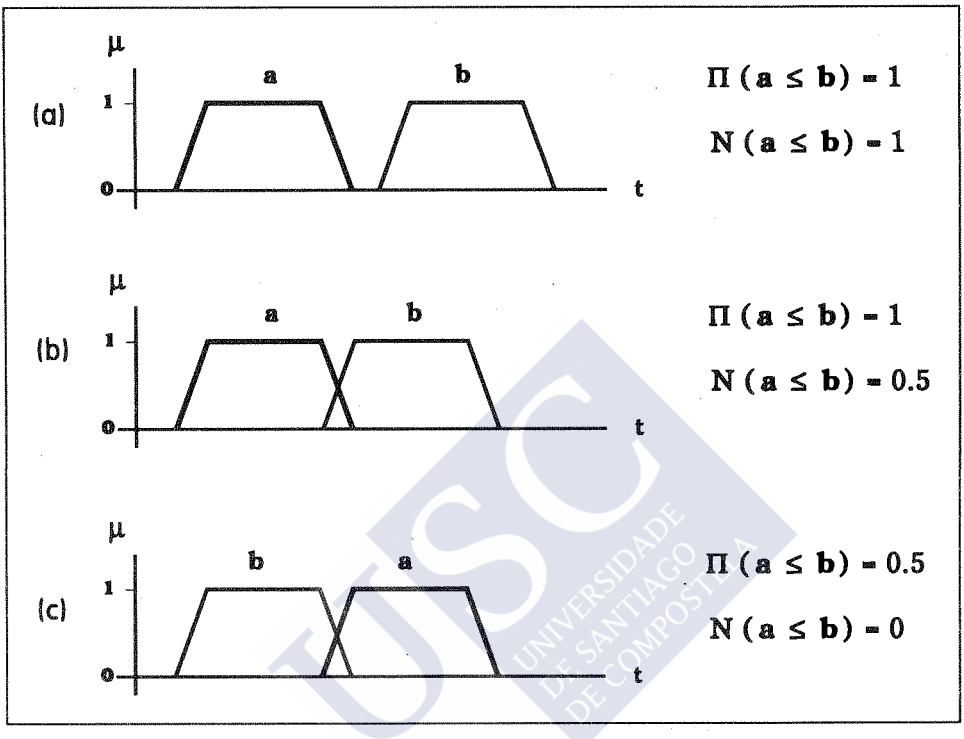


Figura 15. Relaciones temporales cualitativas entre fechas borrosas.

Las relaciones temporales entre intervalos se definen como relaciones entre las fechas extremas de los intervalos. Así, por ejemplo, el intervalo [**a**, **b**] está «antes» que el [**c**, **d**] si y sólo si el instante final del primero está antes que el instante inicial del segundo:

Π ([**a**, **b**] antes [**c**, **d**]) = Π (**b** ≤ **c**)

N ([**a**, **b**] antes [**c**, **d**]) = N (**b** ≤ **c**)

El modelo considera también la aplicación de modificadores lingüísticos a las relaciones temporales (e.g. «mucho antes», «poco después») y aplica la aritmética borrosa al cálculo de duraciones, a partir de lo cual es factible la introducción de relaciones temporales cuantitativas. Sin embargo, es un trabajo de carácter básico que es necesario desarrollar, y en algún caso simplificar, desde la

perspectiva de las aplicaciones prácticas. Un ejemplo en esta línea es el trabajo de Console y col. [1991]. Siguiendo a Dubois y Prade, consideran un caso particular de definición de intervalos borrosos que simplifica el tratamiento posterior de la información temporal. Su aproximación es aplicada al razonamiento diagnóstico basado en modelos causales.

4.4. Redes Temporales Borrosas

Uno de los temas de mayor interés que se derivan del modelo de Dubois y Prade es la representación y manipulación de redes temporales borrosas basadas en relaciones métricas. Como hemos visto antes, las representaciones convencionales mediante relaciones métricas (mapas temporales de Dean y McDermott, y redes temporales de Dechter y col.) engloban a los modelos basados en relaciones cualitativas como las de Allen, y a los modelos exclusivamente basados en fechas absolutas. Es posible una extensión a relaciones métricas borrosas, que generaliza el tratamiento convencional de la imprecisión que se hace en los mapas temporales y, por tanto, engloba a todos los restantes modelos convencionales y borrosos. Además, es la aproximación deseable en dominios como la Medicina, en los cuales el usuario ha de introducir de forma indistinta fechas absolutas y desplazamientos temporales relativos. Comentaremos aquí algunos de los conceptos básicos de esta aproximación, desarrollada por nuestro grupo de trabajo [Barro y col., 1992].

El punto de partida es la definición de distancias entre fechas, como forma de representar duraciones. Si suponemos tiempos discretos, la distancia entre dos fechas **a** y **b** viene dada por un subconjunto borroso del conjunto de los números enteros **I**, obtenido como diferencia borrosa entre los conjuntos **A** y **B** que representan a las fechas:

$$\mathbf{D} = \mathbf{B} \ominus \mathbf{A}$$

Conocidas las distribuciones de posibilidad de **a** y **b**, se puede obtener la distribución de posibilidad de la distancia como:

$$\forall\, n \in I,\ \pi_{d(a,\,b)}(n) = \sup_{n=t-s} \text{mín}\ \{\pi_a(s), \pi_b(t)\}$$

Si **a** y **b** son los extremos de un intervalo, d(**a**, **b**) representa la duración del mismo. La figura 16 muestra algunos de los casos posibles.

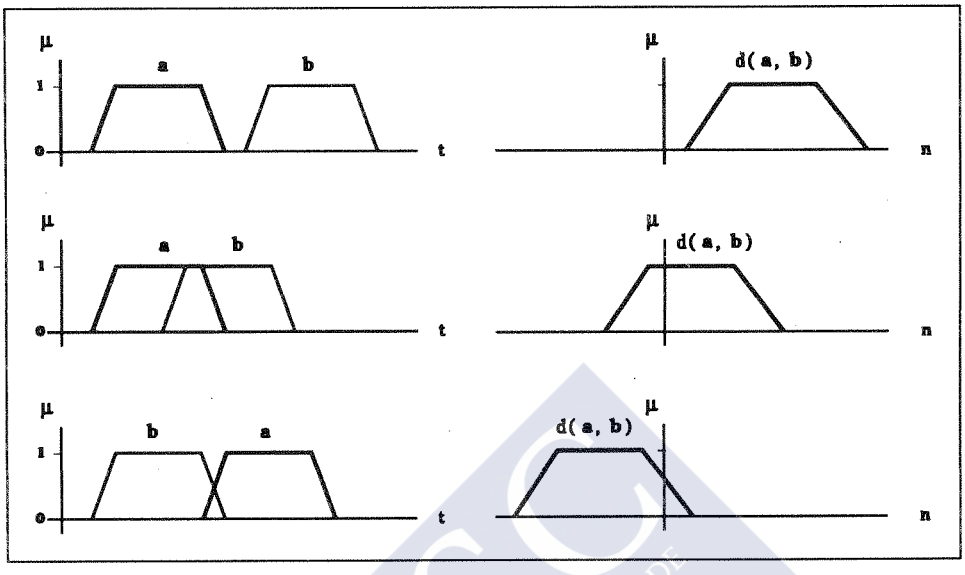


Figura 16. Ejemplos de distancias entre fechas borrosas.

Términos como «duró aproximadamente 1 hora», «duró entre 20 y 30 horas», «duró mucho tiempo», «fue casi instantáneo» u «ocurrió antes» pueden representarse mediante las distribuciones de posibilidad $\pi_{d\approx 1h}$, $\pi_{d=[20,30]}$, $\pi_{d=mucho}$, $\pi_{d>0}$ o $\pi_{d>0}$, dibujadas en la figura 17. Estas distribuciones representan relaciones temporales entre fechas, tanto cuantitativas como cualitativas. En particular, $\pi_{d=[20,30]}$ define restricciones temporales como las de los mapas temporales de Dean y McDermott, y $\pi_{d>0}$ define la relación cualitativa «antes» del modelo de Allen. Estos términos pueden ser usados para comparar fechas entre sí. Por ejemplo, dadas dos fechas **a** y **b**, podemos preguntarnos en qué medida verifican la relación cualitativa «antes», y obtener los correspondientes grados de posibilidad Π y necesidad N como grados de compatibilidad entre las distribuciones $\pi_{d(a,\,b)}$ y $\pi_{d>0}$:

$$\Pi\,(\mathbf{a} < \mathbf{b}) = \sup_{n\in \mathbf{I}} \text{mín}\,\{\pi_{d(a,\,b)}(n),\, \pi_{d>0}(n)\}$$

$$N\,(\mathbf{a} < \mathbf{b}) = 1 - \sup_{n\in \mathbf{I}} \text{mín}\,\{\pi_{d(a,\,b)}(n),\, 1 - \pi_{d>0}(n)\}$$

De forma análoga, podemos calcular cualquier otra relación temporal cualitativa o cuantitativa, sin más que sustituir la distribución π_d apropiada.

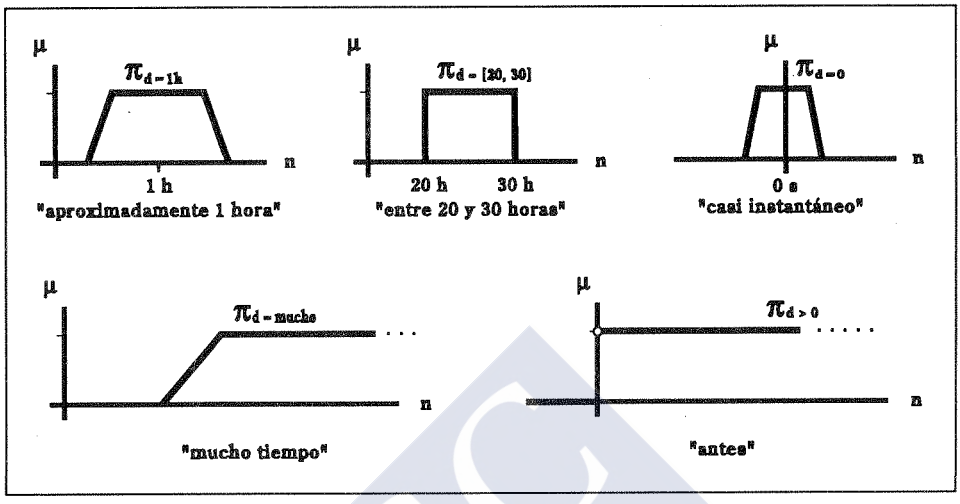


Figura 17. Distribuciones de posibilidad para algunos ejemplos de duraciones.

En un dominio como el que estamos considerando, el usuario puede introducir fechas y relaciones como datos primitivos. Por tanto, la información almacenada en la base de datos deberá representarse como un grafo (*red temporal borrosa*). Como ejemplo podemos considerar el grafo de la figura 18, correspondiente a las siguientes piezas de información:

> «El dolor comenzó hacia la 25ª semana de gestación; unos dos meses después apareció hipertensión arterial, y en la consulta de la 36ª semana se detectó proteinuria»

A partir del grafo, y aplicando expresiones como las anteriores, se pueden realizar inferencias para responder a preguntas como las siguientes:

- ¿Cuando apareció hipertensión? ($¿\pi_B?$)

- ¿Cuanto tiempo transcurrió entre la aparición del dolor y la detección de proteinuria? ($¿\pi_{d(A,\,C)}?$)

- ¿Apareció proteinuria poco después de una hipertensión? ($¿\Pi(\mathbf{B} <_{poco} \mathbf{C})?$ $¿N(\mathbf{B} <_{poco} \mathbf{C})?$)

En particular, preguntas como la última pueden formar parte de las condiciones en las reglas de producción de un sistema experto borroso, y los grados de posibilidad y necesidad devueltos por el especialista temporal se combinarán con los de la regla, de acuerdo con el modelo de la Teoría de la Posibilidad.

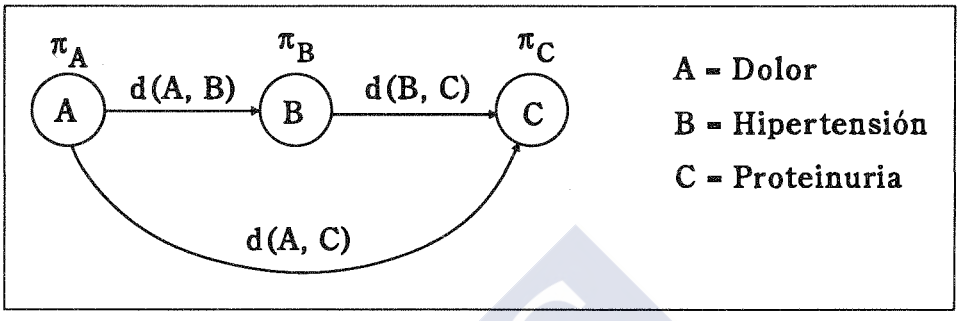


Figura 18. Ejemplo de red temporal borrosa.

5. CONCLUSIONES

Los modelos temporales borrosos que hemos estudiado hasta ahora abordan una pequeña parte del diagrama de bloques inicial presentado en la figura 6. Quedan aún muchas cuestiones por resolver para llegar a la aplicación de modelos borrosos a dominios prácticos. Examinaremos algunas de ellas.

En primer lugar, se puede dotar al especialista temporal de cierto sentido común, mediante la incorporación de conocimiento sobre el tiempo, es decir, de axiomas como los de la Lógica de Allen. Cuestiones como el uso de distintos tipos de persistencias borrosas por defecto o la generalización de las distribuciones de posibilidad a distintos contextos temporales son temas aún pendientes. Es bien conocido el problema de la sensibilidad de las funciones de pertenencia borrosas respecto del contexto. La función que describe a una «mosca grande» no es la misma que la que describe a un «elefante grande» [Lee y col., 1987]. De forma análoga, las funciones que describen tiempos deben depender de la granularidad y de la dinámica intrínseca de cada parámetro.

Por otra parte, hay que tener en cuenta que el especialista temporal trata únicamente con tiempos. Para manejar eventos es necesario combinar hechos con tiempos. El especialista temporal debe ser embebido o conectado a un mecanismo

de inferencias general capaz de manejar conocimiento del dominio. En sistemas expertos borrosos ésto significa que hay que establecer una forma de combinar la imprecisión e incertidumbre en los tiempos con la imprecisión e incertidumbre en los hechos.

Otra cuestión de la máxima importancia en aplicaciones prácticas es la definición de la interfase con el usuario, sacando partido de la capacidad de la Lógica Borrosa para representar términos lingüísticos. Es necesario definir lenguajes para entrada de información y para representación del conocimiento que estén próximos al usuario. En el caso más amplio, estos lenguajes deben tener en cuenta las consideraciones discutidas en la sección 3.2: granularidad, escalas temporales del dominio, tratamiento de tiempos implícitos, etc... Dada la extensa variedad y la riqueza semántica de los términos lingüísticos que usamos para referirnos al tiempo, el tratamiento borroso de estas expresiones es realmente un campo muy amplio y complejo. Aunque existen algunos trabajos en esta línea, han de ser aún integrados en sistemas de razonamiento temporal prácticos.

Como un ejemplo de las posibilidades en esta línea consideremos el caso de los cuantificadores. La Lógica Borrosa introduce cuantificadores borrosos («muchos», «pocos», ...) que rellenan el hueco dejado por los dos únicos cuantificadores de la Lógica de Primer Orden: el cuantificador universal ($\forall$ x) y el cuantificador existencial ($\exists$ x). Cuando el objeto al que hacemos referencia es el tiempo, estos cuantificadores de la lógica convencional corresponden a los adverbios de tiempo «siempre» ($\forall$ t) y «alguna vez» ($\exists$ t). Podemos pensar en muchos otros cuantificadores temporales («a menudo», «a veces», «muy frecuentemente», ...) que no existen en lógica convencional, pero pueden tener un tratamiento adecuado en Lógica Borrosa. De forma similar, términos indicativos de tendencias, repeticiones y patrones temporales complejos admiten representaciones borrosas y, en dominios como la monitorización de señales, pueden permitir el desarrollo de sistemas más cercanos al usuario final.

AGRADECIMIENTOS

Los autores de este trabajo desean agradecer el soporte económico aportado a través de los proyectos TIC-90-0496 y XUGA 20601B92.

REFERENCIAS

Allen, J., «Towards a general theory of action and time», *Artificial Intelligence*, N. 23, (1984), 123-154.

Allen, J., «Time and time again: The many ways to represent time», *International Journal of Intelligent Systems*, N. 6, (1991), 341-355.

Ariav, G., «Design requirements for temporally oriented information systems», en Rolland, C., Bodart, F. y Leonard, M. (Eds.), *Temporal aspects in information systems*, North-Holland, Amsterdam, (1988).

Barro, S., Marín, R., Otero, R.P., Ruiz, R. y Mira, J., «On the handling of time in intelligent monitoring of CCU patients», *14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Paris, Octubre, (1992), 871-873.

Cohn, A., Rosenbaum, S., Factor, M., Sittig, D., Gelernter, D. y Miller, P., «Sequential clinical scenes: A paradigm for computer-based intelligent hemodynamic monitoring», *Proc. of SCAMC 89*, (1989).

Console, L., Rivolin, J. y Torasso, P., «Fuzzy temporal reasoning on causal models», *International Journal of Intelligence Systems*, N. 6, (1991), 107-133.

Cousins, S., Kahn, M. y Frisse, M., «The display and manipulation of temporal information», *Proc. of SCAMC 89*, (1989), 76-80.

Dean, T. y McDermott, A., «Temporal database management», *Artificial Intelligence*, N. 32, (1987), 1-55.

Dechter, R., Meiri, I. y Pearl, J., «Temporal constraint networks», *Artificial Intelligence*, N. 49, (1991), 61-95.

Dubois, D. y Prade, H., «Processing fuzzy temporal knowledge», *IEEE Transactions on Systems, Man and Cybernetics*, 19, 4, (1989), 729-744.

Dutta, S., «Temporal reasoning in medical expert systems», *Proc. of the Symposium on the Engineering of Computer-Based Medical Systems*, Minneapolis, (1988a), 118-122.

Dutta, S., «An event based fuzzy temporal logic», *Proc. 18th IEEE International Symposium on Multiple-Valued Logic*, Palma de Mallorca, España, (1988b), 64-71.

Fagan, L., *VM: Representing time-dependent relations in a medical setting*, Tesis Doctoral, Stanford University, 1980.

Kahn, M. y Gorry, G., «Mechanization of temporal knowledge», *Artificial Intelligence*, N. 9, (1977), 87-108.

Kahn, M., Ferguson, J., Shortliffe, E. y Fagan, L., «Representation and use of temporal information in ONCOCIN», *Proc. IX Annual Symp. in Comp. Appl. in Medical Care*, (1985), 172-176.

Kahn, M., Fagan, L. y Sheiner, L., «Model-based interpretation of time-varying medical data», *Proc. of SCAMC 89*, (1989), 5-10.

Kim, H. y Oh, K., «A new representation model in uncertain temporal knowledge», *Proc. of IFSA'91*, (1991), 113-116.

Lee, N., Grize, Y. y Dehnad, K., «Quantitative models for reasoning under uncertainty in knowledge-based expert systems», *International Journal of Intelligent Systems*, II, (1987), 15-3.

Marín, R., Taboada, M., Mouriño, G., Mira, J., Barreiro, A., Otero, R.P. y Pereira, M., «Representation of temporal knowledge in obstetrics», *Fourth International Symposium on Knowledge Engineering*, Barcelona, España, (1990).

McCarthy, J. y Hayes, P., «Some philosophical problems from the standpoint of artificial intelligence» en Meltzer, B. y Michie, D. (Eds.), *Machine Intelligence*, New York, American Elsevier, (1969).

MacKay, D., *The organization of perception and action*, Springer-Verlag, New York, 1987.

Otero, R., *MEDTOOL: Una herramienta para el desarrollo de sistemas expertos*, Tesis Doctoral, Universidad de Santiago, España, 1991.

Perkins, W. y Austin, A., «Adding temporal reasoning to expert-systems-building environments», *IEEE Expert*, Febrero, 1990.

Russ, T., «Using hindsight in medical decision making», *Proc. of SCAMC 89*, (1989), 5-10.

Sadri, F., «Three recent approaches to temporal reasoning», en Galton, A. (Ed.), *Temporal Logics and their applications*, Academic Press, London, (1987).

Sager, N., Friedman, C. y Lyman, M., *Medical language processing*, Addison-Wesley, Reading, 1987.

Sarda, N., «Extensions to SQL for historical databases», *IEEE Transations on Knowledge and Data Eng.*, Vol. 2, N. 2, (1990), 220-230.

Shoham, Y., *Reasoning about change*, MIT Press, Cambridge, 1988.

Thomson, N., «Human memory: Different stores with different characteristcs», en Monk, A. (Ed.), *Fundamentals of human-computer interaction*, Academic Press, (1985).

Vitek, M., «Fuzzy information and fuzzy time», *Proc. IFAC Symposium on Fuzzy Information, Knowledge Representation and Decision Analysis*, Marseille, Francia, (1983), 159-162.

Zadeh, L., «Coping with the imprecision of the real world: An interview with Lofti A. Zadeh», *Communications of the ACM*, Vol. 27, N. 4, (1984), 304-311.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

COMPUTACIÓN NEURONAL BORROSA

José Mira y Ana E. Delgado
Dpto. Informática y Automática. Facultad de Ciencias. UNED, Madrid.

Roque Marín y Senén Barro
Dpto. Electrónica y Computación. Universidad de Santiago de Compostela

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1. VISIÓN SUBJETIVA DEL CAMPO

La computación neuronal nace ante la evidencia de la calidad del tejido nervioso para procesar información y almacenar conocimiento. La percepción, la memoria, el aprendizaje, la planificación motora y la creatividad son propiedades básicas de los sistemas biológicos.

En contraposición, las máquinas computacionales creadas por el hombre y en particular la programación de calculadoras digitales con arquitectura von Neumann, se han encontrado con enormes dificultades a la hora de duplicar estas funciones básicas del Sistema Nervioso. La Inteligencia Artificial y los Sistemas Basados en Conocimiento, junto con las Arquitecturas Paralelo, son dos fronteras del campo de las ciencias de la computación que tampoco resuelven de forma satisfactoria la síntesis artificial de las funciones más genuinas del cerebro humano.

En un intento de escapar de esta situación límite, la comunidad científica ha vuelto los ojos hacia la anatomía y la fisiología de las redes neuronales y hacia los principios estructurales y organizacionales que están en la base de los sistemas

vivos. Esto ha dado origen a un florecimiento del campo de la computación neuronal como complemento o alternativa al cálculo secuencial convencional. Se ha planteado también la alternativa entre conexionismo (paso de números) y computación simbólica (paso de mensajes), pero esta alternativa no es tal, ya que las computaciones en el dominio propio de los procesadores siempre son subsimbólicas y en el dominio del observador externo siempre son simbólicas. El simbolismo nace en el dominio del observador externo [Mira, 1992].

Se entiende por computación neuronal el cálculo realizado por arquitecturas tridimensionales de un gran número de elementos y con alto grado de conectividad, que realizan localmente una función analógica (suma lineal seguida de comparación no lineal y retardo) y en las que la programación se sustituye por el aprendizaje a través del entrenamiento. Si se acepta esta perspectiva, el problema de la computación neuronal incluye: a) estudio de distintos algoritmos de aprendizaje, junto con procedimientos adicionales para aumentar la velocidad y eficiencia de los ya conocidos, b) desarrollo de arquitecturas y esquemas de realización física, principalmente en microelectrónica analógica, digital o mixta y c) desarrollo de aplicaciones en control, procesamiento de imágenes, reconocimiento de caracteres, etc, donde, de forma más o menos explícita, la red actúa como preprocesador ajustable o co-procesador de otro procesador central, en el que toman cuerpo los algoritmos de aprendizaje para modificar los valores de los pesos y/o del umbral de cada una de las unidades, de acuerdo con ciertos criterios globales, tales como minimizar el error cuadrático medio entre la salida real y la deseada. Con ser importante este campo de investigación, perderíamos los aspectos más genuinos de nuestro sistema nervioso si no volviéramos a considerar las exigencias computacionales de un sistema que pretenda duplicar algunas de las características básicas de las neuronas reales y del tejido nervioso en general, volviendo así a la perspectiva biológica que le dio origen.

En este trabajo pretendemos dar una visión de la computación neuronal, incluyendo la perspectiva biológica del problema, considerada como fuente de inspiración para un nuevo paradigma computacional. La perspectiva biológica es más ambiciosa que la basada en el cálculo analógico no lineal, pues persigue a largo plazo duplicar la capacidad computacional del tejido nervioso, generando un nuevo paradigma basado en los siguientes puntos:

1) Clara distinción entre **procesadores** y **procesos**.

2) Elevado número de procesadores reconfigurables ($\geq 10^{11}$).

3) Gran complejidad en la estructura y función local de estos procesadores, muy lejos siempre del simple cálculo analógico no lineal.

4) Identificación entre estructura y función en el dominio de los procesadores (dominio propio). El software modifica al hardware y viceversa. Al existir identidad computacional de estructura anatómica y macrofunción asociada, no son válidos los conceptos convencionales de hardware y software para explicar y/o duplicar la plasticidad del sistema nervioso.

5) Sustitución del concepto de **programación** por el de **aprendizaje autónomo**, con autoprogramación y autoorganización a nivel local para alcanzar el equilibrio en la interacción con un medio.

6) Redundancia estructural y funcional, resultado de la programación genética que facilita la tolerancia a fallos y la alta seguridad de funcionamiento, incluso frente a la lesión física que elimina un porcentaje importante de la masa de cálculo.

Para abordar el estudio de la computación neuronal de forma que incluya estos puntos de la perspectiva biológica, junto con los otros aspectos más técnicos, se sugiere un esquema que parte del análisis del paradigma computacional clásico y su contrapartida neuronal. A continuación hay que introducir algunos aspectos metodológicos esenciales, tanto en la perspectiva de análisis, como en la de síntesis. En particular, la distinción entre **dominio propio** y **dominio del observador** y el reconocimiento explícito del conocimiento externo inyectado en la red a través de su estructura [Mira y Delgado, 1988; Mira, 1992].

Dentro de este panorama de la perspectiva biológica en computación neuronal hay que volver los ojos hacia el método científico usual en Física y desarrollar un marco teórico-computacional adecuado para describir la fenomenología observada a niveles eléctrico y bioquímico en la función neuronal. No está claro que los potenciales lentos y los *spikes* sean el nivel adecuado. La especificidad bioquímica, los nuevos canales del calcio y la existencia de microradiación en el sistema nervioso, nos obligan a ser cautelosos a este nivel.

Como en toda ciencia, también aquí es esencial la experimentación. Así, tras el establecimiento de la metodología y el marco teórico hay que «bajar» al experimento y estudiar las bases anatomo-fisiológicas y bioquímicas de las redes neuronales. Esto no es sencillo, ni está claro que sea abordable en esta etapa del conocimiento humano. Cuando se está cerca del mundo físico (neuronas sensoriales o motoras) todavía tenemos acceso al experimento. En redes internas han

existido tantas recodificaciones de la información y tanto conocimiento inyectado por la programación evolutiva que es realmente difícil la ruptura del código. Todavía existe la alternativa de estudiar sistemas nerviosos sencillos (caracol, aplixia, límulus, ...).

Sin embargo, no se debe descartar, ni a nivel experimental ni a nivel teórico, la posibilidad de que los aspectos más genuinos de la computación neuronal se encuentren en el dominio subcelular, por lo que es necesario volver los ojos hacia estudios inmunocitoquímicos, de estructuras de proteínas y de otras muchas técnicas que nos permiten enlazar el nivel subcelular con la topología de las redes y sus funcionalidades a nivel eléctrico y de comportamiento global.

Finalmente, como contrapunto al aspecto experimental, hace falta un fuerte desarrollo teórico buscando principios organizacionales y estructurales que, independientemente de lo que «sean» las neuronas, nos permitan reproducir a nivel de síntesis computacional algunas de las propiedades observables en los seres vivos. La descripción a nivel de procesos de estos principios organizacionales ha de ser independiente de las posibles implementaciones físicas a nivel de procesadores. Aprendizaje, memoria asociativa, percepción, autoorganización, autorreparación y tolerancia a fallos, sueño, codificación evolutiva, morfogénesis y autorreproducción, por nombrar unas pocas, son algunas de estas propiedades a investigar.

El panorama de la computación neuronal puede completarse incluyendo un apartado sobre implementaciones y otro sobre aplicaciones. Sin embargo, no puede implementarse nada que antes no haya sido descrito a nivel computacional de forma clara, completa, precisa e inequívoca. Dicho de otra forma, primero es el modelo y después su implementación. Como estamos convencidos de que muchos de los modelos actuales son muy pobres, creemos que es prematuro hablar de su implementación.

Hasta aquí hemos estado mirando esencialmente hacia la biología. Nuestro propósito era comprender la función neuronal. Siguiendo todavía con ese propósito demos ahora un giro y miremos hacia el mundo de la computación y los modelos formales. Pasamos entonces de mirar las neuronas naturales, a buscar modelos matemáticos de esas neuronas. Más exactamente, buscamos descripciones computacionales de las funcionalidades observadas en las redes neuronales. Estamos seguros de que no conocemos todas las funciones reales y todavía así estamos seguros, también, de que no disponemos de la matemática adecuada para su descripción.

Para modelar la función neuronal no podemos usar más que lo que tenemos:

a) Matemática **analógica** lineal o no lineal.

b) Matemática **lógica** (combinacional y secuencial).

c) Matemática **inferencial** (algorítmica o heurística).

d) Versión borrosa de a), b) y c).

e) Aprendizaje en a), b), c) y d).

De aquí nace el campo de las «neuronas» artificiales, al que deberíamos llamar computación analógica, no lineal, distribuida y autoprogramable. Desde este momento, la neurología es sólo una fuente de inspiración.

2. SISTEMAS BORROSOS Y COMPUTACIóN NEURONAL

En los últimos años se ha manifestado un enorme y creciente interés por el estudio de las Redes Neuronales Artificiales (RNA) y por los Sistemas basados en Lógica Borrosa (SLB), a pesar de no ser, sobre todo el primero, ámbitos recientes en el campo de la computación. Es más, existe actualmente un trabajo extenso e intenso enfocado al área, «borrosa» todavía, de posible fusión de los SLB y las RNA.

Pueden plantearse múltiples vías de posible interacción entre RNA y SLB, siendo algunas de ellas las siguientes:

- Aplicación de RNA al diseño y/o refinamiento de reglas condicionales borrosas [Lee, 1991].

- Preestructuración de RNA a partir de su formalización como un SLB, tratando de facilitar un aprendizaje definitivo de la red en la resolución de un problema concreto [Eppler, 1990].

- Representación e implementación de SLB siguiendo esquemas conexionistas, con objeto de aprovechar el paralelismo computacional de éstos últimos [Keller y Tahani, 1992].

- Aplicación de la teoría o formalismo borroso al análisis/síntesis de sistemas naturales/artificiales de computación neuronal. En algunos casos puede demostrarse la equivalencia funcional entre ciertos tipos de RNA y Sistemas de Inferencia

Borrosos [Jang y Sun, 1993; Pedrycz, 1991], lo que permite un trasvase directo de las posibilidades de representación, manipulación y aprendizaje de un modelo al otro.

Enmarcado en este último punto, nuestro trabajo comenta la validez de algunos aspectos de la computación borrosa en el modelado de la función neuronal, junto con la posibilidad de transportar algunos procedimientos de computación analógica distribuida y «autoprogramable», desde el campo de las RNA al campo de los sistemas borrosos, intentando precisar el significado de distintos tipos de neuronas artificiales borrosas (NAB-I, II, III y IV).

El razonamiento de sentido común y, por consiguiente, la computación neuronal que lo soporta, son esencialmente cualitativos y manejan con precisión datos, información y conocimiento imprecisos. Por ello, no parece descabellado explorar la posible utilidad de la matemática borrosa en el modelado de la función neuronal. Junto a los argumentos usados frecuentemente para justificar la idoneidad de esta matemática queremos destacar los siguientes:

1) Capacidad descriptivo-representacional del concepto de función de pertenencia, entendida como curva de compatibilidad de una etiqueta (variable, concepto) con un conjunto de valores (medidas).

2) Idea de razonamiento como combinación y acumulación de evidencia a través de la manipulación de curvas. La posibilidad de «derivar conceptos», por ejemplo, es tremendamente atractiva.

3) Aproximación al lenguaje natural. Intentar hacer computables algunos de los universales del lenguaje, junto con el uso de variables lingüísticas es de nuevo una propuesta atractiva.

No todo son ventajas. Hay inconvenientes entre los que destacamos:

a) Dificultad en la elicitación del conocimiento. Es decir, falta de procedimientos efectivos para obtener las funciones de pertenencia lejos de sensores y efectores.

b) Falta de procedimientos efectivos distintos de los usuales (operadores tipo max, min y complemento) para combinar la evidencia de forma sistemática y constructiva. En particular, falta de procedimiento para cálculo en intenso e intentio, tal como lo hace el lenguaje natural.

c) Falta de procedimientos de autoprogramación (aprendizaje) local, que produzca transformaciones elásticas sobre las funciones de pertenencia, consecuencia del entrenamiento en un medio y la optimización de una función de coste global.

Comentaremos la aplicación de la matemática borrosa en las dos vertientes de la computación neuronal:

A:Vertiente biológica (funciones de sensibilidad modal y espectral).

B:Modelos de computación local.

3. FUNCIONES DE SENSIBILIDAD MODAL Y ESPECTRAL

La matemática borrosa ofrece la mejor capacidad representacional para los datos conocidos sobre la codificación primaria de la información en el sistema nervioso. Esta codificación se realiza en términos de un conjunto de receptores (transductores) para las distintas modalidades sensoriales (visión, audición, tacto, olfato y propiocepción), junto con los termo y nociceptores. Tenemos así una primera codificación de la modalidad mediante «líneas etiquetadas». Sin embargo, estos receptores son multimodales. Es decir, una neurona visual responde principalmente ante estímulos luminosos, pero también, aunque en menor grado, a estimulaciones sonoras o táctiles. A su vez, dentro de una única modalidad sensorial existe sensitividad espectral. Para estímulos luminosos de igual intensidad, hay unidades que responden de forma selectiva a un determinado color. La figura 1.a. (adaptada de Somjen [1975]) muestra los niveles de excitación de tres receptores hipotéticos respondiendo a estímulos de igual energía pero de modalidades sensoriales diferentes. Cuando la estimulación es débil no hay solapamiento en las curvas de respuesta, pero éstas se cruzan al aumentar la intensidad del estímulo.

La figura 1.b. (versión cualitativa de los resultados de Marks y col. [1964]), muestra la absorción relativa de luz en función de su longitud de onda en conos de la retina del mono. Los receptores del azul responden de forma máxima a esta longitud de onda, pero también responden al violeta y al verde.

Se puede obtener un modelo de área primera del córtex sensorial (CS), asumiendo que sobre un substrato anatómico existen neuronas, N(x), en las posiciones x, caracterizadas por su sensibilidad modal. Siguiendo esta representación, tendremos:

$$CS = \{ N(x) ; \Pi^{V}_{N(x)}, \Pi^{A}_{N(x)}, \Pi^{T}_{N(x)}, \Pi^{O}_{N(x)}, \Pi^{S}_{N(x)} \},$$

donde V,A,T,O y S, representan, respectivamente, las modalidades sensoriales Visual, Auditiva, Táctil, Olfativa y Somestética.

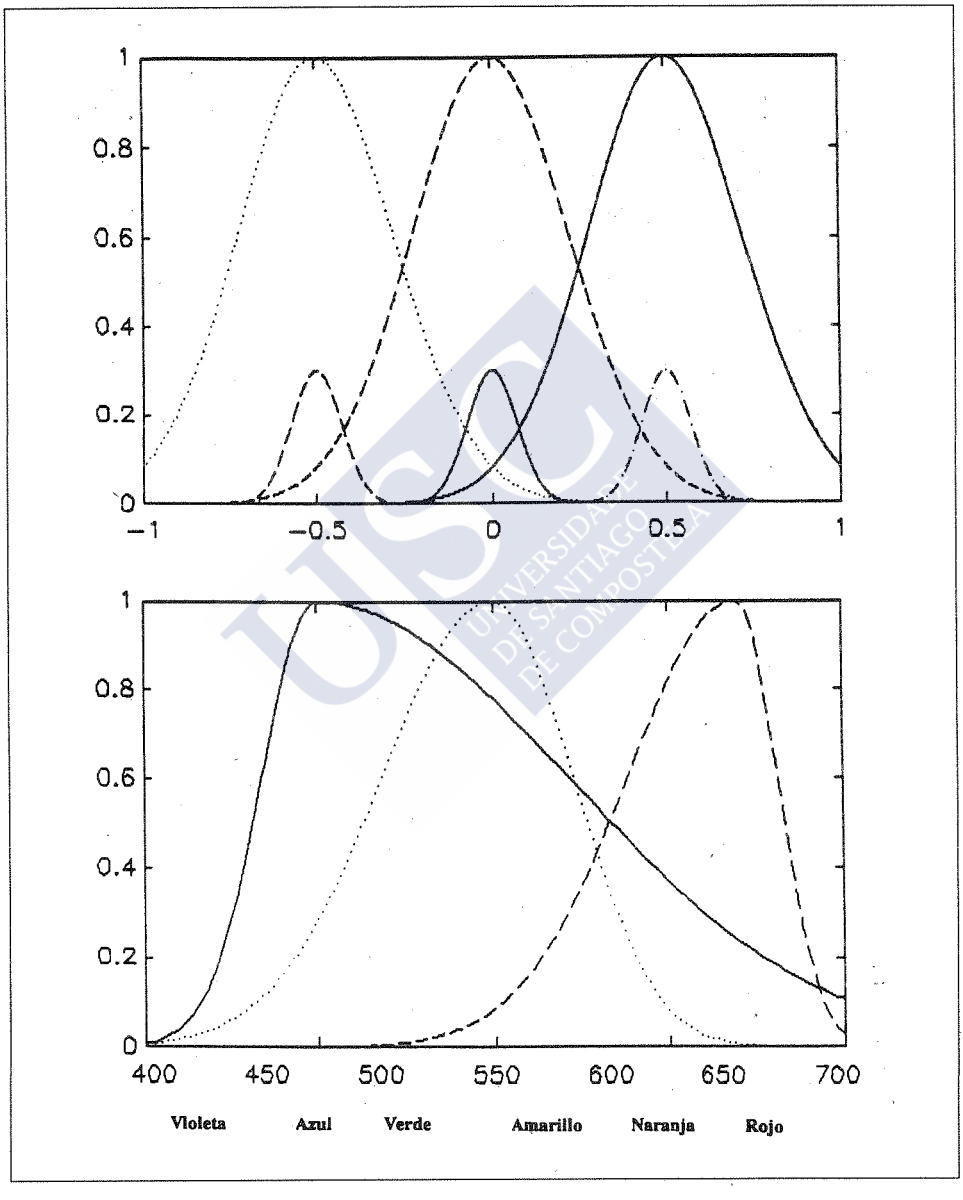


Figura 1. a) Sensibilidad gradual a la excitación de tres tipos de receptores. b) Filtros de color en los conos de la retina del mono.

Para cada modalidad sensorial K $\in$ {V,A,T,O,S}, la sensibilidad modal viene representada por una distribución o función de respuesta modal $\Pi^K_{N(x)}$, definida sobre el universo de discurso formado por el conjunto de posibles estímulos $\{E^i\}$ que participan de las K modalidades sensoriales. De este modo:

$$\Pi^K_{N(x)} = \{\mu_{N(x)}(E^i_K), \forall E^i_K\}$$

donde, $\mu_{N(x)}(E^i_K)$ representa el grado en que la neurona N(x) responde al estímulo E^i de la modalidad K-ésima.

Podemos entender que la «localización» de una modalidad sensorial K, se corresponde con aquellas áreas asociadas a neuronas N(x), en las que $\Pi^K_{N(x)}$ es dominante (especificidad funcional), aunque haya respuesta a excitaciones de otras modalidades sensoriales. No obstante, el problema fundamental de este modelo es la obtención experimental de las $\Pi^K_{N(x)}$. Cuando se supera el nivel de transductores y primera neurona, caso del nervio óptico en el camino visual, casi todo el conocimiento que tenemos de [las $\Pi^K_{N(x)}$] se ha obtenido a partir de experimentos de función residual, medida ésta tras la eliminación de un área de tejido nervioso. Es bien conocido que para procesos no lineales de mediana complejidad, ya es prácticamente imposible obtener densidades de funciones a partir de medidas de función residual [Mira y col., 1978; Delgado y col., 1979].

Sobre esta caracterización de las neuronas sensoriales se puede formular un conjunto de leyes de dinámica sensorial, debidas fundamentalmente a Gonzalo [1945, 1952] y Delgado [1978], que imponen transformaciones en las distribuciones $\Pi^K_{N(x)}$. Algunas de estas leyes son:

1) Existencia de gradientes de integración (G y S).

2) Adaptación.

3) Refuerzo intersensorial.

A partir de las $\Pi^K_{N(x)}$ podemos definir distribuciones $\Pi_K=\{\mu_K(x), \forall x\}$, que caracterizan la disposición anatómica de las distintas modalidades sensoriales, de modo que $\mu_K(x)$ refleja el grado de pertenencia de N(x) a la modalidad sensorial K (un análisis similar nos permitiría avanzar más en la especificación funcional, determinando las $\Pi^i_k=\{\mu_{E^i_k}(x), \forall x\}$). De este modo, la multiplicidad funcional de un área anatómica, G(x), y la especificidad modal correspondiente S(x) = 1-G(x),

pueden ser establecidas a partir de las Π_K. Estos conceptos caracterizan el grado de heterogeneidad de un área de cálculo y generan la etiqueta de la modalidad sensorial dominante en ese área.

La figura 2 muestra la situación anatómica para las modalidades visual (V) y táctil (T), que representamos, respectivamente, por Π_V y Π_T. En puntos como el (b) la especificidad funcional es baja y los módulos participan en las dos modalidades, V y T, por lo que el gradiente de integración es alto. Al pasar de una región a otra se intercambia especificidad por multiplicidad funcional.

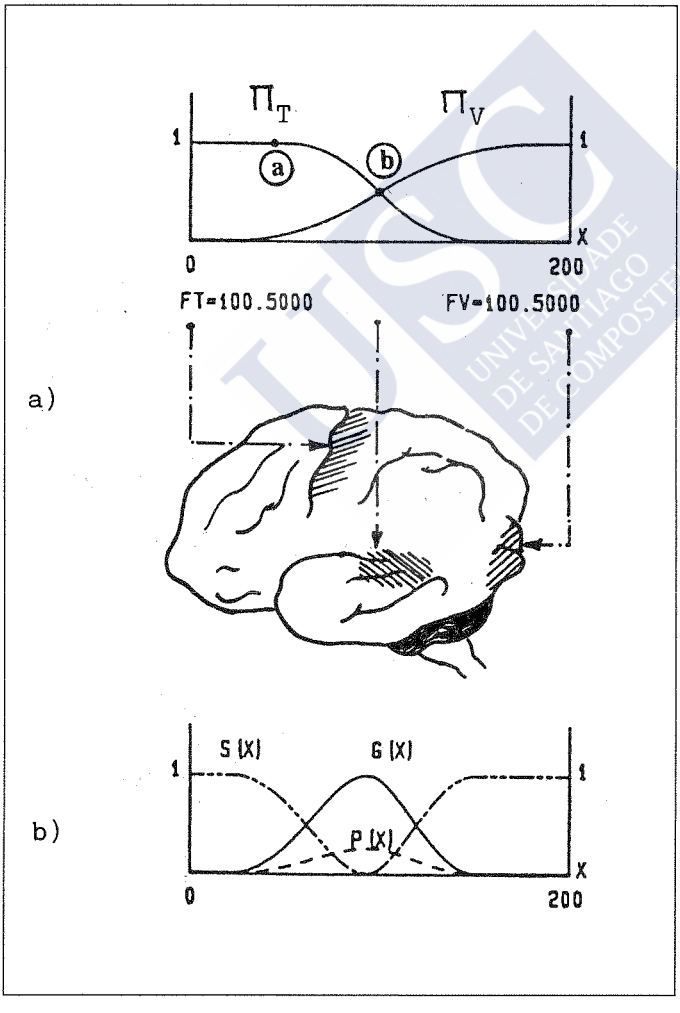


Figura 2. a) Multiplicidad de las áreas visual y táctil. b) Gradiente de integración de J. Gonzalo.

En este nivel, la adaptación es un fenómeno de ajuste de las Π_K (ó Π^i_k), en función de la información aferente, y su derivada (cuando la excitación permanece constante, la respuesta disminuye e, inversamente, una derivada positiva de la excitación aumenta la sensibilidad). De este modo, el grado en el que una neurona N(x) responde a un estímulo E_K en un instante de tiempo t_n, viene dado por la siguiente expresión:

$$\mu_k (x, t_n) = \text{Prog} \left[\mu_k (x, t_{n-1}), E_k, \frac{dE_k}{dt}\right]$$

El fenómeno de refuerzo intersensorial aparece como una facilitación a la «sumación espacio-temporal», que permite suplir la pérdida de excitabilidad elemental del tejido sensorial, debida a una lesión, por ejemplo, mediante el reclutamiento de nueva masa de tejido activo. Para un modelo lineal visuo-táctil, por ejemplo, ello permitiría aumentar $\mu_V (x, t_n)$, como consecuencia de la estimulación concurrente de $\mu_T (x, t_{n-1})$, y viceversa. Concretamente:

$$\mu_V (x, t_n) = \mu_V (x, t_{n-1}) + \alpha_{VT} \cdot \mu_T (x, t_{n-1})$$

y análogamente

$$\mu_T (x, t_n) = \mu_T (x, t_{n-1}) + \alpha_{TV} \cdot \mu_V (x, t_{n-1})$$

4. MODELOS DE COMPUTACIÓN LOCAL: EXTENSIÓN DEL MODELO ANALÓGICO

En el apartado anterior hemos querido ilustrar la capacidad representacional de la matemática borrosa para modelar algunos aspectos del comportamiento de las neuronas biológicas. Queda claro que un tratamiento en extenso de este tema está fuera del propósito y de las posibilidades de este trabajo, que sólo pretende (en su apartado 3) señalar la posibilidad de formulación borrosa de los datos neurofisiológicos.

Veamos ahora alguna de las implicaciones computacionales de la matemática borrosa en el campo de las Redes de Neuronas Artificiales (RNA). Recordemos que ahora deberíamos hablar de modelos de computación local, analógica, lógica

o inferencial y de su contrapartida borrosa. En este trabajo vamos a presentar sólo la extensión del modelo analógico.

4.1. Tipos de Neuronas Borrosas

Una aproximación razonable al problema parte de admitir que hay varios tipos de neuronas artificiales borrosas. Dicho de otra forma, que la formulación borrosa tiene sentido a distintos niveles, que dependen, a su vez, de la tarea y el dominio en el que van a ser usadas las redes y de la arquitectura de las mismas.

Dado que la función más idónea y de uso general es la de clasificación y la arquitectura más adecuada es la de computación por capas, hablaremos de computación borrosa modular por capas en las que, al menos, se pueden distinguir los siguientes tipos de neuronas artificiales borrosas (NAB):

NAB - Tipo-I: Unidades de representación.

NAB - Tipo-II: Unidades de extracción de características.

NAB - Tipo-III: Unidades de decisión.

NAB - Tipo-IV: Unidades efectoras.

4.1.1. NAB-Tipo I

Son NAB de representación, usadas en la primera capa de la red. Cumplen funciones análogas a las neuronas sensoriales que actúan como transductores y codificadores de bajo nivel.

En regiones del córtex visual se ha demostrado la existencia de campos receptivos localmente sintonizados, aunque solapados. Las capas siguientes de la red, en lugar de operar directamente sobre el espacio de estímulos de entrada, lo harán sobre el resultado de una «codificación» realizada por las NAB-I, actuando éstas como unidades de campo receptivo.

La respuesta de estas unidades de primera capa se circunscribe a una región del espacio de estímulos, con centro en un vector de estímulos $\bar{C}$ y de morfología radial (estructuras del tipo de las Funciones de Base Radial [Poggio y Girosi, 1990; Moody y Darken, 1989]). En general, la respuesta de una neurona N_m tendrá

asociada una función sensorial tipo Gaussiana, de modo que su nivel de activación vendrá dado por la siguiente expresión:

$$X_m = A \cdot e^{-\frac{|\bar{E} - \bar{C}_m|^2}{\sigma^2_m}}$$

donde A es una constante, $\bar{E}$ representa el vector de estímulos que alcanza la entrada y $\bar{C}_m$ y σ_m son, respectivamente, el vector centroide y la varianza de la función sensorial asociada a N_m.

Puede demostrarse que este tipo de funciones sensoriales son aproximadores universales, capaces de aproximar con precisión ilimitada cualquier función real continua sobre un conjunto compacto de entradas [Wang, 1992].

En este nivel, la definición de esta primera etapa puede ser de naturaleza empírica y constituir parte del conocimiento que vamos inyectando en la estructura de la red. Posteriormente consideraremos mecanismos de aprendizaje que pueden intervenir en el refinamiento de estas estructuras.

4.1.2. NAB-Tipo II

La figura 3 ilustra el funcionamiento de las NAB-II. Sea $X = (x_1, x_2, \ldots, x_M)$ la ristra de variables de entrada. Cada neurona analógica de esta segunda capa, N_n, realiza una transformación lineal sobre la configuración de entrada, generando la salida siguiente:

$$y_n = \sum_m w_{nm} \cdot X_m \qquad \forall \text{ campo receptivo m asociado a } N_n$$

Los pesos w_{nm} representan los valores relativos asociados por cada neurona N_n a cada variable x_m. Finalmente, obtenido el nivel de excitación y_n asumimos que cada neurona N_n posee una función de activación que podemos identificar con una distribución borrosa que asocia un grado de compatibilidad a cada valor de excitación.

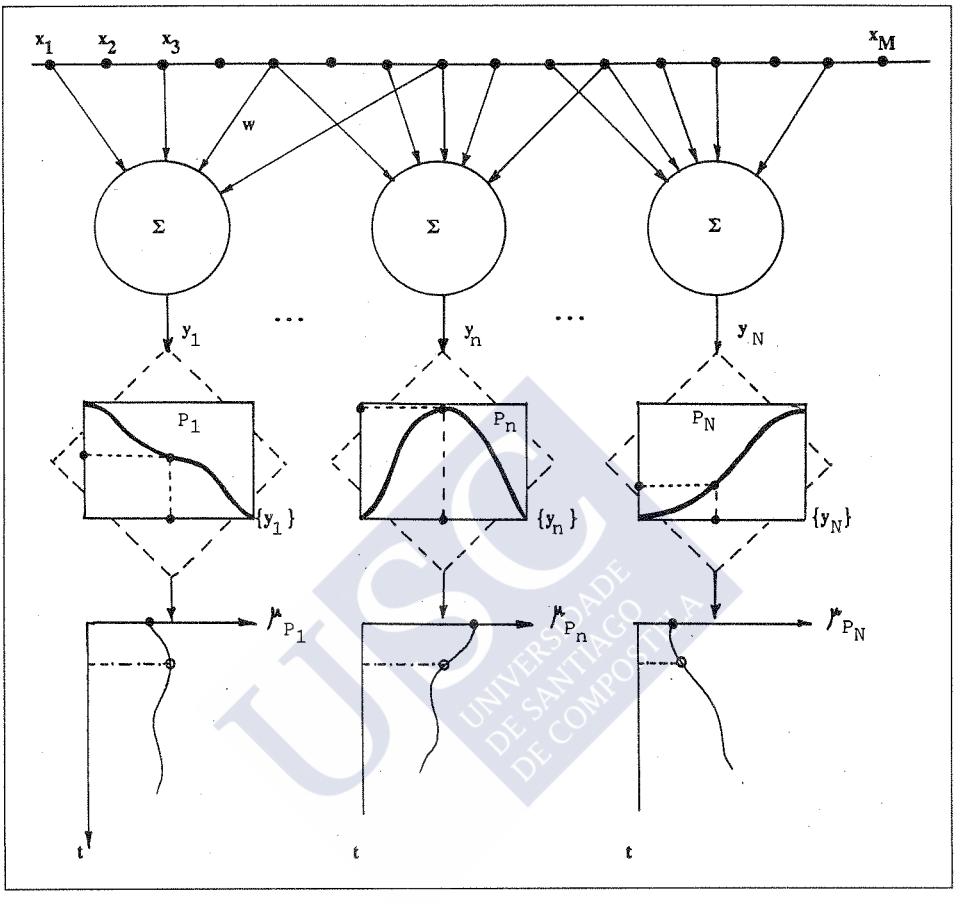


Figura 3. NAB-II. Generan funciones de pertenencia y las correspondientes caracterizaciones del espacio de entradas mediante medidas lineales sobre campos receptivos que se solapan.

La misión de NAB-II es describir el medio en términos de un conjunto de propiedades $\{P_1, P_2,..., P_N\}$, necesitándose tantas NAB-II como propiedades queramos usar para describir el medio [Mira y Moreno-Díaz, 1982] (Figura 4). Posteriormente, se establecerán configuraciones de propiedades, utilizándose como distintivo la multipropiedad. Es decir, una configuración puede poseer en distinto grado todas las propiedades NAB-II.

El aprendizaje en estas dos primeras capas (NAB I y II), implica el especificar los parámetros $\overline{C}_m$, σ_m y w_{nm}. En la literatura se han apuntado soluciones en esta línea [Moody y Darken, 1988; Mel, 1992; Chen y col., 1991]. Por otra parte,

puede verse que, bajo ciertas restricciones, el comportamiento de estas estructuras neuronales es funcionalmente equivalente al de algunos SLB [Jang y Sun, 1993], por lo que podríamos hacer uso de metodologías de identificación estructural y de parámetros propios de este tipo de sistemas. Con ello, podría configurarse la estructura neuronal inicial, asumiendose, en este caso, una cantidad importante de conocimiento a priori.

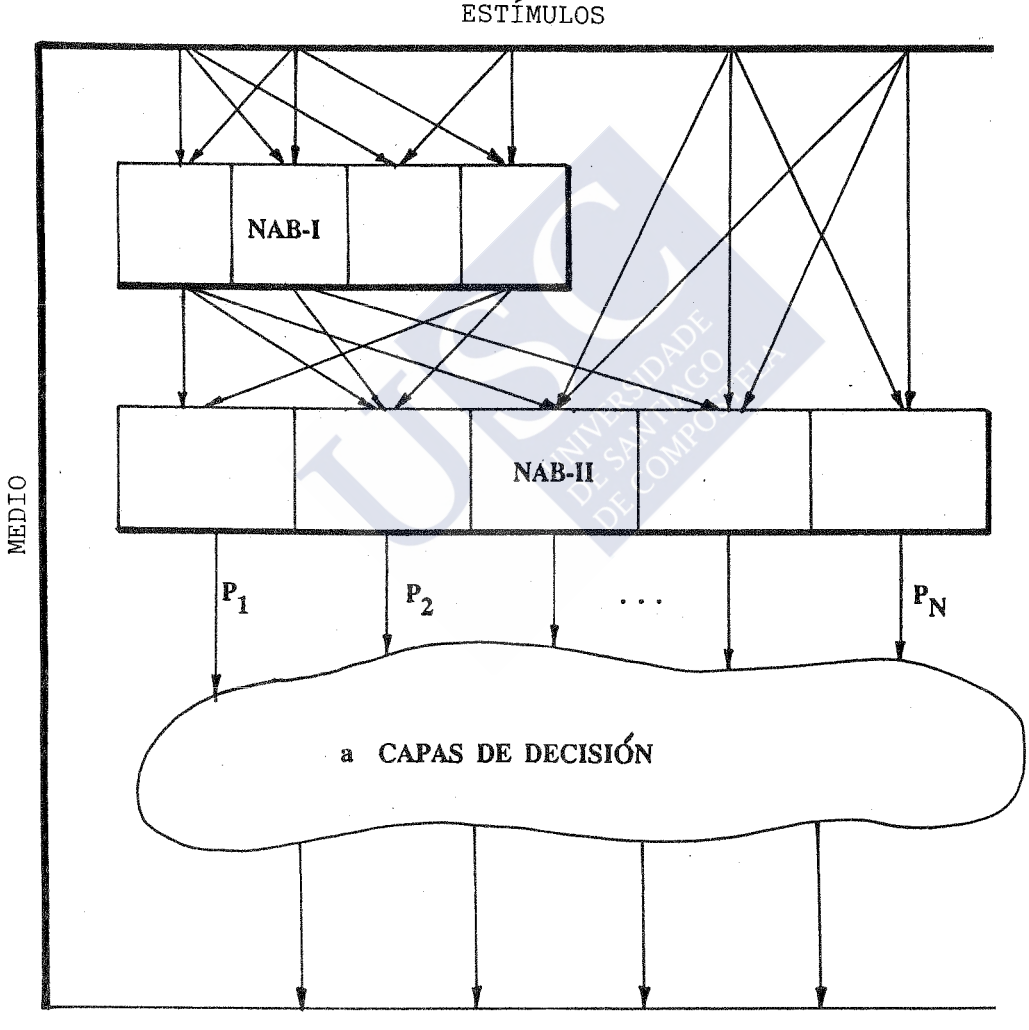


Figura 4. Función de las capas de NAB-I y II que generan una descripción del medio en términos de un conjunto de propiedades (P_1, P_2, ..., P_N).

4.1.3. Neuronas tipo NAB-III

Hasta ahora, las capas NAB-I y NAB-II se han encargado básicamente de producir una representación borrosa del espacio de estímulos, a través de la generación de los valores $\{\mu_{Pn}\}$, n=1,...,N. A partir de aquí el problema que se plantea es el de especificar un proceso de decisión, en general secuencial e iterativo, que produzca un conjunto de inferencias en forma de valores que van llenando el espacio de salidas del nuevo nivel neuronal.

Este proceso de decisión podría realizarse mediante un único procesador que evalúe todas las reglas de decisión y la forma de combinarlas. Estaríamos hablando entonces de un sistema experto con razonamiento aproximado, posiblemente programado por técnicas convencionales en inteligencia artificial aplicada. Sin embargo, sin rechazar esta alternativa, sino más bien complementándola, estamos interesados en la síntesis modular y autoprogramable de este proceso. Es decir, en su implementación mediante una o varias capas de neuronas que admitan cierta autoprogramación, bien directa, a través de las entradas, de acuerdo con las propuestas recientes de Moreno-Díaz (comunicación personal), o bien por los métodos convencionales de asociación, refuerzo o gradiente.

No todas las funciones de decisión son segmentables y, por consiguiente, realizables de forma modular y cooperativa. Admitamos en este contexto que las funciones por las que estamos interesados sí que cumplen esa condición. En tal caso, las capas neuronales siguientes serán capas de condicionales, en las que cada «neurona» soporta un «IF-THEN» en el que se han combinado campos receptivos de condición y acción.

A partir de ahora seguiremos una formulación homogénea para la RNA en términos de un marco de computación por capas. En los modelos analógicos, la función neuronal se interpreta, en el caso más general, en términos de una suma ponderada seguida de umbral y retardo [McCulloch y Pitts, 1943; Blum, 1961; Caianiello, 1966, 1967; Moreno-Díaz y col., 1971; Mira, 1971; Delgado, 1978; Rumelhart, 1986]. En nuestro marco computacional cada neurona artificial de una capa, NA, muestrea en un espacio de entradas R, y en un espacio de salidas R^* (ambos con estructura FIFO) y realiza un cálculo local cuyo resultado descarga en la FIFO de salida de su capa [Mira y Moreno-Díaz, 1982; Mira y col., 1983] (figura 5).

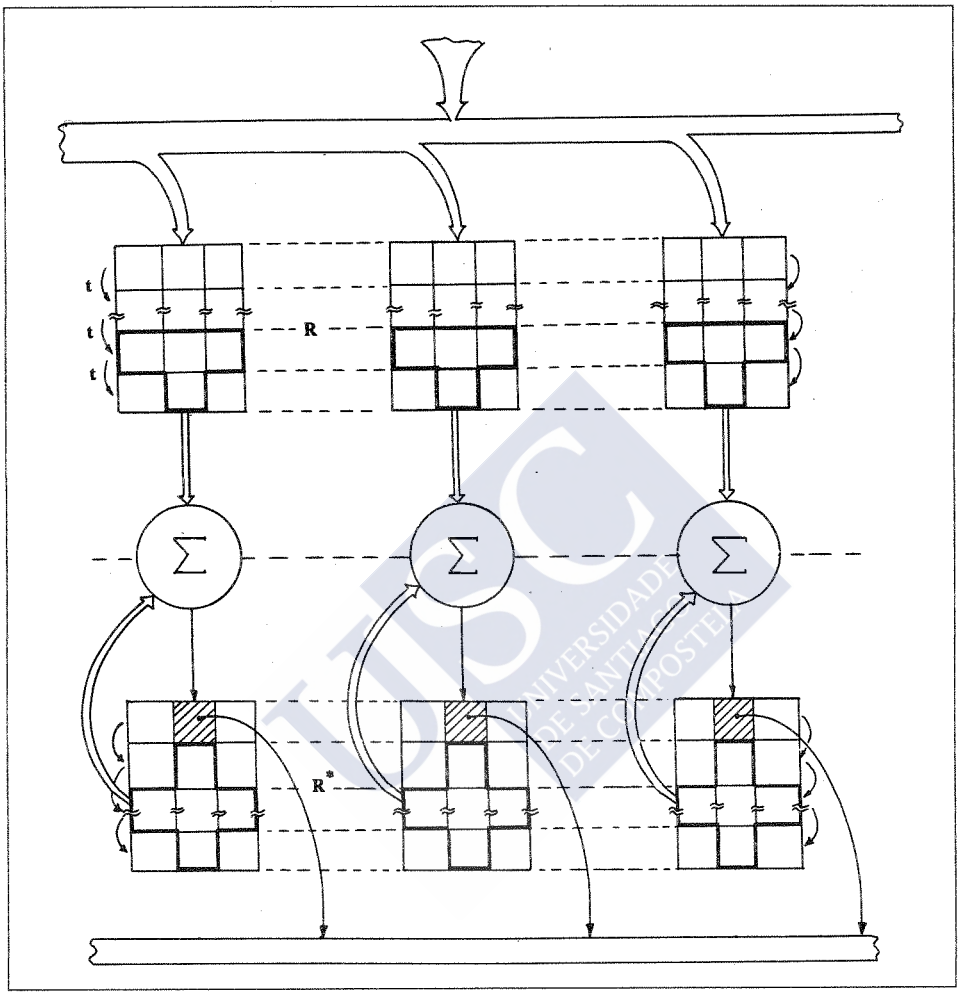


Figura 5. Estructura de cálculo sobre memorias FIFO con realimentación. Ilustración del nivel analógico.

Considerando el caso unidimensional, que en nuestro marco se convierte en bidimensional al considerar al tiempo como variable de cálculo en las FIFO de los espacios de entrada y salida, cada elemento N^i_q de la capa i-ésima toma información de R^i_q en la FIFO de entrada y de R^{*i}_q en la FIFO de salida y calcula la función de excitación antes del umbral. Esta función de excitación puede ser:

a) Lineal o no lineal.

b) Invariante a traslaciones o función de x, siendo x la variable que representa la dimensión considerada.

En el caso lineal y no invariante se trata de sumas ponderadas de la forma:

$$y^i_q(t) = \sum_{l \in R^i_q} w^i_{ql} \cdot x^i_l(t' < t) + \sum_{m \in R^{*i}_q} w^{*i}_{qm} \cdot y^{*i}_m(t' < t)$$

donde las zonas de muestreo R^i_q (asociada a valores de entrada x^i) y R^{*i}_q (asociada a valores de salida y^{*i}) y los valores de los coeficientes, w^i_{ql} y w^{*i}_{qm}, definen la conectividad y los pesos de entrada y realimentación, respectivamente; w^i_{ql} define la contribución (excitadora o inhibidora) de la posición l-ésima del espacio de entradas sobre la unidad N^i_q. Análogamente, w^{*i}_{qm} define la realimentación desde la posición m-ésima del espacio de salidas sobre la unidad N^i_q. En conjunto, las matrices que contienen a los elementos w^i_{ql} y w^{*i}_{qm}, definen la componente lineal de la computación local (figura 6).

La respuesta de cada una de las unidades se obtiene ahora mediante una función de decisión no lineal, $\theta^i_{q,t}$, que en un caso extremo puede ser una función de umbral. Así, el grado de actividad y^{*i}_q, que descarga la unidad N^i_q en su espacio de salidas, es:

$$y^{*i}_q(t+\Delta t) = \theta^i_{q,t}(y^i_q(t))$$

En el caso estático la función de decisión no depende del tiempo ni de la unidad y, al igual que en una NAB-II, puede indentificarse con una distribución borrosa, aunque comúnmente suelen asumirse funciones continuas no decrecientes tipo sigmoide. Es decir:

$$\theta(y) = \frac{1}{C + e^{-y}}$$

La forma específica de la función umbral (su expresión analítica) tiene interés en aquellos casos en los que la ecuación de la neurona va a ser derivada o integrada. En simulación digital este aspecto es menos relevante. Basta con que exista un procedimiento de cálculo de una variación para poder aplicar los algoritmos de aprendizaje que propagan «premios» y «castigos» por la red.

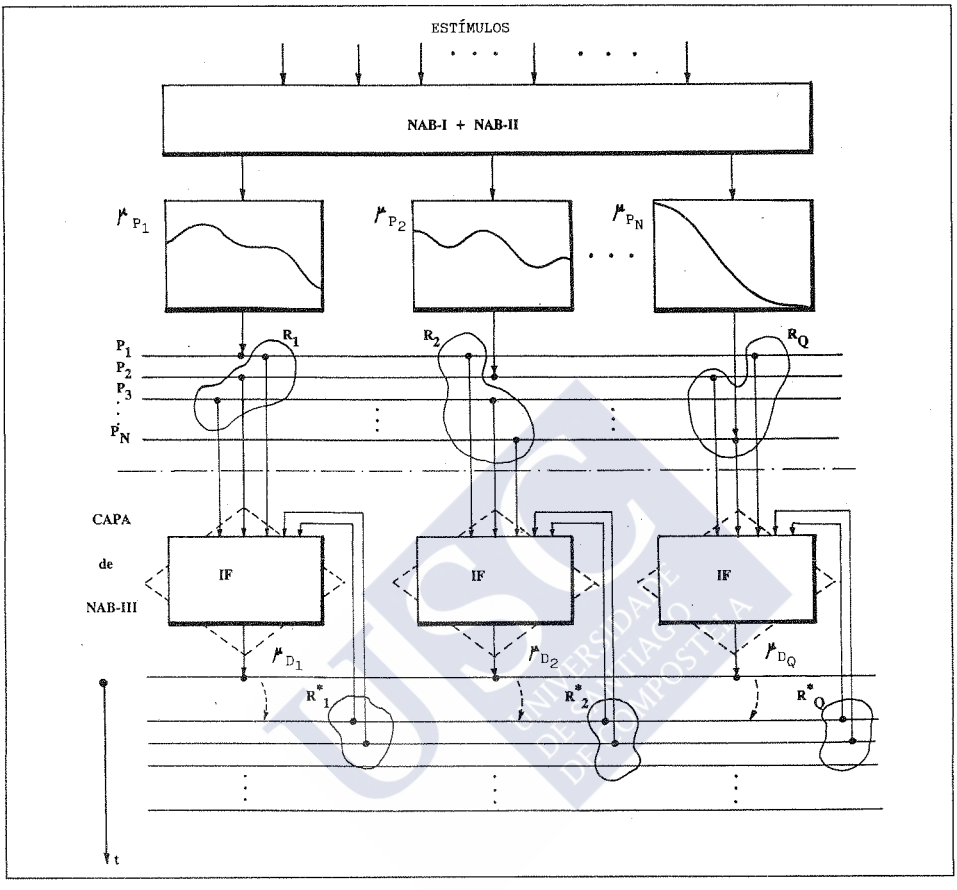


Figura 6. Capa NAB-III.

Finalmente, nos queda el problema del retardo. Por muy elemental que sea un modelo de función neuronal, debe incluir el retardo sináptico como responsable de la «memoria» y de la capacidad de cálculo secuencial. Hay una parte de la memoria de la red que está implícita en el valor de las conexiones y los umbrales. De hecho, lo que aquí se almacena es la función. En la anatomía está el «programa». Sin embargo, en nuestro marco computacional el retardo sináptico está implícito, ya que la forma de R^* define de dónde se toman las respuestas del pasado y dónde se dejan las presentes. El retardo analógico en el campo dendrítico está implícito en R, aunque no suele usarse en los modelos analógicos más extendidos.

El resultado del proceso de decisión de las Q NAB-III rellenará su espacio local de salidas con decisiones (D_1, D_2,...,D_Q), que podemos entender en general como etiquetas lingüísticas de acciones elementales, aunque pueden incluir también respuestas numéricas.

La computación de NAB-III en esta formulación es intrínsecamente cooperativa, ya que en el campo condicional se incluyen no sólo las propiedades del espacio de entradas, (P_1, P_2,..., P_N), sino también las decisiones del espacio de salidas, (D_1, D_2,..., D_Q). Dicho de otra forma, cada N^i_j considera para su decisión en t_n tanto el estado del medio, reflejado en $\{\mu_{Pn} \mid n \in R^i_j\}$ en (t_n, t_{n-1}, t_{n-2},...), como las decisiones individuales de las NAB-III vecinas, $\{\mu_{Dq} \mid q \in R^{*i}_j\}$ en (t_n, t_{n-1}, t_{n-2},...). La forma de la regla será:

Regla j: IF Prog IN $\{\mu_{Pn} \mid n \in R^i_j, \forall t < t_n\}$

AND Prog OUT $\{\mu_{Dq} \mid q \in R^{*i}_j, \forall t < t_n\}$

THEN $\{\mu_{Dj}$ en $t = t_{n+1}\}$

Bajo Prog IN y Prog OUT se incluyen los mecanismos de razonamiento local, que permiten hacer computacional la evolución temporal de las decisiones individuales a partir de la consideración de las funciones de pertenencia del espacio de entradas, en un cierto número de intervalos anteriores, y de las decisiones de los módulos vecinos con los que existe conectividad funcional.

Para completar la descripción de la computación local hay que precisar, al menos, los siguientes puntos:

A. Operadores.

B. Campos receptivos.

C. Capacidad estructural para la autoprogramación.

Prog IN y Prog OUT podrán incluir todos los operadores lógico-relacionales (AND, OR, NOT, MIN, MAX, COMP, $\geq$, $\leq$, . . .), junto con los aritméticos (suma, resta, multiplicación, etc...) y otros especiales, resultado del conocimiento sobre la tarea y el dominio en el que se va a entrenar la red. Por ejemplo, en diagnóstico médico es frecuente la necesidad del operador sinergia, que genera un grado adicional de confirmación de un diagnóstico como consecuencia de la coincidencia espacio-temporal de determinados síntomas.

Lo genuino de la formulación conexionista es la capacidad de aprendizaje, por lo que no es aconsejable excederse en la complejidad de los programas locales. Por eso se usan frecuentemente redes de operador único: redes MAX o redes MIN. Aumentar la complejidad del programa complica el problema del aprendizaje, por lo que se prefiere aumentar el número de capas, haciendo que la propia estructura de conexión transporte parte de la computación.

Otro punto importante en NAB-III es la estructura de los campos receptivos, R y R*. La anatomía de los campos dendríticos y axónicos determina fuertemente la inferencia. Así, en una visión cualitativa, campos dendríticos de un mismo signo (excitador o inhibidor), siempre generan funciones de pertenencia «suaves». Cuanto mayor es la extensión de R y/o R*, más se extiende $\{\mu_{Dj}\}$. Extensión, forma y simetría determinan la computación. La estructura más frecuente en el sistema nervioso es la de centro-periferia, alternando excitación con inhibición («*on-off*»). La inhibición lateral pertenece a este tipo de estructura.

Finalmente, todas las unidades, y en particular NAB-III, necesitan las facilidades estructurales mínimas para admitir autoprogramación local. Cuando se está pensando en simulación de redes borrosas este problema es importante, pero no crucial. En cambio, si aceptamos un conexionismo en términos genuinos, necesitamos pensar en la implementación a nivel de procesadores. Dicho de otro modo, NAB-I, II y III deben pensarse en términos electrónicos porque su destino natural es convertirse en un circuito integrado. En esta perspectiva es evidente la importancia que tiene el rango de operadores, la exigencia de RAM local y la arquitectura interna que debe permitir la autoprogramación. Un aspecto parece evidente: estamos hablando en términos electrónicos de procesadores cuya complejidad es próxima a la de un microprocesador de un bit. Aun así, las redes NAB multicapa deben actuar como coprocesadores de estructuras CPU/RAM convencionales. En otro caso, estaremos obligados a procedimientos muy elementales para combinar la evidencia a nivel local (redes MAX, combinación lineal de $\{\mu_{Pi}\}$, etc...).

La computación local expresada se caracteriza por un esqueleto (suma ponderada seguida de decisión y retardo) y por un conjunto de parámetros (pesos w^{i}_{ql} y w^{*i}_{qm}; funciones de decisión, $\theta(y)$; retardo, Δt), cuyos valores deben ser configurados por un proceso de aprendizaje. En el caso más simple el aprendizaje consistirá únicamente en un ajuste de pesos, siguiendo, por ejemplo, un criterio de minimización de una función de error [Hayashi y col., 1993]. Sin embargo, la complejidad del campo y la influencia de la metodología propuesta, basada en la

necesidad de distinguir tipos de NAB para cada familia de aplicaciones, hace que este apartado sea actualmente objeto de estudio en el grupo, y no lo consideramos todavía con madurez suficiente para introducirlo en un texto con propósito divulgador.

4.1.4. Neuronas tipo NAB-IV

Finalmente, las unidades NAB-IV son, en general, casos particulares de NAB-III ó NAB-I. Su misión es generar un conjunto de acciones elementales concurrentes para cerrar el lazo de la interacción con el medio. En aplicaciones de control las NAB-IV realizan el control de efectores y deben, por consiguiente, presentar la forma de variable numérica o de mensaje específico, resultado de decodificar el proceso borroso de decisión.

5. CONCLUSIONES PRELIMINARES

En este trabajo hemos presentado nuestra visión del campo de la Computación Neuronal Borrosa (CNB), que pretende integrar los esfuerzos en el área de las Redes de Neuronas Artificiales con algunos aspectos de los sistemas basados en Lógica Borrosa. Sobre ambos campos existe en los últimos años un interés casi «excesivo», así como una falta notable de metodología. Por eso, nuestra primera conclusión es la llamada a la reflexión y a los aspectos metodológicos básicos.

Bajo el nombre de Computación Neuronal Borrosa parece legítimo estar interesado al menos en dos tipos de cuestiones: A) Aquellas que son propias de las Ciencias de lo Natural, relacionadas con el análisis de los datos neurofisiológicos y, en definitiva, con la neurología y B) aquellas otras cuestiones propias de las Ciencias de lo Artificial, relacionadas con la síntesis de sistemas computacionales distribuidos para la solución de problemas específicos.

En la primera perspectiva, la matemática borrosa ofrece en nuestra opinión una capacidad representacional excelente para la información disponible sobre las distintas modalidades sensoriales. Las funciones de sensibilidad espectral y modal, la multiplicidad funcional de las neuronas corticales en áreas primarias, secundarias y de asociación, la tolerancia a fallos y la función residual trás lesiones traumáticas de tipo local, son procesos neurofisiológicos que admiten una representación adecuada en términos de Lógica Borrosa. Existe, sin embargo, un

inconveniente difícil de superar. La obtención experimental de las funciones de pertenencia sólo se puede realizar de forma directa en primera neurona o última neurona, es decir, en las fronteras del mundo físico. Para neuronas interiores, trás la segunda y tercera sinapsis, tenemos que basarnos en experimentos de función residual que, como es bien sabido, sólo garantizan la conservación de la información para transformaciones lineales.

En la segunda perspectiva el problema de la CNB nace de la falta de acuerdo acerca de aquello de lo que estamos hablando. Es decir, en oposición a lo que ocurre con campos tales como la electrónica, en el que al hablar de amplificadores operacionales, biestables o contadores binarios reversibles, toda la comunidad científica sabe con precisión de lo que se está hablando, en computación neuronal en general y en CNB en particular, no existe un acuerdo mínimo. Por consiguiente, nuestra propuesta en esta línea intenta aclarar el campo, introduciendo algunos aspectos metodológicos y una cierta clasificación. Es decir, no parece evidente que se pueda hablar de una única neuronal artificial borrosa.

Podemos admitir inicialmente que la CNB tiene que ver con la computación granular, distribuida y autoprogramable, en la que, para la representación y preproceso de la información y/o para la inferencia local, se usan procedimientos propios de la lógica borrosa.

Las exigencias metodológicas mínimas deberían ser del tipo de las siguientes: dado que la CNB nace con el propósito de resolver problemas computacionales concretos, deberíamos contestar al menos a las cuestiones siguientes:

C.1. ¿Cúal es el problema que se pretende resolver?

C.2 ¿Es posible encontrar un procedimiento distribuido para su solución?. Es decir, ¿es segmentable el problema?.

C.3 ¿Es aconsejable usar redes de neuronas artificiales borrosas para su solución?. Es decir, ¿hay exigencias de tiempo real, imprecisión en los datos, ignorancia parcial sobre el medio, necesidad de autoprogramación, ..., que aconsejen una solución distribuida y autoprogramable?.

C.4 ¿Cúal es la arquitectura adecuada? (computación por capas, columnas, barriles, etc.).

C.5 ¿Cúantos tipos de NAB son necesarios para las distintas capas?.

C.6 ¿Qué mecanismos de aprendizaje local son adecuados para cada una de las NAB?.

C.7 Finalmente, si el propósito de la aplicación es conexionista en sentido estricto, ¿cúal debe ser la estructura interna de un procesador borroso que le permita configurarse en cada uno de los tipos de NAB que aparecen en C.5., de acuerdo con los mecanismos de aprendizaje de C.6. y con la arquitectura de C.4.?.

AGRADECIMIENTOS

Los autores agradecen el soporte económico de los proyectos CICYT: TIC 88-0315, TIC 90-0496, TIC 91-0997 y TIC 92-0136 , en cuyo contexto se han realizado parte de los desarrollos presentados, en su aplicación al diagnóstico y a la monitorización en medicina.

REFERENCIAS

Blum, M., «Properties of a neuron with many inputs», en von Foester, H. y Gopf, G. (Eds.), *Principles of self organization*, McMillan, Pergamon, New York, (1961).

Caianiello, E., «Decision equations and reverberations», *Kybernetik*, N. 3, (1966), 98-100.

Caianiello, E., *Reverberations and control of neural networks*, Universita di Napoli, Instituto de Física Teórica, Napoli, 1967.

Chen, S., Cowan, C.F.N. y Grant, P.M., «Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks», *IEEE Transactions on Neural Networks*, Vol. 2, (1991), 302-309.

Delgado, A.E., *Modelos neurocibernéticos de dinámica cerebral*, Tesis Doctoral, ETSIT, Madrid, 1978.

Delgado, A.E., Mira, J. y Moreno-Díaz, R., «Un modelo conceptual de dinámica cerebral», *Revista de Informática y Automática*, N. 39, Enero-Marzo, Madrid, (1979), 1-15.

Eppler, W., «Implementation of Fuzzy Production Systems with Neural Networks», en *Parallel Processing in Neural Systems and Computers*, Eckmiller, R. y col. (Ed.), North Holland, (1990), 249-252.

Gonzalo, J., *Investigaciones sobre la nueva dinámica cerebral*, Tomo I (XVI), C.S.I.C. Madrid, 1945.

Gonzalo, J., «Las funciones cerebrales humanas según nuevos datos y bases fisiológicas», *Trabajos del Instituto Cajal de Investigaciones Biológicas*, Vol. XLIV, Madrid, 1952.

Hayashi, Y., Buckley, J. and Czogala, E., «Fuzzy Neural Network with Fuzzy Signals and Weights», *International Journal of Intelligent Systems*, Vol. 8, (1993), 527-537.

Jang, J.S.R. y Sun, C.T., «Functional Equivalence Between Radial Basis Function Networks and Fuzzy Inference Systems», *IEEE Transactions on Neural Networks*, Vol. 4, N. 1, (1993), 156-159.

Keller, J.M. y Tahani, H., «Implementation of Conjunctive and Disjunctive Fuzzy Logic Rules with Neural Networks», *International Journal of Approximate Reasoning*, N. 6, (1992), 221-240.

Lee, C.C., «A Self-Learning Rule-Based Controller Employing Approximate Reasoning and Neural Net Concepts», *International Journal of Intelligent Systems*, Vol. 6, (1991), 71-93.

Marks, W.B., Dobelle, W.H. y MacNichol, E.F., «Visual Pigments of single primate cones», *Science*, N. 143, (1964), 1181-1183.

McCulloch, W.S. y Pitts, W., «A logical calculus of the ideas immanent in nervous activity», *Bulletin of Mathematical Biophysics*, Vol. 5, Chicago Univ. Press., (1943), 115-133.

Mel, B.W., «Connectionist Robot Motion Planning. A Neurally-Inspired Approach to Visually-Guided Reaching», Academic Press, Inc., 1990.

Mira, J., *Modelos Cibernéticos de aprendizaje*, Tesis Doctoral, Fac. de Ciencias, Universidad de Madrid, 1971.

Mira, J., *Computación Neuronal*, Report interno del Curso de Tercer Ciclo en la UNED, Madrid, 1992.

Mira, J. y Delgado, A.E., «On Processors and Processes in Neural Computing machines», *Proc. International Congress on Knowledge Engineering*, Madrid, (1988), 201-214.

Mira, J. y Delgado, A.E., «Always Trying to Write an Equation for the Brain», en Prieto, A. (Ed.), *Lecture Notes in Computer Science, 540, Artificial Neural Networks*, Springer-Verlag, Berlín, (1991), 93-100.

Mira, J. y Moreno-Diaz, R., «Un marco teórico para interpretar la función neuronal a altos niveles», *Actas de la I Reunion Nacional de Biocibernética*, Real Academia de Ciencias, Madrid, (1982), 151-157.

Mira, J., Moreno-Díaz, R. y Delgado, A.E., «Modelos implícitos en neuropsicología», *Revista de Informática y Automática*, N. 38, Madrid, Octubre-Diciembre, (1978), 30-37.

Mira, J., Delgado, A.E. y Moreno-Díaz, R., «Cooperative processes in cerebral dynamic», en Lainiotis, D.G. y Tzannes, N.S. (Eds.), *Applications of Information and Control Systems*, D. Reidel Pub. Co. Dordrecht, (1980), 273-280.

Mira, J., Moreno, R. y Delgado, A.E., «A theoretical proposal to embody cooperative decision in the nervous system», en Lasker, G. (Ed.), *Systems Learning*, Vol. II, Chapter XXI, Intersystems Pub. Seaside, California, (1983), 687-690.

Mira, J., Delgado, A.E. y Moreno, R., «The fuzzy paradigm for knowledge representation in cerebral dynalics», *Fuzzy Sets and Systems*, N. 23, (1987), 315-330.

Moody, J. y Darken, C., «Learning with Localized Receptive Fields», en Touretzky, D.,

Hinton, G. y Sejnowski, T. (Eds.), *Proc. 1988 Connectionist Models Summer School*, Carnegie Mellon University, Morgan Kaufmann Publishers, (1988).

Moody, J. y Darken, C., «Fast learning in networks of locally-tuned processing units», *Neural Computation*, Vol. 1, (1989), 281-294.

Moreno-Díaz, R., «Deterministic and probabilistic neural nets with loops», *Mathematical Biosciences*, N. 11, (1971), 129-136.

Moreno-Díaz, R., Mira, J. y Roy, A., «Realización de redes no determinísticas», *Automatica*, N. 12, (1972), 5-14.

Moreno-Díaz, R., Mira, J. y Delgado, A.E., «Hacia una teoría social de redes neuronales», *Actas del II Simposium de Ingeniería Biomédica*, Madrid, (1983), 215-220.

Pedrycz, W., «Neurocomputations in Relational Systems», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, N. 3, (1991), 289-297.

Poggio, T. y Girosi, F., «Regularization algorithms for learning that are equivalent to multilayer networks», *Science*, N. 247, (1990), 978-982.

Rumelhart, D.E., Hinton, G.E. y Williams, R.J., «Learning internal representations by error propagation», en Rumelhart, D.E., McClelland, J.L. y the PDP Research Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstucture of Cognition*, Vol. 1: Foundations, MIT Press, Mass., (1986).

Rumelhart, D.E., McClelland y the PDP Research Group, eds., *Parallel distributed processing*, Vol. 1: Foundations, The MIT Press, Mass., 1986.

Santesmases, J.G., Rubio, F. y Mira, J., «Modelo Electrónico de Inhibición Lateral en el Ojo del Límulus», *Real Sociedad Española de Física y Química, Anales de Física*, Tomo LXV, (1969), 219-134.

Santesmases, J.G., Mira, J., Rubio, F. y González, T., «Hacia una Teoría del Aprendizaje en Redes Neuronales», *Revista de Automática*, Año III, N. 10, Madrid, (1970), 11-27.

Somjen, G., *Sensory Coding in the Mammalian Nervous System*, Plenum Pub. Corp., New York, 1975.

Wang, L.X., «Fuzzy Systems are Universal Approximators», *Proc. IEEE International Conference on Fuzzy Systems*, San Diego, EEUU, (1992), 1163-1169.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

UNCERTAINTY IN C & L MACHINES

Panos A. Ligomenides
Cybernetics Research Laboratory, Electrical Engineering Department
University of Maryland, College Park, MD 20742, USA

1. INTRODUCTION: DEALING WITH UNCERTAINTY

Almost all of «real world» human communication, cognitive and reflective activity is characterized by an elusive inexactness which defies efforts for precise mathematical modeling. *Uncertainty* is a fact of life, as it is inevitably involved in our *knowing and dealing* with our sensory world. Common to our everyday experiences of observing, measuring and «computing» the world around us, uncertainty also allows for variability of cause-effect relations and creates the necessary conditions for *adaptation* and *evolution*. We may note that the idea of imperfection in the shapes of the objects of the «real» world of our senses was also central to the Platonic philosophy, which considers «perfect notions» and «exact concepts» to belong to the world of ideas, while «imperfect» or «inexact» structures are everywhere in our sensory world.

We deal with uncertainty by formulating models for its representation, for the assessment of its extend, for its processing and for its propagation. Our commonly used models for representation and processing of uncertainty are expressed in symbolic string-languages of mathematics and logic, and therefore they involve algorithms which are readily executable by digital computer simulation. The most prominent frameworks for uncertainty modeling, quantification, processing

and propagation, are derived from the conceptual premises of Stochasticity Theory, expressed through various types of probability, statistical and entropy concepts and functions, of Fuzzy-set Theory, of Belief Theory, and, recently, of Grey System Theory [Deng, 1989]. In addition, several ad hoc techniques have been devised, notably for various kinds of expert systems, as for example MYCIN's certainty factors, Prospector's subjective Bayesian-based method, and Cohen's theory of endorsements.

Investigations in recent years have established that there is a profound unity between the concepts underlying the subjects of irreversible thermodynamics, bifurcation analysis of nonlinearly coupled rate equations, fluctuation theory, quantum optics, oscillating reactions and fluid mechanics, and the fundamental principles of «Computation and Learning» (C&L). Investigations on the occurrence of nonequilibrium phase transitions and the concomitant *emergence of cooperativity and order* through the creation of self-organizing dissipative structures, have opened new landscapes for big game hunting in our search for more powerful C&L machines. As we continue building computing machines involving larger numbers of locally interacting components of continually decreasing sizes, requiring lesser waste between functionality and hardware complexity, being constrained by three-dimensional connectivity and by finite propagation delays, we see an accelerating convergence between the physical phenomena underlined by energy exchanges and their symbolic perceptions and interpretations underlined by information exchanges. We see the emergence of a new discipline, the **Physics of C&L**, which will deal with the essential uncertainty in a natural manner, and will ask new questions about C&L, and even extend Physics and Biology to new directions.

2. COMPUTATION AND LEARNING

Whether by animals or by machines, «computation» is a well regulated physical activity, whose observed symbolic manifestations are contextually interpreted as a metaphorical and denotative symbolic activity [Ligomenides, 1991b]. In this sense, we say that we have a «computer» which «computes its next state» if we have formulated a *symbolic predictive model* for the physical system, one which will «compute» the physical system's final state when given the initial state. We may note here that models of computation may be formulated at various levels of metaphorical and denotative *symbolic abstraction* of the physical system's behavior, within the contextually chosen reference of «computation» each time.

Based on this generalized concept of computation, we have developed two different computing paradigms:

2.1. Logical Computation Machines

The early computing paradigm of which the conventional digital computer is a product, following the concepts of Turing automata and the von Neumann computing architecture has led to the design of computing physical systems, which perform physical operations that have been metaphorically and denotatively interpreted beforehand as desired symbolic operations in accordance with some pre-defined logical string-grammar. Such computing machines implement the basic functions of *memorization* and *logical operations*. «Well-formed» information-encoded strings, are stored in bistable physical devices, whose measurable «states» –values of some selected physical property – are identified by purely abstract binary symbols which are devoid of any innate meaning. The machine is «programmed» to perform desired computational tasks, i.e. to «process» strings of abstract symbols (externally interpreted) in accordance with the rules of a logical grammar.

In practice, the syntactic logical grammar of a digital computer is conceived in advance of any physical realization. The conceived symbolic representation and grammar of «digital computation» is *subsequently* realized through various physical media, selected appropriately to satisfy speed, power consumption, size, and other requirements. In commercial digital computers, the switching properties and measurable attributes of bistable physical devices are used to represent the *states* and to process the symbolic binary strings of the corresponding logical grammar. Along this line of approach, Turing machines were invented as conceptual tools to make precise the notion of **computability**. Within the conceptual framework of the Turing machine, mathematical propositions are recast as Turing machine algorithms; then, a proposition's decidability can be shown to be equivalent to its algorithm's computability. Thus, an equivalence is established between the notions of algorithmic computability and propositional decidability.

Because of the «logical» basis of its grammar, the «digital» computer operates along state trajectories that are unambiguously and uniquely determined, when given their initial state (program and data). External inputs or interference which redefine the initial state, also redefine the future state trajectory of this logical machine. We may also point out here that, at the level of logical micro-operations, only a notably small subset of selected physical components of the computer are

activated and sequenced in time at any instant. In the symbolic level, *grammatically well-formed strings of abstract symbols* are sequentially transferred and transformed during «programmed» micro-operations in ways that are *contextually meaningful* to the human programmer of the machine.

2.2. Deduction and Induction with Machines

The digital computer, essentially a logical operator, has been designed to replace the human mind in all problems that are solvable by the rules of deductive logic. However, we all know that there are classes of ostensibly deductive problems that the digital computer cannot carry out a solution without man-made extralogical inductive reasoning. The case of theorem proving is one of the most typical deductive operations, where the complexity of solution increases roughly exponentially with the number of processing steps required. Given a set of axioms, a set of logical rules of deduction, and a theorem to be proved, we are supposed to produce a sequence of theorems to lead to the one to be proved (assuming it is provable), such that each theorem is logically derivable from the axioms and/or the preceding theorems. The machine can be programmed to produce all possible theorems that can come about from the given axioms, but in practice will do so only in simple cases, where it does not take prohibitively long time before the machine happens to produce the desired theorem. Any successful theorem-proving program for a digital computer will use some kind of **man-made** heuristic methods to confine the fanning process of the number of produced theorems within tolerable computing times. The principle of minimum entropy, or the introduction of scales of distance or similarity (a seemingly innocuous set of introduced auxiliary data telling the machine *our* value judgment), are such heuristic principles that can be programmed into the logical machine. The heuristic tricks used will produce *inductive generalization* with its concomitant *ambiguity*.

Arguably, the most elementary function of human intelligence is *inductive* in nature. Induction is a process of producing a general proposition on the grounds of a limited number of particular propositions, such that these become special instances of the former. Going from particular to universal, from concrete to abstract, a process that also includes the advancement of argument by analogy, depends not on confirmation but rather on extra-evidential and extra-logical considerations affecting credibility. This places induction beyond the capabilities of purely deductive machines. We can only produce a semblance of inductive inference with a digital computer by formulating our esthetic, judgmental or

conjectural criteria in machine-legible sentences and smuggle them in the programming of the machine under some disguise.

Since the quantitative strength of the extra-logical and extra-evidential considerations used in inductive inference cannot be uniquely determined, induction always is associated with «inductive ambiguity». Uncertainty problems in both categorical and probabilistic cases are usually handled using the tools of Probability, Possibility or Belief theories. We are finding now more and more that induction and generalization are natural capabilities of certain dynamical physical systems with properly coded and interpreted phenomenological behavior. New classes of powerful C&L machines are possible, with the class of Artificial Neural Networks being just an early, rather rudimentary, example.

2.3. Cooperative Computing

As an alternative to the traditional paradigm of finite state machines, an important innovation in «computing» has been developed in recent years, based on the exploitation of the phenomenology of certain kinds of self-organizing nonlinear dynamic systems. Certain dynamical physical systems will indigenously «process» their boundary and initial conditions embedded in their «collective states» into predictably stable dynamic terminal states. Contextually interpreted, such transformations of symbolic «patterns» representing collective states, whose meaning is ingrained in their spatio-temporal forms, constitute a fundamentally different paradigm of «computation», one which is based on the laws of dynamical predictable behavior of certain types of physical phenomena and assemblies.

Diverse investigations in bifurcation analysis of nonlinearly coupled rate equations and fluctuation theory have revealed the occurrence of nonequilibrium phase transitions, the formation of self-organizing dissipative structures and the concomitant *emergence of cooperativity and order* [Nicolis and Prigogine, 1977]. Such self-organizing dissipative structures, capable of *experiencing* their sensory environment, can be used as «Cooperative C&L Machines» [Haken, 1988; Ligomenides, 1992a, 1991a,b, 1987]. While finite state digital machines operate on «well-formed» strings of abstract symbols in accordance with some logical grammar, certain types of nonlinear, self-organizing dynamical systems can offer a different paradigm of computation, one which is associated with emerging collective properties of the physical system. This computation paradigm, which operates on pattern-oriented knowledge representation and processing, and allows

inductive inference and generalization in a natural manner, can deal effectively with such functions as pattern recognition, abstraction and generalization. These basic computational capabilities underscore the brain's most fundamental cognitive activity, that which accounts for the ability to «capture experiential knowledge» [Ligomenides, 1992b] by observing, discerning and processing sensory patterns.

2.4. Learning: Internal Modeling

«Internal modeling» is the act of *compressing and memorizing* sensory information impinging from the environment. Every physical object is a receiver (and a transmitter) of sensory signals from the environment, which, along with energy, also *carry information* about the properties of their sources (other transmitting objects, such as color, shape, intensity variations, etc. Extraction and compression of the information carried by the electromagnetic or sound waves, the information carriers, are achieved by energetic coupling and by altering the internal organization of the receiver, in response to *selective structural and temporal correlations* between the receiver's internal organization and correlated «forms» (called «triggers») in the impinging signals. Internal modeling, the process of abstracting, compressing and memorizing sensory information in response to correlated triggers in the sensory signals, prompts the emergence of *new form* (learning), of diversification and of increased complexity in the receiver.

2.5. C&L in Two «Communicating» Hierarchical Physical Systems

As illustrated in Figure 1, we consider a self-organizing structure, possessing two **hierarchical** levels: An «experiential» lower level (the hardware level) of microscopic physical activity, in which *energy* is the basic commodity of exchange between the component «processors» of the physical assembly, and a «behavioral» upper level (the software level) of metaphorical symbolic manifestation, in which *information* is the basic commodity of exchange. The situation is illustrated in Figure 1(a).

Efferent Computation: Collective properties of the lower level e are manifested, abstracted, observed and interpreted in the higher hierarchical level s, *i.e. they are being* ***computed*** according to our definition of «computation», as illustrated in Figure 1(b). This is the act of «efferent computation», which is associated with the contextual interpretation of a physical system's spatio-temporal symbolic manifestations.

Afferent (Supervised) Learning: The observed and interpreted collective properties at the higher symbolic level s may be forced to change, to suit our computation objectives, **by adjusting the embedded order** in e, i.e. by modifying the internal model, i.e. the structural and functional properties, of the physical system. A way to accomplish this is by *afferent «back propagation of error data»*, as illustrated in Figure 1(c), where «error data» refers to the quantified difference between the actual observed and the desirable collective properties which are manifested at s. The principle of afferent learning is implemented in the procedure used for modifying the wholistic behavior of artificial neural networks by *supervised* back propagation of error, effected by changing the synaptic weights of the network.

Unsupervised Learning by Cross-correlation: Cross-correlating sequences of impinging selective trigger-signals, carrying information about collective properties of their source, are exchanged between two self-organizing hierarchical physical systems, as illustrated in Figure 1(d). This bilateral «communication» by energy exchange and information transaction, where the lower e-levels play the role of «experiential receivers» (the system's hardware) and the higher s-levels play the role of «behavioral transmitters» (the system's software), is an instance of *unsupervised* learning by constant adaptation and accumulation of *experiential knowledge.* The adaptive nature of intersystem communication resides in some interactive control mechanism, which is exercised intersystemically in a feedforward (efferent) manner from high level behavioral transmitters (s and s') to respective lower level experiential receivers (e' and e). We should note, that it is possible to model this adaptive mechanism using only a small number of control parameters [Haken, 1988]).

The *interface language* in each case, relates the mapping between the hierarchical levels, in efferent (feed-forward, e→s) computation, in afferent (back-propagating s→e) supervised learning, and in triggering (cross-correlating s'→e) unsupervised learning.

Uncertainty and ambiguity enter the processes of information transfer s'*e=s, s*e'=s', in various ways. It is present in the emergence of **collective** properties in hierarchical dynamical systems. Also in the interpretation of the phenomenological output «form», as also in the context-driven **abstraction** and interpretation of impinging information. Fuzzy-theoretic and probabilistic methods may be used for the representation and management of uncertainty in these instances.

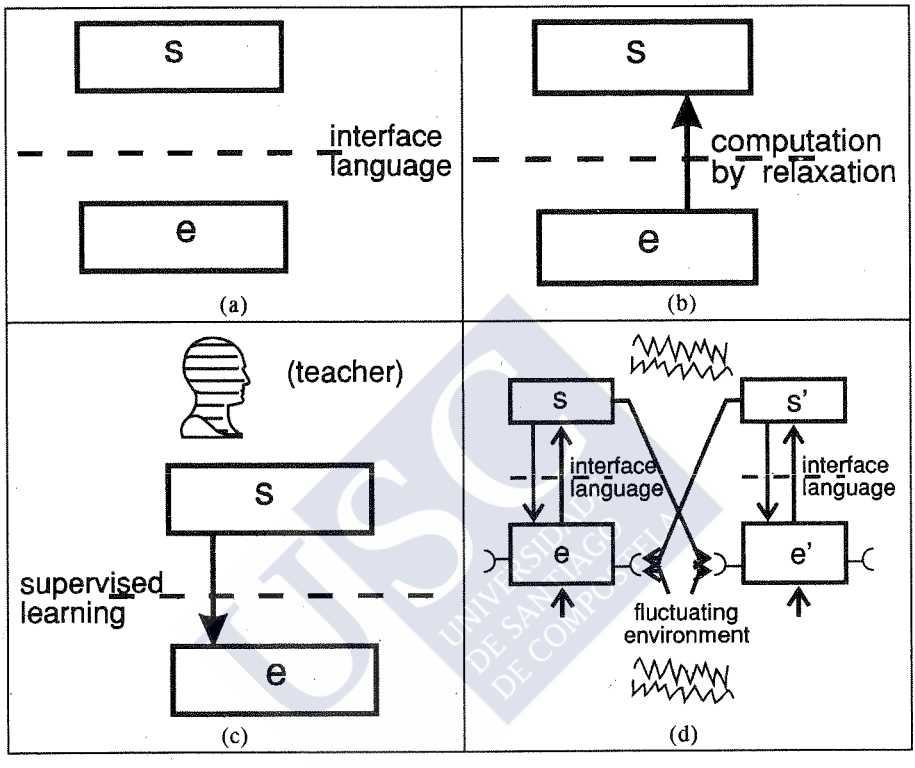


Figura 1. C&L in a hierarchical Physical System.

3. DETERMINISTIC STOCHASTICITY: C&L IN CONTROLLED CHAOS

We have investigated the temporal behavior of classical dynamical physical systems for C&L models. The object is to match the behavior of a regulated hierarchical physical system with the predictive capabilities of a corresponding model, formulated in some symbolic representation system. A rich variety of artificial neural network models have already been developed on the basis of this approach, by simulating Statistical Mechanics of Magnetic Spin Systems.

The dynamics of a hierarchical non-linear system may be modeled by some rate differential equation, like the reaction-diffusion equation,

$$q = f_{NL}(q;r) + D\nabla q + F$$

where q represents the state, r represents the control parameters, f is a non-linear function, and F represents the delta-correlated fluctuations. System dynamics may be visualized as an evolutionary motion in the state-time space [Ligomenides, 1992a,b]. If the physical system possesses a «regular» behavior, i.e. if its state trajectory leads to a fixed dynamically stable «point» in the state space, or to a dynamically stable periodic motion on a «limit cycle», then a «good» model may be formulated for it, one that will provide excellent correspondence, within some acceptable error, between the physical and the simulated behavior.

From a practical point of view, random behavior occurs only to the extend that it can not be predicted, usually because of lack of adequate information. Based on the degree of predictability, some systems are more random than others. Given the physical system's initial conditions, e.g. determined from measurements in the presence of noise and of acceptable precision limitations in the measuring device, the simulation model will compute the future states of the physical system to within some acceptable, and statistically predictable, error. In this manner, the model can serve a «predictive» role in understanding and using such «well-behaving» computing phenomena.

However, not all physical phenomena are so «well-behaved». A large class of very important turbulent - «chaotic» - physical systems are not so well behaved, in that they are characterized by «irregular» motion in their state space. They display seemingly *random* behavior that is inherently unpredictable in the sense that information obtained from making a measurement on the system is lost after some short finite time. Put in other words, the observer's current knowledge of the system's state goes to zero within some, typically short, finite time.

Observed seemingly random or turbulent behavior may be due to one of several different causes: observational uncertainty, inability to resolve many degrees of freedom (as is the case when observing Brownian motion), or *deterministic chaos* [Froehlin et al., 1981; Ruelle and Takens, 1971; Shaw, 1981]. The case of deterministic chaos is of special interest in studying C&L properties in dynamical system behavior. Special geometric structures in the state space, responsible for the chaotic behavior in dissipative dynamical systems, were named «strange attractors» [Ruelle and Takens, 1971]. A strange attractor is an «attractor» because a wide

range of initial conditions (all the initial conditions in its basin of attraction) approach the attractor asymptotically. It is also «strange» because state orbits on the attractor execute neither fixed point nor periodic motion orbits, but, instead, *wander around* erratically and randomly over the attractor. This behavior is attributed to great and sustained sensitivity to initial conditions and to the presence of noise that prevents us from determining the state of the system with precision.

We know that the states of a physical system can not be measured with absolute precision. Given the inherent sensitivity to initial conditions and the concomitant amplification of measurement uncertainty in chaotic systems, there is no hope of using chaotic models to predict physical behavior for times longer than T, the time needed for complete loss of the information gathered by measurement. We may ease somewhat the limitations on the extend of prediction by relaxing the requirements of exact correspondence between model and physical system, and by being content with a correspondence limited to between geometric and statistical characterization of the dynamics, which may be valid only to within some short-range predictive capability. Procedures have been forthcoming in the literature for constructing «closed form» descriptions with statistical and geometric characterizations based on a series of observations.

Deterministic chaos underlines the behavior of a class of physical systems of special interest, whose unpredictability is derived from sensitive dependence on initial conditions in a sustained way that exaggerates uncertainty. It is the swift amplification of our ignorance that *generates* randomness in chaos. Even with only a few degrees of freedom and a marginal source of uncertainty (from measurement or fluctuation), simple chaotic maps produce time-separated chaotic states that do not correlate, i.e. $\langle x_t x_{t+a} \rangle = 0$ unless $a=0$.

Since chaos is defined in the context of *deterministic dynamics*, its uncertainty originates from sources external to its dynamics, such as environmental fluctuations and measurement errors. Due to these causes, even minute uncertainties in the definition of state are amplified with exponential rates into large macroscopic proportions in finite time. In this way, short term determinism becomes long term stochastic behavior. Chaotic trajectories can pass many classical tests of «randomness», and even look random.

Until the definition of chaos, it was assumed that random behavior was caused by complication brought on by the presence of many irreducible degrees of freedom. Based on this assumption, Kolmogorov formulated a theory of random processes, which he defined in terms of d-th order joint probability distributions for

the observed time series. Such dynamic systems, which are not chaotic, may display deterministic predictable behavior if for some delay d, the probability density approaches a delta function in the limit of perfect measurements of state sequences. This fact indicates that the theory of random processes provides the means for coping with inadequate information; however, it makes no statements about the causes of randomness. As far as we know today, save for the only truly fundamental source of randomness is quantum mechanical uncertainty, all other physical processes are deterministic, at least in principle. Nonetheless, we call «random» many phenomena in physics, in social sciences and in economics, which have no apparent connection with quantum mechanics.

For chaotic systems, unpredictability can be quantified and given an information theoretic interpretation. Nearby orbits on a strange attractor diverge exponentially in one or more directions, in ways that rapidly amplify uncertainties in determining the system's state. The positive Lyapunov characteristic exponents quantify the average asymptotic rate of divergence, and can be interpreted as the strange attractor's *information production rate*, or alternatively, the rate at which the information contained in a measurement is lost [Shaw, 1981].

4. INFORMATION GAINED BY OBSERVING A DYNAMICAL SYSTEM

The state-time space is a cellular one because of limited precision in physical measurements of time and attributes. The notion of *information dimension* is linked to the measuring process. Let the measurement's device resolution be ε. Then, there are $1/\varepsilon$ possible numbers available for assessing the value of any variable measured by this device. If we measure the state of n variables using n such measuring devices, then this kind of scale quantization by the measuring device induces a ***partition*** of the n-state space into non-intersecting hypercubes. There are ε^{-n} boxes (partitions) in the measurement state region containing the (strange) attractor, each of dimension n, where each measured state corresponds to a box. Let $N(\varepsilon)$ be the number of boxes covered by the attractor with non-zero probability of occupation (Note that for point attractors, this is the probability distribution function). For strange attractors, these boxes will contain at least one point with nonzero probability, with some boxes having more points than others since they may be visited by the flow more often than others.

Following Farmer [Froehling et al., 1981] let $P_i(\varepsilon)$ be the **density** of probability of occurrence of the attractor somewhere in the ith box. For a set of boxes, the

probability of occurrence of the attractor is $\int_i P(x,i;\varepsilon)dx$. Apparently, $P(x,i;\varepsilon)$, the probability density function for a wide variety of initial conditions x, is a function of the resolution. A collection $\{P_i(\varepsilon)\}$ over the state space is referred to as the «coarse-grained» probability distribution of resolution ε.

Now, concerning the average information contained in a single isolated measurement made at resolution ε we have, (after Shannon),

$$I(\varepsilon) = -\sum_{i=1}^{N(\varepsilon)} P_i(\varepsilon)\log_2 P_i(\varepsilon) \quad bits$$

The ***information dimension*** D_I is the rate at which the information scales as the resolution of the measuring device improves, namely,

$$D_I = \lim_{e\to 0}\left[\frac{I(\varepsilon)}{|\log_2\varepsilon|}\right]$$

Thus, the Information Dimension D_I is given by the asymptotic value of the slope of the graph of $I(\varepsilon)$ versus $\log_2(I/\varepsilon)$. From the above, we see that if the information dimension of an attractor is known, then the information gained from a measurement of the state with resolution ε may be estimated by

$$I(\varepsilon) = D_I|\log_2\varepsilon|$$

Notice that if the probability of occurrence is the same for all the boxes of the state space, i.e. if the attractor can be anywhere –the state space has no structure–, then $I(\varepsilon)=\log_2 N(\varepsilon)$, and the Information Dimension acquires its maximum value, which is equal to the ***fractal*** dimension,

$$D_I = \lim_{\varepsilon\to 0}\left[\frac{\log_2[N(\varepsilon)]}{\log_2\left(\frac{1}{\varepsilon}\right)}\right]$$

The Fractal Dimension is then an upper bound of the Information Dimension.

Besides the accuracy of the measuring device, what limits the resolution ε in practice, is also the level of the internal (intrinsic) fluctuations in the system.

Excluding any external noise, we call these two sources «observational noise». While it does not affect the phenomenological dynamics of the system, it is manifested during a measurement of the state and it determines the resolution of measurement.

A significant aspect here is to quantify the uncertainty in assessing the dynamics of an attractor, i.e. to develop a «metric» for information uncertainty. To express $N(\varepsilon)$ in terms of Lyapunov exponents of the attractor and the resolution length ε, let the number of points on the 1D Poincare return map of the attractor concerned be M. This means that we determine the orbit with strings M bits long. The number of cells representing the attractor with degree of coarseness ε is just $N(\varepsilon)$. If we desire that each of the possible outcomes of measurement of an orbit of length M digits to be determined with some precision, then ε must have a small enough value. If the dynamics is represented adequately in each step by two possible states, say 0 and 1, the number of outcomes is clearly 2^M. Hence, ε should be chosen such that

$$\varepsilon \sim \frac{1}{2^M}$$

Let now t_c be the sampling time, or alternatively, with sampling period $\Delta t=1$, be the number of samplings probing the attractor. For an attractor represented by M points, it is meaningful that

$$t_c \sim \frac{N(\varepsilon)}{M}$$

Since, owing to the chaotic character of the dynamics, the system is completely disengaged from the initial condition x after t_c samplings, t_c can be estimated by

$$\varepsilon \exp(\lambda_+ t_c) \sim 1$$

where is the positive Lyapunov exponent. Therefore

$$t_c \sim \ln\frac{\left(\frac{1}{\varepsilon}\right)}{\lambda_+}$$

and finally,

$$N(\varepsilon) = \frac{\left(\ln \frac{1}{\varepsilon}\right)^2}{\lambda_+ \, ln2}$$

There is a relation between Lyapunov Exponents and Information Dimension, which will not be discussed here for lack of space.

5. NON-LIPSCHITZIAN UNPREDICTABILITY

If the Caushy-Lipschitz condition is satisfied, once the initial and boundary conditions of the receiver are defined in the absence of noise, then, considering an intruding input message, we may determine the future course of the state vector. The state trajectory of the system will be determined *uniquely* in this case, if infinite precision may be achieved in measurements of attribute and time. The receiver will deviate from equilibrium under the influence of the input message and will tend, eventually, to approach a homeostatic «attractor» state. In doing so, the receiver also alters its own internal structure and organization in response to the incoming «messages» from the surrounding world, as we have discussed previously. Prediction in chaotic and in non-Lipschitzian systems may become possible under very special conditions of controllability, offering the possibility of defining new computing architectures.

Contemplate viewing *and interpreting* the physical world as a vast, perpetually computing machine. Consider modeling the structure and behavior of classical (Caushy-Lipschitz condition obeying) dynamical physical systems in the lingo of «computation» [Haken, 1988; Ligomenides, 1992a,b; Nicolis and Prigogine, 1977; Zak, 1991]. The objective would be to «match» the behavior of physical phenomena with the predictive behavior of models formulated and interpreted within some symbolic representation of a procedural computing language. In fact, this has been common practice in Physics research and in the development of related technology, and it has contributed vastly to progress in both these fronts. String languages of symbolic logic and algorithms are formulated and executed on digital computers for this purpose. Also, other extra-logical computational languages may be designed to simulate the behavioral modes of more complex physical systems. Artificial Neural Networks are examples of symbolic approximations and hardware implementations simulating the dynamics of complex cooperative physical systems [Ligomenides, 1991b]. We now seek extensions to today's ANN models.

If a physical system possesses a «regular» behavior, e.g. it is a classical physical system conforming to the Caushy-Lipschitz condition, its state trajectory will lead to some «fixed point attractor» of stability, or to a stable periodic «limit cycle attractor», in its state space. A wide range of initial conditions within the attractor's «basin of attraction», will approach the stable condition asymptotically. A «good» model should provide *acceptable* (within tolerable error) correspondence between the physical and the simulated behavior. The model's representation of state and the determination of state transitions should be compatible with measurements to within acceptable deviations. Then, the computational model can serve a predictive role and can help in understanding the dynamics of related physical phenomena. This is especially important when the dynamics may be interpreted as contextually meaningful «computations».

Modeling physical systems as *«computationally meaningful»* can be extended to physical phenomena with «irregular» dynamics. A large class of interesting turbulent «chaotic» physical systems have been the subject of intensive studies in recent years. These, not so well behaving, systems display irregular motion in their state space, which is inherently unpredictable in the sense that information obtained by measurements at time t is lost some short finite time thereafter; in some such systems, the observer's knowledge about the system *goes abruptly to zero at critical bifurcating points* along the state trajectory.

From an information processing point of view, special interest in these systems has been developed more recently, mainly because of discovering «order within chaos» in phase-coherent chaotic systems. Deterministic chaos and predictive prototypes with turbulent dynamics, dealing with observational uncertainty and with resolving many degrees of freedom, are of interest because of their special «computation» capabilities. Special geometric structures in state space, such as strange attractors, allow information-theoretic interpretation of unpredictability using deterministic models.

Predictive models of computation which are based on chaotic dynamics are limited to predicting behavior only within the time that information is preserved about their past. Often, we must also be content with limited correspondence between the physical system's and the predictive model's behavior to «closed form» descriptions with geometric and statistical characterizations, valid only for short range predictive capabilities. In spite of the many difficulties, we expect that modeling the wide variety of multistable, nonlinear physical behavior, with controllable dynamics, will lead us to designing cooperative computers with

extended computational capabilities; it should also extend our understanding of biological information storage and processing.

Because of space limitations, in following, we will view briefly the C&L properties of non-Lipschitzian systems with unpredictable dynamics [Zak, 1989, 1991].

An «ordinary» attractor would be governed by the equation

$$\tau \frac{dq}{dt} = -(q - q^*)$$

having a solution

$$q(t) = q^* + [q(0) - q^*]\, e^{-\frac{1}{\tau}}$$

showing that the attractor is reached in a decaying exponential manner.

If we modify the dynamics by replacing the right-hand term by a singular function

$$\tau \frac{dq}{dt} = -(q - q^*)^{\frac{1}{3}}$$

we have the solution

$$q(t) = \begin{cases} q^* + \left[(q(0) - q^*)^{\frac{2}{3}} - \dfrac{2t}{3\tau} \right]^{\frac{3}{2}} & \text{if } t < t^* \\ q^* & \text{if } t \geq t^* \end{cases}$$

Now the attractor is reached in a finite time t^* and it is referred to as a «terminal» attractor, with an infinitely stable singular solution $q = q^*$.

Consider now an assembly of «neurons» whose dynamics is given by the following set of non-coupled equations:

$$q_i = \gamma \sin^{\frac{1}{3}} \frac{\omega}{\alpha}\, q_i \sin(\omega t) + \varepsilon_i(t)$$

where γ, ω and α are constants, and $\varepsilon_i(t) \rightarrow 0$, $i = 1, 2, ..., n$ are vanishingly small control inputs (random or prescribed) that regulate the behavioral «personalities» of the neurons. Unpredictability at critical points may be converted to prescribed behavior. Notice the importance of *experiential learning and of adjustment* of the control parameters. The system can, indeed, *forget its past* and adapt its behavior to environmental changes in finite time.

To the above assembly of independent neurons we may impose an inter-neuron coupling objective function in the form of a functional to be minimized:

$$\Phi_i = \int_0^{\tilde{t}} [q_i(t) - f(t)]^2 \, dt$$

where f(t) is a prescribed function. We also let the learning principle be

$$\omega_i = \left[\omega_0 + \Phi_i - \frac{1}{n} \sum_{j=1}^{n} \Phi_j \right] - \omega_i$$

where ω_i is the modifiable frequency of periodic excitations.

As a result, the assembly will acquire different «personalities of activity», depending on the temporal profile of $\varepsilon_i(t)$ of the neurons, and on whether the performance Φ_i of each neuron is above average (will receive higher energy input and will act faster), or below average (will slow down and go out of competition).

Such «neurons» can learn from *experience*, can memorize «useful» behavior and can improve performance. They can also cooperate and compete with each other. Although such capabilities may be achieved also by classical Lipschitzian systems (even by logical digital computers), what is most important in non-Lipschtzian systems is the capability of a uniform assembly of identical neurons to acquire different patterns of behavior without special interference from outside.

Notice that the neurons in the non-Lipschtizian assembly are uncoupled (and therefore they can perform independently parallel tasks), while they become coupled *at the critical points*, where the system can spontaneously alter the location and types of its attractors in response to its learning dynamics. Also, because of the failure of the Lipschitz-condition at the critical points, the transient periods of convergence between critical points is bounded. Thus, the unwelcome limitation of

Lipschitzian dynamics, that of requiring infinite time to approach a regular attractor or to escape a regular repeller, may now be removed.

6. CONCLUSIONS

Modeling a wide variety of multistable, nonlinear physical systems with controllable dynamics, will lead us to designing cooperative C&L systems with extended computational capabilities. We will deal with uncertainty in a natural manner, and we will exploit its controllability. This experience may also provide us with clues about biological C&L processes.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

REFERENCES

Deng, J., «Introduction to Grey System Theory», *The Journal of Grey Systems*, Vol. 1, (1989), 1-24.

Froehling, H., Crutchfield, J.P., Farmer, J.D., Packard, N.H. and Shaw, R.S., «On Determining the Dimension of Chaotic Flows», *Physica*, Vol. 3D, (1981), pág. 605.

Haken, H., *Information and Self-Organization*, Series on Synergetics, Vol. 40, Springer-Verlag, 1988.

Ligomenides, P.A., «Modeling Uncertainty in Human Perception», in Bouchon, B. and Yager, R.R. (Eds.), *Lecture Notes in Computer Science: Uncertainty in Knowledge-Based Systems*, Vol. 286, Springer-Verlag, (1987).

Ligomenides, P.A., «Computation and Uncertainty in Regulated Synergetic Machines», in Bouchon-Meunier, B., Yager, R.R. and Zadeh, L.A. (Eds.), *Lecture Notes in Computer Science: Uncertainty in Knowledge Bases*, Vol. 521, Springer-Verlag, (1991a).

Ligomenides, P.A., «Cooperative Computing and Neural Networks», in Prieto, A. (Ed.), *Lecture Notes in Computer Science: Artificial Neural Networks*, Vol. 540, Springer-Verlag, (1991b).

Ligomenides, P.A., «Learning Paradigms and Intelligent Systems», *Proc. International Symposia on Information Sciences, ISKIT'92*, Kyushu Institute of Technology, Iizuka, Japan, Julio, (1992a), 12-15.

Ligomenides, P.A., «Computation and Learning Paradigms for Biologically Inspired Intelligent Computing», *Proc. 2nd International Conference on Fuzzy Logic and Neural Networks*, Iizuka, Japan, Julio, (1992b), 17-22.

Ligomenides, P.A., «Tools for Experiential Recognition», *International Journal of AI Tools*, Vol. 2, World Scientific Publ., (1992c).

Ligomenides, P.A., «Cooperative Computers with Extended Computational Capabilities», *Proc. European Symposium on ANNs*», Brussels, Belgium, Abril, (1993a), 7-9.

Ligomenides, P.A., «Computational, Adaptive and Evolutionary Processes in Cybernetic Systems», *Proc. 9th International Congress of Cybernetic Systems*, New Delhi, India, Enero, (1993b), 18-23.

Nicolis, G. and Prigogine, I., *Self-Organization in Nonequilibrium Systems*, Wiley, New York, 1977.

Ruelle, D. and Takens, F., «On the Nature of Turbulance», *Communications in Mathematical Physics*, Vol. 20, (1971), pág. 167.

Shaw, R., «Strange Attractors, Chaotic Behavior, and Information Flow», *Zeitsschrift fuer Naturforschung*, Vol. 36a, N. 80, (1981).

Zak, M., «Terminal Attractors in Neural Networks», *Neural Networks*, Vol. 2, (1989).

Zak, M., «An Unpredictable-Dynamics Approach to Neural Intelligence», *IEEE Expert*, Vol. 6, N. 4, (1991).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

El modelo difuso como Programa de Investigación Lingüística[1]

Teresa Moure
Departamento de Filología española, Teoría de la Literatura y Lingüística General
Universidad de Santiago de Compostela

1. INTRODUCCIÓN

Dado el carácter interdisciplinar del modelo que se sugiere en estas páginas, convendría quizás, antes de elaborar sus pautas, ubicar la Lingüística en el conjunto general del conocimiento. En una primera aproximación, la Lingüística es la ciencia que estudia el lenguaje humano, un fenómeno complejo que, al relacionarse con factores muy variados, reporta a la disciplina ciertas peculiaridades.

El lenguaje posee un sustento material ligado a las capacidades fisiológicas de los órganos de articulación, a la naturaleza acústica del sonido o a la riqueza de las conexiones neuronales. En este sentido, la investigación que lo tiene como objeto atiende a una faceta de la realidad observable y puede adquirir carácter experimental, como ocurre en la síntesis de voz o la traducción automática, por citar los casos de mayor impacto tecnológico. Sin embargo, el lenguaje no parece

[1] En la elaboración de este trabajo he contado con el apoyo de tres maestros y amigos. Alejandro Sobrino guió mi interés por la Lógica Borrosa y me invitó a participar en este volumen. Milagros Fernández Pérez y Guillermo Rojo estuvieron, como siempre, atentos a mi trabajo y me proporcionaron valiosas indicaciones desde la óptica lingüística. Por su aliento y por su esfuerzo, he contraído con todos ellos una nueva deuda de gratitud.

someterse a procesos de cuantificación y verificación similares a los que se aplican en otros campos, con lo que admitir sin matices un enfoque empírico sería reconocer entidad a la disciplina únicamente en la medida en que emula procedimientos que gozan del prestigio de la ciencia natural.

El lenguaje es también el vehículo fundamental de la comunicación humana y, por tanto, se liga a los factores sociales y culturales que envuelven una comunidad. Con esta óptica, la Lingüística alcanza una proyección hermenéutica, encaminada a describir las distintas lenguas[2]. Evidentemente, aun cuando esa descripción aporte valoraciones, no se limita a aspectos meramente normativos. El rigor se calibra por la capacidad de recoger toda desviación –social, geográfica o individual– para explicar los índices de variabilidad entre las lenguas y, en general, las constantes universales del lenguaje humano.

2. LA TRADICIÓN DISCRETA EN LINGÜÍSTICA

Con esta perspectiva al fondo, la Lingüística se ve concernida por problemas de categorización en dos niveles distintos. En principio, el lenguaje trasluce nuestra percepción del mundo y es el instrumento de que nos valemos para discernir sus propiedades. Al declarar que un objeto *X es verde*, se pone de manifiesto la capacidad de la retina para discriminar diferencias en las ondas de luz y, también, el procesamiento de esa información física y su posterior clasificación mediante patrones de orden cognitivo. En un sentido parcialmente diferente, la Lingüística, como toda ciencia, pretende reducir alguna dosis de incertidumbre elaborando un aparato conceptual que otorgue regularidad a sus predicciones. La declaración *X es verde* introduce una forma fónica –['berde]– y la asocia a determinada clase de palabras –el adjetivo– con la que ordinariamente expresamos propiedades de los objetos. La investigación lingüística se perfila, entonces, como una sistemática de la lengua, que organiza las categorías propias del objeto de estudio en nuevos moldes o **categorías del metalenguaje**[3].

---

[2] La calificación de *hermenéutica* apunta, en la línea de Itkonen [1978], a comprensión y reflexión como fases ineludibles del quehacer científico. Así entendido, el método hermenéutico resulta compatible con técnicas empíricas porque, como ha indicado Fernández Pérez [1986: 39], la disciplina acoge marcos teóricos parciales plenamente contrastables.

[3] Por supuesto, las categorías del lenguaje y las del metalenguaje guardan relación entre sí y una teoría no estaría bien construida si sus primitivos careciesen absolutamente de correlato objetivo. Las unidades del metalenguaje, como fonema o morfema, no tendrían razón de existir si no hubiese diferencias materiales entre (alo)fonos y (alo)morfos.

Para unas y otras, la categorización lingüística remite a una tradición discreta, es decir, supone unidades de contornos delimitados, sin puntos de contacto o zonas de concurrencia. Las distinciones cualitativas son bruscas ya que cada unidad, según sus atributos, encaja o no en una clase y cualquier duda sobre su asignación sólo puede imputarse a deficiencias de la teoría. Además, esos atributos que garantizan el salto categorial se conciben como propiedades necesarias y suficientes: necesarias por cuanto operan en conjunto y deben estar presentes en su totalidad para asegurar el reconocimiento de la categoría; suficientes porque su validez viene dictada por la oposición a otras propiedades no-esenciales o accesorias. El reparto se verificaría del siguiente modo:

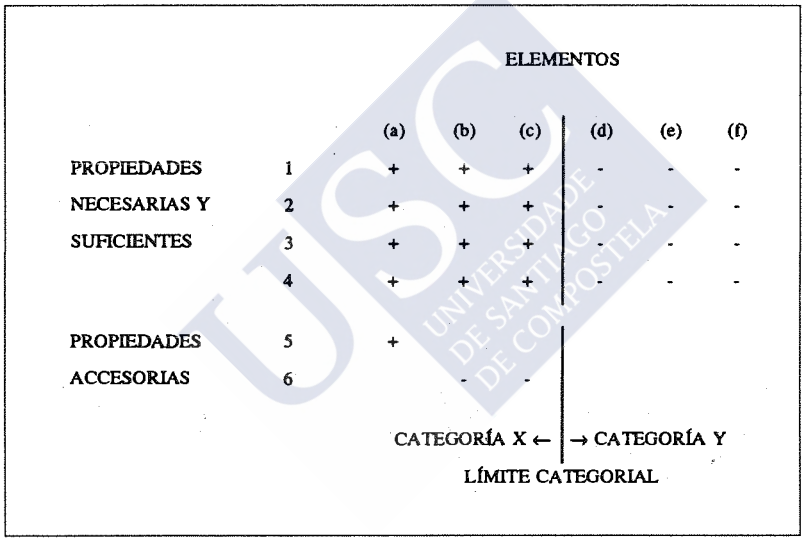


Figura 1

Las propiedades (1-4), y sólo ellas, establecen un corte nítido entre las series (a-c), por un lado, y (d-f) por otro, conformando dos categorías X/Y. Otras propiedades -(5), (6)- no intervienen en la asignación categorial por su condición de rasgos accesorios, que únicamente ofrecen algunos de los ejemplares tomados en cuenta.

Durante el proceso de dilucidación surgen, sin embargo, entidades que combinan rasgos propios de clases enfrentadas, al estilo de lo que ocurre en la figura 2.

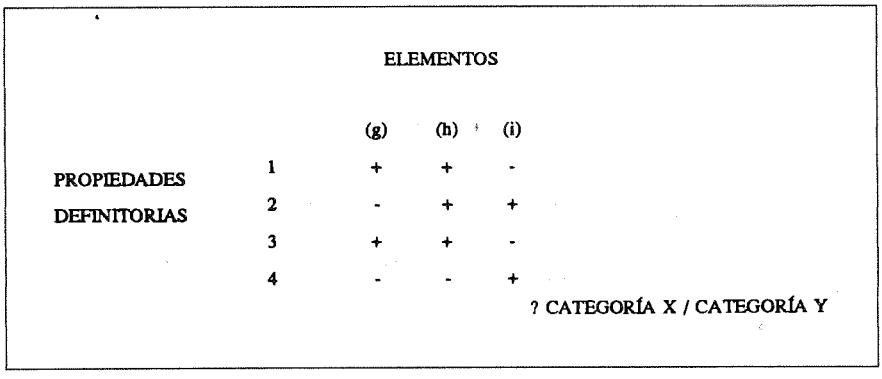


Figura 2

Así, en el inventario de clases de palabras del español los participios disfrutan de rasgos característicos de los verbos y de otros propios de los adjetivos. Del mismo modo, ciertas variantes idiomáticas muestran, cruzados, atributos definitorios de dos lenguas distintas. Y no se trata de mero desatino en la definición sino de una imprecisión intrínseca, para la que se ha probado la imposibilidad de obtener mejores perfiles; en otras palabras, de un ejemplo de vaguedad.

Sólo una sistemática ingenua –y, desde luego, inadecuada– admitiría forzar los casos vagos para mantener el requisito previo de que existen barreras nítidas entre las clases. Quizá los dialectos de una lengua o las clases de palabras como, en general, todas las categorías lingüísticas no permitan establecer condiciones de pertenencia porque no son verdaderos conjuntos, al menos en el sentido usual que el término *conjunto* tiene en Matemáticas. Pero si se pudiera aceptar la noción de **grado de pertenencia** a una clase, se captaría una información adicional que esos casos-límite proporcionan, precisamente por su naturaleza intermedia.

La evidencia de que existen estados de cosas graduales rompe con el prejuicio de la cuantificación en la ciencia. *No-cuantificable* no significa necesariamente *imposible de medir* sino, más bien, *necesitado de técnicas de medida diferentes*. En este sentido, la Lingüística no sería una excepción; casos de vaguedad similares se encuentran en la transición de los elementos químicos o en las taxonomías biológicas y para todos ellos cabe adoptar las técnicas cualitativas del razonamiento ordinario. A fin de cuentas el pensamiento humano no se define por la nitidez de sus contornos: hacemos cálculos aproximados, inferencias difusas, predicciones probabilísticas y nuestro discurso se puebla de entidades que no catalogamos en clases bien definidas. Sin embargo, de esas imprecisas mediciones

surgen categorías depuradas con las que reconocer modelos, procesar información o ejercitar nuestras capacidades abstractas.

En medio de ese panorama general, algunas voces aisladas reclamaron mayor flexibilidad para los supuestos teóricos. Ni siquiera es necesario cuestionar el método para alterarlo de hecho; la historia de los conceptos gramaticales proporciona una amplia gama de casos vacilantes cuya asignación a una clase hubo de ser dudosa por fuerza. Son bien conocidas las dificultades para establecer barreras entre el inventario léxico de una lengua y los morfemas fijos que remiten a propiedades de la estructura gramatical. Los criterios chocan con formas como el prefijo *[re-]* de *reemplazar* o *reproducir*, que posee un significado fijo como los lexemas –en este caso '*volver a*', '*repetir*'– pero que, al igual que los morfemas, depende de otro elemento léxico. También las cópulas como *ser* y, en menor medida, *estar* mantienen una situación intermedia pues, aunque reciben morfemas que las identifican con los verbos, carecen de significado propio. Y concesiones de este tipo cubren todos los ámbitos de la disciplina: las fronteras entre lenguas –las isoglosas– siempre entrañaron continuidad, los matices de contenido que separan sustantivos concretos de abstractos se proyectan en una línea donde resulta difícil establecer el lugar exacto del corte y todas las formas no-personales del verbo participan de una doble naturaleza[4]. Al fin, algunas funciones sintácticas, como complemento directo e indirecto, poseen un perfil tan homogéneo que los hablantes acaban por confundir sus marcas y alteran la norma en beneficio de leísmos, laísmos y loísmos.

Si se reflexiona un momento sobre estas cuestiones, habrá que reconocer que la gramática tradicional era un verdadero caldo de cultivo de la vaguedad porque, al cuestionar siempre sus propias delimitaciones, fomentaba una demolición de fronteras inusual en otros campos del saber. Cuando Sapir [1921][5] sugiere una gradualidad inherente a los conceptos gramaticales se formula por primera vez una

---

[4] [...] ni la gramática ni la lógica han llegado nunca a establecer una división segura entre los nombres o conceptos concretos y los abstractos. [...] hay objetos más y menos concretos, y más y menos abstractos. Y como los distintos grados de concreción y de abstracción corren sin interrupción de un extremo al otro, resulta [...] imposible trazar la división exacta entre los nombres concretos y los abstractos. [Alonso y Henríquez Ureña, 1971].

[5] Las palabras de Sapir son bien elocuentes a este respecto [Sapir, 1921: 129]:

> [...] el lenguaje se debate entre los dos polos de la expresión lingüística -contenido material y relación-, y [...] estos polos tienden a conectarse uno con otro mediante una larga serie de conceptos de transición.

hipótesis de transición categorial capaz de echar por tierra suposiciones simplistas y resolver semejante confusión. Pero estos testimonios, como otros que podrían aducirse, resultan de procedimientos habituales en Lingüística, que no pueden valorarse como antecedentes de un modelo difuso. Su sola presencia apunta, sin embargo, la necesidad de un cambio metodológico.

Desde un punto de vista histórico, el mayor lastre que operó contra la formación de una Lingüística no-discreta fue el papel decisivo del Estructuralismo en su conformación como disciplina científica. En este marco, el signo lingüístico se define como asociación convencional de **significante** y **significado** o, en otras palabras, de una expresión codificada y del valor que se le atribuye en el intercambio comunicativo. El reparto parecería adjudicar al significado los aspectos psicológicos y culturales o las connotaciones derivadas del contexto de uso. Pero, contra cualquier exceso en esta línea, se desglosa cada componente del signo en dos planos: la **sustancia** cognitiva, que recogería las categorías y propiedades del mundo real, y la **forma**, un principio de ordenación inherente a los sistemas y capaz de organizar ese material *universal* de modo diferente en cada lengua.

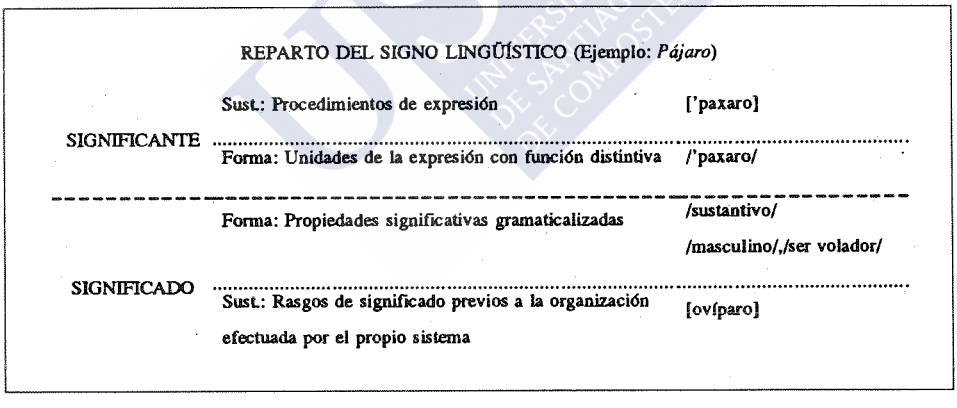


Figura 3

En una hermosa y muy divulgada imagen, Hjelmslev[1943] indica que la forma, al deslindar los cortes pertinentes, se dibuja sobre la sustancia no conformada como una red que proyectase su sombra sobre una superficie indivisa. La lengua queda limitada a forma, en concreto será *una forma entre dos sustancias*, la cognitiva y la fónica. Al calificar de *extralingüístico* todo factor que no provoque variaciones inmediatas en el otro plano, se estatuye un poderoso mecanismo para filtrar

intromisiones de la realidad difícilmente controlables y se prima una visión relacional, donde las unidades se sustentan en el haz de oposiciones que las diferencia. Como consecuencia, la vaguedad del lenguaje queda marginada de la teoría lingüística por su carácter sustancial: concierne a la referencia designada y no al significado de la unidad que la actualiza en el discurso. La pertinencia del significado sólo podrá establecerse en un sistema concreto y siempre que se verifiquen correspondencias entre las dos caras del signo. Así, la oposición animado/inanimado, vigente en español entre *bello* y *bonito* se suspende en otros contextos, como en su antónimo *feo* y no por ello una *acción* es *fea* del mismo modo que una *persona*; de hecho otras lenguas marcan esa diferencia de la realidad. Ni siquiera podría decirse que el significado de *feo* sea vago, a lo sumo sería impreciso por comparación a sistemas que afinasen más.

Además, el Estructuralismo acusó el carácter arbitrario de la relación entre las dos caras del signo bajo el supuesto de que si el vocabulario de las lenguas es diferente no estará fundamentado en relaciones naturales con los objetos. Esta formulación se estilizó hasta el punto de considerar igualmente arbitraria la selección de significados. Cada comunidad tendería a discernir con mayor detalle las entidades más destacadas de su entorno, dejando a un lado aspectos secundarios. Una serie de factores ajenos consolidan la asociación convencional de esos significados con las correspondientes formas fónicas. Así pues, aunque el punto de partida esté en las divisiones cognitivas, la lengua resulta ajena a ellas: el inglés *brown* no tiene el mismo significado que el francés *brun* porque incluye, al menos, los designata que en francés se aluden con las formas *brun*, *marron* y *jaune*.

Esta autonomía del lenguaje se ve afectada por la irrupción de la Gramática Generativo-Transformacional y su concepto de *modularidad de la mente*. Según esta hipótesis, el cerebro humano alberga una serie de subsistemas que operan independientemente. Uno de ellos, la **facultad lingüística**, se contempla como un mecanismo computacional, capaz de generar toda oración de una lengua mediante operaciones recursivas de reglas que asignan a cada secuencia una representación fonética y una interpretación semántica. Las lenguas humanas resultan, al fin, de la interacción entre esa facultad lingüística y otros dos módulos: la **competencia pragmática**, que aporta el uso apropiado de las reglas para distintos fines, y el **sistema conceptual**, encargado de percibir, categorizar y simbolizar; en definitiva, de diferenciar el lenguaje humano de los sistemas de comunicación animal. Una vez más, el estudio evita registrar el modo en que los seres humanos conceptualizan el mundo. Las aproximaciones generativas al significado consagran análisis componenciales que explican la capacidad de los hablantes para emitir o captar un

número potencialmente infinito de oraciones, sin información sobre el contexto y al margen de variaciones individuales.

Con una teoría replegada sobre postulados abstractos la realidad psicológica sólo podría haber penetrado en Lingüística a través de su vertiente aplicada, del estudio de los factores que intervienen en la adquisición del lenguaje o en sus patologías, pero este tipo de investigaciones comenzaría su andadura bastante más tarde. Mientras tanto, el carácter discreto de las unidades garantizaba sobriedad y economía, al apoyar los conceptos en oposiciones binarias del tipo $p\ W\ q$. Por contra, la eliminación de la vaguedad condenaba la investigación a una regularidad falaz, que debería incrementarse con soluciones particulares para los casos que no parecieran ajustarse a ella[6].

3. UNA PROPUESTA ALTERNATIVA: LA TEORÍA DE PROTOTIPOS

Frente a esta tradición de rigidez categorial, surge, al principio como mera intuición, la hipótesis de que las clases poseen confines borrosos. Uno de los más poderosos antecedentes de esta perspectiva arranca de la Filosofía del Lenguaje. En un intento de superar el período analítico anterior, Wittgenstein [1953] equipara el significado de un término a su contexto de uso y, desde entonces, las categorías semánticas no serían ya discretas y absolutas sino, más bien, difusas y contingentes. Una relación de parentesco –un aire de familia– mantiene la cohesión entre varios miembros de una misma categoría y, sin transiciones bruscas, la traspasa y penetra en el área de influencia de otra categoría.

En disposición gráfica, *A* se parecería a *B*, *B* a *C*, *C* a *D* y así sucesivamente. Las diferencias entre los miembros de categorías contiguas en la transición carecen de valor significativo pero, en tramos más extensos, -entre *A* y *D*- las series apenas tienen ya nada en común.

Las primeras huellas de esta nueva perspectiva llegan de la Antropología lingüística y la Psicología Cognitiva que establecen universales semánticos de base biológica, rompiendo con la arbitrariedad que el Estructuralismo había imputado

[6] Debo señalar, sin embargo, que tanto el modelo estructuralista como el generativo perciben continuidades. Para discernir fonemas distintos, la fonología estructural admite un *margen de seguridad*, o camino intermedio entre las realizaciones fónicas. La versión chomskyana, por su parte, se refiere a *grados de adecuación*, *grados de gramaticalidad* y *grados de aceptabilidad*. Parámetros de este tipo denuncian las dificultades de encajar los primitivos en moldes discretos.

al significado. De las comprobaciones experimentales de Berlin y Kay [1969] o Rosch [1975] se deduce la presencia de **focos** en las categorías o ejemplares que los individuos reconocen más deprisa, aprenden antes y emplean con mayor frecuencia. Al tiempo, en Semántica, Putnam [1975] identificaba estereotipos socioculturales para dar cuenta de las divisiones del significado, una visión con antecedentes en Black y su tratamiento de la vaguedad [1949]. Estaba naciendo la Teoría de Prototipos.

Una *ambulancia* es, con toda seguridad, un vehículo. Pero si se nos pidiera opinión sobre la aceptabilidad de considerar vehículo un *patinete*, la duda sería razonablemente mayor. Y de un *tanque* o de un *carro de combate* sólo nos atreveríamos a decir que son vehículos en cierto modo. Las respuestas son escalares porque los miembros de la clase no gozan del mismo estatus. Al basarse en criterios firmes, la perspectiva tradicional era tajante con respecto a la condición de miembro pero, una vez resuelta, los individuos incluidos en una clase adquirían el mismo rango. La teoría de prototipos, en cambio, acepta un espacio continuo tanto para graduar el salto de una a otra categoría como para estructurar el interior de cada una.

Los criterios definitorios de clase incluyen ahora factores extralingüísticos, derivados de nuestro conocimiento del mundo y, paralelamente, pierden los atributos de necesarios y suficientes. Las categorías, lejos de su anterior uniformidad, se ven dotadas de cierta organización interna, con un **centro** ocupado por los ejemplares más representativos y una **periferia** indeterminada, donde cabe mayor heterogeneidad. En lugar de cortes bruscos, la comparación de los miembros de cada clase permite ordenarlos por su **similaridad** al prototipo. En las primeras formulaciones, ese prototipo es, al tiempo, una idealización confeccionada por los criterios definitorios y el ejemplar que se ajusta en mayor medida a ese ideal. Posteriormente, el prototipo se convierte en una combinación de propiedades que no precisan actualizarse conjuntamente; una entidad abstracta, confeccionada sobre las propiedades típicas. Con esta modificación se resolvía el problema de la existencia de varios ejemplares igualmente idóneos -como *ambulancia* y *camión* en el ejemplo anterior- y se eliminaban los rasgos individuales que no fueran pertinentes para la definición de la categoría en conjunto. Se trata, pues, de ordenar las propiedades en términos de importancia relativa para luego insertar cada ejemplar en el lugar oportuno, en función del número y calidad de los requisitos que satisfaga.

Por supuesto, no todos los terrenos son idóneos para este enfoque. Las investigaciones dedican especial atención a los fenómenos de percepción –como el

color–, las especies naturales o los artefactos. En otros dominios, el tratamiento fue deficitario y los análisis multiplicaron innecesariamente el modelo, llegando a sobrepasar su versión estándar[7]. Ante tales deficiencias la teoría de prototipos parecería, más que un prototipo de las teorías sobre la categorización, un mero ejemplar periférico.

4. PROCEDIMIENTOS DE MEDIDA CUALITATIVOS

La impronta cognitiva de los prototipos delataba un origen psicológico que la Lingüística había rechazado como objeto de interés. Con esta desventaja inicial sólo se podría propagar al terreno gramatical apoyándose en un aparato formal muy desarrollado, que encontró en la Teoría de conjuntos borrosos.

En su larga andadura, la Lógica clásica sorteó toda imprecisión apelando al principio del tercio excluso –*tertium non datur*–, que decidía los análisis en favor de dos valores bien establecidos, verdad y falsedad. Las zonas intermedias sólo podían responder a la falta de rigor del lenguaje pero, a partir de las investigaciones de Zadeh [1965], toma cuerpo una teoría sobre la vaguedad capaz de superar el estadio informal e intuitivo en que se hallaba sumida.

La hipótesis de Zadeh arranca de la constatación de que, ante determinados fenómenos, las técnicas cuantitativas resultan imposibles de aplicar, cuando no abiertamente inadecuadas. La única alternativa pasaría por desarrollar modelos semánticamente poderosos, adoptando los mecanismos habituales del razonamiento ordinario. Si en nuestras pesquisas cotidianas basta una caracterización aproximada, bien podría ocurrir que algunos tipos de investigación no reclamaran una alta precisión numérica. Sus categorías serían clases borrosas, para las que la condición de pertenencia no es un parámetro binario al que se pueda responder *sí* o *no*, sino una matriz compleja. Como puede verse, el marco teórico traza un paralelo con la teoría de conjuntos aunque el radio de aplicabilidad sea ahora superior. Es decir, si las clases carentes de límites claros se consideran conjuntos borrosos, los conjuntos definidos sólo podrán interpretarse como una manifestación particular de aquéllos.

Como concepto básico, un conjunto borroso A se caracteriza por una función $\mu_A(x)$ que calibra el grado en que el miembro x pertenece a dicho conjunto. Aunque

[7] Vid. Lakoff [1986].

no es imprescindible, $\mu_A(x)$ suele producir un valor situado en el intervalo [0,1], donde 0 corresponde al rango de *no-miembro* y 1 al de *completamente miembro*. Al asignar valores a los estados vagos, se corre el riesgo de crear precisión donde no la hay, como en la perspectiva tradicional; un riesgo que la lógica borrosa evita primando el proceso seguido sobre el resultado numérico, o bien utilizando etiquetas cualitativas. De acuerdo con ese patrón, el siguiente gráfico expresa la variable *altura* a través de una etiqueta –*alto*–, dos conectivas –*y*, *pero*– y algunos limitadores del tipo *muy*, *bastante*, *mediano*.

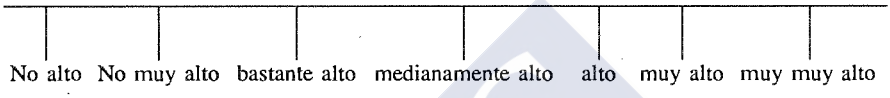


Figura 4

Determinadas disciplinas precisan, desde luego, valoraciones exactas de los datos empíricos, pero no es éste un requisito imprescindible y su generalización provocaría una imagen deformada del conocimiento científico y una epistemología vulgar. La vaguedad elude aproximaciones escrupulosas y reclama especial atención al contexto, como puede apreciarse en Biología o Química, en Ciencias políticas y económicas, en Sociología, Psicología, en el diagnóstico médico o en la gestión y el procesamiento de la información. En general, aparece en aquellos campos donde el sistema es excesivamente complejo o mal definido para admitir análisis numéricos. Semejante perfil hace de la Lingüística un claro candidato a esta aproximación.

Además de su fuerza para organizar distintos tipos de conocimiento o de sus aplicaciones tecnológicas más sofisticadas, otras razones avalan la representación de conceptos no-exactos mediante conjuntos borrosos. De entrada, es una teoría rica en operaciones, que se adapta a los requisitos de las disciplinas y permite que las propiedades matemáticas, construidas para objetos ideales, se proyecten eficazmente sobre objetos naturales o conceptuales. Por otro lado, si la Lógica y la Teoría de Conjuntos han proporcionado lenguajes convincentes para la ciencia natural, cabe esperar que sus formulaciones más amplias construyan modelos útiles para las ciencias no-exactas.

5. LA PENETRACIÓN DE MODELOS NO-DISCRETOS

Aunque prototipos y conjuntos borrosos hicieron del lenguaje uno de sus instrumentos preferidos, no alcanzaron en Lingüística la repercusión que cabría esperar. La penetración de los modelos difusos –como en adelante denominaremos a estas perspectivas no-discretas– se restringe a la vaguedad material del lenguaje, la cual concierne a la sustancia del contenido, no a la forma. Por supuesto, cabría plantearse si los escrúpulos del Estructuralismo merecían un precio tan alto como era desvincular la disciplina de otros estudios sobre el lenguaje. La respuesta ha de ser afirmativa de tener en cuenta, además de la coherencia inevitable en el quehacer científico, los escasos logros que, en un principio, se obtienen en esta dirección.

La teoría de prototipos se vuelca sobre la Lexicografía, que concibe como una faceta específica del problema general que le compete: la categorización. Sus argumentos giran en torno a la incapacidad para aprehender los significados de las palabras desde rasgos semánticos necesarios y suficientes. Ante aproximaciones que definían la forma *bachelor* (=*soltero*) mediante tres atributos: 'varón', 'adulto' y 'no-casado', Fillmore [1975] señala la existencia de ciertos individuos a los que la palabra no se aplica aunque satisfagan los rasgos, como el Papa. No obstante la crítica de Fillmore puede ser asumida sin adoptar su prototipo socio-cultural de *soltero* como el significado de esa forma léxica. De entrada, identifica una técnica particular de descomposición en primitivos con toda la Semántica precedente, lo cual no es exacto[8]. Además, sólo demuestra la inadecuación de ciertos atributos. Si, en vez de 'no-casado', utilizara *'que pudiendo casarse no lo ha hecho o ha dejado de hacerlo'* [Coseriu, 1990: 255], el análisis sería más atinado y no por ello habría cambiado la técnica.

En otra famosa aplicación de los prototipos, Coleman y Kay [1981] defienden que el significado intuitivo del verbo *to lie* (=*mentir*), algo como '*no decir la verdad*', resulta insuficiente puesto que los hablantes no aplican esta forma en caso de mentiras inocentes o errores involuntarios ni en discursos metafóricos. Su estudio experimental les lleva a considerar tres atributos simultáneos: '*falsedad del discurso*'; '*conocimiento por parte del emisor de esa falsedad*' e '*intención del emisor de engañar a su interlocutor*'. En realidad, desde el punto de vista lexicográfico, *mentir* es *no decir la verdad*; la falsedad objetiva del discurso o la intención del

[8] La semántica estructural no prácticó descomposiciones como ésta. Los criterios analíticos y el ejemplo *bachelor* en concreto proceden de la primera semántica generativa (vid. [Katz y Fodor, 1963]), cuya división en trazos de significado mínimos del tipo *markers* y *distinguishers* fue criticada por el propio Chomsky.

hablante forman parte de las condiciones objetivas del mundo o del código moral de una sociedad, no de su lengua. En todo caso, Coleman y Kay estarían demostrando justamente lo contrario de lo que pretendían: que es factible un análisis discreto del significado.

Los prototipos se empobrecen por atender sólo aspectos léxicos y de modo bastante parcial[9]. Si se tratara de una verdadera teoría sobre la categorización del significado, incluiría bajo su alcance unidades superiores. Sólo con una ampliación del modelo que modifique su impronta cognitiva podrá dar cuenta de los fenómenos gramaticales.

La Teoría de conjuntos borrosos, por su parte, despertó mayor interés durante los años '60 y '70, cuando se estaban desarrollando sus fundamentos lógicos. El estudio de Lakoff sobre *cercas semánticas* [1972], el operador α y la noción de nominalidad de Ross [1973] o la debatida cuestión de la gramaticalidad convierten a la Gramática Generativo-Transformacional (GGT) en una buena receptora del modelo. Sorprendentemente, a medida que las aplicaciones borrosas ganan terreno en tecnología o en otras disciplinas teóricas, la Lingüística arrincona lo que se prometía como un fuerte cambio metodológico.

La explicación de este repentino rechazo hay que buscarla, en mi opinión, en dos factores conectados. En primer lugar, la detección de la vaguedad en sí misma no bastaba porque, en sentido estricto, no era un descubrimiento; la Semántica siempre la había presenciado y, tal vez por ello, su fundamento científico había sido bastante más frágil que el de otras ramas de la Lingüística. Era preciso, además, un modelo teórico en donde cupieran los estados de cosas vagos. Ese modelo no existía y, sin él, las aplicaciones borrosas al lenguaje resultaban completamente ajenas a la disciplina, porque, para aceptar un método ajeno importa más la adecuación que el hecho de participar de un objeto común.

La vaguedad en sí misma, por muy habitual que sea en el lenguaje, podría carecer de toda repercusión sobre su sistemática. *Alto*, *calvo* o *de mediana edad*

[9] La clase *bird*, en su versión original en inglés [Lakoff, 1972] se convirtió en un ejemplo de referencia obligada para estos temas. Los defensores de esta perspectiva suelen traducir precipitadamente *bird* como *pájaro*. En español existen dos términos que se ajustan a la designación de *bird*: *ave*, la clase reconocida por la Biología y *pájaro*, procedente de una taxonomía popular. Así que, cuando se sitúa el pingüino en la periferia de los pájaros se obtiene un resultado extraño en español: un pingüino no es en realidad un pájaro porque no vuela, mientras que un murciélago lo es plenamente por su capacidad de volar, aunque se trate, científicamente, de un mamífero.

–como otros ejemplos estudiados por Zadeh– poseen una gradualidad inherente que los diferencia de los predicados *de-sí-o-no*, pero eso no afecta a su comportamiento gramatical. Son, del mismo modo, modificadores de sustantivos y mantienen con ellos relaciones que se expresan en la concordancia. En la gramática del español *alto* se comporta exactamente del mismo modo que *compostelano*, *cazador* o *musulmán*. Y está por demostrar que la gradualidad sea universal; la observamos, v. gr., en español o en inglés pero es plausible que existan lenguas donde *altura*, *calvicie* o *edad* se conciban como parámetros discretos. La estructura lingüística puede transparentar la gradualidad cognitiva pero no tiene necesariamente que impregnarse de ella.

En cambio, los conjuntos borrosos hubieran podido sistematizar la gradualidad del metalenguaje. Los desarrollos en esta dirección se complicaron con cuestiones propias de la Sociología de la Ciencia. Su introducción en Lingüística de la mano de la GGT no propiciaría el asentamiento del cambio, ya que este marco no era el más proclive a la aceptación de una vaguedad epistémica. En primer lugar, en GGT no cabe hablar de grado de aplicación de las reglas o de pertenencia a una categoría gramatical. Las demostraciones de Lakoff [1972] o Ross [1972, 1973] superaban con creces las posibilidades virtuales de la teoría pues su pretensión, al fin, era bloquear las derivaciones y limitar unas reglas que se habían multiplicado innecesariamente; algo que podían conseguir dentro del propio marco.

La ortodoxia generativa dedicó a los conjuntos borrosos críticas contundentes aunque no del todo fundamentadas: atentaban contra el mentalismo y se ocupaban sólo de la *«actuación»*, así que sus productos no eran lingüísticos. En realidad, ni la lógica ni la gramática borrosas entran en el debate racionalismo/empirismo; se podría creer en categorías innatas pero de configuración borrosa. La adscripción de sus resultados a la actuación es, todavía, más peligrosa. El modelo chomskyano plantea una dicotomía tajante entre el conjunto de reglas abstractas que rige la habilidad de los hablantes, su *competencia*, y el uso que hacen de ellas a través de una lengua y en una situación determinada, la *actuación*. En el primer caso, se alude a la facultad lingüística previa a su materialización en una lengua concreta, una dimensión universal que queda convertida en el único objetivo de la Lingüística, mientras la actuación se plantea como parámetro inferior. La distinción se relaciona con dos métodos de trabajo. Los datos que interesan al estudio de la competencia proceden de la introspección del lingüista, que busca en su conocimiento de hablante nativo el modo de hacer aflorar esas reglas. En una línea diferente, el estudio de la actuación supondría analizar un gran elenco de datos reales. Sin intención de insistir en una u otra vía, lo cierto es que una gramática borrosa podría

basarse en las intuiciones de los hablantes nativos y no en un corpus masivo, lo que le permitiría salir ilesa de tales objeciones.

Mientras los dos modelos que hemos contemplado se debaten con las dificultades de incorporar a la Lingüística perspectivas nacidas en otros ámbitos, desde el seno de la propia disciplina se desafiaba también la categorización clásica. A mediados de los años '60, los lingüistas del Círculo de Praga forjan una oposición teórica, similar a la que gobernaba los prototipos, entre *centro* y *periferia*.

La periferia surge en los estudios fonológicos para dar cabida a determinados fonemas que, por la inestabilidad de sus relaciones, no aparecen integrados en el conjunto del sistema. Con las adaptaciones pertinentes, el concepto se exporta a otros ámbitos del estudio gramatical, para oponer el núcleo de las descripciones –más o menos sistemático– a casos de irregularidad, falta de productividad de los mecanismos gramaticales y, en general a las variables excepcionales o eventuales. En esencia, se asume que toda organización simétrica o uniforme del lenguaje es falsa; frente a fenómenos bien caracterizados o de comportamiento regular, otros se desvían de esa sistematicidad. Las periferias surgen en cuanto se haya asentado un criterio de estructuración y los lingüistas de Praga evitan tomar decisiones que no haya tomado el lenguaje mismo[10].

La ductilidad del planteamiento proporciona un proyecto global apto para la teoría lingüística y para la manipulación de datos idiomáticos porque el par centro/periferia, instalado por primera vez en el plano de la forma, es un universal de las lenguas. Pero, tal y como se perfila, parece una mera concesión a la continuidad en medio de un entramado discreto. Si los dos polos no se acompañan de medidas que calibren su gradualidad, la diversidad de comportamientos que pueden acoger en su interior los aboca a una proliferación de estadios intermedios[11].

---

[10] A pesar de que nunca aluden a los trabajos contemporáneos de Zadeh, también los lingüistas de Praga pretenden formalizar sus resultados, según se desprende de las declaraciones de Leska [1966: 54]:

> L'analyse effectuée peut fournir une base pour l'introduction d'une mesure adéquate de ces réalités, mesure qui permettrait de classer les différentes caractéristiques (d'après la caractéristique numérique il serait facile de différencier ce qui se rapproche du *centre* et ce qui s'éloigne du *centre*), en rendant plus précis le tableau d'ensemble et en permettant également la comparaison entre différentes langues.

[11] La oposición Centro/Periferia fue, así, absorbida por la lingüística discreta como un mecanismo cómodo para explicar ciertas discordancias de los sistemas. Sin embargo, el interés del Círculo de Praga por graduar la pertenencia de los conceptos a uno u otro ámbito no fue siempre mantenido con rigor por sus sucesores.

En todo caso, la principal aportación de las visiones difusas a la Lingüística consistió en mitigar las fronteras categoriales. Desde esta base común, sus herramientas fueron confundidas y, al final, completamente asimiladas. Como los prototipos organizaban las categorías por la similaridad de cada miembro respecto a un ejemplar estereotipado, la cuestión de la pertenencia no exigía una respuesta precisa. En la versión estándar de la teoría, cuanto más próximo esté un ejemplar al prototipo, mayor será su grado de pertenencia, es decir, un gorrión pertenecería al conjunto de las aves en mayor medida que un pingüino y eso implica, de algún modo, que es más ave que un pingüino. La prototipicidad se estaba confundiendo con la vaguedad categorial: los pingüinos no serán aves prototípicas pero por eso no lo son menos, pues no existen, de hecho, aves al 30% o al 70%. La cuestión de la calidad de miembro en una categoría y su situación dentro de ella no pueden identificarse.

El lenguaje ofrece dos tipos de vaguedad[12]. El primero surge cuando las unidades lingüísticas remiten a clases para las que la condición de pertenencia es gradual, como los adjetivos *calvo* o *alto*. El segundo tipo procede de la vigencia de varios factores simultáneos para definir la pertenencia, al estilo de lo que ocurre en *ave*. En ambos, las clases son conjuntos borrosos, cuyos miembros se calibran en el intervalo [0,1] o con etiquetas cualitativas que señalen su orden en la escala. Una vez realizada la asignación, aparecen efectos prototípicos: un hombre de 1.85 es mejor ejemplar del conjunto de hombres altos que uno que mida 1.74 y las aves voladoras son más representativas que pingüinos o pollos de corral. El asunto se cierne en torno a la noción psicológica de **representatividad**, porque cualquier organización de individuos en un conjunto produce ejemplares mejores y otros discutibles.

Sin embargo, también se siguen efectos prototípicos de entidades precisas. Lakoff [1986] maneja el prototipo de *madre* que, ya sea biológica o adoptiva, remite en su análisis a un ama de casa. No se puede hablar, sin embargo, de gradualidad inherente al concepto –o se es madre o no se es, sin que quepan estadios intermedios– y ni siquiera confluyen varios criterios en cada tipo de los posibles. Podemos trazar, entonces, una diferencia fundamental entre las dos aproximaciones:

---

[12] Neustupny [1966: 43] señala ya dos tipos de vaguedad, aunque no coinciden totalmente con los que aquí se sugieren. Para este autor la vaguedad del discurso -»cases of a single symbol (word) describing a series of similar objects»- se opone a una vaguedad sistémica entendida como «problem of classing a definite (linguistic) unit of lower order with definite more general (linguistic) units of a higher order».

los prototipos constituyen una visión intuitiva de la realidad y su conformación; los conjuntos borrosos proporcionan un tratamiento formal de la vaguedad.

6. CARACTERIZACIÓN DEL MODELO DIFUSO EN LINGÜÍSTICA

Una de las corrientes más fructíferas de la Lingüística actual ha reavivado el viejo asunto de los universales. Las investigaciones en esta línea agrupan las lenguas del mundo en estructuras tipológicas, un estilo de filiación que relaciona las soluciones dadas a problemas gramaticales concretos.

También la GGT se había ocupado de los universales, aunque con puntos de partida bien diferentes. En la versión chomskyana algunos principios abstractos comunes a todos los seres humanos rigen la facultad lingüística, como las demás capacidades cognitivas, e intervienen en la adquisición de una lengua concreta. Por su carácter mentalista, estos universales del lenguaje podrían apresarse desde cualquier lengua que, al ser humana, participaría necesariamente de ellos. En la perspectiva tipológica, en cambio, se trata de universales de las lenguas o variables estadísticamente mayoritarias que éstas comparten, a pesar de sus divergencias genealógicas o del alejamiento geográfico y cultural de las comunidades en que se desenvuelven. El acopio de gran cantidad de datos servirá para corroborar experimentalmente las hipótesis formuladas y, por vía inductiva, permitirá inferir nuevos parámetros teóricos.

Como es obvio, para establecer conclusiones de gran alcance, que engloben la mayoría de las lenguas, los conceptos gramaticales han de definirse con criterios más amplios. En principio, convenía limar los efectos de inconmensurabilidad entre los paradigmas lingüísticos pero, sobre todo, el modelo debería liberarse de descripciones demasiado apegadas a una lengua concreta. En francés, por ejemplo, el Sujeto se define por su posición relativa en la cláusula y la concordancia en número y persona con el predicado, mientras en español no requiere tales características porque el verbo lleva marcas que delatan su presencia implícita y, en latín o alemán, la concordancia se suple con el uso del caso nominativo. Para eliminar esta variabilidad, Keenan [1976], en un trabajo de gran importancia en la historia de la Tipología, formula criterios con que detectar sujetos en entornos semánticos básicos, al margen de que se materialicen con una marca de caso, una posición, un comportamiento transformacional o una propiedad semántica específica. Como resultado, algunos sujetos cumplen en mayor medida los requisitos porque la subjetividad es un problema de grado [1976: 91]. Aunque no proceda aquí detallar

cuáles fueron sus criterios, sí cabe insistir en que ninguno de ellos era exclusivo o imprescindible, toda vez que *ser sujeto* había pasado a ser una cuestión relativa.

En contra del Estructuralismo, la perspectiva tipológica acoge rasgos sustanciales como base para sus descripciones. Una línea denominada *natural* alude a la versatilidad de las lenguas para codificar matices de la realidad. Como ejemplo podría aducirse la oposición animado/inanimado o la alta frecuencia de primera y segunda personas en el discurso, debida a la preferencia que los seres humanos mostramos por hablar de nosotros mismos. Así, la naturaleza escalar de ciertos ámbitos, como el tiempo, presiona a las lenguas para distinguir grados que cubren desde el pasado remoto hasta el futuro improbable y lejano. Una vez asentadas esas diferencias, el sistema verbal se convierte en vehículo idóneo para expresar otras nociones ligadas de algún modo al tiempo; se ha producido una traslación metafórica[13]. Efectos icónicos de este tipo se manifiestan, a veces, con mecanismos gramaticales mixtos. En la tradición discreta un acto de habla se catalogaba como declarativo cuando enunciaba un mensaje, como imperativo si añadía a la información un matiz de mandato y, como interrogativo, si el mensaje era, en sí mismo, una petición de información. Sin embargo, cada categoría parece ser un estadio dentro de un continuum gradual de este tipo:

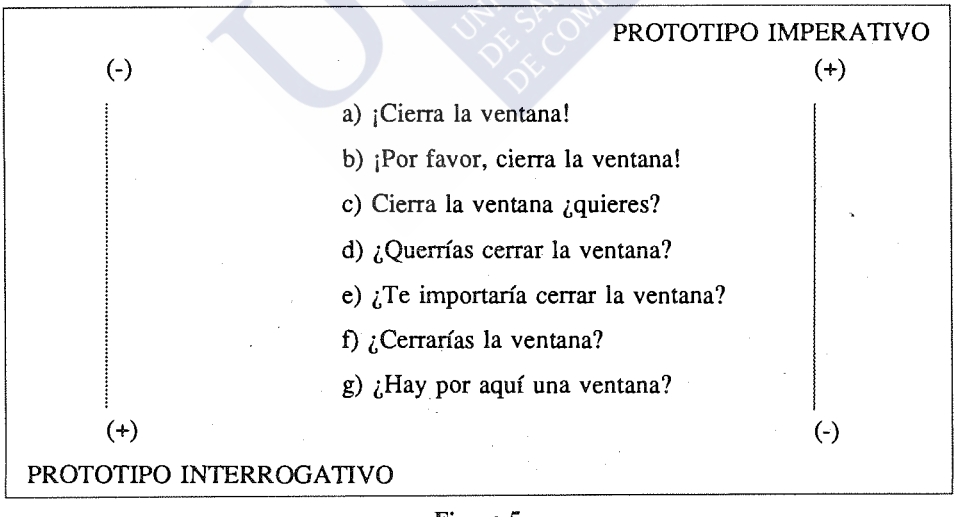


Figura 5

[13] Para la traslación metafórica de los contenidos temporales vid. Silva-Corvalán [1985] o Fleischmann [1989]. Los efectos icónicos de la modalidad (infra) se han tomado de Givón [1986].

Frente a los extremos de la serie, que corresponden a actos de habla prototípicos, los puntos intermedios, como (c), demuestran su condición vacilante en la connivencia de dos modos de expresión bien caracterizados: forma verbal en imperativo y entonación interrogativa. La serie ilustra también la gradualidad de un concepto metalingüístico. Los usos convencionales indican que muchas preguntas no son más que órdenes mitigadas por la cortesía, de forma que (g) podría interpretarse como una construcción eufemística donde el emisor sólo estuviera expresando el deseo de cerrar la ventana con una formulación que atenúa toda carga imperativa.

De este modo, la Tipología revaloriza el concepto de prototipicidad, hasta ahora limitado a teorías que trabajaban con estructuras ideales. Al trasladarse al terreno gramatical, los estereotipos culturales dejan de funcionar: se puede buscar una huella cognitiva para representar el prototipo de *vehículo*, *ave* o *madre* pero no para *complemento directo*, *modalidad interrogativa* o *adverbio de lugar*. Así que la similaridad ha de articularse sobre otros factores, fundamentalmente frecuencia e **integración discursiva**.

La frecuencia supone que el prototipo de una noción gramatical ha de representarse con su exponente más habitual en una lengua, siempre que sea suficientemente neutro, es decir, se halle desprovisto de otros valores. Así en español el complemento directo (CDIR) prototípico carecerá de preposición pese a que existan CDIRs con marca*a*.

La integración en el discurso viene a demostrar que las caracterizaciones gramaticales abstractas resultan parciales. Identificar el prototipo de nombre con la forma que designa un objeto tangible y el de verbo con la que expresa una acción efectiva no bastaría. Aunque esa caracterización básica tenga cierto sustento, los estudios tipológicos [Hopper y Thompson, 1984] subrayan que, en realidad, el nombre prototípico introduce un participante en el discurso; y el verbo expresa lingüísticamente la producción de un suceso. Al otro extremo de la escala, las formas que no se integran en el discurso, no toman los morfemas que corresponden a su clase morfológica. Por eso los segmentos de un nombre compuesto no reciben ambos la marca de plural que poseerían de ser independientes –decimos *bocacalles* y no **bocascalles*– y los predicados nominales aparecen sin determinación en muchas lenguas –*Juan es (*un) profesor*–. Paralelamente, los verbos reducen las oposiciones temporales y aspectuales para expresar irrealidad –*Ojalá estuviera hoy/mañana en la playa*– y pierden completamente sus marcas cuando se nominalizan -*La fabricación de*

metal[14]. En ninguno de estos casos se trata de nombres o verbos típicos puesto que los primeros no introducen una entidad y los segundos actualizan sucesos carentes de producción efectiva.

Con principios similares, otros conceptos renuevan sus formulaciones, casi siempre por intersección de varios parámetros definitorios. Es éste el caso de las estructuras transitivas [Hopper y Thompson, 1980; Givón, 1986] o ergativas [Comrie, 1978], del tópico [Li y Thompson, 1976] o el Objeto [Plank, 1984], por citar los más representativos. La ausencia de fronteras seguras origina conceptos de cuño borroso como la *categoría imprecisa* propugnada por Dahl [1985][15]. Sin embargo, rara vez estos trabajos estipulan los criterios metodológicos con que trazan sus propuestas. Se diría que es tan urgente la necesidad de apelar a la continuidad, que no hay tiempo para entrar a debatirla: su presencia se impone con el uso y, suponiendo una rápida aceptación, se omite cualquier referencia teórica al respecto. Las aproximaciones se refieren, mezclándolos, a procedimientos surgidos en la teoría de prototipos y a técnicas de análisis llegadas de los conjuntos borrosos.

Los códigos son tanto más precisos cuanto mayor sea la biunivocidad entre señal y mensaje –grosso modo, entre significante y significado– pero la adjudicación de una señal inequívoca a cada variación del contenido resultaría antieconómica. Ello explica ciertas deficiencias del lenguaje natural, como generalidad o ambigüedad que, por otra parte, pueden solventarse siempre añadiendo más información. La vaguedad, en cambio, es un tipo de imprecisión resistente al análisis porque procede de los propios estados de cosas, no de nuestra dificultad para categorizarlos.

Como ya se ha señalado, la vaguedad afecta al lenguaje y su estudio podría generar proyectos en Lingüística Aplicada o productos informáticos. Pero, además, la Lingüística debe afrontar la vaguedad epistémica o del metalenguaje que cubre

---

[14] Tras la nominalización, el sustantivo deverbal mantiene la propiedad de regir un argumento pero con marcas propias de su nueva categoría, como son la preposición *de* en español y la preposición *of* o el genitivo sajón en inglés.

[15] By an imprecise category I mean a category which cannot be defined in such a way that for every member *x* of its domain [...] the definition determines a truth-value to the statement that *x* belongs to the category in question. Instead, even if some members of the domain fall under the category and some clearly do not, there is an imprecise category, a zone in the middle where memberships is not clearly defined. [Dahl, 1985: 3]

su aparato conceptual. En este nivel, la aplicación de un modelo difuso enlaza la disciplina con otras facetas del conocimiento.

La vaguedad metalingüística surge allí donde las clases se caractericen regularmente por un buen número de rasgos simultáneos y no por uno solo. Este primer requisito, la *complejidad* de las entidades vagas, selecciona el ámbito de acción del modelo, puesto que cuando la adscripción a una clase es automática, los resultados coinciden con tratamientos discretos. En otras palabras, el esfuerzo teórico que esta innovación reclama no se justificaría si todas las entidades del universo en consideración se definiesen con un solo rasgo. Pero pueden existir clases complejas nítidamente separadas entre sí, sin elementos marginales entre ellas. La vaguedad precisaría, pues, una disposición asimétrica de esos rasgos definitorios, de forma que los atributos de una entidad no coincidan completamente con los de la clase que le corresponde, sino que remitan a otras clases[16].

De este modo, un concepto gramatical como *Complemento directo* (CDIR) se define con rasgos del tipo [+objeto], [+controlado], [+individualizado]. En la estructura *llevamos un paquete al hospital*, *un paquete* responde a esa caracterización y, por tanto, se asigna inequívocamente a la clase de los CDIRs. Pero, en *llevamos a una persona al hospital*, algo ha cambiado: aunque *a una persona* continúa siendo un constituyente individualizado, no se trata ya de un objeto y el control que el agente puede ejercer sobre él es relativo. El español codifica esas alteraciones con la presencia obligatoria de la preposición *a*. La duda radica habida cuenta de que la misma apariencia corresponde a una clase distinta, la del Complemento Indirecto en *llevar algo* ***a alguien***. Es posible que la preposición sólo marque el cambio de animación del referente pero, incluso en ese caso, parece mediar una diferencia de grado. Por otra parte, el estudio de los clíticos que actúan como marcas de esos complementos apunta que, al margen de la asignación que les corresponda en la teoría, los hablantes distinguen entre *lo llevé al hospital* y *le llevé al hospital*. El uso de *lo* implica un complemento altamente controlado por el agente, bien porque se trate de un objeto, bien porque el discurso lo actualice como tal. Así, un efecto icónico permitiría que en determinadas circunstancias –para ser coherentes con el ejemplo, a causa de una enfermedad-, un individuo humano se codificase como objeto–. *Le*, en cambio, se usaría en usos de cortesía *–le llevé al hospital a usted–* o cuando la entidad no se sujetara a control alguno *Le llevé al hospital en mi coche*, no estaba enfermo.

[16] Vid. fig. 2.

La asociación de Complejidad y Asimetría, en casos como éste, puede subsanarse con un tratamiento jerarquizado de los rasgos, es decir, con un relato de prioridades que los organice como fragmentos de un universo más amplio. Al establecer niveles de análisis, se sortean imprecisiones que suelen asociarse a la vaguedad y se considera la importancia relativa de los rasgos. La ordenación responde a una estructuración prototípica, según los aludidos criterios de frecuencia e integración en el discurso. En el trayecto, las propiedades observan una escalaridad progresiva, una gradación susceptible de manifestaciones diversas. Los grados de pertenencia a cada clase pueden ilustrarse con distintos tipos de CDIR asociados al verbo *PONER*.

- En *Puse un libro sobre la mesa, un libro* respondería al prototipo de CDIR ya que se trata de un objeto tangible, individualizado en el discurso con el determinante y que el sujeto animado, yo, puede controlar totalmente.

- En *Puse atención al asunto*, el candidato a CDIR, *atención*, carece de referencia y, por tanto, no actualiza una entidad en el discurso. Ello explica un comportamiento sintáctico diferente del prototipo, como la incompatibilidad con los determinantes (* *Puso la atención*). Al no verse suficientemente individualizado, ese CDIR se incorpora al predicado para constituir un compuesto verbal sinónimo, no ya de *PONER*, sino de *ATENDER*.

- En *Las gallinas ponen* se diría que un CDIR -presuntamente *huevos*- se ha omitido por consabido. El conocimiento del mundo penetra en la estructuración lingüística porque el receptor es capaz de recuperar la información relativa al CDIR, algo que no ocurría en los ejemplos anteriores (**Puse al asunto*, * *Puse sobre la mesa*).

- La alternancia *Puse al niño/el niño en la cuna* muestra dos interpretaciones distintas que se asocian a un mismo esquema sintáctico y a idénticas formas léxicas. Parecería que el hablante maneja recursos comunicativos para actualizar un mismo objeto como más o menos controlado, según sus propósitos y la perspectiva que adopte ante el mensaje. La resistencia del CDIR animado al control del agente se acusa en traslaciones metafóricas del tipo *Poner a alguien en tal situación*.

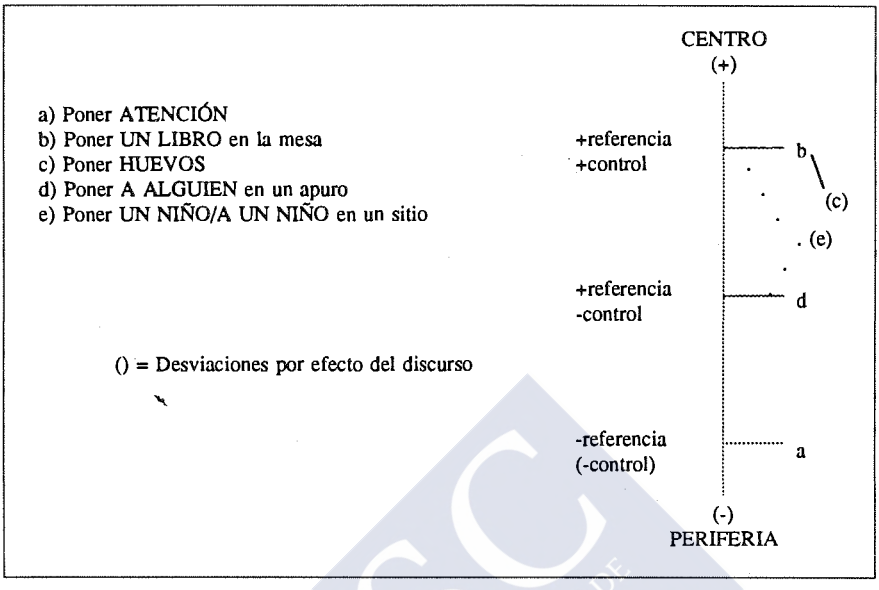


Figura 6

a) Objeto desviado del prototipo por incorporación al predicado: *Poner atención.*

b) Objeto prototípico: *Poner algo en un sitio.*

c) Omisión de un objeto similar a (b) y consiguiente devaluación de su prototipicidad: *Poner (huevos).*

d) Objeto desviado del prototipo por la incapacidad del agente para controlarlo: *Poner a alguien en tal situación.*

e) Objeto de condición vacilante, entre (b) y (d), por efecto del discurso: *Poner un niño / a un niño en un sitio.*

De este modo, se considera una gama, amplia pero finita, de variantes que, aun alejándose progresivamente del prototipo de CDIR, pertenecen a la categoría en alguna medida. La estratificación no precisa moldes fijos pero, evidentemente, existen distintos grados de pertenencia que podrían calibrarse con herramientas borrosas. Sólo así se evita caer en el extremo opuesto y dejarnos deslizar en el

atractivo tobogán de la continuidad. Que los conceptos lingüísticos sean vagos no significa la ausencia de extremos para la trayectoria en que se proyectan: al no haber tenido en cuenta las transiciones, la visión de la Lingüística tradicional fue sesgada, probablemente incompleta pero no errada de raíz. Hace tiempo que en Lingüística se viene gestando este cambio metodológico. Etiquetas como *indeterminación*, *potencialidad*, *matiz* en vez de oposición, *caso-límite*, *combinación*, *vacilación*, *gradación* o *continuidad* se han asentado ya en la bibliografía y no conviene dejar que proliferen sin extremar los cuidados teóricos.

7. CONCLUSIÓN

La vaguedad de los conceptos lingüísticos resulta especialmente persistente porque añade a su imprecisión intrínseca la pluralidad de soluciones que documentan las lenguas humanas. Para obtener rendimiento de ella es necesario contar con herramientas depuradas que permitan reorganizar el entramado teórico. La Tipología lingüística parece el marco apropiado para ampliar la perspectiva tradicional, si bien no basta con adoptar la versión ingenua de los prototipos cognitivos. La solución pasa por combinar la Teoría de Conjuntos Borrosos con una articulación prototípica, un modelo ya existente en Lógica.

La confección de una Lingüística no-discreta repercutiría, por su parte, sobre el propio modelo difuso. La Teoría de Conjuntos Borrosos ha centrado sus logros en aplicaciones experimentales, que cubren un espectro variado de campos. Por su dimensión empírica, la Lingüística no permanece al margen de este interés y el propio Zadeh se ha ocupado de ciertas facetas del lenguaje con vistas a su gestión automática: la Inteligencia Artificial, como las demás Industrias de la lengua, precisan conceptos permeables que puedan resolver sus aplicaciones. Pero aunque la Tecnología nos deslumbre, la dimensión teórica de la Ciencia, previa a cualquier utilidad, ocupa un lugar privilegiado en el pensamiento humano y, en esta vía, prima la adecuación. La Lingüística permitiría medir los alcances de esta nueva perspectiva a salvo de la premura que la tecnología suele imponer. Ni siquiera la naturaleza formalizada del modelo, operaría en su contra, de acuerdo con las indicaciones de Bunge[1967: 134]:

> El progreso de la ciencia ha dependido en gran medida de la capacidad de aprovechar conceptos potentes y desprovistos de contenido empírico, especialmente los que se encuentran ya listos para el uso en la lógica y en la matemática. [...] la teoría de conjuntos, la teoría de retículos y otras teorías

matemáticas cualitativas (no-numéricas), suministran el esqueleto formal propio de la sistemática [...].

Aunque las hipótesis nacen en un marco disciplinar se expanden luego fuera de su contexto originario siempre que resulten fecundas y puedan ser asimiladas por otras teorías. Por ello, no es extraño que el modelo borroso pueda servir como Programa de Investigación en Lingüística. La polémica está servida y es de suponer que no se implante sin debate. En todo caso, los criterios para decidir la adecuación de un modelo deben ser metodológicos: coherencia, simplicidad, economía y capacidad de proyección sobre los datos. No es necesario asumir que sea ésta la única y mejor teoría para tratar los hechos de lengua; basta con que su cohesión interna invite a apostar por el cambio. Después, como cualquier otra teoría, habrá de sujetarse a procesos de verificación experimental y a las consecuentes modificaciones usuales en la investigación científica.

El objetivo del cambio metodológico es, al fin, bien simple. Los principios de actuación borrosa deben fomentar modelos más coherentes, que armonicen con nuestras predicciones teóricas y se adapten, fielmente, a la complejidad de las lenguas. En mi opinión, la investigación lingüística se enriquecería fomentando una óptica continua de sus primitivos, hecha de clases borrosas y proyecciones graduales. Es probable que acogiendo en su seno la vaguedad obtenga conceptos fecundos, que le permitan, no ya dar la respuesta correcta, sino algo de más valor: formular la pregunta adecuada.

REFERENCIAS

Alonso, A. y Henríquez Ureña, P., *Gramática castellana*, Buenos Aires, Losada, 1971, 39-40.

Ballmer, T.T., «Fuzzy Grammatical Categories» en Ballmer, T.T. y Pinkal, M., (Eds.), *Approaching Vagueness*, Elsevier Science Pub., (1983).

Berlin, B. y Kay, P., *Basic color terms. Their universality and evolution*, Univ. California Press, 1969.

Black, M., *Language and Philosophy*, New York, Ithaca, 1949.

Bosque, I., «Perspectivas de una lingüística no discreta» en Abad, F. y col., (Eds.), *Metodología y gramática generativa*, Madrid, SGEL, (1979), 81-111.

Bunge, M., *Scientific Research,* New York, Springer-Verlag, 1967. Trad. esp. de Sacristán, M., *La investigación científica. Su estrategia y su filosofía*, Barcelona, Ariel, 1983².

Coleman, L. y Kay, P., «Prototype Semantics: the English verb 'lie'», *Language* 57/19, (1981), 26-45.

Comrie, B., «Ergativity» en Lehmann, W.P. (Ed.), *Syntactic typology: studies in the phenomenology of language*, Univ. Texas Press, (1978), 329-394.

Coseriu, E., «Semántica estructural y semántica cognitiva» en *Hom. al profesor F. Marsá. Jornadas de Filología*, Univ. de Barcelona, (1990), 239-282.

Craig, C. G., ed., «Noun classes and categorization», *Proc. of a Symposium on Categorization and Noun classification*, Eugene, Oregon, October 1983. *Typological studies in language, 7*, Amsterdam/Philadelphia, J. Benjamins Pub. Co., 1986.

Croft, W.A., *Categories and relations in Syntax: the clause-level organization of information*, UMI, 1986.

Dahl, Ö., *Tense and Aspect systems*, New York, Blackwell, 1985.

Danes, F., «The Relation of Centre and Periphery as a Language Universal», *TLP*, N. 2, (1966), 9-21.

Deaño, A., *Introducción a la lógica formal*, Madrid, Alianza Universidad, 1974.

Fernández Guizzeti, G., «La teoría atomista contextual de la gramaticalidad y las lógicas no standard», *RSEL*, N. 2, (1972), 313-329.

Fernández Pérez, M., *La investigación lingüística desde la filosofía de la ciencia. (A propósito de la lingüística chomskyana)*, Univ. de Santiago de Compostela, anexo 28 de *Verba*, 1986.

Fillmore, C., «An alternative to checklist theories of meaning» en *Proc. of the first annual meeting of the Berkeley Linguistics Society*, (1975).

Fillmore, C., «Topics in Lexical Semantics» en Cole, R.W. (Ed.), *Current Issues in Linguistic Theory*, Indiana Univ. Press, (1977), 76-138.

Fleischman, S., «Temporal distance: a basic linguistic metaphor», *Studies in language*, Vol. 13, N. 1, (1989), 150.

Garrido Medina, J., *Lógica y Lingüística*, Madrid, Síntesis, 1988.

Gentilhomme, Y., «Les ensembles flous en linguistique», *Cahiers de linguistique théorique et appliquée*, N. 5, (1968), 47-63.

Givón, T., *Syntax: a functional-typological introduction*, Amsterdam-Philadelphia, J. Benjamins Pub. Co., 1984.

Givón, T., «Prototypes: between Plato and Wittgenstein» en Craig, C.G. (Ed.), (1986), 77-102.

Goguen, J.A., «The Logic of Inexact Concepts», *Synthese*, N. 19, (1968-9), 325-373.

Hjelmslev, L., *Omkring sprogteoriens grundlaeggelse*, Festskrift udg. af Kobenhavns Univ., Copenhague, 1943. Trad. esp. de Díaz de Liaño, J.L., *Prolegómenos a una teoría del lenguaje*, Madrid, Gredos, 1974.

Hockett, C.F., «Linguistic elements and their relations», *Language*, Vol. 37, (1961), 29-53.

Hopper, P.J. y Thompson, S.A., «Transitivity in Grammar and Discourse», *Language*, Vol. 56, (1980), 251-299.

Hopper, P.J. y Thompson, S.A., «The discourse basis of lexical categories», *Language*, Vol. 60, N. 4, (1984), 703-752.

Itkonen, E., *Grammatical theory and Metascience. A critical investigation into the methodological and philosophical foundations of 'autonomous' linguistic*, Amsterdam, J. Benjamins B. V., 1978.

Katz, J. y Fodor, J., «The structure of a semantic theory», *Language*, Vol. 39, (1963), 170-210.

Kay, P. y McDaniel, C.K., «The linguistic significance of the meanings of basic color terms», *Language*, Vol. 54, N. 3, (1978), 610-647.

Keenan, E.L., «Towards a universal definition of *subject of*» en Li, C.N. (1976), 303-333. Reed. en Keenan, E.L., *Universal Grammar: fifteen essays*, London, Croom Helm, (1987), 89-120.

Kleiber, G., «Prototype, stéréotype: un air de famille?», *DRLAV*, Vol. 38, (1988), 1-61.

Labov, W., *Il continuo e il discreto nel linguaggio*, Bologna, Il Mulino, 1975.

Lakoff, G.A., «Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts», *Papers of the eighth regional meeting of the Chicago Linguistic Society*, (1972), 183-228. Reed. en Hockney, D. y col. (Eds.), *Contemporary Research in Philosophical Logic and Linguistic Semantics*, Dordrecht, D.-Reidel Pub. Co., (1975), 221-271.

Lakoff, G.A., «Classifiers as a reflection of mind» en Craig, C.G. (Ed.), (1986), 13-51.

Leska, O., «Le centre et la périphérie. Des différentes niveaux de la structure linguistique», *TLP*, N. 2, (1966), 53-57.

Li, C.N. y Thompson, S.A., *Subject and topic*, NewYork, Academic Press, 1976.

Martin Vide, C., «Sobre lo abstracto, lo difuso y lo confuso: hacer teoría y hacer imaginería» en *Actas del V CLNLF*, Univ. Barcelona, (1990), pág.2739.

Mendivil Giro, J.L., «Consideraciones sobre el carácter no discreto de las expresiones idiomáticas» en Martin Vide, C. (Ed.), *Actas del VI CLNLF*, Univ. de Barcelona, (1991), 711-735.

Morgan, C.G. y Pelletier, F.J., «Some notes concerning fuzzy Logics», *Linguistics and Philosophy*, Vol. 1, N. 1, (1977), 79-97.

Neustupny, J.V., «On the Analysis of Linguistic Vagueness», *TLP*, N. 2, (1966), 39-51.

Olivas, J.A. y Sobrino, A., «Aproximación lingüística y teoría de prototipos de Zadeh» en *Actas del I CETLF*, Univ. de Granada, (1991), 179-183.

Plank, F., ed., *Objects. Towards a Theory of Grammatical Relations*, London-New York, Academic Press, 1984.

Putnam, H., «The meaning of *meaning*» en Gunderson, K., comp., *Language, Mind and Knowledge, Minnesota Studies in the Philosophy of Science* VII, Univ. of Minnesota Press, (1975). Trad. esp. de Acero, J.J., «El significado de *significado*», *Teorema*, Vol. XIV, N. 3-4, 1984, 345-405.

Rosch, E., «Cognitive reference points», *Cognitive Psychology*, N. 7, (1975), 532-547

Ross, J.R., «The category squish: enstation hauptwort», *Papers of the eighth regional meeting of the Chicago Linguistic Society*, (1972), 316-328.

Ross, J.R., «Nouniness» en Fujimura, D. (Ed.), *Three Dimensions of Linguistic Theory*, Tokio, TEC, (1973), 138-257.

Sapir, E., *Language. An introduction to the study of speech*, New York, Harcourt, Brace & Co., 1921. Trad. esp. de Margit y A. Alatorre, *El lenguaje*, México, FCE, 1954.

Silva-Corvalan, C., «Modality and semantic change» en Fisiak, J. (Ed.), *Historical Semantics, historical word formation*, Berlin, Mouton de Gruyter, (1985), 547-572.

Skala, H.J., Termini, S. y Trillas, E., eds., *Aspects of Vagueness*, Dordrecht-Boston-Lancaster, D.Reidel Pub. Co., 1984.

Sobrino, A., «Is there a prototypical logic for prototypes and classifications founded on them», *Proc. of the IPMU'92*, Univ. de les Illes Balears, en prensa.

Taylor, J.R., *Linguistic categorization*, Oxford, Clarendon Press, 1989.

Wittgenstein, *Philosophical Investigations*, Oxford, Blackwell, 1953.

Yager, R.R., Ovchinnikov, S., Tong, R.M. y Nguyen, H.T., eds., *Fuzzy sets and applications: Selected Papers by L. A. Zadeh*, New York, J. Wiley & Sons, 1987.

Zadeh, L.A., «Fuzzy sets», *Information and Control*, Vol. 8, New York, Academic Press, (1965), 338-353. Reed. en Yager, R., (1987), 29-44.

Zadeh, L.A., «A Fuzzy-Set-Theoretic Interpretation of Linguistic Hedges», *Journal of Cybernetics*, Vol. 2, N. 2, (1972), 4-34. Reed. en Yager, R., (1987), 467-498.

Zadeh, L.A., «Outline of a New Approach to the Analysis of Complex Systems and Decision Processes», *IEEE Transactions Systems, Man and Cybernetics*, Vol. 3, N. 1, (1973), 28-44. Reed. en Yager, R., (1987), 105-146.

Zadeh, L.A., «A Computational Approach of Fuzzy Quantifiers in Natural Language», *Comp. and Maths with Appls.*, N. 9, (1983), 149-184. Reed. en Yager, R., (1987), 569-613.

Zadeh, L.A., «A theory of Commonsense Knowledge» en Skala, H.J. y col. (1984), 257-296. Reed. en Yager, R., (1987), 615-653.

Zinov'ev, A.A., *Filosofskie Problemy Mnogoznacnoj Logiki*, Institut Filosofii Izdatel'stvo Akademii Nauk SSSR, Moscow, 1960. Ed. y trad. ingl. de Küng, G. y Comey, D., *Philosophical Problems of ManyValued Logic*, Dordrecht, D.Reidel Pub. Co., 1963.

.....

- CETLF: *Congreso español sobre Tecnologías y Lógica fuzzy*

- CLNLF: *Congreso de lenguajes naturales y lenguajes formales*

- TLP: *Travaux Linguistiques de Prague*, 2, 1966: *«Lesproblèmes du centre et de la péripherie du sytème de la langue»*.

APLICACIÓN DE LA LÓGICA BORROSA A LA SELECCIÓN DE CARTERAS

José Antonio Redondo
Departamento de Economía Financiera y Contabilidad
Universidad de Santiago de Compostela

José Carlos de Miguel
Departamento de Econometría y Métodos Cuantitativos
Universidad de Santiago de Compostela

1. INTRODUCCIÓN

El objetivo de este trabajo es el incorporar la teoría de los conjuntos borrosos al campo financiero y, más concretamente, a la inversión financiera; campo éste último en el que no existen aplicaciones.

Dentro de este objetivo intentaremos abordar dos aspectos que consideramos importantes:

1. Incorporar la metodología de los conjuntos borrosos a las decisiones de índole financiero.

2. Evaluar la operatividad de la teoría de los conjuntos borrosos en las aplicaciones financieras.

Para alcanzar dicha meta hemos considerado oportuno dividir el trabajo en tres partes. En la primera se introducen los conceptos que consideramos esenciales

de la teoría de los conjuntos borrosos. La segunda parte se dedicará a la aplicación de la teoría a la selección de una cartera de riesgo para el caso particular de dos activos financieros bajo borrosidad. Por último, en la tercera parte se desarrolla una aplicación práctica.

2. FUNDAMENTOS TEÓRICOS

Sea E el espacio universal y A un subconjunto de E, tal que $A \subset E$. Se denomina «función característica de A» a la función que toma el valor 1 para todo elemento de A, $x \in A$; y el valor 0 para aquellos elementos x, tales que $x \notin A$.

Si representamos por $\mu_A(x)$ la función característica de A, tendremos que:

$$\mu_A(x) = 1 \quad \text{si } x \in A$$

$$\mu_A(x) = 0 \quad \text{si } x \notin A$$

Esta definición se corresponde con la definición de conjunto ordinario o nítido, en el cual se da la dicotomía entre la pertenencia estricta y la no pertenencia.

Zadeh generaliza la noción de conjunto a la de conjunto borroso en el cual la pertenencia al conjunto no es dicotómica. De esta forma, un conjunto borroso A en E se caracteriza por una función de pertenencia $\mu_A(x)$ (denominada característica en los números nítidos) que asocia, con cada elemento de E, un número real en el intervalo [0,1], tal que el valor de $\mu_A(x)$ representa el grado de pertenencia de x en A.

$$\mu_A : E \rightarrow [0,1]$$

$$x \rightarrow \mu_A(x)$$

Por tanto, dado un subconjunto borroso A, cualquier elemento de E puede pertenecer a él «un poco», «bastante», «intensamente», etc. En consecuencia, para definir correctamente un conjunto borroso, es necesario utilizar pares de valores, donde el primer término represente el elemento del conjunto considerado y el segundo, $\mu_A(x)$, la intensidad con que tal elemento pertenece al conjunto en cuestión:

$$A = \{(x, \mu_A(x))\}$$

Siguiendo a Kaufmann y Gil Aluja [1986], un número borroso puede interpretarse como una generalización de los intervalos de confianza[1]. De esta forma, un número borroso se halla formado por una secuencia finita o infinita de intervalos de confianza con las siguientes propiedades:

1. Se asocia a cada intervalo de confianza en la secuencia un valor $\alpha \in [0,1]$, de tal manera que dos intervalos de confianza diferentes no pueden tener el mismo valor α. Este valor se llama nivel de presunción.

2. Se designa por $A_\alpha = [\, a_1^{(\alpha)} , a_2^{(\alpha)}]$ el intervalo de confianza de nivel α. Se debe cumplir:

$$(\alpha' < \alpha) \Rightarrow (A_\alpha \subset A_{\alpha'}); \quad \alpha, \alpha' \in [0,1]$$

   Dicho de otro modo : los intervalos de confianza deben encajarse, estrictamente o no, los unos con los otros.

3. Existe un intervalo, y sólo uno, que puede reducirse a un real único (figura 1).

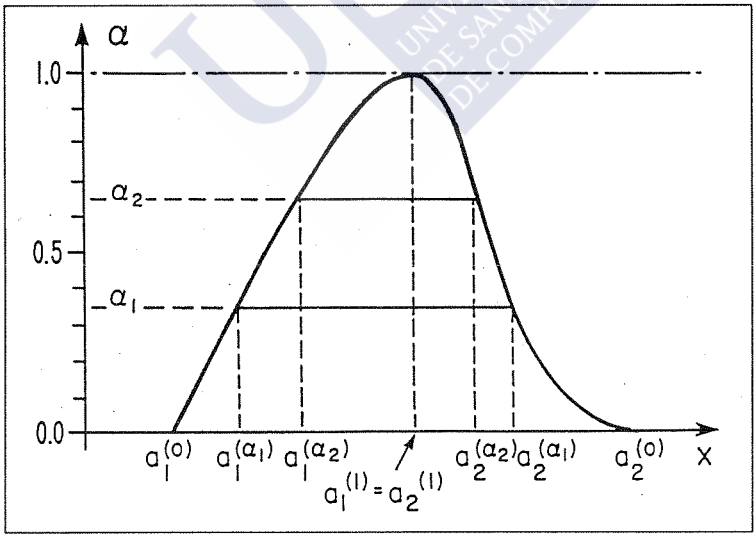


Figura 1. Definición de números borrosos.

[1] Definiciones más generales se pueden encontrar en Buckley [1987].

Existe una segunda forma de definir un número borroso a partir de los conjuntos borrosos. Supuesto el conjunto de los números reales $\mathbb{R}$, un subconjunto A de $\mathbb{R}$ puede definirse utilizando la función característica:

$$\mu_A(x) \in \{0,1\}$$

como ya se ha indicado.

Si efectuamos la modificación para que la función característica tome valores en el intervalo [0,1], lo que hacemos es modular el grado de pertenencia. Para definir un número borroso en este sentido, será necesario definir la *convexidad* y *normalidad* del subconjunto A.

1. Un subconjunto borroso $A \subset \mathbb{R}$ es convexo si y sólo si cada «intervalo de confianza de nivel α", que no tiene por qué ser un intervalo en sentido estricto, es un conjunto convexo:

$$A_\alpha = \{ x / \mu_A(x) \geq \alpha \}, \qquad \alpha \in [0,1]$$

es decir, es un intervalo cerrado de $\mathbb{R}$.

2. Un subconjunto borroso $A \subset \mathbb{R}$ es normal si:

$$\bigvee_{x \in \mathbb{R}} \mu_A(x) = 1$$

Es decir, el máximo de $\mu_A(x)$ $\forall x \in \mathbb{R}$ es 1.

Resumiendo lo expuesto, se puede definir un número borroso en $\mathbb{R}$ como un subconjunto de $\mathbb{R}$ que es convexo y normal (figuras 2, 3 y 4).

Ya hemos dicho que la función de pertenencia expresa la intensidad con que los elementos de un subconjunto borroso pertenecen a él. Sus valores se interpretan también como la compatibilidad de x con el concepto que se representa por el conjunto borroso [Kaufmann, 1977], o como la «complacencia de aceptar» que el elemento x satisface las condiciones necesarias para pertenecer al conjunto borroso [Bellmann y Giertz, 1973].

Uno de los mayores problemas que plantea la teoría de los conjuntos borrosos es el de cómo estimar las funciones de pertenencia. Es en este punto donde se han concentrado las mayores críticas sobre la teoría, fundamentalmente desde la perspectiva bayesiana [Stallings, 1977].

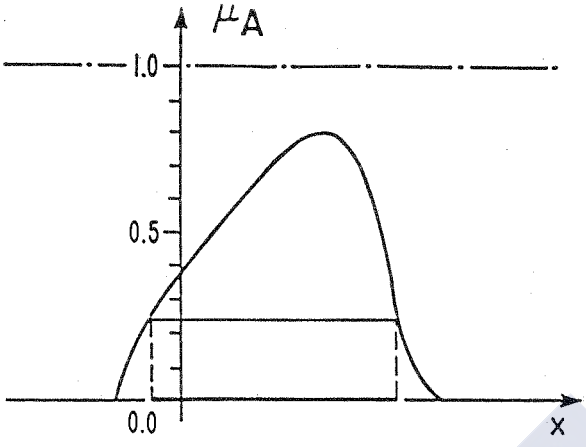


Figura 2. Conjunto borroso convexo (no normal).

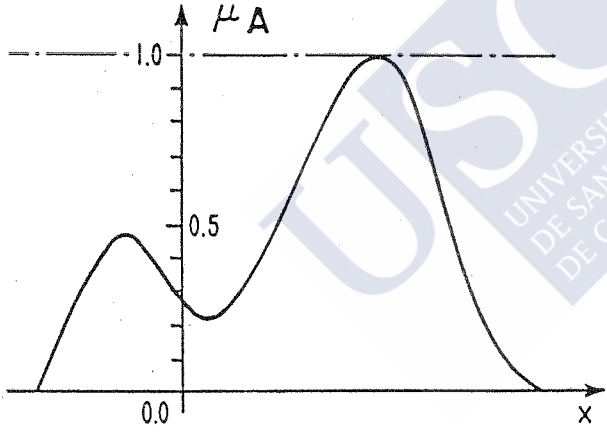


Figura 3. Conjunto borroso (normal).

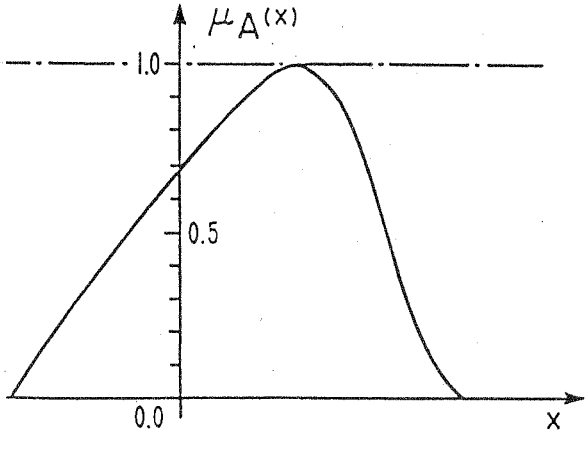


Figura 4. Conjunto borroso convexo y normal.

Es cierto, como señalan Watson y col., [1979], que «la cuestión de cómo valorar estos grados de pertenencia ha recibido sorprendentemente poca atención en la literatura»; pero no está en consonancia con el cuerpo doctrinal de los conjuntos borrosos el preocuparse demasiado por la precisión de estos números. Existe una serie de procedimientos que permiten elegir la función de pertenencia, sin embargo escaparía al objetivo de este trabajo entrar en la consideración de cada uno de ellos. No obstante, podemos señalar como más idóneos para las aplicaciones de índole financiero los desarrollados por Saaty [1978] y Dinh [1984].

Otro aspecto a considerar es el relativo a la forma de las funciones de pertenencia. Teóricamente las formas que pueden adoptar las funciones son infinitas. En la obra de Kaufmann sobre introducción a los subconjuntos borrosos, se pueden encontrar las más usuales[2] [Kauffmann, 1977]. Por su interés en nuestra investigación nos vamos a detener en la función de pertenencia triangular.

Un número borroso triangular está caracterizado por tener una función μ tal que:

$$A_\alpha = [\,(a_2-a_1) * \alpha + a_1\,,\,(a_3-a_2) * \alpha + a_3\,] \quad \alpha \in [0,1]$$

O también, utilizando la función de pertenencia:

$$\forall x \in \mathbb{R} \quad \mu_A(x) = \begin{cases} 0 & \text{si } x \le a_1 \\ \dfrac{x - a_1}{a_2 - a_1} & \text{si } a_1 \le x \le a_2 \\ \dfrac{-x + a_3}{a_3 - a_2} & \text{si } a_2 \le x \le a_3 \\ 0 & \text{si } a_3 \le x \end{cases}$$

[2] La utilización de la función de pertenencia triangular es la más usual dentro del campo de las aplicaciones económicas por su sencillez y flexibilidad en la resolución de problemas. Funciones alternativas pueden verse en Li Calzi [1990].

Un número borroso triangular puede representarse perfectamente por los tres valores asociados a_1 , a_2 y a_3. Por tanto, podrá definirse por [Kaufmann y Gupta, 1986]:

$$A = (a_1, a_2, a_3), \text{ para } a_1, a_2, a_3 \in \mathbb{R}; a_1 \leq a_2 \leq a_3.$$

3. LA SELECCIÓN DE CARTERAS BAJO BORROSIDAD

A efectos expositivos vamos a considerar uno de los casos más sencillos de delimitación de la frontera eficiente; para ello, partiremos de los siguientes supuestos:

1. Se analizan dos activos financieros de los que conocemos su rendimiento esperado y su riesgo asociado (R_j , σ_j); así como la covarianza asociada (σ_{ij}).
2. Existe una tasa de interés sin riesgo R_f para préstamo y endeudamiento.
3. Se permiten las ventas en descubierto.

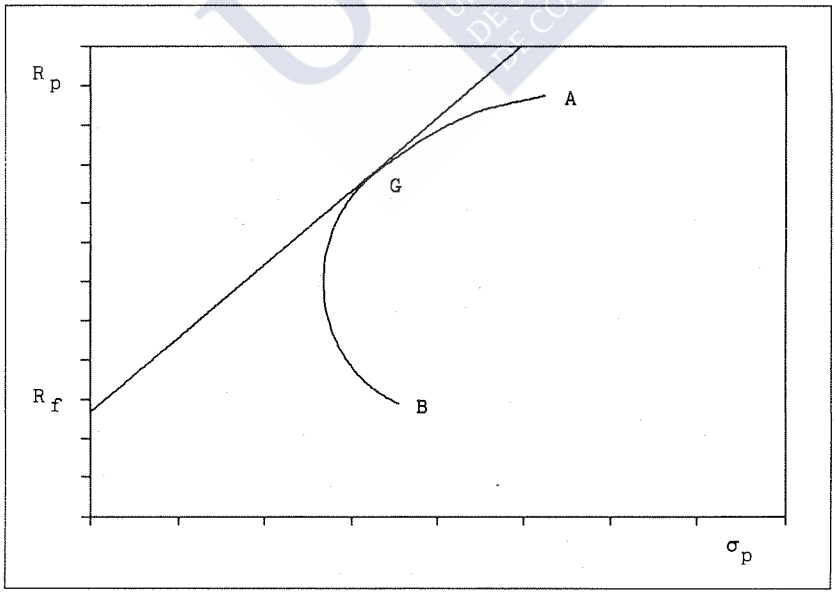


Figura 5

La derivación de la frontera eficiente bajo estos supuestos nos conduce a la selección de la combinación G (figura 5) de activos con riesgo, que se obtiene maximizando la pendiente que, partiendo de R_f, se hace tangente a la curva que representa las combinaciones eficientes de los activos con riesgo A y B [Elton y Gruber, 1984].

Para la determinación de la cartera G bastará con maximizar la función objetivo:

$$\text{Max} \left[\theta = \frac{R_p - R_f}{\sigma_p} \right], \text{ sujeta a la limitación } \sum_i x_i = 1, \text{ con } \sigma_p = \sum_a \sum_b x_a * x_b * \sigma_{ab}$$

donde:

$R_p = R_a * x_a + R_b * x_b = R_a * x_a + R_b * (1-x_a)$;

R_f es igual a la tasa de interés sin riesgo;

σ_p es la desviación típica de las carteras eficientes formadas al combinar los títulos A y B.

La solución, según se puede demostrar [Elton y Gruber, 1984], se encuentra al resolver el sistema de ecuaciones:

$$R_a - R_f = Z_a * \sigma_2 + Z_b * \sigma_{ab}$$
$$R_b - R_f = Z_a * \sigma_{ab} + Z_b * \sigma_2$$

donde Z_a y Z_b son variables intermedias que permiten calcular las proporciones a invertir en A y B a través de la relación:

$$x_a = Z_a/(Z_a + Z_b) \text{ y } x_b = Z_b/(Z_a + Z_b)$$

Consideremos ahora que el tipo de interés para préstamo y endeudamiento es un número borroso triangular al que está asociado una función de pertenencia que:

1. Es continua.
2. Está normalizada; es decir $\exists\, x \in \mathbb{R} / \mu(x) = 1$.
3. Es convexa; si $x,z \in \mathbb{R}$ e $y \in [x,z]$ se cumple que $\mu(y) \geq \min \{\mu(x), \mu(z)\}$.

Un conjunto de α - nivel K, vendrá expresado por un intervalo de confianza que denominaremos:

$$K^{\alpha} = [i^{\alpha}, s^{\alpha}]$$

Es fácil comprobar que este intervalo se puede expresar en función de α:

$$K^{\alpha} = [i^{\alpha}, s^{\alpha}] = [i+(m-i) * \alpha,\ s-(s-m) * \alpha] \text{ para } \alpha \in [0,1].$$

Por consiguiente, si sustituimos la tasa de interés sin riesgo por tasas borrosas en la determinación de la frontera eficiente, tendremos que la determinación de la cartera de riesgo consistirá en obtener:

$$\text{Max}\left[\frac{R_p - R^{\alpha}}{\sigma_p}\right], \text{ sujeto a } \sum_i x_i = 1$$

$$\text{Max}\left[\frac{R_p - [i^{\alpha}, s^{\alpha}]}{\sigma_p}\right], \text{ sujeto a } \sum_i x_i = 1$$

lo que nos va a permitir calcular los niveles de rendimiento y riesgo asociados a cada nivel α; es decir, para cada nivel α podremos determinar el intervalo de variación asociado a la cartera de riesgo, y que cada inversor deberá combinar con el activo sin riesgo de acuerdo con sus preferencias.

Partiendo del problema anterior que nos permite sensibilizar los resultados para cada α, podríamos plantear el problema de selección de la cartera de riesgo en los siguientes términos:
siendo $\alpha^k \in [0,1]$, tendremos que obtener:

$$\text{Max}\left[\frac{R_p^{\alpha} - R_f^{\alpha}}{\sigma_p}\right], \text{ sujeto a } \sum_i x_i = 1.$$

Siguiendo el modelo de Zimmermann [Zimmermann, 1984], la solución precisará resolver:

$$\text{I)} \quad \text{Max}\left[\frac{R_i + (m-i) * \alpha - (i + (m-i) * \alpha}{\sigma_p}\right], \text{ sujeto a } \sum_i x_i = 1$$

II) $\text{Max}\left[\dfrac{R_s + (m-s) * \alpha - (s + (m-s) * \alpha}{\sigma_p}\right]$, sujeto a $\sum_i x_i = 1$

Por tanto, estaremos ante dos soluciones Z_1^* y Z_2^*.

Para cada z_1^* fijamos una w_i, tal que $w_i \leq w_i^*$, para i = 1, 2, que se corresponderá con un hiperplano Γ_i, a partir del cual la función de pertenencia empezará a no ser nula. La elección de ese w_i evidentemente es arbitraria, pero puede ser justificada en determinados casos por la naturaleza del problema. Puede considerarse como el límite de sacrificio que se aceptaría por anticipado, para desviarse de la solución óptima en relación al criterio.

Las w_i pueden venir dadas, por ejemplo, a través de un porcentaje w_i / z_i, o bien $(z_i^* - w_i)/z_i^*$.

La función de pertenencia de cada z_i será entonces:

$$\begin{cases} \mu_i(\beta_i) = 0 & \text{si } \beta_i \leq w_i, \ i = 1, 2 \\ \lambda = 1 - \dfrac{z_i^* - \beta_i}{z_i^* - w_i} & \text{si } w_i \leq \beta_i \leq z_i^* \\ \lambda = 1 & \text{si } z_i^* \leq \beta_i \end{cases}$$

o bien:

$$\begin{cases} \mu(\theta_1) = 0 & \text{si } \beta_1 \leq w_1 \\ \lambda = 1 - \dfrac{z_1^* - \beta_1}{z_1^* - w_1} & \text{si } w_1 \leq \beta_1 \leq z_1^* \\ \lambda = 1 & \text{si } z_1^* \leq \beta_1 \end{cases}$$

$$\begin{cases} \mu(\theta_2) = 0 & \text{si } \beta_2 \leq w_2 \\ \lambda = 1 - \dfrac{z_2^* - \beta_2}{z_2^* - w_2} & \text{si } w_2 \leq \beta_2 \leq z_2^* \\ \lambda = 1 & \text{si } z_2^* \leq \beta_2 \end{cases}$$

O bien:

$$\mu(\theta_{i1}) = \frac{1}{z_i^* - w_i} * z_i - \frac{\beta_i}{z_i^* - w_i} \text{ , sujeto a } \sum_i x_i = 1$$

Establezcamos la hipótesis de que los criterios deben ser satisfechos según una conjunción «y» representada por el operador (min); se deberá considerar entonces

$$\mu^*(\beta_1, \beta_2, ..., \beta) = \text{Max } [\mu_1(\beta_1), \mu_2(\beta_2)].$$

Por tanto, como tratamos de buscar un máximo para el programa lineal:

$$\text{Max } \lambda$$

$$\lambda \leq \frac{1}{z_1^* - w_1} * z_i - \frac{w_1}{z_1^* - w_1}$$

$$\lambda \leq \frac{1}{z_2^* - w_2} * z_i - \frac{w_2}{z_2^* - w_2}$$

$$\sum_i x_i = 1$$

Por tanto, el proceso se articula en tres etapas:

I.-Se determinan los óptimos parciales.

II.-Se establecen las desviaciones admisibles, y las funciones de pertenencia $\mu_i(\beta_r)$ que correspondan.

III.-Se determina la solución óptima:

$$\mu^*(x_1, x_2) = \text{Max } [\mu(\beta_1), \mu(\beta_2)].$$

A modo de ejemplo, consideremos el caso de dos inversiones cuyos rendimientos esperados y riesgos asociados son, respectivamente:

	R_i	σ_2
Activo A	16	9
Activo B	10	4

Supongamos que el tipo de interés sin riesgo viene dado como un número triangular borroso de parámetros (4;6;8), y que el coeficiente de correlación entre los títulos A y B es 0,5.

Siguiendo el planteamiento general expuesto en el apartado anterior, tendremos que la cartera de riesgo se obtiene maximizando:

$$\text{Max}\left[\frac{R_p - (i_\alpha , s_\alpha)}{\sigma_p}\right], \text{ sujeto a } \sum_i x_i = 1$$

Solucionando el sistema para distintos niveles de α - cortes tendremos que, para cada α, se determinan dos valores de rendimiento y riesgo que representan la composición de la cartera de riesgo para las dos alternativas de tasa de interés sin riesgo asociado (figura 6). Los valores asociados están recogidos en la tabla I.

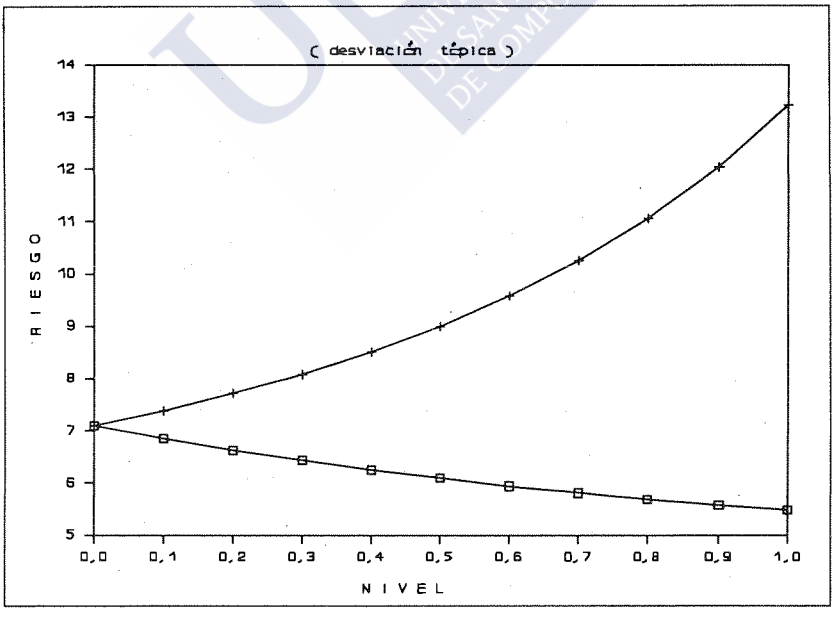


Figura 6. a) Riesgos de carteras por cortes.

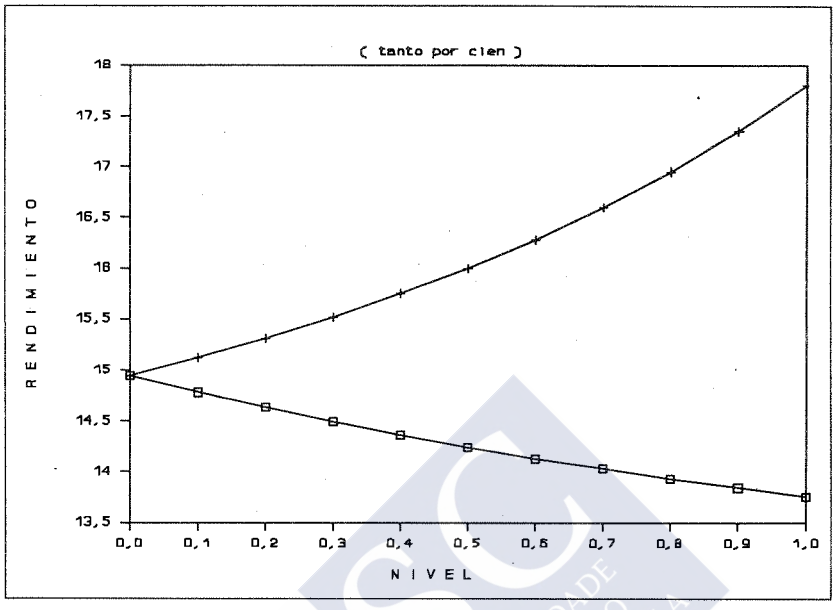


Figura 6. b) Rendimientos de carteras por cortes.

TABLA I. Valores Rendimiento Riesgo por niveles.

Nivel α	R_p A	Riesgo A	R_p B	Riesgo B
0,0	14,94	7,10	14,94	7,10
0,1	14,78	6,85	15,12	7,38
0,2	14,63	6,63	15,31	7,71
0,3	14,49	6,43	15,52	8,08
0,4	14,36	6,25	15,75	8,51
0,5	14,24	6,09	16,00	9,00
0,6	14,13	5,94	16,28	9,58
0,7	14,03	5,81	16,60	10,26
0,8	13,93	5,69	16,95	11,07
0,9	13,84	5,58	17,35	12,04
1,0	13,75	5,48	17,80	13,23

Consideremos, por ejemplo, que el nivel α admitido es 0,2 con lo que el tipo de interés se moverá entre (5,6 ; 6,4), y admitamos que la desviación admisible es del 10 %. La cartera de riesgo bajo estos supuestos se obtiene solucionando el siguiente programa:

$$\text{Max } \lambda$$

$$\lambda \leq 1 - \frac{z_1^* - R_p}{z_1^* - w_1}$$

$$\lambda \leq 1 - \frac{z_2^* - R_p}{z_2^* - w_2}$$

con $\sum_i x_i = 1$ y $0 \leq x_i \leq 1$.

El valor correspondiente a la cartera de riesgo asociada a un nivel $\alpha = 0,2$ sería:

$$R_p = 15,28$$

$$\sigma_p = 7,66$$

con unas proporciones a invertir de:

$$x_a = 0,88$$

$$x_b = 0,12$$

En la figura 7 se recogen los valores asociados a la proporción a invertir en la cartera de riesgo, que se obtiene tomando el valor de x correspondiente al mínimo de los máximos de las ecuaciones correspondientes a las restricciones.

De esta forma, una vez fijado el nivel α, garantizamos la existencia de una solución para la cartera de riesgo de acuerdo con la desviación que consideremos aceptable.

El problema se podría generalizar para el caso de n títulos, o incluso para planteamientos más generales de la teoría de selección de carteras (ventas al descubierto permitidas, diferentes tasas para préstamo y endeudamiento, etc...).

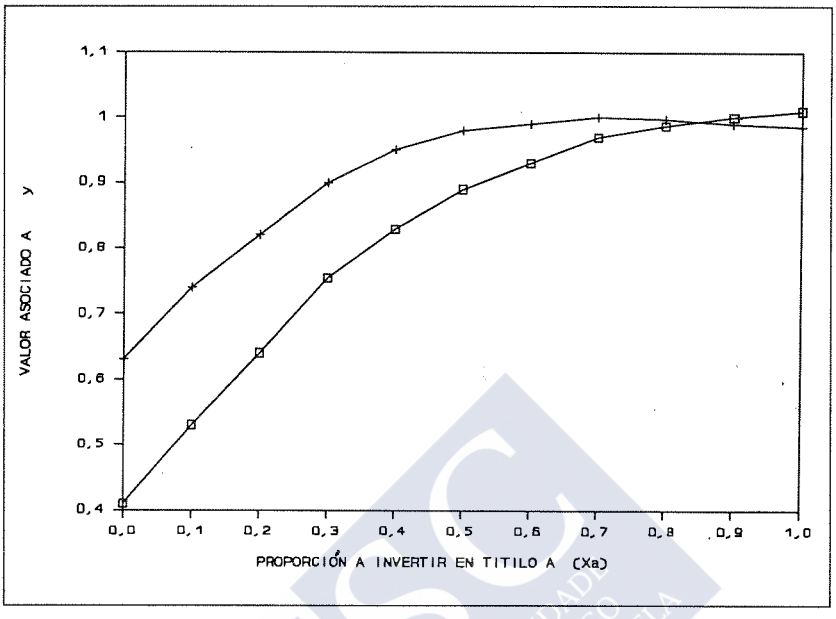


Figura7. Valores asociados a restricciones.

4. CONCLUSIONES

La teoría de los conjuntos borrosos permite, desde una óptica financiera, la modelización de los problemas de selección de carteras cuando los objetivos o restricciones incorporan variables borrosas. A este respecto, hemos corroborado la operatividad de la teoría en la selección de la cartera de riesgo cuando existe una tasa de interés para préstamo y endeudamiento. De la misma forma, el problema se podría plantear bajo hipótesis más generales. Así, se podría incorporar la borrosidad bajo los supuestos generales de Markowitz, o bien bajo las hipótesis simplificadoras de Sharppe.

Pero al mismo tiempo, desde una perspectiva más general, la teoría de conjuntos borrosos permite incorporar una nueva concepción de los problemas financieros más acorde con las dificultades de estimación de las magnitudes económico-financieras en un entorno tan incierto como el actual. Así, la teoría de los conjuntos borrosos puede resultar especialmente interesante en el tratamiento de problemas de planificación financiera (los objetivos y restricciones se pueden

considerar como subconjuntos borrosos [Diez de Castro y Martin Armario, 1985]); la selección de inversiones a través de una modelización multicriterio borrosa [Redondo López y col., 1989]; presupuestos [Benchimol, 1986]; etc. Pero, además, en el desarrollo de los modelos y a través de la función de pertenencia, siempre estaremos en condiciones de controlar el grado de entropía que consideremos adecuado.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

REFERENCIAS

Bellmann, R.E. y Giertz, M., «On the analytical formalism of fuzzy sets», *Information Sciences*, N. 5, (1973).

Benchimol, G., «L'intelligence artificial dans l'enterprise», *Rev. Direction et Gestion*, N.6, (1986).

Buckley, J.J., «The Fuzzy Mathematics of finance», *Fuzzy Sets and Systems*, N. 21, (1987).

Diez de Castro, E. y Martin Armario, E., «La toma de decisiones orientada hacia la optimización bajo restricciones en ambientes borrosos», *Rev. Economía y Empresa*, N. 5, (1985), 61-81.

Dinh, X.B., «A method for estimating the membership function of a fuzzy set», *Rev. Busefal*, LSI Univ. Paul Sabatier, Touluse, (1984), 68-82.

Elton, E.J. y Gruber, M.J., *Modern Portfolio Theory and Investment Analysis*, Ed. John Wiley, 1984, 71-77.

Kaufmann, A., *Introduction a la théorie des sousensembles flous: Elements théoriques de base*, Tomo I, Ed. Masson, París, 1977.

Kaufmann, A. y Gil Aluja, J., *Introducción a la teoría de los subconjuntos borrosos a la gestión de la empresa*, Ed. Milladoiro, 1986.

Kaufmann, A. y Gupta, M., *Introduction to Fuzzy Aritmetic: Theory and Applications*, Ed. Van Mostran Reinhold Co., New York, 1986.

Li Calzi, M., «Towards a general setting for the mathematics of finance», *Fuzzy Sets and Systems*, N. 35, (1990).

Redondo López, J.A. y col., «Modelo multidimensional borroso para la selección de inversiones», Ponencia presentada al congreso Hispano-Luso de Oporto, (1989).

Saaty, T.L., «Exploring the interface between hierarchies, multiple objetives and fuzzy sets», *Fuzzy Sets and Systems*, Vol. 1, N. 1, (1978), 57-68.

Stallings, W., «Fuzzy set theory versus bayesian statistics», *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 7, N. 3, Marzo, (1977).

Watson, S.R., Weiss, J.J. y Donnel, M., «Fuzzy decision analysis», *IEEE Transations on Systems, Man and Cybernetics*, Vol. 9, N. 1, (1979).

Zimmermann, H.J., «Fuzzy Programming with Several Objetive Functions», en Zimmermann, H.J., Zadeh, L.A. y Gaines, B.R., Eds., *Fuzzy Sets and Decision Analysis*, Ed. North Holland, Amsterdam, (1984).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

HARDWARE BORROSO

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

CAP. XII Soluciones Hardware en control borroso
CAP. XIII El proyecto ORBE

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

SOLUCIONES HARDWARE EN CONTROL BORROSO

Alberto Bugarín y Senén Barro
Departamento de Electrónica y Computación. Universidad de Santiago de Compostela

Ramón Ruiz
Departamento de Informática y Automática. Universidad de Murcia

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1. INTRODUCCIÓN

Sino el único, uno de los problemas fundamentales derivados de tratar con sistemas basados en lógica borrosa viene dado por el fuerte coste computacional que suponen. Este hecho resulta especialmente crítico en aplicaciones con restricciones extremas en cuanto a tiempo de respuesta por parte del sistema. Aliviar la carga computacional implica disponer de *hardware* dedicado suficientemente rápido.

El estudio de soluciones *hardware* dedicadas a la ejecución de sistemas basados en lógica borrosa es el objetivo de este capítulo. Más concretamente, nos hemos ceñido al estudio de soluciones en el campo de control de procesos, donde se enmarcan la práctica totalidad de desarrollos recogidos en la literatura, quizás por el hecho de que el control borroso ha sido el área de mayor auge dentro de las aplicaciones derivadas de la teoría de conjuntos borrosos y, sin duda, por su sencillez.

Aunque un sistema de control borroso no es estructuralmente complejo, su ejecución impone la realización de un muy elevado número de operaciones simples (tipo cálculo de mínimos y máximos). Además, la propia dinámica del proceso a controlar puede incrementar la carga computacional demandada. Por ello, dedicar a tareas de control borroso un computador/procesador de propósito general, parece una solución poco interesante desde el punto de vista de la relación coste/prestaciones. Es por ello que ha surgido una intensa labor de investigación en el campo del diseño de arquitecturas para control borroso, que permitan su implementación en un Circuito Integrado de bajo coste, y con la suficiente flexibilidad y potencia como para acomodarse a múltiples aplicaciones. En muchos casos, esta labor ha tenido proyección en soluciones disponibles comercialmente.

Se pueden seguir múltiples criterios taxonómicos en la descripción del *hardware* dedicado a control borroso. Según el modo de ejecución de la BC, pueden distinguirse soluciones que realizan la ejecución de las Sentencias Condicionales Borrosas en serie o en paralelo. Según la tecnología de implementación electrónica, podríamos hablar de soluciones digitales, analógicas o mixtas. Sin embargo, hemos creído más atractivo plantear una taxonomía que deriva de la forma en que se concreta la realización de inferencias durante la ejecución de la BC del controlador borroso. Por ello, aún suponiendo cierta familiaridad del lector con la metodología de diseño y ejecución de la Base de Conocimiento (BC) de un Sistema de Control Borroso, comenzaremos haciendo una descripción detallada del proceso de realización de inferencias en un Controlador Borroso.

2. TOMA DE DECISIONES EN SISTEMAS DE CONTROL BORROSOS

La **base de conocimiento** (BC) de un controlador borroso sintetiza, en forma de variables lingüísticas y relaciones entre valores lingüísticos de dichas variables, un modelo del conocimiento/comportamiento de un experto humano para llevar a cabo una tarea de control. Variables, valores y relaciones se estructuran en forma de reglas del tipo IF-THEN, conocidas como Sentencias Condicionales Borrosas (SCBs) o, simplemente, reglas borrosas. La estructura de dichas reglas queda caracterizada por una clara división entre su parte antecedente, en la que únicamente intervienen variables de estado del proceso, y su parte consecuente, en la que únicamente intervienen variables de control.

La **lógica de toma de decisiones** es el bloque encargado de la ejecución de la BC sobre los valores observados para las variables de estado del sistema. La mayor parte de las implementaciones *hardware* dedicadas al control borroso adoptan como estrategia de ejecución de la BC, la **regla composicional de inferencia** de Zadeh [1973], o regla de composición SUP-* (donde el operador * representa una t-norma).

Además de estos dos bloques, imprescindibles en cualquier controlador borroso, es frecuente que el controlador disponga de bloques de borrosificación y desborrosificación, que actuan de interfase entre el universo de valores puntuales asociado al proceso y el universo de distribuciones asociado al controlador. Aunque múltiples realizaciones *hardware* disponen de estos bloques, nosotros no los trataremos. Considerando que es posible reformular una BC genérica en base a SCBs del tipo Múltiple-Antecedente Único-Consecuente, una SCB de tipo general en un Controlador Borroso será de la forma

$$R_n^r\text{: IF } X_1 \text{ IS } A_1^r \text{ AND } \ldots \text{ AND } X_M \text{ IS } A_M^r \text{ THEN } Y_n \text{ IS } B_n^r \qquad (1)$$

ALSO ...

$$n = 1, \ldots, N, \qquad r = 1, \ldots, R.$$

Además, denotaremos por

$$\tilde{B}_n^r = \left\{ \tilde{b}_{n,j_n}^r \right\}, \qquad j_n = 1,\ldots,J_n, \qquad n = 1, \ldots,N, \qquad (2)$$

la distribución de posibilidad inferida para cada una de las N variables Y_n (que supondremos definidas en los respectivos universos de discurso V_n), tras la aplicación de la regla R_n^r sobre las distribuciones de posibilidad observadas para cada una de las M variables de estado X_m, m=1,...,M. Estas distribuciones de posibilidad vendrán representadas por:

$$\tilde{A}_m = \left\{ \tilde{a}_{m,i_m} \right\}, \qquad i_m = 1,\ldots,I_m, \qquad m = 1,\ldots,M. \qquad (3)$$

y son definidas, respectivamente, sobre los universos de discurso U_m, m=1,...M.

La Regla Composicional de Inferencia interpreta el proceso de ejecución de una SCB como la composición de dos relaciones: una relación de conjunción definida sobre el espacio de antecedentes, inducida por la observación simultánea de un conjunto de valores para las variables de estado (denotaremos esta relación como $S_{(\tilde{A}_1,\ldots,\tilde{A}_M)}$), y una relación de implicación, inducida por la propia existencia de

la SCB, que liga antecedentes y consecuentes y que, por comodidad, denotaremos de igual modo que la SCB correspondiente (R_n^r). La composición de ambas relaciones puede expresarse a través del operador SUP-*, de modo que

$$\tilde{b}_{n,j_n}^{\ r} = \bigvee_{U_1\times\ldots\times U_M} S_{(\tilde{A}_1,\ldots,\tilde{A}_M)}(i_1,\ldots,i_M) * R_n^r(i_1,\ldots,i_M,j_n) \qquad (4)$$

donde $a \vee b = MAX(a,b)$.

Consideramos que la relación asociada a los «valores observados» para las variables de estado se define como:

$$\begin{aligned} S_{(\tilde{A}_1,\ldots,\tilde{A}_M)}: U_1\times\ldots\times U_M &\rightarrow [0,1] \\ i_1,\ldots,i_M &\rightarrow S_{(\tilde{A}_1,\ldots,\tilde{A}_M)}(i_1,\ldots,i_M) \triangleq \tilde{a}_{1,i_1}\wedge\ldots\wedge\tilde{a}_{M,i_M} \end{aligned} \qquad (5)$$

asumiendo como operador conjunción, que define la relación de conjunción, la t-norma mínimo ($\wedge$).

Por su parte, la definición de la relación de implicación será

$$\begin{aligned} R_n^r: U_1\times\ldots\times U_M\times J_n &\rightarrow [0,1] \\ i_1,\ldots,i_M,j_n &\rightarrow R_n^r(i_1,\ldots,i_M,j_n) \end{aligned} \qquad (6)$$

Puesto que la t-norma escogida en la mayoría de las arquitecturas descritas en la literatura es el operador mínimo ($\wedge$) -Regla Composicional de Inferencia SUP-mín-, y utilizando las expresiones (4) y (5), obtenemos la ecuación que describe la distribución de posibilidad inferida punto a punto:

$$\tilde{b}_{n,j_n}^{\ r} = \bigvee_{i_1=1}^{I_1} \ldots \bigvee_{i_1=M}^{I_M} \left[\tilde{a}_{1,i_1}\wedge\ldots\wedge\tilde{a}_{M,i_M}\right] \wedge R_n^r(i_1,\ldots,i_M,j_n) \qquad (7)$$

Por otra parte, la implicación borrosa empleada en la práctica totalidad de las soluciones es la de Mamdani [1977], definida como

$$R_n^r(i_1,\ldots,i_M,j_n) = a_{1,i_1}^{\ r}\wedge\ldots\wedge a_{M,i_M}^{\ r}\wedge b_{n,j_n}^{\ r} \qquad (8)$$

Por tanto, tendremos,

$$\tilde{b}_{n,j_n}^{\ r} = \bigvee_{i_1=1}^{I_1} \ldots \bigvee_{i_M=1}^{I_M} \left[\tilde{a}_{1,i_1}\wedge\ldots\wedge\tilde{a}_{M,i_M}\right] \wedge \left[a_{1,i_1}^{\ r}\wedge\ldots\wedge a_{M,i_M}^{\ r}\wedge b_{n,j_n}^{\ r}\right] \qquad (9)$$

Dado que una BC consta de R reglas, y que la distribución de posibilidad inferida para cada variable de control se obtiene como la unión de las distribuciones parciales (interpretación del conectivo *ALSO*) obtenidas tras la aplicación de cada una de las reglas, tendremos,

$$\tilde{b}_{n,j_n} = \bigvee_{r=1}^{R} \bigvee_{i_1=1}^{I_1} \cdots \bigvee_{i_M=1}^{I_M} \left[\tilde{a}_{1,i_1} \wedge \ldots \wedge \tilde{a}_{M,i_M} \wedge a^{r}_{1,i_1} \wedge \ldots \wedge a^{r}_{M,i_M} \wedge b^{r}_{n,j_n} \right] \qquad (10)$$

donde,

$$\tilde{B}_n = \left\{ \tilde{b}_{n,j_n} \right\} \qquad j_n = 1,\ldots,J_n, \qquad n = 1,\ldots,N \qquad (11)$$

representa la distribución obtenida finalmente para la variable Y_n tras la aplicación de la base de reglas.

A partir de la expresión (10), se derivan dos metodologías de ejecución claramente distintas. Una de ellas, que denominaremos **Regla a Regla**, asumida mayoritariamente en las implementaciones *hardware*, consiste en la obtención de un **grado de compatibilidad** para cada variable de estado entre la distribución de posibilidad para ella observada y la distribución asociada al valor lingüístico vinculado a dicha variable en cada regla. Así, se reduce a un único valor numérico el proceso de ejecución de cada proposición antecedente. Mediante una posterior integración de todos los grados de compatibilidad correspondientes a las distintas premisas que componen la parte antecedente de una misma regla, se obtiene un grado de compatibilidad global para dicha regla, que se asume como grado de cumplimiento o verificación de su parte antecedente. Dicho grado de compatibilidad global modulará la distribución asociada a la variable lingüística reflejada en la parte consecuente de la regla para la variable de control correspondiente. El proceso finaliza tras la ejecución de todas las reglas, lo cual implica la necesidad de componer las distribuciones obtenidas para la misma variable en una única distribución final. Veámoslo formalmente: reordenando la expresión (10), y aplicando la propiedad distributiva entre los operadores $\wedge$ y $\vee$, resulta

$$\tilde{b}_{n,j_n} = \bigvee_{r=1}^{R} b^{r}_{n,j_n} \wedge \bigvee_{i_1=1}^{I_1} (\tilde{a}_{1,i_1} \wedge a^{r}_{1,i_1}) \wedge \ldots \wedge \bigvee_{i_M=1}^{I_M} (\tilde{a}_{M,i_M} \wedge a^{r}_{M,i_M}) \qquad (12)$$

Si definimos

$$\sigma^{r}_{m} \triangleq \bigvee_{i_m=1}^{I_m} \tilde{a}_{m,i_m} \wedge a^{r}_{m,i_m} \qquad (13)$$

habremos obtenido una expresión que proporciona un método de obtención del denominado grado de compatibilidad del antecedente m-ésimo en la regla r-ésima.

Sustituyendo en (12),

$$\tilde{b}_{n,j_n} = \bigvee_{r=1}^{R} (b_{n,j_n}^{r} \wedge \sigma_1^r \wedge \ldots \wedge \sigma_M^r) \tag{14}$$

tenemos que

$$\tilde{b}_{n,j_n} = \bigvee_{r=1}^{R} (b_{n,j_n}^{r} \wedge \sigma^r) \tag{15}$$

donde

$$\sigma^r \triangleq \bigwedge_{m=1}^{M} \sigma_m^r \tag{16}$$

representa la obtención del grado de compatibilidad global correspondiente a la parte antecedente de la regla r-ésima.

En la expresión (15) podemos observar cómo la modulación de la parte consecuente de las reglas, a que hacíamos referencia en el comentario anterior, tiene su implementación práctica como un «truncamiento» de la distribución de posibilidad asociada al valor lingüístico referenciado en dicha parte consecuente.

Una segunda metodología de evaluación de la BC, consiste, básicamente, en la «construcción» de una **Matriz de Implicación Global** (MIG) que liga la parte antecedente con la parte consecuente de las SCBs, de acuerdo con la relación de implicación asumida. Para una aplicación de control concreta, una vez obtenida esta matriz, cuyo cálculo se realiza en tiempo diferido, el proceso de ejecución de la BC presenta, como veremos, una gran similitud con la operación clásica Matriz*Vector.

Retomando las expresiones (7) y (10), tenemos

$$\tilde{b}_{n,j_n} = \bigvee_{r=1}^{R} \bigvee_{i_1=1}^{I_1} \ldots \bigvee_{i_M=1}^{I_M} \left[\tilde{a}_{1,i_1} \wedge \ldots \wedge \tilde{a}_{M,i_M} \wedge R_n^r (i_1,\ldots,i_M,j_n) \right] \tag{17}$$

Intercambiando operadores, y aplicando nuevamente la propiedad distributiva resulta

$$\tilde{b}_{n,j_n} = \bigvee_{i_1=1}^{I_1} \ldots \bigvee_{i_M=1}^{I_M} \left[\tilde{a}_{1,i_1} \wedge \ldots \wedge \tilde{a}_{M,i_M} \wedge R_n (i_1,\ldots,i_M,j_n) \right] \tag{18}$$

donde la expresión

$$R_n(i_1,\ldots,i_M,j_n) \triangleq \bigvee_{r=1}^{R} R_n^r(i_1,\ldots,i_M,j_n) \tag{19}$$

define analíticamente la Matriz de Implicación Global R_n.

Del análisis de la expresión (19), se deduce una diferencia significativa frente a la otra metodología de ejecución –expresión (15)– como es la total independencia del tamaño de la MIG respecto del número total de reglas de la BC. Esta característica hace que, en determinadas situaciones, esta metodología resulte computacionalmente más atractiva.

Siguiendo la metodología que hemos denominado Regla a Regla, se puede comprobar fácilmente que el número de operaciones básicas (NOB) realizadas durante la ejecución de una BC como la descrita, teniendo en cuenta todas las operaciones necesarias para realizar la inferencia sobre N variables de control, será de

$$NOB = R \cdot \left(2 \cdot \sum_{m=1}^{M} I_M + M + 2 \cdot \sum_{n=1}^{N} J_n\right) \tag{20}$$

Si, como sucede en la mayoría de los casos, se supone que todos los universos de discurso están discretizados en el mismo número de intervalos ($I_m=J_n=I$; m=1,...,M, n=1,...,N), la expresión anterior puede reducirse a

$$NOB = 2RI \cdot (M+N) + RM \tag{21}$$

En la expresión (21), se observa la existencia de una relación lineal con R, junto a una dependencia, por otra parte lógica, con otros parámetros relacionados con las características propias de cada problema (número de antecedentes y consecuentes en las reglas que conforman la BC e intervalos de discretización de los universos de discurso). De todos estos parámetros, aquél sujeto a una mayor variabilidad dependiendo de cada aplicación es el número de reglas.

La ejecución de la BC a través de la metodología MIG, considerando el mismo tipo de reglas, implicaría un número de operaciones

$$NOB_{MIG} = 2 \cdot \left(\prod_{m=1}^{M} I_M\right) \cdot \left(\sum_{n=1}^{N} J_n\right) + M \cdot \prod_{m=1}^{M} I_M \tag{22}$$

Suponiendo que el grado de discretización es uniforme (J_n=J=I, n=1,...N), tendremos

$$NOB_{MIG} = 2 \cdot I^M \cdot N \cdot I + M \cdot I^M = (2 \cdot N \cdot I + M) \cdot I^M \quad (23)$$

aunque, en la mayoría de los casos, puede realizarse la aproximación

$$NOB_{MIG} \approx 2 \cdot N \cdot I^M \quad (24)$$

La Fig. 1 muestra, para ambas estrategias de ejecución de la BC, la dependencia del número de operaciones básicas a realizar con el número de reglas R. Por simplicidad, se consideran reglas del tipo Múltiple–Antecedente Único–

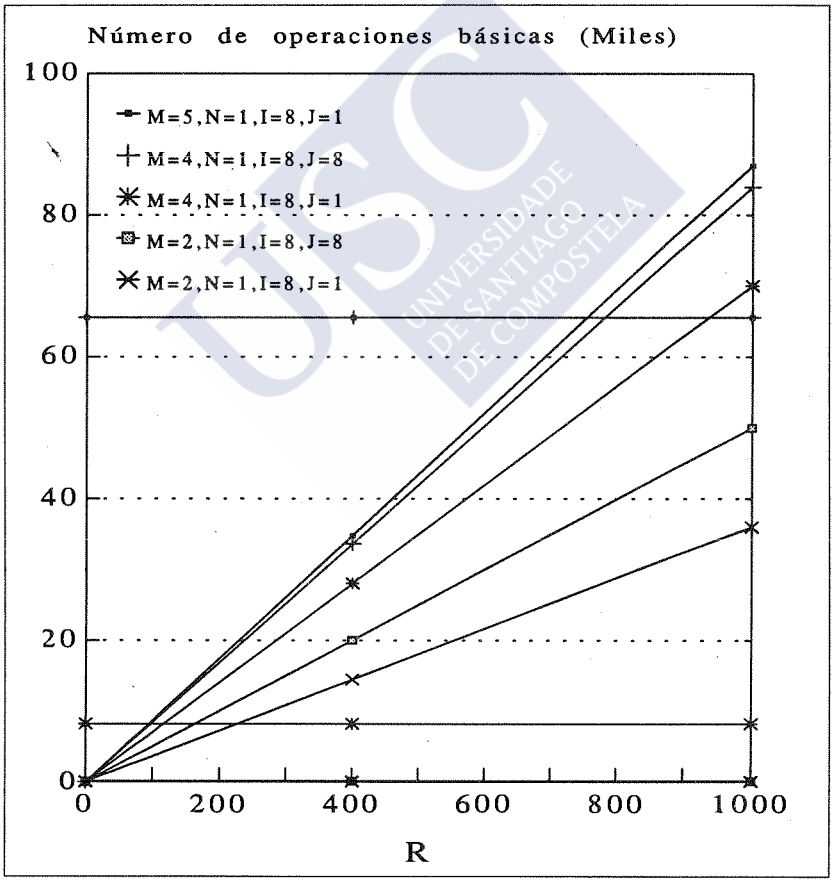


Figura 1. Comparación en distintos Sistemas de Control Borrosos de las estrategias de evaluación regla a regla y MIG.

Consecuente (N=1). Analizando los resultados, se observa que la metodología MIG resulta de mayor interés cuando el número de reglas es claramente dominante frente a los demás parámetros de la BC. La dependencia exponencial del NOB_{MIG} frente al número de antecedentes M –exp. (23)– hace que el proceso parezca poco adecuado para valores elevados de M y para discretizaciones muy finas de los universos de discurso. No obstante, pueden observarse algunas otras ventajas de utilizar la solución MIG: una mayor homogeneidad estructural del proceso de ejecución de la BC, trasladable a posibles arquitecturas *hardware*, facilitando su integración VLSI. Por otra parte, la MIG –cuyo cálculo se realiza en tiempo diferido– resume la relación de implicación borrosa escogida para la realización del proceso de inferencia. De esta forma, sin más que cambiar apropiadamente el valor de los elementos de dicha matriz, la misma arquitectura puede servir para distintas relaciones de implicación borrosa. Esta flexibilidad no es posible con la metodología de evaluación alternativa, dado que en ella la relación de implicación borrosa escogida (normalmente la de Mamdani, como ya se ha comentado) es plasmada directamente en estructuras *hardware*, que deben ser sustituidas (o «reprogramadas») para permitir la implementación de una relación borrosa distinta. Ambos aspectos se harán más evidentes en la exposición que sigue.

3. ESTUDIO DE SOLUCIONES *HARDWARE* PARA CONTROL BORROSO

En este punto expondremos una revisión de la literatura existente en lo que se refiere a soluciones *hardware* dedicadas al Control Borroso. Analizaremos sus arquitecturas, así como algunos de sus parámetros de rendimiento más significativos (complejidad *hardware*, tiempos de operación, modularidad, necesidad de interconexionado, etc.). Además de la obligada etapa de toma de decisiones, algunas soluciones incluyen etapas de Borrosificación y Desborrosificación, a las que, como hemos comentado previamente, no dedicaremos atención.

3.1. Método de evaluación mediante cálculo de grados de compatibilidad

En este apartado vamos a analizar aquellas soluciones *hardware* que abordan el proceso de inferencia según la estrategia de cálculo de grados de compatibilidad entre los valores observados y la parte antecedente de las reglas, así como la posterior modulación de la parte consecuente (expresiones (13) y (15)).

En esta línea, Manzoul [1987] propone una solución sistólica[1], consistente en considerar una BC constituida por una única variable en la parte antecedente o condición y otra en la parte consecuente o acción. Propone una proyección del proceso de ejecución de la BC sobre dos redes sistólicas lineales conectadas en serie, una de ellas dedicada al cómputo de los grados de compatibilidad y la otra a la modulación de la parte consecuente de las reglas. Ambas redes se basan en el mismo Procesador Elemental -PE- (Véase fig. 2a), radicando la diferencia básica entre ambas redes en el interconexionado existente entre los PEs que las componen.

La red sistólica que se encarga de la computación del grado de compatibilidad (fig.2b) está formada por I+R-1 PEs (fig.2c), siendo el número de ciclos consumidos para la obtención del último valor (σ_I^R) de 2I-1+2·(R-1)=2·(R+I)-3. Cada uno de los valores σ_i^r es introducido, a medida que va siendo calculado, en un array sistólico de J PEs, encargado de computar la modulación e integración de la parte consecuente de las reglas. Desde que está disponible el primer grado de compatibilidad (σ_I^1), este segundo array consume un total de 2R+J-2 ciclos hasta calcular el valor final inferido para la variable de control.

Togai y Watanabe [1985, 1986], plantean la ejecución de la BC manteniendo la metodología de ejecución del trabajo de Manzoul [1987], pero con una ejecución en paralelo del conjunto de SCBs. El almacenamiento de las distribuciones de posibilidad asociadas a cada regla se realiza en dos módulos ROM (uno para los antecedentes y otro para los consecuentes) cuyas palabras están formadas por un bit de cada una de las reglas almacenadas. La codificación propuesta permite una cuantificación del rango [0,1] de valores de posibilidad o grados de pertenencia en 16 niveles (4 bits), para cada uno de los 31 puntos en que se discretizan los respectivos universos de discurso. El acceso a las reglas se produce en paralelo; sin embargo, dentro de cada regla, su ejecución es en serie, leyéndose los datos codificados bit a bit. Esta filosofía de acceso y tratamiento bit-serie, permite una considerable reducción en los requerimientos de tamaño de palabra de la memoria y de *hardware* de cálculo, a costa de una mayor lentitud en la ejecución de las reglas. De hecho, en posteriores implementaciones se opta por una filosofía mixta paralelo-serie que reduce los tiempos de lectura [Watanabe y col., 1990].

---

[1] Un array sistólico [Kung y Leiserson, 1978], es básicamente un conjunto de celdas o procesadores muy simples, interconectados entre sí de forma regular y modular, y que alcanza unas altas prestaciones de cálculo mediante la ejecución concurrente de operaciones.

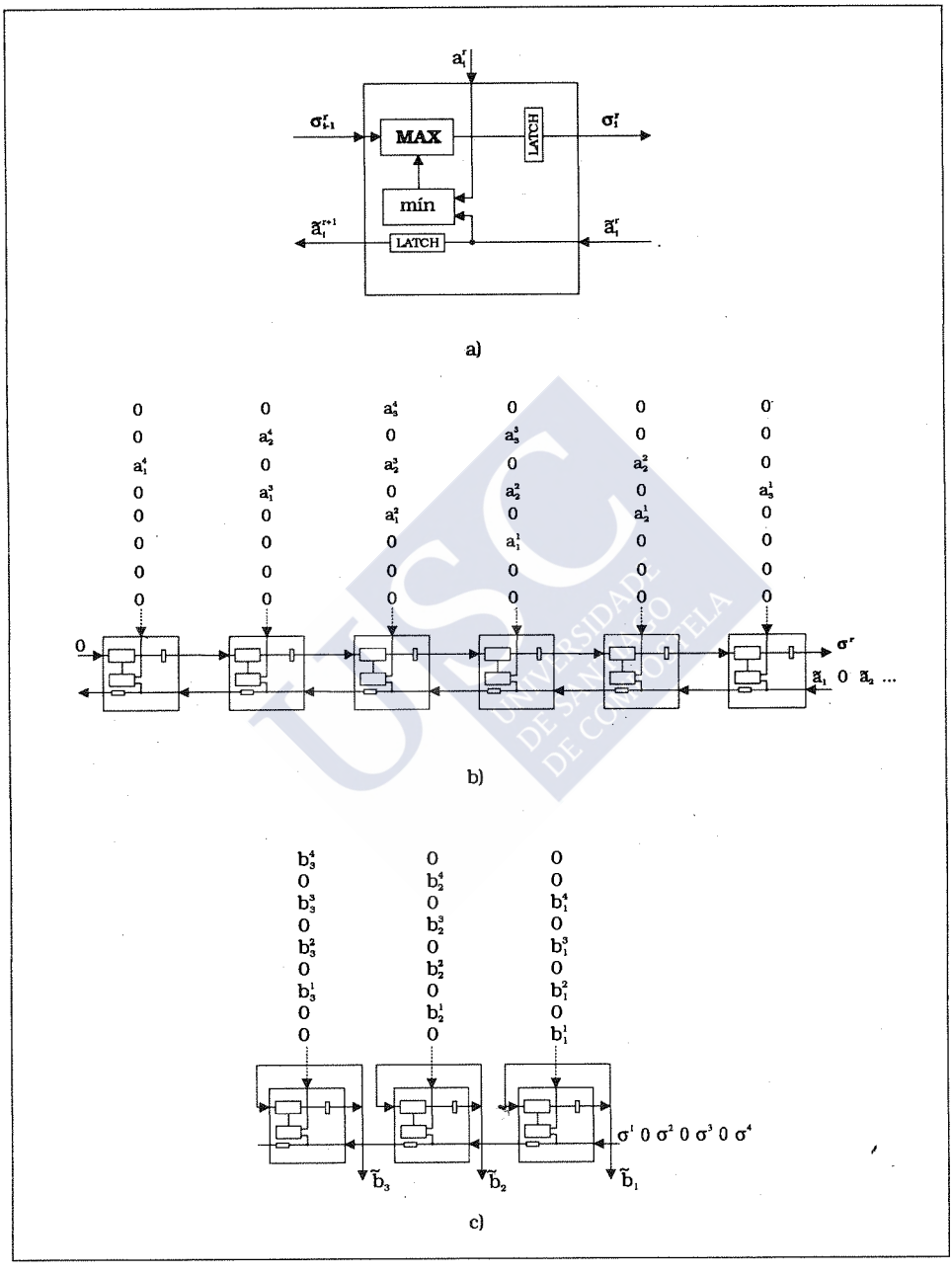


Figura 2. a) PE de la solución sistólica de Manzoul [1987]. b) Red de PEs para la obtención de los grados de compatiblidad. c) Modulación de la parte consecuente de las reglas (R=4, I=3).

En lo que respecta a la arquitectura de evaluación, el esquema propuesto se describe en la Fig. 3, donde se muestra la trayectoria que siguen los datos correspondientes a una regla durante el proceso de evaluación. Esta estructura básica se duplica para cada regla que compone la BC, realizándose posteriormente la unificación de las distribuciones de posibilidad obtenidas para la variable de salida. El proceso de ejecución de cada regla Único-Antecedente Único-Consecuente consume un total de $\Delta_r=\Delta+4\cdot I+4\cdot J=\Delta+4\cdot(I+J)=\Delta+8\cdot I$ ciclos, donde el símbolo Δ representa un retardo inicial inespecificado, y en donde se ha asumido una discretización uniforme de los universos de discurso asociados a las distintas variables de estado y variables de control (I=J).

Una primera implementación VLSI de esta arquitectura, admitía la ejecución de 16 reglas en paralelo, en un circuito integrado de 68 pines, alcanzando aproximadamente 80000 Inferencias Borrosas Por Segundo (IBPS) a 20.8 MHz.

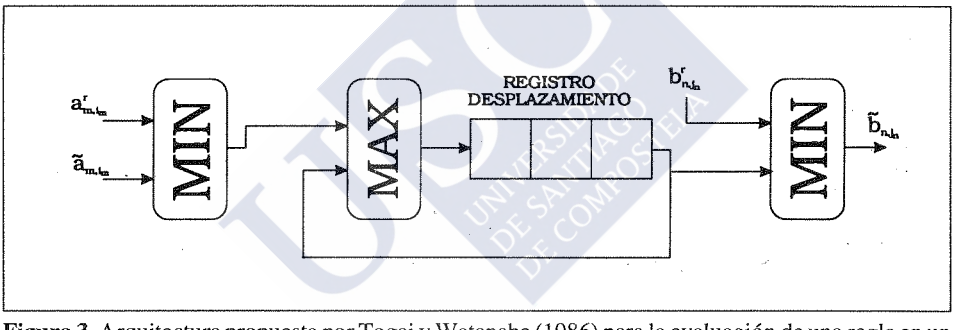


Figura 3. Arquitectura propuesta por Togai y Watanabe (1986) para la evaluación de una regla en un Sistema de Control Borroso.

Partiendo del desarrollo anterior, Watanabe y col. [1990] realizan un circuito integrado que presenta numerosas mejoras. Manteniendo la filosofía de ejecución en paralelo de la base de reglas, se propone una codificación mixta paralelo-serie, que permite un acceso más rápido a los valores que constituyen la base de reglas. De este modo, se accede a cada punto de cada universo de discurso (codificado igualmente en 4 bits), en un único ciclo de reloj. Como contrapartida, debe cuadruplicarse el ancho de palabra de la memoria dedicada al almacenamiento de la base de reglas. Además, en lo que se refiere a la definición de los valores lingüísticos, se incrementa el nivel de discretización de los universos de discurso (pasándose de 31 a 64 intervalos). En cuanto a la estructura de la BC, se incrementa el número de antecedentes de las SCBs, y el número de éstas, permitiéndose la configuración mediante *software* de una base de hasta 102 reglas (del tipo 2-

Antecedentes 1-Consecuente) ó 51 reglas (del tipo 4-Antecedentes 2-Consecuentes).

El circuito es capaz de realizar unas 560000 IBPS, muy superior a la velocidad de procesamiento de la solución anterior, y ello a pesar de poder tratar con BC más complejas. Además, el diseño modular que implementa esta solución hace posible su aplicación en BC de mayor complejidad, mediante la conexión en cascada de los chips.

La solución de Eichfield y col. [1992], además de una alta velocidad de cálculo, persigue la optimización de los requerimientos de memoria del Controlador Borroso. En ella se presenta una arquitectura novedosa, en la que únicamente llegan a ejecutarse aquellas reglas para las que el grado de compatibilidad con los valores observados es no nulo. En la definición de los valores lingüísticos correspondientes a cada variable de control (limitados a 7), se permite un grado dos de máximo solapamiento entre las distribuciones de posibilidad asociadas a los valores lingüísticos. Esto es, dado un valor observado (puntual) para una variable de control, sólo existirán dos valores lingüísticos que tengan asociado un valor no nulo en dicho punto. De esta forma, cada vez que se evalúa la BC, el máximo número de reglas que contribuirán al proceso de inferencia será de 2^M, donde M representa el número de antecedentes de las reglas. Una vez seleccionadas las reglas, se realiza en paralelo el cálculo de los grados de compatibilidad parciales de su parte antecedente y la operación de modulación de su parte consecuente, lo que (incluída una fase de desborrosificación), implica un total de $2^M \cdot J$ ciclos. Este número, como cabría esperar, y al contrario que en las arquitecturas previas, es independiente del número de reglas que componen la BC. El principal inconveniente de esta solución lo constituye la severa discretización propuesta, tanto para el intervalo [0,1] (3 bits), como para los universos de discurso (5 bits).

Una arquitectura disponible comercialmente, y que presenta ciertas variantes respecto de las descritas hasta ahora, es el procesador borroso NLX 230 de American Neuralogix Inc. [1991]. El método de evaluación tradicional, en el que múltiples reglas contribuyen al cálculo del valor de salida, se sustituye aquí por un sistema de selección, según el cual sólo se toma en consideración la regla que posea el mayor grado de compatibilidad con los valores observados. Además, las funciones de pertenencia asociadas a los valores lingüísticos declarados en las reglas son descritas únicamente por dos parámetros: Centro (C) o punto –que se asume único– del universo de discurso para el cual la distribución de posibilidad toma el valor máximo. Ancho (W), indicativo del intervalo donde la distribución de posibilidad toma valores no nulos (Intervalo [C-W, C+W]). El cálculo de los

grados de compatibilidad se realiza mediante una técnica distinta a las reseñadas con anterioridad, consistente en realizar una estimación del alejamiento entre C y el valor de la variable de entrada. La arquitectura del circuito integrado permite la implementación de un máximo de 16 proposiciones distintas para valores de entrada de tipo puntual. Estas proposiciones pueden concatenarse entre sí mediante el conectivo AND para dar lugar a la parte antecedente de las distintas reglas. Se permite almacenar un máximo de 64 reglas, en cada una de las cuales únicamente se puede referenciar una variable de control, de entre 8 posibles. Una característica novedosa de esta arquitectura es que los valores puntuales de acción generados pueden hacerse recircular a la entrada, permitiendo la concatenación de reglas.

Hasta ahora hemos visto soluciones digitales. Sin embargo, también son numerosas las soluciones en el campo del diseño analógico que implementan un amplio abanico de posibilidades: desde etapas concretas dentro del proceso de evaluación [Ruiz Mayor y Gutiérrez, 1991; Sasaki y col., 1990, 1992; Patyra, 1991; Yamakawa, 1987, 1989a, 1989b; Yamakawa y Miki, 1986; Yamakawa y Kabuo, 1988; Yamakawa y Sasaki, 1987], cuya aplicabilidad suele superar el ámbito del control de procesos, hasta sistemas completos que implementan el proceso global de inferencia y obtención de valores puntuales para las variables de control [Yamakawa, 1988a, 1988b]. Sin duda, la más representativa de dichas propuestas es la de Yamakawa [1988a], que plantea una solución *hardware* de evaluación de Sistemas de Control Borrosos, con una filosofía de evaluación similar a la de Watanabe y col. [1990]. Este *hardware* alcanza un rendimiento extraordinario –casi triplica la velocidad de evaluación del procesador de Watanabe y col. [1990]–, aunque considerando SCBs más simples (del tipo 2-Antecedentes 1-Consecuente). La solución de Yamakawa ha sido utilizada con éxito en numerosas aplicaciones [Yamakawa, 1989b], e incluso comercializado como circuito integrado bajo la denominación de OMROM FP-3000.

En el proceso de cálculo de los grados de compatibilidad, se utiliza un bloque que permite considerar distribuciones de posibilidad de morfología trapezoidal o triangular, las cuales pueden ser dispuestas hasta en 7 posiciones distintas en función de su distancia al origen. Para cada distribución seleccionada, y teniendo como entrada un valor de tensión directamente proporcional al valor puntual de la variable de estado considerada, se produce una tensión de salida directamente proporcional al valor que toma en ese punto dicha distribución.

El cómputo del grado de compatibilidad se realiza calculando el valor mínimo entre las tensiones de salida derivadas de los bloques asociados a cada antecedente. Este valor obtenido servirá como modulador de la distribución de

posibilidad descrita para la variable de control. Dicha distribución es proporcionada por un «generador de funciones de pertenencia» (Fig. 4) que proporciona simultáneamente los valores correspondientes a 25 puntos en los que se considera discretizado el universo de discurso. El nivel de cuantificación del intervalo [0,1] es extraordinariamente pobre (4 niveles), lo que, unido a un número máximo de 5 elementos del universo de discurso cuyo valor puede ser no nulo, constituye una severa restricción en cuanto a la representación de valores lingüísticos. Estos 25 valores de tensión sirven como entradas a un bloque en donde serán «truncados» por el grado de compatibilidad derivado del bloque anterior, operando sobre ellos en paralelo a través de una estructura de cálculo de mínimos. La salida (que constituye la distribución de posibilidad inferida para la variable de control) pasa por un bloque de cómputo de valores máximos que realiza la unificación de las distribuciones inferidas al evaluar las distintas reglas.

Algunas ventajas del diseño analógico utilizado son la alta capacidad de integración de la solución propuesta y la velocidad de procesamiento adquirida. En lo que se refiere a este último aspecto, se consigue una ejecución completa de la BC cada 670 ns, lo que permite una velocidad de procesamiento mayor de 1 MIBPS. La contrapartida exigida es el altísimo nivel de adecuación del *hardware* al proceso de evaluación, lo que lo hace muy poco flexible y completamente inadaptable a otros procesos de inferencia. Sin duda, la rigidez frente a posibles cambios es un inconveniente general de las soluciones analógicas frente a las digitales.

Centrándonos en las soluciones a etapas concretas dentro del proceso de evaluación, los trabajos de Ruiz Mayor y Gutiérrez [1991] enfocan el problema de la generación de las funciones de pertenencia o distribuciones a utilizar en la BC de un Sistema de Control Borroso. La técnica empleada para la generación de las funciones de pertenencia profundiza en la propuesta de Yamakawa [1988a], obteniéndose una implementación más eficiente. Básicamente, consiste en un circuito analógico cuya entrada es un nivel de corriente representativo de un elemento del universo de discurso cuyo valor bajo la función de pertenencia se desea obtener y cuya salida es otro nivel de corriente, esta vez proporcional al valor referido. Un problema asociado a esta representación en niveles de corriente es que el *fan-out* se encuentra restringido a 1, lo cual puede plantear problemas de diseño, al existir entradas (por ejemplo, las variables de estado) que deben distribuirse en varios bloques. En este circuito, construído tomando como elementos básicos espejos de corriente implementados en tecnología CMOS, se caracterizan las funciones de pertenencia, cuya morfología genérica es de tipo trapezoidal, mediante cuatro parámetros (anchura de la zona central, pendientes de los flancos izquierdo y derecho, y punto medio de la zona central).

Algunas ventajas del diseño analógico frente al digital, velocidad de procesamiento, ahorro en superficie de integración, reducción en la potencia disipada, etc., no ocultan, sin embargo, el inconveniente de su mayor sensibilidad al ruido, y, sobre todo, su alta dependencia de las características propias de cada transistor, que pueden dar lugar a la aparición de efectos no deseados (efectos de segundo orden en la operación de los espejos de corriente, por ejemplo), pudiendo llegar a producir fallos de funcionamiento. En este sentido, se han presentado soluciones que intentan compensar, al menos parcialmente, este tipo de errores (resulta clarificador el análisis de Patyra [1991]).

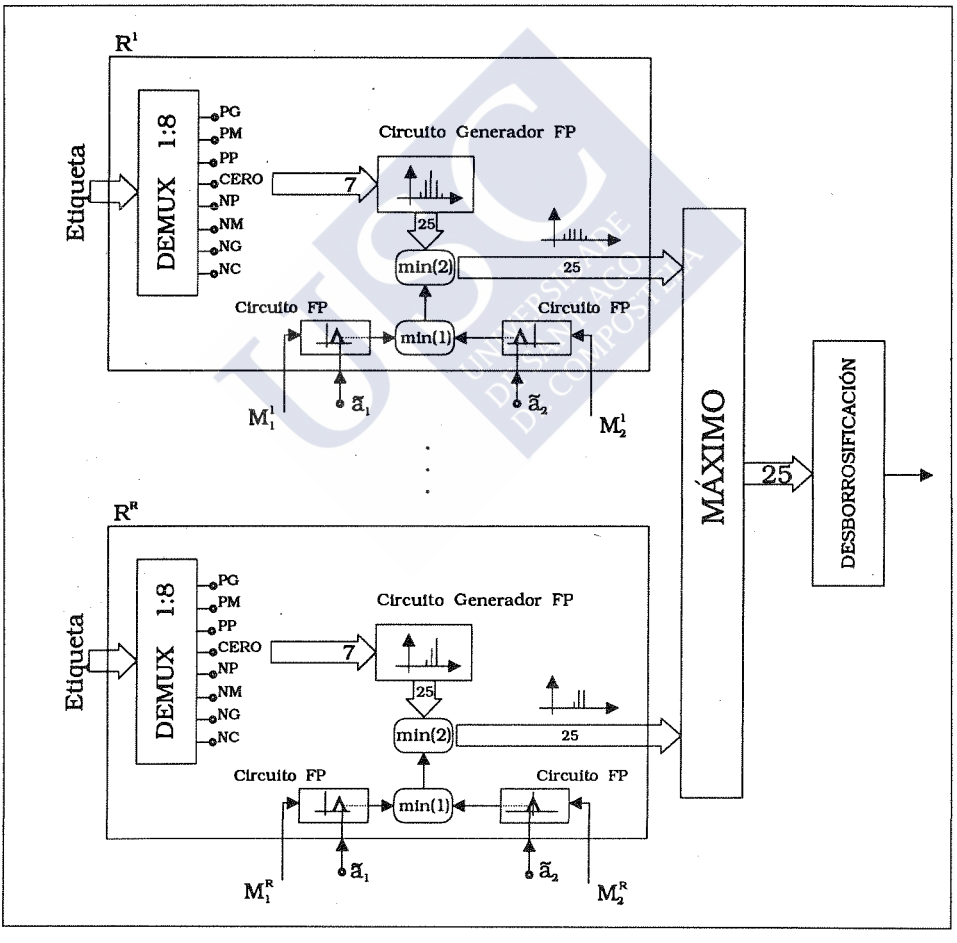


Figura 4. Inferencia de los valores lingüísticos para las variables de control en la solución de Yamakawa [1988].

3.2. Método de evaluación mediante matrices de implicación globales

La segunda estrategia descrita para la evaluación de Sistemas de Control Borrosos, resume en una única Matriz de Implicación Global (MIG) el conjunto de reglas que conforman la BC. La ejecución de la base de reglas según esta estrategia presenta un paralelismo estructural con la operación clásica Matriz*Vector. Por ello, muchas de las soluciones *hardware* que se presentan en esta línea, hacen uso, más o menos explícitamente, de esta analogía.

Manzoul y Serrate [1988] plantean una solución para BC del tipo único-antecedente único-consecuente a través de dos redes sistólicas: la primera dedicada a la obtención de la MIG, poco interesante, ya que puede hacerse en tiempo diferido, y la segunda al proceso de ejecución, a partir de dicha matriz, de la base de SCBs. Esta última es una red sistólica lineal constituída por J PEs (Fig. 5), que necesita I+J ciclos de reloj hasta la obtención del valor final.

La generalización del modo de evaluación basado en MIGs a reglas con múltiples variables, tanto en la parte acción como en la parte condición, nos aproxima un poco más a la realidad de los sistemas de control borrosos, mayoritariamente caracterizados por presentar más de una proposición en las partes *IF* y *THEN* de las reglas que componen su base de conocimiento. Manzoul y Tayal [1990] abordan una arquitectura seudo-sistólica para la evaluación de Sistemas de Control Borroso de este tipo. Su solución, no obstante, no es más que una aproximación a la evaluación estricta de la BC, basada en obtener la MIG para una BC tipo Múltiple-Antecedente Múltiple-Consecuente y considerar las N submatrices que resultan de proyectar la MIG sobre los universos de discurso asociados a las N variables de control. Un planteamiento estricto de la evaluación a través de MIGs de Sistemas de Control Borrosos Múltiple-Antecedente Múltiple-Consecuente, consiste en la reformulación del Sistema de Control Borroso en reglas del tipo Múltiple-Antecedente Único-Consecuente (resultando R*N reglas). Estas reglas pueden agruparse en subsistemas con igual variable de acción en su parte consecuente, dando lugar a N MIGs, que permitirán realizar el proceso de inferencia para cada nuevo valor observado de las variables de estado [Ruiz y col., 1991]. Un trabajo posterior de Manzoul y Jayabharathi [1992], confirma esta conclusión, al abandonar los autores su método de evaluación inicial y utilizar el desdoblamiento y posterior reordenación de la base de reglas que hemos expuesto. No obstante, esta nueva solución propone una implementación que sigue una metodología radicalmente distinta a las anteriores, sustituyendo la arquitectura

sistólica por un dispositivo FPGA que implementa, en forma de tabla (*look-up table*), todas las posibles respuestas del Sistema de Control Borroso. La imposibilidad de operar cuando los valores de entrada a la base de reglas son distribuciones de posibilidad, y no valores puntuales, restringe el ámbito de aplicación de esta solución.

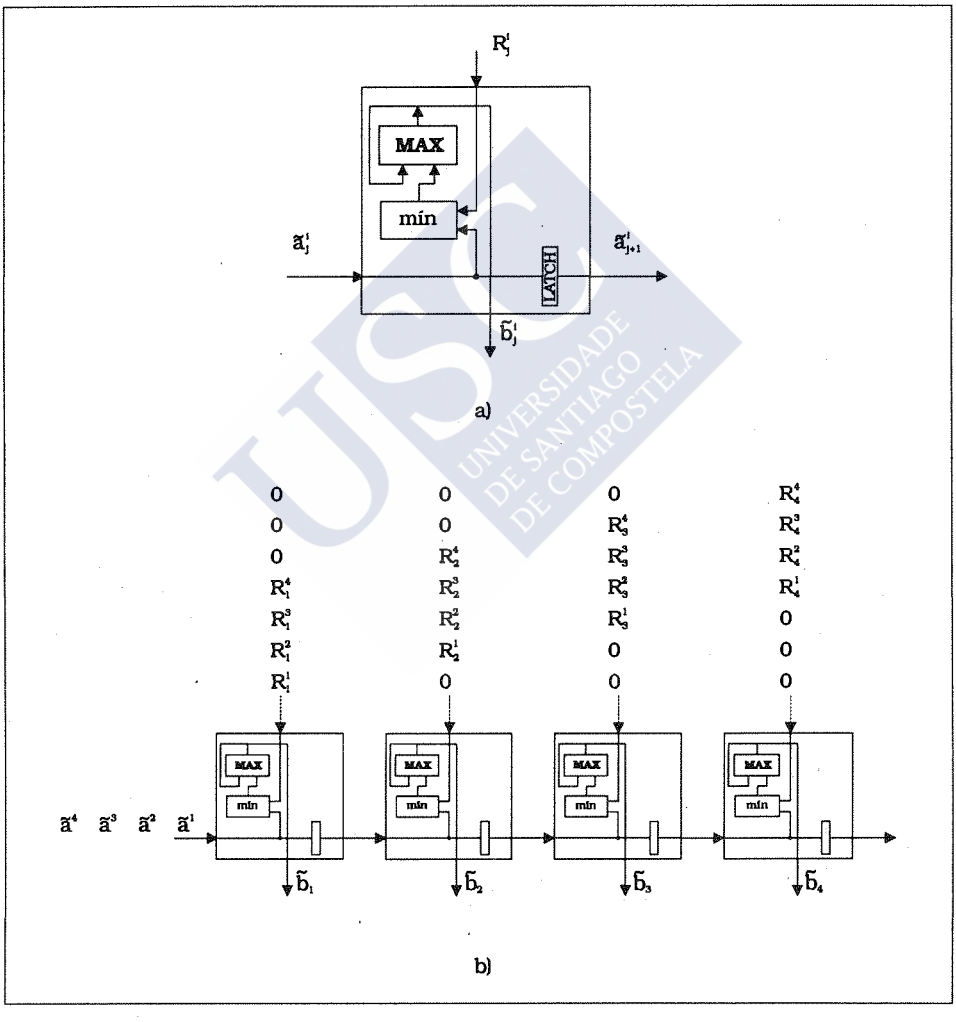


Figura 5. a) PE de la solución sistólica de Manzoul y Serrate [1988]. b) Red de PEs para la evaluación de un Sistema de Control Borroso genérico.

Una arquitectura sistólica de tamaño arbitrario

Nuestro grupo ha diseñado una arquitectura sistólica basada en el método de desdoblamiento y posterior reordenación de la base de reglas [Bugarín y col., 1992]. Siguiendo la metodología de evaluación basada en MIG, hemos propuesto una arquitectura sistólica que se beneficia de la equivalencia estructural existente entre dicha estrategia de ejecución de la BC y el producto clásico Matriz*Vector. La característica más relevante de la solución propuesta es su independencia frente al tamaño de la BC a ejecutar, a diferencia de las demás soluciones presentadas, aun aquellas sistólicas, en las que el número de PEs depende, en general, de las dimensiones del problema (número de reglas, número de proposiciones en la parte antecedente y/o consecuente de cada regla, grado de discretización de los universos de discurso). Por ello, nuestra solución presenta una estructura altamente versátil con respecto a sus posibilidades de integración/aplicación. Antes de describir esta arquitectura, expondremos un proceso previo de transformación de la MIG.

En el apartado 2 se dedujeron las expresiones que caracterizan la metodología de evaluación basada en MIGs: la propia definición de una MIG genérica R_n (expresión (19)) y la realización del proceso de inferencia (expresión (18)). Dado que la dimensión de la MIG es M+1, donde M es el número de antecedentes de cada regla, una implementación directa de la expresión (18) requeriría arquitecturas multidimensionales (hipercubo, sistólicos multidimensionales, ...). Sin embargo, pensando en una posible integración VLSI de la solución, debe forzarse una reducción de la dimensionalidad de dicha expresión que permita su proyección sobre redes de procesadores bidimensionales. La necesaria reducción de la dimensionalidad de la expresión (18) para conseguir una total analogía con la expresión clásica del producto Matriz*Vector, debe realizarse a través de la transformación de la MIG M+1-dimensional en una matriz bidimensional equivalente, así como del conjunto de valores observados en un vector unidimensional equivalente. La realización de ambas transformaciones implicará establecer una aplicación entre el espacio de los subíndices de la MIG y otro espacio de menor dimensión (2-D en el caso de la matriz y 1-D en el caso del vector).

Definimos la MIG 2-D equivalente a R_n como

$$Dn_{j_n s} = R_n(i_1,\ldots,i_M,j_n) \tag{25}$$

donde, a cada coeficiente s le corresponde un vector de índices establecido mediante una aplicación F definida como:

$$\begin{array}{ll} F: \{1,\ldots,S\} & \rightarrow \{1,\ldots,I_1\}\times\ldots\times\{1,\ldots,I_M\} \\ s & F(s) = (i_1,\ldots,i_M) \end{array} \tag{26}$$

con

$$S = \prod_{m=1}^{M} I_m \tag{27}$$

La aplicación F reordena los $S \cdot J_n$ elementos de la matriz (M+1)-dimensional en elementos de una matriz 2-D formada por S filas y J_n columnas.

La correspondencia entre cada subíndice s de la matriz $D_{j_n s}$ y cada vector de índices $(i_1,\ldots,i_M)$ de la MIG puede expresarse analíticamente como solución única de la siguiente ecuacion diofántica

$$s = i_M + \sum_{k=1}^{M-1} (i_k - 1) \cdot \prod_{m=k+1}^{M} I_m \tag{28}$$

La Fig. 6 muestra una representación esquemática del proceso de transformación descrito.

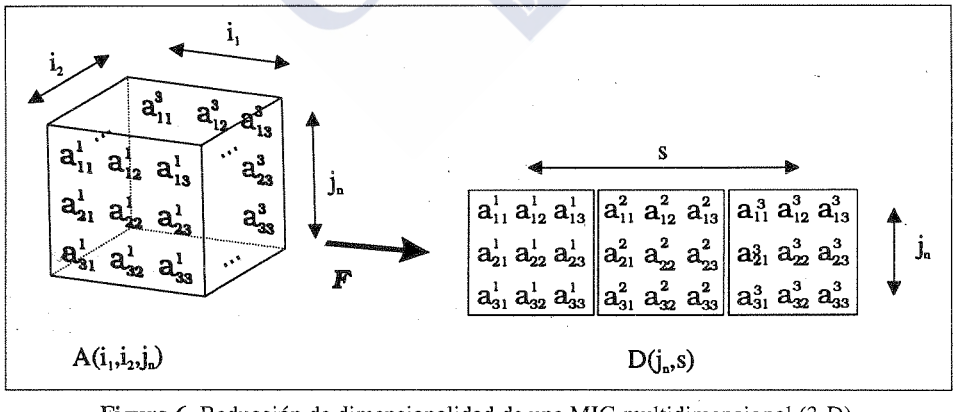


Figura 6. Reducción de dimensionalidad de una MIG multidimensional (3-D).

La transformación del término relacionado con los valores observados en un vector se realiza mediante la misma aplicación F, que reordena los índices de

manera análoga a como opera con la MIG. La definición de dicho vector puede establecerse como:

$$\tilde{a}_s = \tilde{a}_{1,i_1} \wedge \ldots \wedge \tilde{a}_{M,i_M} \qquad s = 1,\ldots,S \tag{29}$$

Finalmente, teniendo en cuenta las definiciones (25), (27) y (29), podemos reescribir (18) como

$$\tilde{b}_{n,j_n} = \bigvee_{s=1}^{S} Dn_{j_n,s} \wedge \tilde{a}_s \ , \qquad j_n = 1,\ldots,J_n \tag{30}$$

expresión que coincide, formalmente, con la expresión clásica del producto Matriz*Vector, sin más que sustituir los operadores "$\wedge$" y "$\vee$" por "$\times$" y "Σ", respectivamente. De esta forma, mediante un simple intercambio de bloques de operación, puede realizarse la evaluación de Sistemas de Control Borrosos a través de arquitecturas diseñadas para el producto de matrices. Sin embargo, dado el tamaño de la matriz, una implementación directa implicaría un número de PEs intratable. Pensando en una solución realizable resulta inevitable aplicar técnicas de particionamiento y proyección de algoritmos adecuadas para este tipo de operación. Nuestra solución resulta de adaptar la técnica DBT *(Dense to Band matrix transformation by Triangular-block partitioning)* propuesta por Navarro y col. [1987]. El objetivo no es otro que conseguir una estructura *hardware* de tamaño no dependiente del tamaño del problema (en nuestro caso, del tamaño de la BC a ejecutar).

En la Fig. 7 se muestra un ejemplo del fundamento teórico de la técnica DBT. La Fig. 7a presenta el cálculo de un producto Matriz*Vector siguiendo la técnica tradicional. Evidentemente, el número de operandos y el número de operaciones a realizar para la obtención de cada componente del vector resultado depende directamente del tamaño de la matriz. En la Fig. 7b se presenta el planteamiento y la resolución de un problema transformado a partir del anterior. Lógicamente, la resolución del problema transformado no constituye directamente la solución del problema inicial, haciéndose necesaria una etapa de composición de las soluciones parciales obtenidas (Fig. 7c). La resolución del problema mediante ambos procedimientos implica la realización del mismo número de operaciones. Sin embargo, la segunda metodología presenta una ventaja de tipo «estructural», consistente en que el número de operaciones a realizar en cada etapa depende exclusivamente del número de elementos no nulos que hay en cada columna o fila de la matriz transformada (lo que se conoce como su ancho de banda) y éste es escogido arbitrariamente. El proceso global podrá realizarse, por tanto, en una estructura de

tamaño independiente de las dimensiones del problema original. Este hecho, añadido a la elevada reiteración y sencillez de las operaciones implicadas, hacen que el procedimiento descrito resulte altamente adecuado para su proyección sobre una arquitectura sistólica. Lógicamente, un punto clave en este método lo constituye la obtención del problema transformado a partir del problema original, que no describiremos en detalle por falta de espacio (véase Navarro y col., [1987]). Básicamente, consiste en el cálculo de una matriz de banda transformada DTn partiendo de la matriz original Dn. En esta matriz de banda resultante los únicos elementos que pueden ser no nulos se agrupan en una banda diagonal de anchura ω. El valor de ω, que se establece arbitrariamente, fija el número de operaciones que se realizarán en cada una de las etapas de cálculo, y habrá de coincidir con el número de PEs que constituirán la arquitectura sobre la cual se va a realizar la proyección del problema. Una vez establecido ω, se obtienen dos valores $S', T' \in \mathbb{N}$, tales que el número de filas y el de columnas de la matriz original Dn (respectivamente, J_n y S) verifiquen

$$\begin{aligned} S &= \omega \cdot S' \\ T \Delta J_n &= \omega \cdot T' \end{aligned} \tag{31}$$

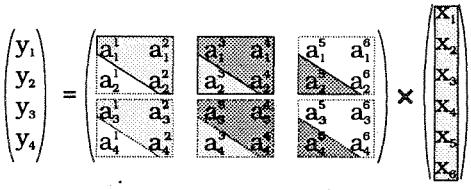
para lo cual puede ser necesario añadir columnas y/o filas de ceros a la matriz original.

La Fig. 8a muestra el esquema de la celda básica (o PE) de la arquitectura sistólica propuesta. Las interconexiones entre PEs, así como el flujo global de datos se muestran en la Fig. 8b, para un ejemplo concreto con 3 PEs y una MIG formada por S=6 columnas y T=3 filas.

Cada PE realiza una operación en cada etapa y transfiere al siguiente PE el resultado parcial, de manera que se concluye dicha etapa cuando se dispone del resultado a la salida del primer PE. Esta dependencia entre el número de PEs y el ancho de banda de la matriz transformada condicionará, como veremos, algunos de los parámetros indicativos del rendimiento de la solución, como son la frecuencia de salida de datos y el tiempo total consumido en la evaluación de la BC.

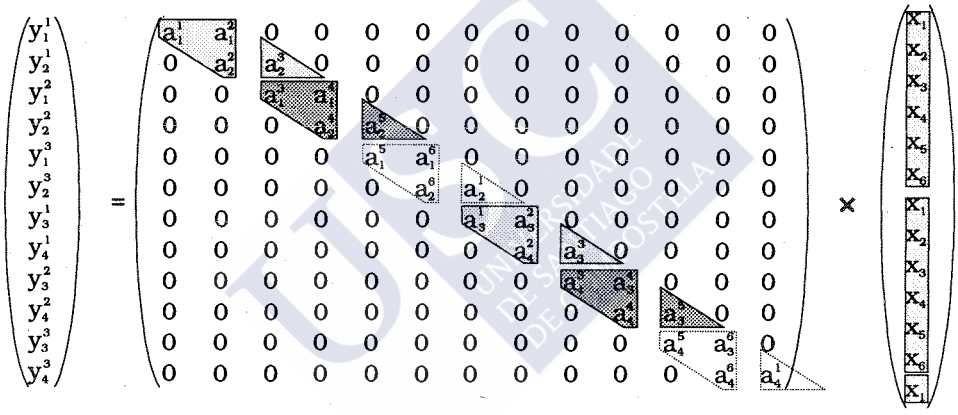
El volcado de los coeficientes de la matriz en la red de PEs sigue una secuencia extremadamente simple, introduciéndose en cada PE una de las diagonales de la matriz de banda, intercalando ciclos de espera a intervalos regulares (cada dos ciclos de reloj), tras los ciclos iniciales de espera y sincronización. Igualmente simple resulta la introducción de las componentes del vector transformado de

valores observados, que se realiza en la entrada X_{in} del primer PE, alternando igualmente ciclos de espera. El cómputo de estas componentes (expresión (29)) se supone realizado en una etapa previa, que, por claridad de la figura, no se ha representado. Este bloque consistirá, básicamente, en una unidad de almacenamiento de las distribuciones de posibilidad observadas (constituida por M bloques –uno por cada uno de los antecedentes– de I palabras –una por cada punto del universo de discurso–) y una unidad de cálculo de mínimos de M entradas. El proceso de generación de las componentes del vector transformado puede realizarse simultáneamente con la evaluación de la matriz, con lo cual se consigue que el tiempo total de computación no se vea alterado. Una atención mayor requiere, por el contrario, el control de flujo de realimentación de los resultados parciales. El lazo de realimentación realiza la función de agregación de los resultados parciales (V. Fig. 7c), proporcionando el sincronismo adecuado para su correcto funcionamiento. Al realizarse el cómputo global de cada componente del vector de valores inferidos en S´ etapas de ω operaciones, los resultados parciales obtenidos durante el cómputo de una etapa deben realimentarse a la red de PEs para su acumulación con los resultados que se obtengan en la siguiente etapa. Realizando consideraciones de carácter temporal, el análisis de la solución indica que el primer resultado parcial se obtiene en el ciclo 2ω-1, instante a partir del cual se van generando, cada dos ciclos de reloj, los siguientes resultados parciales de esta primera etapa, hasta completar el total de ω componentes (que son las que se calculan en cada etapa). Dado que la siguiente etapa no se inicia hasta que han transcurrido ω ciclos (instante 3ω-1), estos resultados parciales deben ser retenidos hasta este instante, en el cual deberán volcarse en la entrada Y_{in} del ω-ésimo PE (que es el que realiza el primer cómputo de cada etapa). El sincronismo necesario se consigue de una manera simple, introduciendo una memoria FIFO de ω palabras en el lazo de realimentación. La primera realimentación tendrá lugar cuando el elemento $\tilde{a}_{\omega+1}$ del vector haya alcanzado la entrada X_{in} del ω-ésimo PE, y, a partir de ese instante, cada dos ciclos de reloj, hasta que se hayan calculado los resultados finales (lo cual sucede en el ciclo 2S-1 y se completa en el 2(S+ω)-3). A partir de aquí, se repite la secuencia de introducción de valores, hasta completar las S´ etapas que componen el cálculo global. La Fig. 8 muestra estas circunstancias para un ejemplo simple en el cual la matriz 2-D equivalente está formada por S=6 columnas y J_n=3 filas, que podría modelar, por ejemplo, una BC constituida por reglas del tipo Doble-Antecedente Único-Consecuente en que se hayan discretizado los universos de discurso sobre los que se definen las variables de estado en 2 y 3 intervalos y en 3 intervalos el universo correspondiente a la variable de control.



$$y_1 = a_1^1 x_1 + a_1^2 x_2 + a_1^3 x_3 + a_1^4 x_4 + a_1^5 x_5 + a_1^6 x_6$$
$$y_2 = a_2^1 x_1 + a_2^2 x_2 + a_2^3 x_3 + a_2^4 x_4 + a_2^5 x_5 + a_2^6 x_6$$
$$y_3 = a_3^1 x_1 + a_3^2 x_2 + a_3^3 x_3 + a_3^4 x_4 + a_3^5 x_5 + a_3^6 x_6$$
$$y_4 = a_4^1 x_1 + a_4^2 x_2 + a_4^3 x_3 + a_4^4 x_4 + a_4^5 x_5 + a_4^6 x_6$$

a)



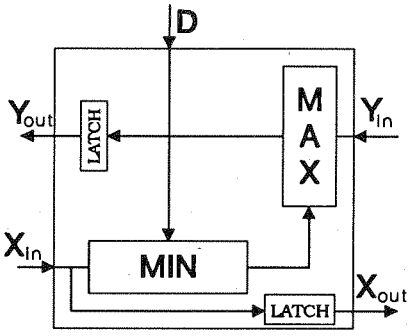
$$y_1^1 = a_1^1 x_1 + a_1^2 x_2$$
$$y_2^1 = a_2^2 x_2 + a_2^3 x_3$$
$$y_1^2 = a_1^3 x_3 + a_1^4 x_4$$
$$y_2^2 = a_2^4 x_4 + a_2^5 x_5$$
$$y_1^3 = a_1^5 x_5 + a_1^6 x_6$$
$$y_2^3 = a_2^6 x_6 + a_2^1 x_1$$
$$y_3^1 = a_3^1 x_1 + a_3^2 x_2$$
$$y_4^1 = a_4^2 x_2 + a_4^3 x_3$$
$$y_3^2 = a_3^3 x_3 + a_3^4 x_4$$
$$y_4^2 = a_4^4 x_4 + a_4^5 x_5$$
$$y_3^3 = a_3^5 x_5 + a_3^6 x_6$$
$$y_4^3 = a_4^6 x_6 + a_4^1 x_1$$

b)

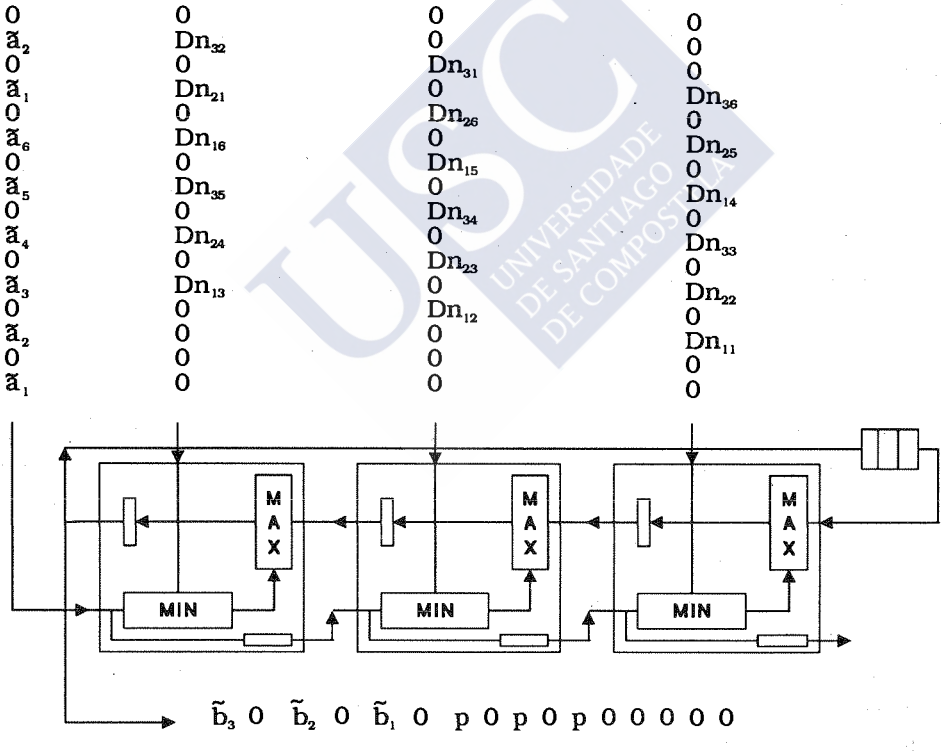
$$y_1 = y_1^1 + y_1^2 + y_1^3$$
$$y_2 = y_2^1 + y_2^2 + y_2^3$$
$$y_3 = y_3^1 + y_3^2 + y_3^3$$
$$y_4 = y_4^1 + y_4^2 + y_4^3$$

c)

Figura 7. Producto Matriz*Vector. a) Expresión clásica. b) Problema transformado y obtención de resultados parciales. c) Composición de resultados parciales.



a)



a)

Figura 8. a) PE básico de nuestra solución [Bugarín y col., 1992]. b) Red de PEs y flujo de datos (ω=3; S=6; T=3).

Un parámetro importante, como indicativo del rendimiento de la red de PEs, es el ciclo de reloj en que se obtiene la última componente del vector de inferencias (tiempo de computación global), puesto que mide la separación temporal mínima necesaria entre dos observaciones sucesivas de las variables de estado. Esto sucederá $2S(T'-1)$ ciclos después de la obtención de la ω-ésima componente de dicho vector. Es decir, en el ciclo

$$\Delta = 2\cdot(S+\omega)-3 + 2S(T'-1) = 2\cdot(ST'+\omega)-3 \qquad (32)$$

Otras soluciones sistólicas orientadas a la evaluación de Sistemas de Control Borrosos, que no presentan particionamiento del problema [Manzoul, 1988; Kung, 1988] (la solución de Kung se plantea para el producto Matriz*Vector, aunque sus resultados son totalmente comparables, al ser ambos procesos estructuralmente idénticos), arrojan el mismo resultado si se verifica, además, que $S'=1$ ó $T'=1$ o, equivalentemente, $S=\omega$ ó $T=\omega$ (recordemos que la solución de Manzoul se propone para sistemas del tipo Único-Antecedente Único-Consecuente, con lo cual la MIG es cuadrada, y, por tanto, $S=T$). Dado que, en general, $S>>T$, la opción $T=\omega$ implicará, manteniendo un mismo tiempo de computación global, la necesidad de un número menor de PEs, por lo cual será preferible frente a las soluciones que no presentan partición, y que necesitan implementar $S+T-1$ PEs. Para que ambos parámetros temporales sean efectivamente idénticos, sería necesario garantizar que las MIG de partida, en uno y otro método, sean exactamente iguales, lo que puede no suceder en algunos casos, dado que el método de particionamiento propuesto puede exigir la adición de columnas de ceros para que sea posible realizar la división de la MIG en submatrices cuadradas. Sin embargo, no será necesario añadir columnas de ceros si $\exists\alpha\in\mathbb{N}$ / $S=\alpha T$, lo cual se verifica en la práctica totalidad de las soluciones implementadas, pues al ser usual considerar grados de discretización uniformes para todos los universos de discurso, tendremos que $S \triangleq \prod_{m=1}^{M} I_m = I^M$, $T \triangleq J_n = I$, con lo que, trivialmente, la condición exigida se satisface.

Pese a todo, en el caso en que esta condición no se verificase, el número de columnas a añadir sería, como máximo, de $T-1$. En una primera aproximación, para la mayoría de los casos en que la BC se pueda estructurar en reglas del tipo Múltiple-Antecedente Único-Consecuente, tendremos $S>>T$, con lo cual

$$S+T-1 \approx S \qquad (33)$$

y la diferencia en cuanto al tiempo de computación global se puede considerar casi nula.

Un inconveniente en la solución propuesta lo constituye la eficiencia de valor 1/2 asociada a los PEs que constituyen la red, impuesta por el propio método de particionamiento. Otra circunstancia negativa, ésta relacionada con el análisis presentado en el trabajo anterior, es que, en determinadas circunstancias (alto grado de discretización, elevado número de antecedentes), el tamaño de la MIG sea excesivamente grande, y la solución poco práctica. En estos casos se hace necesario un planteamiento que se ajuste a la metodología alternativa (regla a regla), incorporando, en la medida de lo posible, técnicas de particionamiento y proyección similares a las descritas aquí. A este trabajo estamos dedicando nuestra atención actualmente.

4. CONCLUSIONES

La tendencia en el diseño de *hardware* borroso ha seguido y, al menos en un futuro inmediato, seguirá la línea del diseño de procesadores *hardware* dedicados a parcelas muy concretas dentro del universo de aplicaciones en las que se demanda tratamiento de información borrosa. Qué duda cabe que el paradigma lo constituyen los procesadores orientados al control borroso, en donde las soluciones actuales tratan de buscar el necesario equilibrio entre eficiencia, flexibilidad de aplicación y adecuación a su integración en silicio. Las ofertas comerciales van desde los circuitos integrados aislados a tarjetas coprocesadoras totalmente compatibles con sistemas de propósito general (PC/AT, Workstation, Macintosh, VME-Bus, ...), pasando por módulos autónomos basados en «chips borrosos» para aplicaciones empotradas. Estas soluciones *hardware* alcanzan prestaciones muy relevantes, explotando al máximo las características particulares del proceso. Resulta difícil, por tanto, conseguir mejoras importantes de velocidad de cálculo, salvo, claro está, a través de mejoras tecnológicas [Morisue y Kogue, 1992].

La gran diversidad de soluciones *hardware* en el limitado dominio del control borroso, surgidas por razones de sencillez, aplicabilidad y rentabilidad, entre otras, deberá abrir paso a soluciones capaces de abarcar dominios más amplios y sistemas mucho más complejos (por ejemplo, los propios de los sistemas expertos). Esto no quiere decir, según la opinión de los autores, que el objetivo esté en diseñar el equivalente «borroso» a un ordenador de propósito general, sino en adecuar ciertos bloques de procesamiento, a modo de coprocesadores dedicados, a la necesidad de tratar con formas más complejas de representar y manipular datos, información y conocimiento borroso.

AGRADECIMIENTOS

Los autores de este trabajo desean agradecer a la Xunta de Galicia el soporte económico aportado a través del proyecto XUGA20601B92.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

REFERENCIAS

American Neuralogix, *NLX 230 Fuzzy MicroController Data Sheet*, 1991.

Bugarín, A., Barro, S., Ruiz, R. y Presedo, J., «Systolic Implementation of DBT-Transformed Fuzzy Control Systems», *Prep. IFAC Symposium SICICA '92*, Benalmádena, Málaga, España, (1992), 135-140.

Eichfeld, H., Löhner, M. y Müller, M., «Architecture of a CMOS Fuzzy Logic Controller with optimized memory organisation and operator design», *Proc. 1st IEEE Conference on Fuzzy Systems*, San Diego, California, EE.UU., (1992), 1317-1323.

Kung, S.Y., *VLSI Array Processors*, Prentice-Hall, Englewood Cliffs, New Jersey, 1988.

Kung, H.T. y Leiserson, C.E., «Systolic Arrays for VLSI», en Duff, I.S. y col. (Eds.), *Proc. Symposium on Sparse Matrices Computations*, (1978), 256-282.

Mamdani, E.H., «Application of Fuzzy Logic to Approximate Reasoning using Linguistic Synthesis», *IEEE Transactions on Computers*, N. 26, (1977), 1182-1191.

Manzoul, M.A., «Fuzzy Inference on a Systolic Array», *Proc. 18th Modelling and Simulation Conference*, (1987), 1103-1108.

Manzoul, M.A. y Jayabharathi, D.,»Fuzzy Controller on FPGA Chip», *Proc. 1st IEEE Interntional Conference on Fuzzy Systems*, San Diego, California, EE.UU., (1992), 1309-1316.

Manzoul, M.A. y Serrate, H.A., «Fuzzy Systolic Arrays», *Proc. 18th International Symposium on Multiple-Valued Logic*, España, (1988).

Manzoul, M.A. y Tayal, S., «Systolic VLSI Array for Multi-Variable Fuzzy Control Systems», *Cybernetics and Systems*, N. 21, (1990), 27-42.

Morisue, M. y Kogue, Y., «A Superconducting Fuzzy Processor», *Proc. 1st IEEE International Conference on Fuzzy Systems*, San Diego, California, EE.UU., (1992), 443-450.

Navarro, J.J., Llabería, J.M. y Valero, M., «Partitioning: A Essential Step in Mapping Algorithms into Systolic Array Processors», *IEEE Computer*, Julio, (1987), 77-89.

Patyra, M.J., «VLSI Implementation of Fuzzy Logic Circuits», *Proc. 4th IFSA Congress*, Bruselas, Bélgica, (1991), 171-174.

Ruiz, R., Barro, S., Bugarín, A. y Presedo, J., «Fuzzy Production Systems on a VLSI Architecture», *Proc. 4th IFSA Congress*, Bruselas, Bélgica, (1991), 175-178.

Ruiz-Mayor, A. y Gutiérrez, J., «Sistema de Codificación de datos fuzzy para su eficaz procesamiento en hardware», *Proc. 1er Congreso Español de Lógica y Tecnologías Fuzzy*, Granada, España, (1991).

Sasaki, M., Inoue, T., Shirai, Y. y Ueno, F., «Fuzzy Multiple-Input Maximum and Minimum Circuits in Current Mode and Their Analyses Using Bounded-Difference Equations», *IEEE Transactions on Computers*, Vol. 6, N. 39, (1990), 768-774.

Sasaki, M., Ishikawa, N., Ueno, F. y Inoue, T., «Current-Mode Analog Fuzzy Hardware with Voltage Input Interface and Normalization Locked Loop», *Proc. 1st IEEE International Conference on Fuzzy Systems*, San Diego, California, EE.UU., (1992), 451-457.

Togai, M. y Watanabe, H., «A VLSI Implementation of Fuzzy Inference Engine: Towards an Expert System on a Chip», *Proc. 2nd IEEE Conference on Artificial Intelligence Applications*, (1985), 192-197.

Togai, M. y Watanabe, H., «Expert System on a Chip: An Engine for Real-Time Approximate Reasoning», *IEEE Expert*, Otoño, (1986), 55-62.

Watanabe, H., Dettloff, W.D. y Yount, K.E., «A VLSI Fuzzy Logic Controller with Reconfigurable, Cascadable Architecture», *IEEE Journal of Solid-State Circuits*, Vol. 2, N. 25, (1990), 376-382.

Yamakawa, T., «Fuzzy Logic Circuits in Current Mode», en Bezdek, J.C. (Ed.), *Analysis of Fuzzy Information*, I, (1987), 241-262.

Yamakawa, T., «High-Speed Fuzzy Controller Hardware System: The Mega FIPS Machine», *Information Sciences*, N. 45, (1988a), 113-128.

Yamakawa, T., «Fuzzy Microprocessors -Rule Chip and Defuzzifier Chip-», *Proc. International Workshop on Fuzzy Sytems Applications*, (1988b), 51-52.

Yamakawa, T., «An Application of a Grade-Controllable Membership Function Circuit to a Singleton-Consequent Fuzzy Logic Controller», *Proc. 3rd IFSA Congress*, (1989a).

Yamakawa, T., «Stabilization of an Inverted Pendulum by a High-speed Fuzzy Logic Controller Hardware System», *Fuzzy Sets and Systems*, N. 32, (1989b), 161-180.

Yamakawa, T. y Kabuo, H., «A Programmable Fuzzifier Integrated Circuit - Synthesis, Design and Fabrication», *Information Sciences*, N. 45, (1988), 75-112.

Yamakawa, T. y Miki, T., «The Current Mode Fuzzy Logic Integrated Circuits Fabricated by the Standard CMOS Process», *IEEE Transactions on Computers*, Vol. 2, N. 35, (1986), 161-167.

Yamakawa, T. y Sasaki, K., «Fuzzy Memory Device», *Prep. 2nd IFSA Congress*, (1987), 551-555.

Zadeh, L.A., «Outline of a New Approach to the Analysis of Complex Systems and Decision Processes», *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 1, N. 3, (1973), 28-44.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

El Proyecto Orbe

Claudio Moraga
Facultad de Informática. Universidad de Dortmund, Alemania

1. INTRODUCCIÓN

El Proyecto de Innovación Tecnológica ORBE (Ordenador Borroso Experimental) tuvo sus comienzos en 1989 por iniciativa del Profesor Dr. Enric Trillas al asumir la Cátedra de Informática e Inteligencia Artificial en la Facultad de Informática de la Universidad Politécnica de Madrid. El proyecto se planteó dos metas principales: desarrollar arquitecturas de ordenadores para dar apoyo eficiente al procesamiento de información borrosa por una parte y por otra, cohesionar a la comunidad de investigadores españoles en el área de Lógica Borrosa proporcionando un desafiante tema para trabajar juntos.

En una primera fase (1989/90) se realizaron estudios básicos de viabilidad del proyecto. Durante esta fase se establecieron contactos personales con el Profesor Takeshi Yamakawa, uno de los pioneros en materia de hardware para el procesamiento de información borrosa. Al término de esta fase se presentó una solicitud de Patrocinio al Ministerio de Industria y Energía, Dirección General de Electrónica y Nuevas Tecnologías, para entrar a la fase experimental. Esta solicitud fue aprobada dentro del Plan Electrónico e Informático Nacional, PEIN II con el número 947/90.

La segunda fase (1991) es la fase experimental, cuyo plan de trabajo fue el siguiente (tomado de la solicitud antes mencionada):

a) Profundizar en la bases teóricas en las que está basado el hardware borroso diseñado por Yamakawa, a fin de comprobar la posibilidad de sustituir sus reglas básicas de funcionamiento por otras más eficientes.

b) Producir nuevos *Chips* borrosos, aprovechando la disponibilidad de tecnologías CMOS de Producción, más avanzadas que las utilizadas por Yamakawa. A su vez investigar sobre nuevos diseños de *Chips* borrosos aprovechando estas mejoras en su producción.

c) Realizar un prototipo de Controlador Borroso y estudiar nuevas aplicaciones.

d) Avanzar en la definición de las especificaciones de un Ordenador Borroso, que deberán definirse desde los distintos puntos de vista: microelectrónica, arquitectura de computadores, sistema operativo, lenguaje y utilidades, requerimiento de los usuarios, prestaciones, etc.

Aceptaron este desafío científicos de los siguientes Centros de Investigación y de Docencia Universitaria: Centro de Estudios Avanzados de Blanes, Centro Nacional de Microelectrónica de Barcelona, Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada, Departamento de Electricidad y Electrónica de la Universidad de Sevilla (CNM Sevilla), Departamento de Ingeniería de Sistemas y Automática de la Universidad de Málaga, Departamento de Ingeniería de Sistemas y de Lenguajes y Sistemas Informáticos de la Universidad de Vigo, Facultad de Informática de la Universidad Politécnica de Madrid e Instituto de Automática Industrial de Madrid (CSIC).

Posteriormente contribuyeron las siguientes Instituciones: Instituto Nacional de Técnica Aeroespacial y Facultad de Informática de la Universidad de Dortmund (Alemania)

2. COLABORACIONES ESPECÍFICAS

a) Bases Teóricas:

Centro de Estudios Avanzados de Blanes (R. López de Mántaras)
Universidad de Granada (M. Delgado, J.L. Verdegay)
Universidad Politécnica de Madrid (E. Trillas)

b) Hardware borroso:

Centro Nacional de Microelectrónica de Barcelona (J. Aguiló)
Universidad Politécnica de Madrid (J. Gutiérrez)
Universidad de Sevilla (J.L. Huertas)

c) Controladores Borrosos:

Instituto de Automática Industrial de Madrid (R. García)
Instituto Nacional de Técnica Aeroespacial (J. Gutiérrez)
Universidades de Málaga y Vigo (A. Ollero)

d) Especificación básica de un Ordenador Borroso:

Representantes de todos los Centros participantes.

3. ALTERNATIVAS DE ARQUITECTURA PARA EL ORBE

Desde el punto de vista de eficiencia la opción aparentemente más adecuada habría sido el diseño de un Procesador de tipo RISC *(Reduced Instruction Set Computer)* convencional, con instrucciones adicionales igualmente soportadas por el hardware del procesador, para realizar operaciones de inferencia en Lógica Borrosa. Esto, sin embargo, habría requerido un plazo de tiempo muy superior al año considerado para la segunda fase del proyecto ORBE. En este contexto, las alternativas más claras son el utilizar un «anfitrión» clásico dotado de una «extensión borrosa» y un entorno de software apropiado. Esto puede ser realizado muy simplemente a nivel de Unidad Central de Proceso en analogía al uso de Co-Procesadores numéricos (ver Figura 1) o bien a nivel de Ordenador, en cuyo caso la extensión borrosa incluye, además de circuitos de inferencia borrosa, una memoria local, una unidad de control y posiblemente un microcomputador o unidad aritmética para la conversión «borroso → nítido» necesaria en procesos de control (ver Figura 2).

Basados en los resultados positivos obtenidos con chips prototipo del Profesor Yamakawa, que se caracterizan por un procesamiento analógico, se optó por esta modalidad. El procesamiento analógico de información borrosa es muy rápido y ofrece una alta funcionalidad por unidad de área de silicio.

Los prototipos del Profesor Yamakawa fueron realizados en tecnología CMOS de 10 µm. Los investigadores de los Centros Nacionales de Microelectrónica de Barcelona y de Sevilla hicieron un rediseño utilizando tecnología ES2 (*European Silicon Structures)* de 1,5 µm, 2 capas de metal y 2 de polisilicio. Ésta es una tarea de una complejida muy superior a un mero cambio de escala a nivel geométrico. En el rango de dimensiones de la tecnología ES2 aumenta notablemente la sensibilidad de los parámetros y las tolerancias se hacen extremadamente críticas. Los investigadores de los centros antes mencionados estudiaron exhaustivamente estos aspectos y lograron construir chips con circuitos estables y robustos. En este mismo período de tiempo, investigadores del Departamento de Tecnología Fotónica de la Universidad Politécnica de Madrid diseñaron un circuito de Función de Pertenencia *programable en tiempo de ejecución* sobre la base de tecnología ES2 de 2 µm. Todos estos resultados demuestran que un Procesador borroso analógico con control digital es perfectamente realizable y que puede constituir la base de las arquitecturas presentadas en el párrafo previo.

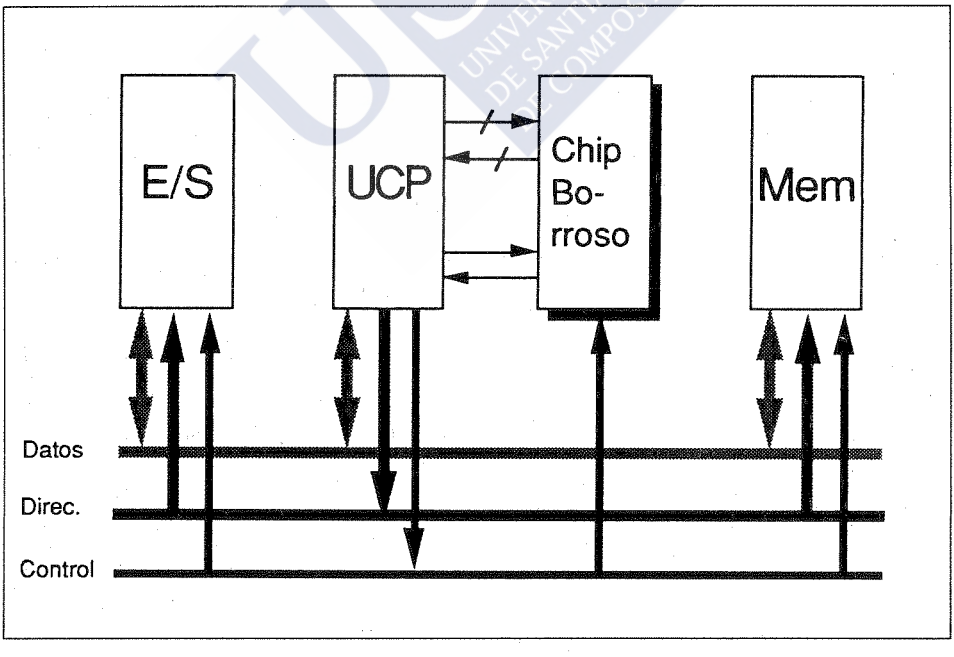


Figura 1. Arquitectura bajo el esquema de coprocesador.

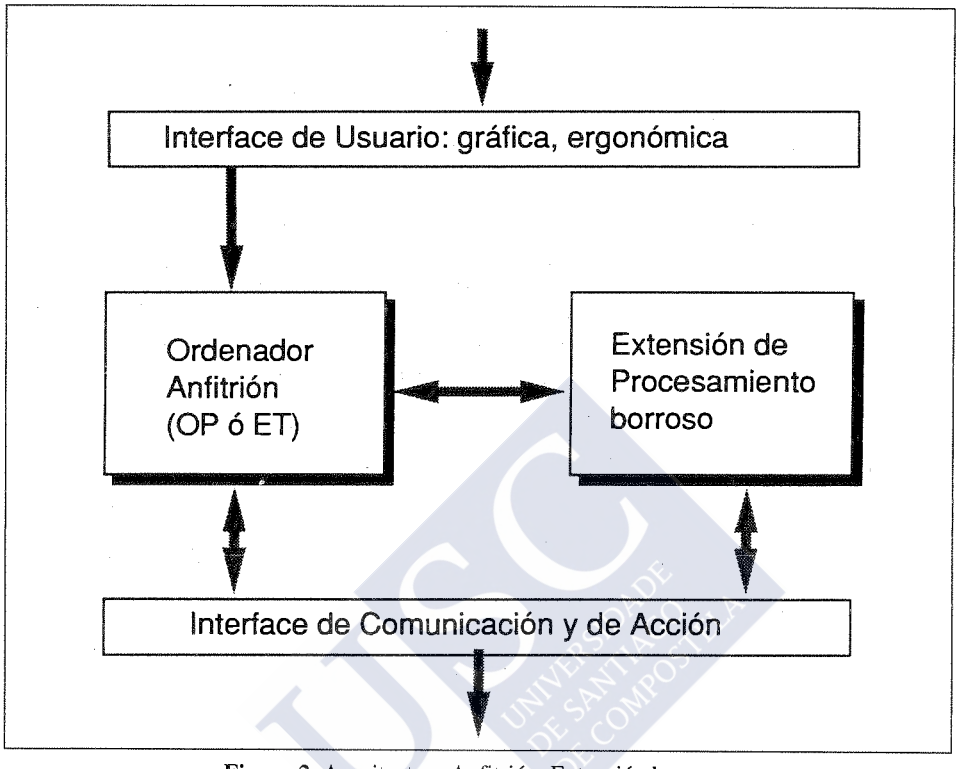


Figura 2. Arquitectura Anfitrión. Extensión borrosa.

4. CONCLUSIONES

ORBE es una alternativa real para arquitectura de sistemas computacionales con capacidad de procesamiento de información borrosa.

El procesamiento analógico es inherentemente rápido, lo cual abre interesantes perspectivas para sistemas borrosos de control con restricciones de tiempo real.

Por último, pero no por ello menos importante: ¡ El Proyecto ORBE ha contribuido a aglutinar un grupo de investigadores en Lógica Borrosa cuyo potencial no tiene igual en Europa!

REFERENCIAS

Gupta, N.M. y Yamakawa, T., *Fuzzy Computing: Theory, Hardware and Applications*, North Holland, 1988.

Felix, R., *Teoría de Decisión con metas cualitativas, basada en la Teoría de Conjuntos Borrosos*, Tesis Doctoral (en alemán), Facultad de Informática, Universidad de Dortmund, 1991.

Martín Romero, J.R., Sánchez Miró, L.F. y Vicente Cuenca, P.D., «Evaluación de un sistema guiado con lógica Difusa», *Actas 1er Congreso español sobre tecnologías y lógica fuzzy*, Granada, (1991).

Ruiz, A. y Gutiérrez, J., «Sistema de codificación de datos fuzzy para su eficaz procesamiento en hardware», *Actas 1er Congreso español sobre tecnologías y lógica fuzzy*, Granada, 1991.

Ollero, A. y García Cerezo, A., «Direct digital control, autotuning and supervision by using fuzzy controllers», *Fuzzy Sets and Systems*, N. 30, (1989), 135-153.

ORBE, Memoria, Relación de Actividades de los Centros Participantes, Noviembre, 1991.

Togai, M. y Watanabe, H., «A VLSI implementation of Fuzzy Inference Engine: toward an Expert System on a chip», *Information Sciences*, N. 38, (1986).

Trillas, E., Delgado, M., Verdegay, J.L. y Vila, M.A., «A review of some aspects on designing fuzzy controllers», *AGARD Conference Proceedings*, 31-1, 31-12, Francia, (1991).

Yamakawa, T., «A simple fuzzy computer hardware system employing min & max operations. A challenge to the 6th Generation Computer», *Proc. 2nd. IFSA Congress*, Tokyo, (1987).

TRABAJOS ESPECIALIZADOS EN LÓGICA BORROSA

CAP. XIV Una aproximación T-Norma a la lógica bungeana de medios y fines
¿Hacia una metodología heurística en Ética?

CAP. XV On some bisymmetric functions on closed intervals

CAP. XVI On some measures of entropy related to aggregation operators

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

UNA APROXIMACIÓN USANDO T-NORMAS A LA LÓGICA BUNGEANA DE MEDIOS Y FINES. ¿HACIA UNA METODOLOGÍA HEURÍSTICA EN ÉTICA?

Enric Trillas
Instituto Nacional de Técnica Aeroespacial. Madrid

«...the way in which de fact/value distinction is hopelessly fuzzy in the real world and in the real language».
H.Putnan, *Reason, Truth and History*, Cambridge University Press, 1981.

«The altruistic values that we acquire by social conditioning and perhaps by heredity are vague and open-ended».
W.V.O.Quine, *Theories and Things*, Harvard University Press, 1981.

1. INTRODUCCIÓN

Hay muchas situaciones, especialmente en ética y en las ciencias sociales, pero también en la historia de la ciencia y la epistemología, en las cuales se buscan ciertos fines a través de determinados medios. En estos procesos aparecen frecuentemente efectos secundarios, de tal forma que llega a ser importante el considerar su existencia cuando se necesita realizar alguna evaluación.

Si representamos por M los medios materiales que tenemos, por E los fines que queremos obtener, y por S las finalidades secundarias, «resultados» o efectos que podemos esperar, podemos resumir esa situación por:

«Si M, entonces E y S» (1)

y describirla en «alguna lógica» como

$$M \rightarrow E \wedge S \qquad (1')$$

Obsérvese que si esta lógica es del tipo del cálculo proposicional bivalorado clásico, sería:

$$M \rightarrow E \wedge S = \neg M \vee (E \wedge S) = (\neg M \vee E) \wedge (\neg M \vee S) = (M \rightarrow E) \wedge (M \rightarrow S) \qquad (2)$$

de modo que (1) es equivalente a

«Si M entonces E y si M entonces S»,

lo cual no resulta demasiado conveniente dado que fácilmente las finalidades secundarias S, podrían no referirse únicamente a M, sino especialmente a E.

Si consideramos otra clase de lógica, como, por ejemplo, la Lógica de la Mecánica Cuántica, entonces el operador $x \rightarrow y$ se expresa como $\neg x \vee (x \wedge y)$, que resulta equivalente a la anterior expresión $\neg x \vee y$ en el caso del álgebra de Boole. Bajo este tipo de lógica, la fórmula (1') es:

$$M \rightarrow E \wedge S = \neg M \vee (M \wedge (E \wedge S)) = \neg M \vee (M \wedge E \wedge S) \qquad (3)$$

siendo $\wedge$ asociativo. Si es también conmutativo y $M \wedge M = M$, entonces

$$M \rightarrow E \wedge S = \neg M \vee ((M \wedge E) \wedge (M \wedge S)).$$

En el caso de que $\vee$ y $\wedge$ fueran, además, distributivas, tendríamos de nuevo:

$$M \rightarrow E \wedge S = (\neg M \vee (M \wedge E)) \wedge (\neg M \vee (M \wedge S)) = (M \rightarrow E) \wedge (M \rightarrow S).$$

Obsérvese que si M, E y S pueden ser identificados con conjuntos borrosos (dotados con las conectivas Max, Min y 1-j), o simplemente si el conjunto que contiene a M, E y S es una álgebra de DeMorgan, entonces ambas definiciones, $M \underset{1}{\rightarrow} E \wedge S = \neg M \vee (E \wedge S)$, y $M \underset{2}{\rightarrow} E \wedge S = \neg M \vee (M \wedge (E \wedge S))$, son posibles y no equivalentes; como la segunda es $(\neg M \vee M) \wedge (\neg M \vee (E \wedge S))$, coincidirá con la primera si $\neg M \vee M = 1$, esto es, si M es un elemento booleano. Denotando por

$$\upsilon(M) = M \wedge \neg M$$

la vaguedad de M, tenemos

$$M \underset{2}{\rightarrow} E \wedge S \leq \neg \upsilon(M) \wedge (M \underset{1}{\rightarrow} E \wedge S)$$

y, en consecuencia,

$$M \underset{2}{\rightarrow} E \wedge S \leq M \underset{1}{\rightarrow} E \wedge S.$$

Antes de continuar con nuestro discurso puede resultar interesante resaltar que (1) podría ser también interpretado como

«Si M, entonces E y S o E» (1'')

o, en notación lógica:

$$M \rightarrow (E \wedge S) \vee E$$

Pero entonces, simplemente con tener $E \wedge S \leq E$, como es usual, obtendríamos

$$M \rightarrow (E \wedge S) \vee E = M \rightarrow E,$$

lo cual significa que bajo la formulación (1'') y en la hipótesis $x \wedge y \leq x$ (¡una hipótesis bastante débil!), los fines secundarios desaparecerían.

Parece, en consecuencia, que en tal clase de situación es no sólo importante el conocimiento de los medios materiales de los que disponemos para obtener E y S, sino también el conocimiento de la descripción lógica que somos capaces de hacer para construir una representación de (1).

Siguiendo a Bunge [1976] hablaríamos de «medios cognitivos» para comprender la forma de la expresión $M \rightarrow E \wedge S$. En ese libro, Bunge trata de modelizar $M \rightarrow E \wedge S$ usando una función V, definida sobre el conjunto que contiene a M, E, S y sus combinaciones lógicas, y tomando sus valores en R, tal que:

$$V(x \vee y) = V(x) + V(y), \text{ si } x \neq y,$$

$$V(\neg x) = -V(x),$$

entendiendo $\neg x$ como la acción opuesta a la acción x. Después de algunos cálculos se llega a la ley $V(M \rightarrow E \wedge S) + V(M) = V(E) + (S)$. El objetivo de este trabajo es el estudio de esta ley en el contexto de una Lógica multivalorada con valores en [0,1].

Para ello supongamos que *A* es el conjunto en el cual podemos «representar» los objetos M, E y S, y que contiene también todas las combinaciones lógicas entre ellos: $M \vee E$, $M \wedge E$, $\neg M$, $M \rightarrow E$, etc. Supondremos que *A* posee algún tipo de estructura lógica que permite traducir todo lo que necesitemos decir sobre M o E, E y S, etc. Quizás algunas de estas operaciones den como resultado 0 ó 1, dependiendo de la interpretación. Por ejemplo, podemos decidir $E \vee S = 1$ ó $\neg E \wedge \neg S = 0$ para reflejar que esperamos únicamente E o S. Por supuesto, el «álgebra» *A* se supone referida a alguna situación real y cambiaría si esta situación cambiase.

Sea, entonces, $M \rightarrow E \wedge S \in A$.

2. LOS CASOS PROBABILÍSTICOS Y DE SUGENO

Supongamos que se define una probabilidad p en A, de modo que p(M), p(E) y p(S) reflejan algo acerca de M, E y S relativo a la situación en que nos encontramos. En tal caso, para que podamos aplicar el cálculo clásico de probabilidades, A necesitará ser un álgebra de Boole.

$p(M\rightarrow E\wedge S)$ es la probabilidad de que, con los medios M, obtengamos los fines E y S. Esto es,

$$c=p(M \rightarrow E\wedge S) = p(\neg M\vee(E\wedge S)) = p(\neg M)+p(E\wedge S),$$

siendo $\neg M\wedge(E\wedge S)=0$, es decir, $E\wedge S \subset M$. Entonces,

$$c=1-pM+p(\neg(\neg E\vee\neg S)) = 2-pM-p(\neg E\vee\neg S),$$

y, si $\neg E\wedge\neg S=0$, esto es si $E\vee S=1$, entonces

$$c = 2\text{-}pM\text{-}p(\neg E)-p(\neg S) = pE+pS-pM,$$

lo cual, escribiendo e=pE, s=pS y m=pM es la ley

$$m+c = s+e$$

los medios son iguales a los fines, supuesto $E\vee S=1$ y $E\wedge S \subset M$. Si llamamos a s+e la utilidad y a m+c el costo, la ley se leería utilidad=costo. Llamaremos a esta ley la ley de Bentham [Ayer, 1979].

Si podemos separar claramente lo que son los fines principales de los secundarios, entonces $E\wedge S=0$ y, por supuesto, $E\wedge S \subset M$. De hecho, esta parece la única posibilidad satisfactoria siendo M los medios materiales. En este caso, E y S son sucesos complementarios y la ley de Bentham queda en la forma «pobre»: m+c=e+s=1, puesto que pE+pS=e+s debe ser 1.

Podría suceder que las cosas no fueran tan claramente definidas como supusimos en el último caso, o que no tuviéramos ninguna probabilidad definida en A. Si «las cosas son lo que son», puede que los valores m, c, s y e no sean obtenibles como probabilidades. Pueden ser únicamente evaluaciones, más o menos subjetivas, hechas por algún observador o agente. Un buen método de evitar, hasta cierto punto, esta clase de problemas, consiste en usar lo que se conoce como medidas de Sugeno [1974], es decir, usando, en lugar de una probabilidad p, funciones

$$v:A \rightarrow[0,1]$$

tales que

$$v(A \vee B)=v(A)+v(B)+\lambda v(A)v(B), \text{ si } A\wedge B=0 \text{ y } -1<\lambda$$

y

$$v(\neg A)=1-v(A)$$

Las operaciones asociativas $G_\lambda(x,y)=x+y+\lambda xy$, $-1\leq\lambda$, tales que $G_\lambda(a,0)=a$ y $G_\lambda(x,y)=G(y,x)$ son t-conormas en [0,1] únicamente si $\lambda=-1$, en cuyo caso G_{-1} es la t-conorma dual de la t-norma $\Pi(x,y)=x\cdot y$ (el producto), t-norma que es continua pero no arquimediana [Schweizer y Sklar, 1983]. Cuando $\lambda=0$, G_0 es la adición y tenemos una probabilidad.

Es estrictamente necesario remarcar que *A* no necesita ser un álgebra de Boole. De hecho, cuando $B=\neg A$, tenemos $A\wedge\neg A=0$ y

$$v(A\vee\neg A)=v(1)=v(A)+1-v(A)+\lambda v(A)(1-v(A))=1+\lambda v(A)(1-v(A)),$$

lo cual conduce a

$$0=\lambda\, v(A)(1-v(A)).$$

Si $\lambda\neq 0$, entonces necesariamente $v(A)=0$ ó $v(A)=1$. Tenemos una medida bivalorada y la vaguedad desaparece.

Es fácil probar que $v(1)=1$ a partir de $v(0)=0$, puesto que $v(1)=v(\neg 0)=1-v(0)$. Como $v(0)=v(0\vee 0)=v(0)+v(0)+\lambda v(0)v(0)$, $-v(0)=\lambda v(0).v(0)$. Si $v(0)=0$, la expresión se verifica, y si $v(0)\neq 0$, $-1=\lambda v(0)$ y $\lambda\in(-1,0)$. Entonces $1=|\lambda|v(0)$ y $v(0)=1/|\lambda|$. Pero $|\lambda|<1$ implica $1<1/|\lambda|$, lo cual es absurdo. Entonces $v(0)=0$. En este caso de las medidas de Sugeno, suponiendo que *A* sea un álgebra de DeMorgan, y, siguiendo los mismos pasos que en el caso probabilístico, después de una sencilla y tediosa computación tenemos:

$$c=-\lambda(2+\lambda)+m(\lambda^2+\lambda-1)+(e+s)(1+\lambda)^2-(me+es+sm)(1+\lambda)\lambda-mes\lambda^2,$$

con la hipótesis de que $\neg M\wedge(E\wedge S)=0$ y $E\vee S=1$. Esta fórmula se reduce al caso probabilístico si $\lambda=0$.

Si m=1 (todos los medios materiales que necesitamos se encuentran disponibles) y s=0 (no aparecerán efectos secundarios), la ley de Bentham nos ofrecería la solución 1+c=e+0, o e=1 y c=0: ¡no necesitamos usar medios cognitivos y tendríamos todos los fines principales que buscábamos!. De hecho

esta podría ser una solución muy satisfactoria. Pero generalmente las cosas no son tan claras, como todo el mundo sabe. Con la nueva ley, si m=1 y s=0, tenemos

$$c=(1+\lambda)((1+2\lambda)e-1)$$

lo que nos permite obtener algunas otras soluciones. Por ejemplo con $\lambda=1/2$ tenemos

$$c=3e-3/2,\ 0\leq e\leq 1/2,$$

como el conjunto de posibles soluciones. En cualquier caso podríamos no tener ninguna solución si exigiéramos también que e=1.

En el interesante caso en que m=c=1 (todos los medios están disponibles) la fórmula quedaría

$$2+\lambda=e(1+\lambda)+s(1+\lambda)-es\lambda,$$

y si s=0, entonces

$$2+\lambda=e(1+\lambda),\ \text{ó}\ e=1+1/1+\lambda,$$

que, como $0<1+\lambda$, nos da e>1: es imposible disponer de todos los medios y no obtener ningún efecto secundario. En el caso de Bentham es 1+1=e+0, y tampoco obtendríamos solución. Históricamente este podría ser un resultado interesante.

NOTA: Considerando $\nu(\neg A)=\dfrac{1-\nu(A)}{1+\lambda\nu(A)}$, el complemento de Sugeno, se obtiene,

$$(c+m+\lambda\, cm^{-1})(\lambda^2+2\lambda+1-\lambda(e+s+\lambda es))=(1+\lambda)^2(e+s+\lambda es-1)$$

Esta fórmula se reduce al caso probabilístico si $\lambda=0$, y, en el caso de que m=c=1 y s=0, quedaría

$$e = 1.25 + \frac{\lambda}{2} + \frac{3}{4\,(2\lambda+1)},\ \text{dando}\ e>1$$

3. ALGUNAS OBSERVACIONES SOBRE LAS ANTERIORES DEFINICIONES

La extraña condición $E\wedge S\subset M$ que deriva de la anterior $\neg M\wedge(E\wedge S)=0$ es realmente problemática y mucho más si se toma conjuntamente con $E\vee S=1$. Ambas son consecuencia de tomar el valor de la unión solo en el caso de que sus componentes sean disjuntos. Pero esto, usual en la Teoría de la Medida, no lo es en la Lógica Multivalorada, en donde, en caso de que $v(A\vee B)=S(v(A),v(B))$ actúe como una definición de $A\vee B$, lo único que se necesita es que A y B sean diferentes porque, si por ejemplo, $S(x,y)=x+y-xy$, entonces $v(A)=v(A\vee A)=S(v(A),v(A))=2v(A)-v(A)^2$, que nos quedaría $0=v(A)\cdot(1-v(A))$ y $v(A)=0$ ó $v(A)=1$ como únicos posibles valores para v. También si $S(x,y)=\min(1,x+y)$, entonces sería $v(A)=\min(1,2v(A))$, dando $v(A)=1$ ó $v(A)=2v(A)$, esto es $v(A)=1$ ó $v(A)=0$

En general, si S es la t-conorma dual de una t-norma T, continua y arquimediana, con generador aditivo t, tendríamos:

$$v(A)=S(v(A),v(A))=1-T(1-v(A),1-v(A))$$

ó

$$1-v(A)=t^{(-1)}(2.t(1-v(A))).$$

Entonces, si $2t(1-v(A))$ no está en $[0,1]$, entonces $1-v(A)=0$ y $v(A)=1$, y si $2t(1-v(A))$ está en $[0,1]$ tendríamos $t(1-v(A))=2t(1-v(A))$ y $t(1-v(A))=0$, de donde $v(A)=1$.

Retomando por un momento nuestra condición de que «$\neg a\wedge b=0$ si $b\leq a$», que se verifica en cualquier álgebra de Boole, es significativo observar que en un álgebra de DeMorgan no es verdad que «$b\leq a$ implique $\neg a\wedge b=0$» como es fácil de ver en el caso usual de los Conjuntos Borrosos. En un álgebra de Boole, de $1=a\vee\neg a$ se sigue que $b=a\wedge b\vee\neg a\wedge b$, y entonces $\neg a\wedge b=0$ implica que $b=a\wedge b$ o $b\leq a$. A la inversa, si $b\leq a$ tendríamos $\neg a\wedge b\leq\neg a\wedge a=0$ ó $\neg a\wedge b=0$. Pero estos argumentos no son reproducibles en una álgebra de DeMorgan pues dependen de las leyes básicas del álgebra de Boole $a\wedge\neg a=0$ y $a\vee\neg a=1$.

¿En qué álgebras de De Morgan se cumple «$\neg a\wedge b=0$ implica $b\leq a$»? No tengo respuesta para esta cuestión. En el caso de que $\underset{\sim}{P}(X)$, con la lógica usual de Zadeh, dada por la t-norma $T(x,y)=M(x,y)=\min(x,y)$, tendríamos que $\neg A\wedge B=0$ sería equivalente a $\min(1-A(x),B(x))=0$ para cualquier $x\in X$. Lo cual significa que $A(x)=1$ ó $B(x)=0$ para cualquier $x\in X$, implicando que $B(x)\leq A(x)$, para cualquier $x\in X$, y, por tanto $B\subset A$.

En el caso de los Conjuntos Borrosos, esto sucede también con otras estructuras que no son distributivas y *a fortiori* no son álgebras de DeMorgan. Por ejemplo, si $T(x,y)=\Pi(x,y)=x.y$ tendríamos $(1-A(x))\cdot B(x)=0$ y se seguiría la misma conclusión. Con la lógica de Łukasiewicz, $T(x,y)=Ł(x,y)=Max(0,x+y-1)$ y entonces tenemos: $0=Max(0,1-A(x)+B(x)-1)=Max(0,B(x)-A(x))$, y entonces $B(x)-A(x)\leq 0$ o $B(x)\leq A(x)$ para cualquier $x\in X$.

Podría ser diferente con alguna otra t-norma. Por ejemplo con[1]

$$Z(x,y) = \begin{cases} x, \text{ si } y = 1 \\ y, \text{ si } x = 1 \\ 0, \text{ e otro caso} \end{cases}$$

es suficiente tomar $A(x)=0.25$ y $B(x)=0.75$, ambos constantes, para obtener $(\neg A\wedge B)(x)=Z(1-0.25,0.75)=0$; esto es, $\neg A\wedge B=0$ pero $A \subset B$.

Todo esto sugiere el uso de una t-conorma sobre el conjunto *A* que contiene a M, S, E... parar sobreimponer una nueva lógica que refleje el proceso de evaluación. Esto es, dada una t-norma T, sea $v:A\rightarrow[0,1]$ una aplicación tal que:

$$v(\neg A)=1-vA$$

$$v(A\vee B)=1-T(1-vA,1-vB), \text{ si } A\neq B.$$

Con ello estamos sobreimponiendo una lógica multivalorada al conjunto de objetos con los que estamos tratando.

Como no usaremos más la condición $A\wedge B=0$ para expresar $v(A\vee B)$ no tendremos el problema con $E\wedge S\wedge\neg M$. Pero, con esta lógica multivalorada, ¿es posible mantener la estructura de álgebra de Boole en *A*? Depende por supuesto de T. Por ejemplo, si T=M, nuestra antigua definición quedaría

$$v(A\vee B)=Max(vA,vB), \text{ si } A\wedge B=0,$$

y tomando $v(1)=1$ y $B=\neg A$, obtenemos

$$1=Max(vA,1-vA) \text{ o } (vA=0 \text{ o } vA=1).$$

[1] Para cualquier t-norma T se cumple $Z\leq T\leq M$. En particular $Z\leq L\leq \Pi\leq M$.

También si T=Π, tendremos 1=1-(1-vA).vA, y vA(1-vA)=0. Pero si T=Ł, entonces 1=Min(1,1-vA+vA)=Min(1,1) es siempre verdadero, y **bajo las «medidas de Łukasiewicz» podemos mantener la estructura de álgebra de Boole en *A***.

Este último resultado es el mismo si T=Z, porque Z(1-a,a)=0 para cualquier a en [0,1]. En general tenemos la fórmula

$$1=1-T(1-a,a), \text{ ó } 0=T(1-a,a)$$

para a=vA∈[0,1]. Dado que T es continua (y arquimediana) con generador aditivo t, la última fórmula sería $t^{(-1)}(t(1-a)+t(a))=0$, que se mantiene si, por ejemplo, t es tal que para cualquier a, t(1-a)+t(a)∈[0,1], en cuyo caso: 1=t(a)+t(1-a), a∈[0,1], con t:[0,1] → [0,1] estrictamente decreciente y con t(1)=0, que implica t(0)=1. Si a=1/2 la ecuación muestra t(1/2)=1/2. La figura 1 representa todas las soluciones de dicha ecuación:

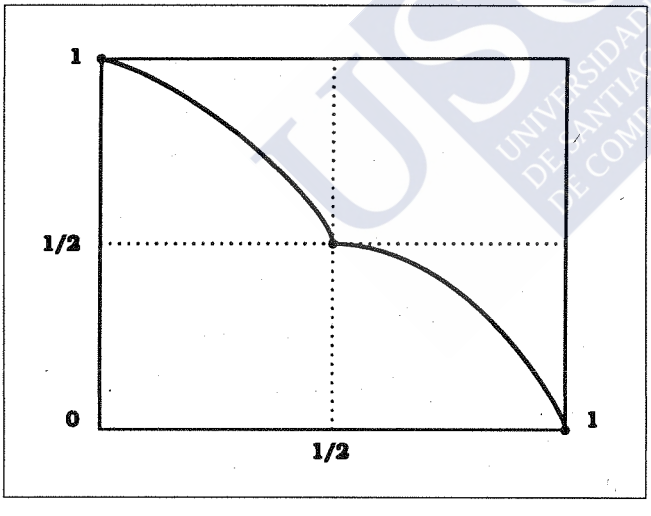


Figura 1

Este gráfico muestra que la solución continua general es obtenida [Trillas, 1979] a través de una función arbitraria w:[0,1/2] → [1/2,1], estrictamente decreciente, y para la cual w(0)=1, w(1/2)=1/2, dando

$$t(a) = \begin{cases} w(a), \text{ si } a\in[0, \frac{1}{2}] \\ 1 - w(1-a), \text{ si } a\in[\frac{1}{2},1] \end{cases}$$

Con tal familia de funciones t, las t-normas

$$T(a,b)=t^{(-1)}(t(a)+t(b)),$$

continuas y arquimedianas, preservan la estructura de álgebra de Boole.

4. LOS CASOS DE T-NORMAS

La fórmula $v(A\vee B)=1-T(1-vA,1-vB)$ si $A\neq B$, quedará:

$v(M\rightarrow E\wedge S)=v(\neg M\vee(E\wedge S))=v(\neg M\vee\neg(\neg E\vee\neg S))=1-T(1-v\neg M,1-v\neg(\neg E\vee\neg S)=$
$=1-T(vM,v(\neg E\vee\neg S)),$

en la hipótesis de que $\neg M\neq\neg E\vee\neg S$ (ó $M\neq E\wedge S$). Si, además, $E\neq S$ entonces:

$$v(M\rightarrow E\wedge S)=1-T(vM,1-T(1-v\neg E,1-v\neg S)=1-T(vM,1-T(vE,vS)),$$

ó

$$c=1-T(m,1-T(e,s))$$

que es la Ley General de Bentham (LGB).

Si **T=Z**, entonces la LGB es $c=1-Z(m,1-Z(e,s))$, que da la realmente pobre solución

$$c=\begin{cases} s, & \text{si } m=e=1 \\ e, & \text{si } m=s=1 \\ 1-m, & \text{en otro caso} \end{cases}$$

Si **T=Ł**, la LGB es (después de una sencilla computación)

$$c=Min(1,1-m+Max(0,e+s-1)),$$

y si $e+s\leq 1$ entonces $c=Min(1,1-m)=1-m$. Si $e+s>1$, $c=Min(1,e+s-m)$, lo cual, si $e+s-m\leq 1$ dará $c=e+s-m$ y si $e+s-m>1$ dará $c=1$. Todos los casos pueden ser resumidos en la solución

$$c=\begin{cases} 1-m & \text{si } e+s\leq 1 \\ e+s-m, & \text{si } 1<e+s\leq 1+m \\ 1, & \text{si } 1+m<e+s \end{cases}$$

Entonces, bajo la lógica de Łukasiewicz, tenemos la ley de Bentham solamente en el caso que el valor del fin esté en el intervalo (1,1+m].

Si **T=∏**, la LGB es

$$c=1-m(1-es),$$

que en el caso de que m·s≠0 nos da

$$e = \frac{1}{s}\,(1 - \frac{1-c}{m}).$$

Si m=0, entonces c=1. Si s=0 entonces c=1-m. Si m=1, entonces c=es. Si s=1, entonces c=1-m(1-e), y

$$e = 1 - \frac{1-c}{m}\ ,\ \text{si } m\neq 0$$

Obsérvese que m=c=1 da es=1, es decir, e=s=1.

Si **T=M**, la LGB es c=1-Max(m,1-Max(e,s))=Min(1-m,Max(e,s)). Entonces, si e≤s, será c=Min(1-m,s), que si 1-m≤s daría c=1-m, y si 1-m>s daría c=s. Todos los casos pueden ser resumidos en la solución:

$$c = \begin{cases} e & \text{si } s<e<1-m \\ s, & \text{si } e\leq s<1-m \\ 1-m & \text{en otro caso} \end{cases}$$

Nótese que estamos trabajando con la definición $x\underset{2}{\rightarrow}y=\neg x\vee y$ y no con la $x\underset{2}{\rightarrow}\neg x\vee(xy)=\neg\upsilon(x)(x\underset{1}{\rightarrow}y)$.

5. COMENTARIO ACERCA DE LOS VALORES DE LOS FINES PRINCIPALES

A través de la fórmula LGB tenemos, en cualquier caso particular, la cantidad de medios cognoscitivos que necesitamos para alcanzar algunos fines principales, algunos efectos secundarios y también algunos medios materiales. Pero frecuentemente estamos interesados en saber los fines principales que podríamos obtener comenzando con determinados medios (materiales y cognitivos) y

esperando ciertos efectos secundarios. Desafortunadamente, las t-normas dan semigrupos pero no grupos, y no es posible, en general, aislar e en una ecuación-LGB.

En cualquier caso, podemos lograr algún resultado si T es continua y arquimediana, en cuyo caso podríamos suponer la existencia de algún generador aditivo $t:[0,1] \rightarrow [0,1]$ tal que $t(0)=1$, y además $t(1)=0$. En otro caso, siendo $t(0)<+\infty$ tomaríamos $t/t(0)$. En esta situación, con algunas restricciones en el rango de los valores m, s y c, tendríamos aislado e. De hecho, como la LGB es

$$T(m,1-T(e,s))=1-c \in [0,1]$$

tenemos

$$t(m)+t(1-T(e,s))=t(1-c) \in [0,1].$$

Entonces si:

$$1 \leq m+c \quad (a)$$

o, equivalentemente:

$$1-c \leq m$$

y

$$t(1-c) \geq t(m)$$

tendremos

$$t(1-T(e,s))=t(1-c)-t(m) \in [0,1].$$

ó

$$1-T(e,s)=t^{(-1)}(t(1-c)-t(m)) \in [0,1].$$

y

$$T(e,s)=1-t^{(-1)}(t(1-c)-t(m)) \in [0,1].$$

de donde

$$t(e)+t(s)=t(1-t^{(-1)}(t(1-c)-t(m))) \in [0,1].$$

Entonces, si

$$t(s) \leq t(1-t^{(-1)}(t(1-c)-t(m))) \quad (b)$$

se sigue

$$t(e)=t(1-t^{(-1)}(t(1-c)-t(m)))-t(s) \in [0,1],$$

y finalmente:

$$e=t^{(-1)}(t(1-t^{(-1)}(t(1-c)-t(m)))-t(s)),$$

cuyos valores están en [0,1] si las condiciones (a) y (b) se mantienen.

Por ejemplo, si T=Ł, el generador t(a)=1-a nos permite usar la última fórmula bajo las restricciones

(a) $1 \leq m+c$

(b) $1-s \leq 1-(1-(1-c)-1+m)=2-c-m$, ó $m+c \leq 1+s$.

Entonces, si $1 \leq m+c \leq 1+s$, tenemos

$$e=1-(t(1-t^{(-1)}(c-1+m))-1+s=1-t(1-t^{(-1)}(c-1+m))-1+s=$$

$$=2-s-t(1-1+c-1+m)=2-s-t(c-1+m)=2-s-1+(c-1+m)=-s+c+m.$$

En otras palabras,

«Bajo T=Ł, si $1 \leq m+c \leq 1+s$ entonces $e+s=m+c$»,

Este resultado completa el obtenido antes y expresa que la ley de Bentham también se cumple si el coste m+c se encuentra en el intervalo esperado [1,1+s] de los efectos secundarios.

6. OBSERVACIONES CRÍTICAS SOBRE LA LEY GENERAL DE BENTHAM

Las observaciones que vienen a continuación tratan de ser críticas en algunos aspectos, pero relacionadas con la metodología que estamos tratando de introducir.

6.1.

Cuando s=0, LGB se transforma en $c=1-m=\neg m$, para cualquier valor de e.

Cuando s=1, LGB se transforma en $c=1-T(m,1-e)=m \rightarrow e$.

Cuando m=0, LGB se transforma en c=1, $\forall e$ y $\forall s$.

Cuando m=1, LGB se transforma en $c=T(e,s)=e \wedge s$.

Cuando e=0, LGB se transforma en $c=1-m=\neg m$, $\forall s$.

Cuando e=1, LGB se transforma en $c=s \rightarrow m$.

6.2.

Cuando m=c=s=1/2, LGB se transforma en 1/2=T(1/2,1-T(e,1/2)). Si T=Z y e=1, la ecuación es falsa, y si e<1 la ecuación es siempre verdadera. Si T=Π, la

ecuación es 1/2=1/2(1-e/2), lo cual es absurdo. Si T=M, la ecuación 1/2=Min(1/2,1-Min(e,1/2)) es equivalente a Min(e,1/2)≤1/2, que es ¡siempre verdadera!.

Si T es continua y arquimediana, de la ecuación se sigue que

$$t(1/2)+(t(1-T(e,1/2))=t(1/2),$$

ó

$$t(1-T(e,1/2))=0,$$

siendo t un generador aditivo de T. Entonces, dado que T(e,1/2)=2, se sigue t(e)+t(1/2)=t(0)=1, y, finalmente,

$$e=t^{(-1)}(1-t(1/2)),$$

caracterizando e como «el valor negado del valor común 1/2» bajo la función de negación $n_t(x)=t^{(-1)}(1-t(x))$,asociada con la t-norma T [Alsina y col., 1983]. Por ejemplo, si T=Ł entonces t=1-j y e=1/2.

6.3.

Cuando m=1 y s=c=1/2, la LGB se transforma

$$1/2=1-T(1,1-T(e,1/2)),$$

que es 1-T(e,1/2)=1/2, o T(e,1/2)=1/2. Si T=Z, esta ecuación es equivalente a e=1; si T=M, es equivalente a 1/2≤e. Si T posee un generador aditivo, la ecuación es t(e)+t(1/2)=t(1/2), o t(e)=0, y e=1. Este es el caso de ser T=Ł o T=Π.

6.4.

Cuando m=1, s=0 y c=1/2, la LGB se transforma en

$$1/2=1-T(1,1-T(e,0))=1-T(1,1)=0$$

lo cual es absurdo.

6.5.

Cuando m=c=1, la LGB se transforma en 1=1-T(1,1-T(e,s)), o T(1,1-T(e,s))=0, es decir T(e,s)=1, significando e∧s=1 e implicando siempre e=s=1 (como se puede también demostrar). En cualquier caso, no es posible tener s=0, pues t(e,0)=0. Entonces, bajo la lógica sobreimpuesta dada por una t-norma, no es

posible la combinación m=c=1-s=1 y este resultado conlleva los mismos comentarios que hicimos en el caso de las medidas de Sugeno.

6.6.

Cuando c=0, LGB se transforma en T(m,1-T(e,s))=1. Si T=Z, esta ecuación implica m=1-Z(e,s)=1 y esto significa que «m=1 y tanto e como s están en [0,1)». Si T=Π, la ecuación es m(1-es)=1, implicando m=1 y e·s=0. Si T=M, la ecuación es Min(m,1-Min(e,s))=1, implicando m=1-Min(e,s)=1, es decir, m=1 y Min(e,s)=0.

Si T es continua y arquimediana con generador t, entonces:

$$t(m)+t(1-T(e,s))=0,$$

lo cual significa

$$t(m)=t(1-T(e,s))=0,$$

ó

«m=1 y T(e,s)=0».

La última ecuación da $s=t^{(-1)}(1-t(e))$.

Entonces, la solución es m=1 y $s=n_t(e)$ para cualquier $e \in [0,1]$. Por ejemplo, si T=Ł, como t=1-j entonces m=1 y s=1-e, $e \in [0,1]$

6.7.

Finalmente consideramos el caso de que c=1. LGB se transforma ahora en

$$T(m,1-T(e,s))=0.$$

Si T=Z, la ecuación Z(m,1-Z(e,s))=0 da las soluciones:

- $0 \neq m \neq 1$ y $Z(e,s) \neq 0$
- m=1 y e=s=1
- m=0 y Z(e,s)=0.

Entonces, si T=Z, hay varias soluciones.

Si T=Ł, la ecuación es Max(0,m-T(e,s))=0 implicando m=T(e,s)= Max(0,e+s-1), que da:

$$m = \begin{cases} 0 & \text{si } e+s \geq 1 \\ e+s-1, & \text{si } e+s<1 \end{cases}$$

Si T=Π, tenemos m(1-es)=0, es decir m=0 o e=s=1. También si T=M, tenemos Min(m,1-Min(e,s))=0 y m=0 o e=s=1.

Si T tiene un generador arquimediano es posible resolver la ecuación $t^{(-1)}(t(m)+t(1-T(e,s))=0$ análogamente al caso que resolvimos en la sección 5, pero con la dificultad adicional de que podríamos tener t(m)+t(1-T(e,s))>1 al principio de la discusión. Este caso conlleva un largo tratamiento que evitamos. De cualquier manera, si t es tal que t(m)+t(1-T(e,s))<1, dado que esto significa que t(m)+t(1-T(e,s))=1, estando 1-t(1-T(e,s)) en [0,1], entonces

$$m=t^{(-1)}(1-t(1-T(e,s)).$$

En particular cuando T=Ł y $t=t^{(-1)}=1-j$ tenemos m=Ł(e,s), como ya hemos obtenido. Además, dado que

$$t(1-T(e,s))=1-t(m),$$

una sencilla computación nos da:

$$e=t^{(-1)}(t(1-t^{(-1)}(1-t(m)))-t(s)).$$

Debería también subrayarse que la ecuación T(m,1-T(e,s))=0 puede leerse $m\wedge\neg(e\wedge s)=0$, o $m\wedge(\neg e\vee\neg s)=(m\vee\neg e)\vee(m\wedge\neg s)=0$, implicando que $m\wedge\neg e=m\wedge\neg s=0$ y mostrando la solución m=0 y e=s=1. Pero no se excluyen otras soluciones.

REFERENCIAS

Bunge, M., *Ética y ciencia*, Ediciones Siglo XXI, Buenos Aires, 1976.

Ayer, A.J., *Ensayos filosóficos*, Editorial Ariel, Barcelona, 1979.

Sugeno, M., *Theory of Fuzzy Integrals and its Applications*, Tesis Doctoral, Tokyo Institute of Technology, Tokio, 1974.

Schweizer, B. y Sklar, A., *Probabilistic Metric Spaces*, North Holland, New York, 1983.

Trillas, E., «Sobre funciones de negación en la teoría de conjuntos difusos», *Stochastica*, Vol. III, N. 1, (1979), 47-59.

Alsina, C., Trillas, E. y Valverde, L., «On some Logical Connectives for Fuzzy Set Theory», *Journal of Mathematical Analysis and its Applications*, Vol. 93, N. 1, (1983), 15-26.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ON SOME BISYMMETRIC FUNCTIONS ON CLOSED INTERVALS

Sergei Ovchinnikov
San Francisco State University. 1600 Holloway Ave., San Francisco, CA 94132, USA

1. INTRODUCTION

In this paper we present some technical results concerning binary operations on closed intervals of real numbers. The main goal of this work is to develop a unified approach to t-norms, t-conorms, and means. Our studies in this area have been strongly motivated by a paper by Trillas et al. [1992].

Let $[a,b]$ be an interval and $\circ$ be a binary operation on this interval which is continuous and increasing in both variables. In other words, we have a (numerical) *relational system* $< [a,b], \leq, \circ >$. If $\circ$ is a *strictly* increasing function in both variables, it's called a reducible or cancelable operation [Aczel, 1966; Pfanzagl, 1971]. Examples of such systems include, but not limited to, t-norms and t-conorms on the unit interval (when modeling logical connectives AND and OR), and various means. Although t-norms, t-conorms, and means are quite different operations with respect to their properties, there is a particular common property which they share. It's called bisymmetry and defined by

$$(x \circ y) \circ (u \circ v) = (x \circ u) \circ (y \circ v) \tag{1}$$

for all $x, y, u, v \in [a, b]$. For instance, any associative and commutative operation is bisymmetric.

If $\circ$ is a strictly increasing function in both variables, then there exists [Aczel, 1966] a continuous strictly increasing real function $f(t) \in [a,b]$ such that

$$x \circ y = f(pf^{-1}(x) + qf^{-1}(y) + r) \tag{2}$$

for all $x,y \in [a, b]$, where $p>0, q>0$. These operations are called *quasi-linear functions* [Aczel, 1966], and include *quasi-linear means* and *quasi-arithmetic means*. Note, that it is not true that *any* operation in the form (2) is bisymmetric. Unfortunately, t-norms and t-conorms are not strictly increasing functions and, therefore, are not covered by representation (2). On the other hand, there is an elaborate theory of t-norms and t-conorms which provides for a similar representation for continuous Archimedean functions [Scweizer and Sklar, 1983]; this theory employs *quasi-inverse* functions instead of inverse functions.

In this paper we study operations $\circ$ on $[a, b]$ which can be represented in the form

$$x \circ y = f(pg(x) + qg(y) + r) \tag{3}$$

for all $x,y \in [a, b]$, $p > 0$, $q > 0$, where $g(x)$ *is* a continuous strictly increasing real function on $[a, b]$ with range $[A, B]$ and $f(x)$ is defined by

$$f(x) = \begin{cases} a & \text{if } x<A \\ g^{-1} & \text{if } A\le x\le B \\ b & \text{if } x>B \end{cases} \tag{4}$$

Thus defined $f(x)$ is a *quasi-inverse* of g. We allow values $A=-\infty$ and $B=\infty$, assuming, therefore, that the domain of f is an extended real line $[-\infty, \infty]$ and using usual conventions for operations like $-\infty + a$ or $\infty + a$ for a finite number a. It's also assumed that $-\infty + \infty$ is not defined. Functions g and f are called *generators* of $\circ$. It follows from (4) that $f(g(x))=x$ for all $x \in [a, b]$ and

$$g(f(x)) = max\{min\{x, B\}, A\} \tag{5}$$

for all real x.

Our goal is to describe all *bisymmetric* operations on $[a, b]$ which can be represented in the form (3).

We begin with one important observation. Suppose $g(x) = g'(x) + \gamma$ where γ is a real number and $g'(x)$ is a strictly increasing function on $[a, b]$ with range $[A', B'] = [A-\gamma, B-\gamma]$. Then

$$x \circ y = f'(pg'(x) + qg'(y) + r')$$

where $r' = r-(p+q-1)\gamma$. Suppose $p + q \neq 1$. Then, by choosing

$$\gamma = \frac{r}{p+q-1}$$

we may consider only representations in the form $x \circ y = f(pg(x) + qg(y))$. If $p + q = 1$ we should consider both cases when $r = 0$ and $r \neq 0$.

In what follows we assume that coefficients p and q are positive numbers.

2. OPERATIONS IN THE FORM : $x \circ y = f(pg(x) + qg(y))$

In this section only operations in the form $f(pg(x) + qg(y))$ are considered.

Obviously, the case $p + q = 1$ is rather special one (quasi-linear means), and we begin with the following

Lemma 1: If $p + q = 1$, then

$$x \circ y = f(pg(x) + qg(y))$$

is a bisymmetric operation on $[a, b]$.

Proof: Since $A \leq min\{g(x), g(y)\} \leq pg(x) + qg(y) \leq max\{g(x), g(y)\} \leq B$ we have

$$x \circ y = g^{-1}(pg(x) + qg(y))$$

on $[a, b]$. Hence,

$$(x \circ y) \circ (u \circ v) = g^{-1}(p(pg(x) + qg(y)) + q(pg(u) + qg(v))) =$$
$$= g^{-1}(p^2g(x) + pqg(y) + qpg(u) + q^2g(v))$$

Since the last expression is symmetric with respect to y and u, we have bisymmetry.

We will consider cases $p + q > 1$ and $p + q < 1$ separately beginning with the case when $p + q > 1$.

Lemma 2: *Suppose* $p+q>1$ and $A<0<B$. Then operation $\circ$ is not bisymmetric.

Proof: Suppose $\circ$ is a bisymmetric operation.

Since $A < 0 < B$ and g is a strictly increasing function there is a unique $c \in (a, b)$ such that $g(c) = 0$. Therefore, c is an idempotent element, i.e., $c \circ c = c$. Since $p + q > 1$, $A < 0$, and $B > 0$, we also have $a \circ a = a$ and $b \circ b = b$.

The product $x \circ y$ of any two idempotents is an idempotent element. Indeed, by bisymmetry, $x \circ y = (x \circ x) \circ (y \circ y) = (x \circ y) \circ (x \circ y)$. We have $b \circ c = f(pB) > c$, since $B>0$. Therefore, $b \circ c = b$. Similarly, $c \circ b = b$, $a \circ c = a$, and $c \circ a = a$.

Consider three cases:

a) $a \circ b = c$. Then

$$a = c \circ a = (a \circ b) \circ (a \circ c) = (a \circ a) \circ (b \circ c) = a \circ b = c,$$

a contradiction.

b) $a \circ b = a$. Then $B \neq \infty$. Consider a continuous increasing function $h(x) = x \circ b$ on $[a,b]$. Then the set $\{x / a < x \circ b < b\}$ is an open interval $(\lambda, \mu) \cup [a, c]$ and $h(x)$ is a strictly increasing function on (λ, μ), since $x \circ b = g^{-1}(pg(x) + qB)$ on (λ, μ). Since $p+q > 1$, there exists $y < b$ such that $y \circ y = b$. Indeed, $pg(y) + qg(y)$ is an increasing continuous function of y and $pB + qB > B$. Therefore, there exists $y < b$ such that $pg(y)+qg(y)>B$ implying $f(pg(y)+qg(y))=b$. Since the range of $h(x)$ is $[a, b]$, there exists $x \in (\lambda, \mu)$ such that $x \circ b = y$, or, equivalently, $(x \circ b) \circ (x \circ b) = b$. By bisymmetry, $(x \circ x) \circ (b \circ b) = (x \circ x) \circ b = b$. On the other hand, we have $x \circ x \leq x$ on (λ, μ), since $f(pg(x)+qg(x)) \leq f(g(x)) \leq x\ (g(x))$ *is* negative on (λ, μ)!). Therefore, $(x \circ x) \circ b \leq x \circ b = y < b$, a contradiction.

c) $a \circ b = b$. Then $A \neq -\infty$. Consider a continuous increasing function $h(x) = a \circ x$ on $[a, b]$. Then the set $\{x / a < a \circ x < b\}$ is an open interval $(\lambda, \mu) \cup [c, b]$ and $h(x)$ is a strictly increasing function on (λ, μ), since $a \circ x = g^{-1}(pA + qg(x))$ on (λ, μ). Since $p+q>1$, there exists $y > a$ such that $y \circ y = a$. Indeed, $pg(y) + qg(y)$ is an increasing continuous function of y and $pA + qA < A$. Therefore, there exists $y > a$ such that $pg(y)+qg(y)<A$ implying $f(pg(y)+qg(y))=a$. Since the range of $h(x)$ *is* $[a, b]$, there exists $x \in (\lambda, \mu)$ such that $a \circ x = y$, or, equivalently, $(a \circ x) \circ (a \circ x) = a$. By bisymmetry, $(a \circ a) \circ (x \circ x) = a \circ (x \circ x) = a$. On the other hand, we have $x \circ x \geq x$ on (λ, μ), since $f(pg(x)+qg(x)) \geq f(g(x)) \geq x\ (g(x))$ *is* positive on (λ, μ)!) Therefore, $a \circ (x \circ x) \geq a \circ x = y > a$, a contradiction.

Lemma 3: Suppose $p + q > 1$ and either $0 \le A < B < \infty$ or $-\infty < B < A \le 0$. Then $\circ$ is bisymmetric if and only if

$$|A| \le \max\left\{ \frac{1-q}{p}, \frac{1-p}{q} \right\} \frac{|B|}{p+q} \tag{6}$$

It is bisymmetric for any A if $0 \le A < B = \infty$ or $-\infty = B < A \le 0$.

(Note the reverse order of A and B when they are negative numbers, i.e., we assume $f(a) = B$ and $f(b) = A$ when $B < A \le 0$.)

Proof: We prove this lemma only for the case when $0 \le A < B$, because the proof in the other case is very similar to this one.

Suppose (6) is true. Then

$$p(p + q)A + qB \ge B \tag{7}$$

and

$$q(p + q)A + pB \ge B. \tag{8}$$

Note, that (7) and (8) are also satisfied when $B = \infty$. Hence, the following argument is valid in this case too.

We have, by (5) and since $pg(x) + qg(y) \ge (p + q)A > A$,

$$(x \circ y) \circ (u \circ v) = f(pg(f(pg(x) + qg(y)) + qg(f(pg(u) + qg(y))) =$$
$$= f(p \max\{\min\{pg(x) + qg(y), B\}, A\} + q \max\{\min\{pg(u) + qg(v), B\}, A\}) =$$
$$= f(p \min\{pg(x) + qg(y), B\} + q \{\min\{pg(u) + qg(v), B\}) =$$
$$= f(\min\{p^2 g(x) + pqg(y), pB\} + \{\min\{qpg(u) + q^2 g(v), qB\}) =$$
$$= f(\min\{p^2 g(x) + pqg(y) + qpg(u) + q^2 g(v), p^2 g(x) + pqg(y) + qB,$$
$$pB + qpg(u) + q^2 g(v), pB + qB\}) =$$
$$= \min\{f(p^2 g(x) + pqg(y) + qpg(u) + q^2 g(v)), f(p^2 g(x) + pqg(y) + qB),$$
$$f(pB + qpg(u) + q^2 g(v)), f(pB + qB)\}).$$

Obviously, $f(pB + qB) = b$. By (7),

$$f(p^2 g(x) + pqg(y) + qB) \ge f(p^2 A + pqA + qB) = b,$$

and, by (8),

$$f(pB + pqg(u) + q^2 g(v)) \ge f(pB + qpA + q^2 A) = b.$$

Therefore,

$$(x \circ y) \circ (u \circ v) = f(p^2 g(x) + pqg(y) + qpg(u) + q^2 g(v)).$$

The right side of this formula is symmetric with respect to x and y. This proves bisymmetry of $\circ$.

Suppose now that (6) is not satisfied (assuming, of course, that $B < \infty$) and consider two cases.

1) $A < \frac{1-q}{p} \cdot \frac{B}{p+q}$ or, equivalently, $p(p+q)A+qB < B$. Then, obviously, $A < pA + qB < B$ and $A < (p+q)\, A < \frac{1-q}{p} \cdot B < B$, since $\frac{B}{p+q} < 1$. We have now

$$a \circ a = f((p+q)A) < b \text{ implying } a \circ a = g^{-1}((p+q)A),$$
$$b \circ b = f((p+q)B) = b \text{ implying } b \circ b = g^{-1}(B),$$
$$a \circ b = f(pA+qB) < b \text{ implying } a \circ b = g^{-1}(pA + qB).$$

Hence,

$$(a \circ a) \circ (b \circ b) = f(pg(a \circ a)+qg(b \circ b)) = f(p(p+q)A+qB) < b.$$

On the other hand

$$(a \circ b) \circ (a \circ b) = f(p(pA+qB)+q(pA+qB)) > f(p(p+q)A+qB)$$

Indeed, $p(pA + qB) + q(pA + qB)) > p(p + q)A + qB$, since $p + q > 1$. Therefore,

$$(a \circ a) \circ (b \circ b) < (a \circ b) \circ (a \circ b)$$

and the bisymmetry condition is not satisfied.

2) $A < \frac{1-q}{p} \cdot \frac{1-p}{q}$. This case is similar to the previous one.

Suppose now that $p + q < 1$.

Lemma 4: Suppose $p + q < 1$ and $A \le 0 \le B$ $(A \neq B)$. Then $\circ$ is a bisymmetric operation.

Proof: Obviously, $A \le pg(x) + qg(y) \le B$ for all $x, y \in [a, b]$. Therefore,

$$x \circ y = g^{-1}(pg(x) + qg(y))$$

and we have

$$(x \circ y) \circ (u \circ v) = f(pg(f(pg(x) + qg(y)) + qg(f(pg(u) + qg(y))) =$$
$$= f(p^2 g(x) + pqg(y) + pqg(u) + q^2 g(v))$$

Since the right side is symmetric with respect to x and y, $\circ$ is a bisymmetric operation.

Lemma 5: Suppose $p + q < 1$ and either $0 < A < B < \infty$ or $-\infty < B < A < 0$. Then $\circ$ is bisymmetric if and only if

$$|B| > \min\left\{\frac{1-q}{p}, \frac{1-p}{q}\right\} \cdot \frac{|A|}{p+q}$$

It is not bisymmetric for any A if $0 \le A < B = \infty$ or $-\infty = B < A \le 0$.

(Note the reverse order of A and B when they are negative numbers, i.e., we assume $f(a)=B$ and $f(b)= A$ when $B < A \le 0$.)

Proof: The proof is very similar to the proof of Lemma 3 and ommited.

The results obtained in Lemmas 1-5 are summarized in the following theorem.

Theorem 1: Suppose $r = 0$. A binary operation $\circ$ is defined by (3)-(4) on a closed interval $[a, b]$ is bisymmetric if and only if one of the following mutually exclusive conditions is satisfied:

1) $p + q = 1$.

2) $p + q > 1$, $AB \ge 0$, $0 \le |A| < |B| < \infty$, and $|A| \ge \max\left\{\frac{1-q}{p}, \frac{1-p}{q}\right\} \frac{|B|}{p+q}$

3) $p + q > 1$, $AB \ge 0$, $0 \le |A| < |B| = \infty$.

4) $p + q < 1$, $A \le 0 \le B$.

5) $p + q < 1$, $AB > 0$, $0 < |A| < |B| < \infty$, and $|B| > \min\left\{\frac{1-q}{p}, \frac{1-p}{q}\right\} \frac{|A|}{p+q}$

6) $p + q < 1$, $AB > 0$, $0 < |A| < |B| = \infty$.

(Note the reverse order of A and B when they are negative numbers, i.e., we assume $f(a)=B$ and $f(b)=A$ when $B < A \leq 0$.)

3. OPERATIONS IN THE FORM $x \circ y=f(pg(x)+qg(y)+r)$

Since any operation (3) with $p+q\neq 1$ can be represented in the form (3) with $r=0$, we may assume that $p+q=1$ in the case $r\neq 0$.

Theorem 2: Suppose $p+q=1$, $r\neq 0$, $A\neq -\infty$ and $B\neq\infty$. Then, operation $\circ$ is bisymmetric if and only if

$$B - A \leq \min\left\{\frac{p+1}{p}, \frac{q+1}{q}\right\} |r| \tag{9}$$

It is bisymmetric if $A=-\infty$ or $B=\infty$.

Proof: We consider only case $r>0$, since the proof in the case $r<0$ is similar to this one.

Suppose (9) is true. Then $p(B-A)\leq pr + r$ and $q(B-A)\leq qr+r$, which is equivalent to

$$pA+pr+qB+r \geq B \tag{10}$$

and

$$qA +qr +pB + r \geq B \tag{11}$$

respectively, since $p + q = 1$.

Note, that (10) and (11) are also satisfied when $B = \infty$. Hence, the following argument is valid in this case too.

We have, by (5) and since $pg(x) + qg(y) + r \geq A+r > A$,

$$(x \circ y) \circ (u \circ v) = f(pg(f(pg(x)+qg(y))+qg(f(pg(u) + qg(y)) + r) =$$

$$=f(p\ max\{min\{pg(x)+qg(y)+r,B\},A\}+q\ max\{min\{pg(u)+qg(v)+r,B\},A\}+r)=$$

$$=f(p\ min\{pg(x)+qg(y)+r,B\}+q\{min\{pg(u)+qg(v)+r,B\}+r)=$$

$$=f(min\{p^2g(x)+pqg(y)+pr,pB\}+\{min\{qpg(u)+q^2g(v)+qr,qB\}+r)=$$

$$=f(min\{p^2g(x)+pqg(y)+qpg(u)+q^2g(v)+r,\ p^2g(x)+pqg(y)+pr+qB+r,$$
$$pB+qpg(u)+q^2g(v)+qr+r,B+r\})=$$
$$=min\{f(p^2g(x)+pqg(y)+qpg(u)+q^2g(v)+r),f(p^2g(x)+pqg(y)+qB+r),$$
$$f(pB+qpg(u)+q^2g(v)+r),f(B+r)\})).$$

Obviously, $f(B+r)=b$. By (10),

$$f(p^2g(x)+pqg(y)+qB+r) \geq f(p^2A+pqA+qB+r) = b,$$

and, by (11),

$$f(pB+qpg(u)+q^2g(v)+r) \geq f(pB+qpA+q^2A+r) = b,$$

since $p + q = 1$.

Therefore,

$$(x \circ y) \circ (u \circ v) = f(p^2g(x)+pqg(y)+qpg(u)+q^2g(v)+r).$$

The right side of this formula is symmetric with respect to x and y. This proves bisymmetry of $\circ$.

Suppose now that (9) is not satisfied. Then

$$B - A > \min\left\{\frac{p+1}{p}, \frac{q+1}{q}\right\}. r.$$

Consider two cases:

1) $B - A > \frac{p+1}{p} \cdot r$. Equivalently,

$$pA+pr+qB+r < B. \qquad (12)$$

We have

$a \circ a = f(A+r) = g^{-1}(A+r)$, since, obviously, $B>A+r$.

$b \circ b = f(B+r) = g^{-1}(B) = b$.

$a \circ b = f(pA+qB+r) = g^{-1}(pA+qB+r)$, since, by (12), $pA+qB+r < B$

Therefore,

$$(a \circ a) \circ (b \circ b) = f(p(A+r)+qB+r) = f(pA+pr+qB+r)$$

and

$$(a \circ b) \circ (a \circ b) = f(p(pA+qB+r)+q(pA+qB+r)=$$
$$=f(p^2A+pqB+pr+qpA+q^2B+qr+r) = f(pA+qB+2r).$$

f is a strictly increasing function on (A, B). Hence, by (12),

$$(a \circ b) \circ (a \circ b) > (a \circ a) \circ (b \circ b),$$

and $\circ$ is not bisymmetric.

2) $B - A > \dfrac{p+1}{p} \cdot r$. Similar to 1).

4. TRIVIAL SOLUTIONS

It is quite possible that an operation defined by (3) is a *trivial* one, in other words, for all x and y, $x \circ y = c$ for some $c \in [a, b]$. Obviously, in this case $\circ$ is a bisymmetric operation.

Theorem 3: An operation $\circ$ is not trivial if and only if $a \circ a < b$ and $b \circ b > a$.

Proof: Suppose $\circ$ is trivial and $x \circ y = c$ for all $x, y \in [a, b]$ and some $c \in (a, b)$. Then $pg(x) + qg(y) + r = g(c)$, which is impossible, since g is a strictly increasing function. Therefore, in this case, either $x \circ y = a$ or $x \circ y = b$ for all $x, y \in [a, b]$ and we conclude that either $b \circ b = a$ or $a \circ a = b$. On the other hand, if, say, $b \circ b = a$, then, by monotonicity of $\circ$, $a \leq x \circ y \leq b \circ b = a$ for all $x, y \in [a, b]$. Hence, $\circ$ is trivial. It is also trivial if $a \circ a = b$. Therefore, $\circ$ is trivial if and only if $a \circ a = b$ or $b \circ b = a$, which is equivalent to the statement of the theorem.

In terms of generators f and g, non-triviality condition can be expressed in the following form

$$(p+q)A + r < B \text{ and } (p+q)B + r > A.$$

By using this condition, we can reformulate our theorems 1 and 2 as follows.

Theorem 4: Suppose $r = 0$. Then $\circ$ is a non-trivial bisymmetric operation if and only if one of the following conditions is satisfied:

1) $p + q = 1$, $B > A$.

2) $p + q > 1$, $AB \geq 0$, $|B| \neq \infty$, and $\max\left\{ \dfrac{1-q}{p}, \dfrac{1-p}{q} \right\} \dfrac{1}{p+q} \leq \dfrac{|A|}{|B|} < \dfrac{1}{p+q}$.

3) $p+q>1$, $AB\geq 0$, $0\leq|A|<|B|=\infty$.

4) $p+q<1$, $A\leq 0\leq B$.

5) $p+q<1$, $AB>0$, $|B|\neq\infty$, and $\min\left\{\frac{1-q}{p},\frac{1-p}{q}\right\}\frac{1}{p+q}<\frac{|B|}{|A|}<\frac{1}{p+q}$

6) $p+q<1$, $AB>0$, $0<|A|<|B|=\infty$.

Theorem 5: Suppose $p+q=1$, $r\neq 0$, $A\neq-\infty$, and $B\neq\infty$. Then operation $\circ$ is a non trivial bisymmetric operation if and only if

$$|r|\leq B-A\leq\min\left\{\frac{p+1}{p},\frac{q+1}{q}\right\}|r|$$

It is bisymmetric if $A=-\infty$ or $B=\infty$.

5. CONCLUSION

In this paper we completely described bisymmetric operations which can be represented by (3). Associative symmetric operations are a special case of bisymmetric operations and important in applications. A complete characterization of such operations is a subject of the author's forthcoming publication.

It is an interesting open problem to describe the class of bisymmetric operations allowing representation (3). This problem will be addressed in our future studies.

REFERENCES

Aczel, J., *Lectures On Functional Equations And Their Applications*, Academic Press, 1966.

Pfanzagl, J., *Theory Of Measurement*, Physica-Verlag, 1971.

Scweizer, B. and Sklar, A., *Probabilistic Metric Spaces*, North-Holland, 1983.

Trillas, E., Alsina, C., Cubillo, S. and Jacas, J., «On "AND" in [0, 1]», *Proc. International Conference on Information Processing and Management of Uncertainty in Knowledge Based Systems*, Mallorca, July, (1992), 329-332.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ON SOME MEASURES OF ENTROPY RELATED TO AGGREGATION OPERATORS

Ronald R. Yager
Machine Intelligence Institute, Iona College. New Rochelle, NY 10801, USA.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

1. INTRODUCTION

The measure of entropy plays a significant role in the field of information theory. Essentially the entropy measures the degree of uncertainty associated with a probability distribution. Assume our set of possible outcomes of a random experiment is the set $X=\{x_1,...,x_n\}$. Let $P=\{p_1,..., p_n\}$ be a probability distribution defined on X, $p_i \in [0,1]$ and $\sum_i p_i=1$. Then we shall denote the entropy associated with this distribution as H(P). In Aczel and Daroczy [1975] and Klir and Folger [1988] a number of important desirable properties associated with this measure are described. One desirable property is that which Klir and Folger [1975] calls symmetry. This property essentially requires that the value of H(P) is indifferent to the indexing of the individual probability values. Another property is that of expansibility. This is the requirement that if we add an element with the probability zero to an experiment this should not change the entropy. It is also desirable that the measure of entropy is continuous with respect to its arguments, the individual probability values. The next property associated with the measure of entropy is one of its major defining properties. An entropy measure should take its maximal value when all the outcomes in the space are equally likely. Thus maximal entropy occurs

when $p_i=1/n$ for all i. The reason for this is that when all the outcomes are equally likely the situation is most uncertain and an entropy measure, which measures the uncertainty associated with a probability distribution, should capture this. A further desired property is that in this equally likely case the entropy value should increase as n increases. Thus if distribution P, has $p_i=1/n$ for all i and $\hat{P}$ has $p_j=1/m$ for all j then m>n we desire $H(\hat{P})>H(P)$. Thus we see the more elements that are equally likely the more uncertainty. This Klir calls monotonicity. Another property associated with a measure of entropy is that its minimal values occur at the probability distribution that provides the most certainty amount in a situation. In particular we desire H(P) is minimal when one of the probabilities has unit value, $p_i=1$ for some i.

Among the other desired characteristics associated with the measure of entropy are the properties of branching and addivity [Klir and Folger, 1975]. These properties are intimately connected with the combining of entropies from two different independent probability distributions. In particular, with these properties we are imposing the condition that the entropy of a probability distribution constructed by the combination of two independent probability distributions be the sum of the entropies of the individual probability distributions. The concern for this required plays a significant role in imposing the logarithmic nature on the entropy concept, the Hartley-Shannon type entropies [Shannon and Weaver, 1964], $-\sum_i p_i \log p_i$. In many applications where we use the concept of entropy without involving the combination of more than one probability distribution the imposition of these requirements just serves to restrict the choice of functional forms that can be used to measure the uncertainty inherent in a probability distribution, its entropy.

One important application where there is no need to impose the addivity requirement is in the use of the principle of maximal entropy [Brillouin, 1956]. In this situation we are interested in selecting the allowable probability distribution that conveys the most uncertainty regarding the elements. In many applications of the principle of maximal entropy we are solving an optimization problem under constraints, finding the probability distribution that maximizes the entropy while solving some conditions. In these kind of applications of the principle of maximal entropy it would be desiderable to have a large class of measures of entropy. This will ease the computational burden in some of these constraint optimization problems.

In the following we shall suggest the measures of entropy.

2. OWA- BASED ENTROPY MEASURES

Yager [1988] introduced the concept of an OWA operator. These operators provide a family of aggregators having the properties of averaging operators. In this section we show how these operators can be used to derive a family of entropy measures.

An OWA operator of dimension n is a function $F:R^n \rightarrow R$ which has associated with it a weighting vector

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_n \end{bmatrix}$$

with

$$1)\ w_i \in [0,1]$$

$$2)\ \sum_i w_i = 1$$

such that

$$F(a_i,\ldots\ldots a_n) = \sum_i w_i b_i$$

where b_i is the i^{th} largest of the a_j.

Yager [1988] shows a number of properties associated with this operator. If $w_i=1$ then $F(a_1,\ldots,a_n)=\text{Max}[a_i]$, if $w_n=1$ then $F(a_i,\ldots,a_n)=\text{Min}[a_i]$. Furthermore if $a_i=a$ for all i, then

$$F(a_i,\ldots a_n) = a.$$

This is called the idempotency property.

It can be easily shown that if arguments lie in the unit interval, $a_i \in I$, then their aggregations lies in the unit interval. In particular this implies for arguments in the unit interval their maximal possible aggregation is one. It also can be easily shown that if the $a_i \in [0,1]$ then their aggregation lies in this interval.

Yager [1993] extended the idea of OWA operators, by allowing the weights to be argument dependent. In particular we can allow

$$w_i = g_i(b_1,b_2,\ldots b_n).$$

Thus the weights depend on the ordered arguments.

One particulary interesting family of argument dependent weights, which are suggested in Yager and Filev [1992] are

$$w_i = \frac{b_i^{\alpha}}{\sum_i b_i^{\alpha}} \quad \alpha \geq 1.$$

We note that these weights satisfy the two required conditions of weights, they sum to one and lie in the unit interval.

Becouse of the close connection of these weights with the BADD transformation [Yager and Filev, 1991] used for defuzzification we call these **BADD-OWA** weights.

In the case of using these BADD-OWA weights we get

$$F(a_1,a_2,\ldots,a_n) = \frac{\sum_i b_i^{\alpha+1}}{\sum_i b_i^{\alpha}} \quad \alpha \geq 1.$$

We further note that in this case no ordering is required amongst the arguments. Hence, we can write

$$F(a_1,a_2,\ldots,a_n) = \frac{\sum_i a_i^{\alpha+1}}{\sum_i a_i^{\alpha}} \quad \alpha \geq 1.$$

It is our objective to show that this formula provides the basis of a family of entropy-like measures that can be used in maximal entropy calculations.

Let us now assume that the a_i are the elements of a probability, P. In particular $a_i=p_i$ where $\sum_i p_i=1$ and $p_i \in [0,1]$. Thus in this case

$$F_{\alpha}(P) = F_{\alpha}(p_1,p_2,\ldots,p_n) = \frac{\sum_i p_i^{\alpha+1}}{\sum_i p_i^{\alpha}} \quad \alpha \geq 1.$$

We can easily identify the two boundary cases of this formulation.
When $\alpha=1$ we get

$$F_{\alpha}(P) = \frac{\sum_i P_i^2}{\sum_i P_i},$$

since $\sum_i P_i=1$ we get for $\alpha=1$

$$F_\alpha(P) = \sum_i P_i^2.$$

When $\alpha \to \infty$ we get

$$F_\alpha(P) = \frac{\sum_i P_i^{\alpha+1}}{\sum_i P_i} = \frac{\sum_i \left(\frac{P_i}{P^*}\right)^\alpha . P_i}{\sum_i \left(\frac{P_i}{P^*}\right)^\alpha}$$

Where $P^*=\text{Max}[P_i]$. Since $\left(\frac{P_i}{P^*}\right)^\alpha \to 0$ for $P_i < P^*$ when $\alpha \to \infty$, and $\left(\frac{P_i}{P^*}\right)^\alpha = 1$ when $\alpha \to \infty$ we get

$$F_\alpha(P) = \frac{\sum_{\substack{i.s.t. \\ P_i=P^*}} (P_i) = P^* = \text{Max}[P_i]}{\sum_{\substack{i.s.t. \\ P_i=P^*}} (1)}$$

Thus for $\alpha \to \infty$, $F_\infty(P) = \text{Max}[P_i]$

In the following we shall call a probability distribution which has one of its constituent probabilities as one and all the others as zero, a singleton probability distribution.

The following theorem shows that the maximal value for $F_\alpha(P)$ occurs, uniquely, when P is a singleton probability distribution.

Theorem: For all α

1. $F_\alpha(P)= 1$ if P is a singleton
2. $F_\alpha(P) < 1$ if P is not a singleton.

Proof:

1. Assume P is a singleton. Without loss of generality assume $p_1=1$. Then

$$F_\alpha(P) = \frac{\sum_i P_i^{\alpha+1}}{\sum_i P_i^\alpha} = \frac{1}{1} = 1.$$

2. Assume P is not a singleton. Then Max[P_i]<1. Furthermore for each $p_i \neq 0$ $p_i^{\alpha+1} < p_i^{\alpha}$ and since there exists at least one $p_i \neq 0$, there exists at least one p_i for wich $p_i^{\alpha+1} < p_i^{\alpha}$. From this it follows that

$$\sum_i P_i^{\alpha+1} < \sum_i P_i^{\alpha}$$

and hence $F_\alpha(P)<1$.

Thus we see that $F_\alpha(P)$ uniquely attains its maximum value when P is a singleton.

Consider next the case when $p_i=1/n$. In this case

$$F_\alpha(P) = \frac{\sum_i (P_i)^{\alpha+1}}{\sum_i (P_i)^{\alpha}} = \frac{n\left(\frac{1}{n}\right)^{\alpha+1}}{n\left(\frac{1}{n}\right)^{\alpha}} = \frac{1}{n}.$$

Actually this result also can be seen to be a result of the idempotency property of OWA weights.

Using the technique of LaGrange multipliers we can show that the minimal $F_\alpha(P)$ occurs uniquely at $p_i=1/n$.

These properties of the BADD-OWA function $F_\alpha(P)$ lead us to suggest a formulation for a measure of entropy as

$$H_\alpha(P) = 1 - F_\alpha(P).$$

We note that since $F_\alpha(P)$ is symmetric so is $H_\alpha(P)$. The continuity of $H_\alpha(P)$ follows directly from the continuity of $F_\alpha(P)$. We also see that $H_\alpha(P)$ is expansive. The addition of a p_i with zero value doesn't affect the calculation. Since $F_\alpha(P)$ takes its maximal value of 1 when P is singleton, then $H_\alpha(P)$ takes its minimal value of zero when P is a single. Furthermore $F_\alpha(P)$ takes its minimal value of 1/n when P is $p_i=1/n$, then $H_\alpha(P)$ takes its maximal value when $p_i=1/n$ hence

Thus

$$H_\alpha(1/n) = 1 - 1/n = \frac{n-1}{n}.$$

$$H_\alpha(P) \in [0, \frac{n-1}{n}].$$

We also note that as n increases the maximal value increases. Thus we see that $H_\alpha(P)$ provides a family of entropy measures. The two particulary interesting cases are when $\alpha = 1$ and $\alpha = \infty$

$$H_1(P) = 1 - \sum_i P_i^2$$

$$H_\infty(P) = 1 - Max[P_i].$$

These two formulations are especially useful in applications of maximal entropy. For the use of H_1 can essentially be replaced by finding the minimal of $\sum_i P_i^2$. Similarly the use of H_∞ can be replaced by finding the minimal of $Max[P_i]$.

We can express these two cases also as

$$H_1(P) = 1 - \sum_i P_i^2 = \sum_i P_i(1-P_i)$$

and

$$H_\infty(P) = 1 - Max[P_i] = Min[\bar{P}_i]$$

It should be noted that in using $H_\infty(P)=1-Max[P_i]$ to solve a maximal entropy problem one can of course use $Max[P_i]$ and minimize this. We note that we would use this in a lexographic manner. Thus if P and P' are two distributions with the same maximal element we would then look to the second maximal element to distinguish them.

3. ALTERNATIVE MEASURES OF ENTROPY

In this section we shall look at some alternative measures of entropy. We shall first look at some metric-based measures of entropy.

In this section we shall view a probability distribution P as a vector in the unit hypercube [Kosko, 1991]. In particular, it is a normal vector since the sum of the components must equal one. Thus the set of all probability distributions are a subset of vectors in the unit hypercube.

Considering the concept of entropy as being a measure of the uncertainty associated with the probability distribution and using this vector space view we can

think about measuring entropy as some measure of distance from the point of maximal entropy, $p_i=1/n$.

In the following we shall denote the probability distribution $p_i=1/n$ by the vector [1/n]. Similarly we shall denote [I_j] as the probability distribution with $p_j=1$. At times we shall not need to distinguish the location of the unitor -we shall then simply denote the probability distribution with a singleton as [I].

Consider a function Q(P) that measures the distance in n space of a probability distribution P from the vector [1/n]. In some sense this distance can be seen as a measure of certainty associated with the probability distribution. We shall refer to Q(P) as a measure of neg-entropy since the further an element is from [1/n] the more certain and the lower its entropy.

We first assume we use the Euclidean metric

$$Q(P) = D(P, [\frac{1}{n}]) = \left(\sum_i (p_i - \frac{1}{n})^2\right)^{1/2}.$$

Consider the term

$$R(P) = \sum_i \left(p_i - \frac{1}{n}\right)^2.$$

We can restate this more expressively as

$$R(P) = \sum_i p_i^2 - \frac{2}{n} p_i + \frac{1}{n^2}.$$

Taking the sum we get

$$R(P) = \sum_i p_i^2 - \frac{2}{n} \sum_i p_i + \frac{1}{n}.$$

Realizing that P is normal, $\sum(P_i)=1$, we get

$$R(P) = \sum_i p_i^2 - \frac{2}{n} + \frac{1}{n} = \sum_i p_i^2 - \frac{1}{n}.$$

Two observations are in order. First we note that $R(I_i)=R(I_j)$ thus the distance of any singleton probability distribution is the same. Furthermore R(P) assumes its maximal value for P=[I]. Thus the singleton is the furthest distance from the most certain, [1/n]. We note that when P=[I] then

$$R([I]) = 1 - \frac{1}{n}.$$

If we construct

$$H(P) = R([I]) - R(P)$$

$$H(P) = (1 - \frac{1}{n}) - (\sum_i p_i^2 - \frac{1}{n})$$

$$H(P) = 1 - \sum_i p_i^2$$

we get a measure of entropy.

More generally if we consider any Minkowski metric

$$D(A,B) = (\sum_i (a_i - b_i)^2)^{1/z} \qquad z \geq 1$$

we develop a measure of neg-entropy as

$$Q(P) = D(P,[\frac{1}{n}]) = \left(\sum_i |p_i - \frac{1}{n}|^z\right)^{1/z}.$$

It is easy to see the symmetry and continuity of this measure. Furthermore we note that $Q(I_i)=Q(I_j)=Q(I)$, the distance to any singleton is the same. It can be shown that the maximal value for Q(P) occurs when P=[I]. In the case where P=[I] we get

$$Q([I]) = \left(\left(\frac{n-1}{n}\right)^z + n^{-1}\left(\frac{1}{n}\right)^z\right)^{1/3} =$$

$$= \frac{1}{n}\left((n-1)^z + n - 1\right)^{1/z} = \frac{n-1}{n}\left(1 + \frac{1}{(n-1)^z}\right)^{1/z}.$$

We can now define a class of measures of entropy as

$$H(P) = Q(I) - Q(P)$$

A particularly noteworthy class of these measures occur when z=1. In this case we get

$$Q(P) = \sum_i |p_i - \frac{1}{n}| \quad \text{and} \quad Q(I) = 2\frac{n-1}{n}.$$

Another interesting case occurs when $z \to \infty$. In this case

$$D(A,B) \to Max\ [|a_i - b_i|].$$

Thus

$$Q(P) = Max\ [(P_i - \frac{1}{n})] = Max\ [P_i] - \frac{1}{n}$$

and

$$Q(I) = 1 - \frac{1}{n}.$$

Thus

$$H(P) = Q(I) - Q(P) = 1 - Max\ [P_i].$$

There appears to be another view of the measure of entropy that exists in the Euclidean space. Consider the vector [1/n] and any other probability P. We can define the cosine of the angle θ between them as

$$\cos\theta = \frac{< P, \left[\frac{1}{n}\right] >}{\|P\| \left\|\frac{1}{n}\right\|}.$$

In the above <P,[1/n]> is the inner product of two vectors. Using the standard inner product

$$<A,B> = \sum_i a_i b_i$$

we get

$$<P, [\frac{1}{n}]> = \sum_i \frac{1}{n}\ p_i = \frac{1}{n}.$$

In addition ||A|| is the associated norm

$$\|A\| = <A,A>^{1/2}.$$

Thus

$$\|\frac{1}{n}\| = \left(\sum_i \left(\frac{1}{n}\right)^2\right)^{1/2} = \left(\frac{1}{n}\right)^{1/2}$$

In addition

$$\|P\| = (\sum_i P_i^2)^{1/2}.$$

Thus

$$\cos\theta = \frac{\frac{1}{n}}{\left(\frac{1}{n}\right)^{1/2}(\sum_i P_i^2)^{1/2}} = \frac{\left(\frac{1}{n}\right)^{1/2}}{(\sum_i P_i^2)^{1/2}}$$

We have shown that the term $\sum P_i^2$ is a measure of neg-entropy. Thus it appears that cos θ, the cosine of the angle between [1/n] and P, can be seen as a measure of entropy.

We have noted the role that the Max[P] plays in measuring entropy, it is a measure of neg-entropy of a probability distribution P. The max operation is an example of what one calls a t-conorm [Dubois and Prade, 1985]. A t-conorm S is an aggregation operator defined as follows:

$$S:[0,1] \times [0,1] \rightarrow [0,1]$$

1) S(a,0) = a
2) S(a,b) = S(b,a) (Commutativity)
3) S(a,b) ≥ S(c,d) if a≥c and b≥d (Monotonicity)
4) S(a,S(b,c)) = S(S(a,b),c) (Associativity)

We note that some examples of this operator besides the max are,

1. S(a,b) = a+b-ab
2. S(a,b) = Min (1,a+b)
3. S(a,b) = Min $(1,(a^z+b^z)^{1/z})$

Closely associated with the t-conorm operators are the t-norm operators. A t-norm T is also mapping [0,1] × [0,1] → [0,1] which while satisfying properties 2-4, instead of property 1 satisfies T(a,1)=a.

The fact that Max[P_i] plays the role of a measure of neg-entropy inspires one to consider looking at the family of t-conorms for other neg-entropy measures. We note that two desirable properties are already assumed in the basic definition of t-conorms. First the commutativity property assures us that the operator will give the property of symmetry. The boundary condition

$$S(a,0) = a$$

along with the associativity guarantees the expansibility. Assume $S(p_1,...,p_n)=q$. Consider $S(p_1,...,p_n, 0)$. Associativity implies

$$S(p_1,...,p_n, 0)=S(S(p_1,...,p_n),0)=S(q,0)$$

then the boundary condition implies $S(q,0)=q$.

Continuity is guaranteed by just restricting ourselves to continuous t-conorm operators.

In looking for neg-entropy measures in the t-conorms we must assure ourselves that the singleton probability distribution is unique in attaining a maximum. We shall say that a t-conorm S is **anti-nilpotent** [Dubois and Prade, 1985] if $S(a,b)=1$ iff $a=1$ or $b=1$. We note that if a t-conorm is anti-nilpotent then $S([I])>S([P])$ for any probability distribution $[P]\neq[I]$. We note that for [I], $s[I]=1$ and for [P] since no element of P equals one, $S([P])<1$. Thus anti-nilpotency guarantees the uniqueness of singleton distribution in attaining the maximum. We note that a+-b-ab is anti-nilpotent.

We must also impose the condition that S[P] uniquely attains its minimum value when [P]=[1/n].

Definition: Assume $a_1, a_2, a_3,..., a_n$ are numbers such that

$$\text{i)} \quad a_i \in [0,1]$$

$$\text{ii)} \quad \sum_i a_i = 1$$

we shall say a t-conorm S is **steadfast** if for any collection a_i satisfying i. and ii. it is the case that

$$S(a_1,a_2,a_3,\ldots,a_n) > S\left(\frac{1}{n}, \frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right).$$

Thus a required condition for a t-conorm S to be a neg-entropy generating function is that S is steadfast. It can be shown that the Max and $S(a,b)=a+b+ab$ are steadfast t-conorms while $S(a,b)=Min(1,a+b)$ is not. Thus not all t-conorms are steadfast. Since any t-conorm is completely defined in terms of its two argument formulation, S(a,b), an interesting open research question is how do we fundamentally characterize steadfast type t-conorms.

At this point we conjecture that what may be the fundamental characterizing property of the steadfast type t-conorms. Let $b\in[.5,1]$ then S is such that $S(b,\bar{b})$ is strictly monotonically increasing function of b, that is as b increases $S(b,\bar{b})$ increases.

Thus as the maximal value of an argument pair (b,$\bar{b}$) increases it is required that the their t-conorm value strictly increases. We shall say that a t-conorm is **resolute** if it satisfies this condition.

A well known [Alsina y col., 1983] duality exist between t-norms and t-conorms,

$$S(a,b) = 1 - T(\bar{a},\bar{b}).$$

Thus if S([P]) is a neg-entropy generating function then

$$T([\bar{P}]), \bar{p}_i = 1 - p_i,$$

can be seen to be an entropy generating funtion. Therefore since Max[p_i] is a neg-entropy we can see that

$$Min\ [p_i]$$

is a measure of entropy. Furthermore since S(a,b)=a+b+ab is a neg-entropy generating function we get that

$$\prod_i \bar{P}_i$$

is a measure of entropy. We note that can express our conjected requirement in terms of t-norms. A t-norm is called resolute if as the minimal value of the argument pair (b, $\bar{b}$) increases it must be the case that T(b,$\bar{b}$) strictly monotonically increases.

4. CONCLUSION

We have suggested some new measures for the calculation of the entropy of a probability distribution. One motivation for this is to provide tools for the application of the maximum entropy principle. The first class of entropy measures are based on the OWA operator. Another class of these measures are based on the distance of probability distribution from the uniform distribution. A third class of entropy measures are based on the t-conorms.

REFERENCES

Aczel, J. and Daroczy, Z., *On Measures of Information and their Characterizations*, Academic, New York, 1975.

Alsina, C., Trillas, E. and Valverde, L., «On some logical connectives for fuzzy set theory», *Journal of Mathematical Analysis and Applications*, N. 93, (1983), 15-26.

Brillouin, L., *Science and Information Theory*, Academic Press, New York, 1956.

Dubois, D. and Prade, H., «A review of fuzzy sets aggregations connectives», *Information Sciences*, N. 36, (1985), 85-121.

Klir, G.J. and Folger, T.A., *Fuzzy Sets, Uncertainty and Information*, Prentice-Hall, Englewood Cliffs, New Jersey, 1988.

Kosko, B., *Neuronal Netsworks and Fuzzy Systems*, Prentice Hall, Englewood Cliffs, New Jersey, 1991.

Shannon, C.L. and Weaver, W., *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1964.

Yager, R.R., «On ordered weighted averaging aggregation operators in multi-criteria decition making», *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 18, (1988), 183-190.

Yager, R.R., *Families of OWA operators*, Technical Report # MII-1301, Machine Intelligence Institute, Iona College, New Rochelle, New York, 1993.

Yager, R.R. and Filev, D.P., «A generalized defuzzification method under BAD distributions», *International Journal of Intelligence Systems*, N. 6, (1991), 687-697.

Yager, R.R. and Filev, D.P., *Parameterized «andlike» and «orlike» OWA operators*, Technical Report # MII-1216, Machine Intelligence Institute, Iona College, New Rochelle, New York, 1992.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CURSOS E CONGRESOS
DA
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA