



Universidad de Valladolid

ESCUELA DE INGENIERÍAS INDUSTRIALES

DEPARTAMENTO DE INGENIERÍA DE SISTEMAS Y AUTOMÁTICA

TESIS DOCTORAL:

**Análisis del Equilibrio entre Precisión e
Interpretabilidad de los Sistemas Basados en
Reglas Difusas Lingüísticos y Aproximativos**

Presentada por Marta Galende Hernández para optar al grado de
doctora por la Universidad de Valladolid

Dirigida por:

Dr. Gregorio I. Sainz Palmero

Dr. Rafael Alcalá Fernández



Universidad de Valladolid

ESCUELA DE INGENIERÍAS INDUSTRIALES

DEPARTAMENTO DE INGENIERÍA DE SISTEMAS Y AUTOMÁTICA

TESIS DOCTORAL:

**Análisis del Equilibrio entre Precisión e
Interpretabilidad de los Sistemas Basados en
Reglas Difusas Lingüísticos y Aproximativos**

Presentada por Marta Galende Hernández para optar al grado de
doctora por la Universidad de Valladolid

Dirigida por:

Dr. Gregorio I. Sainz Palmero

Dr. Rafael Alcalá Fernández

La memoria titulada “*Análisis del Equilibrio entre Precisión e Interpretabilidad de los Sistemas Basados en Reglas Difusas Lingüísticos y Aproximativos*”, que presenta Dña. Marta Galende Hernández para optar al grado de doctor, ha sido realizada dentro del programa de doctorado “*Ingeniería de Procesos y Sistemas (RD1393/2007)*” del Departamento de Ingeniería de Sistemas y Automática de la Universidad de Valladolid bajo la dirección de los doctores D. Gregorio I. Sainz Palmero y D. Rafael Alcalá Fernández.

Valladolid, Noviembre de 2013

El Doctorando

Fdo: Dña. Marta Galende Hernández

El Director

El Director

Fdo: D. Gregorio I. Sainz Palmero

Fdo: D. Rafael Alcalá Fernández

Agradecimientos

Quiero dedicar esta tesis a todas aquellas personas que de una u otra forma han contribuido a la realización de la misma. En especial a mi familia y amigos por su infinita paciencia.

A mis directores Gregorio Sainz y Rafael Alcalá por sus sabios consejos. A Goyo, gracias por dedicar parte de tu tiempo a guiarme en el camino de la investigación. Sin ti esto no hubiera sido posible. Y Rafa, gracias por ayudarme a darle el último empujón a este trabajo.

No puedo olvidarme de Isabel, mi compañera de aventuras durante todos estos años de trabajo y sacrificio. A pesar de las dudas que a veces hemos tenido al final al final conseguimos llegar donde nos proponíamos. Todo el esfuerzo que estamos realizando tiene ahora su recompensa.

Esta tesis me ha dado la oportunidad de realizar viajes que nunca olvidaré. Las experiencias vividas en los congresos serán imborrables. Goyo y María Jesús, gracias por vuestra compañía en esos viajes que nos han llevado a sitios tan lejanos como Japón o la India. Espero que podamos seguir disfrutando de ellos. Y Rafa y María José, gracias por vuestra ayuda y compañía durante mis estancias en Granada.

Por último quiero dar las gracias a mis compañeros de CARTIF, en especial a la división de TIC, por haber escuchado pacientemente mis quejas. A todos mis compañeros de trabajo durante estos últimos años gracias por vuestras aportaciones a esta memoria, vuestros consejos y vuestra compañía en los momentos de relax.

A todos, gracias por haber estado a mi lado en esta larga travesía.

Índice

1. Introducción	1
1.1. Modelos de datos	3
1.2. Técnicas de modelado	5
1.2.1. Interpretando los modelos	9
1.2.2. Optimizando múltiples objetivos	13
1.3. Objetivos	15
1.4. Organización de la Memoria	17
2. Estado del arte	19
2.1. Lógica Difusa: Modelado	19
2.1.1. Lógica Difusa: Hitos y situación actual	20
2.1.2. Modelado difuso	21
2.2. Equilibrio precisión-interpretabilidad	23
2.2.1. Precisión	24
2.2.2. Interpretabilidad	25
2.3. Interpretabilidad: Evolución de un concepto	27
2.3.1. Taxonomías de interpretabilidad	28
2.3.2. Interpretabilidad basada en complejidad: Formulaciones . . .	31
2.3.3. Interpretabilidad basada en semántica: Formulaciones . . .	34
2.3.4. Interpretabilidad basada en semántica: Últimas tendencias .	41

2.4. Sistemas difusos evolutivos	47
3. Reducción de la complejidad y mejora de la interpretabilidad mediante una métrica combinada de interpretabilidad semántica.	51
3.1. Agregación de índices de complejidad e interpretabilidad semántica en una métrica común	52
3.2. Mejora del equilibrio precisión - interpretabilidad en los sistemas basados en reglas difusas	55
3.2.1. Sistemas neuro-difusos	57
3.2.2. Selección de reglas usando algoritmos evolutivos multi-objetivo	58
3.3. Estudio experimental	61
3.3.1. Metodología experimental	62
3.3.2. Características de los sistemas basados en reglas difusas iniciales	64
3.3.3. Resultados y análisis de los sistemas basados en reglas difusas mejorados	66
3.3.4. Análisis Global	84
4. Interpretabilidad de los sistemas basados en reglas difusas lingüísticos y aproximativos: Generalización de GM3M y nueva métrica RMI	87
4.1. Evaluación de la interpretabilidad semántica de los sistemas basados en reglas difusas: Lingüísticos vs aproximativos	88
4.1.1. Generalización de la métrica GM3M para la interpretabilidad a nivel de base de datos	89
4.1.2. Nueva métrica de interpretabilidad semántica a nivel de base de reglas: RMI	97
4.1.3. Evaluando localmente la interpretabilidad relativa entre sistemas basados en reglas difusas: Modelos lingüísticos y pseudo-lingüísticos	103

4.2. Búsqueda del equilibrio en los sistemas basados en reglas difusas lingüísticos y aproximativos: Ajuste y selección de reglas usando AEMOs	104
4.2.1. Sistemas basados en reglas difusas iniciales	106
4.2.2. Post-procesamiento basado en algoritmos evolutivos multi-objetivo	106
4.3. Estudio experimental	113
4.3.1. Metodología experimental	114
4.3.2. Características de los sistemas basados en reglas difusas iniciales	117
4.3.3. Resultados y análisis de los modelos más precisos	119
4.3.4. Resultados y análisis de los modelos situados en la mediana de distintos planos objetivo	124
4.3.5. Resultados y análisis de los modelos más interpretables en los distintos planos objetivo	129
4.3.6. Análisis Global	136
4.3.7. Comparación local de modelos lingüísticos y aproximativos	137
5. Conclusiones	149
5.1. Resumen del trabajo realizado	149
5.2. Contribuciones de esta tesis	151
5.3. Lista de publicaciones relacionadas con esta tesis	153
5.4. Líneas de trabajo en desarrollo	156
Bibliografía	159
Apéndices	179
A. Lógica Difusa y Sistemas Basados en Reglas Difusas	181
A.1. Fundamentos de la Lógica Difusa	181
A.2. Sistemas basados en reglas difusas	183

B. Optimización Genética Multi-Objetivo	187
B.1. Algoritmos Genéticos	187
B.2. Algoritmos Genéticos Multi-Objetivo	190
B.2.1. NSGA-II	190
B.2.2. SPEA2	192

Tabla de Acrónimos

AEMO	—	Algoritmo Evolutivo Multi-Objetivo - <i>Multi-Objective Genetic Algorithm</i>	5
AG	—	Algoritmo Genético - <i>Genetic Algorithm</i>	4
ART	—	Teoría de Resonancia Adaptativa - <i>Adaptive Resonance Theory</i>	58
BC	—	Base de Conocimiento - <i>Knowledge Base</i>	7
BD	—	Base de datos - <i>Data Base</i>	7
BR	—	Base de reglas - <i>Rule Base</i>	7
ECM	—	Error Cuadrático Medio - <i>Mean Squared Error</i>	24
FASART	—	<i>Fuzzy Adaptive System ART based</i>	17
ISI WOK	—	<i>Institute for Scientific Information - Web of Knowledge</i>	9
L-IRL	—	<i>Linguistic Iterative Rule Learning</i>	17
NEFPROX	—	<i>Neuro-Fuzzy Function Approximation</i>	17
RBC	—	<i>Rule-Based Complexity</i>	41
RMI	—	<i>Rule Meaning Index</i>	89
SBRD	—	Sistema basados en reglas difusas - <i>Fuzzy Rule-Based System</i> ..	4
SDE	—	Sistema Difuso Evolutivo - <i>Genetic Fuzzy System</i>	5
SDEMO	—	Sistemas Difusos Evolutivos Multi-Objetivo - <i>Multi-Objective Evolutionary Fuzzy System</i>	13
S-IRL	—	<i>Scatter Iterative Rule Learning</i>	17
TSK	—	Takagi-Sugeno-Kang	8

Capítulo 1

Introducción

La generación y empleo de modelos como forma de capturar y copiar nociones del mundo real es consustancial al progreso del ser humano desde sus orígenes. Los antropólogos consideran que esta capacidad de construir modelos abstractos es una de las características más importantes del ser humano, y lo que le ha diferenciado del resto de sus competidores [1]. Los primeros modelos reconocibles fueron los números: la escritura y recuento de números como marcas en los huesos está documentado 30000 años AC. Tres de las grandes culturas de la antigüedad (babilónica, india y egipcia) poseían conocimientos de matemáticas 2000 años AC. y desarrollaron modelos matemáticos para mejorar su día a día.

Diophantus de Alejandría en el 250 DC. escribió su libro *Aritmética* dónde se esbozaban las primera ideas sobre el álgebra basada en simbolismos y la noción de *variable*. Abu Abd-Allah ibn Musa Al-Ḥwārizmī (siglo VIII) fue uno de los más famosos matemáticos árabes que publicó varios libros conteniendo modelos matemáticos para la resolución de problemas del mundo real referente a comercio, topografía, riego, etc.

En Occidente hasta el siglo XI no se produciría este nivel de desarrollo matemático ni de modelos. Y desde entonces hasta nuestros días la necesidad de generar modelos que ayuden en el día a día continúa: modelos para el regadío y la gestión del agua [2, 3]¹, modelos de cambio climático [4–6], modelos de enfer-

¹<http://www.nal.usda.gov/wqic/modeling.shtml>

medades [7–9], mercadotecnia [10, 11], conducta humana [12, 13], modelos sobre consumo [14, 15], ...

Los modelos capturan una percepción imperfecta de la realidad, son una versión simplificada de la realidad. Estos modelos pueden tener distintas funciones: *Explicar Fenómenos* como ocurre con la mayor parte de las teorías del campo de la Física, *Hacer Predicciones* sobre el desarrollo futuro de aspectos del mundo real, *Toma de Decisiones* y *Comunicación* del conocimiento.

Los conceptos involucrados en el desarrollo de modelos en el ámbito técnico-científico se resumen en: *variables*, *relaciones* y *datos*. Hoy en día un aspecto básico de todos estos modelos es su formulación que debe ser adecuada para su implementación mediante sistemas de computación.

De acuerdo a [16], los modelos pueden llevar a cabo dos funciones representativas diferentes. Por un lado pueden representar un aspecto del mundo real, distinguiéndose entonces dos tipos de modelos distintos en función de la naturaleza del objetivo: modelos de los fenómenos o modelos de datos. Y por otro lado, un modelo también puede representar una teoría. Desde el punto de vista del trabajo realizado en esta tesis, los *modelos de datos* serán el eje central de la misma.

Un modelo de datos es “una versión corregida, rectificada, reglamentada, y en muchos casos idealizada de los datos que obtenemos de la observación inmediata, los llamados *datos en bruto*” [16]. Los modelos de datos tienen un papel esencial en la confirmación de teorías, porque es el modelo de datos y no los datos en bruto lo que se compara con una predicción teórica. La generación de estos modelos puede ser extremadamente compleja, requiriendo avanzadas técnicas y metodologías exhaustivas para alcanzar el objetivo de una manera eficiente.

Por otro lado, los modelos son excelentes formas para aprender y comprender el mundo real que tratan de capturar [17]. En ocasiones el modelo permite descubrir aspectos y características sobre la realidad modelada que el modelo ha sido capaz de “descifrar”.

Por tanto, si somos capaces de conocer el modelo, de tener conocimiento del mismo, podremos transformar este conocimiento en conocimiento sobre la realidad objeto del modelo. El modelo puede enseñar sobre la naturaleza de la realidad. En este punto, la función de cómo comprender esta realidad de los modelos, juega un papel capital.

En esta tesis partiremos de estas ideas claves: modelos, datos, computación y

comprender el modelo para aprender de la realidad, para acercarnos al modelado basado en reglas difusas y a la necesidad de interpretar y comprender dicho modelo que representa con una precisión adecuada la realidad.

La organización del resto del capítulo es: en la Sección 1.1 se realiza una breve presentación de los principales aspectos relacionados con los modelos de datos que han dado lugar a la realización de este tesis. Posteriormente la Sección 1.2 se centra en el modelado de los sistemas difusos, la problemática típica asociada al concepto de interpretabilidad en el ámbito de los Sistemas Difusos Basados en Reglas y el enfoque más utilizado en la actualidad para conseguir el equilibrio entre ambos conceptos de precisión e interpretabilidad en este tipo de modelado. Para finalizar, en la Sección 1.3 se presentan los objetivos a alcanzar con la realización de esta tesis y en la Sección 1.4 la estructura seguida en la elaboración de la presente memoria.

1.1. Modelos de datos

Hoy en día la generación de modelos de cualquier índole es habitual en todos los campos de la actividad humana, y sobre todo en aquellos de actividad económica e industrial: finanzas, medicina, automoción, energía, mantenimiento industrial, mercadotecnia, etc. En el campo de la ingeniería los modelos son necesarios tanto para diseñar nuevos procesos como para analizar procesos ya existentes, lo que tiene un gran impacto a la hora de salvar y optimizar recursos de toda índole. Así por ejemplo tareas como la predicción o simulación del comportamiento de un sistema, el diseño de controladores, la optimización, la supervisión, la detección de fallos y/o el diagnóstico de componentes requieren de la creación de un modelo del proceso. La calidad y prestaciones de la solución final proporcionada dependerá en gran parte de la calidad del modelo generado [18].

Por otro lado, actualmente es muy habitual, cuando no obligatorio, la monitorización de actividades, sean estas de la naturaleza que sean, y el almacenamiento de estos datos/información en formato electrónico en bases de datos o similares. Estos datos reflejan la realidad de esta actividad en cuestión, por tanto la creación de modelos de datos (basados en datos) y la comprensión del modelo puede permitirnos aprender acerca de la realidad. Para llevar a cabo esto podemos contar, entre otras, con las **técnicas de aprendizaje automático a partir de datos**.

El resultado de estas técnicas son modelos que se caracterizan por reproducir el funcionamiento del sistema real de forma precisa [19, 20]. Esta capacidad o característica del modelo es la **precisión** del modelo. El problema es que la precisión no es la única característica deseable de un modelo si deseamos ir más allá, y que el modelo “*nos explique*” su comportamiento, para conocer los principios que guían el mismo.

En muchas ocasiones resulta casi imposible alcanzar este objetivo debido a la propia naturaleza del modelo. Éste es el enfoque de *caja negra* [18] donde es conocida la relación entre entradas y salidas pero no las “razones” de ese funcionamiento. Si pudiésemos conocer dichas razones también podríamos tener opciones de aprender sobre la realidad a partir de este conocimiento y su interpretación. Ésto es lo que usualmente se hace en los procesos de extracción de conocimiento a partir de datos.

A esa capacidad de explicación, y/o comprensión, del modelo se la conoce como **interpretabilidad** del modelo. Es una propiedad deseable, que se convierte en obligatoria en aquellos ámbitos de toma de decisiones donde la seguridad hace que cualquier decisión a tomar esté debidamente explicitada, al igual que el procedimiento o razonamiento que lleva a dicha toma de decisiones. Casos como el ámbito médico, donde a pesar de la precisión de los modelos el resultado puede levantar sospechas [21], o en el ámbito biotecnológico, en los que es importante poder explicar por qué las cosas ocurren de una forma determinada [22].

Una de las aproximaciones de modelado que puede permitir dar una solución a este problema es la Lógica Difusa [23]. Esta lógica da soporte formal al lenguaje natural y enfrenta su capacidad de representar y manejar conocimiento vago e impreciso con la dificultad para adquirir/aprender este conocimiento. Este último aspecto puede ser suplido por otras técnicas de aprendizaje automático, como pueden ser las Redes Neuronales Artificiales, Algoritmos Genéticos (AGs - *Genetic Algorithms*), etc. que si poseen esa capacidad y se pueden hibridar con la Lógica Difusa, aunando de esta forma la capacidad de aprender con la capacidad de manejar y representar información borrosa.

El uso de los llamados **Sistemas Basados en Reglas Difusas** (SBRDs - *Fuzzy Rule-Based Systems*) permite conseguir modelos difusos que son precisos, y en teoría, presentarían una buena interpretabilidad. Estos sistemas siguen un enfoque de modelado de *caja gris*, por lo que en su generación se pueden utilizar tanto datos como conocimiento de los principios básicos. El resultado son modelos capaces de mejorar incluso la precisión obtenida con otros tipos de modelado

matemáticos, pero en la realidad siguen sin ser realmente interpretables, o incluso mínimamente interpretables, en la mayoría de los casos [24]. Ésto se debe al uso de la Lógica Difusa como una mera herramienta matemática para construir modelos, olvidándose de principios básicos de la misma, y es lo que en muchos ámbitos se conoce como *Modelado Difuso*.

Para tratar de obtener SBRDs mediante técnicas de aprendizaje automático (por ejemplo a partir de datos) la Lógica Difusa se híbrida con múltiples técnicas, destacándose las denominadas Inteligencia Computacional o *SoftComputing* [25–27].

Una de las opciones que ofrece mejores perspectivas es la posibilidad de utilizar AGs dentro de los procesos de modelado difuso, más conocidos como **Sistemas Difusos Evolutivos** (SDEs - *Genetic Fuzzy System*) [28]. De esta forma se puede ver la generación de los modelos como un proceso de optimización que mediante el uso de los Algoritmos Evolutivos Multi-Objetivo (AEMOs - *Multi-Objective Genetic Algorithms*), permita obtener modelos con múltiples prestaciones, como puede ser en este dominio el conseguir SBRDs que tengan una buena precisión y un buen nivel de interpretabilidad. Por tanto, el problema consistirá en cómo formular matemáticamente los conceptos, como precisión e interpretabilidad, de forma adecuada para llevar a cabo el proceso de optimización de forma adecuada. Más concretamente la pregunta a responder se puede focalizar en cómo es (definir) y cómo medir (formular) el concepto de interpretabilidad de un SBRD.

Fundiendo todo lo anterior, surge el desarrollo de la presente tesis doctoral que tiene como líneas maestras: **revisar los conceptos de precisión e interpretabilidad** dentro del campo de los sistemas difusos y proponer nuevas fórmulas de evaluación que permitan **generar modelos difusos basados en reglas que sean a la vez suficientemente precisos e interpretables**, mejorando así las legibilidad del conocimiento aprendido y almacenado en el conjunto de reglas difusas.

1.2. Técnicas de modelado

De una forma general, y breve, el modelado de un sistema se puede definir como la generación de un modelo matemático que ante un vector de entrada (\vec{e}), produce una salida (\vec{y}) capaz de expresar con exactitud el comportamiento real del mismo [18]:

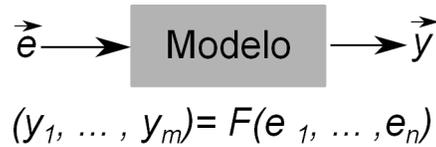


Figura 1.1: Modelado de sistema.

Las prestaciones deseadas para el modelo condicionaran la estructura y los valores de los parámetros concretos de la formulación matemática anterior. Actualmente existen tres enfoques principales a la hora de afrontar el modelado [18]:

- **Caja Blanca**, son aquellos modelos derivados directamente de los principios básicos de la física, la química, la biología, la economía, etc. Todas las ecuaciones y parámetros del modelo representan modelos teóricos. Se caracterizan por tener una baja dependencia de los datos y porque sus parámetros tienen interpretación directa con los principios básicos de la ciencia. Es decir, se conocen los detalles del funcionamiento matemático del sistema.
- **Caja Negra**, son aquellos modelos derivados directamente de los datos. Tanto las entradas como las salidas reales del sistema son conocidas y mediante modelización experimental, sin tener en cuenta explícitamente los principios básicos de la ciencia. Se obtiene un modelo que suele ser muy preciso pero cuyo funcionamiento no es nada fácil de discernir. Es decir, se conocen las decisiones tomadas por el sistema pero no cómo éstas son tomadas.
- **Caja Gris**, son modelos a mitad de camino entre los de caja blanca y caja negra, que tienen en cuenta tanto los principios básicos como la información aportada por los datos. Se caracterizan por ser capaces de integrar distintos tipos de información que está fácilmente accesible. Una forma de abordar este tipo de modelado es mediante el **modelado difuso de sistemas**, que combina la Lógica Difusa y las técnicas clásicas de modelado.

El problema que entonces surge es la forma de evaluar las prestaciones del modelo conseguido, para poder realizar una comparación y determinar aquél de mejores prestaciones. En este sentido, el índice más habitual, y en cierta forma

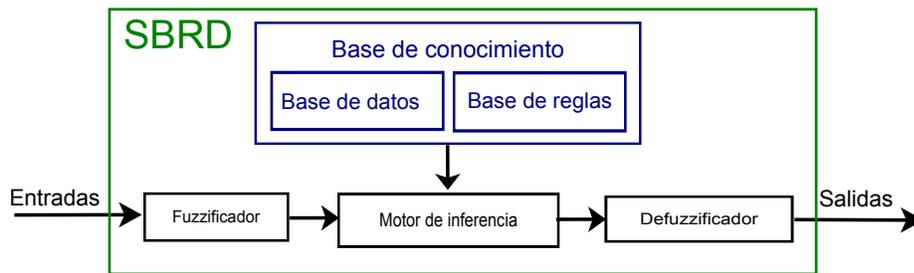


Figura 1.2: Componente de un SBRD.

básico, es la fidelidad con la que el modelo representa la realidad. Esto se ha medido a través de distintas formulaciones del *error* [20]. Posteriormente la evaluación de los modelos se ha ido ampliando a otros conceptos como la complejidad, su capacidad de extrapolación, su escalabilidad o su entendimiento. Concretamente esta última, vista como la capacidad de explicación o interpretabilidad de un modelo, es un factor clave a la hora de evaluar los modelos difusos, tanto desde el punto de vista de los principios de la Lógica Difusa como de la “necesidad” de conocer cómo el modelo toma sus decisiones (genera sus respuestas).

A partir de todo esto, una forma interesante de evaluar un modelo difuso consistirá en la evaluación de su fidelidad (precisión) y de su capacidad de explicación (interpretabilidad) [29–32]:

- **Precisión:** debe ser capaz de simular el comportamiento del sistema real con la mayor exactitud posible.
- **Interpretabilidad:** entendida como la facilidad que tiene la persona de entender el comportamiento del sistema a partir del modelo.

Para esto, un tipo de modelo difuso básico son los SBRDs (Fig. 1.2), cuyo componente más característico es una Base de Conocimiento (BC - *Knowledge Base*) expresada mediante un conjunto de reglas de tipo “SI condición/es ENTONCES consecuencia/s”, Base de Reglas (BR - *Rule Base*), y cuya semántica asociada está definida sobre conjuntos y particiones difusas, Base de Datos (BD - *Data Base*). En función del tipo de reglas se habla de sistemas [33]:

- **Lingüísticos** o Mamdani: tanto los antecedentes como los consecuentes de las reglas son variables lingüísticas representadas mediante conjuntos

difusos que puede ser definidos por un experto de dominio, o de forma alternativa por un sistema de aprendizaje automático, y que tienen un significado asociado en forma de etiqueta o término lingüístico. Este tipo de reglas son, en principio, más interpretables, sobre todo porque existe una semántica común definida que hace que sea más fácil determinar los hechos del mundo real al que hacen alusión las reglas [34].

- **Aproximativos** (o *Scatter*): donde los antecedentes y consecuentes de cada regla se definen utilizando distintos conjuntos difusos, cada uno de ellos con una semántica diferente. El uso de una semántica particular para cada regla dificulta determinar su relación con el mundo real [35].
- Takagi-Sugeno-Kang o **TSK**: en los que el antecedente de la regla esta definido por un conjunto difuso mientras que el consecuente es un valor numérico que se define en función de los antecedentes [36]. En este tipo los antecedentes también pueden seguir a su vez un enfoque lingüístico o aproximativo.

Todas esas reglas pueden ser definidas o generadas a partir de conocimiento experto, a partir de los datos asociados al sistema en cuestión, o mediante una combinación de ambos. Simplemente por utilizar conjuntos y reglas difusas en el modelado no se obtienen necesariamente modelos interpretables capaces de explicar de forma entendible el comportamiento de un sistema. Sin embargo esto no es necesariamente así, ya que el uso de la Lógica Difusa no garantiza *per se* la interpretabilidad del modelo final.

Se plantea entonces el dilema de generar SBRDs capaces de obtener buenas prestaciones para características como precisión e interpretabilidad pero, tal y como establece el Principio de Incompatibilidad de Zadeh, dichas características son excluyentes [37], lo que conlleva alcanzar soluciones de compromiso razonablemente balanceadas: desde el momento en que ambos aspectos entran en conflicto, el objetivo final del modelado difuso debe ser obtener un buen balance para ambos parámetros de forma simultánea. Es decir, es necesario buscar un equilibrio entre ambas características contradictorias si se quiere ser coherente con los conceptos básicos de la Lógica Difusa y aprovechar el potencial práctico de la misma [38–41]. Cómo medir estos conceptos y cómo alcanzar modelos que combinen de forma adecuada estas características es el reto que se va a abordar en este trabajo de tesis.

1.2.1. Interpretando los modelos

Como ya se ha comentado previamente conseguir que un modelo difuso sea interpretable es una característica muy deseable desde distintos puntos de vista. Sobre todo en aquellos casos que se necesita una explicación sobre el por qué de la conducta del modelo. Pero, ¿cómo saber cuando un modelo es interpretable o suficientemente interpretable? Para poder evaluar si un modelo cumple en alguna medida esta característica es necesario definir el concepto para posteriormente formularlo matemáticamente.

Como tal, el término **interpretabilidad** proviene de la palabra inglesa *interpretability* y, aunque no está recogido en la Real Academia Española², es ampliamente usado en el contexto científico y académico. Una primera y simple definición del concepto de interpretabilidad puede ser: “La capacidad de ser entendido o comprendido por una persona”.

La interpretabilidad en los modelos difusos es uno de los “topics” más relevantes en los últimos años. Una búsqueda en la *Thompson Corporation ISI Web of Knowledge* (ISI WOK) localiza 5748 publicaciones científicas sobre la temática³ distribuidas anualmente tal y como muestra la Fig. 1.3, que han sido citadas una media de 6,52 veces. Un factor determinante a la hora de justificar el elevado número de publicaciones científicas realizadas es la naturaleza subjetiva que tiene por definición la interpretabilidad.

Haciendo un rápido barrido por estas publicaciones se puede comprobar que, aunque las áreas de investigación asociadas a la mayoría de ellas están relacionadas con las ciencias de la computación y la ingeniería en distintos campos y dominios, también se han realizado publicaciones en áreas tan variadas como la del transporte, los combustibles de energía, la geología, la óptica o los recursos hídricos pasando por la robótica, la economía de la empresa o la ciencia de los materiales. Realizando un estudio más detallado de las publicaciones relacionadas con la interpretabilidad de los SBRD cabe destacar la temática asociada a cómo definir y cómo evaluar el concepto de interpretabilidad, que sigue siendo tema de debate entre los investigadores [29, 30, 38].

²<http://www.rae.es>

³Fecha de consulta: 13 Mayo 2013. Términos clave propuestos en [38]

TS=[("interpretab*") OR ("understandab*") OR ("comprehensib*") OR ("intelligib*") OR ("transpar*") OR ("readab*") OR ("complexity")) AND ("fuzzy")]

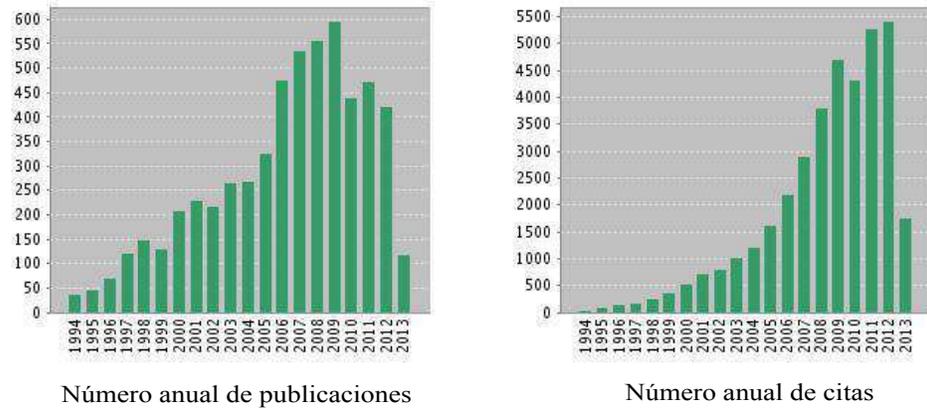


Figura 1.3: Estadísticas de publicaciones sobre interpretabilidad dentro del ámbito de la Lógica Difusa según ISI WOK.

Métricas de interpretabilidad

Las primeras formulaciones utilizadas para evaluar la interpretabilidad de un SBRD se basaban principalmente en medir la **complejidad** y/o en imponer **restricciones semánticas** sobre las particiones difusas [42–44]. La forma clásica de evaluar la complejidad del SBRD era a través del número de elementos (reglas, variables, condiciones de las reglas, etc.), mientras que como restricciones semánticas se imponían entre otras la normalización de los conjuntos difusos o el solapamiento adecuado entre particiones.

La mayoría de las publicaciones realizadas en estos primeros años se centraban en evaluar estos dos aspectos previos de interpretabilidad y no fue hasta diez años después de las primeras publicaciones, debido principalmente a la diversidad de propuestas existentes, cuando surgió la necesidad de establecer un marco de trabajo que ayudara a organizar tanto el concepto de interpretabilidad como las múltiples medidas utilizadas hasta el momento por la comunidad científica. Esto quedó materializado en la aparición de una serie de *revisiones del estado del arte*:

- Un completo análisis de las principales **restricciones de interpretabilidad** utilizadas hasta ese momento, así como su formulación asociada, fue realizado por Mencar y Fanelli en [45]. Una de las conclusiones de los autores es que la interpretabilidad no se puede limitar sólo a imponer restricciones

sobre los granulos de información, sino que es necesario tener en cuenta el concepto a lo largo de todo el proceso de diseño de los modelos difusos, algo que en ocasiones hace que la precisión empeore. Es necesario por tanto encontrar un equilibrio.

- Una taxonomía de interpretabilidad a dos niveles, centrada en el modelado de los SBRD, es realizada por Zhou y Gan en [46]. La forma de denominar a cada una de las categorías propuestas es: **Alto Nivel** para definir la interpretabilidad a nivel de regla difusa y **Bajo Nivel** para definir la interpretabilidad a nivel de conjunto difuso.

Algunas de las conclusiones expuestas por los autores establecen que, aunque hasta ese momento las publicaciones se habían centrado solo en formular la interpretabilidad de los sistemas a Bajo Nivel, una buena medida de interpretabilidad debería tener en cuenta ambos niveles, además de poder calcularse sobre cualquier tipo de modelo, sea Lingüístico, Aproximativo o TSK.

- Otra revisión centrada en la interpretabilidad de los SBRD en general fue publicada por Alonso y otros en [30] donde, tomando como base lo publicado en [46], se presenta otro marco de trabajo con elementos y conceptos adicionales. Los autores extienden la taxonomía de Alto/Bajo nivel de interpretabilidad estableciendo nuevos niveles llamados **Descripción** o Legibilidad de la Estructura del Sistema (que engloba los dos niveles propuestos en [46]) y **Explicación** o Comprensión del Sistema. Es decir, la interpretabilidad depende tanto de la legibilidad individual de cada uno de sus componentes como de la legibilidad conjunta de todos ellos, mecanismo de inferencia incluido.

La principal conclusión de los autores es que no existe una única medida capaz de cuantificar la interpretabilidad de un sistema, sino que es necesario definir un índice difuso capaz de adaptarse al contexto de cada problema así como a los criterios de calidad que establezca el usuario.

- Por último, una revisión centrada en la interpretabilidad de los SBRDs lingüísticos es presentada por Gacto y otros en [29], donde se propone una taxonomía particular basada en la estructura de dichos modelos lingüísticos. En concreto distingue dos tipos de interpretabilidad, una basada en **Complejidad** y otra basada en **Semántica**, que puede ser evaluadas sobre dos componentes principales del SBRD, la **BR** y la **BD**. La Tabla 1.1 muestra,

a modo de ejemplo, los cuatro cuadrantes de interpretabilidad obtenidos en la publicación y algunas medidas asociadas a dichos cuadrantes.

Tras el estudio realizado los autores concluyen que no existe una medida única y completa para cuantificar la interpretabilidad de los modelos lingüísticos. En su opinión, para cuantificar de manera global la interpretabilidad de un modelo sería necesario considerar las medidas adecuadas de todos los cuadrantes de manera independiente (o combinada en aquellos aspectos en los que tenga sentido), a fin de tener en cuenta las diferentes propiedades de interpretabilidad necesarias para este tipo de problemas (ya que son en muchos casos contradictorias entre sí).

	Nivel de BR	Nivel de BD
Complejidad	C_1 Número de reglas Número de condiciones	C_2 Número de funciones de pertenencia Número de características
Semántica	C_3 Consistencia de las reglas Reglas disparadas al mismo tiempo Transparencia de la estructura de reglas Cointension	C_4 Medidas absolutas: Completitud o cobertura, Normalidad, Distinguidad, Complementariedad Medidas relativas

Tabla 1.1: Cuadrantes de interpretabilidad según [29].

Analizadas las publicaciones previas se constata que las medidas basadas en la complejidad son, además de las medidas más clásicas, las más utilizadas en la mayoría de las publicaciones. Sin embargo, las medidas de interpretabilidad basadas en semántica son hoy en día un problema abierto.

Por otro lado, un aspecto a considerar a la hora de evaluar la interpretabilidad de un SBRD, y que casi no ha sido analizado hasta el momento, es el mecanismo de inferencia concreto utilizado por el SBRD, así como los operadores difusos utilizados tanto de conjunción, disyunción o defuzzificación. Ya en [47] se menciona

que la interpretabilidad del SBRD dependía, entre otros factores, de la sencillez del razonamiento difuso a aplicar. Aun así, pocos son los autores que han propuesto medidas teniendo en cuenta la influencia de estos mecanismos, a pesar de que si afectan a la interpretabilidad semántica a nivel de BR puesto que influyen en la forma en que interaccionan las reglas [48, 49].

Las métricas de interpretabilidad más utilizadas actualmente, así como un análisis más exhaustivo de las taxonomías propuestas en las revisiones previas, serán analizadas en la Sección 2.3 del Capítulo 2.

1.2.2. Optimizando múltiples objetivos

Actualmente multitud de problemas del mundo real involucran criterios que entran en conflicto: reducir el precio de un producto sin afectar a su calidad, diseñar dispositivos electrónicos que ocupen menos espacio y tengan más capacidad o realizar un viaje maximizando la velocidad sin aumentar en exceso el gasto de combustible. En el caso concreto del diseño de SBRDs las prestaciones en conflicto son la **precisión** y la **interpretabilidad**. En todos estos casos la mejora de uno de los criterios suele provocar que el otro empeore, por lo que es necesario encontrar un equilibrio o al menos contemplar distintas alternativas.

Una de las opciones más utilizadas para esta cuestión entre criterios contradictorios es utilizar los llamados **Algoritmos Evolutivos Multi-Objetivo** (AEMOs), capaces de optimizar varios objetivos simultáneamente generando un conjunto de soluciones finales no dominadas conocido como *Frente de Pareto*. A modo descriptivo, en la Fig. 1.4 se puede ver un ejemplo de optimización bi-objetivo de dos criterios f_1 y f_2 que se quieren minimizar de forma simultánea y donde puede verse aquellas soluciones no dominadas que forman parte del *Frente de Pareto*.

En el Apéndice B de esta memoria se pueden ver brevemente los principales fundamentos teóricos necesarios para entender el funcionamiento de los AEMOs. Destacar aquí que una parte fundamental de este tipo de algoritmos es la definición de las funciones objetivo, ya que son las encargadas de guiar el proceso de búsqueda por un espacio complejo de posibles soluciones. Por esta razón deben estar bien diseñadas para ser capaces de realizar una búsqueda eficiente y eficaz.

La incorporación de este tipo de técnicas a los sistemas de aprendizaje automático de los SBRDs se conoce con el nombre de **SDEs** y, más recientemente, Sistemas Difusos Evolutivos Multi-Objetivo (SDEMOs - *Multi-Objective Evolu-*

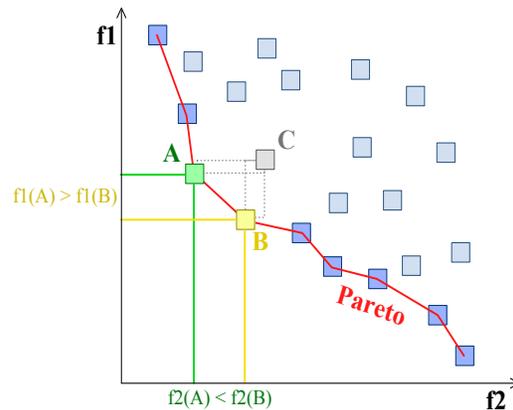


Figura 1.4: Ejemplo de un *Frente de Pareto*.

tionary Fuzzy Systems), siendo una de las metodologías más populares para el diseño de estos SBRDs [28, 33, 50–53].

Una amplia serie de publicaciones especializadas sobre este tipo de sistemas [39, 54–60], así como la existencia de diversas páginas web con completas revisiones bibliográficas sobre la temática ⁴ muestran el interés existente en este campo de investigación. Dicho interés queda también confirmado al realizar una búsqueda bibliográfica en la *Thompson Corporation ISI Web of Knowledge* que localiza un total de 8994 publicaciones científicas⁵ distribuidas anualmente tal y como muestra la Fig. 1.5, y que han sido citadas una media de 6,41 veces.

Actualmente los SDEs están siendo ampliamente utilizados para tratar la problemática del equilibrio precisión-interpretabilidad de los SBRDs [33, 50, 51]. Su éxito se debe tanto a la flexibilidad que presentan los algoritmos (permiten codificar en un mismo cromosoma una o varias partes del SBRD) como a su capacidad de tener en cuenta, de forma simultánea, múltiples criterios de optimización (evolución simultánea de múltiples soluciones en un mismo *Frente de Pareto*).

Es necesario señalar que la búsqueda de dicho equilibrio entre precisión e interpretabilidad en el modelado difuso no es algo ligado exclusivamente al uso de AGs. Otros muchos autores han planteado técnicas muy diversas como la metodología HILK++ (*Highly Interpretable Linguistic Knowledge*), reducciones

⁴<http://sci2s.ugr.es/gfs/>, <http://sci2s.ugr.es/moefs.review/>,
<http://www.iet.unipi.it/m.cococcioni/emofrbss.html>

⁵Fecha de consulta: 13 Mayo 2013. Términos clave propuestos en [33]

TS = [("GA-" OR "GA based" OR evolutionary OR "genetic algorithm*" OR "genetic programming" OR "evolution strate*" OR "genetic learning") AND ("fuzzy rule*" OR "fuzzy system*" OR "fuzzy neural" OR "neuro-fuzzy" OR "fuzzy control*" OR "fuzzy logic cont*" OR "fuzzy class*" OR "fuzzy if" OR "fuzzy model*")]

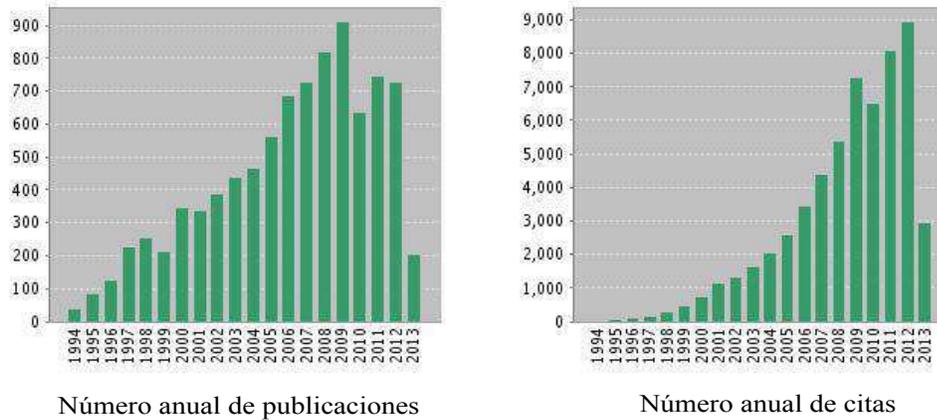


Figura 1.5: Estadísticas de publicaciones sobre SDE según ISI WOK.

basadas en transformaciones ortogonales y algunas otras [61–65].

1.3. Objetivos

A partir de lo visto en las secciones previas, que trata de ilustrar un primer mapa de situación sobre el diseño de los SBRDs y el dilema asociado a la búsqueda de equilibrio entre precisión e interpretabilidad, en esta tesis se busca poner en valor las capacidades interpretativas heredadas de la Lógica Difusa por este tipo de sistemas, manteniendo o incluso mejorando la precisión.

Se plantean así dos líneas maestras de esta tesis:

1. **Diseño de medidas y metodologías para mejorar el modelado de los SBRDs en términos de su interpretabilidad**, sin renunciar a su exactitud, o incluso tratando de mejorarla. Dado que ambos aspectos son contradictorios se planteará la necesidad de buscar soluciones balanceadas entre ambos aspectos. De esta forma el problema se convertirá, en parte, en un problema de optimización multi-objetivo.
2. **Diseño de medidas que permitan comparar las semánticas utili-**

zadas en los SBRDs de distinta naturaleza, en concreto lingüísticos y aproximativos. De esta forma se podrá relacionar la semántica utilizada en un SBRD aproximativo con la definida por un experto para un SBRD lingüístico, mejorando así los procesos de extracción de conocimiento a partir de datos. Todo ello sin perder de vista la precisión.

Esta tesis se centra por tanto en el desarrollo de metodologías y medidas que permitan refinar, optimizar y comparar los sistemas SBRDs, lingüísticos o aproximativos, consiguiendo en cualquier caso una mejora en el equilibrio precisión-interpretabilidad.

En los siguientes puntos se presentan **objetivos más específicos** de investigación:

- Revisión crítica de las ideas, definiciones y formulaciones acerca de los conceptos de interpretabilidad y precisión en el ámbito de la Lógica Difusa y, más concretamente, de los SBRDs.
- Revisión de las métricas existentes para evaluar la interpretabilidad semántica de una BR difusa.
- Revisión y análisis de las métricas existentes para evaluar la interpretabilidad semántica de la BD.
- Definición de métricas globales que puedan ser utilizadas en cualquier tipo de SBRD, independientemente de su naturaleza lingüística o aproximativa, y en combinación con las métricas de precisión.
- Definición de métricas de interpretabilidad semántica de un SBRD que no sean computacionalmente muy costosas, de forma que puedan ser integradas en los procesos de optimización multi-objetivo sin que represente un problema.
- Establecer un marco de análisis y comparación de distintos tipos de SBRDs, en base a las medidas de interpretabilidad establecidas en esta tesis, que de soporte al usuario en un proceso de toma de decisiones y le permita decidir en cada momento qué SBRD es más conveniente para su problema.
- Formulación de los algoritmos que permitan validar las métricas propuestas. Ajuste de sus parámetros para obtener SBRDs con un buen equilibrio entre precisión, complejidad e interpretabilidad semántica.

1.4. Organización de la Memoria

El desarrollo de los objetivos definidos para esta tesis, y materializados en la presente memoria, se ha organizado en 5 capítulos y 2 apéndices, pasándose a describir brevemente el contenido de cada uno de ellos.

En el **Capítulo 2** se hace un repaso al estado del arte del modelado difuso y se hace un repaso de los diferentes enfoques existentes en la literatura científica sobre qué es y cómo evaluar el concepto de interpretabilidad.

En el **Capítulo 3** se define una nueva métrica de interpretabilidad semántica a nivel de BR, llamada *Inter_C* y definida como la agregación de medidas simples de interpretabilidad, basadas en conceptos de similitud y complejidad. Se presenta una metodología general de post-procesamiento, basada en selección genética de reglas difusas, que va a permitir reducir la complejidad de los SBRDs basados en datos, mejorando así su interpretabilidad sin renunciar a su exactitud, e incluso mejorándola.

La metodología propuesta se valida utilizando múltiples conjuntos de datos del repositorio del proyecto KEEL y diversos algoritmos de modelado neuro-difusos, uno lingüístico (NEFPROX - Neuro-Fuzzy Function Approximation) y otro aproximativo (FASART - Fuzzy Adaptive System ART based), que permiten generar SBRDs de distinta naturaleza, cada uno de ellos con sus propias características de precisión e interpretabilidad.

Posteriormente en el **Capítulo 4** se presentan nuevas medidas de interpretabilidad semántica a nivel de BR y a nivel de BD, llamadas RMI y GM3M EXTENDIDA respectivamente. Ambas métricas pueden ser utilizadas tanto en SBRDs lingüísticos como aproximativos, lo que permite generar un marco de comparación local que permitirá al usuario seleccionar en cada momento el tipo de modelado que más se ajuste a sus necesidades. El post-procesamiento presentado en el capítulo 3 será ampliado para poder realizar, junto con la selección de reglas, un ajuste de las particiones difusas, formulando un problema de optimización multi-objetivo.

La metodología propuesta se valida utilizando los mismos conjuntos de datos del repositorio KEEL que en el Capítulo 3, y cuatro algoritmos de modelado, dos lingüísticos (NEFPROX y L-IRL - *Linguistic Iterative Rule Learning*), y dos aproximativos (FASART y S-IRL - *Scatter Iterative Rule Learning*). Dichos algoritmos generan SBRDs de distinto tipo, que permitirán validar la propuesta en distintos entornos de aplicación.

Finalmente se incluye un **Capítulo 5** de conclusiones que resume los resul-

tados obtenidos en este trabajo de tesis. Además se comentan algunos aspectos sobre trabajos actualmente en marcha.

También se incluye una **recopilación bibliográfica** que recoge las contribuciones más destacadas en la materia estudiada y una serie de **apéndices** con aspectos teóricos básicos sobre la Lógica Difusa, los SBRDs y los AEMOs más representativos.

Capítulo 2

Estado del arte

El contenido de este capítulo, tras una breve introducción al campo de la Lógica Difusa, se centra en describir el modelado de los SBRDs y la problemática asociada a la búsqueda de un buen equilibrio entre las propiedades de precisión e interpretabilidad de dicho sistemas. Se analizará el concepto de interpretabilidad difusa y se revisarán algunas de las formulaciones más novedosas propuestas recientemente con el objetivo de poder evaluar el alcance de dicho concepto. Finalmente se presenta una de las estrategias de modelado difuso más utilizada en la actualidad para conseguir SBRDs que sean a la vez precisos e interpretables: los SDEs.

2.1. Lógica Difusa: Modelado

La Real Academia Española ¹ define la palabra Lógica como: “Ciencia que expone las leyes, modos y formas del conocimiento científico”. Distingue a su vez distintos tipos de lógica entre los que se encuentra la **Lógica Difusa o Borrosa** que define como: “La que admite una cierta incertidumbre entre la verdad o falsedad de sus proposiciones, a semejanza del raciocinio humano”.

La Lógica Difusa es por tanto una disciplina capaz de manejar información imprecisa, que otorga grados de cumplimiento (o de verdad) de manera muy similar a la utilizada en el lenguaje natural. En la primera sección de este capítulo

¹<http://lema.rae.es/drae/?val=logica>

se va a realizar un breve repaso histórico a los principales hitos de la Lógica Difusa, hasta llegar a la relevancia e impacto que tiene hoy en día en los distintos campos de actividad tanto técnico-científica como industrial.

2.1.1. Lógica Difusa: Hitos y situación actual

Formalmente la **Lógica Difusa** es una extensión de la lógica booleana en la cual se emplean conjuntos difusos [66] capaces de representar grados de verdad o posibilidad, ya que en el mundo hay hechos que no siempre son completamente falsos o verdaderos. El Apéndice A de esta memoria presenta brevemente los fundamentos teóricos básicos de esta disciplina.

En el desarrollo de la Lógica Difusa desde sus orígenes hasta ahora se pueden destacar una serie de hitos:

1965 Lofty A. Zadeh, ingeniero en la Universidad de Berkeley (California), establece los fundamentos de la teoría de conjuntos difusos [66].

1975 Mamdani y Assilian desarrollan el primer controlador difuso para una máquina de vapor en el Reino Unido [67].

1980 F.L. Smidth & Co. realizan la primera implantación real de un controlador difuso en una planta cementera en Dinamarca [68].

1985 Takagi y Sugeno desarrollan la primera aproximación para construir reglas difusas a partir de datos de entrenamiento que será el punto de partida para investigar la identificación de modelos difusos [36].

1987 y siguientes Se desarrollan diversas aplicaciones industriales que utilizan Lógica Difusa: aires acondicionados (Mitsubishi), aspiradoras (Panasonic), sistemas de transmisión automática de coches (Nissan, Subaru y Mitsubishi), lavadoras (Matsushita, Hitachi), videocamaras (Panasonic, Sanyo, Fisher, Canos), ascensores (Fujitec, Toshiba), televisiones (Sony), [26] etc.

Actualidad A partir de los años 90 y hasta nuestros días la Lógica Difusa se ha establecido como un campo de investigación multidisciplinar, con gran actividad en el desarrollo teórico y de aplicación. La Fig. 2.1 muestra los resultados de la búsqueda bibliográfica realizada en la *Thompson Corporation ISI Web of Knowledge* sobre la aparición de publicaciones con el término

“*Fuzzy Logic*”. La figura muestra la distribución anual de las 25668 publicaciones científicas ² realizadas desde el año 1990, así como la distribución de las 6121 relacionadas con sus aplicaciones. Por otra lado la Fig. 2.2 muestra los resultados obtenidos cuando la búsqueda se realiza en la web de patentes *Espacenet*, en la que se obtiene que desde 1990 el número total de patentes solicitadas que contienen el término *Fuzzy Logic* es de 2380 ³.

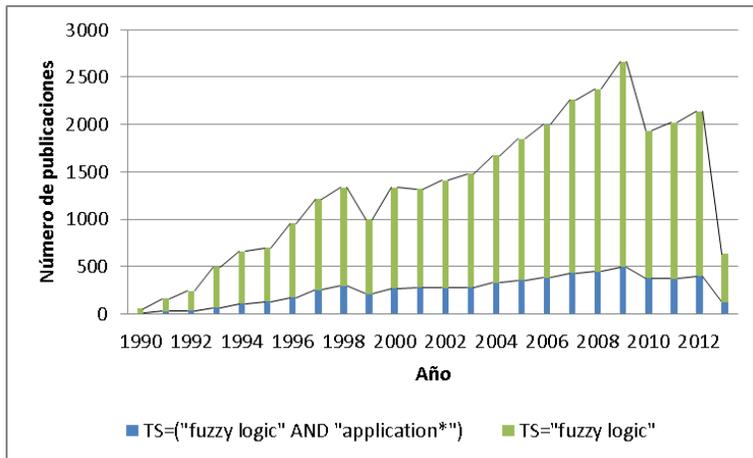


Figura 2.1: Impacto actual de la Lógica Difusa según publicaciones en ISI WOK.

2.1.2. Modelado difuso

Como ya se ha visto en el Capítulo 1 existen múltiples alternativas para modelar un sistema, entre las cuales se encuentran el **Modelado Difuso**. Este tipo de modelado busca aplicar los principios de la Lógica Difusa al modelado de sistemas con el objetivo de obtener modelos que no solo sean precisos sino que, de acuerdo a sus principios teóricos, también sean capaces de explicar el por qué del comportamiento del sistema modelado de forma comprensible.

Una alternativa para realizar modelado difuso es mediante el uso de los llamados SBRDs. Se trata de “*sistemas expertos*” capaces de efectuar un razonamiento difuso basado en un conjunto de reglas del tipo SI-ENTONCES. Estos sistemas

²Fecha de consulta: 23 Junio 2013

³Fecha de consulta: 23 Febrero 2013

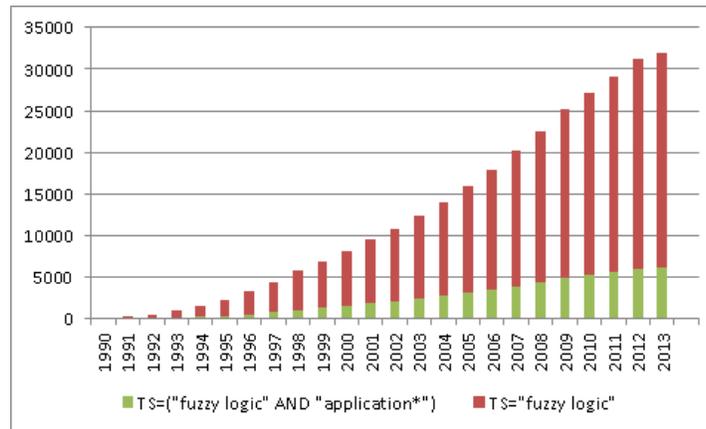


Figura 2.2: Impacto actual de la Lógica Difusa según patentes solicitadas en ESPACENET.

deben gran parte de su éxito a la capacidad de representar conocimiento difuso o impreciso que es, o debería ser, acorde a como el ser humano maneja las ideas y como razona de forma aproximada. El Apéndice A presenta brevemente los principales componentes que pueden formar parte de la estructura de un SBRD, así como una breve descripción de cada uno de ellos.

Analizando las propiedades que son deseables obtener en un SBRD, está claro que para evaluar sus prestaciones es fundamental medir su calidad en términos de [29]:

- **Precisión**, o capacidad de representar fielmente un sistema real con la mayor exactitud posible.
- **Interpretabilidad**, o capacidad de ser entendido o comprendido por un experto.

Los algoritmos de modelado de los SBRDs pueden ser clasificados en dos categorías, en función del objetivo principal que guía el proceso de modelado difuso [28,31–33]:

1. **Modelado difuso preciso** cuyo principal objetivo es obtener modelos tan precisos como sea posible. En general las reglas son aproximadas a partir de datos, aunque el conocimiento experto también puede ser considerado.

Estos sistemas minimizan el error pero suelen tener un bajo nivel de interpretabilidad.

2. **Modelado difuso lingüístico** cuyo principal objetivo es obtener sistemas con un buen nivel de interpretabilidad aunque se pierda cierta precisión. En general las reglas son generadas utilizando tanto conocimiento experto como datos.

Ambos tipos de modelado tienen sus propias ventajas y desventajas en cuanto a los niveles de precisión e interpretabilidad que son capaces de alcanzar, por lo que un aspecto fundamental es llegar a conseguir un equilibrio o compromiso entre ambas características [31, 32, 38, 39]. Es decir, conseguir un modelo difuso con una precisión suficiente e interpretabilidad adecuada.

Si se analizan los trabajos publicados se puede comprobar que hay tres grandes líneas de trabajo que buscan conseguir SBRDs con estas prestaciones (Fig. 2.3):

1. Dotando de mayor precisión al modelado difuso lingüístico, sin perder su interpretabilidad [32]. Una forma de conseguirlo es extendiendo el diseño de los modelos [69–75] o cambiando la estructura de las reglas [44, 76–78] haciendo que sean más flexibles.
2. Mejorando la interpretabilidad del modelado difuso preciso, sin disminuir su grado de exactitud [31]. Una forma de obtener dicha mejora es reduciendo la complejidad de los modelos [47, 64, 79–89].
3. Desarrollando nuevos algoritmos de modelado que sean capaces de generar modelos con un buen equilibrio entre ambos conceptos [61, 63, 65, 90–103].

A continuación, en la Sección 2.2 se presenta un análisis más extenso de la problemática que rodea la búsqueda de equilibrio precisión-interpretabilidad en los SBRDs.

2.2. Equilibrio precisión-interpretabilidad

Como se ha visto anteriormente, las principales características asociadas y deseadas para un SBRD son su precisión y su interpretabilidad, dos aspectos contradictorios/opuestos entre los cuales es necesario alcanzar un equilibrio. Es por tanto necesario ver como definir y evaluar/medir cada una de esas propiedades.

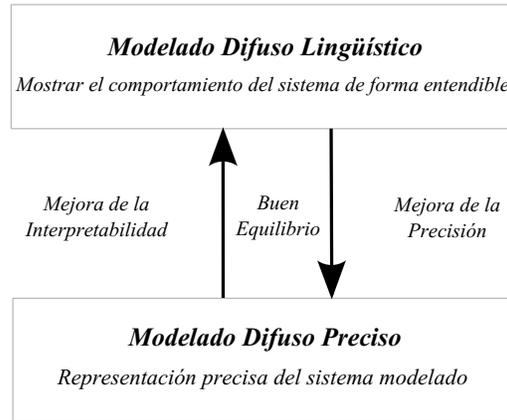


Figura 2.3: Búsqueda del equilibrio precisión-interpretabilidad en modelado difuso [40, 41].

2.2.1. Precisión

Un SBRD es tanto más **preciso** cuanto más se aproxima su comportamiento al sistema real que modela. La medida más utilizada para caracterizar la precisión es el *error*, de forma que un buen conjunto de reglas difusas, junto con el resto de componentes que aparecen en un SBRD (fuzzificador, sistema de inferencia, etc.), deben minimizar las formulaciones que evalúan los errores cometidos. Según el tipo de modelos sobre el que se desea calcular el error las fórmulas a aplicar pueden variar:

- **Regresión:** para este tipo de modelos la fórmula más ampliamente utilizada es la del Error Cuadrático Medio (*ECM* - *Mean Squared Error*):

$$ECM = \frac{1}{N} \sum_{i=1}^N (Y_i - Y_i')^2 \quad (2.1)$$

donde N es el número de ejemplos, Y_i la salida obtenida e Y_i' la salida esperada. Para mejorar la precisión de un modelo es necesario minimizar el valor del error. Otras fórmulas utilizadas también para evaluar la precisión de un modelo de regresión son las mostradas en la Tabla 2.1 [20].

Algunas de las publicaciones que utilizan la fórmula mostrada en la Ec. 2.1

Medida de precisión	Fórmula
Raíz Cuadrada del Error Cuadrático Medio (<i>Root Mean Squared Error</i>)	$\sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - Y'_i)^2}$
Error Absoluto Medio (<i>Mean Absolute Error</i>)	$\frac{1}{N} \sum_{i=1}^N Y_i - Y'_i $
Error Cuadrático Relativo (<i>Relative Squared Error</i>)	$\frac{\sum_{i=1}^N (Y_i - Y'_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y}_i)^2}, \bar{Y}_i = \frac{1}{N} \sum_{i=1}^N Y_i$
Error Absoluto Relativo (<i>Relative Absolute Error</i>)	$\frac{\sum_{i=1}^N Y_i - Y'_i }{\sum_{i=1}^N Y_i - \bar{Y}_i }, \bar{Y}_i = \frac{1}{N} \sum_{i=1}^N Y_i$

Tabla 2.1: Medidas de precisión para regresión.

para evaluar la precisión de un SBRD predictor son [69–72, 76–78, 80, 85, 86, 88, 90, 92–95, 97, 100, 102, 104, 105].

- **Clasificación:** en este tipo de problemas las métricas más utilizada son el porcentaje de ejemplos Clasificados Correctamente (*CC*):

$$CC = \frac{\sum_{i=1}^N (corr_i | corr_i = 1 \text{ si } \hat{C}_i = C_i, 0 \text{ en otro caso})}{N} * 100 \quad (2.2)$$

o el porcentaje de ejemplos Clasificados Incorrectamente (*CI*):

$$CI = \frac{\sum_{i=1}^N (err_i | err_i = 1 \text{ si } \hat{C}_i \neq C_i, 0 \text{ en otro caso})}{N} \quad (2.3)$$

donde C_i es la clase observada, \hat{C}_i la clase inferida y N el número total de ejemplos. Para mejorar la precisión de un modelo es necesario maximizar el valor del número de clases bien clasificadas o minimizar las clases mal clasificadas [20].

Algunas de las publicaciones que utilizan las fórmulas mostradas en las Ecs. 2.2 o 2.3 para evaluar la precisión de un SBRD clasificador son [43, 44, 47, 61, 62, 82, 91, 98, 103, 106].

2.2.2. Interpretabilidad

La otra característica principal deseable de un SBRD es su capacidad de explicación sobre el comportamiento del sistema que está modelando de una forma

comprensible por una persona, es decir su **interpretabilidad**.

Históricamente los autores han utilizado distinta terminología para referirse al concepto de interpretabilidad dentro de los sistemas difusos. Así, palabras como legibilidad, transparencia, entendibilidad, comprensibilidad, inteligibilidad, etc. han sido ampliamente usadas para referirse a un mismo concepto [38].

Por otro lado, tal y como está definido el concepto de interpretabilidad es subjetivo, ya que depende directamente de la persona que hace la interpretación del sistema, de su conocimiento del mismo y de la experiencia previa. Por lo tanto, no es nada fácil ni definir, ni formalizar de forma única el concepto de interpretabilidad [29, 30, 45, 46].

Centrando el análisis en el ámbito de la interpretabilidad de los SBRDs en la literatura científica se pueden encontrar, entre otras, las siguientes definiciones sobre el concepto de interpretabilidad:

- Bodenhofer y Bauer en [107] definen formalmente la interpretabilidad de un sistema como “*la posibilidad de estimar su comportamiento a partir de la lectura y comprensión de la descripción de su base de reglas*” y establecen una relación directa entre la interpretabilidad y las propiedades de las particiones de la variable lingüística.
- Mencar y Fanelli en [45] establecen que “*Un modelo es interpretable si su comportamiento es inteligible, es decir, que puede ser fácilmente percibido y comprendido por un usuario*”.
- Mencar y otros en [62] establecen que una base de conocimiento es interpretable si “*la semántica explícita incluida en el modelo es cointensiva con la semántica implícita inferida por el usuario mientras lee las reglas*”. El concepto de *cointensión* se define en [108] como “*una medida de la proximidad de las relaciones de entrada/salida del objeto de la modelización y el modelo. Un modelo es cointensivo si la proximidad es alta*”.
- Posteriormente Gacto y otros en [29] definen la interpretabilidad como “*La capacidad de expresar el comportamiento del sistema de una manera entendible*”. Establecen además que “*es una propiedad subjetiva que depende de varios factores*”, por lo que la analizan desde dos puntos de vista, la complejidad del modelo y la semántica asociada a las funciones de pertenencia.

Aunque no exista una única definición del concepto de interpretabilidad para un SBRD, algo que si parece estar claro es que su evaluación va a estar muy

relacionada con las características, tanto individual como de conjunto, de sus componentes: reglas, variables, conjuntos difusos, sistema de inferencia, operadores difusos, etc.

De forma individual, un primer aspecto clave de la interpretabilidad de un SBRD es el **tipo de reglas** que utiliza, ya que la forma del consecuente y/o el tipo de semántica del modelo repercute directamente sobre el grado de interpretabilidad. En este sentido se distinguen tres tipos principales de reglas: lingüísticas, aproximativas y TSK [33]. La estructura concreta de las mismas puede encontrarse en el Apéndice A de esta memoria.

Los dos tipos de reglas difusas más populares son las *reglas lingüísticas* tipo Mamdani [34] y las *reglas TSK* [36]. La principal diferencia entre ambos tipos de reglas está en la forma del consecuente: en las primeras es una clase o conjunto difuso que se corresponde con una variable lingüística, mientras que en las segundas es una función polinomial dependiente de los antecedentes. Un tercer tipo son las *reglas basadas en particiones aproximativas* [35], que se diferencian de las primeras en que cada regla tiene su propia semántica.

De entre todas ellas las reglas lingüísticas o Mamdani son reconocidas universalmente como las más interpretables, al utilizar expresiones lingüísticas tanto en los antecedentes como en los consecuentes de las reglas. Le siguen las reglas aproximativas, siendo las reglas TSK las menos interpretables.

Sin embargo, a la hora de formular métricas de evaluación de interpretabilidad de un SBRD no existe aún consenso en la comunidad científica. Es por tanto necesario realizar un estudio más extenso sobre los distintos índices de interpretabilidad propuestos en la literatura actualmente, estudiando su planteamiento y buscando qué mejoras podrían realizarse sobre ellos para conseguir SBRDs que sean a la vez precisos e interpretables.

2.3. Interpretabilidad: Evolución de un concepto

Una vez revisada la dificultad para definir de forma única el concepto de interpretabilidad de un SBRD, en esta sección se van a analizar las principales métricas de interpretabilidad propuestas actualmente en la literatura científica, así como algunos trabajos de especial interés publicados sobre la temática.

A lo largo de los últimos años las métricas utilizadas para evaluar la interpretabilidad de los SBRDs han evolucionado desde unas métricas muy simples

que asociaban el concepto de interpretabilidad directamente con la complejidad del sistema [43, 44] hasta métricas más complejas relacionadas con la semántica [62, 76, 77, 109].

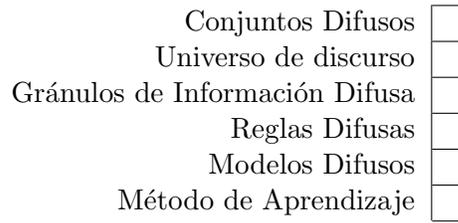
Se presentan a continuación algunas de las taxonomías sobre interpretabilidad presentes en la literatura científica actual, que serán utilizadas posteriormente para clasificar las métricas más utilizadas para su evaluación. La revisión de las principales métricas utilizadas y las formulaciones asociadas a cada una de las métricas será realizada a continuación, tomando primero en consideración las métricas basadas en complejidad y posteriormente aquellas basadas en semántica.

2.3.1. Taxonomías de interpretabilidad

Teniendo en cuenta que el objetivo es poder clasificar las distintas métricas y formulaciones más utilizadas en la literatura científica para evaluar el concepto de interpretabilidad dentro del campo de la Lógica Difusa, se van a presentar a continuación las principales taxonomías publicadas recientemente sobre la temática. Se trata concretamente de los cuatro trabajos básicos, ya introducidos en la Sección 1.2.1, que tratan de organizar todas las ideas sobre interpretabilidad presentadas hasta ese momento, recopilando las ideas más representativas y los distintos puntos de vista de la literatura especializada [29, 30, 45, 46]. A continuación se amplía la descripción de las ideas presentadas en dichas publicaciones.

Mencar y Fanelli en [45] presentan un completo estudio sobre las **restricciones de interpretabilidad** más utilizadas para realizar modelado difuso, recopilando las ideas más representativas existentes en la literatura especializada. En función del componente al que afectan las restricciones se clasifican en uno de los siguientes seis niveles: (1) Conjuntos Difusos; (2) Marco del Conocimiento; (3) Granularidad de la Información Difusa; (4) Reglas Difusas; (5) Sistemas Difusos y (6) Adaptación del Modelo Difuso.

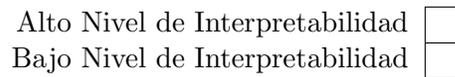
Con el objetivo de representar visualmente la clasificación de las métricas de interpretabilidad a explicar posteriormente se va a usar el siguiente esquema para representar la taxonomía presentada en [45]:



Zhou y Gan en [46] centran su estudio en la interpretabilidad del modelado de los sistemas difusos y establecen dos niveles de interpretabilidad en función del componente del SBRD sobre el que se esté hablando:

- **Bajo Nivel de Interpretabilidad**, que engloba todos los aspectos directamente relacionados con los conjuntos difusos.
- **Alto Nivel de Interpretabilidad**, que contiene todas las propiedades relacionadas con las reglas difusas.

Al igual que en el caso anterior el esquema a utilizar en este caso para clasificar posteriormente las métricas es el siguiente:



Otro punto de vista complementario a los anteriores y centrado en la interpretabilidad de los SBRDs es descrito por Alonso y otros en [30]. La estructura conceptual que se presenta en este artículo extiende la clasificación mostrada en [46], definiendo dos nuevos niveles con elementos y conceptos adicionales:

- **Descripción** (Legibilidad de la estructura del sistema): relacionado con la descripción global de la estructura del SBRD y su facilidad de lectura, que engloba los dos niveles de interpretabilidad (alto y bajo) establecidos en [46]. En función del componente involucrado las métricas de esta categoría se clasifican en uno de los siguientes seis grupos: (1) Base de Reglas; (2) Regla Difusa Lingüística; (3) Proposición Lingüística; (4) Partición Lingüística; (5) Partición Difusa; y (6) Conjunto Difuso.
- **Explicación** (Comprensión del sistema): tiene en cuenta aquellos factores que afectan al comportamiento del SBRD, como el mecanismo de inferencia, los operadores de agregación, unión y disyunción, tipo de defuzzificación y tipo de regla.

El esquema utilizado para representar posteriormente la taxonomía propuesta en [30] es el siguiente:

Descripción	Explicación
Base de Reglas	<input type="checkbox"/>
Regla Difusa Lingüística	<input type="checkbox"/>
Proposición Lingüística	<input type="checkbox"/>
Partición Lingüística	<input type="checkbox"/>
Partición Difusa	<input type="checkbox"/>
Conjunto Difuso	<input type="checkbox"/>

Finalmente Gacto y otros en [29] presentan una taxonomía centrada en la interpretabilidad de los SBRDs lingüísticos, basada en la estructura de este tipo de sistemas. Los autores clasifican las medidas en base a dos criterios:

1. Tipo de interpretabilidad que se quiere medir. Se distingue así entre:
 - a) Interpretabilidad basada en **complejidad** para la cual una reducción en la complejidad del sistema hace que este sea más fácil de entender, y por tanto más interpretable.
 - b) Interpretabilidad basada en **semántica** que tiene en cuenta las características semánticas asociadas a las particiones difusas, es decir a su significado en el mundo real.
2. Elemento del sistema sobre el que se mide. Se distinguen dos niveles: entre métricas calculadas sobre:
 - a) Partición Difusa o **Base de Datos** (BD).
 - b) **Base de Reglas** (BR).

Se consigue así separar las medidas de interpretabilidad del SBRD en cuatro cuadrantes, por lo que el esquema de representación de la taxonomía será:

	BR	BD
Complejidad	<input type="checkbox"/>	<input type="checkbox"/>
Semántica	<input type="checkbox"/>	<input type="checkbox"/>

2.3.2. Interpretabilidad basada en complejidad: Formulaciones

A continuación se revisan algunas de las formulaciones más frecuentemente utilizadas en las publicaciones científicas para evaluar el concepto la interpretabilidad de un SBRD, medida a través de la **complejidad** de sus componentes.

Número de reglas

Una de las primeras métricas utilizadas para medir la facilidad de comprensión de un SBRD fue el **número de reglas** del sistema, el cual debía ser reducido pero adecuado [43, 44]:

$$\text{Interpretabilidad} = \text{NumeroDeReglas} \quad (2.4)$$

En este sentido se supone que cuanto menor sea el número de reglas más fácil de interpretar es el sistema y mejor se entendía su funcionamiento, aunque a medida que se reducía este valor se comprobaba que la precisión empeoraba generalmente. Por ello era necesario alcanzar cierto grado de equilibrio entre ambos conceptos, para lo cual algunos autores utilizaban AG de primera generación, mono-objetivo (utilizando sumas con pesos para integrar ambos objetivos de precisión y complejidad en una única función de adaptación) o multi-objetivo (utilizando una función de adaptación para cada objetivo) [43].

En la Fig. 2.4 se puede ver la clasificación de este índice en función de las taxonomías de interpretabilidad vista previamente.

Actualmente el número de reglas es uno de los índices más utilizados a la hora de analizar la interpretabilidad de un SBRD [70–72, 76–78, 80, 81, 93].

Número total de condiciones

Otra métrica basada en la complejidad del SBRD es el **número total de condiciones** en los antecedentes de las reglas, ya que una regla cuyo antecedente está definido por un subconjunto de las variables del sistema suele ser más fácil de entender [106]:

$$\text{Interpretabilidad} = \text{NumeroTotalDeCondiciones} \quad (2.5)$$

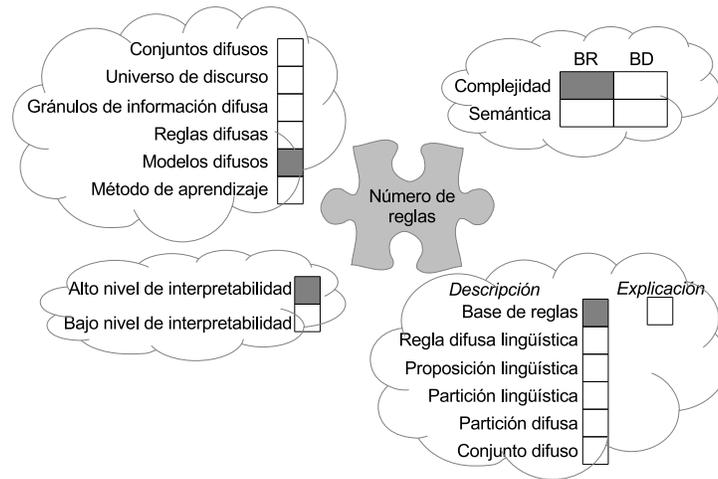


Figura 2.4: Clasificación del índice Número de Reglas según distintas taxonomías.

Para conseguir alcanzar un sistema con una buena precisión y un bajo nivel de complejidad, algunos autores utilizaban un AEMO de primera generación con tres funciones objetivo que evaluaban respectivamente: el error, el número de reglas y el número total de condiciones.

La Fig. 2.5 muestra la clasificación del índice en función de las taxonomías de interpretabilidad. La clasificación es similar a la del número de reglas, salvo en [45] donde el número total de condiciones se considera que es una restricción de interpretabilidad asociada a la regla difusa y no al modelo difuso.

El número total de condiciones en los antecedentes se sigue utilizando para evaluar la interpretabilidad del SBRD, tanto de forma aislada [69,90,92,94,95,99] como en combinación con el número de reglas [47, 79, 82, 96, 98].

Número de variables y Número de conjuntos difusos

Además de las anteriores otros índices utilizados para medir la complejidad de un SBRD, esta vez a nivel de BD, son el **número de variables** o características y el **número de conjuntos difusos** o funciones de pertenencia por variable:

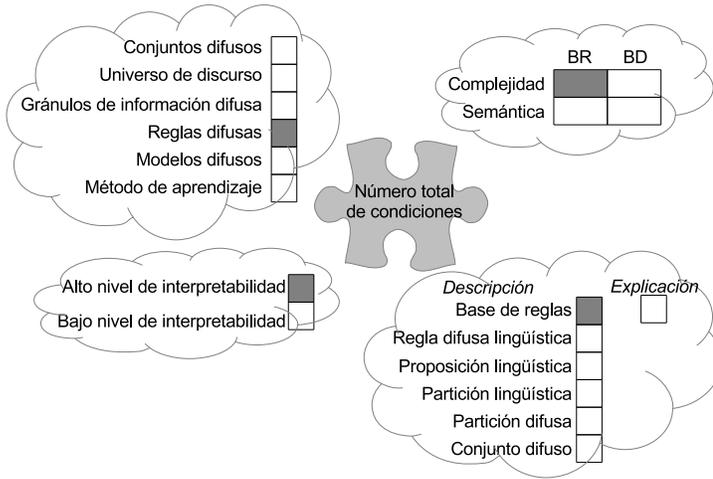


Figura 2.5: Clasificación del índice Número Total de Condiciones según distintas taxonomías.

$$\text{Interpretabilidad} = \text{NumeroDeVariables} \quad (2.6)$$

$$\text{Interpretabilidad} = \text{NumeroDeConjuntosDifusos} \quad (2.7)$$

En ambos casos el número de elementos debe ser reducido pero adecuado, ya que según aumenta el valor del índice más difícil resulta interpretar el SBRD. Se ha establecido como recomendación que el número de conjuntos difusos no debería superar el límite de 7 ± 2 [42, 110].

La Fig. 2.6 muestra la clasificación del índice número de variables en función de las taxonomías de interpretabilidad, y la Fig. 2.7 hace lo mismo para el índice número de conjuntos difusos. Aunque en [29] ambos índices son clasificados en el mismo cuadrante no ocurre lo mismo en el resto de taxonomías, que clasifican cada uno de los índices en una categoría distinta.

La mayoría de los autores utilizan estos índices en combinación con algunos de los anteriores:

- Número de variables combinado con el número de reglas: [61, 63].

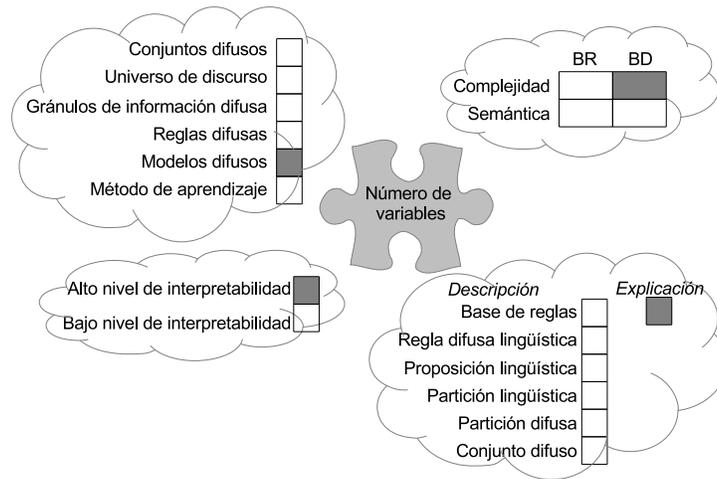


Figura 2.6: Clasificación del índice Número de Variables según distintas taxonomías.

- Número de conjuntos difusos combinado con el número de reglas: [79, 83–86, 88, 89, 97, 102, 105, 111].
- Número de conjuntos difusos combinado con el número total de condiciones: [65, 103].

2.3.3. Interpretabilidad basada en semántica: Formulaciones

Además de los índices basados en complejidad, presentados en la sección anterior, también hay trabajos y autores que incorporan o se centran en evaluar la interpretabilidad semántica del SBRD, tanto a nivel de BD como a nivel de BR.

Restricciones Semánticas en la Base de Datos

Con el objetivo de facilitar la asociación entre los términos lingüísticos y las funciones de pertenencia del SBRD, mejorando así su interpretabilidad, algunos algoritmos de modelado imponen ciertas **restricciones semánticas** a la hora de diseñar las **particiones difusas**.

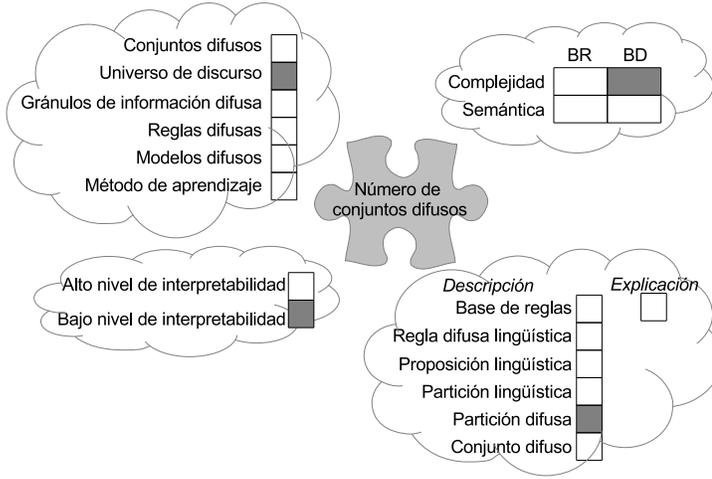


Figura 2.7: Clasificación del índice Número de Conjuntos Difusos según distintas taxonomías.

Uno de los primeros en proponerlo fue Oliveira en [42], donde se establecían, además de un número de funciones de pertenencia reducido (Ec. 2.7), las siguientes propiedades a tener en cuenta en la BD:

- **Normalización** de las funciones de pertenencia, de forma que el grado de pertenencia tome siempre valores en el dominio $[0,1]$, existiendo al menos un punto con grado de pertenencia igual a 1. La formulación asociada a este concepto es:

$$\exists v \in V \forall i=1,2,\dots,n \mu_i(v) = 1,0 \quad (2.8)$$

siendo v una variable definida en V , n el número de funciones de pertenencia y μ_i la i -ésima función de pertenencia.

- **Cobertura y solapamiento** del universo de discurso. Es necesario utilizar funciones de pertenencia que cubran todo el rango de las variables, con un solapamiento entre ellas que garantice que para cualquier punto del rango siempre exista un conjunto borroso tal que su grado de pertenencia al mismo

sea mayor que un grado de cobertura prefijado ϵ . En este caso la formulación asociada es:

$$\forall v \in V \exists i: 1 \leq i \leq n \mu_i(v) > \epsilon \quad (2.9)$$

siendo v una variable definida en V , n el número de funciones de pertenencia y μ_i la i -ésima función de pertenencia.

Posteriormente en [84] se propone un índice de cobertura CI basado en medir la proporción de ejemplos activados (n) respecto al número total de ejemplos disponibles (N) dado un umbral de activación $\alpha \in [0, 1]$:

$$CI_\alpha = \frac{n}{N} \quad (2.10)$$

- **Posición cero natural**, con una función de pertenencia unimodal, convexa y centrada en cero, representando el término lingüístico “cercano a cero”.
- **Distinguibilidad**, de forma que al asociar de forma clara un término lingüístico a cada función de pertenencia los conceptos lingüísticos representados por estas funciones sean distintos unos de otros. Esto se consigue evitando que en un mismo punto del universo de discurso el grado de pertenencia a dos conjuntos distintos sea elevado de forma simultánea. La formulación asociada al concepto de distinguibilidad es:

$$\forall v \in V M_p(L_v(v)) \leq 1 \quad (2.11)$$

siendo v una variable definida en V , L_v todos los conjuntos difusos definidos en V y M_p la suma del grado de pertenencia a todos esos conjuntos difusos.

Otra forma de garantizar la distinguibilidad fue propuesta en [62], donde se establece un umbral σ máximo de solapamiento entre los conjuntos difusos s_i y s_j :

$$\forall v \forall s_i, s_j \in L_v : \frac{s_i \cap s_j}{s_i \cup s_j} \leq \sigma \quad (2.12)$$

Por todo ello, las particiones difusas fuertes con funciones de pertenencia equidistribuidas suelen considerarse habitualmente las más interpretables independientemente de lo que pudiese decir un experto o cuando no lo hay. Un ejemplo de este tipo de particiones puede verse en la Fig. 2.8.

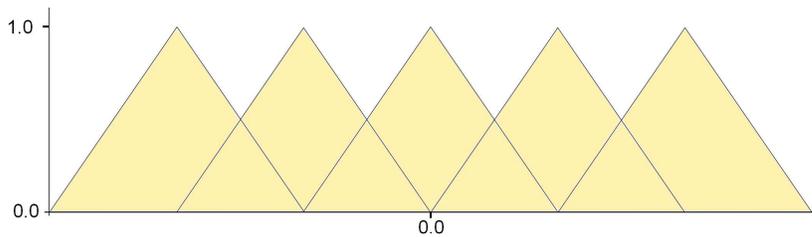


Figura 2.8: Partición difusa fuerte uniformemente distribuida.

Todas estas restricciones semánticas son clasificadas según se muestra en la Fig. 2.9. La doble clasificación en algunas taxonomías las produce la propiedad de normalización que es considerada una propiedad asociada al conjunto difuso en vez de a la partición difusa [30] o al marco de conocimiento [45].

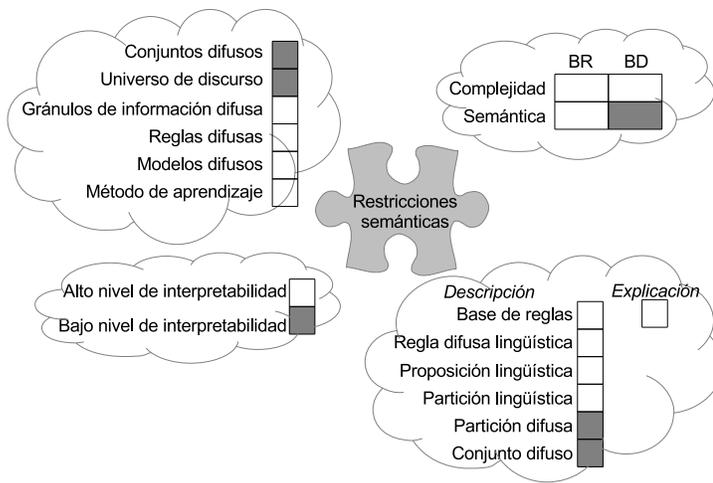


Figura 2.9: Clasificación de las Restricciones semánticas según distintas taxonomías.

Algunas de las publicaciones científicas que buscan garantizar algunas de estas propiedades son: [62, 69, 74, 75, 91, 100, 104, 111, 112].

Consistencia de la Base de Reglas

Por otro lado, con el objetivo de obtener BRs interpretables semánticamente, los índices más utilizados están basados en evaluar su **consistencia**, en términos de número de reglas redundantes y/o incoherentes. En este sentido, y normalmente derivado de un excesivo número de reglas, aparecen las reglas redundantes, cuyas acciones ya están cubiertas por otras reglas, y las reglas incoherentes, que distorsionan el funcionamiento del sistema. Ambos tipos de reglas provocan que el significado de las reglas difusas se vea alterado, por lo que es necesario eliminarlas [30, 61, 75, 83, 86, 88, 89, 96, 100, 101, 111, 113].

Esta propiedad de consistencia se clasifica dentro de las taxonomías de interpretabilidad según se muestra en la Fig. 2.10.

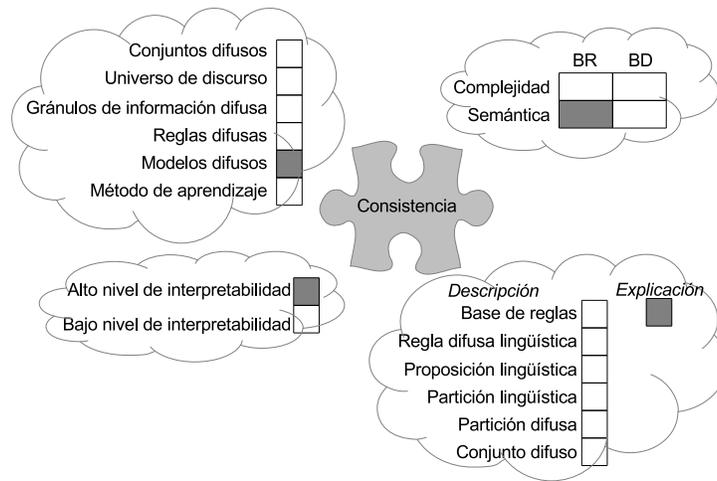


Figura 2.10: Clasificación de la Consistencia según distintas taxonomías.

Las estrategias más utilizadas en los algoritmos de modelado para conseguir localizar y eliminar reglas redundantes e incoherentes se basan en el uso de **medidas de similitud** entre los conjuntos difusos, aunque no es la única forma. Así por ejemplo en [64, 114, 115] se utilizan **transformaciones ortogonales** con el objetivo de establecer un orden dentro del conjunto de reglas que permita posteriormente realizar la selección de reglas, y en [101] se añade un parámetro al proceso de aprendizaje capaz de detectar reglas redundantes a ser eliminadas.

Medidas de Similitud y Consistencia

Una de las medidas de similitud entre conjuntos difusos más utilizada por los algoritmos de modelado difuso para fusionar conjuntos difusos similares, con el objetivo de garantizar la consistencia de la BR, es [89]:

$$Similitud(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{\sum_{j=1}^m [\mu_A(x_j) \wedge \mu_B(x_j)]}{\sum_{j=1}^m [\mu_A(x_j) \vee \mu_B(x_j)]} \quad (2.13)$$

donde A, B son conjuntos difusos, $|\cdot|$ es la cardinalidad del conjunto difuso, \cap y \cup representan respectivamente la intersección y la unión, m es el número de funciones de pertenencia, y por último, \wedge y \vee son los operadores de mínimo y máximo respectivamente.

Posteriormente en [88] la Ec. 2.13 fue extendida para poder evaluar la similitud total del modelo *Similitud**:

$$Similitud^* = \frac{1}{n} \sum_{i=1}^n \left(\frac{\text{máx}(S(A_{ij}, A_{ik}))}{\eta_i - 1} \right) \quad (2.14)$$

donde η_i es el número de conjuntos difusos por cada variable, $j, k \in 1, 2, \dots, \eta_i, j \neq k$, y n es el número de variables de entrada.

La misma estrategia de usar medidas de similitud para evaluar la completitud y la consistencia del sistema, pero adaptando la Ec. 2.13 al dominio continuo:

$$Similitud(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} = \int_{-\infty}^{\infty} \mu_A(x) dx \quad (2.15)$$

fue utilizada en [75], donde se presenta un proceso de generación de reglas basado en algoritmos genéticos y estrategias evolutivas que utiliza, además del número de reglas para medir la complejidad, la medida de similitud anterior para localizar y eliminar reglas redundantes e incoherentes. La misma medida vuelve a ser utilizada por los autores en [111], combinando la búsqueda de consistencia en la BR con la unificación de conjuntos difusos similares para mejorar la distinguibilidad de la BD.

Finalmente, la forma de evaluar la consistencia de la BR a partir de las medidas de similitud es [75, 111]:

$$\begin{aligned}
Cons(R(i), R(k)) &= \exp\left\{-\frac{\left(\frac{SRP(i,k)}{SRC(i,k)} - 1, 0\right)^2}{\left(\frac{1}{SRP(i,k)}\right)^2}\right\} \\
SRP(i, k) &= \min_{j=1}^n S(A_{ij}, A_{kj}) \\
SRC(i, k) &= S(B_i, B_k)
\end{aligned} \tag{2.16}$$

donde $SRP(i, k)$ es definido como la similitud en los antecedentes, $SRC(i, k)$ como la similitud en los consecuentes, siendo R_i, R_k dos reglas difusas y n el número total de variables de entrada.

Otras publicaciones en las que se han utilizado algunas de las medidas de similitud previa para garantizar la consistencia han sido: [64, 86, 102].

Medidas de Similitud y Restricciones Semánticas

Por otro lado las medidas de similitud entre conjuntos difusos no han sido utilizadas sólo para mejorar la consistencia de la BR, sino también para conseguir o mantener algunas de las restricciones semánticas descritas previamente (normalización, cobertura, solapamiento o distinguibilidad) [84, 85, 87, 111].

Así en [111] se unifican conjuntos difusos similares utilizando la siguiente medida de similitud:

$$Similitud(A, B) = \frac{1}{1 + d(A, B)} \tag{2.17}$$

donde d es una medida de distancia entre los conjuntos, con el objetivo de conseguir distinguibilidad. Otro ejemplo lo encontramos en [85] donde se propone la siguiente métrica para medir la completitud y distinguibilidad del SBRD:

$$sim = \frac{\sum_{i=1}^n \hat{\beta}_i + \sum_{l=1}^m \sum_{k=1}^{M_l^a - 1} \check{\beta}_k^l}{n + \sum_{l=1}^m (M_l^a - 1)} \tag{2.18}$$

siendo M_l^a el número de conjuntos difusos de cada variable de entrada x_l , m el número de variables de entrada definidas por más de un conjunto difuso, β_k el número de veces que la similitud de dos conjuntos difusos consecutivos (definidas sobre dichas variables) está dentro de unos límites predefinidos, n el número de variables de entrada definidas por un solo conjunto difuso y β_i el número de

veces que la similitud entre el conjunto difuso (que define dicha variable) y su conjunto difuso universal está dentro de dichos límites. En este caso los autores, para conseguir el equilibrio entre la precisión y la interpretabilidad, utilizan un AEMO de primera generación con cinco funciones objetivo.

2.3.4. Interpretabilidad basada en semántica: Últimas tendencias

En la sección anterior se han mostrado las formulaciones más clásicas utilizadas en la literatura científica para evaluar la interpretabilidad semántica, tanto a nivel de regla como a nivel de BD. Sin embargo la propuesta de nuevas métricas en este ámbito sigue siendo hoy en día un problema abierto, lo que ha dado lugar a que en los últimos años se hayan realizado diversas propuestas de métricas más complejas [29].

En relación a la interpretabilidad semántica a nivel de BR la propiedad de **consistencia** sigue siendo la propiedad más buscada [61, 83, 96]. Sin embargo recientemente se han empezado a utilizar **nuevas medidas**, algunas muy simples de calcular como el número de reglas activadas simultáneamente [76, 116], y otras más complejas, y por tanto computacionalmente costosas, como el índice RBC (*Rule-Based Complexity* [61] para evaluar la consistencia o la cointension [62, 91].

Por su parte a la hora de conseguir la interpretabilidad semántica a nivel de BD, se siguen imponiendo **restricciones semánticas** con el objetivo de conseguir que las particiones difusas sean completas, cubiertas, normales, distinguibles y complementarias. Sin embargo las métricas absolutas asociadas a la forma de evaluar estas propiedades han ido ganando en complejidad [83, 109] a pesar de que la mayoría de ellas obtienen su mejor valor cuando la BD está compuesta por particiones fuertes equidistribuidas, que cumplen con prácticamente todas las propiedades anteriormente indicadas. Por otro lado la principal novedad ha estado en la aparición de **medidas relativas** como GM3M [77] y el índice de integridad I [90].

Se muestra a continuación la formulación asociada a algunas de estas métricas complejas, relacionadas tanto con la BD como con la BR.

Número de reglas activadas simultáneamente

El **número de reglas activadas simultáneamente** es una medida a la que los investigadores no le han prestado especial atención hasta hace solo unos

años y que actualmente vuelve a ser de interés principalmente por su facilidad de cálculo [76, 116].

La definición de esta métrica, orientada a medir la interpretabilidad semántica de la BR, es:

$$AvR_{TG} = \frac{\sum_j^M R_{TG}^j}{P} \quad (2.19)$$

siendo P el número de ejemplos y R_{TG}^j el número de reglas disparadas para el ejemplo j . Al clasificar esta medida según las taxonomías de interpretabilidad previas se obtiene la Fig. 2.11.

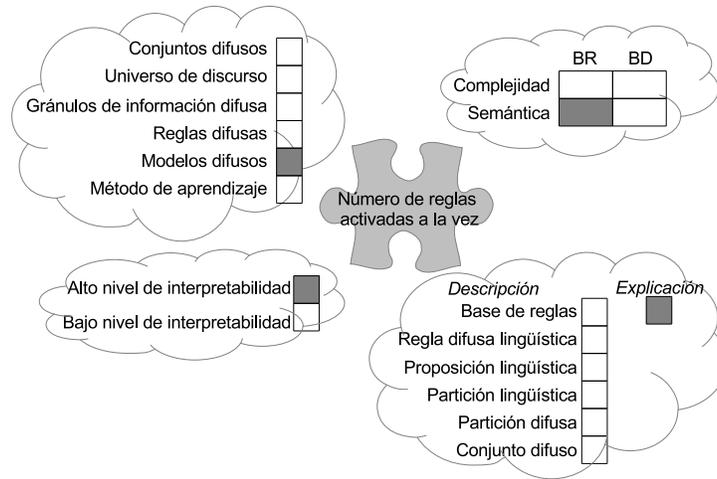


Figura 2.11: Clasificación del Número de Reglas Activadas Simultáneamente según distintas taxonomías.

Índice RBC

Alonso y Magdalena en [61] presentan una nueva métrica para medir la **consistencia** de la BR, llamado RBC. En un análisis previo realizado por los autores en [30] se detectó que las premisas de tipo NOT no eran fácilmente interpretables

por los usuarios, dando preferencia a las tipo OR. Por ello el cálculo de la nueva métrica RBC se basa en la presencia de etiquetas NOT o OR en las premisas difusas:

$$RBC = \sum_{j=1}^{NR} \prod_{a=1}^{NI} \left(2 - \frac{LT_a^j}{NL_a}\right) \quad (2.20)$$

siendo NR el número total de reglas, NI el número total de variables de entrada, LT_a^j el número de términos elementales incluidos en A_a^i (término lingüístico asociado a la variable I_a) y NL_a el número de términos elementales definidos en la partición difusa (de la misma variable I_a). La forma de asignar el valor al parámetro LT_a^j se realiza siguiendo cuatro reglas: (1) Uno para los términos elementales; (2) número de términos elementales combinados con OR; (3) NL_a menos un medio para los términos combinados con NOT; y (4) NL_a cuando la variable I_a no es considerada en la regla.

La clasificación de esta nueva métrica dentro de las taxonomías coincide con la presentada en la Fig. 2.10 para la consistencia de la BR.

Cointensión

El término **cointension** fue introducido por Zadeh en [108], y ha sido posteriormente formalizado por Mencar y otros para ser utilizado en clasificadores basados en reglas difusas [62, 91].

El objetivo de este índice es evaluar como de “cointensiva”⁴ es la semántica explícita contenida en las reglas difusas (funciones de pertenencia, operadores difusos, etc.) al compararla con la semántica implícita inferida por el usuario al leer la BR.

Para ello los autores hacen una minimización lógica basada en tablas de verdad para posteriormente, haciendo uso de la llamada vista lógica y comparando la capacidad de clasificación de la BC original y la minimizada, se calcula la cointensión semántica (SC) según la siguiente ecuación:

$$SC = 1 - \frac{\sum_{i=1}^N (1 - \chi(f(x_i), f'(x_i)))}{N} \quad (2.21)$$

⁴La palabra “cointensiva” se utiliza aquí como traducción del término inglés “cointensive”, aunque como tal el término no existe en la lengua española.

donde $f(x_i)$ y $f'(x_i)$ es la clasificación proporcionada para la entrada x_i por la BC original y la minimizada respectivamente, y N es el número de ejemplos. Si los errores de clasificación obtenidos por ambas BCs son similares (valores altos de SC) eso significa que la semántica explícita de los valores lingüísticos es coherente con los operadores lógicos utilizados en el proceso de minimización, por lo que la BR es “cointensiva” con el conocimiento del usuario. Es decir que, desde el punto de vista de la cointensión, el sistema tiene un alto grado de interpretabilidad.

La clasificación del concepto de cointensión dentro de las taxonomías de interpretabilidad coincide con la presentada en la Fig. 2.10 de la Sección 2.3.3 para la consistencia de la BR y con la presentada en la Fig 2.11 de esta misma sección para el número de reglas activadas simultáneamente. De momento esta métrica solo ha sido definida para su uso en problemas de clasificación, pero no de regresión.

Medidas Absolutas a nivel de Base de Datos

Una vez vistas las nuevas medidas de interpretabilidad semántica a nivel de BR, se van a repasar las relacionadas con la BD, comenzando con las métricas absolutas presentadas en [83,109]. Se consideran medidas absolutas ya que tratan de satisfacer restricciones semánticas como la cobertura, la distinguibilidad o la complementariedad de forma independiente. Casi todas estas medidas absolutas suelen encontrar sus valores máximos en la partición difusa fuerte uniformemente distribuida.

Concretamente Pulkkinen y otros en [83] proponen la siguiente medida para evaluar la **transparencia** de una partición:

$$T = P_{OL} + P_{DC} + P_{MV} \quad (2.22)$$

$$P_{OL} = \frac{1}{n_s} \sum_{i=1}^{n_s} \frac{1}{N_{OV}^i} \sum_{j=1}^{N_{OV}^i} \frac{\lambda_{i,j}}{\chi_i}$$

$$N_{OV}^i = \frac{M_i!}{2(M_i - 2)!}$$

$$P_{DC} = \frac{1}{n_s} \sum_{i=1}^{n_s} \sum_{j=1}^{G_i} \frac{\psi_{i,j}}{\chi_i}$$

$$P_{MV} = \frac{1}{n_s} \sum_{i=1}^{n_s} \delta_i$$

donde P_{OL} es una medida de penalización sobre el solapamiento entre conjuntos difusos, P_{DC} es una penalización a la discontinuidad y P_{MV} es otra penalización relacionada con la cobertura. Además $\lambda_{i,j}$ es la longitud del j -ésimo solapamiento entre dos particiones difusas de la variable de entrada i , n_s es el número de variables seleccionadas y N_{OV}^i es el número de pares de particiones difusas que tienen solapamiento. Por su parte G_i es el número de discontinuidades y $\psi_{i,j}$ es la longitud de la j -ésima discontinuidad presente en la variable de entrada i . Además δ_i es $\frac{\delta_i^* - \alpha_L}{1 - \alpha_L}$ si $\delta_i^* > \alpha_L$ y 0 en otro caso, siendo δ_i^* el máximo valor medio presente en la variable de entrada i y α_L un umbral de cobertura prefijado por el usuario.

En dicha publicación los autores utilizan un AEMO de segunda generación con tres funciones objetivo: dos para evaluar aspectos de precisión y una para evaluar la semántica según el índice T , para conseguir SBRDs con un buen equilibrio precisión-interpretabilidad.

Por su parte Botta y otros en [109] proponen la siguiente medida de interpretabilidad para garantizar el **ordenamiento difuso** y la **distinguibilidad**:

$$\Phi_Q(P) = \frac{\sum_{1 \leq i \leq N-1} \frac{1}{d_{j,i}} \cdot \mu_Q^{d_{j,i}}(Q \leq (A_i, A_j))}{\sum_{1 \leq i \leq N-1} \frac{1}{d_{j,i}}}, i \leq j \leq N \quad (2.23)$$

siendo N el número de conjuntos difusos, $d_{j,i} = |j - i|$ la distancia semántica entre los conjuntos difusos A_j y A_i , Q un índice de ordenamiento difuso y $\mu_Q^{d_{j,i}}$ los grados de pertenencia a los conjuntos difusos definidos para los valores de Q .

Aquí también se utiliza un AEMO de segunda generación con dos funciones objetivo para realizar la adaptación al contexto (adaptación de las funciones de pertenencia usando funciones de escala) en la etapa de post-procesamiento en la que se mejora la BC generada inicialmente.

Medidas relativas a nivel de Base de Datos

Y como principal novedad en los últimos años ha surgido el uso de medidas relativas para medir la interpretabilidad semántica a nivel de BD [77, 90]. Estas medidas son relativas ya que para su cálculo se tiene en cuenta una definición previa de las funciones de pertenencia que se considera interpretable y que puede

ser proporcionada bien por un experto o bien por un proceso de aprendizaje automático que utilice alguna de las medidas absolutas. Estas medidas relativas son compatibles con las medidas absolutas al tener en cuenta como partición difusa de referencia aquella que maximice dichas medidas, cuando no hay conocimiento experto, por ejemplo considerando la partición difusa fuerte uniformemente distribuida.

Gacto y otros en [77] proponen el uso de una medida relativa denominada **GM3M**, que mide la interpretabilidad semántica de una partición difusa ajustada respecto a una partición difusa interpretable previamente definida. Concretamente la métrica se define como la media geométrica de tres valores: el desplazamiento (δ), la amplitud lateral (γ) y la similitud del área (ρ) entre los conjuntos difusos:

$$GM3M = \sqrt[3]{\delta \cdot \gamma \cdot \rho} \quad (2.24)$$

Dicha métrica de interpretabilidad es utilizada junto con un AEMO de segunda generación que busca mejorar simultáneamente tres aspectos de los SBRDs: aumentar su precisión, disminuir su complejidad y mejorar la interpretabilidad semántica. Tal y como está definida la métrica puede ser utilizada sólo con modelos difusos lingüísticos con particiones difusas triangulares.

Posteriormente Antonelli y otros en [90] proponen el siguiente índice relativo I para evaluar de **integridad** de la partición:

$$I = 1 - \frac{D}{D_{MAX}} \quad (2.25)$$

donde $D = \frac{1}{F+1} \sum_f^{F+1} \sum_{j=2}^{T_f-1} |b_{f,j} - \tilde{b}_{f,j}|$ es una medida de cómo de diferentes son, en media, las particiones difusas fuertes iniciales (T) y las modificadas de la variable f . En este caso, b y \tilde{b} son los puntos centrales de las particiones, F el número de variables de entrada y D_{MAX} es el mayor valor de D .

Este índice es similar al índice GM3M explicado previamente, pero adaptado a la metodología de trabajo utilizada en el grupo de investigación. Los autores también utilizan un AEMO de segunda generación, con tres funciones objetivo (precisión, complejidad y semántica) para generar sistemas difusos con un buen equilibrio precisión-interpretabilidad.

Todos estas métricas de interpretabilidad semántica de las BD, tanto absolutas o relativas, se clasifican en las taxonomías de interpretabilidad igual que las restricciones semánticas vistas previamente (Fig. 2.9, Sección 2.3.3).

2.4. Sistemas difusos evolutivos

Un enfoque interesante para alcanzar el tan deseado equilibrio o compromiso precisión- interpretabilidad consiste en utilizar SDEs [28,33,50–53]. Un SDE no es más que un sistema difuso en cuyo proceso de aprendizaje, ajuste o sintonización se utilizan computación evolutiva, que incluye AGs, programación genética y estrategias evolutiva.

En [33] se presenta una taxonomía general de los SDE (Fig. 2.12) en función del componente del SBRD que se codifica en el cromosoma, y sobre el cual se quiere actuar (creando y/o mejorando el componente). Se distinguen así las siguientes líneas de trabajo:

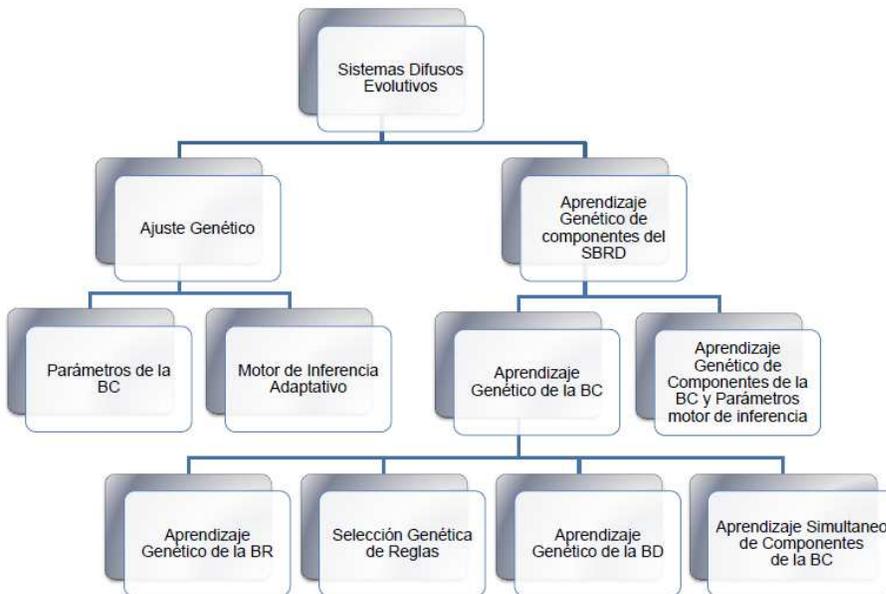


Figura 2.12: Taxonomía de los SDEs [33].

- **Ajuste Genético:** para mejorar las características de los SBRDs sin modificar la BR existente. Normalmente, dada una BC, se aplican los AGs con el objetivo de ajustar la BD [70–72, 77, 80, 82, 83, 85, 109] o alguno de los parámetros del motor de inferencia [76], manteniendo siempre la misma BR.

- **Aprendizaje Genético:** para aprender las reglas difusas o cualquier otro componente del SBRD. Normalmente se distingue entre Selección Genética de Reglas [43, 44, 47, 98, 106], Aprendizaje Genético de la BR [99], Aprendizaje Genético de la BD [97, 117] y Aprendizaje Simultáneo de Componentes de la BC [69, 90, 92–96, 103].

En [50] dicha taxonomía ha sido ampliada, centrandolo en los SDEMOs, es decir en el uso de AEMOs para diseñar sistemas difusos. Dicha taxonomía añade un nivel superior de clasificación referido a la naturaleza del problema a resolver, uno de los cuales es el equilibrio precisión-interpretabilidad.

Relacionado con el problema del equilibrio el artículo concluye que en la mayor parte de los trabajos se utilizan AEMOs de segunda generación, ya que el concepto de “elitismo” es fundamental a la hora de conseguir la convergencia del algoritmo. También concluye que se utilizan más los sistemas Mamdani que los TSK y que la forma de evaluar el concepto de interpretabilidad se ha vuelto cada vez más compleja y completa.

Ajuste de las funciones de pertenencia

Dentro de la categoría de ajuste genético la opción más utilizada es el **ajuste de las funciones de pertenencia**, que se realiza una vez que todos los demás componentes del SBRD han sido definidos. Dicho ajuste consiste en refinar las funciones de pertenencia previamente definidas, aunque no el número de términos lingüísticos en la partición, que permanece fijo desde el principio del proceso. La primera de las contribuciones, realizada en el ámbito de ajuste genético, y la más clásica, fue [118]. Otras contribuciones posteriores que realizan ajuste de particiones difusas son [75, 119–121].

La forma clásica de ajustar las funciones de pertenencia es modificando los parámetros característicos que la definen. Por ejemplo, si consideramos que la función de pertenencia tiene forma triangular (ver Fig. 2.13) se podrían modificar sus parámetros básicos a , b , y c , para cambiar la forma del conjunto difuso asociado a la función de pertenencia, e influir así en el rendimiento del SBRD. Esto mismo ocurre con cualquier otro tipo de función de pertenencia (trapezoidal, gaussiana, etc.).

Sin embargo esta no es la única forma de realizar el ajuste. Por ejemplo en [109] se presupone que la BR está bien definida y que para mejorar el sistema lo único

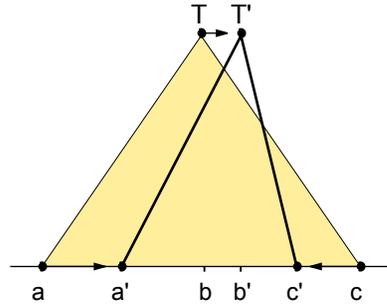


Figura 2.13: Ajuste de una función de pertenencia modificando sus parámetros característicos.

que hay que hacer es ajustar los conceptos de la BD al contexto de aplicación (por ejemplo, el concepto temperatura alta es distinto según el país en el que nos encontremos).

Selección genética de reglas

Otro de los enfoques utilizados dentro de la categoría de aprendizaje genético para mejorar el equilibrio es la **selección genética de reglas**. Dicho enfoque se basa en usar AGs para realizar una selección de reglas difusas cooperativas a partir de un conjunto de reglas candidatas, con el objetivo de obtener un conjunto de reglas más reducido capaces de mejorar el equilibrio precisión-interpretabilidad. Eliminando las reglas irrelevantes, redundantes, erróneas y conflictivas se obtiene un nuevo conjunto de reglas que es más compacto e interpretable, e incluso en ocasiones más preciso.

La primera de las contribuciones realizada en este ámbito, y la más clásica, fue propuesta por Ishibuchi y otros en [44]. En dicha contribución los autores utilizaban una función objetivo mono-objetivo que combina dos medidas para maximizar el número de patrones clasificados correctamente y minimizar el número de reglas. Posteriormente Ishibuchi en [43] utiliza un AEMO con el mismo propósito, y en [106] introduce un tercer objetivo para minimizar el número de condiciones.

Una combinación de ambos enfoques, selección de reglas difusas y ajuste de las funciones de pertenencia, ya ha sido utilizada exitosamente en varias contribuciones [70, 71, 77, 78] con el objetivo de obtener sistemas con un buen equilibrio precisión-interpretabilidad.

Capítulo 3

Reducción de la complejidad y mejora de la interpretabilidad mediante una métrica combinada de interpretabilidad semántica.

Teniendo en cuenta los capítulos anteriores sobre la búsqueda del equilibrio precisión-interpretabilidad en el modelado de los SBRDs y las métricas de interpretabilidad ya propuestas, en este capítulo se va a proceder a desarrollar una primera aproximación de **mejora de la interpretabilidad** de los SBRDs basada en una **reducción de su complejidad**, que permita obtener modelos que sean simultáneamente más interpretables y precisos.

Para guiar la búsqueda, y teniendo en cuenta que dicha reducción puede afectar a la interpretabilidad semántica a nivel de BR, se va a definir una nueva métrica de interpretabilidad basada en la agregación de índices sencillos y fáciles de entender, como son la compactitud, la similitud, la redundancia, la incoherencia y la completitud. Adicionalmente, se va a validar la propuesta bajo diferentes condiciones, utilizando múltiples conjuntos de datos y diversos algoritmos para

el modelado, tanto lingüísticos como aproximativos. Sobre los distintos SBRDs generados se va a aplicar una estrategia evolutiva bi-objetivo (precisión e interpretabilidad) de selección de reglas difusas que va a permitir mejorar sus características y obtener sistemas con un buen equilibrio precisión-interpretabilidad.

El capítulo se organiza como sigue: en la Sección 3.1 se presentan las métricas de complejidad e interpretabilidad utilizadas, esta última generada como una agregación de índices sencillos de interpretabilidad. A continuación, en la Sección 3.2 se describe la metodología utilizada y en la Sección 3.3 se presentan los experimentos realizados y se analizan los resultados obtenidos.

3.1. Agregación de índices de complejidad e interpretabilidad semántica en una métrica común

A continuación se va a definir una nueva métrica de interpretabilidad que va a permitir evaluar la interpretabilidad de un SBRD, teniendo en cuenta tanto la semántica como la complejidad del sistema. La métrica se va a definir mediante la agregación de diversos índices sencillos de interpretabilidad que evalúan, a nivel de BR, conceptos relacionados tanto con la complejidad como con la similitud de sus componentes.

Como ya se ha visto en el Capítulo 2, una forma de obtener SBRD interpretables es consiguiendo que sus componentes, tanto los conjuntos difusos, como las reglas, particiones difusas, estructura, operadores, etc., también sean interpretables. Por ello, el cálculo de la nueva métrica de interpretabilidad se va a plantear como la media ponderada de múltiples índices sencillos de interpretabilidad:

$$Inter_C = MediaAritmetica(\lambda_j * IndiceInterpretabilidad_j) \quad (3.1)$$

Todos los índices de interpretabilidad se ponderan individualmente usando λ_j , que puede tomar cualquier valor entre 0 y 1. 1 significa que el índice tiene una relevancia elevada dentro de la métrica $Inter_C$ y 0 significa que el índice es ignorado. Posteriormente todos los índices individuales se combinan utilizando la *MediaAritmetica* para estimar un único valor de interpretabilidad global del SBRD llamado $Inter_C$.

Así definida la métrica está abierta tanto a la eliminación de índices como a la inclusión de otros nuevos, de forma que se pueda adaptar a las propiedades

de interpretabilidad concretas deseadas en cada momento por el usuario. Además puede ser utilizada tanto con cualquier tipo de SBRD (lingüísticos, aproximativos, etc.) como de partición difusa (triangular, trapezoidal, gaussiana, etc.).

Concretamente en el contexto de los SBRDs, una reducción en la complejidad del sistema implica, en cierta forma, una mejora de su interpretabilidad. Igualmente una reducción en la similitud entre reglas, la eliminación de reglas redundantes e incoherentes y la completitud de las variables son otras formas de mejorar su interpretabilidad. Por lo tanto los índices sencillos de interpretabilidad considerados en esta propuesta, conceptualmente unidos con los marcos de trabajo descritos en [29, 30, 46], son los siguientes:

- **Número de reglas**, como medida de la compactitud del SBRD:

$$Compactitud = Numero\ de\ reglas \quad (3.2)$$

Al reducir el número de reglas el modelo es más simple y más fácil de entender. Es decir, se disminuye la complejidad y se aumenta la interpretabilidad de la BR.

- **Similitud entre reglas**, de forma que a menor similitud mejor *distinguibilidad* en la BR y más fácil de interpretar. El cálculo de este índice se realiza en 2 pasos:

1. Primero la similitud entre dos reglas cualesquiera se calcula aplicando la medida propuesta en [89] sobre los antecedentes de las reglas:

$$S_k(R_i, R_j) = \frac{\sum_{i,j} R_{ik}(x) \wedge R_{jk}(x)}{\sum_{i,j} R_{ik}(x) \vee R_{jk}(x)} = \frac{\sum_{i,j} \min(R_{ik}, R_{jk})}{\sum_{i,j} \max(R_{ik}, R_{jk})} \quad (3.3)$$

$$\forall 1 \leq i < j \leq NumeroDeReglas$$

$$\forall 1 \leq k \leq NumeroDeAntecedentes$$

2. Posteriormente se calcula la similitud global de la BR como la media aritmética de la similitud en todos los antecedentes:

$$Similitud = F_{i,j}(F_k(S_k(R_i, R_j)))$$

$$F \Rightarrow MediaAritmetica \quad (3.4)$$

$$\forall 1 \leq i < j \leq NumeroDeReglas$$

$$\forall 1 \leq k \leq NumeroDeAntecedentes$$

- **Redundancia** total de la BR, calculada como el número de pares de reglas ($Card$) cuya similitud es mayor que un umbral de redundancia β_R ($0 < \beta_R < 1$):

$$\begin{aligned}
 Redundancia &= \frac{Card(S_{kA}(R_i, R_j) > \beta_R \text{ Y } S_{kC}(R_i, R_j) > \beta_R)}{(NumeroDeReglas-1)!} \\
 \forall 1 \leq i < j \leq NumeroDeReglas \\
 \forall 1 \leq kA \leq NumeroDeAntecedentes \\
 \forall 1 \leq kC \leq NumeroDeConsecuentes
 \end{aligned} \tag{3.5}$$

siendo S_{kA} la similitud en los antecedentes y S_{kC} la similitud en los consecuentes, ambas similitudes calculadas siguiendo la Ec. 3.3. Las reglas redundantes, similares tanto en sus antecedentes como en sus consecuentes, deben ser eliminadas.

- **Incoherencia** del conjunto de reglas, calculada de forma similar a la redundancia. Se establece un umbral de incoherencia β_I , definido como $\beta_I = 1 - \beta_R$, para medir la “no similitud” en los consecuentes y se calcula la incoherencia como:

$$\begin{aligned}
 Incoherencia &= \frac{Card(S_{kA}(R_i, R_j) > \beta_R \text{ Y } S_{kC}(R_i, R_j) < \beta_I)}{(NumeroDeReglas-1)!} \\
 \forall 1 \leq i < j \leq NumeroDeReglas \\
 \forall 1 \leq kA \leq NumeroDeAntecedentes \\
 \forall 1 \leq kC \leq NumeroDeConsecuentes
 \end{aligned} \tag{3.6}$$

siendo S_{kA} la similitud en los antecedentes y S_{kC} la similitud en los consecuentes. Al eliminar reglas incoherentes, similares en sus antecedentes y muy diferentes en sus consecuentes, se mejora la consistencia haciendo que el SBRD sea más interpretable.

- **Completitud** o *No-Cobertura* de las particiones difusas, calculada como la media aritmética de la “no cobertura” de cada una de las variables:

$$\begin{aligned}
 NoCobertura &= MediaAritmetica(NoCoberturaParticion_k) \\
 NoCoberturaParticion_k &= \frac{PuntosSinCobertura}{PuntosTotales} \\
 PuntosSinCobertura \text{ si Nivel de Activacion} &\leq \beta_C \\
 \forall 1 \leq k \leq NumeroDeAntecedentes + NumeroDeConsecuentes
 \end{aligned} \tag{3.7}$$

Las particiones difusas completas son aquellas que minimizan la no cobertura del universo de discurso, de forma que para cualquier posible valor de

la variable existe en la partición un conjunto difuso para el cual su valor de pertenencia es mayor que un umbral de cobertura β_C .

En todos los casos un valor pequeño de los índices propuestos mejora la interpretabilidad del SBRD, reduciendo su complejidad, disminuyendo su similitud y aumentando su distinguibilidad. La métrica $Inter_C$ queda entonces definida como:

$$\begin{aligned}
 Inter_C = & MediaAritmetica(\lambda_{nr} * NumeroDeReglas_{nor}, \\
 & \lambda_s * Similitud_{nor}, \lambda_r * Redundancia_{nor}, \\
 & \lambda_i * Incoherencia_{nor}, \lambda_{nc} * NoCobertura_{nor}) \\
 & \lambda_j \in (0, 1)
 \end{aligned} \tag{3.8}$$

Tal y como se ha definido esta nueva métrica va a ser muy dependiente de la influencia de cada uno de los índices individuales de complejidad considerados. Por ejemplo, el número de reglas frente a la completitud de las variables ¿deben tener la misma relevancia? $Inter_C$ permite que este aspecto pueda ser definido por el usuario según su forma de entender la complejidad y la interpretabilidad de los SBRDs.

Por temas de cálculo cada uno de los índices debe ser normalizado a un rango común, utilizando la siguiente función de normalización:

$$Indice_{nor} = 1 - \frac{Indice_{Original} - Indice_{Actual}}{Indice_{Original}} \tag{3.9}$$

donde $Indice$ se refiere al índice concreto que va a ser normalizado, por ejemplo el $NumeroDeReglas$, la $Similitud$, etc. Se puede utilizar cualquier otra función de normalización, aunque la aquí propuesta es la más adecuada de cara al proceso de optimización genético a realizar posteriormente, ya que tiene en cuenta los valores del modelo difuso inicial.

3.2. Mejora del equilibrio precisión - interpretabilidad en los sistemas basados en reglas difusas

Una vez definida la nueva métrica de interpretabilidad $Inter_C$ en esta sección se va a proponer una metodología que, haciendo uso de dicha métrica, permita

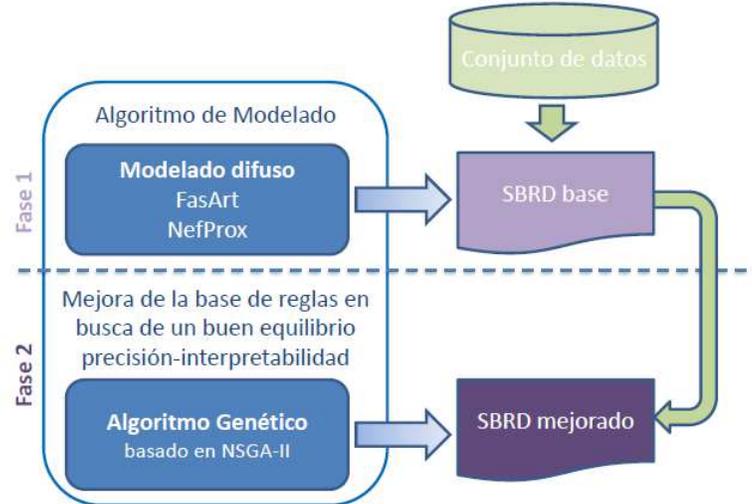


Figura 3.1: Proceso de mejora en dos fases.

generar SBRDs con mejores propiedades desde el punto de vista del equilibrio precisión-interpretabilidad, mediante la reducción de la complejidad del SBRD.

El objetivo final del proceso es obtener un SBRD preciso más interpretable, a través de un post-procesamiento de selección de reglas difusas que utiliza un enfoque genético bi-objetivo y un conjunto de índices de precisión e interpretabilidad sencillos y fácilmente interpretables.

Concretamente se va a proponer un proceso de mejora en dos fases (Fig. 3.1) donde:

1. Se genera un SBRD base con un buen nivel de precisión, que contenga un conjunto de reglas difusas candidatas a optimizar.
2. Se realiza un post-procesamiento de selección de reglas difusas utilizando un AEMO con el objetivo de mejorar las prestaciones del SBRD base en términos de interpretabilidad mientras se conserva, o incluso se mejora, su precisión.

Ambas fases del método son abiertas, de forma que en la primera fase podría utilizarse cualquier algoritmo de modelado difuso existente y en la segunda fase

cualquier técnica de optimización. Se explican a continuación cada una de las fases detalladamente.

3.2.1. Sistemas neuro-difusos

Los sistemas neuro-difusos son muy populares a la hora de generar SBRDs con un buen nivel de precisión, debido principalmente a las ventajas que presenta la capacidad de aprendizaje de las Redes Neuronales Artificiales y la capacidad de explicación de la Lógica Difusa.

En la primera fase del método propuesto se van a generar los **SBRD base** utilizando dos algoritmos de modelado neuro-difuso cuyo único objetivo es generar modelos difusos tan precisos como sea posible: NEFPROX (Neuro-Fuzzy Function Approximation) [122] como algoritmo de modelado **lingüístico** y FA-SART (Fuzzy Adaptive System ART based) [123] como algoritmo de modelado **aproximativo**. La principal diferencia existente entre ambos algoritmos es el tipo de reglas difusas que genera, lingüísticas en el primer caso (se define una semántica común para todas las reglas) y aproximativas en el segundo (cada regla tiene su propia semántica). Ambos algoritmos generan SBRDs precisos pero con prestaciones finales muy diferentes, ya que la forma de generar los SBRDs también lo es.

Con cada uno de estos algoritmos se generan dos SBRDs base distintos, uno al que llamaremos **compacto** y otro al que llamaremos **complejo**. Cada SBRD generado tendrá sus propias características en cuanto a precisión, complejidad, interpretabilidad y naturaleza difusa, de forma que se pueda validar la metodología propuesta en distintos entornos.

A continuación se describen las principales características de cada uno de los algoritmos de modelado:

NEFPROX Es un algoritmo de modelado neuro-difuso de aproximación de funciones que utiliza un algoritmo de *aprendizaje supervisado*. Es decir, el usuario fija ciertos parámetros y el algoritmo de aprendizaje genera las reglas difusas a partir de los datos, minimizando el error [122]. Concretamente el usuario define las particiones difusas iniciales de las variables uniformemente distribuidas y el algoritmo genera tantas reglas como sea necesario para cubrir todos los datos de entrenamiento. En este caso, al ser los conjuntos difusos de las variables definidos a priori por el usuario, las particiones difu-

sas generadas tienen una interpretabilidad inherente que no se da en todos los casos.

En esta tesis NEFPROX¹ es utilizado para generar un SBRD con funciones de pertenencia triangulares, inferencia max-min y defuzzificación mediante media de máximos.

FASART Es un algoritmo de modelado neuro-difuso basado en la Teoría de Resonancia Adaptativa (ART - *Adaptive Resonance Theory*). El sistema FASART se caracteriza por usar un algoritmo de *aprendizaje no supervisado* basado en datos y por establecer una equivalencia entre la función de activación de cada neurona y una función de pertenencia, de forma que el sistema final es equivalente a un SBRD donde: la fuzificación se realiza por punto simple, la inferencia a través del producto y la defuzzificación por media de los centros de los conjuntos difusos [123, 124].

El sistema FASART ya ha sido utilizado en diversos trabajos de modelado, detección de fallos, reconocimiento de patrones, etc. con resultados razonables en cuanto a la precisión de los modelos difusos; pero cuando estos modelos se evalúan bajo otros puntos de vista como la interpretabilidad, complejidad, etc. entonces aparecen problemas como el de la proliferación de reglas y/o conjuntos difusos [24, 125].

3.2.2. Selección de reglas usando algoritmos evolutivos multi-objetivo

La segunda fase del proceso de mejora propuesto se basa en aplicar un enfoque genético bi-objetivo de selección de reglas que hace uso del algoritmo genético NSGA-II. Gracias a la selección de reglas se pretende obtener un SBRD más compacto e interpretable, eliminando reglas irrelevantes, redundantes e incoherentes, particiones difusas incompletas, etc.

En las siguientes secciones se presenta la formulación matemática de las funciones objetivos y el resto de parámetros necesarios para ejecutar NSGA-II [126].

¹<http://fuzzy.cs.uni-magdeburg.de/nefprox/>

Funciones objetivo

El objetivo del proceso propuesto es obtener un SBRD con un mejor equilibrio precisión-interpretabilidad, por lo que las dos funciones objetivo definidas son:

- Maximizar la **precisión**, minimizando el error cuadrático medio (*ECM*) (Ec. 2.1):

$$\text{máx}(Precision) = \text{mín}(ECM) = \text{mín} \frac{1}{N} \sum_{i=1}^N (Y_i - Y'_i)^2 \quad (3.10)$$

El *ECM* es un índice de precisión ampliamente aceptado para medir la exactitud en problemas de modelado.

- Maximizar la **interpretabilidad**, minimizando la métrica *Inter_C* propuesta en la Sección 3.1 como la agregación de diversas características de interpretabilidad ponderadas por el usuario:

$$\begin{aligned} \text{máx}(Interpretabilidad) &= \text{mín}(Inter_C) = \\ &= \text{mín}(MediaAritmetica(\lambda_j * IndiceInterpretabilidad_j)) \end{aligned} \quad (3.11)$$

Para poder calcular los índices de interpretabilidad propuestos para *Inter_C* es necesario establecer determinados umbrales. Teniendo en cuenta [64, 88] los umbrales aquí utilizados son:

- $\beta_R = 0,8$ como umbral de redundancia.
- $\beta_I = 1 - \beta_R = 0,2$ como umbral de incoherencia.
- $\beta_C = 0$ como nivel de activación para evaluar la cobertura.
- $\lambda_i = 1$ para todos los índices individuales de *Inter_C*, ya que en principio se considera que la relevancia de cada una de las características es la misma.

Codificación, Poblaciones y Operadores Genéticos

A continuación se especifican varios aspectos del algoritmo NSGA-II para que el algoritmo realice una buena selección de reglas:

- Individuos codificados mediante **código binario**: $S = s_1s_2\dots s_N$ (N es el número de reglas iniciales), donde $s_q = 0$ indica que la regla candidata R_q no está incluida en el conjunto final, mientras que $s_q = 1$ significa que la regla está incluida. De esta forma cada individuo representa un subconjunto de reglas candidatas.
- Todos los genes toman valor 1 para todos los individuos de la **población inicial**, de forma que todos los individuos de la población inicial contienen todas las reglas candidatas. Esto permitirá realizar posteriormente una extracción progresiva de las peores reglas.
- **Operadores genéticos** seleccionados en función del objetivo final (ver Tabla 3.1):
 - Torneo binario para el operador de *selección*.
 - HUX (Eshelman 1991) como operador de *cruce* con probabilidad P_c . El cruce HUX intercambia exactamente la mitad de los alelos que son diferentes en los padres (los genes a ser cruzados son seleccionados aleatoriamente entre todos aquellos que son diferentes en los padres). Este operador asegura la máxima distancia de los descendientes respecto a sus padres (proceso de exploración).
 - *Mutación* clásica con probabilidad P_m . Este operador cambia aleatoriamente el valor de un gen, poniéndolo a 0 con probabilidad P_m y a 1 con probabilidad $1 - P_m$. Este operador fue propuesto para realizar selección de reglas en [43]. Para favorecer la eliminación de reglas, algo adecuado cuando todos los individuos de la población inicial contienen todas las reglas candidatas, el valor de P_m debe ser elevado.
- Además, si un individuo (subconjunto de reglas candidatas) deja sin cubrir ejemplos que ya estaban cubiertos es **penalizado** en ambos objetivos. De esta forma se asegura que el ranking de dichos individuos es (al menos) 2, enfrentándose entre ellos en frentes secundarios.
- El *criterio de parada* es el número de evaluaciones realizadas.

La *implementación* inicial del algoritmo NSGA-II ha sido obtenida de la página web de Kanpur Genetic Algorithms Laboratory ². Dicha implementación ha

²<http://www.iitk.ac.in/kangal/codes.shtml>

Tabla 3.1: Parametros NSGA-II.

Operadores genéticos	
Selección	Torneo binario
Cruce	HUX $P_c=0,9$
Mutación	Clásica $P_m=0,7$
Otras opciones	
Tamaño de la población	100
Número de evaluaciones	50000

sido modificada de forma que tanto los operadores genéticos como las funciones objetivo han sido adaptadas a lo explicado previamente. La Tabla 3.1 muestra los parámetros de ejecución utilizados en el algoritmo.

3.3. Estudio experimental

A fin de validar la métrica de agregación y la metodología propuesta, se van a utilizar nueve conjuntos de datos reales obtenidos del repositorio del proyecto KEEL³ [127, 128]. La Tabla 3.2 muestra las principales características de dichos conjuntos de datos.

El resto de la sección se organiza como sigue:

- La Sección 3.3.1 presenta la metodología seguida en la realización de los experimentos.
- La Sección 3.3.2 describe los parámetros usados para generar los SBRDs iniciales con FASART y NEFPDROX, así como sus características de precisión e interpretabilidad. Concretamente se generan dos modelos para cada algoritmo y conjunto de datos disponible, uno más *compacto* y otro más *complejo*. El único criterio tenido en cuenta en el proceso de generación de todos los SBRDs es el error cometido.
- La Sección 3.3.3 muestra los resultados obtenidos con los SBRDs de tipo

³<http://www.keel.es/>

Tabla 3.2: Conjuntos de datos considerados en el estudio experimental.

Conjunto de datos	Nombre	Variables	Registros
Plastic Strength	PLA	3	1650
Quake	QUA	4	2178
Electrical Maintenance	ELE	5	1056
Abalone	ABA	9	4177
Stock prices	STP	10	950
Weather Ankara	WAN	10	1609
Weather Izmir	WIZ	10	1461
Mortgage	MOR	16	1049
Treasury	TRE	16	1049

Disponible en <http://sci2s.ugr.es/keel/datasets.php>

compacto y los test estadísticos no paramétricos de los rangos con signo de Wilcoxon [129, 130].

- La Sección 3.3.3 muestra los resultados y los test estadísticos no paramétricos obtenidos con los SBRDs *complejo*, siguiendo el mismo esquema que en la sección 3.3.3.
- Finalmente, la Sección 3.3.4 presenta algunas conclusiones globales.

3.3.1. Metodología experimental

El Procedimiento 1 muestra la metodología de experimentación, siguiendo las fases ya presentadas en la Sección 3.2. En todos los experimentos se utiliza una **validación cruzada** 5-fold (cada grupo contiene un 20% de los registros) utilizando cuatro grupos para entrenamiento y una para test ⁴. Para cada una de las posibles 5 particiones diferentes (entrenamiento/test) el proceso (algoritmo de modelado y optimización genética) es ejecutado 6 veces, considerando en cada ejecución una semilla de generación de números aleatorios diferente. De esta forma se realizan 30 ejecuciones diferentes a partir de las cuales se generaran los valores medios finales.

⁴Las correspondientes particiones (5-fold) de los conjuntos de datos están disponibles en la página web del proyecto KEEL [128]: <http://sci2s.ugr.es/keel/datasets.php>

El **análisis de los Frentes de Pareto** de los SBRDs mejorados finales se realiza sólo sobre *tres SBRDs representativos* del Frente, seleccionados en función de los objetivos planteados [77, 92]:

1. SBRD más interpretable: Mejor $Inter_C$.
2. SBRD más preciso: Mejor Pre .
3. SBRD situado en la mediana precisión-interpretabilidad: Mediana $Pre - Inter_C$.

Procedimiento 1 Metodología experimental.

```

for Algoritmo de Modelado Neuro-Difuso=Lingüístico:Aproximativo do
  for Modelo=Compacto:Complejo do
    for ConjuntoDeDatos=1 to 9 do
      for ValidacionCruzada=1 to 5 do
        Generar SBRD base inicial
        for Ejecucion=1 to 6 do
          Ejecutar Algoritmo Genético NSGA-II modificado para selección
          de reglas bi-objetivo:
            Precisión  $\rightarrow \min(ECM_{tra})$ 
            Interpretabilidad  $\rightarrow \min(Inter_C)$ 
        end for
      end for
      Análisis Frentes de Pareto {SBRDs mejorados finales}
    end for
    Test Estadísticos No Paramétricos
  end for
end for

```

Por último, para poder conocer la significación estadística de los resultados, se calculan los valores medios (sobre 30 ejecuciones) de los tres SBRDs representativos y se realizan **test estadísticos no paramétricos** [131–134]⁵. Concretamente se va a ejecutar el test de Wilcoxon, teniendo en cuenta los tres modelos

⁵<http://sci2s.ugr.es/sicidm>

característicos de los frentes de Pareto generados con FASART+S_{NSGAII-InterC} y NEFPROX+S_{NSGAII-InterC} para cada SBRD base *compacto* y *complejo*. Este test se utiliza para detectar si existen diferencias significativas entre dos medias muestrales, y es análogo al *t-test* *pareado* en los procedimientos estadísticos no paramétricos. En general, el test se pregunta acerca de (H_0): las dos muestras provienen de poblaciones con la misma distribución?, y se basa en evaluar la diferencia entre pares de datos.

Adicionalmente, para poder tener diferencias bien definidas en los índices *ECM* y *NR* se van a utilizar diferencias normalizadas *DIFF* según la siguiente ecuación:

$$DIFF = \frac{Media(Otro) - Media(AlgoritmoDeReferencia)}{Media(Otro)} \quad (3.12)$$

donde $Media(x)$ representa tanto el *ECM* medio como el *NR* medio obtenido por el algoritmo x . De esta forma la diferencia representa el porcentaje de mejora obtenido respecto al algoritmo de referencia [77, 92]. Para el resto de índices (similitud, redundancia, incoherencia, no cobertura e *InterC*) esta normalización no es necesaria.

3.3.2. Características de los sistemas basados en reglas difusas iniciales

Con cada uno de los conjuntos de datos considerados se generan dos tipos de SBRDs base iniciales utilizando los algoritmos NEFPROX y FASART, uno *compacto* y otro *complejo*. Los parámetros utilizados para generar dichos SBRDs base iniciales, así como la codificación empleada para identificar los modelos, son los mostrados en la Tabla 3.3 donde:

- nLT : número de conjuntos difusos utilizados en cada partición de cada variable de entrada y salida utilizado por NEFPROX.
- ρ : parámetro de vigilancia utilizado por FASART.
- γ : tasa de fuzzificación utilizado por FASART.

Como ya se ha comentado previamente ambos algoritmos permiten generar SBRDs con buenos niveles de precisión, ya que durante el proceso de aprendizaje sólo tienen en cuenta el error cometido. La Tabla 3.4 resume las principales características de los SBRDs base iniciales generados utilizando la metodología

Tabla 3.3: Métodos considerados y parámetros de los SBRDs base iniciales.

Método	Ref.	Descripción y Parámetros	Objetivos
<i>Métodos para generar los modelos iniciales</i>			
NEFPROX	[122]	Neuro-Fuzzy Function Approximation (Genera SBRD lingüístico inicial)	—
FASART	[123]	Neuro fuzzy system based on ART (Genera SBRD aproximativo inicial)	—
<i>Algoritmo Evolutivo Multi-Objetivo para el Post-procesamiento</i>			
$S_{NSGAI-InterC}$	Nuevo	Selección de reglas bi-objetivo con NSGA-II	$ECM / InterC$
<i>Combinaciones analizadas (Una Lingüística y Una Aproximativa)</i>			
LINGA	—	NEFPROX Compacto ($nLT = 5$) + $S_{NSGAI-InterC}$	$ECM / InterC$
LINGB	—	NEFPROX Complejo ($nLT = 7$) + $S_{NSGAI-InterC}$	$ECM / InterC$
SCATA	—	FASART Compacto ($\rho = 0,3, \gamma = 10$) + $S_{NSGAI-InterC}$	$ECM / InterC$
SCATB	—	FASART Complejo ($\rho = 0,9, \gamma = 10$) + $S_{NSGAI-InterC}$	$ECM / InterC$

descrita previamente. Los índices mostrados en las tablas son el error cuadrático medio en fase de entrenamiento (ECM_{tra}) y test (ECM_{tst}), el número de reglas (NR), la similitud (S), la redundancia (R), la incoherencia (I) y el porcentaje de completitud ($C(\%)$) calculado como 1 menos el valor de no cobertura ($1 - NC$).

Al analizar las características de los SBRDs base iniciales se observa que:

- Los SBRDs LINGA y SCATA, generados respectivamente con NEFPROX y FASART, son más compactos al obtener buenos niveles de precisión con un número bajo de reglas. La similitud entre reglas puede ser disminuida, tienen algo de redundancia e incoherencia, y las particiones difusas son completas.
- Los otros dos SBRDs (LINGB y SCATB) son más complejos, obteniendo una mejor precisión al incluir reglas adicionales. Los modelos *complejos* NEFPROX tienen menos similitud, redundancia e incoherencia que los modelos

compactos, mientras que los modelos *complejos* FASART tienen unas veces menos similitud, redundancia e incoherencia que los modelos *compactos* y otras veces tienen más, con una completitud similar en todos los casos.

La Tabla 3.5 compara las propiedades de los SBRDs base iniciales con las obtenidas por el algoritmo Wang & Mendel [135] en [77]. En general, los SBRDs NEFPROX y FASART muestran una mayor precisión, obtenida utilizando más reglas, y, en la mayoría de los casos, peores propiedades de interpretabilidad, algo típico en los algoritmos de modelado neuro-difusos.

3.3.3. Resultados y análisis de los sistemas basados en reglas difusas mejorados

Una vez generados los SBRDs base iniciales el siguiente paso es mejorarlos con el objetivo de obtener SBRDs finales mejorados con un equilibrio precisión-interpretabilidad adecuado.

Se presenta a continuación el análisis realizado sobre los resultados obtenidos, agrupando por un lado los SBRDs *compactos* y por otro los SBRDs *complejos*. En ambos casos se presentan las métricas *Pre* e *Inter_C* para los SBRDs mejorados, y se llevan a cabo los *test estadísticos no paramétricos* para comprobar si los resultados obtenidos son estadísticamente relevantes.

Tabla 3.4: Características iniciales para LINGA, LINGB, SCATA y SCATB.

Método	Conjunto de datos	ECM_{tra}	ECM_{tst}	NR	S	R	I	$C(\%)$
LINGA	PLA	3.208	3.222	17.0	0.210	0	0	100
	QUA	0.039	0.041	55.8	0.272	0	0	100
	ELE	620411	622331	79.6	0.286	0	0	99.4
	ABA	6.653	7.231	272.2	0.433	0.012	0.007	99.3
	STP	2.248	2.493	303.4	0.285	0.017	0.007	100
	WIZ	9.958	13.837	500.0	0.503	0.024	0.009	99.7
	WAN	12.227	15.920	500.0	0.449	0.017	0.006	98.3
	MOR	0.716	0.729	170.0	0.359	0.046	0.011	100
	TRE	1.029	1.087	170.6	0.350	0.047	0.011	100
LINGB	PLA	2.606	2.636	31.0	0.161	0	0	100
	QUA	0.035	0.037	98.0	0.229	0	0	100
	ELE	556228	598472	100.2	0.227	0	0	99.4
	ABA	5.636	6.370	500.0	0.350	0.006	0.003	95.5
	STP	1.307	1.727	433.8	0.215	0.009	0.003	100
	WIZ	10.103	17.471	500.0	0.442	0.011	0.004	100
	WAN	21.836	33.620	500.0	0.390	0.008	0.003	98.5
	MOR	0.337	0.510	301.6	0.255	0.026	0.006	99.9
	TRE	0.491	0.673	305.4	0.243	0.028	0.004	100
SCATA	PLA	3.483	3.621	48.6	0.238	0.001	0.016	99.9
	QUA	0.050	0.054	119.8	0.220	9e-5	0.002	98.0
	ELE	117867	158820	92.6	0.225	0.004	0.001	96.3
	ABA	6.872	7.683	122.8	0.277	32e-5	0.001	98.6
	STP	2.091	2.270	101.8	0.195	0	16e-5	100
	WIZ	5.452	16.555	221.6	0.335	3e-5	0	99.9
	WAN	9.813	21.970	231.8	0.304	0	0	99.9
	MOR	1.041	1.258	52.6	0.299	0	0	99.9
	TRE	0.908	1.339	49.6	0.292	50e-5	0	99.8
SCATB	PLA	2.783	2.821	96.6	0.162	43e-5	0.006	100
	QUA	0.046	0.050	243.8	0.265	20e-5	0.002	98.1
	ELE	56584	100229	129.8	0.266	0.003	0.001	96.9
	ABA	5.033	6.247	298.0	0.331	41e-5	0.001	100
	STP	0.426	0.698	163.6	0.185	3e-5	16e-5	100
	WIZ	1.788	21.934	466.4	0.360	0.6e-5	0	100
	WAN	2.593	28.312	537.6	0.312	0.3e-5	0.1e-5	99.9
	MOR	0.085	0.352	92.2	0.284	63e-5	8e-5	100
	TRE	0.150	0.552	76.6	0.289	8e-5	13e-5	100

SBRDs *Compactos*

Los resultados medios obtenidos para los SBRDs base *compactos* (LINGA y SCATA) son los mostrados en la Tabla 3.6. Concretamente se muestran los valores medios de la métrica $Inter_C$ y del error ECM_{tst} para cada uno de los tres SBRDs a tener en cuenta: Mejor $Inter_C$, Mediana $Pre - Inter_C$ y Mejor Pre . La primera línea muestra el SBRD base inicial, mientras que la segunda muestra las características del SBRD mejorado final.

Según los valores mostrados en la tabla la interpretabilidad mejora hasta en un 71,53% (PLA, SCATA, SBRD Mejor $Inter_C$, $Inter_C$ se reduce de 0,800 a 0,228), salvo para el SBRD de Mejor Pre del conjunto de datos MOR con el algoritmo SCATA en el que la interpretabilidad es reducida al aumentar el valor de $Inter_C$ desde 0,480 a 1,548. Por su parte ECM_{tst} se mantiene en el mismo orden de magnitud en todos los casos.

Las Tablas 3.7, 3.8 y 3.9 muestran, para cada uno de los SBRDs analizados del Frente de Pareto, los valores medios de cada uno de los índices individuales: el error cuadrático medio en entrenamiento (ECM_{tra}) y test (ECM_{tst}), el número de reglas medio (NR), la similitud media (S), la redundancia media (R), la incoherencia media (I) y el porcentaje de completitud medio ($C(\%)$) calculado como 1 menos el valor de no cobertura ($1 - NC$).

Evidentemente, cuando la redundancia o la incoherencia son nulas en el SBRD base inicial también lo van a ser en los distintos SBRDs que se puedan generar a partir del inicial. Desde el punto de vista del proceso de optimización es como si dichas características no fueran tenidas en cuenta a la hora de calcular el valor de $Inter_C$. Es decir, es como si para dichos índices se seleccionara $\lambda_j = 0$. Por su parte, a la hora de realizar los test estadísticos, se considera que tanto el valor del índice en el SBRD inicial como en el SBRD final es 0.

Tabla 3.6: SBRDs compactos mejorados: Características principales.

Método	Conjunto de datos	Mejor $Inter_C$		Mediana $Pre - Inter_C$		Mejor Pre	
		ECM_{1st}	$Inter_C$	ECM_{1st}	$Inter_C$	ECM_{1st}	$Inter_C$
LINGA	$PLA_{inicial}$	3.222	0.400	3.222	0.400	3.222	0.400
	PLA_{final}	4.391	0.287	3.572	0.344	3.222	0.400
	$QUA_{inicial}$	0.041	0.400	0.041	0.400	0.041	0.400
	QUA_{final}	0.042	0.303	0.041	0.327	0.041	0.363
	$ELE_{inicial}$	622331	0.600	622331	0.600	622331	0.600
	ELE_{final}	700774	0.474	620292	0.505	614691	0.551
	$ABA_{inicial}$	7.231	0.920	7.231	0.920	7.231	0.920
	ABA_{final}	6.377	0.642	6.271	0.680	6.205	0.725
	$STP_{inicial}$	2.493	0.800	2.493	0.800	2.493	0.800
	STP_{final}	2.339	0.528	2.164	0.557	2.055	0.588
	$WIZ_{inicial}$	13.837	0.840	13.837	0.840	13.837	0.840
	WIZ_{final}	13.271	0.630	12.998	0.644	12.804	0.667
	$WAN_{inicial}$	15.920	1.000	15.920	1.000	15.920	1.000
	WAN_{final}	16.661	0.808	16.161	0.827	16.034	0.852
	$MOR_{inicial}$	0.729	0.800	0.729	0.800	0.729	0.800
	MOR_{final}	0.975	0.475	0.660	0.529	0.632	0.615
	$TRE_{inicial}$	1.087	0.800	1.087	0.800	1.087	0.800
	TRE_{final}	1.350	0.482	0.978	0.543	0.942	0.608
SCATA	$PLA_{inicial}$	3.621	0.800	3.621	0.800	3.621	0.800
	PLA_{final}	3.718	0.228	3.073	0.260	2.688	0.472
	$QUA_{inicial}$	0.054	0.840	0.054	0.840	0.054	0.840
	QUA_{final}	0.042	0.448	0.039	0.483	0.038	0.546
	$ELE_{inicial}$	158820	1.000	158820	1.000	158820	1.000
	ELE_{final}	166133	0.884	159943	0.916	160842	1.017
	$ABA_{inicial}$	7.683	1.000	7.683	1.000	7.683	1.000
	ABA_{final}	6.881	0.494	5.902	0.565	5.788	0.633
	$STP_{inicial}$	2.270	0.520	2.270	0.520	2.270	0.520
	STP_{final}	2.815	0.346	2.745	0.368	2.405	0.537
	$WIZ_{inicial}$	16.555	0.520	16.555	0.520	16.555	0.520
	WIZ_{final}	17.571	0.366	17.401	0.376	16.701	0.387
	$WAN_{inicial}$	21.970	0.560	21.970	0.560	21.970	0.560
	WAN_{final}	23.233	0.439	22.970	0.451	22.567	0.465
	$MOR_{inicial}$	1.258	0.480	1.258	0.480	1.258	0.480
	MOR_{final}	1.266	0.411	1.216	0.469	1.178	1.548
	$TRE_{inicial}$	1.339	0.680	1.339	0.680	1.339	0.680
	TRE_{final}	1.463	0.498	1.351	0.512	1.335	0.652

Tabla 3.7: SBRDs compactos mejorados: Características Mejor $Inter_C$.

Método	Conjunto de datos	Mejor $Inter_C$							
		ECM_{tra}	ECM_{1st}	NR	S	R	I	$C(\%)$	
LINGA	$PLA_{inicial}$	3.208	3.222	17.00	0.2098	0	0	100	
	PLA_{final}	4.350	4.391	10.60	0.1701	-	-	100	
	$QUA_{inicial}$	0.039	0.041	55.80	0.2723	0	0	100	
	QUA_{final}	0.041	0.042	28.40	0.2742	-	-	100	
	$ELE_{inicial}$	620411	622331	79.60	0.2860	0	0	99.44	
	ELE_{final}	659619	700774	37.20	0.2575	-	-	99.44	
	$ABA_{inicial}$	6.653	7.231	272.20	0.4330	0.0118	0.0069	99.25	
	ABA_{final}	5.783	6.377	153.40	0.4199	0.0071	0.0033	99.25	
	$STP_{inicial}$	2.248	2.493	303.40	0.2848	0.0174	0.0066	100	
	STP_{final}	1.881	2.339	174.47	0.2766	0.0126	0.0025	100	
	$WIZ_{inicial}$	9.958	13.837	500.00	0.5031	0.0237	0.0092	99.67	
	WIZ_{final}	8.988	13.271	357.80	0.4847	0.0187	0.0045	99.67	
	$WAN_{inicial}$	12.227	15.920	500.00	0.4492	0.0174	0.0060	98.33	
	WAN_{final}	11.627	16.661	369.60	0.4346	0.0140	0.0032	98.33	
	$MOR_{inicial}$	0.716	0.729	170.00	0.3587	0.0460	0.0112	100	
	MOR_{final}	0.890	0.975	76.60	0.3510	0.0316	0.0029	100	
	$TRE_{inicial}$	1.029	1.087	170.60	0.3498	0.0465	0.0106	100	
	TRE_{final}	1.260	1.350	76.80	0.3440	0.0293	0.0037	99.55	
	SCATA	$PLA_{inicial}$	3.483	3.621	48.60	0.2379	0.0011	0.0156	99.90
		PLA_{final}	3.742	3.718	18.80	0.1314	0	0	99.90
$QUA_{inicial}$		0.050	0.054	119.80	0.2202	0.0001	0.0023	98.03	
QUA_{final}		0.039	0.042	56.57	0.1692	0	0	98.03	
$ELE_{inicial}$		117867	158820	92.60	0.2251	0.0043	0.0010	96.34	
ELE_{final}		124646	166133	87.20	0.2178	0.0038	0.0005	96.34	
$ABA_{inicial}$		6.872	7.683	122.80	0.2770	0.0003	0.0008	98.61	
ABA_{final}		6.187	6.881	67.80	0.2536	0	0	98.61	
$STP_{inicial}$		2.091	2.270	101.80	0.1952	0	0.0002	100	
STP_{final}		2.239	2.815	64.60	0.1756	-	0.0001	99.15	
$WIZ_{inicial}$		5.452	16.555	221.60	0.3354	3e-5	0	99.88	
WIZ_{final}		5.341	17.571	115.97	0.3041	0	-	99.67	
$WAN_{inicial}$		9.813	21.970	231.80	0.3039	0	0	99.92	
WAN_{final}		9.686	23.233	109.80	0.2794	-	-	99.87	
$MOR_{inicial}$		1.041	1.258	52.60	0.2986	0	0	99.93	
MOR_{final}		1.007	1.266	36.80	0.2847	-	-	99.60	
$TRE_{inicial}$		0.908	1.339	49.60	0.2920	0.0005	0	99.83	
TRE_{final}		1.030	1.463	36.00	0.2816	0	-	99.74	

Tabla 3.8: SBRDs compactos mejorados: Características Mediana $Pre - Inter_C$.

Método	Conjunto de datos	Mediana $Pre - Inter_C$						
		ECM_{tra}	ECM_{tst}	NR	S	R	I	$C(\%)$
LINGA	$PLA_{inicial}$	3.208	3.222	17.00	0.2098	0	0	100
	PLA_{final}	3.559	3.572	13.00	0.2008	-	-	100
	$QUA_{inicial}$	0.039	0.041	55.80	0.2723	0	0	100
	QUA_{final}	0.039	0.041	34.00	0.2790	-	-	100
	$ELE_{inicial}$	620411	622331	79.60	0.2860	0	0	99.44
	ELE_{final}	616480	620292	47.20	0.2659	-	-	99.44
	$ABA_{inicial}$	6.653	7.231	272.20	0.4330	0.0118	0.0069	99.25
	ABA_{final}	5.627	6.271	161.20	0.4222	0.0079	0.0039	99.25
	$STP_{inicial}$	2.248	2.493	303.40	0.2848	0.0174	0.0066	100
	STP_{final}	1.671	2.164	181.40	0.2770	0.0129	0.0032	100
	$WIZ_{inicial}$	9.958	13.837	500.00	0.5031	0.0237	0.0092	99.67
	WIZ_{final}	8.189	12.998	363.00	0.4860	0.0188	0.0049	99.67
	$WAN_{inicial}$	12.227	15.920	500.00	0.4492	0.0174	0.0060	98.33
	WAN_{final}	10.791	16.161	376.00	0.4359	0.0146	0.0035	98.33
	$MOR_{inicial}$	0.716	0.729	170.00	0.3587	0.0460	0.0112	100
	MOR_{final}	0.618	0.660	82.80	0.3521	0.0321	0.0053	100
	$TRE_{inicial}$	1.029	1.087	170.60	0.3498	0.0465	0.0106	100
	TRE_{final}	0.848	0.978	84.20	0.3447	0.0291	0.0065	99.77
SCATA	$PLA_{inicial}$	3.483	3.621	48.60	0.2379	0.0011	0.0156	99.90
	PLA_{final}	2.818	3.073	22.00	0.1535	0	0	99.90
	$QUA_{inicial}$	0.050	0.054	119.80	0.2202	0.0001	0.0023	98.03
	QUA_{final}	0.035	0.039	64.13	0.1760	0	0.0002	98.02
	$ELE_{inicial}$	117867	158820	92.60	0.2251	0.0043	0.0010	96.34
	ELE_{final}	118905	159943	87.60	0.2198	0.0042	0.0007	96.34
	$ABA_{inicial}$	6.872	7.683	122.80	0.2770	0.0003	0.0008	98.61
	ABA_{final}	5.185	5.902	80.00	0.2614	0.0001	5e-5	98.61
	$STP_{inicial}$	2.091	2.270	101.80	0.1952	0	0.0002	100
	STP_{final}	2.100	2.745	69.20	0.1802	-	0.0001	98.97
	$WIZ_{inicial}$	5.452	16.555	221.60	0.3354	3e-5	0	99.88
	WIZ_{final}	5.096	17.401	124.40	0.3083	0	-	99.67
	$WAN_{inicial}$	9.813	21.970	231.80	0.3039	0	0	99.92
	WAN_{final}	9.298	22.970	119.20	0.2858	-	-	99.87
	$MOR_{inicial}$	1.041	1.258	52.60	0.2986	0	0	99.93
	MOR_{final}	0.945	1.216	38.80	0.2961	-	-	98.91
	$TRE_{inicial}$	0.908	1.339	49.60	0.2920	0.0005	0	99.83
	TRE_{final}	0.917	1.351	39.00	0.2833	0	-	99.74

Tabla 3.9: SBRDs *compactos* mejorados: Características Mejor *Pre*.

Método	Conjunto de datos	Mejor <i>Pre</i>							
		ECM_{vtra}	ECM_{tst}	NR	S	R	I	$C(\%)$	
LINGA	PLA _{inicial}	3.208	3.222	17.00	0.2098	0	0	100	
	PLA _{final}	3.208	3.222	17.00	0.2098	-	-	100	
	QUA _{inicial}	0.039	0.041	55.80	0.2723	0	0	100	
	QUA _{final}	0.038	0.041	43.80	0.2799	-	-	100	
	ELE _{inicial}	620411	622331	79.60	0.2860	0	0	99.44	
	ELE _{final}	612113	614691	62.20	0.2791	-	-	99.44	
	ABA _{inicial}	6.653	7.231	272.20	0.4330	0.0118	0.0069	99.25	
	ABA _{final}	5.579	6.205	174.40	0.4232	0.0089	0.0045	99.25	
	STP _{inicial}	2.248	2.493	303.40	0.2848	0.0174	0.0066	100	
	STP _{final}	1.616	2.055	191.00	0.2773	0.0137	0.0037	100	
	WIZ _{inicial}	9.958	13.837	500.00	0.5031	0.0237	0.0092	99.67	
	WIZ _{final}	8.004	12.804	375.40	0.4880	0.0199	0.0053	99.67	
	WAN _{inicial}	12.227	15.920	500.00	0.4492	0.0174	0.0060	98.33	
	WAN _{final}	10.654	16.034	388.80	0.4380	0.0153	0.0037	98.33	
	MOR _{inicial}	0.716	0.729	170.00	0.3587	0.0460	0.0112	100	
	MOR _{final}	0.582	0.632	98.00	0.3556	0.0366	0.0079	100	
	TRE _{inicial}	1.029	1.087	170.60	0.3498	0.0465	0.0106	100	
	TRE _{final}	0.816	0.942	97.80	0.3446	0.0331	0.0081	99.77	
	SCATA	PLA _{inicial}	3.483	3.621	48.60	0.2379	0.0011	0.0156	99.90
		PLA _{final}	2.549	2.688	27.20	0.2207	0.0006	0.0037	98.63
QUA _{inicial}		0.050	0.054	119.80	0.2202	0.0001	0.0023	98.03	
QUA _{final}		0.034	0.038	79.90	0.1925	0	0.0004	97.97	
ELE _{inicial}		117867	158820	92.60	0.2251	0.0043	0.0010	96.34	
ELE _{final}		116196	160842	88.60	0.2208	0.0045	0.0011	96.34	
ABA _{inicial}		6.872	7.683	122.80	0.2770	0.0003	0.0008	98.61	
ABA _{final}		5.079	5.788	91.20	0.2630	0.0002	0.0001	98.61	
STP _{inicial}		2.091	2.270	101.80	0.1952	0	0.0002	100	
STP _{final}		2.037	2.405	78.20	0.1965	-	0.0002	99.12	
WIZ _{inicial}		5.452	16.555	221.60	0.3354	3e-5	0	99.88	
WIZ _{final}		5.021	16.701	134.53	0.3129	0	-	99.81	
WAN _{inicial}		9.813	21.970	231.80	0.3039	0	0	99.92	
WAN _{final}		9.151	22.567	131.80	0.2903	-	-	99.87	
MOR _{inicial}		1.041	1.258	52.60	0.2986	0	0	99.93	
MOR _{final}		0.924	1.178	41.40	0.3075	-	-	98.05	
TRE _{inicial}		0.908	1.339	49.60	0.2920	0.0005	0	99.83	
TRE _{final}		0.898	1.335	41.40	0.2877	0.0005	-	99.72	

Los resultados del test de Wilcoxon sobre los modelos *compactos* de LINGA (Tabla 3.10) muestran que:

- **SBRDs Mejor *Inter_C***: mejoran los índices de interpretabilidad (mejor interpretabilidad y menor complejidad) aunque la precisión se reduce un poco. Observando el valor exacto del error (i.e. en el peor caso el ECM_{tst} aumenta de 3,222 a 4,391) se observa que este se mantiene suficientemente bajo.
- **SBRDs Mediana *Pre – Inter_C***: tienen una precisión similar a los modelos difusos base y los índices de interpretabilidad aumentan, por lo que su complejidad se reduce.
- **SBRDs Mejor *Pre***: mejoran tanto la precisión como la interpretabilidad. Todos los índices individuales de error de test (ECM_{tst}), número de reglas (NR), similitud (S), redundancia (R) e incoherencia (I) reducen su valor mientras se mantiene la no cobertura (NC).

El test de Wilcoxon sobre los modelos *compactos* de SCATA (Tabla 3.11) acepta que:

- **SBRDs Mejor *Inter_C***: mejoran los índices y la métrica de interpretabilidad, aunque la precisión es ligeramente peor. Ésto es algo habitual, debido al dilema contradictorio de Precisión frente a Interpretabilidad-Complejidad. Aunque en algunos casos el error aumenta, su valor sigue siendo razonablemente bajo, por lo que la precisión final es aceptable.
- **SBRDs Mediana *Pre – Inter_C***: tienen una precisión similar a los modelos difusos base, y los índices de interpretabilidad se mejoran. La precisión se mantiene en el mismo orden de magnitud aunque el valor del error es algo mayor.
- **SBRDs Mejor *Pre***: casi ninguno de los índices, ni de precisión ni de interpretabilidad, muestran diferencias estadísticas significativas. En este caso, la no variación de la interpretabilidad puede considerarse paradójica si se analizan individualmente los valores concretos del número de reglas, similitud, redundancia, incoherencia y completitud de la Tabla 3.9. Este resultado es consecuencia, básicamente, de la forma en la que FASART genera las particiones difusas, y de la igualdad de la relevancia dada a cada índice en la métrica global definida: número de reglas frente a completitud.

Tabla 3.10: Test de Wilcoxon para los SBRDs *Compactos* LINGA: SBRD base (R+) y SBRD mejorado (R-).

SBRD Mejor <i>Inter_C</i>				
Medidas	R+	R-	Hipótesis ($\alpha = 0,10$)	<i>p</i> -valor
<i>ECM_{tst}</i>	11.0	34.0	Aceptada	0.173
<i>NR</i>	45.0	0.0	Rechazada	0.008
<i>S</i>	44.0	1.0	Rechazada	0.011
<i>R</i>	41.0	2.0	Rechazada	0.028
<i>I</i>	41.0	2.0	Rechazada	0.028
<i>NC</i>	18.0	27.0	Aceptada	0.655
<i>Inter_C</i>	45.0	0.0	Rechazada	0.008
SBRD Mediana <i>Pre – Inter_C</i>				
Medidas	R+	R-	Hipótesis ($\alpha = 0,10$)	<i>p</i> -valor
<i>ECM_{tst}</i>	35.0	10.0	Aceptada	0.139
<i>NR</i>	45.0	0.0	Rechazada	0.008
<i>S</i>	42.0	3.0	Rechazada	0.021
<i>R</i>	41.0	2.0	Rechazada	0.028
<i>I</i>	41.0	2.0	Rechazada	0.028
<i>NC</i>	18.0	27.0	Aceptada	0.655
<i>Inter_C</i>	45.0	0.0	Rechazada	0.008
SBRD Mejor <i>Pre</i>				
Medidas	R+	R-	Hipótesis ($\alpha = 0,10$)	<i>p</i> -valor
<i>ECM_{tst}</i>	42.0	2.0	Rechazada	0.017
<i>NR</i>	44.0	0.0	Rechazada	0.012
<i>S</i>	38.0	6.0	Rechazada	0.066
<i>R</i>	41.0	2.0	Rechazada	0.028
<i>I</i>	41.0	2.0	Rechazada	0.028
<i>NC</i>	18.0	27.0	Aceptada	0.655
<i>Inter_C</i>	44.0	0.0	Rechazada	0.011

Tabla 3.11: Test de Wilcoxon para los SBRDs *Compactos* SCATA: SBRD base (R+) y SBRD mejorado (R-).

SBRD Mejor <i>Inter_C</i>				
Medidas	R+	R-	Hipótesis ($\alpha = 0,10$)	<i>p-valor</i>
<i>ECM_{tst}</i>	15.0	30.0	Aceptada	0.374
<i>NR</i>	45.0	0.0	Rechazada	0.008
<i>S</i>	45.0	0.0	Rechazada	0.008
<i>R</i>	41.0	2.0	Rechazada	0.028
<i>I</i>	40.0	5.0	Rechazada	0.043
<i>NC</i>	5.0	40.0	Rechazada	0.018
<i>Inter_C</i>	45.0	0.0	Rechazada	0.008
SBRD Mediana <i>Pre – Inter_C</i>				
Medidas	R+	R-	Hipótesis ($\alpha = 0,10$)	<i>p-valor</i>
<i>ECM_{tst}</i>	26.0	19.0	Aceptada	0.678
<i>NR</i>	45.0	0.0	Rechazada	0.008
<i>S</i>	45.0	0.0	Rechazada	0.008
<i>R</i>	41.0	2.0	Rechazada	0.028
<i>I</i>	40.0	5.0	Rechazada	0.043
<i>NC</i>	2.0	41.0	Rechazada	0.012
<i>Inter_C</i>	45.0	0.0	Rechazada	0.008
SBRD Mejor <i>Pre</i>				
Medidas	R+	R-	Hipótesis ($\alpha = 0,10$)	<i>p-valor</i>
<i>ECM_{tst}</i>	31.0	14.0	Aceptada	0.314
<i>NR</i>	45.0	0.0	Rechazada	0.008
<i>S</i>	40.0	5.0	Rechazada	0.038
<i>R</i>	33.0	10.0	Aceptada	0.249
<i>I</i>	29.0	16.0	Aceptada	0.225
<i>NC</i>	1.5	43.5	Rechazada	0.008
<i>Inter_C</i>	33.0	12.0	Aceptada	0.214

SBRDs *Complejos*

Los resultados medios obtenidos para los SBRDs base *complejos* (LINGB y SCATB) son los mostrados en la Tabla 3.12. Concretamente se muestran los valores medios de la métrica $Inter_C$ y del error ECM_{tst} para cada uno de los tres SBRDs a tener en cuenta: Mejor $Inter_C$, Mediana $Pre - Inter_C$ y Mejor Pre . La primera línea muestra el SBRD base inicial, mientras que la segunda muestra las características del SBRD mejorado final.

Como puede verse la interpretabilidad mejora entre un 2,02 % (PLA, LINGB, SBRD Mejor Pre , $Inter_C$ pasa de 0,400 a 0,392) y un 71,52 % (PLA, SCATB, SBRD Mejor $Inter_C$, $Inter_C$ pasa de 0,760 a 0,216), mientras que la precisión aumenta entre un 0,82 % (ELE, LINGB, SBRD Mejor Pre , ECM_{tst} pasa de 598472 a 593548) y un 26,14 % (MOR, LINGB, SBRD Mejor Pre , ECM_{tst} pasa de 0,510 a 0,377). En el peor caso $Inter_C$ aumenta de 0,560 a 1,123 (TRE, SCATB, SBRD Mejor Pre) y ECM_{tst} aumenta de 0,552 a 0,725 (TRE, SCATB, SBRD Mejor $Inter_C$). Al igual que antes un valor inicial nulo de la redundancia o la incoherencia hace que dichos índices no sean tenidos en cuenta en el proceso de optimización.

Las Tablas 3.13, 3.14 y 3.15 muestran los valores medios de cada uno de los índices individuales: el error cuadrático medio en entrenamiento (ECM_{tra}) y test (ECM_{tst}), el número de reglas medio (NR), la similitud media (S), la redundancia media (R), la incoherencia media (I) y el porcentaje de completitud medio ($C(\%)$) calculado como 1 menos el valor de no cobertura ($1 - NC$).

Para los SBRDs *complejos* de LINGB los resultados del test de Wilcoxon (Tabla 3.16) muestran que:

- **SBRDs Mejor $Inter_C$** : en algunos casos la precisión es peor aunque esta pérdida se compensa con la reducción de complejidad, y por tanto con una mejora de la interpretabilidad.
- **SBRDs Mediana $Pre - Inter_C$** : comenten un error similar pero con mejor interpretabilidad (menos reglas, similitud, redundancia e incoherencia con igual completitud).
- **SBRDs Mejor Pre** : los índices de interpretabilidad se mejoran, al igual que en los modelos previos, mientras que el error se mantiene bajo.

Tabla 3.12: SBRDs *complejos* mejorados: Características principales.

Método	Conjunto de datos	Mejor $Inter_C$		Mediana $Pre - Inter_C$		Mejor Pre	
		ECM_{1st}	$Inter_C$	ECM_{1st}	$Inter_C$	ECM_{1st}	$Inter_C$
LINGB	$PLA_{inicial}$	2.636	0.400	2.636	0.400	2.636	0.400
	PLA_{final}	3.286	0.296	2.824	0.335	2.655	0.392
	$QUA_{inicial}$	0.037	0.400	0.037	0.400	0.037	0.400
	QUA_{final}	0.037	0.285	0.037	0.302	0.037	0.331
	$ELE_{inicial}$	598472	0.600	598472	0.600	598472	0.600
	ELE_{final}	612939	0.502	598735	0.520	593548	0.543
	$ABA_{inicial}$	6.370	1.000	6.370	1.000	6.370	1.000
	ABA_{final}	6.128	0.782	6.093	0.799	6.019	0.827
	$STP_{inicial}$	1.727	0.800	1.727	0.800	1.727	0.800
	STP_{final}	1.812	0.560	1.693	0.580	1.664	0.611
	$WIZ_{inicial}$	17.471	0.840	17.471	0.840	17.471	0.840
	WIZ_{final}	19.099	0.603	18.829	0.624	18.860	0.653
	$WAN_{inicial}$	33.620	0.880	33.620	0.880	33.620	0.880
	WAN_{final}	35.156	0.658	35.003	0.681	34.684	0.744
	$MOR_{inicial}$	0.510	0.880	0.510	0.880	0.510	0.880
	MOR_{final}	0.415	0.620	0.382	0.647	0.377	0.681
	$TRE_{inicial}$	0.673	0.800	0.673	0.800	0.673	0.800
	TRE_{final}	0.611	0.528	0.584	0.551	0.572	0.585
SCATB	$PLA_{inicial}$	2.821	0.760	2.821	0.760	2.821	0.760
	PLA_{final}	3.172	0.216	2.567	0.248	2.374	0.590
	$QUA_{inicial}$	0.050	1.000	0.050	1.000	0.050	1.000
	QUA_{final}	0.041	0.492	0.038	0.521	0.038	0.581
	$ELE_{inicial}$	100229	1.000	100229	1.000	100229	1.000
	ELE_{final}	97416	0.740	87004	0.788	85157	0.869
	$ABA_{inicial}$	6.247	0.880	6.247	0.880	6.247	0.880
	ABA_{final}	5.385	0.431	5.273	0.470	5.240	0.535
	$STP_{inicial}$	0.698	0.680	0.698	0.680	0.698	0.680
	STP_{final}	0.755	0.353	0.749	0.549	0.686	0.721
	$WIZ_{inicial}$	21.934	0.440	21.934	0.440	21.934	0.440
	WIZ_{final}	23.251	0.360	23.182	0.365	23.067	0.377
	$WAN_{inicial}$	28.312	0.680	28.312	0.680	28.312	0.680
	WAN_{final}	31.191	0.510	30.416	0.516	30.070	0.580
	$MOR_{inicial}$	0.352	0.680	0.352	0.680	0.352	0.680
	MOR_{final}	0.447	0.391	0.381	0.415	0.373	0.642
	$TRE_{inicial}$	0.552	0.560	0.552	0.560	0.552	0.560
	TRE_{final}	0.725	0.354	0.548	0.379	0.534	1.123

Tabla 3.13: SBRDs *complejos* mejorados: Características Mejor *Inter_C*.

Método	Conjunto de datos	Mejor <i>Inter_C</i>							
		ECM_{tra}	ECM_{tst}	NR	S	R	I	$C(\%)$	
LINGB	PLA _{inicial}	2.606	2.636	31.00	0.1610	0	0	100	
	PLA _{final}	3.284	3.286	18.00	0.1423	-	-	98.68	
	QUA _{inicial}	0.035	0.037	98.00	0.2286	0	0	100	
	QUA _{final}	0.035	0.037	46.00	0.2180	-	-	100	
	ELE _{inicial}	556228	598472	100.20	0.2266	0	0	99.35	
	ELE _{final}	577220	612939	59.40	0.2082	-	-	99.35	
	ABA _{inicial}	5.636	6.370	500.00	0.3496	0.0062	0.0030	95.52	
	ABA _{final}	5.247	6.128	356.40	0.3355	0.0046	0.0015	95.52	
	STP _{inicial}	1.307	1.727	433.80	0.2145	0.0093	0.0026	100	
	STP _{final}	1.213	1.812	296.00	0.2098	0.0070	0.0010	100	
	WIZ _{inicial}	10.103	17.471	500.00	0.4418	0.0106	0.0042	99.97	
	WIZ _{final}	9.819	19.099	373.60	0.4257	0.0069	0.0019	99.97	
	WAN _{inicial}	21.836	33.620	500.00	0.3899	0.0080	0.0032	98.50	
	WAN _{final}	21.239	35.156	388.20	0.3763	0.0050	0.0017	98.50	
	MOR _{inicial}	0.337	0.510	301.60	0.2548	0.0260	0.0062	99.88	
	MOR _{final}	0.226	0.415	171.60	0.2510	0.0217	0.0019	99.70	
	TRE _{inicial}	0.491	0.673	305.40	0.2426	0.0275	0.0044	100	
	TRE _{final}	0.333	0.611	171.40	0.2434	0.0220	0.0012	99.69	
	SCATB	PLA _{inicial}	2.783	2.821	96.60	0.1622	0.0004	0.0058	100
		PLA _{final}	3.206	3.172	36.20	0.1147	0	0	100
QUA _{inicial}		0.046	0.050	243.80	0.2649	0.0002	0.0022	98.13	
QUA _{final}		0.036	0.041	145.80	0.2240	0	3e-5	98.12	
ELE _{inicial}		56584	100229	129.80	0.2661	0.0032	0.0010	96.85	
ELE _{final}		54089	97416	111.80	0.2541	0.0006	0.0007	96.84	
ABA _{inicial}		5.033	6.247	298.00	0.3312	0.0004	0.0012	99.96	
ABA _{final}		4.276	5.385	208.20	0.3048	3e-5	0.0001	99.96	
STP _{inicial}		0.426	0.698	163.60	0.1854	3e-5	0.0002	100	
STP _{final}		0.451	0.755	133.60	0.1749	0	0	99.45	
WIZ _{inicial}		1.788	21.934	466.40	0.3603	0.6e-5	0	100	
WIZ _{final}		1.743	23.251	358.00	0.3472	0.2e-5	-	100	
WAN _{inicial}		2.593	28.312	537.60	0.3119	0.3e-5	0.1e-5	99.93	
WAN _{final}		2.505	31.191	415.40	0.3053	0	0	99.93	
MOR _{inicial}		0.085	0.352	92.20	0.2844	0.0006	0.0001	99.99	
MOR _{final}		0.097	0.447	64.13	0.2759	0	4e-5	99.82	
TRE _{inicial}		0.150	0.552	76.60	0.2893	0.0001	0.0001	99.99	
TRE _{final}		0.208	0.725	46.73	0.2760	0	0	99.46	

Tabla 3.14: SBRDs complejos mejorados: Características Mediana $Pre - Inter_C$.

Método	Conjunto de datos	Mediana $Pre - Inter_C$						
		ECM_{tra}	ECM_{tst}	NR	S	R	I	$C(\%)$
LINGB	$PLA_{inicial}$	2.606	2.636	31.00	0.1610	0	0	100
	PLA_{final}	2.733	2.824	23.00	0.1499	-	-	100
	$QUA_{inicial}$	0.035	0.037	98.00	0.2286	0	0	100
	QUA_{final}	0.035	0.037	52.20	0.2236	-	-	100
	$ELE_{inicial}$	556228	598472	100.20	0.2266	0	0	99.35
	ELE_{final}	555512	598735	67.60	0.2093	-	-	99.35
	$ABA_{inicial}$	5.636	6.370	500.00	0.3496	0.0062	0.0030	95.52
	ABA_{final}	5.151	6.093	362.60	0.3362	0.0047	0.0016	95.52
	$STP_{inicial}$	1.307	1.727	433.80	0.2145	0.0093	0.0026	100
	STP_{final}	1.090	1.693	300.20	0.2100	0.0071	0.0012	100
	$WIZ_{inicial}$	10.103	17.471	500.00	0.4418	0.0106	0.0042	99.97
	WIZ_{final}	9.497	18.829	380.60	0.4270	0.0072	0.0021	99.97
	$WAN_{inicial}$	21.836	33.620	500.00	0.3899	0.0080	0.0032	98.50
	WAN_{final}	20.772	35.003	392.80	0.3779	0.0052	0.0019	97.22
	$MOR_{inicial}$	0.337	0.510	301.60	0.2548	0.0260	0.0062	99.88
	MOR_{final}	0.196	0.382	179.00	0.2514	0.0217	0.0026	99.88
$TRE_{inicial}$	0.491	0.673	305.40	0.2426	0.0275	0.0044	100	
TRE_{final}	0.284	0.584	176.60	0.2428	0.0217	0.0017	99.69	
SCATB	$PLA_{inicial}$	2.783	2.821	96.60	0.1622	0.0004	0.0058	100
	PLA_{final}	2.409	2.567	45.80	0.1240	0	0	100
	$QUA_{inicial}$	0.046	0.050	243.80	0.2649	0.0002	0.0022	98.13
	QUA_{final}	0.033	0.038	154.80	0.2315	0	0.0002	98.12
	$ELE_{inicial}$	56584	100229	129.80	0.2661	0.0032	0.0010	96.85
	ELE_{final}	41899	87004	113.60	0.2561	0.0012	0.0007	96.84
	$ABA_{inicial}$	5.033	6.247	298.00	0.3312	0.0004	0.0012	99.96
	ABA_{final}	4.087	5.273	216.20	0.3075	5e-5	0.0002	99.96
	$STP_{inicial}$	0.426	0.698	163.60	0.1854	3e-5	0.0002	100
	STP_{final}	0.409	0.749	147.40	0.1806	3e-5	0.0001	99.74
	$WIZ_{inicial}$	1.788	21.934	466.40	0.3603	0.6e-5	0	100
	WIZ_{final}	1.655	23.182	365.87	0.3488	0.2e-5	-	100
	$WAN_{inicial}$	2.593	28.312	537.60	0.3119	0.3e-5	0.1e-5	99.93
	WAN_{final}	2.374	30.416	428.60	0.3062	0	0	99.93
	$MOR_{inicial}$	0.085	0.352	92.20	0.2844	0.0006	0.0001	99.99
	MOR_{final}	0.081	0.381	69.33	0.2806	0	0.0001	99.80
$TRE_{inicial}$	0.150	0.552	76.60	0.2893	0.0001	0.0001	99.99	
TRE_{final}	0.151	0.548	54.93	0.2822	0	0	99.58	

Tabla 3.15: SBRDs *complejos* mejorados: Características Mejor *Pre*.

Método	Conjunto de datos	Mejor <i>Pre</i>						
		ECM_{tra}	ECM_{tst}	NR	S	R	I	$C(\%)$
LINGB	PLA _{inicial}	2.606	2.636	31.00	0.1610	0	0	100
	PLA _{final}	2.606	2.655	29.60	0.1618	-	-	100
	QUA _{inicial}	0.035	0.037	98.00	0.2286	0	0	100
	QUA _{final}	0.035	0.037	62.80	0.2313	-	-	100
	ELE _{inicial}	556228	598472	100.20	0.2266	0	0	99.35
	ELE _{final}	550608	593548	77.60	0.2127	-	-	99.35
	ABA _{inicial}	5.636	6.370	500.00	0.3496	0.0062	0.0030	95.52
	ABA _{final}	5.105	6.019	374.80	0.3388	0.0051	0.0018	95.52
	STP _{inicial}	1.307	1.727	433.80	0.2145	0.0093	0.0026	100
	STP _{final}	1.048	1.664	312.20	0.2106	0.0073	0.0015	100
	WIZ _{inicial}	10.103	17.471	500.00	0.4418	0.0106	0.0042	99.97
	WIZ _{final}	9.400	18.860	392.80	0.4295	0.0078	0.0024	99.97
	WAN _{inicial}	21.836	33.620	500.00	0.3899	0.0080	0.0032	98.50
	WAN _{final}	20.639	34.684	402.20	0.3798	0.0055	0.0020	96.46
	MOR _{inicial}	0.337	0.510	301.60	0.2548	0.0260	0.0062	99.88
	MOR _{final}	0.191	0.377	191.80	0.2528	0.0224	0.0032	99.88
	TRE _{inicial}	0.491	0.673	305.40	0.2426	0.0275	0.0044	100
	TRE _{final}	0.275	0.572	187.40	0.2430	0.0219	0.0023	99.84
SCATB	PLA _{inicial}	2.783	2.821	96.60	0.1622	0.0004	0.0058	100
	PLA _{final}	2.249	2.374	55.40	0.1452	0.0007	0.0013	100
	QUA _{inicial}	0.046	0.050	243.80	0.2649	0.0002	0.0022	98.13
	QUA _{final}	0.033	0.038	163.60	0.2355	3e-5	0.0004	98.12
	ELE _{inicial}	56584	100229	129.80	0.2661	0.0032	0.0010	96.85
	ELE _{final} ELE-2(F)	39839	85157	115.20	0.2540	0.0014	0.0011	96.84
	ABA _{inicial}	5.033	6.247	298.00	0.3312	0.0004	0.0012	99.96
	ABA _{final}	4.028	5.240	223.00	0.3096	0.0001	0.0003	99.96
	STP _{inicial}	0.426	0.698	163.60	0.1854	3e-5	0.0002	100
	STP _{final}	0.396	0.686	149.60	0.1871	4e-5	0.0002	99.86
	WIZ _{inicial}	1.788	21.934	466.40	0.3603	0.6e-5	0	100
	WIZ _{final}	1.614	23.067	380.53	0.3513	0.3e-5	-	100
	WAN _{inicial}	2.593	28.312	537.60	0.3119	0.3e-5	0.1e-5	99.93
	WAN _{final}	2.326	30.070	447.00	0.3074	0.2e-5	0	99.93
	MOR _{inicial}	0.085	0.352	92.20	0.2844	0.0006	0.0001	99.99
	MOR _{final}	0.074	0.373	79.47	0.2848	0.0007	0.0001	99.94
	TRE _{inicial}	0.150	0.552	76.60	0.2893	0.0001	0.0001	99.99
	TRE _{final}	0.138	0.534	65.93	0.2940	0.0001	0.0002	99.76

Tabla 3.16: Test de Wilcoxon para los SBRDs *Complejos* LINGB: modelo base (R+) y modelo mejorado (R-).

SBRD Mejor <i>Inter_C</i>				
Medidas	R+	R-	Hipótesis ($\alpha=0.10$)	<i>p</i> -valor
<i>ECM_{tst}</i>	17.0	28.0	Aceptada	0.515
<i>NR</i>	45.0	0.0	Rechazada	0.008
<i>S</i>	44.0	1.0	Rechazada	0.011
<i>R</i>	41.0	2.0	Rechazada	0.028
<i>I</i>	41.0	2.0	Rechazada	0.028
<i>NC</i>	6.0	36.0	Rechazada	0.028
<i>Inter_C</i>	45.0	0.0	Rechazada	0.008
SBRD Mediana <i>Pre – Inter_C</i>				
Medidas	R+	R-	Hipótesis ($\alpha=0.10$)	<i>p</i> -valor
<i>ECM_{tst}</i>	27.0	18.0	Aceptada	0.594
<i>NR</i>	45.0	0.0	Rechazada	0.008
<i>S</i>	44.0	1.0	Rechazada	0.011
<i>R</i>	41.0	2.0	Rechazada	0.028
<i>I</i>	41.0	2.0	Rechazada	0.028
<i>NC</i>	12.0	29.0	Rechazada	0.080
<i>Inter_C</i>	45.0	0.0	Rechazada	0.008
SBRD Mejor <i>Pre</i>				
Medidas	R+	R-	Hipótesis ($\alpha=0.10$)	<i>p</i> -valor
<i>ECM_{tst}</i>	33.0	12.0	Aceptada	0.214
<i>NR</i>	45.0	0.0	Rechazada	0.008
<i>S</i>	38.0	7.0	Rechazada	0.066
<i>R</i>	41.0	2.0	Rechazada	0.028
<i>I</i>	41.0	2.0	Rechazada	0.028
<i>NC</i>	12.0	29.0	Rechazada	0.080
<i>Inter_C</i>	45.0	0.0	Rechazada	0.008

Tabla 3.17: Test de Wilcoxon para los SBRDs *Complejos* SCATB: SBRD base (R+) y SBRD mejorado (R-).

SBRD Mejor <i>Inter_C</i>				
Medidas	R+	R-	Hipótesis ($\alpha=0.10$)	<i>p-valor</i>
<i>ECM_{tst}</i>	14.0	31.0	Aceptada	0.314
<i>NR</i>	45.0	0.0	Rechazada	0.008
<i>S</i>	45.0	0.0	Rechazada	0.008
<i>R</i>	45.0	0.0	Rechazada	0.008
<i>I</i>	44.0	0.0	Rechazada	0.012
<i>NC</i>	5.0	40.0	Rechazada	0.028
<i>Inter_C</i>	45.0	0.0	Rechazada	0.008
SBRD Mediana <i>Pre – Inter_C</i>				
Medidas	R+	R-	Hipótesis ($\alpha=0.10$)	<i>p-valor</i>
<i>ECM_{tst}</i>	31.0	14.0	Aceptada	0.314
<i>NR</i>	45.0	0.0	Rechazada	0.008
<i>S</i>	45.0	0.0	Rechazada	0.008
<i>R</i>	42.0	3.0	Rechazada	0.021
<i>I</i>	44.0	0.0	Rechazada	0.012
<i>NC</i>	5.0	40.0	Rechazada	0.028
<i>Inter_C</i>	45.0	0.0	Rechazada	0.008
SBRD Mejor <i>Pre</i>				
Medidas	R+	R-	Hipótesis ($\alpha=0.10$)	<i>p-valor</i>
<i>ECM_{tst}</i>	33.0	12.0	Aceptada	0.214
<i>NR</i>	45.0	0.0	Rechazada	0.008
<i>S</i>	38.0	7.0	Rechazada	0.066
<i>R</i>	26.0	19.0	Aceptada	0.678
<i>I</i>	29.0	15.0	Aceptada	0.401
<i>NC</i>	5.0	40.0	Rechazada	0.028
<i>Inter_C</i>	34.0	11.0	Aceptada	0.173

El test de Wilcoxon para los SBRDs *complejos* de SCATB (Tabla 3.17) muestra que:

- **SBRDs Mejor *Inter_C***: mejoran los índices de interpretabilidad aunque empeoran su precisión. Teniendo en cuenta la alta precisión de los modelos difusos base, la disminución existente es asumible.
- **SBRDs Mediana *Pre – Inter_C***: mejoran los índices de interpretabilidad mientras conservan la precisión de los modelos difusos base.
- **SBRDs Mejor *Pre***: no muestran diferencias estadísticas significativas en ninguno de los índices considerados, ni de precisión ni de interpretabilidad. En este caso la explicación de no variación de la interpretabilidad de los modelos es análoga al caso de los modelos *compactos* comentado previamente.

3.3.4. Análisis Global

En general, los resultados obtenidos muestran una mejora razonable de la interpretabilidad en los modelos neuro-difusos (precisos), tanto lingüísticos como aproximativos. Esta mejora de la interpretabilidad, que implica una reducción de la complejidad de los modelos, se alcanza con una pérdida aceptable de la precisión la cual, en algunos casos, incluso aumenta. En general la precisión se mantiene en los mismos órdenes de magnitud que la precisión de los modelos difusos base. De hecho en algunos casos una mejora de la interpretabilidad no ha implicado una pérdida de precisión (ver Tablas 3.6 y 3.12). La mejora del equilibrio precisión-interpretabilidad, esta última basada en índices de complejidad, ha sido alcanzada mediante el uso de un enfoque genético centrado en los objetivos finales.

Basándose en el análisis de las mejoras obtenidas en los SBRDs característicos analizados en el Frente de Pareto (Mejor *Pre*, Mejor *Inter_C* y Mediana *Pre – Inter_C*) se puede concluir que:

- **SBRD Mejor *Inter_C***: se mejora la interpretabilidad y se reduce la complejidad, aunque el deterioro en la precisión es mayor que en otros SBRDs (Mediana *Pre – Inter_C* y Mejor *Pre*). En estos casos el error alcanzado en los modelos es más significativo, pero teniendo en cuenta el hecho de que los SBRDs base tienen en general un nivel de precisión elevado, los

SBRDs mejorados siguen siendo suficientemente precisos, incluso a pesar del empeoramiento del error.

- **SBRD Mediana *Pre* – *Inter_C***: se reduce la complejidad, mejorando el valor de los índices considerados en la métrica *Inter_C*, sin una pérdida significativa del error. De hecho la mejora de la interpretabilidad se alcanza manteniendo el nivel de precisión.
- **SBRD Mejor *Pre***: la mejora es superior en el caso de los SBRDs *compactos* LINGA, aunque en los SBRDs *complejos* LINGB también se consigue mejorar. Esto no ocurre con los SBRDs SCATA y SCATB en los que tanto la precisión como la interpretabilidad de los modelos base y los modelos mejorados son similares.

Si se comparan de nuevo los SBRDs finales, a pesar de los distintos enfoques de ambos métodos, con el algoritmo Wang & Mendel mostrado en [77], se observa que la precisión es mejor y que el equilibrio precisión-interpretabilidad ha sido altamente mejorado para estos modelos neuro-difusos, lingüísticos o aproximativos, y que aunque se ha conseguido disminuir el número de reglas en los SBRDs mejorados, este sigue siendo más elevado que en Wang & Mendel.

Por otro lado, cuando la interpretabilidad de las particiones difusas es tenida en cuenta desde un punto de vista clásico, si se observa una diferencia significativa entre los dos algoritmos lingüísticos y aproximativos aquí utilizados. Así con el primero de los algoritmos este aspecto está garantizado, aunque podría mejorarse, a través de la inicialización por parte del usuario. Sin embargo, con el algoritmo aproximativo esto no está garantizado, sobre todo en los SBRDs *complejos*, y es recomendable realizar algún otro tipo de post-procesamiento centrado en los conjuntos difusos generados mediante técnicas de agrupamiento.

Como ya se comentó inicialmente la influencia en la métrica final de cada uno de los índices individuales de complejidad considerados es muy diferente. En esta propuesta, las características propias de los algoritmos de modelado difuso considerados hacen que la influencia final de los índices sea distinta según se trabaje con algoritmos lingüísticos o aproximativos. Por ejemplo, en los SBRDs del conjunto de datos ELE el porcentaje de reducción conseguido en relación al *NR* es mayor cuando se usa el algoritmo lingüístico, ya que en estos casos no se está teniendo en cuenta la influencia ni de la redundancia ni de la incoherencia, al tener estas un valor de partida igual a 0.

La propuesta presenta también otro pequeño inconveniente que es posible mejorar, el elevado tiempo de ejecución necesario para realizar el post-procesamiento, sobre todo por la complejidad del cálculo de algunas de las medidas de interpretabilidad al tratar grandes conjuntos de datos.

Resumiendo, la propuesta ha permitido comprobar que es posible mejorar el equilibrio precisión-interpretabilidad de los SBRDs, evaluando esta última en términos de complejidad e interpretabilidad semántica. En el siguiente capítulo se presentan las mejoras realizadas en la etapa de post-procesamiento, que tomando como base las conclusiones aquí obtenidas, buscan solventar algunos de los inconvenientes actuales proponiendo nuevas medidas de interpretabilidad que permitan trabajar con un ajuste de las funciones de pertenencia y que sean computacionalmente menos costosas.

Capítulo 4

Interpretabilidad de los sistemas basados en reglas difusas lingüísticos y aproximativos: Generalización de GM3M y nueva métrica RMI

En el capítulo anterior se ha visto que es posible mejorar el equilibrio precisión-interpretabilidad de los SBRDs a través de un proceso de post-procesamiento utilizando estrategias evolutivas y diversos índices de interpretabilidad. A partir de esta primera aproximación se plantea un siguiente paso, proponer métricas de interpretabilidad que puedan ser utilizadas con cualquier SBRDs, independientemente de su naturaleza (lingüística o aproximativa) y del tipo de función de pertenencia utilizada (triangular, trapezoidal, gaussiana, etc.). Además, se pretende que dichas medidas tengan en cuenta la semántica que un experto pudiese proporcionar respecto a los términos lingüísticos que es capaz de entender cuando dicho experto este disponible, así como el sistema de inferencia utilizado, ya que podría afectar al comportamiento de las reglas. Todo ello mediante el diseño de

medidas que sean intuitivas para que puedan ayudar al análisis o selección del modelo final, teniendo en cuenta que no deben ser computacionalmente costosas.

Para abordar dichos objetivos, esta nueva aproximación propone las siguientes dos ideas principales: generalizar la métrica relativa de interpretabilidad semántica de la BD GM3M y presentar una nueva medida de interpretabilidad semántica de la BR llamada RMI. Para mostrar la utilidad de ambas medidas, y para ver cómo afectan a los procesos de modelado, ambas métricas son utilizadas para guiar el proceso de búsqueda de un AEMO que realiza simultáneamente selección de reglas y ajuste de las particiones difusas con el objetivo de generar SBRDs que sean a la vez precisos e interpretables.

Adicionalmente, se pretende que mediante las nuevas métricas propuestas, que van a poder ser utilizadas con cualquier tipo de SBRD, se pueda realizar un análisis local del equilibrio precisión-interpretabilidad entre distintos SBRDs. La idea es que una aplicación local de las métricas sobre los distintos elementos del SBRD ayude al usuario a seleccionar en cada momento el SBRD más adecuado a sus necesidades.

El capítulo se organiza como sigue: en la Sección 4.1 se presentan las dos métricas de interpretabilidad semántica utilizadas, la generalización de GM3M y la nueva métrica RMI. Posteriormente, en la Sección 4.2 se describen la metodología y los AEMOs utilizados en la etapa de post-procesamiento. Finalmente, en la Sección 4.3 se presentan los experimentos realizados y se analizan los resultados obtenidos.

4.1. Evaluación de la interpretabilidad semántica de los sistemas basados en reglas difusas: Lingüísticos vs aproximativos

Para la evaluación y comparación de los niveles de interpretabilidad de cualquier SBRD es necesario definir métricas que puedan ser aplicadas tanto en sistemas de naturaleza lingüística, como en sistemas de naturaleza aproximativa.

Tomando como punto de partida el análisis realizado en el Capítulo 2 de esta memoria, así como la taxonomía propuesta en [29], se tiene que:

- Las métricas más novedosas de interpretabilidad semántica a nivel de BD

son las medidas relativas que evalúan la semántica de una partición difusa respecto a otra, entre las que se encuentra la métrica GM_{3M} [77]. Tal y como está definido actualmente dicho índice puede ser utilizado con SBRDs lingüísticos con funciones de pertenencia triangulares, por lo que es necesario extender su definición para su utilización con cualquier tipo de funciones de pertenencia y con SBRD basados en reglas aproximativas.

- Donde menos contribuciones se han realizado hasta el momento es dentro de la categoría de métricas de interpretabilidad semántica a nivel de BR, con medidas en algunos casos computacionalmente costosas que no tienen en cuenta el sistema de inferencia del modelo, por lo que parece necesario proponer alguna nueva métrica de interpretabilidad que no sólo evalúe la consistencia de la BR sino que, a la hora de evaluar la cooperación entre reglas, tenga también en cuenta la influencia del mecanismo de inferencia concreto utilizado por el SBRD y que sea fácil de calcular.

Se presenta a continuación la generalización de la métrica de interpretabilidad semántica a nivel de BD GM_{3M} y la definición de una nueva métrica de interpretabilidad semántica a nivel de BR llamada RMI (*Rule Meaning Index*), para finalmente introducir una metodología de análisis y comparación de BRs lingüísticas frente a aproximativas, considerando estos últimos como SBRDs pseudo-lingüísticos.

4.1.1. Generalización de la métrica GM_{3M} para la interpretabilidad a nivel de base de datos

GM_{3M} [77] es una conocida métrica relativa de interpretabilidad utilizada en SBRDs lingüísticos para evaluar la semántica de funciones de pertenencia triangulares cuando son ajustadas respecto a una partición difusa interpretable previamente definida. Dicha partición interpretable puede ser definida bien por un experto, por un método de aprendizaje automático o, tal y como se considera en [77], como una partición difusa fuerte uniformemente distribuida.

La métrica fue propuesta con el objetivo de ayudar a conservar la interpretabilidad semántica inicial de un SBRD lingüístico dentro de un proceso de optimización de las funciones de pertenencia. En el caso de los SBRDs lingüísticos, el ajuste implica una modificación de la definición de las funciones de pertenencia asociadas con los términos lingüísticos considerados inicialmente en el sistema.

Es decir, se está modificando el significado del término lingüístico previamente definido (una BD inicial compuesta por conceptos semánticos y las funciones de pertenencia asociadas que dan significado a dichos conceptos).

A continuación se muestra brevemente como se realiza el cálculo de la métrica GM3M, para posteriormente extender su definición de forma que pueda ser utilizada para evaluar la interpretabilidad semántica de cualquier partición difusa, independientemente de su forma (triangular, trapezoidal, gaussiana,...) o tipo de SBRD (lingüístico o aproximativo).

Métrica GM3M

Formalmente, la métrica GM3M se define como la media geométrica de tres índices complementarios que sirven para cuantificar la proximidad entre una función de pertenencia dada y la función inicialmente definida para el término lingüístico asociado. Cada índice tiene en cuenta, respectivamente, aspectos tales como el desplazamiento (δ), diferencias en la amplitud lateral (γ) y diferencias en el área (ρ) entre funciones de pertenencia:

$$GM3M = \sqrt[3]{\delta \cdot \gamma \cdot \rho} \quad (4.1)$$

Dichos índices se definieron para medir la interpretabilidad cuando se realiza un proceso de ajuste en las funciones de pertenencia, que implicaría una modificación de su definición. En general, dicho ajuste es necesario para conseguir mejorar el nivel de precisión del sistema. Con el uso de la media geométrica se pretende que, en el caso de que uno de los índices tome valores muy pequeños (baja interpretabilidad), el valor de GM3M tome también valores pequeños. Con todo, GM3M toma valores en el rango $[0, 1]$ siendo 0 el nivel más bajo de interpretabilidad y 1 el nivel más alto.

Dada una función difusa triangular F_j , se identifican los parámetros de definición de la función original asociada como (a_j, b_j, c_j) y los parámetros de la función ajustada como (a'_j, b'_j, c'_j) , los cuales pueden moverse dentro de unos intervalos de variación identificados respectivamente como $[I_{a_j}^l, I_{a_j}^r]$, $[I_{b_j}^l, I_{b_j}^r]$ y $[I_{c_j}^l, I_{c_j}^r]$ (ver Fig. 4.1). Estos intervalos establecen la variación máxima de cada uno de los parámetros, y pueden ser definidos de forma diferente en cada problema.

De forma individual, cada uno de los índices anteriores se calcula como sigue:

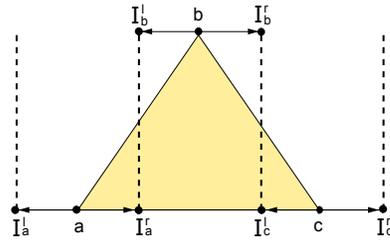


Figura 4.1: Intervalos de variación de una función de pertenencia triangular.

- El índice δ permite controlar el desplazamiento de las funciones de pertenencia en su punto central. Su proceso de cálculo se basa en: medir la distancia normalizada entre el punto central de la función de pertenencia ajustada y el punto central de la función de pertenencia original asociada, para finalmente quedarse con el desplazamiento máximo de todas las funciones de pertenencia del sistema. Para cada F_j de la partición difusa lingüística se define $\delta_j = \frac{|b_j - b'_j|}{I}$, donde $I = \frac{I_{b_j}^r - I_{b_j}^l}{2}$ representa la máxima variación de cada parámetro central. Entonces δ^* se define como $\delta^* = \max_j \{\delta_j\}$ (el peor caso). δ^* toma valores entre 0 y 1 (valores cercanos a 1 muestran que la función de pertenencia presenta un gran desplazamiento). Finalmente se realiza la siguiente transformación, de forma que el índice represente proximidad (maximización): Maximizar $\delta = 1 - \delta^*$.
- El índice γ permite controlar la forma de la función de pertenencia. Su proceso de cálculo se basa en medir la relación entre las partes derecha e izquierda del soporte de las funciones de pertenencia original y ajustada. Se define $leftS_j = |a_j - b_j|$ como la amplitud de la parte izquierda del soporte de la función de pertenencia original, $rightS_j = |b_j - c_j|$ como la amplitud de la parte derecha, $leftS'_j = |a'_j - b'_j|$ y $rightS'_j = |b'_j - c'_j|$ como las amplitudes correspondientes a la función de pertenencia ajustada. Entonces, el valor γ_j para cada función de pertenencia se calcula como:

$$\gamma_j = \frac{\min\{leftS_j/rightS_j, leftS'_j/rightS'_j\}}{\max\{leftS_j/rightS_j, leftS'_j/rightS'_j\}}. \quad (4.2)$$

Valores cercanos a 1 indican que la relación de aspecto entre la parte izquierda y derecha del soporte de la función de pertenencia ajustada se mantiene

respecto a la función de pertenencia original. Finalmente γ se calcula obteniendo el valor mínimo de las γ_j (el peor caso) con un claro objetivo de maximización: Maximizar $\gamma = \min_j\{\gamma_j\}$.

- El índice ρ permite controlar el área de las funciones de pertenencia. Su proceso de cálculo se basa en medir la relación entre las áreas de las funciones de pertenencia original y ajustada. Si se define A_j como el área del triángulo correspondiente a la función de pertenencia original, y A'_j como el nuevo área, entonces el valor ρ_j para cada función de pertenencia se calcula utilizando la siguiente ecuación: $\rho_j = \frac{\min\{A_j, A'_j\}}{\max\{A_j, A'_j\}}$.

Valores cercanos a 1 indican que el área de la función de pertenencia original y el área de la función de pertenencia ajustada son muy similares (cambios mínimos). Finalmente el índice ρ se calcula obteniendo el valor mínimo de las ρ_j (el peor caso) con un claro objetivo de maximización: Maximizar $\rho = \min_j\{\rho_j\}$.

Aunque inicialmente estos índices fueron propuestos para medir la interpretabilidad de funciones de pertenencia triangulares definidas dentro de un marco de trabajo lingüístico, su definición tiene que ser extendida realizando cambios en la formulación que permitan su aplicación para funciones de pertenencia trapezoidales o gaussianas, así como en sistemas con particiones difusas aproximativas.

Extensión de GM3M

Siguiendo la filosofía establecida al definir la métrica GM3M original (presentada en el apartado anterior), lo primero que es necesario establecer es la asociación entre una función de pertenencia cualquiera y su función de pertenencia de referencia (función de pertenencia original interpretable definida bien por un experto, bien por un proceso de aprendizaje automático o bien como una partición difusa fuerte uniformemente distribuida) respecto a la cual medir el desplazamiento (δ), la amplitud lateral (γ) y la similitud del área (ρ). Posteriormente, tras parametrizar distintos tipos de funciones de pertenencia, se presentan las modificaciones realizadas en los cálculos de los índices.

Una vez extendida la definición de la métrica GM3M se podrá comparar la interpretabilidad semántica a nivel de partición difusa de distintos SBRDs, independientemente del tipo de funciones de pertenencia que utilicen o de su naturaleza. Es decir, se podría tener una función de pertenencia original de forma

triangular definida por un experto y cuantificar cómo de diferente es de una función de pertenencia trapezoidal obtenida con un algoritmo de aprendizaje. Si la función de pertenencia trapezoidal no es muy diferente de la triangular entonces GM3M tendrá valores altos, indicando que la nueva función de pertenencia representa un concepto similar al que el experto sería capaz de entender.

Asociación entre funciones de pertenencia

Dado que ahora el entorno de trabajo no tiene por qué implicar necesariamente un ajuste de las funciones de pertenencia, la asociación entre funciones de pertenencia ya no está tan clara como antes, por lo que es necesario redefinirla. Además, tiene que poder usarse con distintos tipos de funciones de pertenencia (trapezoidal, gaussiano, etc.) y SBRDs (lingüísticos o aproximativos). Por ello, la forma de realizar esta nueva asociación es utilizando el punto medio del conjunto α -corte de ambas funciones de pertenencia, con $\alpha = 0.5$. De esta forma dada una función de pertenencia cualquiera se busca la función de pertenencia original de referencia cuyo punto medio del conjunto 0.5-corte esté más cercano al punto medio del conjunto 0.5-corte de la función de pertenencia dada.

En la Fig. 4.2 se pueden ver algunos ejemplos de cómo realizar dicha asociación. Concretamente en el primer ejemplo se puede observar que la función de pertenencia etiquetada como “?” es asociada con la función de pertenencia original etiquetada como “M”, ya que la distancia entre los puntos medios de sus 0.5-corte es menor que la distancia al resto de funciones de pertenencia candidatas. Esta asociación parece más razonable/adecuada que la obtenida al usar directamente los puntos centrales de las funciones de pertenencia, ya que la similitud entre la función de pertenencia dada y “M” es mayor que la existente con la función de pertenencia etiquetada como “S”. Es por esto por lo que se propone el uso del 0.5-corte y no el uso de los puntos centrales. En el segundo ejemplo la función de pertenencia “?” es asociada con “M”.

Así, a cada F' presente en el SBRD para el cual queremos calcular su valor GM3M se le asocia su correspondiente función de pertenencia original o interpretable. Debido a la posible naturaleza aproximativa del modelo a comparar, o al posible uso de una granularidad distinta varias F' dadas podrían ser asociadas con la misma función de pertenencia original o interpretable, por lo que la asociación ya no sería uno a uno. Por ello en el resto de la sección se van a eliminar los subíndices de las formulaciones, utilizando sólo las primas para identificar las funciones de pertenencia del SBRD aprendido o ajustado, respecto a la función de pertenencia interpretable que le ha sido asociada.

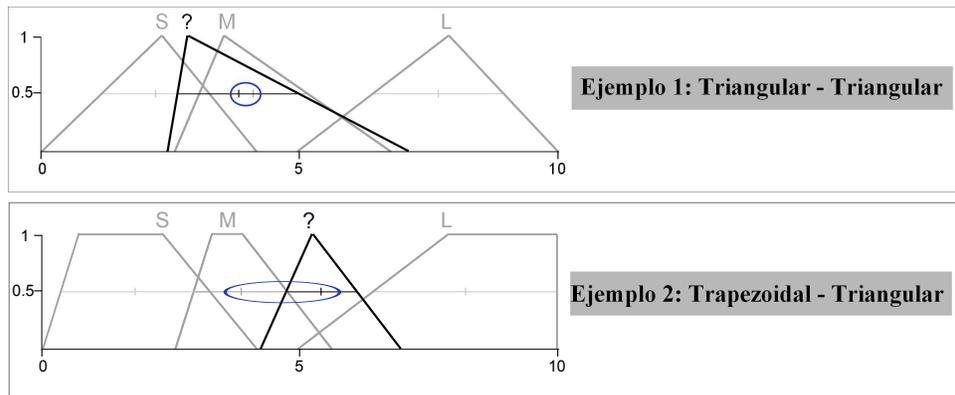


Figura 4.2: Ejemplos de cálculo de GM3M.

Parametrización de distintas funciones de pertenencia

El siguiente paso consiste en determinar los puntos característicos de las funciones de pertenencia (ver Fig. 4.3) y los intervalos de variación de dichos puntos que permitirán calcular el índice δ y finalmente la métrica GM3M.

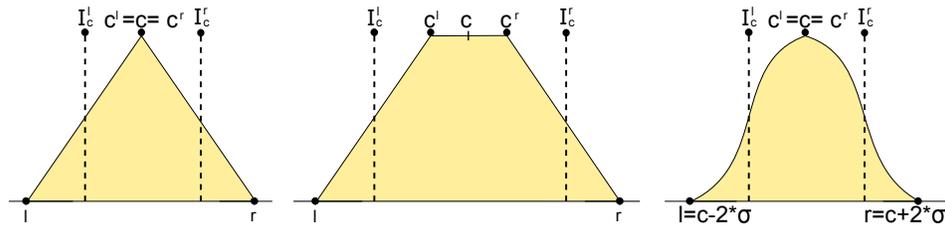


Figura 4.3: Definición de puntos característicos e intervalos de variación para distintos tipos de funciones de pertenencia.

La forma de parametrizar las funciones de pertenencia, en función de la forma que tenga, es como se describe a continuación:

- Triangular:** (l, c, r) son los puntos izquierdo, central y derecho característicos de la función de pertenencia, siendo c^l y c^r los puntos izquierdo y derecho característicos del núcleo de la función de pertenencia. En este caso los puntos característicos del núcleo coinciden con el punto central de la función de pertenencia $c^l = c^r = c$, es decir, la posición central característica de la

función de pertenencia.

- **Trapezoidal:** (l, c^l, c^r, r) son respectivamente los puntos izquierdo, central izquierdo, central derecho y derecho característicos de la función de pertenencia, siendo $c = (c^l + c^r)/2,0$ el centro del núcleo, es decir, la posición central característica de la función de pertenencia.
- **Gaussiana:** (c, σ) son respectivamente el punto central y la amplitud de la función de pertenencia, siendo c^l y c^r los puntos izquierdo y derecho característicos del núcleo de la función de pertenencia. En este caso los puntos característicos del núcleo coinciden con el punto central de la función de pertenencia $c^l = c^r = c$, es decir, la posición central característica de la función de pertenencia. De forma adicional, en este caso se define $l = c - 2\sigma$ y $r = c + 2\sigma$.

Una vez que las funciones de pertenencia han sido parametrizadas se define el correspondiente intervalo de variación sobre la función de pertenencia de referencia (original) $[I_c^l, I_c^r]$, con $I_c^l = \frac{l+c^l}{2,0}$ y $I_c^r = \frac{c^r+r}{2,0}$. Este intervalo representa el rango en el que el punto central de la función de pertenencia debería encontrarse para que no se pierda por completo su significado.

Cálculo de los índices

Por último se realizan las siguiente modificaciones en el cálculo de los índices complementarios necesarios para calcular la métrica:

- **Desplazamiento** (δ): dada una determinada F' , $\delta' = \frac{|c-c'|}{I}$, con $I = \max(c - I_c^l, I_c^r - c)$ representando la máxima variación posible para la posición central característica. Por lo tanto, δ^* se define como $\delta^* = \max_k \{\delta'\}$ (el peor caso). Finalmente se realiza la siguiente transformación, de forma que el índice represente proximidad (maximización): Maximize $\delta = 1 - \delta^*$. Así definido δ^* puede tomar valores negativos cuando la posición central característica esté fuera del rango de variación permitido, lo que implica que δ puede ser mayor que 1, por lo que para cuantificar la interpretabilidad de una función de pertenencia concreta se va a considerar un límite inferior de 0 para el índice δ^* , aunque de cara al proceso de optimización se mantienen los valores negativos.
- **Amplitud lateral** (γ): el valor γ' se calcula de la misma manera, teniendo en cuenta los parámetros asociados según las funciones de pertenencia correspondientes. Se definen:

- $leftS = |c^l - l|$ como la amplitud del soporte de la parte izquierda de la función de pertenencia asociada, y $rightS = |r - c^r|$ como la amplitud de la parte derecha.
- $leftS' = |c'^l - l'|$ y $rightS' = |r' - c'^r|$ como las amplitudes correspondiente en la nueva F' .

Posteriormente, γ se calcula utilizando la misma fórmula que la presentada en la Sección 4.1.1: Maximizar $\gamma = \min_k \{\gamma'\}$.

- **Área (ρ):** por otra parte el cálculo del índice ρ se modifica para permitir la existencia de núcleos no *singleton*, como en el caso de las funciones de pertenencia trapezoidales. En este caso se tienen dos áreas que representan partes conceptuales diferentes: A_s como el área de las partes laterales y A_c como el área del núcleo, que tienen que ser consideradas por separado de cara a detectar cambios en cualquiera de las dos. Si se considerase directamente el área total como en la definición inicial, podría darse el caso de que A_c aumentase y A_s decrementase de manera que al área total fuese la misma. En ese caso si además se mantiene el mismo punto característico y se preserva la relación de aspecto lateral, no se detectarían dichos cambios en la función. En el caso de las funciones de pertenencia triangulares y gaussianas el área A_c será igual a 0.

Por lo tanto, para cada F' , ρ' se calcula usando la siguiente ecuación: $\rho' = \frac{\min\{A_s, A'_s\} + \min\{A_c, A'_c\}}{\max\{A_s, A'_s\} + \max\{A_c, A'_c\}}$. Valores cercanos a 1 indican que el área de la función de pertenencia original y el área de la función de pertenencia dada son muy similares (cambios mínimos). Finalmente el índice ρ se calcula obteniendo el valor mínimo de las ρ' (el peor caso): Maximizar $\rho = \min_k \{\rho'\}$.

La Fig. 4.4 muestra dos ejemplos detallados del cálculo de la métrica GM3M en diferentes casos para una mejor comprensión de lo expuesto anteriormente. En el Ejemplo 1, el valor de GM3M es de 0,545 con una proximidad del punto característico de $\delta = 0,562$, una similitud en la relación de aspecto lateral de $\gamma = 0,330$ y una similitud del área de $\rho = 0,872$: estos valores indican que la función de pertenencia dada no es tan diferente de la original aunque presenta ciertos problemas. Por otro lado, en el Ejemplo 2, el valor de GM3M es de 0,000, lo que muestra que la función de pertenencia M' tiene una interpretabilidad semántica muy baja, debido principalmente al gran desplazamiento (distancia) entre las funciones de pertenencia consideradas ($\delta = 0,000$).

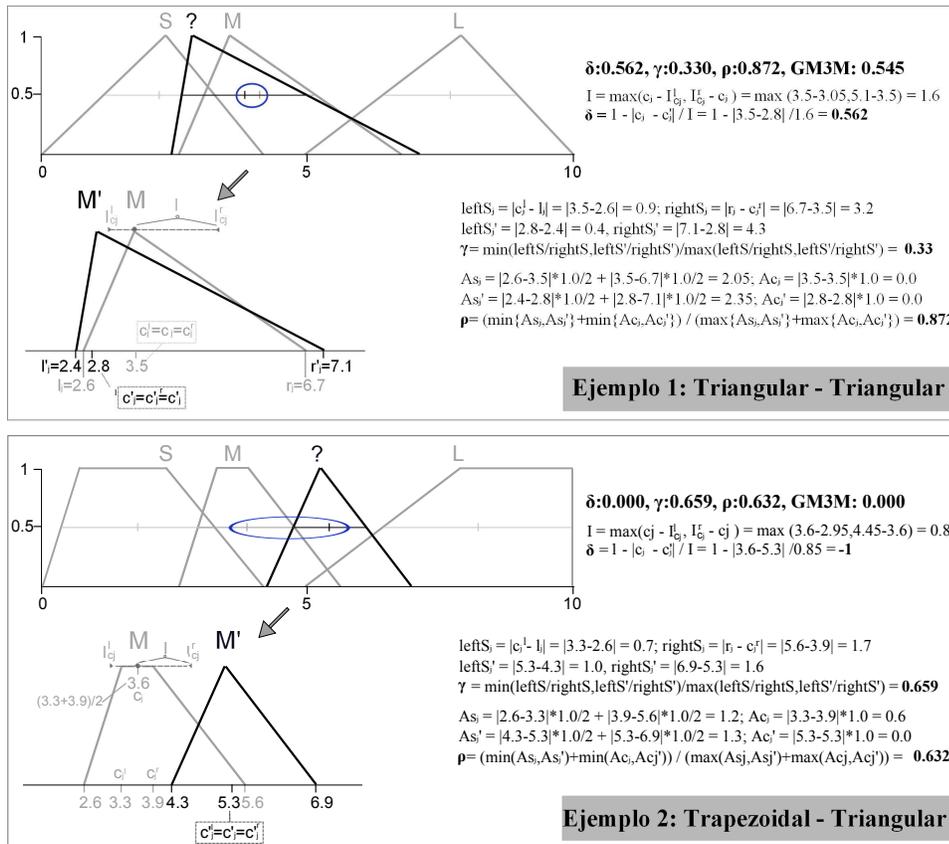


Figura 4.4: Ejemplos de cálculo de GM3M.

Para mostrar la bondad de esta propuesta, el resto del estudio se realizará utilizando funciones de pertenencia triangulares. Además, como partición difusa original se considera una partición difusa fuerte uniformemente distribuida, al ser consideradas en general como las más interpretables [29].

4.1.2. Nueva métrica de interpretabilidad semántica a nivel de base de reglas: RMI

Una vez realizada la extensión de la métrica GM3M, de forma que ya puede ser utilizada en distintos entornos de trabajo, el siguiente paso es proponer una

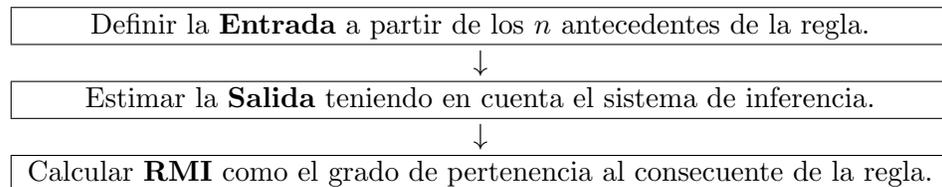
métrica de interpretabilidad semántica a nivel de BR que pueda ser utilizada junto con GM3M para comparar la interpretabilidad de distintos SBRDs.

En general la interpretabilidad semántica de la BR suele ser evaluada a través de propiedades como la consistencia o el número de reglas disparadas simultáneamente [29]. El objetivo aquí es proponer una nueva métrica de evaluación de este tipo de interpretabilidad que pueda ser utilizada con cualquier tipo (lingüística o aproximativa) de BR, teniendo en cuenta que de cara a la interpretabilidad el principal problema es la existencia de reglas contradictorias (antecedentes similares pero consecuentes diferentes) y no la interacción simultánea entre reglas que expresan lo mismo. Es decir, múltiples reglas disparadas al mismo tiempo con antecedentes y consecuentes similares no representan un problema de interpretabilidad semántico sino más bien un problema de complejidad, ya que el significado de esas reglas al mirarlas individualmente es consistente con la salida que daría el sistema para su correspondiente zona de activación.

Además, estas medidas deberían tener en cuenta los operadores utilizados en la inferencia, ya que dichos operadores afectan en el cálculo de la salida del sistema y podrían dar lugar a que lo que dice una regla no sea consistente con lo que dice el sistema al inferir en su zona de activación. Finalmente, el coste computacional para el cálculo de la medida debe ser bajo para que pueda ser integrada dentro de los procesos de modelado de los SBRDs. La nueva métrica de interpretabilidad semántica de la BR propuesta a continuación tiene en cuenta todos estos aspectos y recibe el nombre de RMI (*Rule Meaning Index*).

El cálculo de la métrica RMI se basa en medir el valor RMI individual de cada regla de la BR, al que llamaremos $RMI(R_i)$, para posteriormente calcular un valor RMI global para todo el SBRD. De forma individual el objetivo de $RMI(R_i)$ es evaluar el grado de fiabilidad de la regla R_i respecto a lo que diría el modelo completo para la zona de activación de dicha regla, teniendo en cuenta para su cálculo el sistema de inferencia concreto utilizado por el SBRD a través de la salida estimada.

De forma muy resumida, la forma de calcular $RMI(R_i)$ es como sigue:



Formalmente, si la estructura típica de una regla difusa R_i es:

$$R_i : \text{Si } X_{i1} \text{ es } A_{i1} \text{ y } \dots \text{ y } X_{in} \text{ es } A_{in} \text{ entonces } Y \text{ es } B_i \quad (4.3)$$

donde X e Y son las variables de entrada y salida respectivamente, n es el número de antecedentes y $A_{i1}, \dots, A_{in}, B_i$ son conjuntos difusos, se define O_{R_i} como la salida del SBRD cuando la entrada se define a partir de los núcleos de los conjuntos difusos de los antecedentes A_{ik} ($k = 1..n$) de la regla R_i . Dicha salida se obtiene infiriendo sobre todas las reglas presentes en la BR. Entonces el valor RMI para una regla R_i es:

$$\text{RMI}(R_i) = \mu_{\text{Cons}_{R_i}}(O_{R_i}) \quad (4.4)$$

donde $\mu_{\text{Cons}_{R_i}}(O_{R_i})$ es el grado de pertenencia de O_{R_i} al conjunto difuso B_i del consecuente de la regla R_i .

Finalmente, el valor global para un SBRD de la métrica RMI es el valor mínimo entre todos los posibles valores $\text{RMI}(R_i)$:

$$\begin{aligned} \text{RMI} &= \min_i(\text{RMI}(R_i)) \\ \forall 1 \leq i \leq \text{NumeroDeReglas} \end{aligned} \quad (4.5)$$

La métrica RMI esta definida en el intervalo $[0, 1]$, donde 0 indica el menor nivel de interpretabilidad y 1 el mayor. De esta forma:

- un valor alto de RMI significa que la regla esta bien definida ya que la salida estimada, al utilizar los antecedentes de la regla como entrada, tiene un alto grado de pertenencia al conjunto difuso del consecuente de la regla.
- un valor bajo de RMI significa que la regla no esta bien definida y que es incoherente con lo que realmente haría el sistema, ya que la salida estimada tiene un bajo nivel de pertenencia al conjunto difuso del consecuente definido en dicha regla.

Una vez descritos cada uno los pasos necesarios para calcular RMI se explican en detalle cada uno de ellos:

1 Definición de la entrada. Para calcular el valor RMI de una regla, la entrada al FBRS debe ser definida a partir de los núcleos de los conjuntos difusos definidos en los n antecedentes de la regla R_i . Si se observa la Fig. 4.3 entonces:

- Función de pertenencia triangular o gaussiana: el núcleo del conjunto difuso coincide con el punto central c .
- Función de pertenencia trapezoidal: el núcleo del conjunto difuso es un nuevo conjunto difuso de forma rectangular definido entre los puntos c^l y c^r .

La entrada así definida es a su vez un conjunto difuso de cualquier tipo: *singleton*, rectangular, etc. Al nuevo conjunto difuso definido a partir del núcleo del antecedente k de la regla i , A_{ik} , lo denotaremos por $Nucleo(A_{ik})$.

2 Estimar la salida O_{R_i} , a partir de la entrada generada en el paso anterior. Se distinguen dos posibles tipos de entradas al SBRD:

- Si la entrada es un conjunto *singleton* entonces la salida se estima de forma clásica, considerando cada conjunto *singleton* como un valor preciso.
- En el caso de tener como entrada un nuevo conjunto difuso no *singleton*, entonces el primer paso es calcular el grado de activación de cada antecedente k de cada regla de la BR como un valor preciso para luego estimar la salida como en cualquier otro SBRD.

La forma de calcular el grado de activación es a través de la altura máxima del conjunto intersección entre:

1. la entrada definida a partir del núcleo del antecedente k de la regla i , $Nucleo(A_{ik})$.
2. el conjunto difuso que define la función de pertenencia del antecedente k de la regla j , A_{jk} .

La Fig. 4.5 muestra gráficamente el grado de activación para dos funciones de pertenencia distintas, una triangular y otra trapezoidal. La definición formal asociada es:

$$\begin{aligned}
 &GradoActivacion(Nucleo(A_{ik}), A_{jk}) = \\
 &= Altura(Nucleo(A_{ik}) \wedge A_{jk}) = \\
 &= \text{máx}\{h|h = \mu_{Nucleo(A_{ik}) \wedge A_{jk}}\} = \\
 &= \text{máx}\{h|h = (\text{mín}\{\mu_{Nucleo(A_{ik})}, \mu_{A_{jk}}\})\} \\
 & i, j \in BaseDeReglas
 \end{aligned} \tag{4.6}$$

En este segundo paso es cuando la métrica RMI tiene en cuenta el sistema de inferencia concreto del SBRD, de forma que si el sistema de inferencia cambia también lo hace el valor de la métrica.

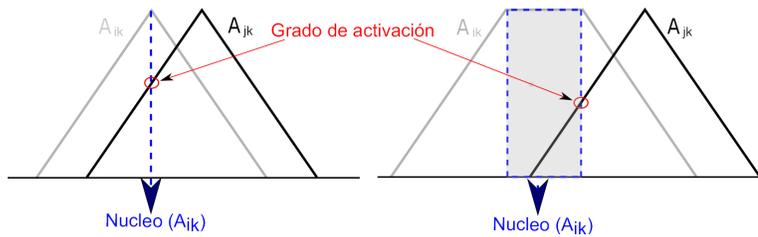


Figura 4.5: Grado de activación entre $Nucleo(A_{ik})$ y A_{jk} .

3 Calcular el valor de la métrica RMI como el grado de pertenencia de la salida estimada en el paso 2 a la función de pertenencia definida en el consecuente de la regla R_i : evaluando así como de diferentes son la salida global proporcionada por el SBRD y la salida concreta de R_i .

La Fig. 4.6 muestra un ejemplo del cálculo de RMI para un SBRD de tres reglas cuando el sistema de inferencia es min-max y la defuzzificación se realiza por centros de gravedad. Como se observa en la figura: R_1 entra en conflicto con R_2 ya que tienen antecedentes similares pero consecuentes algo diferentes, por lo que su nivel de fiabilidad es de $RMI(R_1) = 0,5$; R_3 es coherente con todas las demás reglas, por lo que su nivel de fiabilidad es de $RMI(R_3) = 1$. El valor global de la métrica para el conjunto de las tres reglas es el valor mínimo de las métricas individuales: $RMI(R_1, R_2, R_3) = 0,5$.

Como se ha podido ver, la métrica RMI se calcula teniendo en cuenta el sistema de inferencia concreto utilizado para indicar el peor caso de interacción, por lo que va a ser capaz de detectar posibles problemas derivados de la elección de un mal operador [48, 49] y/o incluso aquellos que pudieran aparecer como consecuencia del uso de pesos en las reglas [136].

Por supuesto la definición previa de la métrica RMI también puede ser generalizada (extendida) haciendo uso de los α -cortes en los núcleos de los conjuntos difusos que definen los antecedentes, de forma que se puedan tener en cuenta distintos niveles de intensidad a la hora de definir la entrada al SBRD. En estos casos el universo de discurso de la entrada se define a partir del núcleo del conjunto difuso del antecedente de la regla con una intensidad α como se muestra en la Ec. 4.7:

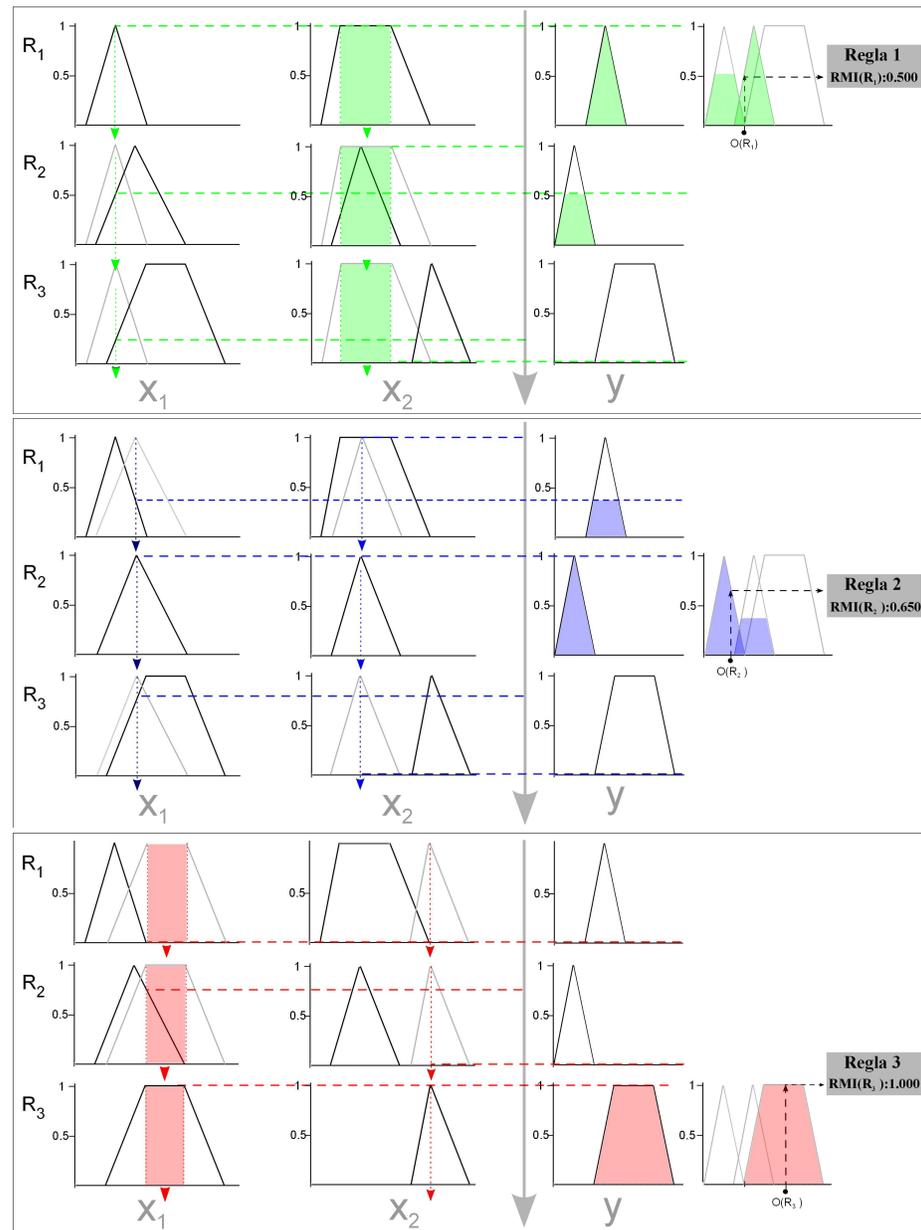


Figura 4.6: Ejemplos de cálculo de RMI.

$$\mu_{Nucleo^\alpha(A_{ik})} = \begin{cases} \mu_{A_{ik}} & \text{If } \mu_{A_{ik}} \geq \alpha \\ 0 & \text{If } \mu_{A_{ik}} < \alpha \end{cases} \quad (4.7)$$

En los sistemas normalizados el valor de α debería ser superior a 0.5, ya que no tiene mucho sentido usar valores inferiores. La Fig. 4.7 muestra algunos ejemplos cuando la entrada se define a partir del núcleo del conjunto difuso con una intensidad de 0,8.

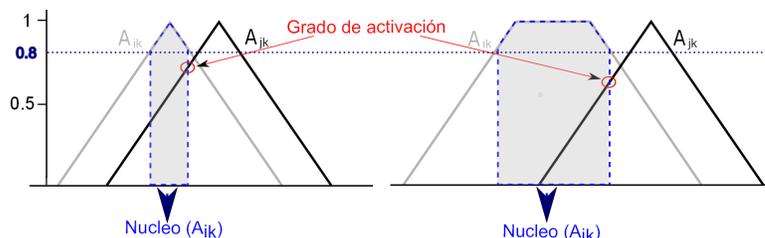


Figura 4.7: Grado de activación entre $Nucleo(A_{ik})$ y A_{jk} con $\alpha = 0,8$.

De cara a evaluar la métrica RMI en el resto del estudio se considerarán funciones de pertenencia triangulares con intensidad 1,0.

4.1.3. Evaluando localmente la interpretabilidad relativa entre sistemas basados en reglas difusas: Modelos lingüísticos y pseudo-lingüísticos

Tal y como se han definido las medidas de interpretabilidad semántica GM3M y RMI, ambas pueden ser utilizadas para comparar localmente distintos SBRDs. Por un lado GM3M, que mide la interpretabilidad semántica a nivel de BD, puede ser utilizada para cuantificar de forma individual la semántica de cada función de pertenencia. Es decir, se puede asignar a cualquier función de pertenencia, lingüística o aproximativa, un significado lingüístico definido inicialmente por un experto indicando su grado de interpretabilidad respecto a dicho término lingüístico. Si tomamos el ejemplo 1 de la Fig. 4.4 se puede decir que la semántica de la partición difusa aproximativa “M” es equivalente a la semántica de la partición difusa lingüística “M” en un grado de 0,545. Formalmente la variable de la regla difusa podría ser descrita de dos formas:

1. usando directamente la notación de la partición difusa aproximativa: x_1 es M' .
2. usando la variable de la partición difusa lingüística interpretable equivalente, cuya semántica asociada habrá sido definida previamente por un experto: x_1 es $M^{0,545}$. Igualmente usando representación gráfica se podrían representar las funciones de pertenencia junto con su grado para facilitar la interpretación local (véase la Fig. 4.8).

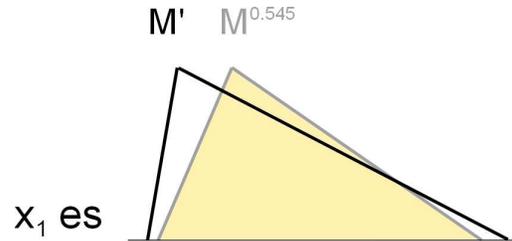


Figura 4.8: Representación gráfica de una variable.

Por otro lado, RMI, que mide la interpretabilidad semántica a nivel de BR, puede ser utilizada para cuantificar de forma individual la confianza/veracidad de una regla respecto a las demás, tal y como se ha visto en la sección anterior. Llamaremos a los modelos aproximativos representados de esta forma, modelos pseudo-lingüísticos.

Gracias a estas medidas se van a poder comparar cuantitativamente BRs de todo tipo, lo que ayudará al usuario en su toma de decisiones a la hora de seleccionar la BC que más le convenga en cada caso: lingüístico o pseudo-lingüístico.

4.2. Búsqueda del equilibrio en los sistemas basados en reglas difusas lingüísticos y aproximativos: Ajuste y selección de reglas usando algoritmos evolutivos multi-objetivo

Una vez presentadas las nuevas métricas de interpretabilidad se propone una metodología que, haciendo uso de dichas métricas, consiga de forma simultánea

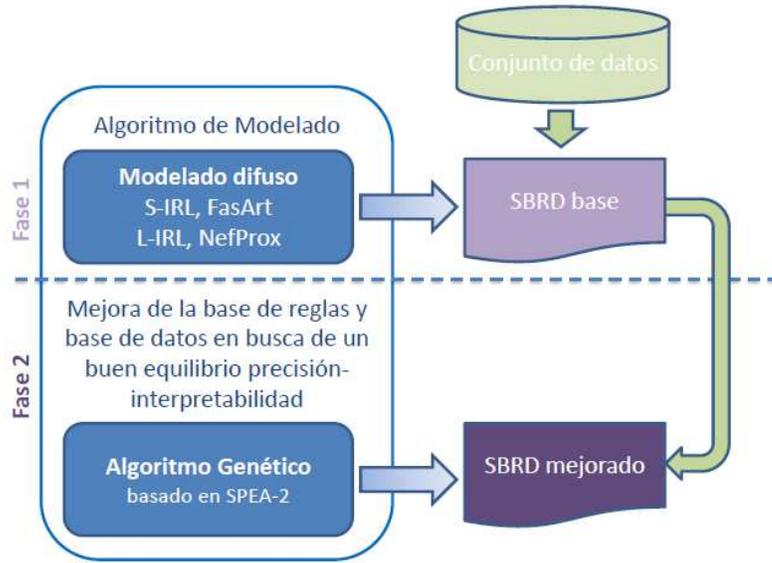


Figura 4.9: Proceso de mejora en dos fases.

mejorar la precisión del SBRD, reducir su complejidad y conservar la interpretabilidad semántica tanto de las reglas como de las funciones de pertenencia. El objetivo no es en sí tener un algoritmo o enfoque ganador sino tener una serie de enfoques de distinto tipo, que siguiendo los estándares de los algoritmos que forman parte del estado del arte, permitan el uso de las medidas propuestas para generar SBRDs con un buen equilibrio interpretabilidad-precisión y su comparación posterior tanto de manera global como local.

Para ello se propone realizar una selección de reglas difusas y un ajuste de las funciones de pertenencia mediando el uso de AEMOs. Al igual que en el capítulo anterior se sigue un proceso de mejora en dos fases (Fig. 4.9) consiste en:

1. Generar un SBRD base que contenga un conjunto de reglas difusas candidatas a optimizar.
2. Realizar un post-procesamiento de selección de reglas y ajuste de las funciones de pertenencia utilizando un AEMO avanzado diseñado específicamente para este problema con el objetivo de mejorar las prestaciones del SBRD base.

4.2.1. Sistemas basados en reglas difusas iniciales

En la primera fase se generan los SBRDs base iniciales. En este caso se hace uso de cuatro algoritmos de modelado, generando un SBRD inicial distinto con cada uno de ellos:

L-IRL (*Linguistic Iterative Rule Learning*) [119] para generar un SBRD **lingüístico** siguiendo un enfoque basado en el aprendizaje iterativo de reglas. Dicho enfoque se caracteriza porque en cada ejecución del AG se obtiene como solución sólo la mejor regla aprendida en el proceso evolutivo, por lo que el SBRD final se forma mediante la unión de las reglas obtenidas en una serie de ejecuciones sucesivas. El algoritmo maneja particiones difusas lingüísticas, como operador de inferencia usa la t-norma del mínimo y el método de defuzzificación está basado en el cálculo del centro de gravedad.

NEFPROX (*Neuro-Fuzzy Function Approximation*) [122] para generar un SBRD **lingüístico** siguiendo la estrategia explicada en la Sección 3.2.1 del capítulo anterior.

S-IRL (*Scatter Iterative Rule Learning*) [137] para generar un SBRD **aproximativo** siguiendo el enfoque basado en el aprendizaje iterativo de reglas explicado previamente. La única diferencia respecto a L-IRL es que ahora el algoritmo maneja particiones difusas aproximativas, aplicando el mismo operador de inferencia (t-norma del mínimo) y método de defuzzificación (basado en el cálculo del centro de gravedad).

FASART (*Fuzzy Adaptive System ART based*) [123] para generar un SBRD **aproximativo** siguiendo la estrategia explicada en la Sección 3.2.1 del capítulo anterior.

4.2.2. Post-procesamiento basado en algoritmos evolutivos multi-objetivo

Una vez generadas las reglas candidatas, el siguiente paso consiste en aplicar un post-procesamiento *ad-hoc* basado en AEMOs. Concretamente se han implementado dos nuevas versiones del algoritmo *SPEA2* [138], ya utilizado previamente para garantizar la interpretabilidad semántica de las funciones de pertenencia [77]. Las versiones aquí implementadas han sido diseñadas en función del

tipo de modelado sobre el cual se quiere actuar: **L – TS_{SP2}–SEM** para SBRDs con reglas lingüísticas y **S – TS_{SP2}–SEM** para SBRDs con reglas aproximativas.

Se presenta a continuación una descripción de ambas versiones. Dado que las dos versiones están basadas en componentes comunes especialmente diseñados, estas serán descritas como un único algoritmo, proporcionando una descripción individual solo para aquellas partes en las que ambas versiones son diferentes.

Funciones objetivo

Dado que el objetivo principal del proceso es conservar y/o mejorar la interpretabilidad de los SBRDs, varias son las métricas que se van a utilizar para evaluar cada SBRD. Como índice de precisión se utiliza el error cuadrático medio y como índice de complejidad el número de reglas, ya que ambos son los índices más utilizados en la literatura científica para evaluar dichos conceptos. En cuanto a la interpretabilidad semántica esta será evaluada a través de las dos nuevas métricas propuestas previamente.

Por ello cada cromosoma de la población, que representa un SBRD, va a ser evaluado en función del grado de cumplimiento en cada uno de los siguientes cuatro objetivos:

1. **Maximizar** el valor de la métrica extendida de **GM3M**: para conservar o mejorar la interpretabilidad semántica de las funciones de pertenencia.
2. **Maximizar** el valor de la métrica **RMI**: para conservar o mejorar la interpretabilidad semántica de las reglas.
3. **Minimizar** el número de reglas **NR**: para reducir la complejidad del modelo.
4. **Minimizar** el error cuadrático medio **ECM**: para reducir el error cometido por el sistema. Dicho error se define como $ECM = \frac{1}{|D|} \sum_{l=1}^{|D|} (F(x^l) - y^l)^2$, donde $|D|$ es el tamaño del conjunto de datos, $F(x^l)$ es la salida del SBRD cuando la entrada es el l -ésimo ejemplo, y y^l es la salida deseada conocida.

Además, si alguna de las soluciones a evaluar no cubre alguno de los ejemplos de entrada disponibles, entonces las cuatro funciones objetivo son penalizadas de forma que se garantiza que la solución va a ser dominada por otras en el Frente de Pareto.

Codificación

Se va a utilizar un doble esquema de codificación, uno para la *selección de reglas* (C_S) y otro para el *ajuste* (C_T): $C^p = C_S^p C_T^p$.

El esquema de codificación para la parte $C_S^p = (c_{S1}, \dots, c_{Sm})$ consiste en un cadena de código binario de tamaño m (número de reglas iniciales). En función de si una regla está o no seleccionada, el gen correspondiente toma los valores '1' o '0' respectivamente.

Para la parte C_T se utiliza un esquema de codificación real que representa los valores característicos de las funciones de pertenencia, y cuya interpretación es diferente en función de tipo de SBRD considerado:

- **Codificación de C_T para modelado lingüístico (L-TS_{SP2-SEM}).** El esquema de codificación utilizado es:

$$\begin{aligned} C_T^p &= C_1 C_2 \dots C_n; \\ C_i &= (a_1^i, b_1^i, c_1^i, \dots, a_{m^i}^i, b_{m^i}^i, c_{m^i}^i), \quad i = 1, \dots, n \end{aligned} \quad (4.8)$$

donde m^i es el número de etiquetas de cada una de las n variables que forman parte de la BD y a, b, c los puntos característicos de definición de las funciones de pertenencia triangulares.

- **Codificación de C_T para modelado aproximativo (S-TS_{SP2-SEM}).** El esquema de codificación utilizado es similar al anterior, pero teniendo en cuenta que ahora hay que codificar por separado cada función de pertenencia de cada variable de cada regla. Si el SBRD tiene m reglas iniciales con n variables ($n - 1$ variables de entrada y 1 variable de salida), entonces el esquema de codificación es:

$$\begin{aligned} C_T^p &= C_1 C_2 \dots C_m; \\ C_i &= (a_1^i, b_1^i, c_1^i, \dots, a_n^i, b_n^i, c_n^i), \quad i = 1, \dots, m \end{aligned} \quad (4.9)$$

Población inicial

La población inicial se obtiene siguiendo las siguientes reglas:

- C_S : todos los genes de todos los individuos toman valor '1'.

- C_T del primer individuo: se introduce la codificación de las funciones de pertenencia del modelo inicial (lingüístico o aproximativo).
- C_T del segundo individuo: se introduce la BD de referencia inicial utilizada para calcular el valor de la métrica GM3M.
- C_T del resto de individuos: se inicializa aleatoriamente, teniendo en cuenta los intervalos de variación definidos previamente en el Sección 4.1.1.

Indicar aquí que, como consecuencia de la naturaleza de los SBRDs (lingüística o aproximativa), pueden darse situaciones en las que los puntos característicos de una función de pertenencia estén fuera de los intervalos de variación definidos en el sección 4.1.1 (definidos a partir de las funciones de pertenencia interpretables iniciales) por lo que es necesario ampliar dichos límites de forma que incluyan dichos puntos característicos. Tal y como se explica posteriormente, dicho intervalo de variación será reajustado con cada re-inicio del algoritmo.

Por otro lado, en algunas ocasiones muy concretas, el modelo basado en reglas aproximativas inicial deja sin cubrir un número muy pequeño de datos de entrenamiento. En dichas ocasiones los modelos son reparados añadiendo reglas a la BR. La forma de generar esas nuevas reglas es ejecutando el algoritmo Wang & Mendel [135] sobre aquellos datos no cubiertos, utilizando la BD de referencia inicial. El proceso se repite hasta que todos los datos de entrenamiento son cubiertos por alguna regla. Por supuesto, estas reglas son tenidas en cuenta posteriormente a la hora de realizar la selección de reglas y el ajuste de las funciones de pertenencia del modelo como reglas aproximativas.

Cruce y Mutación

El operador de cruce inteligente y el operador de mutación que se van a utilizar en la propuesta han sido seleccionados en base a experiencias previas ([70]) obtenidas al tratar el problema concreto de selección de reglas y ajuste de funciones de pertenencia.

Al utilizar dos tipos distintos de codificaciones en el cromosoma es necesario definir operadores específicos para cada una de las partes. Se muestran a continuación los pasos a seguir para obtener cada uno de los descendientes:

- La parte C_T del descendiente se obtiene aplicando el operador de cruce BLX-0,5 [139].

- Una vez que la parte real C_T del descendiente ha sido obtenida, la parte binaria C_S se obtiene a partir de las partes C_T (funciones de pertenencia) de los padres y del descendiente. Así, para cada gen en la parte C_S que representa una regla concreta:
 1. Se obtienen las funciones de pertenencia presentes en dicha regla a partir de las funciones de pertenencia codificadas en las partes C_T de los individuos involucrados en el cruce (dos padres y un descendiente). De esta forma se pueden obtener las reglas concretas que representan cada uno de estos tres individuos.
 2. Posteriormente se calculan las distancias euclídeas normalizadas entre el descendiente y cada uno de los padres, en base a los puntos centrales de las funciones de pertenencia obtenidas para cada uno de ellos. Las diferencias entre cada par de puntos centrales se normalizan en función de la amplitud de sus respectivos intervalos de variación.
 3. El padre con la regla más cercana a la obtenida en el descendiente es el que determina si una regla es seleccionada o no para el descendiente, copiando directamente su valor en la parte binaria C_S para el gen correspondiente.

El proceso se repite hasta que se asignan valores a toda la parte C_S del descendiente. En concreto el proceso se repite cuatro veces hasta conseguir los cuatro descendientes (después de aplicar el operador de mutación solo los dos descendientes con mejor precisión son finalmente seleccionados).

Una vez que los descendientes han sido generados se aplica el operador de mutación. Dicho operador se aplica de forma independiente en cada una de las partes del cromosoma. Para la parte real C_T el operador cambia el valor de un gen seleccionado aleatoriamente, haciendo que su valor sea igual al valor que tiene dicho gen en el cromosoma que representa la BD de referencia inicial. Por otro lado en la parte C_S el operador pone directamente a '0', con probabilidad P_m , un gen seleccionado también aleatoriamente. De esta forma se modifica un gen en cada una de las partes del cromosoma.

Al aplicar estos operadores tal y como se ha descrito previamente se resuelven dos problemas. Primero, al cruzar individuos que contienen configuraciones de reglas diferentes el resultado es más productivo. Segundo, se favorece la extracción de reglas desde el momento en que la mutación se enfoca solo a eliminar reglas innecesarias.

Mecanismos especiales para manejar aspectos propios del equilibrio precisión-interpretabilidad

Las dos versiones del algoritmo aquí propuestas utilizan el mecanismo de selección utilizado en *SPEA2* [138]. Sin embargo, para mejorar sus habilidades de búsqueda, se proponen las siguientes modificaciones:

- Un **mecanismo de prevención de incesto** basado en los conceptos de CHC [140] que va a evitar una convergencia prematura en la parte C_T :

$L - TS_{SP2-SEM}$ usa el mecanismo tal y como se describe en [77]: solo se cruzan aquellos padres cuya distancia *Hamming* dividida por 4 sea mayor que un umbral. Debido a que se utiliza un esquema de codificación real (solo se considera la parte C_T), es necesario transformar cada gen a código *Gray* con un número fijo de bits por gen ($BGene$), determinado por el experto. De esta forma, el valor del umbral se inicializa como $L = \frac{\#C_T * BGene}{4}$, donde $\#C_T$ es el número de genes de la parte C_T del cromosoma. En cada generación del algoritmo, el valor del umbral decrece en una unidad, lo que permite que cada vez se puedan cruzar soluciones más cercanas.

$S - TS_{SP2-SEM}$ usa el mismo mecanismo pero con dos diferencias mínimas respecto a lo anterior: el valor del umbral se inicializa como $L = \frac{maxD}{4} + 1$, donde $maxD$ es la máxima distancia de *Hamming* en la parte C_T de la población externa, y en cada generación del algoritmo el valor del umbral decrece de forma proporcional al número de variables. De esta forma se evita que el algoritmo realice muchas iteraciones antes del primer cruce.

Además en este algoritmo, a la hora de calcular la distancia de *Hamming* de este mecanismo, así como el valor de la métrica GM3M, solo se van a tener en cuenta las funciones de pertenencia de las reglas seleccionadas en la parte C_S .

- El **operador de reinicialización**, que fuerza a vaciar la población externa, genera una nueva población inicial que contiene alguna de las mejores soluciones ya localizadas por el algoritmo. En concreto, la nueva población externa contiene una copia del individuo con mejor precisión y copias de los dos mejores individuos con mejor valor en cada uno del resto de objetivos (GM3M, RMI y NR). En total la nueva población mantiene a 7 de sus

individuos, mientras que el resto de individuos se reinician tomando en la parte C_S los mismos valores que el individuo con mejor precisión y en la parte C_T valores aleatorios.

Este proceso de reinicialización se aplica cada vez que se consigue mejorar la solución más precisa o si el 50 por ciento de los cruces son realizados en una generación (dicha proporción se define como $\%_{Required} = 0,5$). Dicho valor se actualiza cada vez que se produce un reinicialización de la siguiente manera: $\%_{Required} = (1 + \%_{Required})/2$. Además, este proceso de reinicialización es desactivado en las últimas evaluaciones del algoritmo y si no se ha aplicado nunca antes de la mitad del número total de evaluaciones.

Después de cada reinicialización, los correspondientes intervalos de variación (para el cálculo de la métrica GM3M) son recalculados tomando como base la solución más precisa.

- Durante el proceso de evolución, cuando el tamaño del conjunto de soluciones no dominadas supera el tamaño máximo de la población externa (\bar{N}), el **procedimiento de corte** de SPEA2 se modifica según lo siguiente: se marcan como soluciones no eliminables las $\frac{\bar{N}}{2}$ soluciones más precisas, ya que este objetivo es el más difícil, y la mejor y la segunda mejor en cada uno de los otros objetivos. Si para estos objetivos existen varias soluciones que tienen el mismo valor, entonces se selecciona solo aquella con mejor precisión. De esta forma, la segunda mejor solución representa un valor diferente del objetivo.
- En cada etapa del algoritmo (entre puntos de reinicialización), el número de soluciones presentes en la población externa (\bar{P}_{t+1}) que forman parte de la **población de padres** se reduce progresivamente, centrándose sólo en aquellas soluciones que son más precisas. Para hacer esto, las soluciones se ordenan de mejor a peor (siendo la precisión el criterio de ordenación) y el número de soluciones consideradas en el proceso de selección se reduce progresivamente desde el 100 % del principio al 50 % al final de cada etapa. Este proceso se lleva a cabo teniendo en cuenta el valor de L . En las últimas evaluaciones, cuando el mecanismo de reinicialización se desactiva, el mecanismo anterior de ordenación de las soluciones se modifica, dejando de ser la precisión el único criterio de ordenación para pasar a considerar el resto de objetivos. De esta forma las soluciones se ordenan según un objetivo distinto cada vez.

Tabla 4.1: Parámetros $L - TS_{SP2-SEM}$ y $S - TS_{SP2-SEM}$.

Mutación	$P_m=0,2$
Tamaño de la población	200
Tamaño de la población externa	61
Número de evaluaciones	100000
Bits por gen	$BGene = 30$

Los valores finales que toman los parámetros necesarios para ejecutar los algoritmos son los mostrados en la Tabla 4.1.

4.3. Estudio experimental

Para evaluar la bondad de las nuevas métricas y de la metodología propuesta, se van a utilizar nueve conjuntos de datos reales obtenidos del repositorio del proyecto KEEL ¹ [128]. La Tabla 4.2 muestra las principales características de dichos conjuntos de datos.

Tabla 4.2: Conjuntos de datos considerados en el estudio experimental.

Conjunto de datos	Nombre	Variables	Registros
Plastic Strength	PLA	3	1650
Quake	QUA	4	2178
Electrical Maintenance	ELE	5	1056
Abalone	ABA	9	4177
Stock prices	STP	10	950
Weather Ankara	WAN	10	1609
Weather Izmir	WIZ	10	1461
Mortgage	MOR	16	1049
Treasury	TRE	16	1049

Disponible en <http://sci2s.ugr.es/keel/datasets.php>

¹<http://www.keel.es/>

El resto de la sección se organiza como sigue:

1. La Sección 4.3.1 presenta la metodología seguida en la realización de los experimentos.
2. La Sección 4.3.2 muestra los parámetros usados para generar los SBRDs base iniciales y las características de los mismos en términos de precisión e interpretabilidad.
3. La Sección 4.3.3 analiza los resultados en la solución más precisa del Frente de Pareto.
4. La Sección 4.3.4 presenta un análisis en la solución mediana del Frente de Pareto, analizando por separado cada uno de los planos objetivo.
5. La Sección 4.3.5 analiza las soluciones más interpretables del Frente de Pareto en cada uno de los tres planos de interés.
6. La Sección 4.3.6 incluye un análisis global en el que se exponen las conclusiones más relevantes.
7. Finalmente, la Sección 4.3.7 muestra a modo de ejemplo el análisis local de algunas de las BCs obtenidas durante la experimentación.

4.3.1. Metodología experimental

El Procedimiento 2 muestra la metodología de experimentación, siguiendo las fases ya presentadas en la Sección 4.2.

Como ya se ha comentado previamente se van a utilizar diferentes algoritmos de modelado para generar, con cada uno de los conjuntos de datos considerados, varios SBRDs base iniciales. Dado que el objetivo final es evaluar el equilibrio precisión-interpretabilidad de aproximaciones tanto lingüísticas como aproximativas, se ha decidido utilizar dos algoritmos de modelado para cada una de las aproximaciones: L-IRL, NEFPPOX, S-IRL y FASART. De esta forma se generan cuatro conjuntos iniciales de reglas candidatas a optimizar, sobre los cuales validar la metodología propuesta.

Una vez que los SBRDs base iniciales han sido generadas se aplican los distintos algoritmos de post-procesamiento. Siguiendo las especificaciones mostradas en la Sección 4.2 se han desarrollado dos versiones del AEMO: $L - TS_{SP2-SEM}$

para el modelado basado en reglas lingüísticas, y $S - TS_{SP2-SEM}$ para modelado basado en reglas aproximativas. La aplicación de estas dos versiones del AEMO sobre los cuatro algoritmos de modelado da lugar a las **cuatro posibles combinaciones** mostradas en la Tabla 4.3: **Ling1**, **Ling2**, **Scat1** y **Scat2**.

Procedimiento 2 Metodología experimental.

```

for Algoritmo de Modelado=Lingüísticos:Aproximativos do
  for ConjuntoDeDatos=1 to 9 do
    for ValidacionCruzada=1 to 5 do
      Generar SBRD base inicial
      for Ejecucion=1 to 6 do
        Ejecutar Algoritmo Genético SPEA2 modificado para selección de
        reglas y ajuste de funciones de pertenencia con cuatro objetivos:
          Precisión  $\rightarrow$   $\text{máx}(ECM_{tra})$ 
          Complejidad  $\rightarrow$   $\text{máx}(NR)$ 
          Interpretabilidad semántica BD  $\rightarrow$   $\text{mín}(GM3M)$ 
          Interpretabilidad semántica BR  $\rightarrow$   $\text{mín}(RMI)$ 
      end for
    end for
  for Plano=Precisión- $NR$ :Precisión- $GM3M$ :Precisión- $RMI$  do
    Análisis Frentes de Pareto {SBRDs mejorados finales}
  end for
end for
for Plano=Precisión- $NR$ :Precisión- $GM3M$ :Precisión- $RMI$  do
  Test Estadísticos No Paramétricos
end for
end for

```

Para realizar los experimentos se utiliza un modelo de **validación cruzada 5-fold**, es decir, el conjunto de datos se divide aleatoriamente en 5 subconjuntos, cada uno de ellos conteniendo un 20% de los registros totales, de los cuales 4 subconjuntos son utilizados en el entrenamiento y uno en el test ². El algoritmo

²Las correspondientes particiones (5-fold) de los conjuntos de datos están disponibles en la página web del proyecto KEEL [128]: <http://sci2s.ugr.es/keel/datasets.php>

Tabla 4.3: Métodos considerados para el análisis.

Método	Ref.	Descripción	Objetivos
<i>Métodos para generar los modelos iniciales</i>			
L-IRL	[119]	Linguistic Iterative Rule Learning (Genera SBRD lingüístico inicial)	—
NEFPROX	[122]	Neuro-Fuzzy Function Approximation (Genera SBRD lingüístico inicial)	—
S-IRL	[137]	Scatter hard constrained Iterative Rule Learning (Genera SBRD aproximativo inicial)	—
FASART	[123]	Neuro fuzzy system based on ART (Genera SBRD aproximativo inicial)	—
<i>Algoritmo Evolutivo Multi-Objetivo para el Post-procesamiento</i>			
L-TS _{SP2-SEM}	Nuevo	Ajuste lingüístico y selección de reglas con SPEA2 semántico	<i>ECM</i> / <i>NR</i> / GM3M / RMI
S-TS _{SP2-SEM}	Nuevo	Ajuste aproximativo y selección de reglas con SPEA2 semántico	<i>ECM</i> / <i>NR</i> / GM3M / RMI
<i>Combinaciones analizadas (Dos Lingüísticas y Dos Aproximativas)</i>			
LING1	—	L-IRL + L-TS _{SP2-SEM}	<i>ECM</i> / <i>NR</i> / GM3M / RMI
LING2	—	NEFPROX + L-TS _{SP2-SEM}	<i>ECM</i> / <i>NR</i> / GM3M / RMI
SCAT1	—	S-IRL + S-TS _{SP2-SEM}	<i>ECM</i> / <i>NR</i> / GM3M / RMI
SCAT2	—	FASART + S-TS _{SP2-SEM}	<i>ECM</i> / <i>NR</i> / GM3M / RMI

de post-precesamiento se aplica 6 veces sobre cada una de las 5 particiones disponibles (entrenamiento/test), considerando una semilla aleatoria diferente cada vez. Finalmente, se analizan los resultados medios de las 30 ejecuciones para los siguientes índices: error de entrenamiento y test ($ECM_{tra/tst}$), número de reglas NR , valor de GM3M y/o valor de RMI.

Dado que todos los algoritmos utilizados en la experimentación utilizan un enfoque multiobjetivo con cuatro objetivos, las **soluciones** que se obtengan van a ser **proyectadas en tres planos**:

1. Precisión-Complejidad (NR).
2. Precisión-Semántica de la BD (GM3M).
3. Precisión-Semántica de la BR (RMI).

Todas las soluciones obtenidas serán proyectadas en cada uno de los planos sucesivamente, eliminando en cada caso las soluciones que pasen a ser dominadas

por otras. De esta forma se podrán analizar mejor las relaciones existentes entre cada uno de los objetivos de interpretabilidad y el objetivo de precisión. Este tipo de proyecciones ya ha sido utilizado previamente por otros investigadores para realizar representaciones gráficas y análisis estadísticos cuando el objetivo de la propuesta era optimizar tres objetivos simultáneamente [77, 106].

Además, para facilitar tanto el análisis como la aplicación posterior de los test estadísticos, para cada ejecución y conjunto de datos se van a considerar sólo **tres puntos representativos** del Frente de Pareto **en cada uno de los planos**:

1. SBRD más preciso en entrenamiento.
2. SBRD situado en la mediana precisión-interpretabilidad.
3. SBRD más interpretable.

Dado que estas tres soluciones (más precisa, mediana y más interpretable) se consideran características en cada uno de los planos, serán las que se utilizarán para realizar el **análisis estadístico**. No obstante el usuario final podrá seleccionar la solución más apropiada para cada caso, de entre todas las presentes en el Frente de Pareto final, en función de las propiedades para las cuales se desea obtener el equilibrio: entre precisión y número de reglas, GM3M o RMI.

Finalmente, con el objetivo de determinar si existen o no diferencias significativas en los resultados obtenidos, se va a realizar un análisis estadístico [131–134] basado en la ejecución de **test no paramétricos**, siguiendo las recomendaciones realizadas en [133] y [134], donde se introducen un conjunto simple, seguro y robusto de test no paramétricos utilizados para comparar estadísticamente múltiples clasificadores. En concreto se van a utilizar los siguientes test no paramétricos de comparación múltiple: test de Friedman [141], test de Iman-Davenport [142] y el test de Holm [143]. La página web <http://sci2s.ugr.es/sicidm/> contiene una descripción detallada de dichos test. Para realizar los test se va a utilizar un nivel de confianza de $\alpha = 0,1$.

4.3.2. Características de los sistemas basados en reglas difusas iniciales

Como ya se ha comentado previamente se van a utilizar cuatro algoritmos de modelado diferentes cada uno de ellos con sus propios parámetros de configuración y operadores del sistema de inferencia:

L-IRL y S-IRL El operador de defuzzificación sigue una estrategia basada en calcular el *Centro de gravedad ponderado por el grado de emparejamiento*, mientras que como operadores de implicaciones y conjunción se utiliza la *t-norma del mínimo*.

Los parámetros de configuración del sistema son: nLT (número de términos lingüísticos para las particiones lingüísticas iniciales), ϵ (grado mínimo de cobertura), ω (cobertura para ejemplos positivos), K (porcentaje de ejemplos negativos), P (tamaño de la población), Gen (número de generaciones), a y b (cruce y mutación), P_c (probabilidad de cruce) y P_m (probabilidad de mutación). Además, en el caso de S-IRL la estrategia evolutiva (ES) se aplica hasta que, tras 50 generaciones, dejan de producirse mejoras en un $\alpha = 20\%$ de los individuos de la población.

NEFPROX Inferencia *max-min* y defuzzificación por *media de máximos*. El único parámetro del modelado es: nLT (número de términos lingüísticos para las particiones lingüísticas iniciales).

FASART Fuzzificación por *punto simple*, inferencia por *producto*, y defuzzificación por *media de los centros de los conjuntos difusos*. Los parámetros de modelado son: ρ (parámetro de vigilancia) y γ (tasa de fuzzificación).

Tabla 4.4: Parámetros de generación de los SBRDs base iniciales.

#var	L-IRL y S-IRL	NEFPROX	FASART
< 9	$nLT = 5$	$nLT = 5$	$\rho = 0,7, \gamma = 8$
≥ 9	$nLT = 3$	$nLT = 3$	$\rho = 0,7, \gamma = 6$
en cualquier caso	$\epsilon = 1,5, \omega = 0,05$ $K = 0,1, P = 61$ $Gen = 100$ $a = 0,35, b = 5$ $P_c = 0,6, P_m = 0,1$ $ES = 50, \alpha = 30\%$		

La Tabla 4.4 muestra los parámetros concretos utilizados para generar los SBRDs iniciales, que van a tomar diferentes valores en función del número de variables que contenga el conjunto de datos. En concreto se van a distinguir dos situaciones: conjuntos de datos con menos de 9 variables, y el resto. Al establecer

este criterio se consigue que las BCs obtenidas inicialmente (y que serán posteriormente optimizadas) contengan un número razonable de reglas para aquellos conjuntos de datos más complejos.

La Tabla 4.5 muestra los resultados medios obtenidos con los modelos iniciales generados para cada uno de los cuatro casos a analizar: LING1, LING2, SCAT1 y SCAT2. En concreto se muestran los valores de los siguientes índices: error cuadrático medio en entrenamiento y test ($ECM_{tra/tst}$), número de reglas (NR), interpretabilidad semántica de las funciones de pertenencia (GM3M) e interpretabilidad semántica de las reglas (RMI).

4.3.3. Resultados y análisis de los modelos más precisos

La Tabla 4.6 muestra los resultados obtenidos con los SBRDs más precisos del frente de Pareto para cada uno de los distintos métodos considerados en el estudio. En este caso el punto a analizar es el mismo independientemente del plano objetivo considerado, por lo que el análisis realizado es único para los tres planos de estudio propuestos en la metodología.

La tabla agrupa los resultados por columnas en función del método considerado, mostrando en cada caso los valores medios obtenidos por cada uno de ellos para cada uno de los conjuntos de datos utilizados. Concretamente se muestran los valores medios de los siguientes índices: error cuadrático medio en entrenamiento y test ($ECM_{tra/tst}$), número de reglas (NR), valor GM3M y valor RMI. La penúltima fila muestra el valor medio para todos los conjuntos de datos y la última fila muestra en número de veces que se consigue que el SBRD final mejore al inicial.

Según los resultados medios obtenidos se puede decir que:

- En todos los casos tanto el error como el número de reglas de los SBRDs optimizados se reducen de forma significativa respecto a los valores de los índices en los SBRDs iniciales. También se mejora la interpretabilidad semántica de las reglas y de las funciones de pertenencia en los SBRDs aproximativos, aunque no se puede decir lo mismo en los SBRDs lingüísticos, en los cuales la mejora en la precisión se obtiene a costa de empeorar la semántica del SBRD inicial.

Tabla 4.5: Características iniciales para LING1, LING2, SCAT1 y SCAT2.

Conjunto de datos	Ling1				Ling2				Scat1				Scat2			
	$ECM_{tra/test}$	NR	GM3M	RMI	$ECM_{tra/test}$	NR	GM3M	RMI	$ECM_{tra/test}$	NR	GM3M	RMI	$ECM_{tra/test}$	NR	GM3M	RMI
PLA	5.25/5.26	75.4	1.00	0.00	3.40/3.38	14.8	0.77	0.83	4.05/4.16	91.4	0.00	0.00	3.79/3.82	46.8	0.46	0.00
QUA	0.06/0.06	227.6	1.00	0.00	0.04/0.04	53.6	0.73	0.96	259/260	71.4	0.00	0.00	0.05/0.05	107.4	0.42	0.00
ELE	129401/133565	88.8	1.00	0.25	407398/410115	65.0	0.97	0.99	239184/242827	38.6	0.06	0.46	109178/153749	81.8	0.39	0.64
ABA	24.79/24.72	50.2	1.00	0.25	18.63/18.59	72.0	0.77	0.86	20.66/20.70	21.5	0.02	0.62	8.13/8.59	45.6	0.41	0.11
STP	16.75/16.91	45.6	1.00	0.70	10.47/10.62	123.2	0.64	0.86	20.20/20.15	16.1	0.02	0.63	2.07/2.19	36.2	0.39	0.87
WIZ	38.41/39.48	52.4	1.00	0.60	14.43/15.13	105.4	0.77	0.90	87.03/88.25	18.7	0.00	0.58	7.02/9.97	83.4	0.38	0.72
WAN	52.74/53.53	45.6	1.00	1.00	26.17/26.96	157.4	0.79	0.78	88.31/90.32	19.8	0.00	0.56	8.96/11.65	93.6	0.35	0.75
MOR	2.00/2.00	31.4	1.00	1.00	1.99/2.01	78.2	0.90	0.97	25.17/24.84	15.5	0.00	0.58	0.45/0.50	22.6	0.41	0.83
TRE	2.67/2.68	33.0	1.00	0.53	3.68/3.72	74.4	0.54	0.82	20.53/21.38	15.9	0.01	0.64	0.82/0.86	25.0	0.40	0.82

Tabla 4.6: SBRDs más precisos: Características para cada método.

Planos Precisión/NR, Precisión/GM3M y Precisión/RMI																
Conjunto de datos	Ling1			Ling2			Scat1			Scat2						
	ECM_{tra}/tst	NR	GM3M RMI	ECM_{tra}/tst	NR	GM3M RMI	ECM_{tra}/tst	NR	GM3M RMI	ECM_{tra}/tst	NR	GM3M RMI				
PLA	2.24/2.38	32.0	0.71	0.04	2.39/2.52	13.3	0.47	0.09	2.23/2.45	27.7	0.39	0.24	2.11/2.27	21.6	0.52	0.35
QUA	0.03/0.04	96.5	0.47	0.00	0.03/0.04	26.2	0.41	0.77	0.04/0.04	21.7	0.43	0.34	0.03/0.04	65.0	0.45	0.00
ELE	28247/37644	32.5	0.54	0.54	23140/30249	24.4	0.56	0.73	16079/23613	27.3	0.47	0.76	15907/19419	58.6	0.41	0.87
ABA	5.42/5.56	13.0	0.47	0.66	5.20/5.46	17.5	0.46	0.89	4.81/5.07	13.6	0.48	0.86	4.49/4.83	25.9	0.44	0.93
STP	2.98/3.28	14.1	0.47	0.76	1.44/1.94	25.8	0.48	0.88	1.07/1.46	12.6	0.51	0.91	0.63/0.87	26.7	0.39	0.99
WIZ	2.51/3.05	13.0	0.61	0.93	2.48/3.26	29.8	0.53	0.83	1.81/3.13	12.0	0.51	0.93	1.07/2.02	52.9	0.34	0.96
WAN	3.86/5.65	9.5	0.57	0.91	4.27/5.29	29.2	0.53	0.89	2.47/4.68	12.3	0.53	0.92	1.38/2.54	62.5	0.33	0.95
MOR	0.07/0.09	9.0	0.60	0.97	0.03/0.05	15.5	0.61	0.93	0.06/0.08	10.3	0.52	0.94	0.05/0.06	16.8	0.43	1.00
TRE	0.10/0.11	9.0	0.63	0.98	0.06/0.08	15.4	0.64	0.99	0.08/0.10	10.5	0.55	0.95	0.07/0.09	17.4	0.43	1.00
Media	-	25.40	0.56	0.64	-	21.89	0.52	0.78	-	16.45	0.49	0.76	-	38.61	0.41	0.78
Ganador	9/9	9	0	7	9/9	9	1	4	9/9	9	9	9	9/9	9	6	9

Por otro lado, la Tabla 4.7 muestra las puntuaciones obtenidas con el **test de Friedman** para cada uno de los métodos considerados en las cuatro medidas de interés (MST_{tst} , NR, GM3M y RMI), cuando se analizan los resultados obtenidos con los SBRDs más precisos. En este test los métodos aproximativos son los que obtienen mejor puntuación, excepto en el caso de la métrica GM3M para la cual el método mejor situado es LING1.

Tabla 4.7: SBRDs más precisos: Test de Friedman en cada una de las medidas.

Planos Precisión/NR, Precisión/GM3M y Precisión/RMI				
Método	ECM _{tst}	NR	GM3M	RMI
LING1	3.3333	2.1111	1.6667	2.4444
LING2	2.5556	2.4444	2.1111	2.1111
SCAT1	2.5556	1.7778	2.6667	3.6667
SCAT2	1.5556	3.6667	3.5556	1.7778

Por su parte, los *p-valores* obtenidos en el **test de Iman-Davenport** [(0.0246), (5.081E-3), (6.132E-3) y (5.081E-3) para ECM_{tst} , NR, GM3M y RMI respectivamente] indican que existen diferencias significativas en los resultados obtenidos con los conjuntos de datos considerados.

Por último la Tabla 4.8 muestra los resultados del **test de Holm** en las mismas cuatro medidas de interés. Según los resultados de este test:

- Se acepta la hipótesis de equivalencia entre los tres primeros métodos para todas las medidas. En términos de complejidad e interpretabilidad semántica (NR, GM3M y RMI) uno de los métodos aproximativos (SCAT1, SCAT1 y SCAT2 respectivamente) y los dos métodos lingüísticos pueden ser considerados equivalentes. En términos de precisión (ECM_{tst}) los dos métodos aproximativos y uno de los métodos lingüísticos (LING2) son también equivalentes.

De forma general para estos SBRDs se observa que:

- Con SCAT2 se obtienen los SBRDs más precisos con una interpretabilidad semántica alta a nivel de regla.
- Según los test estadísticos LING2 es equivalente a SCAT2 en términos de

Tabla 4.8: SBRDs más precisos: Test de Holm en cada una de las medidas con $\alpha = 0,1$.

Planos Precisión/NR, Precisión/GM3M y Precisión/RMI											
Holm en ECM _{test}			Holm en NR			Holm en GM3M			Holm en RMI		
i	Met.	z	p	α/i	Hip.	i	Met.	z	p	α/i	Hip.
3	LING1	2.92	0.003	0.03	Rech.	3	SCAT2	3.10	0.002	0.03	Rech.
2	SCAT1	1.64	0.100	0.05	Acep.	2	LING2	1.10	0.273	0.05	Acep.
1	LING2	1.64	0.100	0.10	Acep.	1	LING1	0.55	0.584	0.10	Acep.
						3	SCAT1	3.10	0.002	0.03	Rech.
						2	SCAT1	1.64	0.100	0.05	Acep.
						1	LING2	0.73	0.465	0.10	Acep.
						2	LING1	1.10	0.273	0.05	Acep.
						1	LING2	0.55	0.584	0.10	Acep.

error y RMI, aunque LING2 obtiene mejor valores de interpretabilidad en términos de *NR* y GM3M.

- Además LING2 es equivalente a SCAT1 en términos de *NR* y GM3M, aunque SCAT1 introduce sobre-aprendizaje (error de test superior al error de entrenamiento) en algunos conjuntos de datos (WIZ y WAN).

Por lo tanto, según estos resultados, no hay una regla general que permita definir qué método, lingüístico o aproximativo, es más aconsejable utilizar para alcanzar SBRDs con la mejor precisión posible. De hecho en algunos casos el algoritmo de modelado basado en reglas lingüísticas obtiene niveles de precisión similares a los alcanzados por los algoritmos basados en reglas aproximativas.

4.3.4. Resultados y análisis de los modelos situados en la mediana de distintos planos objetivo

Esta sección analiza, para todos los métodos propuestos, los resultados obtenidos para los SBRDs situados en la mediana del frente de Pareto en los tres planos considerados. Las Tablas 4.9 y 4.10 muestran los valores medios de los distintos índices considerados en cada plano, siguiendo el esquema utilizado en la Tabla 4.6 de la sección previa (para una descripción más detallada ver la sección 4.3.3).

Según los resultados medios obtenidos se puede decir que:

- Las mejoras alcanzadas en cada uno de los planos analizados son, al comparar los SBRDs optimizados con los iniciales, similares a las obtenidas con los SBRDs más precisos: tanto el error de entrenamiento y test como el número de reglas se reducen en todos los casos, excepto para algunos conjuntos de datos en SCAT2.
- En general, en este caso obtienen menor número de reglas, así como una mejor interpretabilidad semántica, tanto de reglas como de funciones de pertenencia, en todos y cada uno de los planos. La interpretabilidad semántica de las reglas y las funciones de pertenencia también se mejora en los métodos aproximativos, aunque no se puede decir lo mismo para los métodos lingüísticos.

Tabla 4.9: SBRDs en la mediana: Características para cada método (1/2).

Plano Precisión/NR												
Conjunto de datos	Ling1			Ling2			Scat1			Scat2		
	$ECM_{traq/tst}$	NR	Gm3M RMI	$ECM_{traq/tst}$	NR	Gm3M RMI	$ECM_{traq/tst}$	NR	Gm3M RMI	$ECM_{traq/tst}$	NR	Gm3M RMI
PLA	2.46/2.56	21.5	0.64	0.21	2.82/2.95	10.1	0.53	0.57	2.70/2.89	21.2	0.38	0.36
QUA	0.03/0.04	87.9	0.51	0.00	0.03/0.04	22.8	0.44	0.69	0.04/0.04	20.3	0.46	0.36
ELE	51770/62262	24.5	0.54	0.56	40586/49636	20.3	0.49	0.73	58102/69947	22.5	0.49	0.69
ABA	5.87/5.99	9.5	0.49	0.67	5.69/5.88	11.5	0.46	0.93	6.13/6.31	9.5	0.50	0.81
STP	3.85/4.07	10.9	0.45	0.81	2.33/2.80	19.6	0.51	0.86	2.83/3.26	9.1	0.52	0.90
WIZ	3.54/3.90	8.9	0.59	0.90	3.35/4.01	21.7	0.46	0.87	3.73/5.16	8.8	0.52	0.92
WAN	4.97/6.54	7.0	0.57	0.89	5.13/6.07	23.8	0.47	0.89	6.14/7.82	8.9	0.51	0.90
MOR	0.22/0.24	6.4	0.55	0.95	0.10/0.13	10.5	0.60	0.93	0.23/0.26	7.5	0.53	0.93
TRE	0.23/0.26	6.0	0.56	0.97	0.11/0.13	10.9	0.61	0.97	0.29/0.30	7.5	0.52	0.92
Media	-	20.27	0.55	0.66	-	16.79	0.51	0.83	-	12.80	0.49	0.75
Ganador	9/9	9	0	7	9/9	9	1	3	9/9	9	9	9
Plano Precisión/GM3M												
Conjunto de datos	Ling1			Ling2			Scat1			Scat2		
	$ECM_{traq/tst}$	NR	Gm3M RMI	$ECM_{traq/tst}$	NR	Gm3M RMI	$ECM_{traq/tst}$	NR	Gm3M RMI	$ECM_{traq/tst}$	NR	Gm3M RMI
PLA	2.26/2.39	30.8	0.75	0.05	2.46/2.58	13.0	0.53	0.30	2.40/2.61	26.9	0.44	0.36
QUA	0.03/0.04	95.5	0.57	0.00	0.04/0.04	25.9	0.57	0.71	0.00/0.04	23.1	0.53	0.29
ELE	29165/38261	32.3	0.61	0.54	36944/45671	25.5	0.63	0.71	29680/40013	27.0	0.55	0.75
ABA	5.57/5.71	12.6	0.57	0.68	5.55/5.76	16.5	0.53	0.91	6.21/6.38	11.7	0.58	0.85
STP	3.40/3.65	13.8	0.57	0.78	1.65/2.10	26.7	0.52	0.86	3.26/3.63	11.1	0.59	0.87
WIZ	2.79/3.24	13.0	0.68	0.94	2.83/3.57	32.0	0.60	0.83	4.59/6.05	10.6	0.59	0.91
WAN	4.41/5.86	9.4	0.65	0.92	5.20/6.20	33.2	0.60	0.90	4.52/6.30	11.8	0.59	0.91
MOR	0.07/0.09	9.0	0.66	0.95	0.06/0.07	16.8	0.67	0.95	0.20/0.22	9.4	0.60	0.93
TRE	0.11/0.12	9.0	0.69	0.98	0.08/0.11	17.4	0.68	0.97	0.31/0.35	9.6	0.61	0.92
Media	-	25.06	0.64	0.65	-	23.00	0.59	0.79	-	15.68	0.56	0.75
Ganador	9/9	9	0	7	9/9	9	1	4	9/9	9	9	9

Tabla 4.10: SBRDs en la mediana: Características para cada método (2/2).

		Plano Precisión/RMI														
Conjunto de datos	$ECM_{traq/1st}$	Ling1			Ling2			Scat1			Scat2					
		NR	GM3M	RMI	NR	GM3M	RMI	NR	GM3M	RMI	$ECM_{traq/1st}$	NR	GM3M	RMI		
PLA	2.36/2.49	28.5	0.65	0.27	2.40/2.52	13.3	0.48	0.21	2.31/2.51	27.3	0.39	0.39	2.45/2.55	18.3	0.52	0.53
QUA	0.03/0.04	96.5	0.47	0.00	0.03/0.04	26.1	0.40	0.86	0.04/0.04	22.6	0.42	0.49	0.03/0.04	61.4	0.47	0.02
ELE	30563/39637	32.0	0.53	0.58	28442/36966	23.7	0.52	0.82	28150/34894	26.8	0.49	0.82	20759/25373	57.6	0.42	0.89
ABA	5.44/5.58	12.9	0.47	0.72	5.20/5.46	17.5	0.45	0.91	5.20/5.41	13.5	0.49	0.92	4.50/4.84	25.9	0.44	0.94
STP	3.16/3.44	13.8	0.60	0.82	1.45/1.95	25.8	0.47	0.92	1.08/1.48	12.6	0.52	0.95	0.74/0.97	26.2	0.38	0.99
WIZ	2.54/3.11	13.0	0.60	0.96	2.49/3.24	29.7	0.51	0.87	2.16/3.60	11.9	0.51	0.96	2.06/2.98	48.3	0.35	0.97
WAN	3.96/5.50	9.6	0.56	0.95	4.28/5.31	29.2	0.51	0.92	2.60/4.71	12.2	0.53	0.96	9.55/10.59	40.2	0.43	0.97
MOR	0.07/0.09	9.0	0.59	0.99	0.03/0.05	15.5	0.60	0.95	0.06/0.08	10.2	0.52	0.96	0.05/0.06	16.7	0.43	1.00
TRE	0.10/0.11	9.0	0.63	0.99	0.06/0.09	15.4	0.64	0.99	0.14/0.16	10.3	0.56	0.98	0.07/0.09	17.4	0.41	1.00
Media	-	24.93	0.55	0.70	-	21.81	0.51	0.83	-	16.38	0.49	0.82	-	34.67	0.43	0.81
Ganador	9/9	9	0	7	9/9	9	1	4	9/9	9	9	9	8/9	9	7	9

La Tabla 4.11 muestra las puntuaciones obtenidas con el **test de Friedman** de los distintos métodos considerados en las cuatro medidas de interés (MST_{tst} , NR , $GM3M$ and RMI), cuando se analizan los resultados de los SBRDs situados en la mediana. Los *p-valores* del **test de Iman-Davenport** en cada uno de los planos indican que:

- Plano Precisión/ NR \rightarrow *p-valores* 1.081E-4, 0.0213, 0.0126 y 0.1175 en las medidas ECM_{tst} , NR , $GM3M$ y RMI respectivamente. Existen diferencias significativas en los resultados, en todos los casos menos para RMI .
- Plano Precisión/ $GM3M$ \rightarrow *p-valores* 0.0712, 0.0394, 8.214E-5 y 0.6321 en las medidas ECM_{tst} , NR , $GM3M$ y RMI respectivamente. Existen diferencias significativas en los resultados, excepto para RMI .
- Plano Precisión/ RMI \rightarrow *p-valores* 0.4551, 5.081E-3, 5.081E-3 y 6.249E-4 en las medidas ECM_{tst} , NR , $GM3M$ y RMI respectivamente. Existen diferencias significativas en los resultados, excepto para ECM_{tst} .

Los resultados obtenidos muestran que en algunos planos las diferencias entre los métodos no son estadísticamente significativas (ECM_{tst} en el plano RMI y RMI en los planos NR y $GM3M$), aunque para el resto de medidas y planos (donde las diferencias entre métodos si son estadísticamente significativas) todos los métodos alcanzan, para alguna de las medidas, la mejor puntuación del test de Friedman: los métodos aproximativos para las medidas NR ($SCAT1$) y RMI ($SCAT2$) y los métodos lingüísticos para las medidas ECM_{tst} ($LING1$ en el plano $GM3M$, $LING2$ en el plano NR) y $GM3M$ ($LING1$).

La Tabla 4.12 muestra los resultados del **test de Holm**. Analizando los resultados en función de cada una de las medidas de interés se puede concluir que:

- En la medida ECM_{tst} de precisión el test rechaza la hipótesis de igualdad entre métodos sólo para el plano NR , donde los métodos lingüísticos superan a los aproximativos. Para el resto de los planos todos los métodos deben ser considerados equivalentes.
- En cuanto a la medida RMI de interpretabilidad semántica de las reglas, la hipótesis de igualdad sólo es rechazada para el plano RMI , por lo que se puede decir que el método con mayor puntuación ($SCAT2$) supera a todos

Tabla 4.11: SBRDs en la mediana: Test de Friedman en cada una de las medidas.

Plano Precisión/NR				
Método	ECM _{tst}	NR	GM3M	RMI
LING1	2.0000	2.1667	1.5556	3.2778
LING2	1.4444	2.6667	2.5556	1.8889
SCAT1	2.8889	1.7222	2.4444	2.2222
SCAT2	3.6667	3.4444	3.4444	2.6111
Plano Precisión/GM3M				
Método	ECM _{tst}	NR	GM3M	RMI
LING1	1.8889	2.2222	1.4444	2.8333
LING2	2.0000	2.8889	2.1111	2.6667
SCAT1	3.1111	1.6667	2.6667	2.4444
SCAT2	3.0000	3.2222	3.7778	2.0556
Plano Precisión/RMI				
Método	ECM _{tst}	NR	GM3M	RMI
LING1	3.0000	2.1111	1.5556	3.2222
LING2	2.2222	2.4444	2.3333	3.1111
SCAT1	2.6667	1.7778	2.5556	2.4444
SCAT2	2.1111	3.6667	3.5556	1.2222

los demás, aunque todos ellos presentan buenos resultados en términos de RMI. Para el resto de los planos todos los métodos deben ser considerados equivalentes.

- En términos de complejidad medida a través de NR , la hipótesis de igualdad sólo es rechazada en algunos casos (SCAT2 para los planos NR y RMI; SCAT2 y LING2 para el plano GM3M), por lo que el método aproximativo con mayor puntuación (SCAT1) debe ser considerado equivalente a uno o a ambos métodos lingüísticos (LING1 y LING2 para los planos NR y RMI; LING1 para el plano GM3M).
- Por último, en cuando a la interpretabilidad semántica de las funciones de pertenencia GM3M, el método lingüístico con mayor puntuación (LING1) deber ser considerado equivalente al otro método lingüístico (LING2) para todos los planos, y equivalente a uno de los métodos aproximativos (SCAT1)

en los planos NR y RMI .

De forma general se puede decir que con cualquiera de los métodos aquí considerados se pueden conseguir SBRDs con un buen equilibrio precisión-interpretabilidad: la elección final vendrá fijada por las preferencias del usuario. Así por ejemplo LING2 obtiene buenos SBRDs en el plano Precisión/ NR , aunque para el resto de planos obtiene SBRDs con menor nivel de interpretabilidad en comparación con los SBRDs obtenidos por los métodos aproximativos.

4.3.5. Resultados y análisis de los modelos más interpretables en los distintos planos objetivo

Las Tablas 4.13 y 4.14 muestran los resultados obtenidos con los SBRDs más interpretables. En la tabla se muestran los valores medios de todos los índices en cada uno de los planos, siguiendo el esquema ya utilizado en la Tabla 4.6 de la sección previa (para una descripción más detallada ver la sección 4.3.3).

Según los resultados medios obtenidos se puede decir que:

- Los SBRDs obtenidos con los métodos LING1 y SCAT1 muestran comportamientos similares a los analizados previamente, mejorando las prestaciones de los SBRDs base iniciales de la misma forma. Sin embargo, la interpretabilidad semántica de las reglas se mejora, respecto a los SBRDs iniciales, al utilizar el método LING2, y el error aumenta con SCAT2. El grado de mejora o empeoramiento es diferente según el plano analizado.

La Tabla 4.15 muestra las puntuaciones obtenidas con el **test de Friedman** de los distintos métodos considerados en las cuatro medidas de interés (MST_{tst} , NR , $GM3M$ y RMI), cuando se analizan los resultados de las soluciones más interpretables. Por otro lado en el **test de Iman-Davenport**:

- Plano Precisión/ NR \rightarrow *p-valores* 1.431E-7, 0.0532, 0.9592 y 0.0459 en ECM_{tst} , NR , $GM3M$ y RMI respectivamente. Existen diferencias significativas en los resultados, en todos los casos menos para $GM3M$.
- Plano Precisión/ $GM3M$ \rightarrow *p-valores* 6.132E-3, 6.132E-3, 0.0394 y 4.043E-6 en ECM_{tst} , NR , $GM3M$ y RMI respectivamente. Existen diferencias significativas en todos los resultados.

Tabla 4.12: SBRDs en la mediana: Test de Holm en cada una de las medidas con $\alpha = 0,1$.

Plano Precisión/NR																							
Holm en ECM _{st}			Holm en NR			Holm en GM3M			Holm en RMI														
Met.	z	p	α/i	Hip.	Met.	z	p	α/i	Hip.	Met.	z	p	α/i	Hip.									
3	SCAT2	3.65	0.000	0.03	Rech.	3	SCAT2	2.83	0.005	0.03	Rech.	3	SCAT2	3.10	0.002	0.03	Rech.	3	LING1	2.28	0.022	0.03	Rech.
2	SCAT1	2.37	0.018	0.05	Rech.	2	LING2	1.55	0.121	0.05	Acep.	2	LING2	1.64	0.100	0.05	Acep.	2	SCAT2	1.19	0.235	0.05	Acep.
1	LING1	0.91	0.361	0.10	Acep.	1	LING1	0.73	0.465	0.10	Acep.	1	SCAT1	1.46	0.144	0.10	Acep.	1	SCAT1	0.55	0.584	0.10	Acep.
Plano Precisión/GM3M																							
Holm en ECM _{st}			Holm en NR			Holm en GM3M			Holm en RMI														
Met.	z	p	α/i	Hip.	Met.	z	p	α/i	Hip.	Met.	z	p	α/i	Hip.									
3	SCAT1	2.01	0.045	0.03	Acep.	3	SCAT2	2.56	0.011	0.03	Rech.	3	SCAT2	3.83	0.000	0.03	Rech.	3	LING1	1.28	0.201	0.03	Acep.
2	SCAT2	1.83	0.068	0.05	Acep.	2	LING2	2.01	0.045	0.05	Rech.	2	SCAT1	2.01	0.045	0.05	Rech.	2	LING2	1.00	0.315	0.05	Acep.
1	LING2	0.18	0.855	0.10	Acep.	1	LING1	0.91	0.361	0.10	Acep.	1	LING2	1.10	0.273	0.10	Acep.	1	SCAT1	0.64	0.523	0.10	Acep.
Plano Precisión/RMI																							
Holm en ECM _{st}			Holm en NR			Holm en GM3M			Holm en RMI														
Met.	z	p	α/i	Hip.	Met.	z	p	α/i	Hip.	Met.	z	p	α/i	Hip.									
3	LING1	1.46	0.144	0.03	Acep.	3	SCAT2	3.10	0.002	0.03	Rech.	3	SCAT2	3.29	0.001	0.03	Rech.	3	LING1	3.29	0.001	0.03	Rech.
2	SCAT1	0.91	0.361	0.05	Acep.	2	LING2	1.10	0.273	0.05	Acep.	2	SCAT1	1.64	0.100	0.05	Acep.	2	LING2	3.10	0.002	0.05	Rech.
1	LING2	0.18	0.855	0.10	Acep.	1	LING1	0.55	0.584	0.10	Acep.	1	LING2	1.28	0.201	0.10	Acep.	1	SCAT1	2.01	0.045	0.10	Rech.

Tabla 4.13: SBRDs más interpretables: Características para cada método.

Plano Precisión/NR												
Conjunto de datos	Ling1			Ling2			Scat1			Scat2		
	$ECM_{tra/tst}$	NR	GM3M RMI	$ECM_{tra/tst}$	NR	GM3M RMI	$ECM_{tra/tst}$	NR	GM3M RMI	$ECM_{tra/tst}$	NR	GM3M RMI
PLA	2.95/3.09	15.0	0.70 0.17	3.57/3.72	7.9	0.57 0.74	3.38/3.50	15.3	0.48 0.60	4.78/4.82	8.4	0.69 0.58
QUA	0.04/0.04	76.2	0.61 0.00	0.04/0.04	19.0	0.48 0.81	0.04/0.04	18.9	0.47 0.43	0.04/0.04	43.3	0.61 0.01
ELE	88331/98883	18.0	0.58 0.71	68855/79716	15.4	0.57 0.78	116518/131559	18.4	0.55 0.79	185888/194169	37.8	0.51 0.57
ABA	6.77/6.85	6.2	0.55 0.86	7.20/7.32	7.1	0.53 0.98	10.25/10.27	6.3	0.59 0.85	9.44/9.47	13.4	0.53 0.69
STP	5.25/5.50	7.3	0.52 0.83	4.18/4.61	13.3	0.57 0.97	9.56/9.90	6.3	0.62 0.90	5.68/5.86	12.4	0.52 0.96
WIZ	5.38/7.93	4.9	0.61 0.95	6.36/6.91	12.8	0.56 0.93	9.95/11.34	6.0	0.60 0.93	18.44/19.66	11.7	0.67 0.98
WAN	6.58/7.93	4.6	0.58 0.92	8.37/9.32	16.0	0.54 0.96	17.91/19.53	6.2	0.61 0.92	32.03/33.07	15.0	0.65 0.96
MOR	0.58/0.60	4.4	0.58 0.97	0.46/0.50	6.5	0.61 0.98	0.90/0.94	5.2	0.58 0.92	2.04/2.08	5.9	0.60 0.99
TRE	1.17/1.18	4.2	0.59 0.96	0.57/0.58	6.5	0.60 1.00	1.33/1.33	5.2	0.60 0.92	2.83/2.82	7.5	0.52 0.95
Media	-	15.64	0.59 0.71	-	11.61	0.56 0.91	-	9.77	0.57 0.81	-	17.28	0.59 0.74
Ganador	9/9	9	0 7	8/8	9	1 6	9/9	9	9 9	1/1	9	9 8

Plano Precisión/GM3M												
Conjunto de datos	Ling1			Ling2			Scat1			Scat2		
	$ECM_{tra/tst}$	NR	GM3M RMI									
PLA	2.54/2.59	33.6	0.94 0.01	3.69/3.74	14.2	0.95 0.93	3.97/4.01	35.8	0.86 0.25	3.97/4.02	14.6	0.94 0.25
QUA	0.04/0.04	120.9	0.95 0.00	0.04/0.04	34.0	0.89 0.94	0.05/0.05	32.0	0.78 0.24	0.04/0.04	57.5	0.91 0.00
ELE	97468/106125	53.8	0.90 0.40	354767/365093	45.1	0.96 0.99	169819/178258	26.0	0.87 0.83	219667/226300	49.5	0.86 0.47
ABA	10.05/10.19	20.8	0.87 0.49	15.66/15.89	29.4	0.93 0.98	19.53/19.46	11.5	0.93 0.86	7.17/7.28	25.3	0.76 0.31
STP	9.56/9.75	23.7	0.97 0.86	5.54/5.76	33.2	0.94 0.98	15.68/15.94	10.0	0.94 0.99	5.77/5.57	25.2	0.82 0.69
WIZ	10.01/10.92	23.8	0.94 0.89	9.06/10.30	52.8	0.94 0.97	40.24/42.96	10.5	0.87 0.95	18.52/19.75	17.9	0.91 0.80
WAN	22.67/23.19	16.0	0.92 0.97	15.39/16.16	63.6	0.94 0.97	39.48/40.32	9.9	0.87 0.93	35.86/36.95	24.1	0.89 0.74
MOR	0.81/0.83	12.8	0.89 0.98	0.83/0.82	26.4	0.87 0.97	2.16/2.20	8.6	0.86 0.97	2.94/2.90	11.3	0.90 0.76
TRE	0.79/0.79	11.6	0.82 0.87	1.29/1.33	20.5	0.79 0.97	3.03/3.16	8.3	0.80 0.88	3.12/3.15	15.3	0.82 0.55
Media	-	35.23	0.91 0.61	-	35.47	0.91 0.97	-	16.96	0.86 0.77	-	26.73	0.87 0.51
Ganador	9/9	9	0 7	7/8	9	7 7	9/9	9	9 9	2/2	9	9 4

Tabla 4.14: SBRDs más interpretables: Características para cada método (2/2).

		Plano Precisión/RMI														
Conjunto de datos	$ECM_{tra/1st}$	Ling1			Ling2			Scat1			Scat2					
		NR	GM3M	RMI	NR	GM3M	RMI	NR	GM3M	RMI	$ECM_{tra/1st}$	NR	GM3M	RMI		
PLA	2.63/2.72	26.6	0.72	0.55	2.71/2.82	12.6	0.59	1.00	3.82/3.92	27.1	0.73	0.90	4.93/4.98	11.7	0.81	0.92
QUA	0.03/0.04	96.5	0.47	0.00	0.04/0.04	28.3	0.48	1.00	0.05/0.05	29.0	0.66	0.65	0.03/0.04	59.3	0.46	0.04
ELE	92340/101194	29.3	0.64	0.89	274481/284773	40.7	0.82	1.00	154191/163598	26.2	0.84	0.97	68156/75396	54.0	0.52	0.92
ABA	6.74/6.80	11.4	0.55	0.98	5.23/5.49	17.7	0.45	1.00	18.63/18.55	11.5	0.82	0.99	6.20/6.56	24.7	0.43	0.95
STP	8.90/9.13	22.9	0.91	1.00	1.48/1.97	26.5	0.48	1.00	14.65/14.94	10.2	0.90	1.00	5.63/5.64	18.8	0.61	1.00
WIZ	6.08/7.06	16.9	0.76	1.00	3.73/4.70	43.5	0.57	1.00	30.48/31.83	10.3	0.77	1.00	17.99/19.32	17.3	0.79	1.00
WAN	15.40/16.44	12.7	0.78	1.00	4.66/5.69	31.7	0.50	1.00	29.14/30.68	10.7	0.77	0.99	36.03/37.64	20.2	0.80	1.00
MOR	0.32/0.33	9.2	0.68	1.00	0.04/0.05	15.6	0.59	1.00	1.47/1.55	8.7	0.73	1.00	0.82/0.90	10.4	0.46	1.00
TRE	0.31/0.32	8.1	0.64	1.00	0.06/0.09	15.5	0.64	1.00	1.77/1.90	8.9	0.69	1.00	1.97/2.03	13.7	0.42	1.00
Media	-	25.95	0.68	0.83	-	25.78	0.57	1.00	-	15.84	0.77	0.94	-	25.55	0.59	0.87
Ganador	9/9	9	0	7	9/9	9	1	9	9/9	9	9	9	3/3	9	9	9

- Plano Precisión/RMI \rightarrow p -valores 6.249E-4, 0.0394, 0.0246 y 3.450E-5 en ECM_{tst} , NR, GM3M y RMI respectivamente. Existen diferencias significativas en todos los resultados.

La mejor puntuación del test de Friedman la obtienen, en general, los métodos lingüísticos, aunque los métodos aproximativos son mejores para algunas medidas en algunos planos (NR en el plano GM3M; NR y GM3M en el plano RMI).

Tabla 4.15: SBRDs más interpretables: Test de Friedman en cada una de las medidas.

Plano Precisión/NR				
Método	ECM _{tst}	NR	GM3M	RMI
LING1	1.5556	1.7778	2.4444	3.0000
LING2	1.5556	2.8889	2.6667	1.5556
SCAT1	3.1111	2.1111	2.5556	3.0000
SCAT2	3.7778	3.2222	2.3333	2.4444
Plano Precisión/GM3M				
Método	ECM _{tst}	NR	GM3M	RMI
LING1	1.6667	2.7778	1.7778	3.0000
LING2	2.1111	3.2222	2.1111	1.2222
SCAT1	3.5556	1.3333	3.3333	2.1111
SCAT2	2.6667	2.6667	2.7778	3.6667
Plano Precisión/RMI				
Método	ECM _{tst}	NR	GM3M	RMI
LING1	2.1111	2.2222	2.3333	2.8889
LING2	1.4444	3.2222	3.2222	1.0000
SCAT1	3.5556	1.6667	1.5556	3.2222
SCAT2	2.8889	2.8889	2.8889	2.8889

La Tabla 4.16 muestra los resultados del **test de Holm**:

- Se rechaza la hipótesis de igualdad entre los métodos lingüísticos y aproximativos en términos de precisión (ECM_{tst}) para todos los planos. En estos casos los métodos lingüísticos (LING1, LING2) superan a los aproximativos, excepto en el plano GM3M donde uno de los métodos aproximativos

(SCAT2) debe ser considerado equivalente a los métodos lingüísticos, aunque a costa de la precisión.

- En términos de complejidad (NR) los resultados observados son muy diferentes: el método lingüístico con mayor puntuación en el plano NR (LING1) superó sólo a uno de los métodos aproximativos (SCAT2), el método aproximativo con mayor puntuación en el plano GM3M (SCAT1) supera al resto de métodos considerados. Y el método aproximativo con mayor puntuación en el plano RMI (SCAT1) supera a uno de los métodos aproximativos y a uno lingüístico (SCAT2, LING2).
- En términos de interpretabilidad semántica de las funciones de pertenencia se observa el mismo comportamiento: todos los métodos son equivalente para el plano NR , el método lingüístico con mayor puntuación en el plano GM3M (LING1) sólo supera a uno de los métodos aproximativos (SCAT1) y el método aproximativo con mayor puntuación en el plano RMI (SCAT1) supera al otro método aproximativo y a uno lingüístico (SCAT2, LING2).
- En términos de interpretabilidad semántica de las reglas (RMI) el método lingüístico con mayor puntuación (LING2) superó al otro método lingüístico (LING1) y a uno de los métodos aproximativos en los planos NR y GM3M (SCAT1 y SCAT2 respectivamente). Sólo en el plano RMI SCAT2 supera al resto de métodos considerados.

En general los métodos lingüísticos parecen ser los más apropiados para obtener SBRDs con una buena interpretabilidad semántica, aunque los métodos aproximativos son capaces de obtener también SBRDs finales con una buena interpretabilidad en algunos casos. De forma general, para este tipo de SBRDs tampoco hay un método que sea claramente ganador.

Por ejemplo, en algunos conjuntos de datos (STP,WIZ,MOR,TRE) ha sido posible obtener valores máximos de RMI en el plano Precisión/RMI, utilizando un bajo número de reglas y obteniendo valores altos de GM3M, utilizando los métodos SCAT1 o SCAT2. De todos modos, cuando se utiliza el método SCAT2 el error aumenta con respecto a los SBRDs base iniciales. La principal desventaja es que en estos casos la precisión obtenida por los métodos LING1 o LING2 es mejor que la obtenida con SCAT1 o SCAT2.

Tabla 4.16: SBRDs más interpretables: Test de Holm en cada una de las medidas con $\alpha = 0,1$.

Plano Precisión/NR																	
Holm en ECM _{test}			Holm en NR			Holm en GM3M			Holm en RMI								
i	Met.	z	p	α/i	Hip.	i	Met.	z	p	α/i	Hip.	i	Met.	z	p	α/i	Hip.
3	SCAT2	3.65	0.000	0.03	Rech.	3	SCAT2	2.37	0.018	0.03	Rech.	3	LING1	2.37	0.018	0.03	Rech.
2	SCAT1	2.56	0.011	0.05	Rech.	2	LING2	1.83	0.068	0.05	Acep.	2	SCAT1	2.37	0.018	0.05	Rech.
1	LING1	0.00	1.000	0.10	Acep.	1	SCAT1	0.55	0.584	0.10	Acep.	1	LING1	0.18	0.855	0.10	Acep.
Plano Precisión/GM3M																	
Holm en ECM _{test}			Holm en NR			Holm en GM3M			Holm en RMI								
i	Met.	z	p	α/i	Hip.	i	Met.	z	p	α/i	Hip.	i	Met.	z	p	α/i	Hip.
3	SCAT1	3.10	0.002	0.03	Rech.	3	LING2	3.10	0.002	0.03	Rech.	3	SCAT1	2.56	0.011	0.03	Rech.
2	SCAT2	1.64	0.100	0.05	Acep.	2	LING1	2.37	0.018	0.05	Rech.	2	SCAT2	1.64	0.100	0.05	Acep.
1	LING2	0.73	0.465	0.10	Acep.	1	SCAT2	2.19	0.028	0.10	Rech.	1	LING2	0.55	0.584	0.10	Acep.
Plano Precisión/RMI																	
Holm en ECM _{test}			Holm en NR			Holm en GM3M			Holm en RMI								
i	Met.	z	p	α/i	Hip.	i	Met.	z	p	α/i	Hip.	i	Met.	z	p	α/i	Hip.
3	SCAT1	3.47	0.001	0.03	Rech.	3	LING2	2.56	0.011	0.03	Rech.	3	LING2	2.74	0.006	0.03	Rech.
2	SCAT2	2.37	0.018	0.05	Rech.	2	SCAT2	2.01	0.045	0.05	Rech.	2	SCAT2	2.19	0.028	0.05	Rech.
1	LING1	1.10	0.273	0.10	Acep.	1	LING1	0.91	0.361	0.10	Acep.	1	LING1	1.28	0.201	0.10	Acep.

4.3.6. Análisis Global

Finalmente se puede hacer un análisis global de las conclusiones obtenidas en las secciones previas. En la mayoría de los casos se ha conseguido reducir tanto el error como el número de reglas, por lo que se han obtenido SBRDs más precisos y menos complejos. Además se ha conseguido mejorar la interpretabilidad semántica de los SBRDs finales con el enfoque aproximativo, aunque con el enfoque lingüístico no pasa lo mismo. La razón por la que esto ocurre se encuentra en el SBRD de partida, cuya interpretabilidad semántica a nivel de funciones de pertenencia y reglas es máxima en la mayoría de los conjuntos de datos utilizados en la experimentación, lo que hace complicado conseguir las mejoras.

A pesar de ello, al centrarse solo en los SBRDs más precisos y en los más interpretables, se pueden extraer las siguientes conclusiones:

SBRDs más precisos Los métodos aproximativos obtiene mejor precisión, número de reglas e interpretabilidad semántica. El principal problema de estos SBRDs es la pérdida, en un alto grado, de la interpretabilidad semántica de las funciones de pertenencia.

SBRDs más interpretables Con los métodos aproximativos es posible obtener SBRDs con un bajo número de reglas y un valor alto de interpretabilidad GM3M. Por su parte, con los métodos lingüísticos se obtienen SBRDs con un error bajo y un valor alto de interpretabilidad RMI.

En realidad, los resultados obtenidos muestran que es posible obtener SBRDs precisos con métodos lingüísticos y modelos interpretables con métodos aproximativos, lo que choca con la idea preconcebida del rendimiento de los métodos aproximativos y lingüísticos.

SBRDs en la mediana Las conclusiones obtenidas con el análisis de estas soluciones han sido muy diversas, por lo que en estos casos es complicado establecer una regla común de funcionamiento.

Un punto clave de todo el análisis realizado es la existencia de algunos conjuntos de datos en los que las reglas generales preestablecidas se rompen y/o no funcionan como se esperaba. Es decir, no siempre un algoritmo de modelado basado en reglas lingüísticas va a generar un SBRD más interpretable y con menos precisión que un aproximativo, ni viceversa. Por ejemplo en los conjuntos de datos aquí analizados, si el usuario quiere obtener un SBRD con alta precisión para

el conjunto de datos TRE deberá utilizar un algoritmo de modelado lingüístico, y no aproximativo como se podría pensar inicialmente.

La siguiente sección muestra un ejemplo ilustrativo de cómo comparar modelos obtenidos con distintos enfoques, lingüístico y aproximativo, utilizando las métricas aquí propuestas.

4.3.7. Comparación local de modelos lingüísticos y aproximativos

Como se ha podido comprobar previamente en el análisis de resultados, los test estadísticos realizados muestran que lo que cabría esperar dependiendo del tipo de modelado escogido no siempre se da en la práctica dependiendo de muchos factores como los datos o naturaleza del propio problema, o los algoritmos utilizados para el modelado. Así, no siempre un algoritmo de modelado basado en reglas aproximativas va a generar un SBRD más preciso y menos interpretable que un lingüístico, y viceversa. Igualmente, los algoritmos utilizados para un tipo de modelado, bien lingüístico o bien aproximativo, difieren en su comportamiento debido al complejo espacio de búsqueda al que se enfrenta, que hace que en la práctica no siempre lleguen a los niveles de precisión e interpretabilidad óptimos.

Por lo tanto, cuando nos enfrentamos a un nuevo problema del mundo real, es necesario poder analizar y comparar los modelos con los distintos grados de precisión e interpretabilidad obtenidos para una o incluso distintas técnicas de modelado difuso, de manera que un experto pueda decidir qué modelo prefiere de acuerdo a sus preferencias. Las métricas propuestas también se pueden aplicar de manera local para analizar los distintos componentes de un SBRD siguiendo el esquema de representación local y/o pseudo-lingüística presentado en la Sección 4.1.3, ayudando al experto a tomar la decisión final. Se muestra a continuación como, haciendo uso de las métricas aquí propuestas, el usuario puede analizar y comparar distintos modelos para tomar su decisión final.

Ejemplo 1

En primer lugar se han seleccionado dos SBRDs obtenidos con la misma partición del conjunto de datos ELE, uno utilizando el método LING1 y otro el método SCAT1. Ambos sistemas corresponden al SBRD de máxima precisión en cualquiera de los planos analizados. La Tabla 4.17 muestra, para cada uno de

Tabla 4.17: SBRDs del conjunto de datos ELE.

Método	E_{tra}	E_{tst}	NR	GM3M	RMI
LING1	29301.67	36266.97	29	0.314	0.649
SCAT1	17102.27	19865.52	27	0.490	0.730

los sistemas, los valores concretos de los índices de precisión e interpretabilidad considerados en esta propuesta.

El problema de modelado ELE [144] trata de relacionar el coste de mantenimiento de una red eléctrica de media tensión en una ciudad (y) en función de las siguientes cuatro características:

- x_1 : Suma de las longitudes de todas las calles de la ciudad.
- x_2 : Área total de la ciudad.
- x_3 : Área ocupada por edificios.
- x_4 : Energía suministrada a la ciudad.

A continuación se muestran las BCs concretas de cada uno de estos SBRDs. La Fig. 4.10 muestra la BD obtenida con LING1, comparada con la partición difusa interpretable inicial (en nuestro caso una partición difusa uniforme fuertemente distribuida). La Tabla 4.18 muestra la BR de dicho SBRD descrita tal y como se propone en la Sección 4.1.3, en función de dicha partición difusa interpretable inicial, así como el valor RMI individual de cada regla.

Por su parte la Fig. 4.11 muestra las funciones de pertenencia obtenidas con SCAT1, comparada con la partición difusa interpretable inicial y la Tabla 4.18 muestra la BR de dicho SBRD descrita de la misma forma (representación pseudo-lingüística) y el valor RMI de cada regla.

Observando las características generales de los SBRDs desde nuestro punto de vista (obviamente subjetivo) la mejor elección es seleccionar la obtenida con el método SCAT1 ya que no solo es más precisa, sino que en general obtiene también mejores valores en todas las métricas de interpretabilidad.

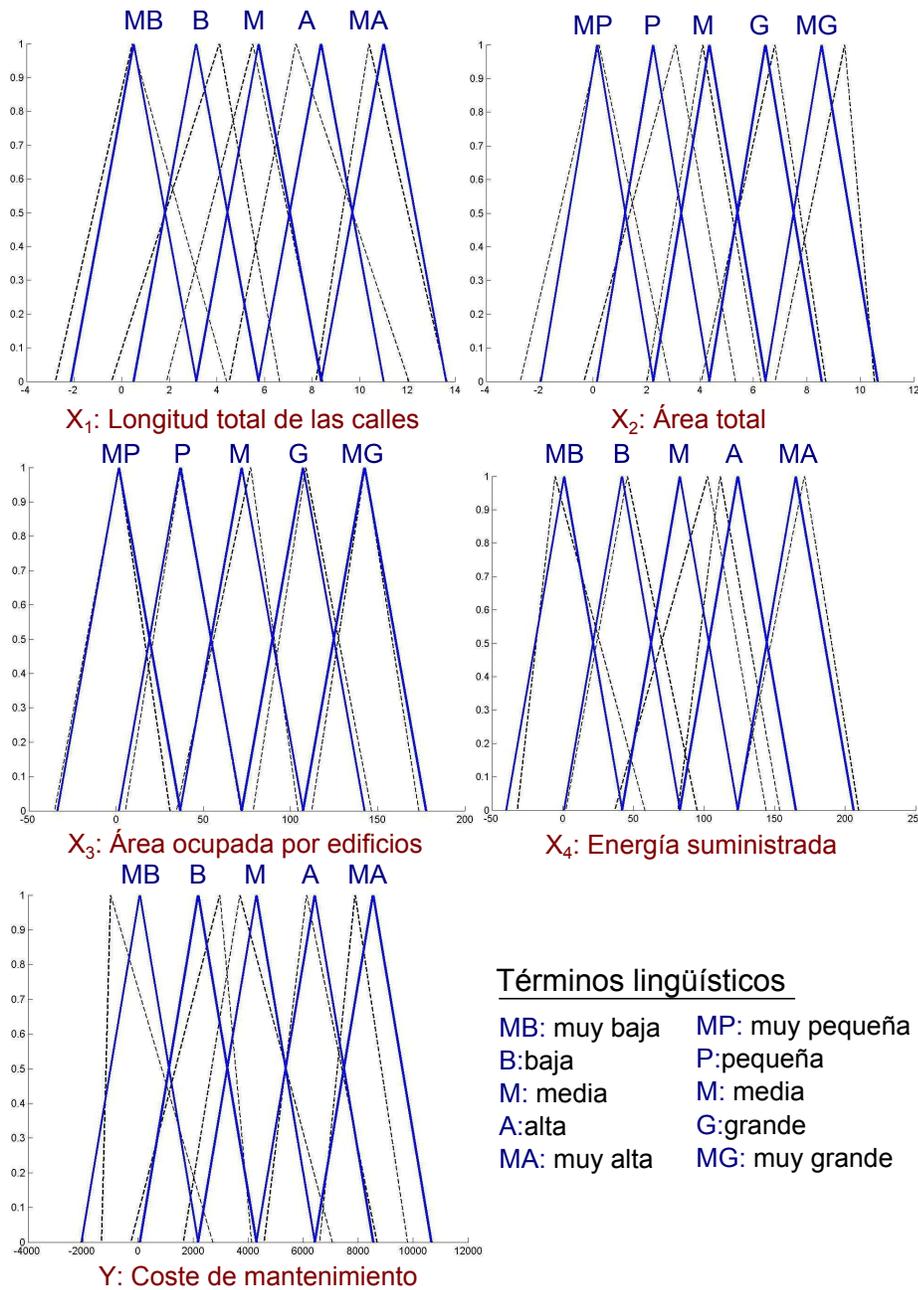


Figura 4.10: BD obtenida para ELE con el método LING1.

Tabla 4.18: BR obtenida para ELE con el método LING1.

Regla	X ₁	X ₂	X ₃	X ₄	Y	RMI	
R ₁	MB ^{0,83}	MP ^{0,86}	MP ^{0,90}	MB ^{0,68}	MB ^{0,34}	0.68	Precisión <i>ECM_{tra}</i> : 29301.67 <i>ECM_{tst}</i> : 36266.97
R ₂	MB ^{0,83}	MP ^{0,86}	MP ^{0,90}	MB ^{0,68}	MB ^{0,34}	0.68	
R ₃	B ^{0,63}	MP ^{0,86}	MP ^{0,90}	MB ^{0,68}	MB ^{0,34}	0.67	
R ₄	B ^{0,63}	MP ^{0,86}	MP ^{0,90}	B ^{0,89}	B ^{0,60}	0.70	
R ₅	B ^{0,63}	MP ^{0,86}	MP ^{0,90}	B ^{0,89}	B ^{0,60}	0.70	
R ₆	B ^{0,63}	P ^{0,66}	P ^{0,95}	MB ^{0,68}	MB ^{0,34}	0.69	
R ₇	B ^{0,63}	P ^{0,66}	P ^{0,95}	B ^{0,89}	B ^{0,60}	0.99	Variables X ₁ : Longitud total de las calles X ₂ : Área total X ₃ : Área edificios X ₄ : Energía suministrada Y: Coste mantenimiento
R ₈	B ^{0,63}	P ^{0,66}	P ^{0,95}	B ^{0,89}	M ^{0,70}	0.64	
R ₉	M ^{0,83}	P ^{0,66}	MP ^{0,90}	M ^{0,62}	B ^{0,60}	0.78	
R ₁₀	M ^{0,83}	M ^{0,93}	P ^{0,95}	B ^{0,89}	B ^{0,60}	0.96	
R ₁₁	M ^{0,83}	M ^{0,93}	P ^{0,95}	M ^{0,62}	M ^{0,70}	0.82	
R ₁₂	M ^{0,83}	M ^{0,93}	M ^{0,81}	B ^{0,89}	M ^{0,70}	0.83	
R ₁₃	M ^{0,83}	M ^{0,93}	M ^{0,81}	B ^{0,89}	M ^{0,70}	0.83	
R ₁₄	M ^{0,83}	M ^{0,93}	M ^{0,81}	B ^{0,89}	M ^{0,70}	0.83	
R ₁₅	A ^{0,62}	P ^{0,66}	P ^{0,95}	B ^{0,89}	B ^{0,60}	0.97	
R ₁₆	A ^{0,62}	P ^{0,66}	P ^{0,95}	M ^{0,62}	M ^{0,70}	0.83	
R ₁₇	A ^{0,62}	M ^{0,93}	P ^{0,95}	MB ^{0,68}	B ^{0,60}	0.78	
R ₁₈	A ^{0,62}	M ^{0,93}	P ^{0,95}	M ^{0,62}	M ^{0,70}	0.83	
R ₁₉	A ^{0,62}	M ^{0,93}	P ^{0,95}	A ^{0,75}	M ^{0,70}	0.83	Términos lingüísticos MB: muy baja B: baja M: media A: alta MA: muy alta MP: muy pequeña P: pequeña M: media G: grande MG: muy grande
R ₂₀	A ^{0,62}	M ^{0,93}	M ^{0,81}	M ^{0,62}	A ^{0,79}	0.86	
R ₂₁	A ^{0,62}	G ^{0,80}	M ^{0,81}	B ^{0,89}	M ^{0,70}	0.78	
R ₂₂	A ^{0,62}	G ^{0,80}	M ^{0,81}	M ^{0,62}	A ^{0,79}	0.86	
R ₂₃	A ^{0,62}	G ^{0,80}	G ^{0,90}	B ^{0,89}	A ^{0,79}	0.86	
R ₂₄	A ^{0,62}	MG ^{0,60}	MG ^{0,95}	M ^{0,62}	MA ^{0,70}	0.89	
R ₂₅	MA ^{0,79}	P ^{0,66}	P ^{0,95}	MA ^{0,86}	A ^{0,79}	0.86	
R ₂₆	MA ^{0,79}	P ^{0,66}	M ^{0,81}	B ^{0,89}	M ^{0,70}	0.82	
R ₂₇	MA ^{0,79}	P ^{0,66}	M ^{0,81}	B ^{0,89}	M ^{0,70}	0.82	
R ₂₈	MA ^{0,79}	G ^{0,80}	M ^{0,81}	MA ^{0,86}	MA ^{0,70}	0.89	
R ₂₉	MA ^{0,79}	G ^{0,80}	M ^{0,81}	B ^{0,89}	M ^{0,70}	0.79	

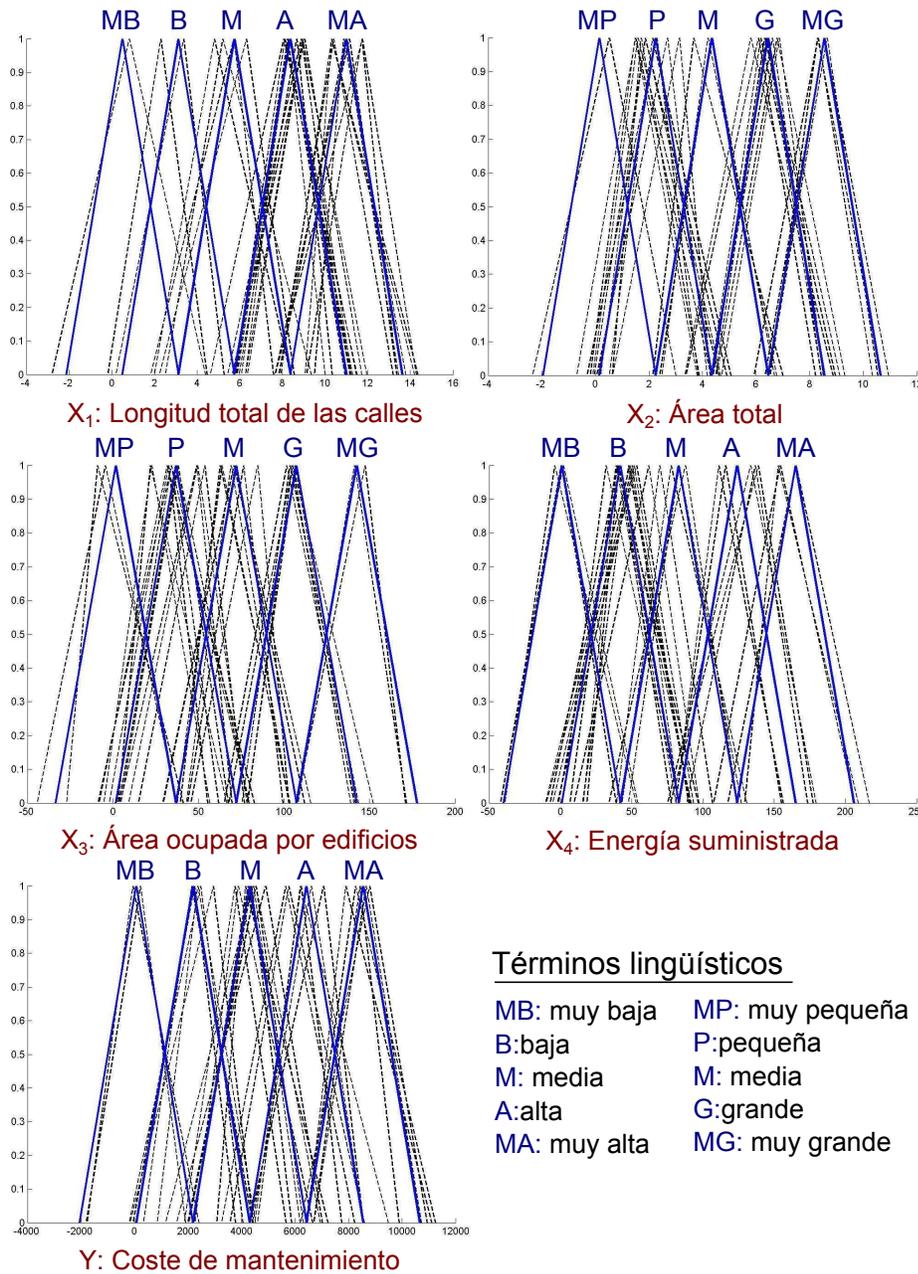


Figura 4.11: BD obtenida para ELE con el método SCAT1.

Tabla 4.19: BR obtenida para ELE con el método SCAT1.

Regla	X ₁	X ₂	X ₃	X ₄	Y	RMI	
R ₁	B ^{0,80}	P ^{0,76}	P ^{0,65}	B ^{0,78}	B ^{0,65}	0.88	Precisión <i>ECM_{tra}</i> : 17102.27 <i>ECM_{tst}</i> : 19865.52
R ₂	MB ^{0,85}	MP ^{0,82}	MP ^{0,87}	MB ^{0,88}	MB ^{0,86}	0.89	
R ₃	B ^{0,80}	P ^{0,75}	P ^{0,84}	B ^{0,87}	B ^{0,82}	0.88	
R ₄	M ^{0,70}	P ^{0,78}	P ^{0,75}	B ^{0,75}	B ^{0,79}	0.97	
R ₅	M ^{0,74}	P ^{0,73}	P ^{0,64}	B ^{0,68}	M ^{0,70}	0.73	Variables <i>X</i> ₁ : Longitud total de las calles <i>X</i> ₂ : Área total <i>X</i> ₃ : Área edificios <i>X</i> ₄ : Energía suministrada <i>Y</i> : Coste mantenimiento
R ₆	A ^{0,79}	M ^{0,84}	P ^{0,76}	M ^{0,80}	M ^{0,96}	0.97	
R ₇	M ^{0,73}	G ^{0,80}	M ^{0,78}	B ^{0,71}	M ^{0,87}	0.82	
R ₈	MA ^{0,88}	M ^{0,93}	P ^{0,84}	B ^{0,77}	B ^{0,83}	0.99	
R ₉	A ^{0,81}	P ^{0,65}	P ^{0,60}	A ^{0,75}	M ^{0,74}	0.90	
R ₁₀	MA ^{0,71}	P ^{0,79}	P ^{0,90}	B ^{0,83}	M ^{0,90}	0.98	
R ₁₁	A ^{0,80}	G ^{0,94}	G ^{0,88}	M ^{0,92}	A ^{0,92}	0.98	
R ₁₂	A ^{0,87}	G ^{0,75}	M ^{0,62}	M ^{0,76}	A ^{0,76}	0.97	
R ₁₃	A ^{0,90}	P ^{0,97}	P ^{0,90}	MB ^{0,95}	B ^{0,90}	0.94	
R ₁₄	MA ^{0,81}	G ^{0,82}	M ^{0,84}	B ^{0,75}	M ^{0,83}	0.95	
R ₁₅	A ^{0,86}	G ^{0,87}	M ^{0,78}	A ^{0,68}	A ^{0,79}	0.97	
R ₁₆	MA ^{0,61}	P ^{0,75}	P ^{0,77}	A ^{0,86}	M ^{0,76}	0.78	
R ₁₇	A ^{0,91}	MG ^{0,86}	MG ^{0,92}	M ^{0,81}	MA ^{0,77}	0.97	
R ₁₈	A ^{0,95}	G ^{0,83}	G ^{0,84}	B ^{0,72}	A ^{0,84}	0.95	
R ₁₉	MA ^{0,70}	P ^{0,69}	M ^{0,90}	MA ^{0,91}	A ^{0,72}	0.86	
R ₂₀	A ^{0,86}	MG ^{0,94}	G ^{0,84}	A ^{0,79}	MA ^{0,84}	0.94	
R ₂₁	M ^{0,75}	P ^{0,82}	MP ^{0,62}	MB ^{0,78}	MB ^{0,87}	0.94	
R ₂₂	A ^{0,81}	MG ^{0,92}	G ^{0,85}	B ^{0,90}	M ^{0,84}	0.99	
R ₂₃	A ^{0,84}	MG ^{0,89}	MG ^{0,77}	B ^{0,78}	MA ^{0,87}	0.82	
R ₂₄	MA ^{0,70}	M ^{0,64}	M ^{0,72}	B ^{0,92}	M ^{0,67}	0.99	
R ₂₅	MA ^{0,71}	G ^{0,89}	M ^{0,92}	MA ^{0,65}	MA ^{0,65}	0.87	
R ₂₆	A ^{0,76}	G ^{0,82}	G ^{0,88}	A ^{0,78}	MA ^{0,94}	0.98	
R ₂₇	MA ^{0,89}	P ^{0,61}	P ^{0,80}	MA ^{0,81}	A ^{0,72}	0.85	
							Términos lingüísticos MB: muy baja B: baja M: media A: alta MA: muy alta MP: muy pequeña P: pequeña M: media G: grande MG: muy grande

Comparando las propiedades particulares de ambos SBRDs se puede decir que:

- En el SBRD SCAT1 casi todas las funciones de pertenencia utilizadas por las reglas aproximativas se pueden asociar con las correspondientes funciones de pertenencia dadas en la partición inicial con un alto grado de similitud, por lo que su interpretabilidad semántica a nivel de funciones de pertenencia también es elevada.
- En el SBRD SCAT1 todas las reglas tienen un valor RMI de confianza alto, por lo que casi no hay inconsistencia entre ellas. La interpretabilidad semántica a nivel de BR es por tanto elevada.

Por su parte dicho valor RMI en LING1 es algo más pequeño en casi todas las reglas, mostrando que el significado de algunas reglas no es totalmente representativo del comportamiento real del sistema.

- El número de reglas de ambos SBRDs es similar, por lo que no es un factor determinante a la hora de seleccionar un modelo u otro.

Ejemplo 2

Posteriormente se seleccionan dos SBRDs obtenidos con la misma partición del conjunto de datos PLA, uno utilizando el método LING2 y otro el método SCAT2. Ambos sistemas corresponden al SBRD situado en la mediana del plano Precision/NR. La Tabla 4.20 muestra, para cada uno de los sistemas, los valores concretos de los índices de precisión e interpretabilidad considerados en este caso.

Tabla 4.20: SBRDs del conjunto de datos PLA.

Método	E_{tra}	E_{tst}	NR	GM3M	RMI
LING2	2.59	2.59	14	0.613	1.000
SCAT2	2.10	2.30	23	0.466	0.380

El problema de modelado PLA³ trata de calcular cuanta presión (y) puede soportar una determinada pieza de plástico cuando se aplica sobre ella una fuerza (x_1) a una determinada temperatura (x_2).

³<http://sci2s.ugr.es/keel/dataset.php?cod=74>

Al igual que antes la Fig. 4.12 muestra la BD obtenida con LING2, comparada con la partición difusa interpretable inicial, la Tabla 4.21 muestra la BR descrita en función de dicha partición y el valor RMI individual de cada regla, la Fig. 4.13 muestra las funciones de pertenencia obtenidas con SCAT2 y la Tabla 4.22 muestra la BR.

En este caso la elección del SBRD, aunque dependerá de las necesidades finales de un usuario o experto para el problema PLA, según nuestro propio criterio (de nuevo subjetivo) debería inclinarse más hacia el método LING2 ya que el sistema muestra mejores propiedades de interpretabilidad a todos los niveles y la pérdida de precisión con respecto al modelo SCAT2 parece mínima.

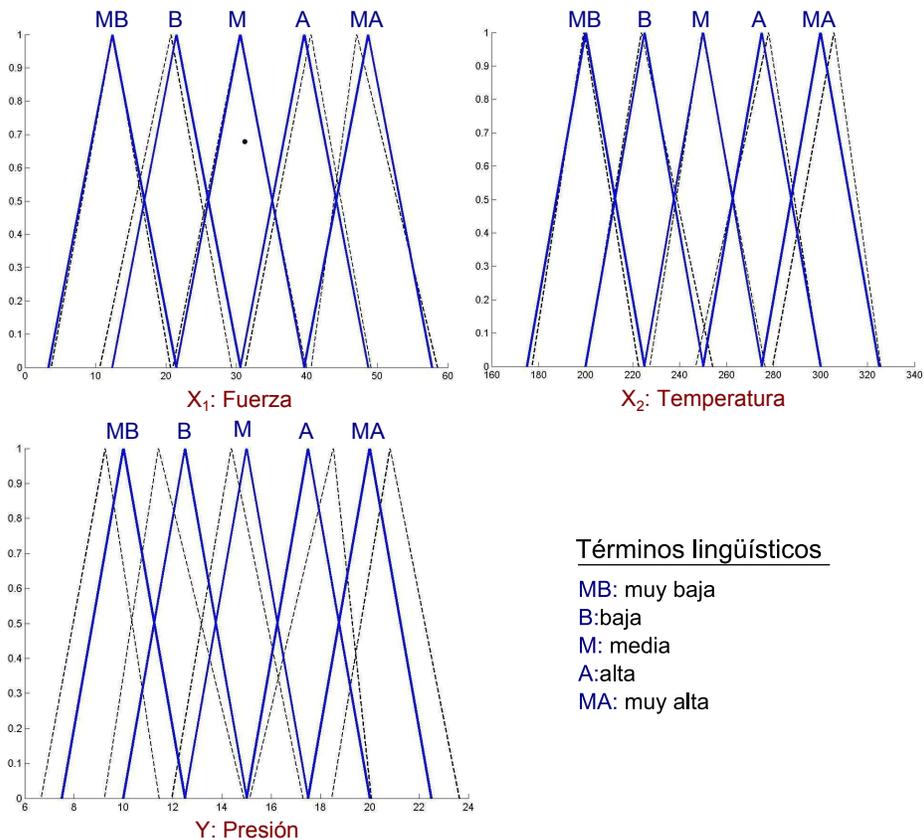


Figura 4.12: BD obtenida para PLA con el método LING2.

Tabla 4.21: BR obtenida para PLA con el método LING2.

Regla	X_1	X_2	Y	RMI	
R ₁	MB ^{0,95}	MB ^{0,94}	MA ^{0,81}	1.00	Precisión <i>ECM_{tra}</i> : 2.59 <i>ECM_{tst}</i> : 2.59
R ₂	MB ^{0,95}	B ^{0,88}	MA ^{0,81}	1.00	
R ₃	B ^{0,91}	MB ^{0,94}	M ^{0,83}	1.00	Variables X_1 :Fuerza X_2 :Temperatura Y:Presión
R ₄	B ^{0,91}	B ^{0,88}	A ^{0,64}	1.00	
R ₅	B ^{0,91}	M ^{0,93}	MA ^{0,81}	1.00	Términos lingüísticos MB: muy baja B: baja M: media A: alta MA: muy alta
R ₆	M ^{0,98}	MB ^{0,94}	MB ^{0,82}	1.00	
R ₇	M ^{0,98}	B ^{0,88}	B ^{0,68}	1.00	
R ₈	M ^{0,98}	A ^{0,83}	A ^{0,64}	1.00	
R ₉	M ^{0,98}	MA ^{0,81}	MA ^{0,81}	1.00	
R ₁₀	A ^{0,93}	M ^{0,93}	MB ^{0,82}	1.00	
R ₁₁	A ^{0,93}	A ^{0,83}	B ^{0,68}	1.00	
R ₁₂	A ^{0,93}	MA ^{0,81}	M ^{0,83}	1.00	
R ₁₃	MA ^{0,77}	A ^{0,83}	MB ^{0,82}	1.00	
R ₁₄	MA ^{0,77}	MA ^{0,81}	MB ^{0,82}	1.00	

Al comparar la interpretabilidad de ambos SBRDs se tiene que:

- El número de reglas es mucho menor en el caso de LING2, siendo un conjunto lo suficientemente pequeño como para que pueda ser interpretado por una persona.
- La interpretabilidad semántica de la BR en LING2 es máxima, siendo esta la situación ideal a obtener en todos los modelos. Dicho valor indica que no hay inconsistencias entre las reglas. En SCAT2 el valor RMI les asigna a muchas de las reglas un nivel de confianza medio, por lo que no son tan representativas del comportamiento real del modelo.
- Las funciones de pertenencia en LING2 son en general más próximas a las de la partición inicial interpretable que en el caso de las obtenidas por SCAT2.

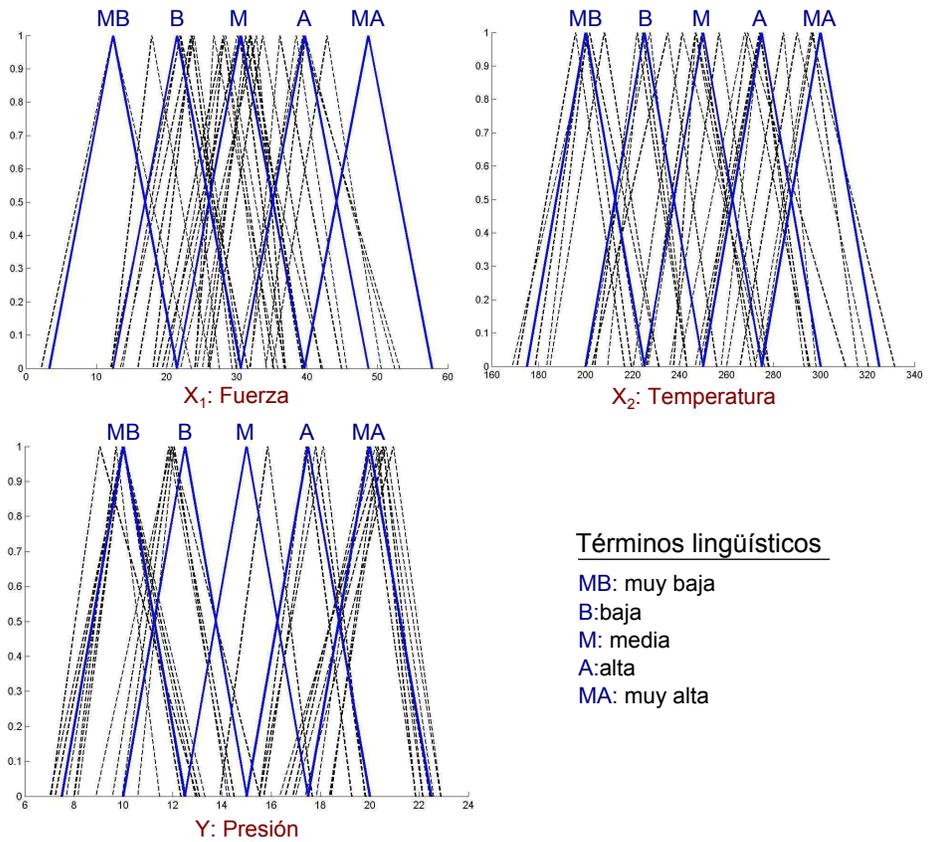


Figura 4.13: BD obtenida para PLA con el método SCAT2.

Tabla 4.22: BR obtenida para PLA con el método SCAT2.

Regla	X_1	X_2	Y	RMI	
R ₁	B ^{0,68}	MB ^{0,85}	MA ^{0,74}	0.50	Precisión <i>ECM_{tra}</i> : 2.10 <i>ECM_{tst}</i> : 2.30
R ₂	M ^{0,47}	M ^{0,92}	A ^{0,89}	0.99	
R ₃	M ^{0,61}	B ^{0,71}	MB ^{0,63}	0.69	
R ₄	M ^{0,84}	MA ^{0,85}	MA ^{0,76}	0.97	
R ₅	B ^{0,76}	MB ^{0,85}	B ^{0,74}	0.70	
R ₆	A ^{0,81}	MA ^{0,76}	MB ^{0,86}	0.56	
R ₇	A ^{0,80}	A ^{0,85}	MB ^{0,90}	0.81	Variables X_1 : Fuerza X_2 : Temperatura Y: Presión
R ₈	A ^{0,59}	MA ^{0,78}	A ^{0,75}	0.86	
R ₉	B ^{0,79}	MB ^{0,83}	MB ^{0,82}	0.57	
R ₁₀	M ^{0,84}	A ^{0,85}	A ^{0,89}	0.69	
R ₁₁	B ^{0,94}	M ^{0,69}	MA ^{0,78}	0.83	
R ₁₂	M ^{0,63}	A ^{0,66}	MA ^{0,66}	0.59	
R ₁₃	M ^{0,88}	MB ^{0,87}	MB ^{0,84}	0.81	Términos lingüísticos MB: muy baja B: baja M: media A: alta MA: muy alta
R ₁₄	M ^{0,88}	A ^{0,64}	MA ^{0,90}	0.64	
R ₁₅	B ^{0,79}	B ^{0,76}	MA ^{0,68}	0.49	
R ₁₆	M ^{0,80}	M ^{0,86}	MB ^{0,90}	0.57	
R ₁₇	M ^{0,81}	B ^{0,83}	B ^{0,86}	0.55	
R ₁₈	M ^{0,73}	M ^{0,84}	MA ^{0,72}	0.38	
R ₁₉	A ^{0,75}	A ^{0,88}	MB ^{0,89}	0.75	
R ₂₀	M ^{0,76}	M ^{0,59}	B ^{0,83}	0.62	
R ₂₁	B ^{0,76}	B ^{0,80}	M ^{0,77}	0.69	
R ₂₂	A ^{0,84}	A ^{0,83}	B ^{0,73}	0.76	
R ₂₃	MB ^{0,91}	MB ^{0,90}	MA ^{0,89}	0.99	

Capítulo 5

Conclusiones

La memoria de esta tesis concluye con un resumen breve de los resultados y las contribuciones más importantes obtenidas a lo largo del desarrollo de la misma. Por otro lado se presentan las publicaciones asociadas a este trabajo y se comentan diversos aspectos sobre otros trabajos en desarrollo que prosiguen las líneas de trabajo aquí expuestas.

5.1. Resumen del trabajo realizado

La presente tesis doctoral trata una parte de la problemática asociada a la búsqueda de equilibrio entre los conceptos de precisión e interpretabilidad a la hora de diseñar SBRDs, concretamente la definición de nuevas métricas de interpretabilidad y su aplicación al modelado de sistemas lingüísticos y aproximativos.

Tras la revisión crítica realizada (Capítulo 2) sobre las ideas, definiciones y formulaciones acerca de los conceptos de interpretabilidad y precisión en el ámbito de la Lógica Difusa y, más concretamente de los SBRDs, los capítulos siguientes proponen metodologías para conseguir SBRDs con mejores propiedades tanto en precisión como en interpretabilidad.

En primer lugar (Capítulo 3) se ha presentado e implementado una nueva métrica de **interpretabilidad semántica de la BR** denominada **Inter_C**, resultado de la combinación de un conjunto de índices capaces de evaluar la compactitud, la similitud, la redundancia, la consistencia y la completitud de los

conjuntos y reglas difusas. Dicha métrica de interpretabilidad es utilizada, junto con un índice de precisión, por un AG bi-objetivo para realizar selección de reglas difusas en una etapa de post-procesamiento.

La validación de la propuesta se ha llevado a cabo utilizando nueve casos de estudio presentes en el proyecto KEEL y dos algoritmos de modelos difuso capaces de generar SBRD con buena precisión: NEFPROX como algoritmos de modelado lingüístico y FASART como algoritmo de modelado aproximativo. Con cada uno de estos algoritmos se generan dos tipos de modelos con diferentes características de complejidad y propiedades, de forma que la propuesta es validada en distintos entornos difusos.

Los resultados obtenidos confirman que la metodología de mejora propuesta, que puede ser aplicada tanto a SBRD lingüísticos como aproximativos, independientemente del tipo de partición difusa que se utilice (triangular, trapezoidal, gaussiana, etc.), consigue obtener SBRDs con un buen equilibrio precisión-interpretabilidad.

Además, (Capítulo 4) a partir del análisis de los resultados obtenidos previamente y tomando como referencia el estudio realizado en [29], se han propuesto dos nuevas métricas de interpretabilidad semántica, una a nivel de BR y otra a nivel de BD:

- **RMI** como nueva métrica de **interpretabilidad semántica a nivel de BR** basada en evaluar el grado en que el significado de cada una de las reglas del SBRD representa lo que el sistema completo diría para sus respectivas zonas de activación, teniendo en cuenta para su cálculo el sistema de inferencia concreto utilizado.
- Generalización de **GM3M**, medida relativa de **interpretabilidad semántica a nivel de BD** propuesta inicialmente por Gacto y otros en [77], para que pueda ser aplicada a otros tipos de particiones difusas (no sólo triangulares) y a los SBRDs aproximativos (no sólo lingüísticos).

Ambas métricas son utilizadas dentro de una etapa de post-procesamiento que, utilizando AEMOs desarrollados específicamente para esta propuesta, realiza simultáneamente selección de reglas y ajuste de las particiones difusas. La validación de la propuesta se ha llevado a cabo utilizando nueve casos de estudio presentes en el proyecto KEEL y cuatro algoritmos de modelado difuso, dos lingüísticos (NEFPROX y L-IRL) y dos aproximativos (FASART y S-ILR),

capaces de generar modelos con distintas propiedades iniciales de precisión e interpretabilidad, de forma que la propuesta es validada en contextos diferentes.

En este caso los resultados han confirmado que es posible mejorar tanto la interpretabilidad semántica, como la complejidad y la precisión de los SBRDs de forma simultánea, mejorando así tanto la exactitud como las capacidades lingüísticas de los sistemas.

Finalmente, como ambas métricas GM3M y RMI pueden ser utilizadas con cualquier tipo de SBRD (lingüístico o aproximativo) y partición difusa (triangular, trapezoidal, gaussiana, etc.), se ha propuesto un método de comparación local de modelos lingüísticos y/o aproximativos que ayuda al usuario a elegir, en cada momento, el sistema de reglas que más le interese.

5.2. Contribuciones de esta tesis

Las contribuciones principales aportadas por esta tesis, y las publicaciones asociadas a cada una de ellas, se comentan a continuación:

- **Definición de la métrica $Inter_C$** de interpretabilidad semántica a nivel de BR, obtenida mediante la agregación de diversos índices de interpretabilidad sencillos de entender, que permiten evaluar las capacidades lingüísticas tanto de los SBRDs lingüísticos como aproximativos.

Los resultados experimentales presentados (Capítulo 3) muestran que el uso de la métrica $Inter_C$ permite reducir la complejidad, por lo que se mejora la interpretabilidad, en todos los SBRDs. Además, dicha mejora en la interpretabilidad no implica necesariamente una pérdida de precisión, y cuando dicha pérdida ocurre esta se produce de forma moderada y aceptable. Incluso en no pocos casos la precisión se conserva o incluso se mejora.

La publicación consecuencia de esta parte del trabajo es:

- ✓ M. Galende-Hernández, G.I. Sainz-Palmero and M.J. Fuente-Aparicio. Complexity Reduction and Interpretability Improvement for fuzzy rule systems based on simple interpretability measures and indices by bi-objective evolutionary rule selection. *Soft Computing*, 16, Issue 3, pp. 451-470. March 2012 [145].

Parcialmente parte de esta propuesta está utilizada también en:

- ✓ M.I. Rey, M. Galende, M. Fuente, G. I. Sainz Palmero. Checking Orthogonal Transformations and Genetic Algorithms for Selection of Fuzzy Rules based on Interpretability-Accuracy Concepts. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 20, Suppl. 2, pp. 159 - 186. October 2012 [146].
- **Generalización de la métrica GM3M** de interpretabilidad semántica a nivel de BD, de forma que pueda ser utilizada con cualquier SBRD (lingüístico o aproximativo) y partición difusa (triangular, trapezoidal, gaussiana, etc.). De esta forma se puede llegar a cuantificar la mucha o poca diferencia a nivel de interpretabilidad semántica que hay entre SBRDs de distinta naturaleza.

La publicación asociada a esta propuesta es:

- ✓ M. Galende, M.J. Gacto, G. Sainz, R. Alcalá. On the Interpretability of Linguistic and Scatter FRBSs: GM3M Generalization and New Rule Meaning Index for Global Assessment and Local Pseudo-Linguistic Representation. *Information Sciences*. Enviado, 2013 [147].
- **Definición de la métrica RMI** de interpretabilidad semántica a nivel de BR, que: (1) tiene en cuenta la forma en que interactúan las reglas, (2) es lo suficientemente general como para ser aplicable a cualquier tipo de partición difusa y tipo de sistemas, lingüísticos y/o aproximativos, y (3) no resulta computacionalmente costosa, de forma que se ha podido integrar en un proceso de optimización evolutivo sin que represente un problema. El índice, al calcularse teniendo en cuenta el sistema de inferencia, va a ser capaz de detectar tanto los problemas derivados de la elección de un mal operador como aquellos que pudieran aparecer como consecuencia del uso de pesos en las reglas.

Los resultados presentados en el Capítulo 4 muestran que el uso conjunto de las métricas RMI, GM3M, número de reglas y error cuadrático medio, dentro de un proceso de optimización genética multi-objetivo, permite mejorar las propiedades de los SBRDs, tanto lingüísticos como aproximativos, en términos de precisión e interpretabilidad.

La publicación asociada es:

- ✓ M. Galende, M.J. Gacto, G. Sainz, R. Alcalá. On the Interpretability of Linguistic and Scatter FRBSs: GM3M Generalization and New Rule Meaning Index for Global Assessment and Local Pseudo-Linguistic Representation. *Information Sciences*. Enviado, 2013 [147].
- Definición de un **marco de análisis y comparación de sistemas lingüísticos y aproximativos** en base a las nuevas medidas de interpretabilidad semántica RMI y GM3M y a las medidas ya conocidas de complejidad y precisión. De esta forma el usuario podrá seleccionar en cada momento la BC más conveniente para su problema.

Los resultados presentados (Capítulo 4) muestran que: no existen tendencias generales en cuanto al comportamiento de los algoritmos considerados, siendo difícil afirmar que alguno de dichos algoritmos es mejor que el resto en alguno de los objetivos considerados. Las reglas generales preestablecidas no se cumplen siempre, por lo que es necesario analizar cada caso de forma particular.

La publicación asociada es:

- ✓ M. Galende, M.J. Gacto, G. Sainz, R. Alcalá. On the Interpretability of Linguistic and Scatter FRBSs: GM3M Generalization and New Rule Meaning Index for Global Assessment and Local Pseudo-Linguistic Representation. *Information Sciences*. Enviado, 2013 [147].

A raíz del trabajo realizado se ha generado también la siguiente publicación:

- ✓ M.J. Gacto, M. Galende, R. Alcalá, F. Herrera. METSK-HD^e: A multiobjective Evolutionary Algorithm to learn accurate TSK-fuzzy Systems in High-Dimensional and Large-Scale Regression Problems. *Information Sciences*. En Prensa, 2013 [148].

5.3. Lista de publicaciones relacionadas con esta tesis

A continuación se presenta un listado completo de las publicaciones relacionadas con el trabajo realizado en esta tesis. En concreto 4 artículos en revistas

internacionales con índice de impacto (una de ellas por aparecer y una en proceso de revisión), 7 contribuciones a congresos internacionales y 3 en congresos nacionales.

■ Publicaciones en revistas internacionales:

- M. Galende, M.J. Gacto, G. Sainz, R. Alcalá. On the Interpretability of Linguistic and Scatter FRBSs: GM3M Generalization and New Rule Meaning Index for Global Assessment and Local Pseudo-Linguistic Representation. *Information Sciences*. Enviado, 2013 [147].
- M.J. Gacto, M. Galende, R. Alcalá, F. Herrera. METSK-HD^e: A Multiobjective Evolutionary Algorithm to learn accurate TSK-fuzzy Systems in High-Dimensional and Large-Scale Regression Problems. *Information Sciences*. En Prensa, 2013 [148].
- M.I. Rey, M. Galende, M. Fuente, G. I. Sainz Palmero. Checking Orthogonal Transformations and Genetic Algorithms for Selection of Fuzzy Rules based on Interpretability-Accuracy Concepts. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 20, Suppl. 2, pp. 159 - 186. October 2012. [146].
- M. Galende-Hernández, G.I. Sainz-Palmero and M.J. Fuente-Aparicio. Complexity Reduction and Interpretability Improvement for fuzzy rule systems based on simple interpretability measures and indices by bi-objective evolutionary rule selection. *Soft Computing*, 16, Issue 3, pp. 451-470. March 2012 [145].

■ Publicaciones en congresos internacionales:

- M. I. Rey, M. Galende, G. I. Sainz, M. J. Fuente. Selection of Rules by Orthogonal Transformations and Genetic Algorithms to Improve the Interpretability in Fuzzy Rule Based Systems. In *Proceedings of 2013 IEEE International Conference on Fuzzy Systems (Fuzz-IEEE 2013)*, Hyderabad, India, July 2013 [149].
- M. J. Gacto, M. Galende, R. Alcalá, F. Herrera. Obtaining Accurate TSK Fuzzy Rule-Based Systems by Multi-Objective Evolutionary Learning in High-Dimensional Regression Problems. In *Proceedings of 2013 IEEE International Conference on Fuzzy Systems (Fuzz-IEEE 2013)*, Hyderabad, India, July 2013 [150].

- M. Galende, G. Sainz, M. Fuente. Accuracy-Interpretability Trade-Off for Precise Fuzzy Modeling using simple indices. Application to Industrial Plants. In *Proceedings of the 18th IFAC World Congress*, pp. 12656 - 12661, Milano, Italy, September 2011 [151].
- M. I. Rey, M. Galende, G.I. Sainz, M. Fuente. Checking Orthogonal Transformations and Genetic Algorithms for Selection of Fuzzy Rules based on Interpretability-Accuracy Concepts. In *Proceedings of 2011 IEEE International Conference on Fuzzy Systems (Fuzz-IEEE 2011)*, pp 1271 - 1278, Taipei, Taiwan, June 2011 [152].
- M. Galende, G. Sainz, M. Fuente. Accuracy-Interpretability Balancing in Fuzzy models based on Multiobjective Genetic Algorithm. In *Proceedings of European Control Conference 2009 (ECC'09)*, pp. 3915 - 3920, Budapest, Hungary, August 2009 [153].
- M.I. Rey, M. Galende, G. Sainz. Criteria for Linguistic Improvement of Precise Fuzzy Models by Orthogonal Transforms. Application to ART Based Models. In *Proceedings of European Control Conference 2009 (ECC'09)*, pp 348 - 353, Budapest, Hungary, August 2009 [154].
- M. Galende, G. I. Sainz, M. J. Fuente and A. Herreros. Interpretability-accuracy improvement in a neuro-fuzzy art based model of a DC motor. In *Proceedings of the 17th IFAC World Congress*, pp. 7034 - 7039, Seoul, Korea, July 2008 [155].
- Publicaciones en congresos nacionales:
 - M. Galende, G. Sainz. Mejora del equilibrio precisión-interpretabilidad en modelos difusos. In *XV Congreso Español Sobre Tecnologías y Lógica Fuzzy (ESTYLF2010)*, pp. 79 - 84, Huelva, Spain, February 2010 [156].
 - M.I. Rey, M. Galende, G. Sainz. Criterios para el empleo de transformaciones ortogonales en la mejora de modelos difusos. Aplicación a modelos basados en ART. In *XV Congreso Español Sobre Tecnologías y Lógica Fuzzy (ESTYLF2010)*, pp. 91 - 96, Huelva, Spain, February 2010 [157].
 - M. Galende and G. I. Sainz. Mejora lingüística de modelos neurodifusos mediante algoritmos genéticos: Aplicación a un motor DC. In *Proceedings of the XXVIII Jornadas de Automáticas*, Huelva, Spain, September 2007 [158].

5.4. Líneas de trabajo en desarrollo

Tras el trabajo realizado en esta tesis y las conclusiones obtenidas de la experimentación realizada, actualmente se encuentran en marcha las siguientes ideas que esperamos sirvan para seguir mejorando el modelado de los SBRDs en términos de precisión e interpretabilidad:

- Dado que el proceso de mejora en dos fases aquí empleado es bastante flexible, se están realizando nuevas experimentaciones que combinan las nuevas métricas aquí propuestas con la información proporcionada por las transformaciones ortogonales, capaces de evaluar la importancia de cada regla en función de la información que contienen.
- Se pretende aplicar los métodos aquí descritos a problemas reales de la industria, como por ejemplo los procesos biotecnológicos existentes en el tratamiento de aguas residuales. El objetivo aquí es doble: generar un buen modelo difuso que permita realizar un pequeño proceso de extracción de conocimiento, gracias al conjunto de reglas interpretables capaces de describir las complejas relaciones existentes que rigen dichos tratamientos.
- Visto que en los últimos años, a medida que las nuevas tecnologías han ido evolucionando, la cantidad de datos almacenados también ha ido aumentando el tema de la alta dimensionalidad es otro aspecto que no podemos olvidar a la hora de diseñar SBRD. La metodología propuesta, basada en el uso de algoritmos genéticos, requiere en algunos casos mucho tiempo de computación por lo que es necesario modificar alguno de sus elementos para que el proceso de modelado no requiera demasiado tiempo. En este sentido ya se han realizado algunas propuestas [148, 150].

Igualmente una idea futura de trabajo se muestra a continuación:

- Teniendo en cuenta la generalización del índice GM3M, un aspecto que no se ha tratado es el uso de particiones alternativas cuando un usuario/experto pueda proporcionarlas. Si un experto es capaz de definir una partición con 3 etiquetas y otro con 5 para la misma variable, se amplían las posibilidades por ejemplo para obtener SBRDs aproximativos que se puedan acercar a una o a otro. GM3M se calcularía sobre cada uno de los posibles casos (combinaciones de particiones originales) y nos quedaríamos con el mejor

valor. Esta forma de uso se podría utilizar para permitir varias semánticas en problemas reales, para modificar o aprender granularidades, etc., y debería estudiarse.

Bibliografía

- [1] H. Schichl, “Models and history of modeling,” in *Modeling languages in mathematical optimization*, vol. 88 of *Applied Optimization*, ch. 2, Springer, 2004.
- [2] W. G. Knisel *et al.*, “CREAMS: A field-scale model for chemicals, runoff and erosion from agricultural management systems,” *USDA Conservation Research Report*, no. 26, 1980.
- [3] R. Clarke *et al.*, *Mathematical models in hydrology*. FAO, 1973.
- [4] J. H. Seinfeld and S. N. Pandis, *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons, 2012.
- [5] J. T. Houghton, *Climate change 1995: The science of climate change: contribution of working group I to the second assessment report of the Intergovernmental Panel on Climate Change*, vol. 2. Cambridge University Press, 1996.
- [6] K.-N. Liou, *Radiation and cloud processes in the atmosphere. Theory, observation, and modeling*. New York, NY (United States); Oxford University Press, 1992.
- [7] R. M. Granich, C. F. Gilks, C. Dye, K. M. De Cock, and B. G. Williams, “Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model,” *The Lancet*, vol. 373, no. 9657, pp. 48–57, 2009.
- [8] O. Diekmann, J. Heesterbeek, and J. A. Metz, “On the definition and the computation of the basic reproduction ratio r_0 in models for infectious

- diseases in heterogeneous populations,” *Journal of mathematical biology*, vol. 28, no. 4, pp. 365–382, 1990.
- [9] T. S. Parker, L. O. Chua, and T. S. Parker, *Practical numerical algorithms for chaotic systems*. Springer New York, 1989.
- [10] J. Casillas and F. J. Martínez-López, “Knowledge discovery by genetic fuzzy systems applied to consumer behavior modelling,” *Romanian Marketing Review*, vol. 3, pp. 111–142, 2007.
- [11] G. L. Lilien, P. Kotler, and K. S. Moorthy, *Marketing models*. Prentice-Hall Englewood Cliffs, 1992.
- [12] H. A. Simon, “A behavioral model of rational choice,” *The quarterly journal of economics*, vol. 69, no. 1, pp. 99–118, 1955.
- [13] H. A. Simon, *Models of man; social and rational*. Wiley, 1957.
- [14] T. T. W. Hertel and T. W. Hertel, *Global trade analysis: modeling and applications*. Cambridge university press, 1999.
- [15] P. D. Berger and N. I. Nasr, “Customer lifetime value: marketing models and applications,” *Journal of interactive marketing*, vol. 12, no. 1, pp. 17–30, 1998.
- [16] R. Frigg and S. Hartmann, “Models in science,” in *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.), <http://plato.stanford.edu/archives/fall2012/entries/models-science/>, Fall 2012 ed., 2012.
- [17] R. Hughes, “Models and representation,” *Philosophy of Science*, pp. S325–S336, 1997.
- [18] O. Nelles, *Nonlinear system identification: from classical approaches to neural networks and fuzzy models*. Springer, 2001.
- [19] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [20] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, third ed., January 2011.
- [21] J. Wyatt, “Nervous about artificial neural networks?,” *The Lancet*, vol. 346, no. 8984, pp. 1175 – 1177, 1995.

-
- [22] M. J. Araúzo-Bravo, J. M. Cano-Izquierdo, E. Gómez-Sánchez, M. J. Lopez-Nieto, Y. A. Dimitriadis, and J. Lopez-Coronado, “Automatization of a penicillin production process with soft sensors and an adaptive controller based on neuro fuzzy systems,” *Control Engineering Practice*, vol. 12, no. 9, pp. 1073 – 1090, 2004.
- [23] L. A. Zadeh, “The concept of a linguistic variable and its applications to approximate reasoning. Parts I, II and III,” *Information Sciences*, vol. 8, 8 and 9, pp. 199–249, 301–357 and 43–80, 1975.
- [24] G. I. Sainz, J. Juez, E. J. Moya, and J. R. Peran, “Fault detection and fuzzy rule extraction in AC motors by a neuro-fuzzy ART-based system,” *Engineering Applications of Artificial Intelligence*, vol. 18, pp. 867–874, 2005.
- [25] A. Konar, *Computational Intelligence: Principles, techniques and applications*. Springer-Verlag, Berlin, 2005.
- [26] F. O. Karray and C. d. De Silva, *Soft Computing and Intelligent Systems Design. Theory, Tools and Applications*. Addison Wesley, 2004.
- [27] P. P. Bonissoene, Y.-T. Chen, K. Goebel, and P. Khedkar, “Hybrid soft computing systems: industrial and commercial applications,” *Proceedings of the IEEE*, vol. 87, pp. 1641–1667, September 1999.
- [28] O. Cordon, F. Herrera, F. Hoffmann, and L. Magdalena, *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*, vol. 19 of *Advances in Fuzzy Systems - Applications and Theory*. World Scientific, 2001.
- [29] M. J. Gacto, R. Alcala, and F. Herrera, “Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures,” *Information Sciences*, vol. 181, pp. 4340 – 4360, 2011.
- [30] J. M. Alonso, L. Magdalena, and G. Gonzalez-Rodriguez, “Looking for a good fuzzy system interpretability index: An experimental approach,” *International Journal of Approximate Reasoning*, vol. 51, pp. 115 – 134, December 2009.
- [31] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena, eds., *Accuracy Improvements in Linguistic Fuzzy Modelling*, vol. 129 of *Studies in Fuzziness and Soft Computing*. Berlin Heidelberg: Springer-Verlag, 2003.

- [32] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, eds., *Interpretability Issues in Fuzzy Modeling*, vol. 128 of *Studies in Fuzziness and SoftComputing*. Berlin Heidelberg: Springer-Verlag, 2003.
- [33] F. Herrera, “Genetic fuzzy systems: Taxonomy, current research trends and prospects,” *Evolutionary Intelligence*, vol. 1, pp. 27 – 46, 2008.
- [34] E. H. Mamdani, “Application of fuzzy algorithms for control of simple dynamic plant,” *Electrical Engineers, Proceedings of the Institution of*, vol. 121, pp. 1585 – 1588, Dec. 1974.
- [35] R. Alcalá, J. Casillas, O. Cordón, and F. Herrera, “Building fuzzy graphs: Features and taxonomy of learning for non-grid-oriented fuzzy rulebased systems,” *Journal of Intelligent Fuzzy Systems*, vol. 11, pp. 99 – 119, 2001.
- [36] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-15, pp. 116–132, January/February 1985.
- [37] L. A. Zadeh, “Outline of a new approach to the analysis of complex systems and decision processes,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, pp. 28 – 44, January 1973.
- [38] J. M. Alonso and L. Magdalena, “Special issue on interpretable fuzzy systems,” *Information Science Volume*, vol. 181, Issue 20, pp. 4331 – 4822, 2011.
- [39] J. Casillas, F. Herrera, R. Pérez, M. J. d. Jesus, and P. Villar, “Special issue on genetic fuzzy systems and the interpretability-accuracy trade-off,” *International Journal of Approximate Reasoning*, vol. 44, pp. 1 – 90, 2007.
- [40] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, “Accuracy improvements to find the balance interpretability-accuracy in fuzzy modeling: An overview,” in *Accuracy Improvements in Linguistic Fuzzy Modelling* (J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, eds.), vol. 129 of *Studies in Fuzziness and SoftComputing*, pp. 3–24, Berlin Heidelberg: Springer-Verlag, 2003.
- [41] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, “Interpretability improvements to find the balance interpretability-accuracy in fuzzy modeling: An overview,” in *Interpretability Issues in Fuzzy Modelling* (J. Casillas,

- O. Cordón, F. Herrera, and L. Magdalena, eds.), vol. 128 of *Studies in Fuzziness and SoftComputing*, pp. 3–22, Berlin Heidelberg: Springer-Verlag, 2003.
- [42] J. d. Oliveira, “Semantic constraints for membership function optimization,” *IEEE Transactions on Systems, Man and Cybernetics. Part A: Systems and Humans*, vol. 29, pp. 128 – 138, January 1999.
- [43] H. Ishibuchi, T. Murata, and I. B. Türksen, “Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems,” *Fuzzy Sets and Systems*, vol. 89, pp. 135 – 150, July 1997.
- [44] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, “Selecting fuzzy if-then rules for classification problems using genetic algorithms,” *IEEE Transactions on Fuzzy Systems*, vol. 3, pp. 260 – 270, August 1995.
- [45] C. Mencar and A. Fanelli, “Interpretability constraints for fuzzy information granulation,” *Information Sciences*, vol. 178, pp. 4585 – 4618, December 2008.
- [46] S.-M. Zhou and J. Q. Gan, “Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling,” *Fuzzy Sets and Systems*, vol. 159, pp. 3091 – 3131, 2008.
- [47] H. Ishibuchi and T. Yamamoto, “Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining,” *Fuzzy Sets and Systems*, vol. 141, pp. 59 – 88, January 2004.
- [48] E. Trillas, A. R. de Soto, and S. Cubillo, “A glance at implication and t-conditional functions,” in *Discovering the world with fuzzy logic* (V. Novák and I. Perfilieva, eds.), pp. 126–147, Heidelberg, Germany, Germany: Physica-Verlag GmbH, 2000.
- [49] O. Cordón, F. Herrera, and A. Peregrín, “Applicability of the fuzzy operators in the design of fuzzy logic controllers,” *Fuzzy Sets and Systems*, pp. 15 – 41, 1997.
- [50] M. Fazzolari, R. Alcalá, Y. Nojima, H. Ishibuchi, and F. Herrera, “A review of the application of multiobjective evolutionary fuzzy systems: Current status and further directions,” *IEEE Transactions on Fuzzy Systems*, vol. 21, pp. 45 – 65, 2013.

-
- [51] O. Cordon, “A historical review of evolutionary learning methods for mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems,” *International Journal of Approximate Reasoning*, vol. 52, pp. 894 – 913, 2011.
- [52] F. Herrera, “Genetic fuzzy systems: Status, critical considerations and future directions,” *International Journal of Computational Intelligence Research*, vol. 1, pp. 59 – 67, 2005.
- [53] O. Cordon, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena, “Ten years of genetic fuzzy systems: current framework and new trends,” *Fuzzy sets and systems*, vol. 141, no. 1, pp. 5–31, 2004.
- [54] R. Alcalá, Y. Nojima, H. Ishibuchi, and F. Herrera, “Special issue on evolutionary fuzzy systems,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 20, pp. v – x, October 2012.
- [55] R. Alcalá, Y. Nojima, H. Ishibuchi, and F. Herrera, “Special issue on evolutionary fuzzy systems,” *International Journal of Computational Intelligence Systems*, vol. 5, pp. 209 – 211, April 2012.
- [56] Y. Nojima, R. Alcalá, H. Ishibuchi, and F. Herrera, “Special issue on evolutionary fuzzy systems,” *Soft Computing*, vol. 15, pp. 2299–2301, December 2011.
- [57] R. Alcalá and Y. Nojima, “Special issue on genetic fuzzy systems: New advances,” *Evolutionary Intelligence*, vol. 2, pp. 1 – 3, April 2009.
- [58] J. Casillas and B. Carse, “Special issue on genetic fuzzy systems: recent developments and future directions,” *Soft Computing*, vol. 13: 5, pp. 417 – 418, 2009.
- [59] O. Cordon, R. Alcalá, J. Alcalá-Fdez, and I. Rojas, “Special issue on genetic fuzzy systems: What’s next?,” *IEEE Transactions on Fuzzy Systems*, vol. 15: 4, pp. 533 – 535, April 2007.
- [60] B. Carse and A. G. Pipe, “Special issue on genetic fuzzy systems,” *International Journal of Intelligent Systems*, vol. 22: 9, pp. 905 – 907, 2007.
- [61] J. M. Alonso and L. Magdalena, “HILK++: an interpretability-guided fuzzy modeling methodology for learning readable and comprehensible fuzzy rule-based classifiers,” *Soft Computing*, vol. 15, pp. 1959 – 1980, 2011.

-
- [62] C. Mencar, C. Castiello, R. Cannone, and A. Fanelli, “Interpretability assessment of fuzzy knowledge bases: A cointension based approach,” *International Journal of Approximate Reasoning*, vol. 52, pp. 501 – 518, 2011.
- [63] R. Mikut, J. Jäkel, and L. Gröll, “Interpretability issues in data-based learning of fuzzy systems,” *Fuzzy Sets and Systems*, vol. 150, pp. 179 – 197, March 2005.
- [64] M. Setnes, “Simplification and reduction of fuzzy rules,” in *Interpretability Issues in Fuzzy Modelling* (J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, eds.), vol. 128 of *Studies in Fuzziness and SoftComputing*, pp. 278–302, Berlin Heidelberg: Springer-Verlag, 2003.
- [65] S. Guillaume and B. Charnomordic, “A new method for inducing a set of interpretable fuzzy partitions and fuzzy inference systems from data,” in *Interpretability Issues in Fuzzy Modelling* (J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, eds.), vol. 128 of *Studies in Fuzziness and SoftComputing*, pp. 148–175, Berlin Heidelberg: Springer-Verlag, 2003.
- [66] L. A. Zadeh, “Fuzzy sets,” *Information and Control*, vol. 8, pp. 338–353, 1965.
- [67] E. H. Mamdani and S. Assilian, “An experiment in linguistic synthesis with a fuzzy logic controller,” *Int. J. Mach. Studies*, vol. 7, no. 1, pp. 1 – 13, 1975.
- [68] L. P. Holmblad and J.-J. Østergaard, *Control of a cement kiln by fuzzy logic*. Fuzzy Information and Decision Process, North Holland, Amsterdam: Smidth, 1982.
- [69] P. Pulkkinen and H. Koivisto, “A dynamically constrained multiobjective genetic fuzzy system for regression problems,” *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 1, pp. 161 – 177, 2010.
- [70] M. J. Gacto, R. Alcalá, and F. Herrera, “Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems,” *Soft Computing*, vol. 13, pp. 419 – 436, March 2009.
- [71] R. Alcalá, M. J. Gacto, F. Herrera, and J. Alcalá-Fdez, “A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and com-

- pact linguistic fuzzy rule-based systems,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 15, no. 5, pp. 539 – 557, 2007.
- [72] R. Alcalá, J. Alcalá-Fdez, and F. Herrera, “A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection,” *IEEE Transactions on Fuzzy Systems*, vol. 15, pp. 616 – 635, August 2007.
- [73] R. Alcalá, J. Alcalá-Fdez, J. Casillas, O. Cordón, and F. Herrera, “Hybrid learning models to get the interpretability-accuracy trade-off in fuzzy modeling,” *Soft Computing*, vol. 10, no. 9, pp. 717 – 734, 2006.
- [74] J. Espinosa and J. Vandewalle, “Constructing fuzzy models with linguistic integrity from numerical data-AFRELI algorithm,” *IEEE Transactions on Fuzzy Systems*, vol. 8, pp. 591 – 600, October 2000.
- [75] Y. Jin, W. Von Seelen, and B. Sendhoff, “On generating FC^3 fuzzy rule systems from data using evolution strategies,” *IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics*, vol. 29, pp. 829–845, December 1999.
- [76] A. A. Márquez, F. A. Márquez, and A. Peregrín, “A mechanism to improve the interpretability of linguistic fuzzy systems with adaptive defuzzification based on the use of a multi-objective evolutionary algorithm,” *International Journal of Computational Intelligence Systems*, vol. 5, pp. 297 – 321, April 2012.
- [77] M. J. Gacto, R. Alcalá, and F. Herrera, “Integration of an index to preserve the semantic interpretability in the multi-objective evolutionary rule selection and tuning of linguistic fuzzy systems,” *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 3, pp. 515 – 531, 2010.
- [78] J. Casillas, O. Cordón, M. J. d. Jesus, and F. Herrera, “Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction,” *IEEE Transactions on Fuzzy Systems*, vol. 13, pp. 13 – 29, February 2005.
- [79] J. Chen and M. Mahfouf, “Improving transparency in approximate fuzzy modeling using multi-objective immune-inspired optimisation,” *International Journal of Computational Intelligence Systems*, vol. 4, pp. 322 – 342, April 2012.

-
- [80] O. Guenounou, A. Belmehdi, and B. Dahhou, “Multi-objective optimization of TSK fuzzy models,” *Expert Systems with Applications*, vol. 36, pp. 7416 – 7423, 2009.
- [81] L. Yu and J. Xiao, “Trade-off between accuracy and interpretability: Experience-oriented fuzzy modeling via reduced-set vectors,” *Computers and Mathematics with Applications*, vol. 57, pp. 885 – 895, 2009.
- [82] P. Pulkkinen and H. Koivisto, “Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms,” *International Journal of Approximate Reasoning*, vol. 48, no. 2, pp. 526 – 543, 2008.
- [83] P. Pulkkinen, J. Hytönen, and H. Koivisto, “Developing a bioaerosol detector using hybrid genetic fuzzy systems,” *Engineering Applications of Artificial Intelligence*, vol. 21, no. 8, pp. 1330–1346, 2008.
- [84] S. Destercke, S. Guillaume, and B. Charnomordic, “Building an interpretable fuzzy rule base from data using orthogonal least squares - Application to a depollution problem,” *Fuzzy Sets and Systems*, vol. 158, pp. 2078 – 2094, September 2007.
- [85] H. Wang, S. Kwong, Y. Jin, W. Wei, and K. F. Man, “Multi-objective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction,” *Fuzzy Sets and Systems*, vol. 149, pp. 149–186, January 2005.
- [86] M. Y. Chen and D. A. Linkens, “Rule-base self-generation and simplification for data-driven fuzzy models,” *Fuzzy Sets and Systems*, vol. 142, pp. 243–265, March 2004.
- [87] R. P. Paiva and A. Dourado, “Interpretability and learning in neuro-fuzzy systems,” *Fuzzy Sets and Systems*, vol. 147, pp. 17 – 38, 2004.
- [88] H. Roubos and M. Setnes, “Compact and transparent fuzzy models and classifiers through iterative complexity reduction,” *IEEE Transactions on Fuzzy Systems*, vol. 9, pp. 516–524, August 2001.
- [89] M. Setnes, R. Babuška, U. Kaymak, and H. van Nauta Lemke, “Similarity measures in fuzzy rule base simplification,” *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, vol. 28, pp. 376 – 386, June 1998.

- [90] M. Antonelli, P. Ducange, B. Lazzerini, and F. Marcelloni, “Learning knowledge bases of multi-objective evolutionary fuzzy systems by simultaneously optimizing accuracy, complexity and partition integrity,” *Soft Computing*, vol. 15, pp. 2335 – 2354, December 2011.
- [91] C. Mencar, C. Castiello, R. Cannone, and A. Fanelli, “Design of fuzzy rule-based classifiers with semantic cointension,” *Information Science*, vol. 181:20, pp. 4361 – 4377, 2011.
- [92] R. Alcalá, P. Ducange, F. Herrera, B. Lazzerini, and F. Marcelloni, “A multiobjective evolutionary approach to concurrently learn rule and data bases of linguistic fuzzy-rule-based systems,” *IEEE Transactions on Fuzzy Systems*, vol. 17, pp. 1106 – 1122, October 2009.
- [93] J. Casillas, P. Martínez, and A. D. Benítez, “Learning consistent, complete and compact sets of fuzzy rules in conjunctive normal form for regression problems,” *Soft Computing*, vol. 13, pp. 451 – 465, 2009.
- [94] M. Antonelli, P. Ducange, B. Lazzerini, and F. Marcelloni, “Learning concurrently partition granularities and rules bases of mandani fuzzy systems in a multi-objective evolutionary framework,” *International Journal of Approximate Reasoning*, vol. 50, pp. 1066 – 1080, 2009.
- [95] M. Antonelli, P. Ducange, B. Lazzerini, and F. Marcelloni, “Multi-objective evolutionary learning of granularity, membership function parameters and rules of mamdani fuzzy systems,” *Evolutionary Intelligence*, vol. 2, pp. 21 – 37, 2009.
- [96] J. M. Alonso, L. Magdalena, and S. Guillaume, “HILK: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism,” *International Journal of Intelligent Systems*, vol. 23, no. 7, pp. 761 – 794, 2008.
- [97] R. Alcalá, J. Alcalá-Fdez, F. Herrera, and J. Otero, “Genetic learning of accurate and compact fuzzy rule based systems based on the 2-tuples linguistic representation,” *International Journal of Approximate Reasoning. Special Issue on Genetic Fuzzy Systems and the Interpretability-Accuracy Trade-off*, vol. 44, pp. 45 – 64, 2007.
- [98] H. Ishibuchi and Y. Nojima, “Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning,”

- International Journal of Approximate Reasoning. Special Issue on Genetic Fuzzy Systems and the Interpretability-Accuracy Trade-off*, vol. 44, pp. 4 – 31, January 2007.
- [99] M. Cococcioni, P. Ducange, B. Lazzerini, and F. Marcelloni, “A pareto-based multi-objective evolutionary approach to the identification of mamdani fuzzy systems,” *Soft Computing*, vol. 11, pp. 1013 – 1031, 2007.
- [100] M. Delgado, F. Von Zuben, and F. Gomide, “Hierarchical genetic fuzzy systems: Accuracy, interpretability and design autonomy,” in *Interpretability Issues in Fuzzy Modelling* (J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, eds.), vol. 128 of *Studies in Fuzziness and SoftComputing*, pp. 379–405, Berlin Heidelberg: Springer-Verlag, 2003.
- [101] A. Fiordaliso, “About the trade-off between accuracy and interpretability of takagi-sugeno models in the context of nonlinear time series forecasting,” in *Interpretability Issues in Fuzzy Modelling* (J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, eds.), vol. 128 of *Studies in Fuzziness and SoftComputing*, pp. 406–430, Berlin Heidelberg: Springer-Verlag, 2003.
- [102] F. Jimenez, A. F. Gómez-Skarmeta, G. Sanchez, H. Roubos, and R. Babuška, “Accurate, transparent and compact fuzzy models by multi-objective evolutionary algorithms,” in *Interpretability Issues in Fuzzy Modelling* (J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, eds.), vol. 128 of *Studies in Fuzziness and SoftComputing*, pp. 431–451, Berlin Heidelberg: Springer-Verlag, 2003.
- [103] O. Cordón, M. J. del Jesus, F. Herrera, L. Magdalena, and P. Villar, “A multiobjective genetic learning process for joint feature selection and granularity and context learning in fuzzy rule-based classification systems,” in *Interpretability issues in fuzzy modeling* (J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, eds.), pp. 79–99, Springer-Verlag, 2003.
- [104] T. Suzuki and T. Furuhashi, “Conciseness of fuzzy models,” in *Interpretability Issues in Fuzzy Modelling* (J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, eds.), vol. 128 of *Studies in Fuzziness and SoftComputing*, pp. 569–586, Berlin Heidelberg: Springer-Verlag, 2003.
- [105] O. Cordón, F. Herrera, L. Magdalena, and P. Villar, “A genetic learning process for the scaling factors, granularity and contexts of the fuzzy rule-based system data base,” *Information Science*, vol. 136, pp. 85–107, 2001.

- [106] H. Ishibuchi, T. Nakashima, and T. Murata, “Three-objective genetics-based machine learning for linguistic rule extraction,” *Information Sciences*, vol. 136, pp. 109 – 133, August 2001.
- [107] U. Bodenhofer and P. Bauer, “A formal model of interpretability of linguistic variables,” in *Interpretability Issues in Fuzzy Modelling* (J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, eds.), vol. 128 of *Studies in Fuzziness and SoftComputing*, pp. 524–545, Berlin Heidelberg: Springer-Verlag, 2003.
- [108] L. A. Zadeh, “Is there a need for fuzzy logic?,” *Information Sciences*, vol. 178, pp. 2751 – 2779, 2008.
- [109] A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu, “Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index,” *Soft Computing*, vol. 13, no. 5, pp. 437 – 449, 2009.
- [110] G. A. Miller, “The magical number seven, plus or minus two: Some limits on our capacity for processing information,” *The Psychological Review*, vol. 63, no. 2, pp. 81 – 97, 1956.
- [111] Y. Jin, “Fuzzy Modeling of High-Dimensional Systems: Complexity Reduction and Interpretability Improvement,” *IEEE Transactions on Fuzzy Systems*, vol. 8, pp. 212–221, April 2000.
- [112] P. Fazendeiro, J. V. de Oliveira, and W. Pedrycz, “A multiobjective design of a patient and anaesthetist-friendly neuromuscular blockade controller,” *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 9, pp. 1667–1678, 2007.
- [113] W. Pedrycz, “Expressing relevance interpretability and accuracy of rule-based systems,” in *Interpretability Issues in Fuzzy Modelling* (J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, eds.), vol. 128 of *Studies in Fuzziness and SoftComputing*, pp. 547–567, Berlin Heidelberg: Springer-Verlag, 2003.
- [114] M. Setnes and R. Babuška, “Rule base reduction: Some comments on the use of orthogonal transforms,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 31, pp. 199 – 206, May 2001.

-
- [115] J. Yen and L. Wang, "Simplifying fuzzy rule-based models using orthogonal transformation methods," *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, vol. 29, pp. 13–24, February 1999.
- [116] D. P. Pancho, J. M. Alonso, O. Cordon, A. Quirin, and L. Magdalena, "FINGRAMS: Visual representations of fuzzy rule-based inference for expert analysis of comprehensibility," *IEEE Transactions on Fuzzy Systems*, 2013.
- [117] O. Cordon, F. Herrera, and P. Villar, "Generating the knowledge base of a fuzzy rule-based system by the genetic learning of data base," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 667–674, 2001.
- [118] C. Karr, "Genetic algorithms for fuzzy controllers," *AI Expert*, vol. 6, no. 2, pp. 26 – 33, 1991.
- [119] O. Cordon and F. Herrera, "A three-stage evolutionary process for learning descriptive and approximate fuzzy logic controller knowledge bases from examples," *International Journal of Approximate Reasoning*, vol. 17, no. 4, pp. 369–407, 1997.
- [120] H. B. Gürocak, "A genetic-algorithm-based method for tuning fuzzy logic controllers," *Fuzzy Sets and Systems*, vol. 108, no. 1, pp. 39–47, 1999.
- [121] F. Herrera, M. Lozano, and J. L. Verdegay, "Tuning fuzzy logic controllers by genetic algorithms," *International Journal of Approximate Reasoning*, vol. 12, pp. 299–315, 1995.
- [122] D. Nauck and R. Kruse, "Neuro-fuzzy systems for function approximation," *Fuzzy Sets and Systems*, vol. 101, pp. 261–271, January 1999.
- [123] J. M. Cano Izquierdo, Y. A. Dimitriadis, E. Gómez Sánchez, and J. López Coronado, "Learning from noisy information in FasArt and Fas-back neuro-fuzzy systems," *Neural Networks*, vol. 14, pp. 407–425, May 2001.
- [124] G. I. Sainz, Y. Dimitriadis, J. Cano, E. Gómez, and E. Parrado, "ART based model set for pattern recognition: FasArt family," in *Neuro-fuzzy pattern recognition* (H. Bunke and A. Kandel, eds.), ch. 1, pp. 147–177, World Scientific Pub. Co., December 2000.

- [125] G. I. Sainz, M. Fuente, and P. Vega, "Recurrent neuro-fuzzy modelling of a wastewater treatment plant," *European Journal of Control*, vol. 10, pp. 83–95, 2004.
- [126] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multi-objective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, pp. 182–197, April 2002.
- [127] J. Alcalá-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera, "KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework," *Journal of Multiple-Valued Logic and Soft Computing*, vol. 17:2-3, pp. 255–287, 2011.
- [128] J. Alcalá-Fdez, L. Sánchez, S. García, M. J. del Jesus, S. Ventura, J. M. Garrell, J. Otero, C. Romero, J. Bacardit, V. M. Rivas, J. C. Fernández, and F. Herrera, "KEEL: a software tool to assess evolutionary algorithms for data mining problems," *Soft Computing*, vol. 13, pp. 307 – 318, February 2009.
- [129] D. J. Sheskin, *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall/CRC, 2003.
- [130] J. Zar, *Biostatistical Analysis*. Prentice Hall, 1999.
- [131] S. García, A. Fernández, J. Luengo, and F. Herrera, "A study of statistical techniques and performance measures for genetics-based machine learning: Accuracy and interpretability," *Soft Computing*, vol. 13, no. 10, pp. 959–977, 2009.
- [132] S. García, D. Molina, M. Lozano, and F. Herrera, "A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: A case study on the CEC'2005 special session on real parameter optimization," *Journal of Heuristics*, vol. 15, pp. 617 – 644, 2009.
- [133] S. García and F. Herrera, "An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons," *Journal of Machine Learning Research*, vol. 9, pp. 2677 – 2694, 2008.
- [134] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

-
- [135] L.-X. Wang and J. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, pp. 1414–1427, November/December 1992.
- [136] A. Riid and E. Rüstern, "Adaptability, interpretability and rule weights in fuzzy rule-based systems," *Information Sciences*, vol. in press, 2013.
- [137] O. Cordón and F. Herrera, "Hybridizing genetic algorithms with sharing scheme and evolution strategies for designing approximate fuzzy rule-based systems," *Fuzzy Sets and Systems*, vol. 118, pp. 235 – 255, 2001.
- [138] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization," in *Proc. Evolutionary Methods for Design, Optimization and Control with App. to Industrial Problems*, (Barcelona, Spain), pp. 95–100, 2001.
- [139] L. J. Eshelman and J. D. Schaffer, "Real-coded genetic algorithms and interval-schemata," *Foundations of Genetic Algorithms*, vol. 2, pp. 187–202, 1993.
- [140] L. J. Eshelman, "The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination," in *Foundations of genetic Algorithms* (G. Rawlin, ed.), vol. 1, pp. 265–283, Morgan Kaufman, 1991.
- [141] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association*, vol. 32, pp. 675–701, 1937.
- [142] R. L. Iman and J. H. Davenport, "Approximations of the critical region of the friedman statistic," *Comm. Statist. Part A Theory Methods*, vol. 9, pp. 571–595, 1980.
- [143] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian journal of statistics*, vol. 6, pp. 65–70, 1979.
- [144] O. Cordón, F. Herrera, and L. Sánchez, "Solving electrical distribution problems using hybrid evolutionary data analysis techniques," *Applied Intelligence*, vol. 10, pp. 5–24, 1999.

- [145] M. Galende, G. I. Sainz, and M. J. Fuente, “Complexity reduction and interpretability improvement for fuzzy rule systems based on simple interpretability measures and indices by bi-objective evolutionary rule selection,” *Soft Computing*, vol. 16, pp. 451 – 470, March 2012.
- [146] M. I. Rey, M. Galende, M. Fuente, and G. I. Sainz, “Checking orthogonal transformations and genetic algorithms for selection of fuzzy rules based on interpretability-accuracy concepts,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 20, pp. 159 – 186, October 2012.
- [147] M. Galende, M. J. Gacto, G. Sainz, and R. Alcalá, “On the interpretability of linguistic and scatter FRBSs: GM3M generalization and new rule meaning index for global assessment and local pseudo-linguistic representation,” *Information Sciences*, Enviado. 2013.
- [148] M. J. Gacto, M. Galende, R. Alcalá, and F. Herrera, “METSK-HD^e: A multiobjective evolutionary algorithm to learn accurate TSK-fuzzy systems in high-dimensional and large-scale regression problems,” *Information Sciences*, En prensa. 2013.
- [149] M. I. Rey, M. Galende, G. I. Sainz, and M. J. Fuente, “Selection of rules by orthogonal transformations and genetic algorithms to improve the interpretability in fuzzy rule based systems,” in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2013)*, (Hyderabad, India), July 2013.
- [150] M. J. Gacto, M. Galende, R. Alcalá, and F. Herrera, “Obtaining accurate TSK fuzzy rule-based systems by multi-objective evolutionary learning in high-dimensional regression problems,” in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2013)*, (Hyderabad, India), July 2013.
- [151] M. Galende, G. I. Sainz, and M. J. Fuente, “Accuracy-interpretability trade-off for precise fuzzy modeling using simple indices. Application to industrial plants,” in *Proceedings of the 18th IFAC World Congress*, (Milan, Italy), pp. 12656 – 12661, August 28 - September 2 2011.
- [152] M. I. Rey, M. Galende, G. I. Sainz, and M. J. Fuente, “Checking orthogonal transformations and genetic algorithms for selection of fuzzy rules based on interpretability-accuracy concepts,” in *Proceedings of 2011 IEEE International Conference on Fuzzy Systems (Fuzz-IEEE)*, (Taipei, Taiwan), pp. 1271 – 1278, 27 - 30 June 2011.

- [153] M. Galende, G. I. Sainz, and M. J. Fuente, “Accuracy-interpretability balancing in fuzzy models based on multiobjective genetic algorithm,” in *Proceedings of European Control Conference 2009 (ECC’09)*, (Budapest, Hungary), pp. 3915 – 3920, 23 – 26 August 2009.
- [154] M. I. Rey, M. Galende, and G. I. Sainz, “Criteria for linguistic improvement of precise fuzzy models by orthogonal transforms. Application to art based models,” in *Proceedings of European Control Conference 2009 (ECC’09)*, (Budapest, Hungary), pp. 348 – 353, 23 – 26 August 2009.
- [155] M. Galende, G. I. Sainz, M. J. Fuente, and A. Herreros, “Interpretability-accuracy improvement in a neuro-fuzzy ART based model of a DC motor,” in *Proceedings of the 17th IFAC World Congress*, (Seoul, Korea), pp. 7034 – 7039, 6-11 July 2008.
- [156] M. Galende and G. I. Sainz, “Mejora del equilibrio precisión-interpretabilidad en modelos difusos,” in *XV Congreso Español Sobre Tecnologías y Lógica Fuzzy (ESTYLF2010)*, (Huelva), pp. 79 – 84, 3-5 February 2010.
- [157] M. I. Rey, M. Galende, and G. Sainz, “Criterios para el empleo de transformaciones ortogonales en la mejora de modelos difusos. Aplicación a modelos basados en ART,” in *XV Congreso Español Sobre Tecnologías y Lógica Fuzzy (ESTYLF2010)*, (Huelva), pp. 91 – 96, 3-5 February 2010.
- [158] M. Galende and G. I. Sainz, “Mejora Lingüística de Modelos Neurodifusos mediante Algoritmos Genéticos: Aplicación a un Motor DC.,” in *Proceedings of the XXVIII Jornadas de Automáticas*, (Huelva, Spain), September 2007.
- [159] M. M. Gupta and J. Qi, “Design of fuzzy logic controllers based on generalized t-operators,” *Fuzzy Sets and Systems*, vol. 40, pp. 473–489, 1991.
- [160] E. Trillas and L. Valverde, “On implication and indistinguishability in the setting of fuzzy logic,” in *Management Decision Support Systems Using Fuzzy Logic and Possibility Theory* (J. Kacprzyk and R. Yager, eds.), pp. 198–212, Verlag TUV Rheinland, 1985.
- [161] Z. Cao and A. Kandel, “Applicability of some fuzzy implication operators,” *Fuzzy Sets and Systems*, vol. 31, pp. 151–186, 1989.

- [162] J. Kiszka, M. Kochanska, and D. Sliwinska, "The influence of some fuzzy implication operators on the accuracy of a fuzzy model," *Partes I y II Fuzzy Sets and Systems*, vol. 15,15, pp. 111–128,223–240, 1985.
- [163] H. Hellendoorn and C. Thomas, "Defuzzification in fuzzy controllers," *Journal of Intelligent Fuzzy Systems*, vol. 1, pp. 109–123, 1993.
- [164] M. Sugeno and T. Yasukawa, "A fuzzy-logic-based approach to qualitative modeling," *IEEE Transactions on Fuzzy Systems*, vol. 1, pp. 7–31, February 1993.
- [165] J. H. Holland, *Adaptation in natural and artificial systems*. The University of Michigan Press, 1975.
- [166] D. E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, 1989.
- [167] Z. Michalewicz, *Genetic algorithms + data structures = evolution programs*. Springer-Verlag, 1996.
- [168] V. Pareto, *Cours d'economie Politique*. F. Rouge, 1896.
- [169] C. A. Coello Coello, "A Short Tutorial on Evolutionary Multiobjective Optimization," in *First International Conference on Evolutionary Multi-Criterion Optimization* (E. Zitzler, K. Deb, L. Thiele, C. A. C. Coello, and D. Corne, eds.), pp. 21–40, Springer-Verlag. Lecture Notes in Computer Science No. 1993, 2001.
- [170] N. Srinivas and K. Deb, "Multiobjective optimization using nondominated sorting in genetic algorithms," *Evolutionary Computation*, vol. 2, pp. 221–248, 1994.
- [171] J. Horn, N. Nafpliotis, and D. Goldberg, "A niched pareto genetic algorithm for multiobjective optimization," in *Proc. First IEEE Conf. on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, (Piscataway, NJ), pp. 82–87, 1994.
- [172] C. Fonseca and P. Fleming, "Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization," in *Proc. 5th Int. Conf. on Genetic Algorithms*, (San Mateo, CA), pp. 416–423, 1993.

-
- [173] E. Zitzler and L. Thiele, “Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach,” *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 257–271, 1999.

Apéndices

Apéndice A

Lógica Difusa y Sistemas Basados en Reglas Difusas

Este apéndice está dedicado a presentar, de forma muy breve, los componentes principales de los Sistemas Basados en Reglas Difusas, parte fundamental del desarrollo de esta tesis, así como algunos conceptos fundamentales muy utilizados al hablar de Lógica Difusa.

A.1. Fundamentos de la Lógica Difusa

La Lógica Difusa fue propuesta por Zadeh en 1965 como una extensión de la lógica booleana con el objetivo de poder representar información vaga o imprecisa [66]. A continuación se recuerdan las definiciones de algunos conceptos básicos de esta lógica:

- **Conjunto Difuso.** Conjunto caracterizado porque los elementos del universo de discurso en el que está definido pueden pertenecer a él en un cierto grado de pertenencia [66].
- **Partición Difusa.** Serie de conjuntos difusos definidos sobre una misma variable.

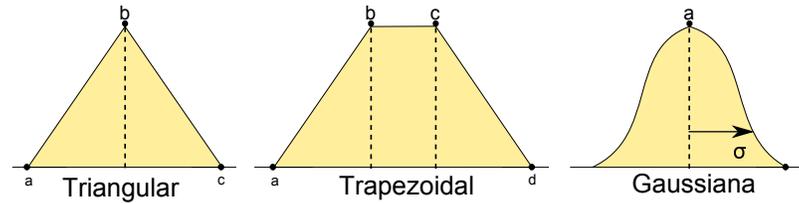


Figura A.1: Funciones de pertenencia más utilizadas.

- **Variable Difusa.** Variable definida sobre un universo de discurso y representada por una partición difusa.
- **Función de Pertenencia.** Función que define un conjunto difuso en el que cada elemento del universo de discurso tiene asignado un grado de pertenencia al conjunto en el rango $[0, 1]$. Un valor 0 indica que el elemento no está en el conjunto, mientras que un valor 1 indica justamente lo contrario. Las funciones de pertenencia puede tener cualquier forma, aunque las más utilizadas son las triangulares, las trapezoidales y las gaussianas (Fig. A.1).
- **Operadores lógicos difusos.** Operadores de la lógica booleana, como la unión, la intersección o la negación, definidos sobre las funciones de pertenencia de los conjuntos difusos. Los más utilizados son:

- *Unión*, que se implementa normalmente a través de *s-norma* del máximo:

$$ASB = \text{máx}(A, B) = A \vee B \quad (\text{A.1})$$

- *Intersección*, implementada a través de la *t-norma* del mínimo:

$$ATB = \text{mín}(A, B) = A \wedge B \quad (\text{A.2})$$

- **Variable Lingüística.** Variable cuyos valores son términos del lenguaje natural representados mediante conjuntos difusos caracterizados por funciones de pertenencia definidas sobre un universo de discurso [23].
- **Regla Difusa.** Regla de tipo SI-ENTONES que combina uno o más conjuntos difusos de entrada, antecedentes o premisas de la regla, y le infiere un conjunto difuso de salida, consecuencia o consecuente. Los conjuntos difusos de entrada se asocian mediante operadores lógicos difusos.

En función de como sea el consecuente se habla de distintos tipos de reglas:

- *Mamdani o lingüísticas*, donde el consecuente de la regla es un conjunto difuso que corresponde con una variable lingüística [34]:

SI x_1 es $A_1 \wedge x_2$ es $A_2 \wedge \dots \wedge x_n$ es A_n ENTONCES y es B

o una clasificación:

SI x_1 es $A_1 \wedge x_2$ es $A_2 \wedge \dots \wedge x_n$ es A_n ENTONCES y es C_k en grado w_k

siendo x_1, \dots, x_n las variables difusas de entrada, y la variable difusa de salida y A_1, \dots, A_n, B las funciones de pertenencia asociadas a las variables de entrada y salida respectivamente.

- *Aproximativas o scatter*, donde el consecuente de la regla es un conjunto difuso que presenta su propia semántica, al igual que todos los demás conjuntos difusos que componen la regla [35]. Se representan como:

SI x_1 es $\hat{A}_1 \wedge x_2$ es $\hat{A}_2 \wedge \dots \wedge x_n$ es \hat{A}_n ENTONCES y es \hat{G}

siendo x_1, \dots, x_n las variables difusas de entrada, y la variable difusa de salida y $\hat{A}_1, \dots, \hat{A}_n, \hat{G}$ las funciones de pertenencia asociadas a las variables de entrada y salida respectivamente.

- *Takagi-Sugeno-Kang o TSK* donde el consecuente de la regla es un valor numérico función de la entrada [36]. Se representan como:

SI x_1 es $F_1 \wedge x_2$ es $F_2 \wedge \dots \wedge x_n$ es F_n ENTONCES $y = f(x)$

siendo x_1, \dots, x_n las variables difusas de entrada, y la variable difusa de salida, F_1, \dots, F_n las funciones de pertenencia asociadas a las variables de entrada y f la función que relaciona las entradas y la salida.

A.2. Sistemas basados en reglas difusas

Un Sistema Basado en Reglas Difusas (SBRD) es un sistema que ante una entrada precisa o difusa, empleando un sistema de inferencia basado en un razonamiento difuso y teniendo en cuenta la información contenida en una base

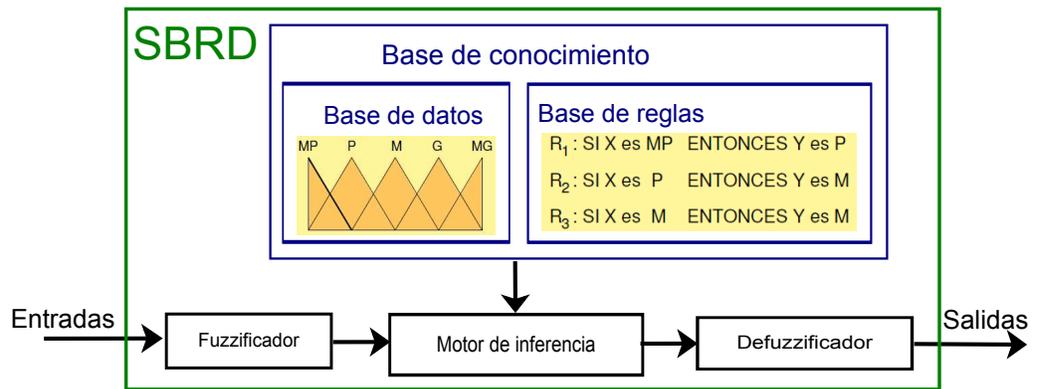


Figura A.2: Sistema Basado en Reglas Difusas.

de conocimiento, es capaz proporcionar una salida precisa o difusa. La estructura básica de un SBRD es la mostrada en la Fig. A.2, siendo sus componentes principales los siguientes:

- **Base de Conocimiento (BC).** Es la parte principal de un SBRD y la que contiene la colección de reglas difusas que guían el comportamiento del sistema. Esta formada a su vez por otros dos componentes:
 - **Base de Datos (BD).** Es la parte que almacena la estructura de los conjuntos difusos que describen las variables contenidas en las reglas.
 - **Base de Reglas (BR).** Es la parte que almacena las reglas difusas. Todas las reglas están unidas por el operador ADEMÁS, por lo que ante una misma entrada pueden dispararse múltiples reglas. Esto da lugar a la aparición de:
 - *Reglas Redundantes.* Son reglas cuyas acciones ya esta cubiertas por otras reglas. Se caracteriza por tener antecedentes y consecuentes similares, de forma que ante una misma entrada generan salidas similares.
 - *Reglas Incoherentes:* Son reglas que distorsionan el funcionamiento del sistema. Se caracterizan por tener antecedentes similares pero distintos consecuentes, de forma que ante una misma entrada generan distintas salidas.

- **Interfaz de Fuzzificación**¹ que convierte las entradas no difusas al sistema en valores difusos que puedan ser empleados como antecedentes de las reglas.
- **Motor de Inferencia** capaz de procesar la información de entrada (conocimiento difuso previo) y dar una salida (nuevo conocimiento difuso). Su funcionamiento está basado en la aplicación del *modus ponens generalizado* propuesto por Zadeh en 1973 [37]:

$$\begin{array}{l}
 \text{REGLA :} \quad \quad \quad \text{SI } X \text{ es } A \text{ ENTONCES } Y \text{ es } B \\
 \text{PREMISA :} \quad \quad \quad X \text{ es } A' \\
 \hline
 \text{CONCLUSION :} \quad Y \text{ es } B'
 \end{array}$$

El mecanismo de inferencia requiere de la elección de los diversos operadores que intervienen en el proceso:

- Operador para realizar la intersección entre los conjuntos difusos de los antecedentes, normalmente una *t-norma* [159, 160].
 - Operador de implicación difusa, correspondiente al condicional ENTONCES. Para ello se suelen usar *t-normas* [34, 159, 160] o *t-conormas* [49, 161, 162].
 - Operador de agregación ADEMÁS que, junto con el método de defuzzificación combina las salidas individuales de múltiples reglas en una salida final. Los más utilizados son la *media*, la *media ponderada* o la selección de algún *valor característico* de los conjuntos difusos en función del grado de importancia de la regla que los ha generado en el proceso de inferencia [49].
- **Interfaz de Defuzzificación**² que convierte las conclusiones difusas obtenidas en salidas no difusas del sistema. El operador más empleado es la *media ponderada por el grado de emparejamiento*, que se suele combinar con el *centro de gravedad* como valor característico del conjunto difuso [49, 163, 164].

¹El término ‘fuzzificación’ es una traducción directa del vocablo inglés *fuzzification* cuyo uso en la comunidad científica esta ampliamente extendido.

²El término ‘defuzzificación’ es una traducción directa del vocablo inglés *defuzzification* cuyo uso en la comunidad científica esta ampliamente extendido.

Apéndice B

Optimización Genética Multi-Objetivo

Puesto que los métodos de optimización propuestos en esta memoria de tesis están basados en el uso de Algoritmos Genéticos Multi-Objetivo, este apéndice está dedicado a describir, en líneas generales, algunos fundamentos teóricos de este tipo de algoritmos.

B.1. Algoritmos Genéticos

Los Algoritmos Genéticos (AGs) son una familia de modelos computacionales que imita a la evolución biológica como estrategia para resolver problemas. Fueron propuestos inicialmente por Holland [165], aunque posteriormente han sido estudiados en profundidad por otros autores [166, 167].

Formalmente hablando los AGs son un tipo de algoritmos evolutivos enmarcados dentro de la familia de los algoritmos aproximativos estocásticos, que se caracterizan por proporcionar una solución aproximada del problema de optimización basándose en una búsqueda con criterio variable. Este tipo de algoritmos solventan el problema de la convergencia prematura de los algoritmos deterministas y son capaces de realizar búsquedas robustas en entornos complejos, aunque tienen como principal desventaja el alto coste computacional debido al retraso en la convergencia del algoritmo.

La idea básica de estos algoritmos consiste en hacer evolucionar, mediante un proceso iterativo, una población compuesta por individuos que codifican soluciones del problema. Generalmente cada individuo recibe el nombre de *cromosoma*, y cada una de sus componentes el de *gen*. El funcionamiento básico de un AG es el mostrado en el Procedimiento 3:

Procedimiento 3 Algoritmo Genético.

```
Generar Población Inicial  $P_0$ 
Evaluar  $P_0$ 
 $t = 0$ 
while no se cumpla el criterio de parada do
  Seleccionar  $P_{t+1}$  a partir de  $P_t$ 
  Cruzar y Mutar  $P_{t+1}$ 
  Evaluar  $P_{t+1}$ 
   $t = t + 1$ 
end while
```

Inicialmente, a partir de la información que se tiene del problema a resolver o de forma aleatoria, se genera un **población inicial** de individuos que se hace evolucionar a lo largo del tiempo mediante un proceso de competición y alteración controlada que emula los procesos genéticos que tienen lugar en la naturaleza. La forma en la que los individuos codifican las soluciones depende del problema concreto a resolver.

Una parte importante del algoritmo, a la que hay que prestar especial atención, es la definición de la **función de adaptación** o función objetivo, encargada de evaluar numéricamente el grado de adecuación de cada individuo a la solución del problema. Dicha función es la que va a guiar al algoritmo por el espacio de búsqueda, por lo que debe estar bien diseñada para que sea capaz, no solo de evaluar la adaptación de los individuos al problema, sino también de ordenarlos en función de su capacidad para resolverlo.

Una vez que cada individuo de la población inicial ha sido evaluado mediante la función de adaptación comienza el proceso iterativo representado en la Fig. B.1:

- **Selección:** es el mecanismo encargado de obtener una nueva población formada por copias de los mejores individuos de la población anterior, es

decir, aquellos que obtienen un mejor valor en la función de adaptación.

- **Cruce:** este operador implica combinar dos individuos (padres) de la nueva población para generar dos nuevos individuos (descendientes) con la intención de que estos últimos, obtenidos mediante la recombinación de los primeros, estén mejor adaptados al problema. Este operador solo se aplica sobre una selección aleatoria de la nueva población, en función de un parámetro denominado probabilidad de cruce (P_c).

El objetivo del operador de cruce es explotar el espacio de búsqueda, compartiendo información entre individuos con la intención de mantener las buenas características que presentan.

- **Mutación:** este segundo operador altera aleatoriamente, en función de una probabilidad de mutación (P_m), uno o más genes de los individuos de la nueva población.

El objetivo del operador de mutación es explorar el espacio de búsqueda, ya que el cambio aleatorio suele conllevar un salto a otra zona del espacio de búsqueda que puede ser más prometedora.

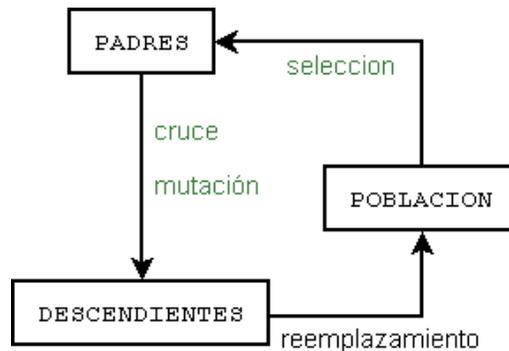


Figura B.1: Proceso iterativo de un AG.

Al conjunto de individuos generados en cada uno de los ciclos de evolución del algoritmo se le conoce con el nombre de *generación*. El proceso iterativo finaliza cuando se cumple el criterio de parada.

De forma adicional un AG puede ser completado con el concepto de **elitismo**, basado en mantener entre generaciones algunos de los individuos mejor evaluados.

Es decir, aquellos individuos mejor adaptados de una generación se mantienen exactamente igual en la siguiente, sin tener que aplicarles los operadores de cruce y mutación [166, 167].

B.2. Algoritmos Genéticos Multi-Objetivo

Se conoce con el nombre de Algoritmos Evolutivos Multi-Objetivo (AEMOs) a aquellos AGs en los que se definen múltiples objetivos relevantes para un mismo problema que, en general, están en conflicto. Es necesario por tanto definir múltiples funciones de adaptación a evaluar, cada una de ellas asociada a un objetivo diferente.

La mejor forma de tratar con este tipo de problemas es mediante el concepto de **Óptimo de Pareto**, que permite realizar una búsqueda en paralelo para identificar múltiples soluciones en una única ejecución del algoritmo. Formalmente un Frente de Pareto se define como el conjunto de soluciones no-dominadas para las cuales las correspondiente funciones objetivo no pueden ser mejoradas simultáneamente por ninguna otra solución [168, 169].

Los primeros AEMOs desarrollados por los investigadores, conocidos como algoritmos de primera generación, se caracterizan por su sencillez y basan su funcionamiento en combinar un buen mecanismo de selección con un buen mecanismo para mantener la diversidad. Los algoritmos más representativos de esta generación son “*Nondominated Sorting Genetic Algorithm*” (NSGA) [170], “*Niched-Pareto Genetic Algorithm*” (NPGA) [171] y “*Multi-Objective Genetic Algorithm*” (MOGA) [172].

Posteriormente, cuando el elitismo se convirtió en un mecanismo estándar tras ser propuesto en SPEA [173], aparecieron los AEMOs de segunda generación. Los algoritmos más representativos de esta etapa son “Strength Pareto Evolutionary Algorithm 2” (SPEA2) [138] y “Nondominated Sorting Genetic Algorithm II” (NSGA-II) [126].

Se describen a continuación el funcionamiento de estos dos últimos AEMOs.

B.2.1. NSGA-II

Propuesto por Deb y otros en el año 2002 [126] es uno de los AEMOs más utilizados por la comunidad científica.

Su funcionamiento se basa en el concepto de no-dominancia para realizar una clasificación de la población por frentes. Así los individuos que pertenecen al primer frente son los no dominados; los que pertenecen al segundo frente son los no dominados en ausencia de los del frente anterior, y así sucesivamente. A cada individuo dentro de cada frente se le asigna un rango equivalente a su nivel de no dominancia. Los mejores individuos son aquellos que tienen rangos menores y, por tanto, más posibilidades de reproducirse en la siguiente generación.

Este AEMO incorpora también el cálculo de una distancia de “crowding” que va a permitir mantener la diversidad de la población, con el fin de mejorar la selección por torneo binario. El Procedimiento 4 muestra su esquema de funcionamiento.

Procedimiento 4 Esquema del algoritmo NSGA-II.

Entradas: N (tamaño de la población), T (máximo número de generaciones)

Salida: Q (conjunto de soluciones no dominadas)

Generar población inicial P_0 de tamaño N

P_0 se ordena según la no-dominancia

A cada individuo se le asigna un objetivo igual a su nivel de no-dominancia

Los operadores de selección, cruce y mutación se utilizan para crear la población de descendientes Q_0 de tamaño N

$t=0$

repeat

$R_t = P_t \cup Q_t$

$F =$ ordenar R_t según la no dominancia

$P_{t+1} = \emptyset$ y $i = 1$

while $|P_{t+1}| + |F_i| \leq N$ **do**

 Calcular la distancia de cruce en F_i

$P_{t+1} = P_{t+1} \cup F_i$

$i = i + 1$

end while

Ordenar F_i de forma descendiente utilizando el operador de “crowding”

Seleccionar los primeros $(N - |P_{t+1}|)$ elementos de F_i

Usar los operadores de selección, cruce y mutación sobre P_{t+1} para generar la nueva población Q_{t+1}

$t = t + 1$

until se alcanza el criterio de parada, $t > T$

B.2.2. SPEA2

Propuesto por Zitler y otros en el año 2001 [138] como mejora al algoritmo SPEA [173], se caracteriza por: (1) modificar el proceso de la evaluación de los individuos, teniendo en cuenta para cada individuo el número de individuos a los que domina y por los que es dominado, (2) añadir una técnica de estimación de densidad poblacional, penalizando a los individuos en zonas muy pobladas para mejorar la exploración del espacio de búsqueda y (3) aplicar un método de truncamiento que evita eliminar las soluciones extremas del conjunto de individuos no dominados. El Procedimiento 5 muestra su esquema de funcionamiento.

Procedimiento 5 Esquema del algoritmo SPEA2.

Entradas: N (tamaño de la población), \bar{N} (tamaño de la población externa), T (máximo número de generaciones)

Salida: A (conjunto de soluciones no dominadas)

Generar población inicial P_0 y un fichero vacío con la población externa $\bar{P}_0 = \emptyset$.

$t = 0$

while no se cumpla el criterio de parada ($t < T$) **do**

 Evaluar todos los individuos de P_t y \bar{P}_t según las funciones objetivo

 Copiar todos los individuos no dominados de $P_t \cup \bar{P}_t$ en \bar{P}_{t+1} .

if $|\bar{P}_{t+1}| > N$ **then**

 Reducir $|\bar{P}_{t+1}|$ aplicando el operador de truncamiento

else

 Rellenar \bar{P}_{t+1} con individuos dominados de $P_t \cup \bar{P}_t$

end if

if se cumpla el criterio de parada **then**

$A = NoDominados(\bar{P}_{t+1})$ y terminar

end if

 Selección por torneo binario con reemplazo, teniendo en cuenta el nuevo criterio de evaluación de los individuos.

 Cruzar y Mutar P_{t+1}

$t = t + 1$

end while

Índice de figuras

1.1. Modelado de sistema.	6
1.2. Componente de un SBRD.	7
1.3. Estadísticas de publicaciones sobre interpretabilidad dentro del ámbito de la Lógica Difusa según ISI WOK.	10
1.4. Ejemplo de un <i>Frente de Pareto</i>	14
1.5. Estadísticas de publicaciones sobre SDE según ISI WOK.	15
2.1. Impacto actual de la Lógica Difusa según publicaciones en ISI WOK.	21
2.2. Impacto actual de la Lógica Difusa según patentes solicitadas en ESPACENET.	22
2.3. Búsqueda del equilibrio precisión-interpretabilidad en modelado difuso [40,41].	24
2.4. Clasificación del índice Número de Reglas según distintas taxonomías.	32
2.5. Clasificación del índice Número Total de Condiciones según distintas taxonomías.	33
2.6. Clasificación del índice Número de Variables según distintas taxonomías.	34
2.7. Clasificación del índice Número de Conjuntos Difusos según distintas taxonomías.	35
2.8. Partición difusa fuerte uniformemente distribuida.	37
2.9. Clasificación de las Restricciones semánticas según distintas taxonomías.	37

2.10. Clasificación de la Consistencia según distintas taxonomías.	38
2.11. Clasificación del Número de Reglas Activadas Simultáneamente según distintas taxonomías.	42
2.12. Taxonomía de los SDEs [33].	47
2.13. Ajuste de una función de pertenencia modificando sus parámetros característicos.	49
3.1. Proceso de mejora en dos fases.	56
4.1. Intervalos de variación de una función de pertenencia triangular. . .	91
4.2. Ejemplos de cálculo de GM3M.	94
4.3. Definición de puntos característicos e intervalos de variación para distintos tipos de funciones de pertenencia.	94
4.4. Ejemplos de cálculo de GM3M.	97
4.5. Grado de activación entre $Nucleo(A_{ik})$ y A_{jk}	101
4.6. Ejemplos de cálculo de RMI.	102
4.7. Grado de activación entre $Nucleo(A_{ik})$ y A_{jk} con $\alpha = 0,8$	103
4.8. Representación gráfica de una variable.	104
4.9. Proceso de mejora en dos fases.	105
4.10. BD obtenida para ELE con el método LING1.	139
4.11. BD obtenida para ELE con el método SCAT1.	141
4.12. BD obtenida para PLA con el método LING2.	144
4.13. BD obtenida para PLA con el método SCAT2.	146
A.1. Funciones de pertenencia más utilizadas.	182
A.2. Sistema Basado en Reglas Difusas.	184
B.1. Proceso iterativo de un AG.	189

Índice de tablas

1.1. Cuadrantes de interpretabilidad según [29].	12
2.1. Medidas de precisión para regresión.	25
3.1. Parametros NSGA-II.	61
3.2. Conjuntos de datos considerados en el estudio experimental.	62
3.3. Métodos considerados y parámetros de los SBRDs base iniciales.	65
3.4. Características iniciales para LINGA,LINGB, SCATA y SCATB.	67
3.5. Wang & Mendel vs Sistemas Neuro-Difusos. Los resultados mostrados para Wang & Mendel han sido obtenidos de [77].	68
3.6. SBRDs <i>compactos</i> mejorados: Características principales.	70
3.7. SBRDs <i>compactos</i> mejorados: Características Mejor $Inter_C$	71
3.8. SBRDs <i>compactos</i> mejorados: Características Mediana $Pre-Inter_C$	72
3.9. SBRDs <i>compactos</i> mejorados: Características Mejor Pre	73
3.10. Test de Wilcoxon para los SBRDs <i>Compactos</i> LINGA: SBRD base (R+) y SBRD mejorado (R-).	75
3.11. Test de Wilcoxon para los SBRDs <i>Compactos</i> SCATA: SBRD base (R+) y SBRD mejorado (R-).	76
3.12. SBRDs <i>complejos</i> mejorados: Características principales.	78
3.13. SBRDs <i>complejos</i> mejorados: Características Mejor $Inter_C$	79
3.14. SBRDs <i>complejos</i> mejorados: Características Mediana $Pre-Inter_C$	80
3.15. SBRDs <i>complejos</i> mejorados: Características Mejor Pre	81

3.16. Test de Wilcoxon para los SBRDs <i>Complejos</i> LINGB: modelo base (R+) y modelo mejorado (R-).	82
3.17. Test de Wilcoxon para los SBRDs <i>Complejos</i> SCATB: SBRD base (R+) y SBRD mejorado (R-).	83
4.1. Parámetros $L - TS_{SP2-SEM}$ y $S - TS_{SP2-SEM}$	113
4.2. Conjuntos de datos considerados en el estudio experimental.	113
4.3. Métodos considerados para el análisis.	116
4.4. Parámetros de generación de los SBRDs base iniciales.	118
4.5. Características iniciales para LING1, LING2, SCAT1 y SCAT2.	120
4.6. SBRDs más precisos: Características para cada método.	121
4.7. SBRDs más precisos: Test de Friedman en cada una de las medidas.	122
4.8. SBRDs más precisos: Test de Holm en cada una de las medidas con $\alpha = 0,1$	123
4.9. SBRDs en la mediana: Características para cada método (1/2).	125
4.10. SBRDs en la mediana: Características para cada método (2/2).	126
4.11. SBRDs en la mediana: Test de Friedman en cada una de las medidas.	128
4.12. SBRDs en la mediana: Test de Holm en cada una de las medidas con $\alpha = 0,1$	130
4.13. SBRDs más interpretables: Características para cada método.	131
4.14. SBRDs más interpretables: Características para cada método (2/2).	132
4.15. SBRDs más interpretables: Test de Friedman en cada una de las medidas.	133
4.16. SBRDs más interpretables: Test de Holm en cada una de las medidas con $\alpha = 0,1$	135
4.17. SBRDs del conjunto de datos ELE.	138
4.18. BR obtenida para ELE con el método LING1.	140
4.19. BR obtenida para ELE con el método SCAT1.	142
4.20. SBRDs del conjunto de datos PLA.	143
4.21. BR obtenida para PLA con el método LING2.	145

4.22. BR obtenida para PLA con el método SCAT2. 147