Study and Design of Multi-Objective Evolutionary Fuzzy Systems

for Improving the Interpretability-Accuracy Trade-Off of

Linguistic Fuzzy Rule-Based Systems when Dealing with

High-Dimensional and Large Scale Problems

Michela Fazzolari

Department of Computer Science and Artificial Intelligence

University of Granada

A thesis submitted for the degree of

PhilosophiæDoctor (PhD) in
Information and Communication Technologies

16th September 2013

This work, submitted by Michela Fazzolari to obtain the degree of PhilosophiæDoctor, has been realized within the PhD program "Information and Communication Technologies" of the University of Granada, under the supervision of Prof. Rafael Alcalá Fernandez and Prof. Francisco Herrera Triguero.

El doctorando Michela Fazzolari y los directores de la tesis Rafael Alcalá y Francisco Herrera garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la direccin de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realizacin del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, 16th September 2013

Directors   PhD fellow

*"You do not really understand something
unless you can explain it to your grandmother."
(A. Einstein)*

Acknowledgements

Todo empezó hace 3 años en aquel caluroso 8 de Julio 2010, cuando Ryanair operaba su último vuelo Bolonia-Granada. Como pisé el suelo andaluz, nada más llegar a la terminal del aeropuerto que casi me pillé una insolación.

No tenía ni idea de donde me había metido.

Desde aquel día mucha agua ha corrido bajo el puente. Italia perdió unas cuantas copas de fútbol, un to loco se lanzó desde 39.045m de altura, Gaddafi fue matado, William se casó con Kate, Steave Jobs se murió, el señor Schettino consiguió hundir un barco inhundible, el papa dimitió, el maratn de Nueva York no se corrió por primera vez en 30 años, el *tag* se difundió como la plaga del nuevo milenio, no obstante Berlusconi sigue en su sitio.

Mientras tanto yo he vivido una de las experiencias más importantes de mi vida, y esto fue posible también gracias a algunas personas que en estos tres años han cruzado su camino con el mio. Gracias a tod@s por el apoyo y la diversión que me ofrecisteis y que mucho me ayudaron durante estos años, sobre todo en los momentos difíciles. Así que gracias a quien estuvo conmigo en el quiz del Piaf, las fiestas sexy/robot, las noches en Hamelyn, la vuelta a casa pasando por los juguetes de la Caleta, las tapas en Plaza de Toros, las pizzas en Pizzametro, las fiestas de mojitos, los viajecitos por Andalucía, las exquisitas comidas de los comedores, las conferencias y sus noches locas, las clases de baile, el Triunfo en primavera, las barbacoas, los findes en la playa, las meriendas del domingo, las clases de granaino/atarfeño/manchego, las broncas incrementales, las puestas en San Miguel Alto......

Grazie anche a voi, parenti e amici italici, che mi avete tenuto compagnia da lontano durante questi tre anni e mi avete fatto trascorrere piacevolissimi momenti in Skype parlando di sogni, di ricordi, di ragazzi, di paper, di viaggi, di microcontrollori, di alberi a camme, di vacanze, di parenti, di progetti futuri. Sopratto grazie a te, mamma, che con la videochiamata ti diverti a giocare al Grande Fratello.

[NerdModeOn]
Thanks to the WEP internet connections of my neighbors, the Debian Project, the LaTeX community, Hercules, Atolon and Isaac.
[NerdModeOff]

Contents

Resumen y Conclusiones en Español

En este apartado se incluye el resumen en castellano de la tesis, para cumplir con los requisitos necesarios para poder acceder al título de doctor en Tecnologías de la Información y la Comunicación de la Universidad de Granada.

Introducción

El objetivo primario de esta memoria es el estudio del equilibrio entre precisión y complejidad en Sistemas Basados en Reglas Difusas (SBRDs) para el caso de problemas de alta dimensionalidad y/o con un gran número de ejemplos. Para ello, se proponen el uso de Algoritmos Evolutivos Multi-Objetivos (AEMOs), que permiten generar un conjunto de soluciones no-dominadas con distintos equilibrios para ambos objetivos. El contenido de este apartado está organizado en cinco secciones principales:

- En la primera sección se presenta el planteamiento del problema, introduciendo este con detalle y describiendo las técnicas utilizadas para resolverlo.

- A continuación, en la segunda sección, definimos algunos de los problemas abiertos en este marco de trabajo que justifican la realización de esta tesis.

- En la tercera sección se describe el objetivo general y se desglosa en objetivos específicos que constituyen el tema de cada propuesta presentada en la memoria.

- La cuarta sección incluye un resumen de las propuestas y de los resultados obtenidos en las distintas partes en que se divide el estudio.

- Finalmente, la última sección presenta algunos comentarios finales junto con algunas conclusiones sobre los resultados obtenidos y se comentan algunos aspectos sobre trabajos futuros que quedan abiertos.

Planteamiento

En las últimas décadas los sistemas basados en lógica difusa han tenido una gran difusión, debido a su flexibilidad para su aplicación en distintos problemas. En particular, los sistemas difusos más estudiados y utilizados son los SBRDs (130), que han sido aplicados con éxito a diferentes campos, como el control, el modelado, la clasificación, la predicción de datos, etc.

Los SBRDs están constituidos por un conjunto de reglas que permiten representar conceptos imprecisos e incluir conocimiento previo. Tradicionalmente, el diseño de un SBRD considera como principal objetivo la mejora de la precisión del modelo.

En los últimos años, la investigación en este campo ha evolucionado integrando los sistemas difusos con otras técnicas tales como los algoritmos evolutivos (AEs). La hibridación de ambos sistemas da origen a los sistemas difusos evolutivos (SDEs), y en particular a los sistemas basados en reglas difusas evolutivos (SBRDEs), donde el proceso de diseño del SBRD se considera como un problema de optimización o búsqueda y el AE es utilizado para resolver este problema.

En el marco de los AEs, los algoritmos genéticos (AGs) se consideran actualmente como una de las técnicas de búsqueda global más conocida y empleada, siendo procedimientos adaptativos para buscar soluciones en espacios complejos. Los AGs ofrecen soluciones aproximadas válidas a problemas donde las técnicas clásicas de búsqueda no resultan eficientes. La principal ventaja del uso de los AGs es su capacidad para explotar la información acumulada sobre un espacio de búsqueda y dirigir las siguientes búsquedas hacia los mejores sub-espacios.

Por esta razón, numerosos autores han trabajado en el desarrollo de AGs para el aprendizaje automático de SBRDs. En particular, en las últimas décadas, el enfoque se ha centrado en los modelos que permiten manejar múltiples objetivos, dando lugar a los Sistemas Difusos Evolutivos Multi-Objetivos (SDEMOs). Este tipo de sistemas permite obtener un conjunto de soluciones no dominadas, es decir que no existe una solución óptima con respecto a todos los objetivos, sino que cada solución del conjunto representa un equilibrio entre los objetivos considerados.

Uno de los objetivos adicionales comúnmente usados es la interpretabilidad del SBRD (40), es decir la posibilidad de entender fácilmente el modelo. El problema está en la definición de índices universalmente aceptados para evaluar dicha propiedad.

Actualmente los investigadores concuerdan en distinguir dos formas de manejar la interpretabilidad:

- Controlar la complejidad del modelo por medio de medidas como el número de reglas, de variables, de etiquetas por regla, etc.

- Medir la interpretabilidad de las particiones difusas usando medidas de interpretabilidad semántica (cobertura, distinguibilidad, consistencia, etc).

Con respecto a la interpretabilidad semántica, no existen todavía medídas universalmente aceptadas por los investigadores, por lo tanto en la memoria nos hemos centrado en la complejidad de los modelos.

Justificación

Después de introducir brevemente los principales conceptos a los que se refiere la memoria, nos planteamos una serie de problemas abiertos que justifican el desarrollo del proyecto de tesis.

Actualmente el área de investigación sobre SDEMOs se puede considerar madura. Sin embargo, aún quedan muchos problemas sin resolver cuando se utilizan dichos sistemas en aplicaciones especificas, por ejemplo cuando nos encontramos en presencia de problemas de alta dimensionalidad (118). Este tipo de problemas se da con muchas frecuencia en aplicaciones reales hoy en día debido a la facilidad para realizar mediciones, recuperar información y coleccionar grandes cantidades de datos.

En estos casos, el proceso de aprendizaje de un SBRD se enfrenta a un espacio de búsqueda que se incrementa al aumentarse el número de variables y de ejemplos del conjunto de datos. sto dificulta el proceso de aprendizaje y, en la mayoría de los casos, lleva a la generación de un SBRD con un gran número de reglas, lo que reduce el nivel de interpretabilidad del sistema.

Por otro lado, cuando se tratan problemas de alta dimensionalidad, el tiempo de ejecución de los algoritmos estándar aumenta tanto que a veces no es posible utilizarlos en estas aplicaciones. Por lo tanto, se necesita el diseño de algoritmos específicos a partir de las versiones estándar para manejar el complejo espacio de búsqueda de manera adecuada.

Una forma de afrontar ambos problemas, mejorar el equilibrio entre interpretabilidad y precisión y abordar problemas de alta dimensionalidad es mediante el aprendizaje de la granularidad (número de etiquetas lingüísticas asociadas a cada variable) (59). A lo largo de los años, se han formulado varias propuestas en la literatura. La estrategia más simple consiste en fijar previamente una única granularidad y generar particiones difusas uniformes para todas las variables (105, 113). A pesar de su simplicidad, esta estrategia en algunos casos no resulta adecuada, porqué no considera en absoluto la información sobre un determinado problema incluida en los datos disponibles y normalmente conlleva la generación de un gran número de reglas. Otras estrategias

intentan determinar automáticamente granularidades adecuadas a partir de un conjunto de datos. Sin embargo los métodos propuestos (13, 115) presentan algunas limitaciones cuando se aplican a problemas de alta dimensionalidad o con un gran número de ejemplos, debido al incremento de la dimensión del espacio de búsqueda y al aumento del tiempo necesario por la evaluación de la función de fitness en el MOEA.

Objetivos

Para dar solución a los distintos problemas que se acaban de mencionar en la sección anterior, la memoria de tesis se desarrolla entorno a los siguientes objetivos, que implican el estudio del comportamiento de los SDEMOs considerando el equilibrio entre precisión y complejidad de los SBRDs en presencia de problemas de alta dimensionalidad. En concreto, los objetivos que proponemos son:

1. Realizar un estudio sobre los SDEMOs existentes en la literatura. Proponer una taxonomía que permita organizar las contribuciones en distintas categorías para conocer cuales son los problemas abiertos relacionados con el diseño de los SDEMOs.

2. Realizar igualmente un estudio sobre las propuestas que hayan tratado el tema del equilibrio entre interpretabilidad y precisión en los SDEMOs, centrándonos en aquellos algoritmos que alcanzan los mejores resultados.

3. Realizar un estudio sobre la influencia del aprendizaje de la granularidad de la base de datos de los SBRDs, desarrollando distintos algoritmos con el objetivo de mejorar el balance entre precisión y complejidad mediante el aprendizaje del número adecuado de funciones de pertenencias (es decir la granularidad) de cada variable. Para ello, se pretenden desarrollar varios SDEMOs considerando dos objetivos contradictorios: minimizar el error del modelo (precisión) y minimizar el numero de reglas (complejidad).

4. Estudiar la combinación de las técnicas de pre-procesamiento de selección de instancias con los SDEs. Combinar ambas técnicas con el objetivo de determinar si la reducción del número de ejemplos es efectiva para reducir también la complejidad de los SBRDs y el tiempo de calculo requerido para manejar grandes cantidades de datos, manteniendo una precisión aceptable.

5. Validar los resultados obtenidos mediante una comparación con los métodos existentes, mediante el uso de técnicas estadísticas.

Desarrollo de la tesis

A continuación se describen brevemente los capítulos de la tesis, resumiendo las propuestas incluidas en la memoria y presentando para cada una de ellas una breve discusión sobre los resultados obtenidos.

Resumen del capítulo 1: Estudio sobre las aplicaciones de SDEMOs: estado-del-arte y problemas abiertos

En esta sección presentamos un resumen del primer capítulo de la tesis, donde se propone una taxonomía en dos niveles para clasificar las contribuciones más importantes sobre el tema de los SDEMOs. El primer nivel está basado en el tipo de problema multi-objetivos abordado, mientras que en el segundo nivel se organizan las propuestas basándose en el tipo de componentes del SBRD que se optimizan durante el proceso evolutivo. En el primer nivel se han identificado tres distintas categorías:

1. SDEMOs diseñados para problemas que consideran la interpretabilidad en los SBRDs: el primer objetivo es la precisión, los demás objetivos están relacionados con el tema de la interpretabilidad.

2. SDEMOs diseñados para problemas de control: los objetivos considerados dependen del problema de control que se aborda.

3. SDEMOs diseñados para extracción de reglas de asociación difusas: se consideran objetivos que describen la calidad de las reglas extraídas.

SDEMOs diseñados para mejorar el equilibrio entre precisión e interpretabilidad

La primera categoría reúne contribuciones en las cuales los SDEMOs se han diseñado para generar modelos difusos que presentan un buen equilibrio entre precisión e interpretabilidad. En estos casos, por lo menos uno de los objetivos está siempre relacionado con la interpretabilidad del modelo obtenido. La mayoría de todos los trabajos que pertenecen al ámbito de los SDEMOs se ocupan de este problema, ya que la interpretabilidad es uno de los aspectos más importantes de los SBRDs.

El problema de la mejora de la precisión, manteniendo o mejorando la interpretabilidad de un modelo difuso, se abordó por primera vez a mediados de 1990 (105), integrando la interpretabilidad en el proceso de optimización, gracias a la aplicación de los AEMOs a sistemas difusos. Desde entonces, la interpretabilidad ha adquirido una importancia creciente.

Debido a su subjetividad, el problema principal es encontrar una definición universalmente reconocida y una manera objetiva de medir esta característica. En la última década, varios trabajos han analizado el problema de la interpretabilidad en los SBRDs (54), en busca de medidas que podrían ser universalmente aceptadas por la comunidad científica (61, 95, 151) y este esfuerzo ha continuado en los últimos años, como demuestran los numerosos trabajos de revisión publicados (21, 84, 138, 168). Actualmente los investigadores concuerdan en considerar dos tipos de medidas:

- Medidas basadas en complejidad: se utilizan para disminuir la complejidad del modelo difuso (número de reglas, número de antecedentes en una regla, etc).

- Medidas de interpretabilidad semántica: se utilizan para preservar la semántica asociada con las funciones de pertenencia (distinguibilidad, cobertura, etc) y reglas (consistencia, etc.)

Generalmente al evaluar la interpretabilidad de un modelo difuso, los índices se centran en el primer tipo de medidas, mientras que la definición de buenas medidas de interpretabilidad semántica es todavía un problema abierto (21, 25, 83). Por lo tanto, en esta tesis nos centraremos en las medidas de complejidad.

Considerando la importancia del equilibrio entre precisión e interpretabilidad, en esta primera categoría se incluyen las contribuciones en las que los SDEMOs están diseñados para manejar estos objetivos. Debido al gran número de trabajos existentes, se propone también un segundo nivel de clasificación, según la taxonomía de SDEs presentada en (97), teniendo en cuenta los componentes del SBRD gestionados por el proceso de optimización:

- Ajuste de componentes del SBRD, combinado o no con un proceso de selección de reglas: el proceso de optimización ajusta una base de conocimiento predefinida, es decir, los parámetros del sistema se modifican para obtener sistemas más precisos. Con el fin de mantener el sistema simple y de reducir la complejidad, en algunos casos, un proceso de selección de la reglas se puede integrar en la optimización. Las contribuciones que pertenecen a esta categoría se dividen en dos sub-categorías, nombradas *ajuste de las funciones de pertenencia* y *ajuste de los parámetros de inferencia*.

- Aprendizaje de la base de conocimiento: los trabajos que pertenecen a esta categoría consideran el aprendizaje de la base de datos y/o de reglas. Este grupo se divide en tres sub-categorías: *aprendizaje por selección de reglas* (en este caso el proceso de selección de reglas se utiliza para llevar a cabo un aprendizaje de

la base de reglas), *aprendizaje de la base de reglas* y *aprendizaje simultáneo de los componentes de la base de conocimiento*.

La mayoría de las contribuciones utilizan un modelo difuso lingüístico, ya que es el tipo más interpretable de SBRD. Sin embargo, hay un número limitado de trabajos que consideran el tema de la interpretabilidad incluso en los SBRDs de tipo Takagi-Sugeno-Kang (TSK).

SDEMOs diseñados para problemas de control

Tradicionalmente la calidad de un sistema de control está relacionada con la precisión en el modelado de la dinámica del sistema. En el diseño de un sistema de control un primer problema surge cuando falta un completo conocimiento de los procesos físicos involucrados y los procesos se describen de manera imprecisa. Otro problema consiste en diseñar modelos adaptables, es decir sistemas inteligentes que sean capaces de actuar un proceso de aprendizaje y adaptación cuando los parámetros del sistema real cambian. Por lo tanto, puede resultar difícil identificar un modelo dinámico preciso para diseñar un controlador tradicional.

En estos casos, la lógica difusa representa una herramienta para afrontar el problema de la representación del conocimiento en un entorno de incertidumbre e imprecisión. Además, en el diseño del sistema de control, frecuentemente es necesario considerar múltiples objetivos que pueden estar en conflicto entre ellos. Por lo tanto no existe una única solución de diseño que se pueda considerar como la mejor con respecto a todos los objetivos. Estas consideraciones han conducido a la aplicación de los AEMOs para el diseño de Controladores Basados en Lógica Difusa (CBLDs).

El diseño de un CBLD presupone determinar tanto la estructura del controlador como los parámetros numéricos correspondientes. Los AEMOs pueden tratar los dos problemas mediante la codificación de la estructura y de los parámetros en un cromosoma que representa el CBLD. Por lo tanto, esta segunda categoría reúne los trabajos que aplican los AEMOs a los CBLDs, considerando las siguientes sub-categorías (79):

- identificación de los parámetros del controlador y/o de las reglas (por ejemplo el ajuste de los parámetros de la función de pertenencia);

- aprendizaje de la estructura del controlador (por ejemplo el aprendizaje de la base de reglas).

En esta categoría se han incluido también algunos trabajos que describen una hibridación entre AEMOs, lógica difusa y redes neuronales.

SDEMOs diseñados para la extracción de reglas de asociación difusas

Los SDEMOs se pueden utilizar para la extracción automática de conocimiento a partir de datos, por lo tanto los problemas de minería de datos son uno de los dominios de aplicación más importantes para los SDEMOs. La minería de datos se ha tratado como un sinónimo de Knowledge Discovery in Databases (KDD) (75, 146), aunque en realidad es la etapa de análisis de KDD. El objetivo general de un proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior.

Generalmente las técnicas de minería de datos se dividen en dos categorías: predictivas o descriptivas. Un enfoque predictivo se centra en la capacidad de predicción y genera modelos que se pueden utilizar para predecir valores, basándose en patrones determinados a partir de datos conocidos. Uno de los métodos utilizados en los modelos predictivos es el *aprendizaje supervisado*, que determina una función a partir de los datos de entrenamiento. Consiguientemente, dicha función se utiliza para predecir el valor de salida por cualquier dato de entrada válido. La técnicas predictivas se aplican en problemas de clasificación, de regresión y en algunos casos también se pueden utilizar en problemas de control.

Por otro lado, el enfoque descriptivo se centra en la comprensión del proceso de generación de datos, buscando de patrones interesantes en los datos existentes, sin tener ningún objetivo predefinido. El método utilizado en este modelo es normalmente el *aprendizaje no supervisado*, que difiere del aprendizaje supervisado en el hecho de que no se conoce la salida de los datos de entrenamiento. Esta estrategia se aplica principalmente a modelos que funcionan con reglas asociativas.

Por último, existen también aplicaciones de minería de datos que requieren un cierto grado de predicción como de descripción. Un método que combina el enfoque predictivo con el descriptivo es el Descubrimiento se Subgrupos (DS) (128).

En minería de datos, una forma de representar el conocimiento extraído desde una base de datos es por medio de reglas de asociación (167), cuyo concepto básico es descubrir asociaciones significativas entre los valores de pares de atributos, por ejemplo, si la presencia de un valor para un determinado atributo implica la presencia de otro valor para otro atributo. Como los sistemas difusos pueden tratar con conocimiento impreciso, se pueden aplicar con éxito también para la representación de este tipo conocimiento, extendiendo las reglas de asociación a reglas de asociación difusas (65).

En la minería de reglas de asociación difusas, los objetivos se basan en la calidad de las reglas extraídas: estas reglas deben ser precisas, generales o suficientemente específicas, interesantes, etc. Debido a la gran cantidad de objetivos, los AEMOs se

han utilizado con éxito para extraer reglas de asociación difusas. Los trabajos incluidos en este grupo utilizan principalmente un enfoque descriptivo, es decir, se centran en producir modelos de los datos comprensibles e interpretables. Además, pertenecen a esta categoría algunos trabajos que utilizan el enfoque de descubrimiento de subgrupos.

Problemas abiertos

A continuación se presentan algunas tendencias actuales en el campo de los SDEMOs y se plantean algunas cuestiones para dirigir la atención de los investigadores hacia nuevos problemas que surgen cuando se utilizan dichos sistemas en aplicaciones reales.

Una cuestión importante está relacionada con el hecho de que los AEMOs no están específicamente diseñados para ser integrados en los SDEMOs, donde un cromosoma representa partes de un SBRD y consecuentemente asume una estructura compleja que puede comprender distintas codificas.

Además, los SDEMOs deben tener en cuenta el error sobre los datos de test, que normalmente no se considera en los problemas estándar de la optimización evolutiva multi-objetivo. Debido a ésto, es posible que los AEMOs existentes no sean adecuados para optimizar la estructura de los SBRDs, generando así soluciones sub-óptimas.

Considerando estos asuntos y el estado-del-arte, hemos evidenciado los siguientes temas relacionados con los SDEMOs en los cuales todavía se puede investigar.

- Evaluación de la calidad de los SDEMOs para que se puedan comparar.

- Medidas de interpretabilidad fiables.

- Dimensionalidad de los objetivos.

- Problemas de escalabilidad.

- Aplicación a conjuntos de datos imbalanceados.

- Selección automática de la solución más adecuada entre las soluciones que pertenecen al frente de Pareto.

- Integración de preferencias en el proceso evolutivo.

- Diseño de SDEMOS que incluyen sistemas difusos de tipo 2.

Conclusiones

El estudio presentado en el primer capítulo de la memoria ha evidenciado claramente que el interés de la comunidad científica se centra en el equilibrio entre interpretabilidad y precisión, de hecho la primera categoría de la taxonomía incluye la mayoría de los trabajos existentes. Sin embargo, quedan aún algunos problemas abiertos que ofrecen nuevas líneas de investigación en este campo.

Considerando que los problemas reales se vuelven más complejos, implicando problemas de escalabilidad para los algoritmos que los tratan, en el desarrollo de la tesis decidimos centrarnos en el estudio de nuevos SDEMOs para mejorar el equilibrio entre interpretabilidad y precisión en este tipo de problemas, que requieren de un diseño especifico de los algoritmos propuestos.

Resumen del capítulo 2: Mejora del equilibrio entre precisión y complejidad de SDEMOs por medio de un aprendizaje de la granularidad basado en medidas heurísticas sobre granularidad múltiple: MO-FARCG

Una manera de mejorar el equilibrio entre precisión y complejidad es por medio de la aplicación de un proceso evolutivo de selección de reglas junto con un proceso de ajuste. En este caso, no se conoce con antelación la granularidad, es decir el número apropiado de funciones de pertenencia para cada variable.

Para enfrentarse al problema existen dos posibilidades: fijar previamente una granularidad (105, 113) o adoptar múltiples granularidades (115). La primera estrategia es más simple, aunque la elección de la granularidad se realiza a menudo a mano, y por lo tanto no es optimizada e induce la generación de un elevado número de reglas difusas. Por otro lado, el enfoque de múltiple granularidades es útil para reducir el número de reglas, pero se criticó por la pérdida de interpretabilidad de los modelos obtenidos.

En la memoria asociada a este resumen se han presentado dos propuesta para abordar el problema de como determinar la granularidad, puesto que granularidades adecuadas contribuyen a la mejora del equilibrio entre precisión y complejidad.

La primera propuesta se ha descrito en el segundo capítulo y utiliza algunos conceptos presentados en (13), donde los autores han propuesto un mecanismo para identificar granularidades individuales apropiadas durante un proceso evolutivo multi-objetivo de selección de reglas difusas, basado en la propuesta presentada en (115). Dicho mecanismo incluye cuatro pasos: a) se utiliza un procedimiento heurístico para generar un número pre-determinado de reglas difusas prometedoras con granularidades múltiples,

b) para cada variable se asigna una sola granularidad, teniendo en cuenta la frecuencia de las particiones usadas y la importancia de las reglas extraídas en el paso anterior; c) las granularidades individuales determinadas en el paso previo se utilizan para extraer otra vez un número pre-determinado de reglas difusas; d) por último se utiliza un AEMO para realizar un proceso de selección de reglas.

El método propuesto (MO-FARCG) combina el mecanismo de especificación de granularidades individuales descrito anteriormente con una nueva versión multi-objetivo de un algoritmo de clasificación basado en reglas de asociación difusas para problemas de alta dimensionalidad, llamado FARC-HD y propuesto en (15). El objetivo es evitar el uso de múltiples granularidades, proporcionando aun así una reducción de la complejidad de los clasificadores obtenidos y manteniendo al mismo tiempo una alta capacidad de generalización, considerando los dos objetivos en un marco evolutivo multi-objetivo.

El método consta de tres etapas:

1. Fase de configuración: aprendizaje de las granularidades adecuadas. El fin de un problema de clasificación es determinar para cada ejemplo de ingreso una clase de salida. En esta etapa, para cada clase se genera un número pre-determinado de reglas con múltiples granularidades, evaluando las reglas según medidas comúnmente reconocidas en el campo de la minería de datos (1). Después se elige un sola granularidad para cada variable, según la frecuencia de las reglas extraídas y según unas medidas de calidad.

2. Fase de aprendizaje: extracción de reglas de asociación difusas candidatas. Todos los posibles conjuntos de ítems frecuentes se enumeran en un árbol de búsqueda que se utiliza para generar reglas de asociación difusas. Por último, las reglas son evaluadas y ordenadas según un criterio y sólo las mejores reglas se mantienen, con el fin de reducir el número de reglas candidatas.

3. Fase de post-procesamiento: aplicación del proceso evolutivo multi-objetivo para selección de reglas y ajuste de las funciones de pertenencia. Se seleccionan las reglas mejores y se ajustan las funciones de pertenencia, utilizando un AEMO basado en el Strength Pareto Evolutionary Algorithm (SPEA2) (169), aprovechando la sinergia positiva de ambas técnicas en el mismo proceso.

Resultados y conclusiones

El método se ha comparado con el FARC-HD original (15), considerando 24 problemas reales con distintos tamaños. El método propuesto se ha evaluado comparando

sus resultados con los resultados obtenidos por FARC-HD. Este algoritmo pertenece actualmente al estado del arte de los algoritmos de clasificación, y se ha demostrado que supera en precisión algunos de los algoritmos de clasificación más conocidos.

Se han considerado dos versiones de MO-FARCG, utilizando dos criterios diferentes para la selección de la granularidad individual: el producto (MO-FARCG-prod) y la confianza (MO-FARCG-conf), respectivamente. Debido al enfoque multi-objetivo del algoritmo SPEA2 incluido en MO-FARCG, para la comparación se ha considerado el promedio de la solución más precisa de todos los frentes de Pareto. Para investigar la presencia de diferencias estadísticas entre los métodos, se ha aplicado el test no paramétrico de Rangos y Signos de Wilcoxon (153, 162) con un nivel de confianza del 95% ($\alpha = 0.05$).

Una primera comparación se ha realizado entre las dos versiones diferentes de MO-FARCG aplicando el test de Wilcoxon sobre los porcentajes medios de clasificación obtenidos en test. Los resultados han mostrado que los dos métodos no son equivalentes y la versión MO-FARCG-conf resulta mejor. Por lo tanto, esta versión de MO-FARCG se ha elegido para la comparación con FARC-HD y se ha aplicado otra vez el test de Wilcoxon sobre los porcentajes medios de clasificación obtenidos en test por las soluciones más precisas. Los resultados evidencian que los dos métodos no son equivalente y que FARC-HD consigue obtener soluciones más precisas. Una comparación adicional se ha realizado con respecto al número medio de reglas que constituyen las soluciones más precisas. En este caso el método propuesto MO-FARCG-conf resulta mejor que FARC-HD.

Observando las conclusiones de los test estadísticos y los resultados medios obtenidos por los algoritmos, se puede concluir que MO-FARCG-conf es superado por el FARC-HD con respecto a la precisión en el test, mientras que se verifica lo contrario cuando se considera la complejidad de los sistemas obtenidos. Sin embargo, frente a una pérdida de menos del 3% en precisión del test, la complejidad se reduce en más de un 50%. Por lo tanto, el aprendizaje de las granularidades combinado con MO-FARCG-conf produce modelos con una precisión ligeramente disminuida, lo que se compensa con una reducción muy considerable de la complejidad.

Resumen del capítulo 3: Mejora del equilibrio entre precisión y complejidad de SDEMOs por medio de un aprendizaje de la granularidad basado en discretización difusa: D-MOFARC

Como se ha afirmado en la sección anterior, la precisión y la complejidad de los SBRDs dependen de la definición de la base de datos asociada con la base de reglas, por lo

tanto el diseño de una base de datos adecuada es una tarea crucial, particularmente la definición de la granularidad. El método presentado en la sección anterior incluye un aprendizaje de la granularidad y consigue generar modelos significativamente más simples a costa de una ligera perdida de precisión.

En esta sección se resume el tercer capítulo de la memoria, donde se propone un nuevo método con el fin de mejorar al mismo tiempo los dos objetivos. El método se basa en un algoritmo de discretización difusa que extrae granularidades adecuadas a partir de los datos, para generar una base de datos inicial. Este mecanismo se ha integrado dentro de un SDEMO que evoluciona la base de conocimiento inicial. Aunque el objetivo principal es la precisión y el AEMO se ha diseñado como una herramienta para mejorarla, al mismo tiempo ayuda a reducir la complejidad de los modelos (82). Este método, llamado D-MOFARC, comprende los siguientes pasos:

- Un algoritmo de discretización difusa se ha diseñado para aprender automáticamente granularidades adecuadas para cada variable y para generar las correspondientes particiones difusas. Este algoritmo se basa en el concepto de discretización (44, 76), que es el proceso de transformación del conjunto de valores continuos de un atributo en un conjunto de intervalos y la asignación de un valor discreto a cada intervalo. Este concepto se ha extendido al caso de particiones difusas, teniendo en cuenta las interdependencias entre las variables. El algoritmo resultante determina un conjunto de intervalos para cada variable y después asigna un conjunto difuso a cada intervalo obtenido, en lugar de un valor discreto. De esta manera se determinan las etiquetas difusas asociadas a cada variable. Este proceso se ha integrado en un mecanismo de generación basado en árboles que considera las interdependencias entre las variables.

- Se genera una base de reglas inicial asociada a las particiones difusas generadas en la etapa anterior. La base de reglas se genera extrayendo reglas de asociación difusas que constituyen las reglas candidatas. Para ello, se han utilizado los dos primeros pasos del modelo FARC-HD presentado en (15). Las reglas extraídas no utilizan todas las etiquetas generadas en el paso inicial, por lo tanto las particiones difusas iniciales se ajustan mediante la eliminación de las etiquetas no utilizadas.

- Un nuevo AEMO específico se ha diseñado para realizar al mismo tiempo un proceso de ajuste de las funciones de pertenencia en la base de datos y un proceso de selección de reglas en la base de reglas. Este algoritmo representa una versión modificada del algoritmo SPEA2 (169), y tiene como propósito la mejora de la

precisión, reduciendo la complejidad del modelo inicial siempre que no vaya en detrimento de la precisión. Se ha preferido un enfoque multi-objetivo respecto a uno mono-objetivo porqué se ha demostrado ser útil en la generación de modelos más precisos, mientras que el uso del número de reglas como segundo objetivo ayuda en la limitación de la complejidad (82).

Resultados y conclusiones

El método propuesto se ha evaluado comparando sus resultados con los resultados obtenidos por FARC-HD (15). Como se ha comentado en la sección anterior, este algoritmo pertenece actualmente al estado del arte de los algoritmos de clasificación, y se ha demostrado que supera en precisión algunos de los algoritmos de clasificación más conocidos. Los experimentos se han realizado considerando 35 problemas reales y los dos métodos se han comparado mediante la aplicación del test no paramétrico de Rangos y Signos de Wilcoxon con un nivel de confianza del 95% ($\alpha = 0.05$). El test se ha aplicado considerando el porcentaje medio de acierto sobre los datos de test para la precisión y considerando el número medio de reglas obtenidas para la complejidad.

Los resultados muestran que el método propuesto genera soluciones más precisas respecto FARC-HD. Por otro lado, comparando los dos métodos con respecto al número medio de reglas de las soluciones más precisas, se nota que la diferencia estadística entre los dos no es tan evidente. Sin embargo, al observar los valores medios obtenidos por los dos métodos, se puede ver que el número medio de reglas generado aplicando el método D-MOFARC es ligeramente menor que el valor obtenido aplicando el algoritmo FARC-HD.

Por lo tanto se puede concluir que el D-MOFARC supera claramente a FARC-HD considerando los resultados en la precisión del test, mientras que se observa una diferencia estadística menos significativa cuando se considera la complejidad. Aun así, el número de reglas se reduce en media aplicando el D-MOFARC.

Resumen del capítulo 4: Estudio sobre la aplicación de técnicas de selección de instancias en los sistemas de clasificación evolutivos basados en reglas difusas para mejorar el equilibrio entre precisión y complejidad

El proceso de aprendizaje de un SDE se ve afectado negativamente cuando crece el número de instancias utilizado para generar los SBRDs. Un primer problema está relacionado con el tiempo de cálculo necesario para la evaluación de la función de fitness

durante el proceso evolutivo, ya que es directamente proporcional a la cantidad de instancias. Un segundo problema está relacionado con la complejidad de los modelos obtenidos: con el fin de cubrir lo más posible las instancias del conjunto de datos, el proceso de aprendizaje suele generar un alto número de reglas.

Para enfrentarse al problema se pueden aplicar técnicas de pre-procesamiento sobre los conjuntos de datos. El objetivo perseguido por el pre-procesamiento es obtener conjuntos de datos tales que al aplicar técnicas de aprendizaje automático sobre ellos se generan modelos representativos con mejores prestaciones.

Con este objetivo, en la literatura se han propuesto varios métodos de reducción de datos. En particular, cuando se consideran problemas de alta dimensionalidad se pueden aplicar técnicas de Selección de Instancias (SIs) (68, 86, 133, 163), que tienen como objetivo extraer desde el conjunto inicial un pequeño subconjunto que sea representativo del conjunto original.

Disminuyendo el conjunto inicial de datos se consigue reducir tanto el tiempo de cálculo, como la complejidad de los modelos obtenidos. La eliminación de instancias no tiene porqué producir una degradación de los resultados, ya que se pueden estar eliminando ejemplos repetidos, datos inconsistentes, redundantes, etc. Es interesante notar que reduciendo el número de ejemplos se puede en algunos casos superar situaciones de sobre-aprendizaje.

Las técnicas de SIs se pueden agrupar en dos categorías, en función del objetivo perseguido después de obtener el conjunto reducido:

- Selección de prototipos (SP) (145): el conjunto reducido viene utilizado por clasificadores basados en prototipos (por ejemplo, K-NN) para clasificar nuevas instancias. Estos tipos de clasificadores asumen que las instancias no etiquetadas se puedan clasificar basándose en las instancias etiquetadas, de acuerdo con una cierta similitud o función de distancia. En este caso el subconjunto seleccionado debe proporcionar el mejor compromiso entre la precisión de la clasificación y la reducción del número de instancias.

- Selección de conjuntos de entrenamiento (SCE): (27, 37): el subconjunto de instancias viene utilizado por un algoritmo de aprendizaje automático para construir un modelo de predicción (por ejemplo, redes neuronales, SBRDs, árboles de decisión, etc.)

En esta sección se resume el cuarto capítulo de la memoria, donde nos centramos en el uso de técnicas de SCE como método de pre-procesamiento antes de aplicar un SDE para la generación de SBRDs. Nuestro objetivo es investigar si las técnicas de SCE

ayudan a reducir la complejidad de los SBRDs generados, preservando o incluso aumentando su precisión. Para ello, hemos presentado un primer estudio en (91), donde se han considerado una serie de 20 conjuntos de datos de tamaño pequeño. Este estudio ha evidenciado como una familia especifica de métodos de SCE permite obtener modelos menos complejos manteniendo sustancialmente la precisión. Dicho estudio se ha extendido considerando también conjuntos de datos de tamaño medio-grande, que aparecen con frecuencia en los problemas del mundo real. Cuando se consideran estos conjuntos, el número de reglas de los SBRDs generados puede ser muy grande y por lo tanto su interpretabilidad puede ser muy baja.

Resultados y conclusiones

Para el estudio completo se han considerado 36 técnicas de SCE, 20 conjuntos de datos pequeños y 17 conjuntos de datos adicionales de tamaño medio-grande. Las técnicas de SCE se han aplicado a cada conjunto de datos y se han obtenido los conjuntos reducidos. Después, los conjuntos de datos reducidos se han utilizado para generar SBRDs mediante el uso del algoritmo FARC-HD (15), que se ha demostrado ser eficaz cuando se manejan conjuntos de datos de alta dimensionalidad.

El objetivo es investigar si las técnicas de SCE son capaces de disminuir el número de instancias de un conjunto de datos sin perder la información necesaria para permitir que FARC-HD genere SBRDs con alto porcentaje de clasificación y una complejidad reducida, con un tiempo de computación mínimo. El estudio se ha realizado teniendo en cuenta la combinación de los conjuntos de datos de tamaño pequeño y medio-grande para obtener resultados más fiables cuando se aplican los tests estadísticos y para investigar si existen técnicas de SIs que pueden ser utilizadas eficazmente con conjuntos de datos de cualquier tamaño. El análisis ha evidenciado que se deben utilizar diferentes técnicas en función de la dimensión del conjunto de datos considerado. Combinando distintas técnicas para conjuntos pequeños y medio-grandes se obtiene un buen equilibrio entre la reducción de la complejidad y la precisión de los SBRDs generados, con una reducción media del número de reglas del 38% y una reducción media de la precisión del 2%.

Por último, se ha realizado un análisis del tiempo de cálculo requerido para la aplicación de técnicas de SCE y para la ejecución del SED en los conjuntos de datos reducidos, con el fin de evaluar si los subconjuntos seleccionados conducen a una reducción en el tiempo empleado por el SED para generar modelos de clasificación. Sin embargo, los tiempos de cálculo totales requeridos para ejecutar tanto el pre-procesamiento como el algoritmo FARC-HD son más largos que el tiempo necesario para la ejecución de

FARC-HD sobre el conjunto de entrenamiento inicial. Por lo tanto, el objetivo de utilizar una técnica de SCE no es reducir el tiempo de cálculo en general, sino más bien mejorar el equilibrio entre la precisión y la complejidad, al reducir el número de reglas, preservando al mismo tiempo la mayor parte de la precisión.

Comentarios finales

A continuación se presentan unas conclusiones globales sobre el trabajo desarrollado durante el proyecto de tesis y algunas lineas de investigación futuras relacionadas con los SDEMOs.

Conclusiones

En la memoria asociada a este resumen se ha tratado el tema de la mejora del equilibrio entre precisión y complejidad en SDEMOs para clasificación, centrándonos en problemas de alta dimensionalidad. Se han presentado dos estrategias para abordar el problema:

1. La primera estrategia se centra en el aprendizaje de la granularidad de las variables involucradas en el problema, como manera de obtener una base de datos más adecuada y por lo tanto de mejorar el equilibrio entre precisión y complejidad. Se han presentado dos propuestas con los relativos estudios. En la primera propuesta se ha presentado un SDEMO que aprende granularidades adecuadas y consigue una reducción de la complejidad, prejuzgando ligeramente la precisión. En la segunda propuesta se ha presentado un SDEMO que aprende las granularidades por medio de un algoritmo de discretización difusa. En este caso el método logra una mejora de la precisión del modelo manteniendo la complejidad al mismo nivel o en algunos casos reduciéndola ligeramente.

2. La segunda estrategia aborda el problema por medio de un pre-procesa-miento de los datos de ingreso, mediante la aplicación de técnicas de selección de instancias y la aplicación sucesiva de un SDE. El estudio que se ha realizado ha evidenciado como los modelos obtenidos presentan una complejidad reducida sin afectar demasiado a la precisión. Por otro lado, el análisis de los tiempos de computación ha evidenciado como las técnicas de SCE no resultan útiles desde el punto de vista de la reducción del tiempo de computación si se considera el conjunto de pre-procesamiento y aplicación del método de clasificación, ya que el tiempo total aumenta considerablemente.

Líneas de investigación futuras

Las conclusiones presentadas en las distintas secciones nos indican que el aprendizaje de la granularidad y la selección de instancias representan herramientas validas para la mejora del equilibrio entre precisión y complejidad. De todos modos, existen otras problemas a tener en cuenta para mejorar el equilibrio entre interpretabilidad y precisión.

- Una posible línea de investigación se podría desarrollar considerando en los SDEMOs no sólo medidas de complejidad, sino también medidas de interpretabilidad semántica. Sin embargo, debemos recordar que todavía no hay medidas comúnmente aceptadas por la comunidad científica. Algunos trabajos recientes han propuesto nuevas medidas para describir la interpretabilidad semántica (25, 83) aplicando dichas medidas en problemas de regresión. Una posible línea de investigación futura podría aplicar estas medidas al caso de los SDEMOs para clasificación, con el objetivo de obtener bases de datos más interpretables bajo la perspectiva de la semántica.

- Por lo que concierne la selección de instancias, el estudio efectuado evidencia que en presencia de problemas de alta dimensionalidad las técnicas de SCE no consiguen reducir el tiempo de cálculo total, sino más bien mejorar el equilibrio entre la precisión y la complejidad, por medio de una reducción de las reglas y sin afectar demasiado a la precisión. Por lo tanto, en este ámbito un posible campo de investigación sería el diseño de SCE específicos para ser utilizados en combinación con los SDEs, con el fin de mejorar ambos objetivos al mismo tiempo.

- A partir de los estudios presentados en la memoria es evidente la importancia de aprender las granularidades adecuadas para obtener modelos precisos. Una posible línea de investigación futura podría contemplar el desarrollo de un SDEMO de envoltura (*wrapper*), donde la granularidad y las funciones de pertenencia asociadas se aprenden dentro del proceso evolutivo, basándose en la ejecución de un método rápido de inducción de reglas para evaluar cada nueva base de datos. Adicionalmente, se podría utilizar un algoritmo de discretización difusa para inicializar el cromosoma que será evolucionado por el AEMO.

- Una alternativa al uso de los métodos de selección de instancias como preprocesamiento es el uso de mecanismos de reducción de ejemplos integrados en el AEMO. Un enfoque de este tipo se ha utilizado en problemas de regresión

en (27). Su ventaja es que la elección del conjunto reducido se adapta durante el proceso evolutivo, proporcionando modelos con precisión parecida y permitiendo una reducción drástica del tiempo de calculo.

List of Figures

List of Tables

Introduction

A   Proposal

Nowadays, the development of mathematical models is one of the most relevant topics, since models help in better understanding a system through the simplification of reality. Modeling approaches are used to solve several problems, such as the analysis of systems behaviour, the design of control processes, the prediction of trends, etc.

The scientific and technological development of the last decades brought to the identification of more and more complex problems in several fields, such as engineering, biochemical applications, social science, economic-financial environments, etc. The recurring characteristics that make these systems difficult to be modeled usually deal with high-dimensionality, non-linear behaviors, dynamic components, uncertainty and ambiguity.

To this end, fuzzy systems received a growing attention by the scientific community since their first introduction in 1965 (165). In fact, this kinds of systems present high generalization abilities, robustness with respect to imprecision and vagueness, high expressive power. In particular, the most popular types of fuzzy systems are those based on fuzzy rules, namely, Fuzzy Rule-Based Systems (FRBSs).

The design of FRBSs is usually performed by an expert, by defining the linguistic variables associated to the problem, the set of fuzzy rules and the fuzzy inference system. Nevertheless, to ease this task, several methods have been proposed to automatically design fuzzy systems that match a given set of input-output data. Among them, Evolutionary and in particular Genetic Algorithms (EAs/GAs) has been successfully applied, exploiting their ability in finding near optimal solutions in complex search spaces and their capability to incorporate a priori knowledge. The hybridization between FRBSs and GAs is currently known as Genetic Fuzzy System (GFSs) (58, 97).

In order to consider more than one objective during the design or optimization process, GFSs have been extended to a multi-objective approach: Multi-Objective Evo-

lutionary Algorithms (MOEAs) have been used to design FRBSs with different trade-offs between two or more objectives. In fact, MOEAs represent an effective tool due to their ability in finding a set of compromise solutions when dealing with multi-criteria optimization problems (53, 62). Nowadays, the hybridization between MOEAs and FRBSs is often referred to as Multi-Objective Evolutionary Fuzzy Systems (MOEFSs) (78).

One of the main advantage of FRBSs with respect to other non-linear models is their intrinsic *interpretability*, that is the possibility to understand the model description to a certain extent (40). For this reason, FRBSs are commonly indicated as *gray-box models*. Nevertheless, the interpretability usually conflicts with the model precision, thus the use of MOEAs for the design is particularly suitable, in order to consider both measures at the same time.

FRBSs can be broadly divided into two main categories. The first one includes linguistic fuzzy models (also known as Mamdani FRBSs (134)), which attempt to describe phenomena in a more human-like way, by using the concept of linguistic variable. These kinds of models can be clearly understood by human beings, but sometimes suffer from lack of precision when modeling very complex systems, since the linguistic variables impose strong constraints on the rule structure. In addiction, when dealing with multi-dimensional problems with a non-linear relationship between input and output, the set of rules that describe the system can include redundant, inconsistent and contradictory rules, which decrease both the model precision and interpretability. The second category comprehends precise fuzzy models, which include approximative FRBSs (39, 57) and Takagi-Sugeno-Kang (TSK) (157) models. In these kinds of systems the rules' structure allows the models to achieve more precision, but on the other hand the interpretability is heavily affected since the rules are semantic free.

The trade-off between interpretability and accuracy was pointed out for the first time in the mid 1990s (105) and it has acquired more an more importance in the last decades, as demonstrated by the considerable number of studies carried out on this topic.

In the field of linguistic FRBSs, two kinds of approaches are used to describe the interpretability: complexity-based interpretability approaches, which aim to decrease the complexity of the obtained models (usually expressed by the number rules, variables, antecedents per rules, etc.) and semantic-based approaches, which aim to maintain the semantics associated with the membership functions of each variable (usually expressed by the distinguishability, coverage, etc.) (84).

With respect to semantic interpretability, there still not exists commonly accepted measures, therefore in this thesis we will focus on complexity measures, in particular

taking into account the accuracy-complexity trade-off when dealing with large and high dimensional problems. This kind of problems involves datasets that present a large number of variables (high dimensional problems) and/or example patterns (large scale problems), as they are increasingly appearing in real world.

Nevertheless, when dealing with large and high dimensional datasets, some problems arise: the large number of patterns cause an increase in the computational time needed to perform the evolutionary learning process, due to the time necessary to evaluate the fitness function. Moreover, when the number of variables increases, the search space grow, inducing in most of the cases the generation of more rules. Resulting models can be very complex and the interpretability extremely reduced.

The aim of this thesis is to investigate new methods which can help in improving the accuracy while maintaining or even reducing the complexity of FRBSs, especially when dealing with the high-dimensional and large scale datasets.

B   Objectives

As stated above, the main objective of this thesis is to improve the balance between accuracy and complexity in MOEFSs, when considering problems that involve high dimensional and/or large scale problems. In order to solve the problems previously stated, the following specific objectives have been identified:

- Performing a study of the proposals dealing with MOEFSs, in order to define the state-of-the-art. To this end, we should propose a taxonomy to focus on a certain branch and easily find the related works. This would help to identify the current trends and the problems that are still open and should be addressed in this field.

- Performing a study of the current techniques that cope with the interpretability-accuracy trade-off in linguistic fuzzy models, focusing on the analysis of existent algorithms that obtain the best performance in terms of accuracy and interpretability.

- Taking into account the influence of granularity learning as a way to improve the accuracy-interpretability trade-off when dealing with high dimensional or large scale problems, by developing different algorithms with the aim to improve the linguistic variables' representation. Usually, a standard set of linguistic variables is used by the associated fuzzy rules to describe the behavior of a system. This

approach is simple and straightforward, but the linguistic variables are generated without considering the knowledge we have about a problem. As a consequence, sub-optimal fuzzy partitions for each variable are created, provoking in some cases a reduction of the accuracy. Moreover, the generated models can be so complex that sometimes it is impossible to apply the algorithms. A better approach would be to identify suitable fuzzy partitions for the variables stating from the knowledge we have about a problem, i.e, from the datasets. To this end, two methods have been proposed, which consider two contradictory objectives: to minimize the models' error (precision) and to minimize the number of rules (complexity).

- Performing a study of the combination of instance selection preprocessing with GFSs, in order to study how it affects the trade-off between accuracy and complexity in linguistic FRBSs. The aim is to reduce the number of patterns included in the dataset, by removing all unnecessary data so that two aims can be achieves: 1) obtaining a reduced dataset, thus decreasing the time required by the MOEFS in large scale problems to generate a model, by processing a reduced number of data; 2) obtaining a dataset that better describe a certain phenomenon by removing confusing or noisy or useless data. This approach, when used as a preprocessing method to further apply an algorithm is commonly known as Training Set Selection (TSS).

- Validating the obtained results by comparing the proposed methods towards the existing ones, by using reliable statistic techniques.

C   Summary

To develop the topics proposed, this thesis has been organized into four chapters plus a section containing final conclusions. In the following, the structure and a brief description for each chapter are presented.

In chapter 1 a review of the application of MOEFS is presented. The main contributions on the topic are grouped according to their application and sorted chronologically. A brief description of each proposal is given and finally some open problem related with the topic are discussed. A special attention is paid to the branch of the application of MOEFSs to improve the accuracy-interpretability trade-off.

The remaining part of the thesis is divided into two parts: in the first part two methods to improve the accuracy-complexity trade-off are proposed, based on the learning

of suitable granularities. The first method, presented in chapter 2, aims to learn single granularities by using heuristic measures over multiple granularities. The second method, described in chapter 3, is designed to learn granularities by using a fuzzy discretization algorithm.

In the second part of the thesis, we present a study on the application of instance selection techniques when used in combination with Genetic Fuzzy Rule-Based Classification Systems, with the aim of investigating if they are useful to improve the accuracy-complexity trade-off (chapter 4).

Finally, in chapter 5 some final remarks are pointed out and some future works in this topic are proposed.

Chapter 1

A Review of the Application of Multi-Objective Evolutionary Fuzzy Systems: Current Status and Further Directions

1.1 Introduction

As briefly introduced, MOEFSs have been proposed to generate automatically fuzzy systems that adapt to a given set of data, by exploiting the search ability of MOEAs.

An evolutionary algorithm is basically an algorithm which considers a structure which is codified into an individual called chromosome. This individual is associated to one or more fitness functions and evolved by using concepts inspired to the genetic, such as mutation and crossover. The new individuals that are obtained at each step are evaluated with respect to the fitness functions and the best individuals are promoted. Therefore, evolutionary algorithms perform an exploration and an exploitation of a certain search space, related with the structure codified into the individual. It has been demonstrated that this approach is useful to solve search or optimization problems. This property has been exploited to automatically generate FRBSs, by codifying the FRBS or parts of it into an individual and then associating one or more fitness functions to it.

In the last decades several contributions have been published, proposing the use of MOEAs to generate FRBSs for several applications. In this chapter, the most important of these contributions are gathered and classified by means of a two-level taxonomy. In the first level, the proposals are grouped according to the multi-objective nature

of the problem faced, i.e the type of objectives used by the evolutionary process to promote the best individuals, whereas the second level is based on the type of FRBS components optimized during the evolutionary process.

This taxonomy is useful to easily find existing proposals related to a particular branch and to focus on significant further developments. In the final part of the chapter, some of the problems related to the application of MOEFSs are described, focusing in particular on the ones that represent current trends in the field.

1.2　A taxonomy based on the application of MOEFSs

In this section we describe a two-level taxonomy, proposed to organize the most important contributions that deal with MOEFSs (Figure 1.1). In the first level the contributions are grouped according to the multi-objective nature of the handled problem, i.e. the type of the objectives optimized. The second level groups papers according to the type of FRBS components optimized during the evolutionary process. Both criteria affect the type and the complexity of the search space, and therefore the way in which MOEFSs are applied.
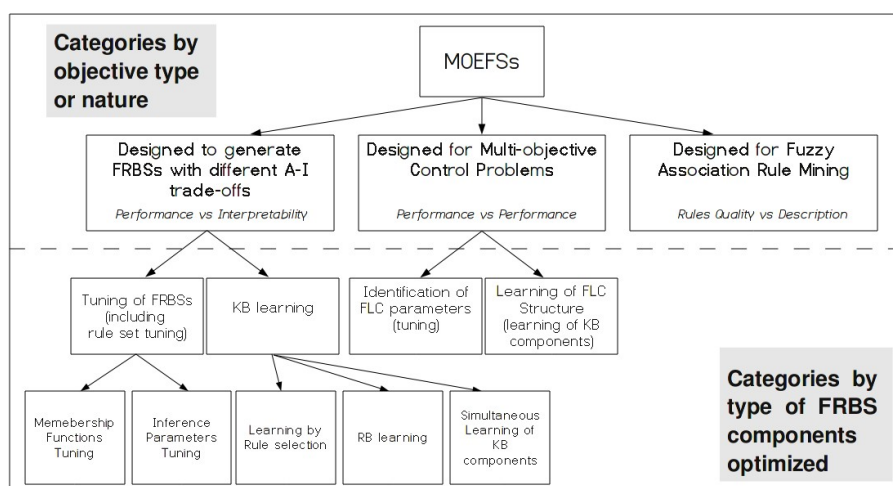
Figure 1.1: A two-level taxonomy based on the type of the objectives optimized (1st level) and on the type of GFS used (2nd level).

In the first level, three main categories have been identified. The first one includes contributions in which MOEFSs are designed to generate FRBSs with different trade-offs between accuracy and interpretability. In this case, at least one of the objectives

is always related to the interpretability of the obtained model, regardless of the problem considered. A considerable number of papers can be found in this group, since interpretability is one of the most important aspects of FRBSs. While the accuracy is difficult to improve, interpretability is easy to obtain, since interpretable models can even be provided by hand. These differences between both types of objectives influence the optimization process.

The second main category gathers contributions in which MOEFSs are applied to multi-objective control problems. The considered objectives strictly depend on the particular kind of problem that is taken into account and usually all of them are related to performance issues of the control system. Therefore, the trade-off and the search space will be different for each problem and dependent on the problem itself.

The third main category groups contributions in which MOEFSs are applied to fuzzy association rule mining. The aim of rule mining is to find a set of fuzzy association rules that reliably represents the knowledge hidden in a data base. In this case, the objectives are used to describe the quality of the obtained rules, i.e. their accuracy and interestingness. To this end, support and confidence are the major factors in measuring the quality of an association rule, although other metrics exist. The aim of the optimization process is not only to improve the general trade-off between objectives for the whole set of rules, but also to obtain a large number of rules, each of them satisfying the objectives to different degrees.

This section illustrates the proposed taxonomy and includes the description of sub-categories for each main category.

1.2.1 MOEFSs designed to generate FRBSs with different accuracy-interpretability trade-offs

One of the main uses of FRBSs is in the approximation of a real system with a fuzzy model, which can be used to explain, simulate or predict the behavior of the original system. Of course, the higher the accuracy is, the more reliable the model will be.

Initially, the interpretability of the obtained models was neglected, since single-objective EAs permit the optimization of only a single metric. The problem of improving accuracy while maintaining or even improving the interpretability of a fuzzy model was first faced in the mid 1990s by Ishibuchi and his group (105) and the comprehensibility of fuzzy models began to be integrated into the optimization process, thanks to the application of MOEAs to fuzzy systems.

Ever since, interpretability has acquired an increasing importance in the field of MOEFSs. Because of its subjectivity, the main problem is to find a shared definition of

9

interpretability and to measure this characteristic in the obtained models, since several issues need to be taken into account to obtain a human-interpretable model.

Over the course of the last decade, several works have analyzed the interpretability problem in FRBSs (54), looking for interpretability measures that could be universally accepted by the research community (61, 95, 151). This effort has continued in recent years, as demonstrated by the review papers presented in (21, 84, 138, 168), which aim to propose a well-established framework to characterize and classify these measures.

Despite this, there are still no commonly accepted measures, and even the terms used in the area (comprehensibility, readability, completeness, consistency, etc.) are confusing and used as synonyms, even if they refer to different concepts. Nowadays, researchers agree on the need to consider two groups of interpretability measures:

- Complexity-based interpretability measures, which are used to decrease the complexity of the fuzzy model (number of rules, number of antecedents in a rule, etc).

- Semantic-based interpretability measures, which are used to preserve the semantics associated with membership functions (distinguishability, coverage, etc) and rules (consistency, etc).

Classically, interpretability indexes have only focused on the former group, when evaluating the overall interpretability of a fuzzy model. On the other hand, the definition of good semantic interpretability measures is still an open problem, since they are strongly affected by subjectivity. To this end, several indexes have been proposed recently (21, 25, 83).

Considering the importance of the Accuracy-Interpretability trade-off for the research community, this first category includes contributions in which MOEFSs are designed to handle this trade-off. that deal with this concept. Due to the huge number of existent works, we organized them into a second-level grouping, according to the taxonomy of GFSs presented in (97) (see Figure 1.1), and thus considering the components of the FRBS that are managed by the optimization process (for further information on the types of FRBSs and KB components see the associated web page `http://sci2s.ugr.es/moefs-review/`):

- Tuning of FRBS components, combined or not with a rule set tuning process: a predefined KB is tuned by the optimization process, i.e. the parameters of the system (shape of membership functions in the Data Base (DB), inference parameters, etc) are modified to obtain more accurate systems. In order to keep the

system simple or to reduce complexity, in some cases a rule selection process, used as a post-processing method, can be integrated in the optimization: from the initial RB only necessary rules are selected. This approach can be considered a rule set tuning process. The contributions belonging to this category are further divided into two subcategories, named Membership Functions Tuning and Inference Parameters Tuning.

- KB learning: papers belonging to this category consider the learning of the DB and/or Rule Base (RB). This group is further divided into three subcategories: learning by rule selection, RB learning and simultaneous learning of KB components. In this case the rule selection process is used to perform a learning of the RB.

The majority of works use a linguistic fuzzy model, since it is the most interpretable type of FRBS. However, there are a small number of works in which interpretability is considered even in a TSK-type FRBS. Because of their particularities, these contributions will be described at the end of this section.

1.2.2 MOEFSs designed for multi-objective control problems

The performance of traditional controllers depends on their accuracy in modeling the system's dynamics. When designing a controller, a first problem appears if the processes are imprecisely described or are controlled by humans, without recourse to mathematical models, algorithms or a deep understanding of the physical processes involved. A further problem concerns how to design adaptive models, i.e. intelligent control systems that involve a learning or adaptation process when system parameters change.

Thus, it can be difficult to identify an accurate dynamic model to design a traditional controller. In these cases fuzzy logic represents a powerful tool to deal with the problem of knowledge representation in an environment of uncertainty and imprecision. Furthermore, in control system design, there are often multiple objectives to be considered. These objectives are sometimes conflicting, causing an inevitable trade-off among them and no single design solution emerges as the best with respect to all objectives. These considerations have led to the application of MOEAs in the design of Fuzzy Logic Controllers (FLCs).

The design of an FLC includes obtaining a structure for the controller and the corresponding numerical parameters. MOEAs can manage these problems by encoding both structure and parameters in one chromosome that represents the whole FLC.

Therefor, in this second group, works will be explained considering the following two categories (79):

a) identification of controller parameters and/or rules (e.g. tuning of membership function parameters, rule selection as a post-processing method);

b) learning of controller structure (e.g. learning of the RB).

At the end of the corresponding section, some works are described that represent a hybridization of MOEAs, fuzzy logic and neural networks.

1.2.3   MOEFSs designed for fuzzy association rule mining

The hybridization of MOEAs and fuzzy systems permits automatic knowledge extraction from data, therefore data mining problems are one of the most important application domains for MOEFSs. Data mining has been treated as a synonym of Knowledge Discovery in Databases (KDD) (75, 146), although it is a step of KDD. Data mining techniques usually fall into two categories: predictive or descriptive.

A predictive approach focuses on accuracy in predictive ability and generates models that can be used to predict explicit values, based on patterns determined from known results. In prediction, a user may not care whether the model reflects reality as long as it has predictive power. One of the methods used in predictive models is supervised learning, which can create a function from training data, used to predict the output value for any valid input object. The predictive approach is applied in classification and regression and in some cases it can also be used in control problems.

On the other hand, the descriptive approach focuses on understanding the implicit data-generating process, searching for interesting patterns in existing data, without having any predefined target. The method used in this model is usually unsupervised learning, which differs from supervised learning in that there is no a priori output to train the model. This method is mainly applied to models that work with associative rules.

Finally, in some cases there are data mining applications demanding some degree of both predictive and descriptive approaches. A method which combines the mixed approach between descriptive and predictive is Subgroup Discovery (128).

A possibility to represent knowledge extracted with data mining techniques is by means of association rules (167), whose basic concept is to discover meaningful associations between different pairs of sets of attribute values. For example, the presence of a value of some set in a database element implies the presence of another value in

another set. Since fuzzy systems can deal with imprecise knowledge, they can be successfully applied in the representation of this knowledge using fuzzy association rules (65).

In mining fuzzy association rules, the objectives are based on the quality of the extracted rules: these rules should be precise, general or specific enough, interesting, etc. Due to the large amount of metrics, MOEAs have been used successfully to mine fuzzy association rules.

The works included in this group mainly use a descriptive approach, i.e. description sets focused on making the data comprehensible and interpretable. Additionally, some works using the Subgroup Discovery approach will be described.

1.3 MOEFSs designed to generate FRBSs with different accuracy-interpretability trade-offs

The problem of improving accuracy while maintaining or even improving the interpretability of a fuzzy system is widely acknowledged in the community of MOEFSs, its presence noted in the mid 1990s (105). It is known that there is a point at which it is not possible to improve both the accuracy and interpretability of a fuzzy system at the same time. Therefore, in this framework an MOEA aims to find a set of feasible fuzzy systems with different trade-offs between accuracy and interpretability (see Figure 1.2).

Hereinafter, we describe contributions in which MOEFSs are designed to generate FRBSs with a good trade-off between Accuracy and Interpretability, and we group
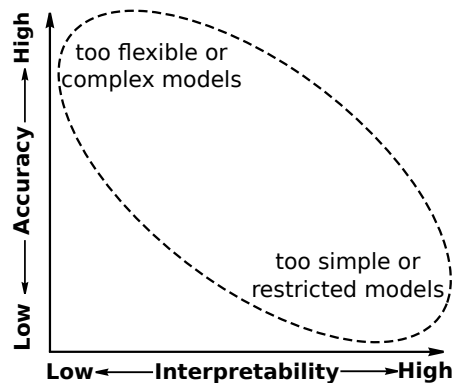
Figure 1.2: Accuracy-interpretability trade-off.

them by following the second level of the taxonomy presented in section 1.2 and explained in section 1.2.1.

1.3.1 Approaches to performing tuning

MOEAs can be used to perform the genetic tuning of FRBS components. Genetic tuning is applied as a post processing method, once the RB has been obtained, to refine the KB parameters (8, 57, 98, 121) or to adapt the parameters of the inference engine (17), therefore the works belonging to this category have been divided into two subcategories: Membership Functions Tuning and Inference Parameters Tuning. Moreover, in some cases the tuning process can be combined with a rule selection process, to improve the interpretability of the obtained model by removing unnecessary rules. This approach can be seen as a rule set tuning process, since it is applied to a previously defined RB.

1.3.1.1 Tuning of membership functions

An example of membership functions tuning process combined with a rule selection process can be found in (11), in which the authors present a post-processing algorithm to improve the performance of linguistic FRBSs for regression problems. A specific MOEA is used to achieve a good balance between accuracy and complexity, improving accuracy by the tuning of membership functions, while reducing complexity by removing unnecessary rules. The proposed algorithm, called Accuracy-Oriented Strength Pareto Evolutionary Algorithm 2 ($SPEA2_{ACC}$), is based on a particular modification of SPEA2 (169) and takes into account two objectives: accuracy, expressed by computing the Mean Squared Error (MSE) and complexity, expressed as the number of selected rules. Rule selection and the tuning of membership functions are performed together, by coding both of them in the same chromosome. The $SPEA2_{ACC}$ concentrates the search on the Pareto zone that have the most accurate solutions with the least number of possible rules.

The same algorithm is extended in (82), in which six algorithms are considered to perform a rule selection from a given fuzzy rule set along with the tuning of the membership function parameters applied to regression problems. The Nondominated Sorting Genetic Algorithm II (NSGA-II) (64) and SPEA2 are used, along with two versions of NSGA-II proposed for general use, which concentrate the search on the Pareto knees. Two MOEAs for specific application to this concrete problem are applied. The first one is the $SPEA2_{ACC}$ proposed in (11), the second one is its extension,

$SPEA2_{ACC^2}$. All these algorithms improve two objectives: MSE and the number of rules.

In (148) a hybrid method for the identification of a Pareto-optimal Fuzzy Rule-Based Classifier (FRBC) is presented. The initial population is created in two steps: firstly a decision tree, generated through the classical C4.5 algorithm, is transformed into an FRBC. In this way, relevant variables are selected and an initial partition of the input space is performed. Afterwards, the remaining population is created by randomly replacing some parameters of the initial FRBC. The tuning process is performed by applying the well-known NSGA-II, with polynomial mutation and simulated binary cross-over (SBX) (63) as genetic operators. Three objectives are minimized: the number of misclassified patterns, the number of rules and the total number of conditions in the rules. Each chromosome codifies an FRBC, including antecedents of the rules and parameters of the fuzzy sets.

An adaptation of the previous framework can be found in (147), in which the authors use FRBCs to model a bioareosol detector. As the metrics of accuracy, true positive (TP) and false positive (FP) rates were used instead of the commonly used misclassification rate, because of the uneven misclassification costs and class distributions of the collected data. Interpretability of the model is also a requirement, since it allows the bioareosol detector to be subsequently adjusted. Therefore, NSGA-II is applied to find FRBCs with a good trade-off between objectives. The FP rate and the complement of the TP rate measure the accuracy, whereas transparency of fuzzy partitions is used for interpretability. The latter objective is expressed by the sum of three interpretability measures, the length of overlap and the length of discontinuity between fuzzy sets, proposed by Kim (126), and the middle value penalty.

Another contribution to the tuning of DB parameters of FRBSs for regression problems can be found in (32). In this work the concept of context adaptation is used: context adaptation is a tuning process that exploits context-specific information to adapt a context-free model to a context-adapted FRBS. NSGA-II has been applied to the tuning of DB parameters, to maximize both the accuracy and interpretability of a linguistic FRBS. A novel index is therefore proposed, to provide a measure of interpretability, considering ordering, coverage and distinguishability. The proposed index and the MSE are used as objectives of the EA.

The tuning of membership function parameters is tackled again in (83), in the framework of linguistic fuzzy models for regression problems. A novel relative index is proposed to help preserve the semantic interpretability of FRBSs while the tuning of membership functions is performed. The index, called GM3M, is the aggregation

of three metrics that aim to maintain the original meanings of the membership functions as much as possible. In this work a tuning of membership function parameters is combined with a rule-selection mechanism, in order to also reduce the complexity of the fuzzy models. Therefore, an improved specific version of the well-known SPEA2, namely SPEA2-SI, including incest prevention and restarting, is proposed and three objectives are considered: accuracy maximization, semantic interpretability maximization and complexity reduction.

1.3.1.2 Tuning of inference parameters

Few works have taken into account the tuning of the inference engine (17). In (135) a method is presented to concurrently learn the fuzzy inference operators and the RB of linguistic FRBSs, in order to obtain simpler, more compact yet still accurate linguistic fuzzy models. To this end, two MOEAs were used and adapted, SPEA2 and NSGA-II. The proposed MOEAs generate a set of FRBSs with different trade-offs between interpretability and accuracy: the two objectives are expressed by the number of rules and the MSE, respectively.

In (136) an approach is proposed to tackle the interpretability-accuracy trade-off in linguistic FRBSs with adaptive defuzzification. Adaptive defuzzification methods improve the accuracy of the system, but cause a loss of interpretability and increase complexity, due to the introduction of parameters in the defuzzification operator and weights associated with each rule. To quantify the interpretability of FRBSs with adaptive defuzzification, a novel index is proposed, which is the aggregation of two metrics: number of rules with weight and average number of rules triggered by each example. Afterwards, an adaptation of NSGA-II is exploited in order to obtain a set of accurate and interpretable linguistic fuzzy models with adaptive defuzzification. Three objectives are minimized: the MSE, the number of final rules in the system and the proposed interpretability index.

1.3.2 Approaches to performing KB learning

Besides the tuning of FRBS components, another possibility is to learn the KB or a part of it by means of MOEAs. We identify three approaches within this category: learning by rule selection, RB learning and the simultaneous learning of KB components.

1.3.2.1   Approaches to learning by rule selection

The first contributions to the application of MOEAs to linguistic FRBS generation
with a good interpretability-accuracy trade-off were proposed by Ishibuchi's group on
multi-objective rule selection applied to learning. In their earlier works (104, 105), the
authors use 1st-generation MOEAs (i.e. MOEAs without elitism) to perform a rule
selection on an initial set of candidate rules as a two-stages learning process: candi-
date rule set generation and multi-objective rule selection. In the second stage they
consider two different objectives: maximization of the number of correctly classified
training patterns and minimization of the number of selected rules, therefore the ob-
tained classification systems consist of a small number of linguistic rules. In (104)
this rule selection method is extended to the case of classification problems with many
continuous attributes, by using a prescreening procedure of candidate rules based on
the number of antecedent conditions of each rule.

To better control the dimensionality problem, the authors add a third objective in
(115). An MOEA is used to extract a small number of fuzzy rules from numerical data,
taking into account three objectives: to maximize the number of correctly classified
training patterns, to minimize the number of fuzzy rules and to minimize the total
number of antecedent conditions. The MOEA presented in (105) is extended to a
Multi-Objective Genetic Local Search (MOGLS) algorithm, in which a local search
procedure adjusts the selection process. Moreover, it is combined with a learning
algorithm to obtain rule weights.

In (107), two multi-objective genetic-based approaches are applied, to obtain FR-
BCs with a good trade-off between accuracy and complexity. The first approach was
presented in (105), while the second one is a hybrid multi-objective Genetics-Based
Machine Learning (GBML) algorithm, a hybridization between the Michigan (31, 101)
and Pittsburgh (155) approaches. It considers the same three objectives as the previous
model (115).

The same multi-objective GBML algorithm is used in (111), but in this contribution
it is implemented taking advantage of the well-known NSGA-II and again consists
of a hybrid version of the Michigan and Pittsburgh approaches: each fuzzy rule is
represented by its antecedent fuzzy sets as an integer string of fixed length, then the
concatenation of these strings represents an FRBC. The objectives remain the same as
in (107).

In (110), NSGA-II is applied to the design of FRBCs belonging to the accuracy-
complexity Pareto optimal front. The accuracy of each classifier is measured as the
number of correctly classified training patterns, whereas the complexity is computed

as the number of fuzzy rules and the total number of antecedent conditions. Finally, an ensemble classifier (also called a multi-classifier) is designed by combining non-dominated FRBCs and its performances are analyzed by performing computational experiments on six benchmark datasets taken from the UCI machine learning repository. The authors observe that the effect of combining several FRBCs is problem dependent and that an ensemble of classifiers with high diversity usually has better performances.

1.3.2.2 Approaches to performing RB learning

Most of the approaches proposed to automatically learn the KB from numerical information focus on RB learning using a predefined DB.

In (152) an MOEA is used to generate FRBCs with a good trade-off between the complexity of the rule systems and their reflection of the data. This MOEA uses a measure based on Area Under the receiver operating characteristic Curve (AUC) to determine how well the classifier reflects the data. Moreover, some concepts taken from SPEA2 are included: the fitness assignment of SPEA2 is used to avoid premature convergence and an external archive is maintained to store the best individuals from all the solutions considered. In addition, a tailor-made representation scheme is used to preserve the comprehensibility of the rule systems and a self-adaptation mechanism is included to reduce the number of free parameters. Three objectives are optimized: the accuracy, expressed as a measure based on the AUC, and complexity, computed as the number of rules and conditions.

An example of rule learning for regression problems is presented in (51), in which the authors propose a modified version of the well-known (2+2)Pareto Archived Evolution Strategy (PAES), called (2+2)M-PAES, introduced in (129). Unlike classical (2+2)PAES, which only uses mutation to generate new candidate solutions, (2+2)M-PAES exploits both crossover and mutation. This approach considers a predefined DB uniformly distributed and enables a large set of RBs to be derived, concurrently minimizing the accuracy and the complexity. The accuracy is computed as the Root Mean Squared Error (RMSE), whereas complexity is measured as the sum of the conditions which compose each of the antecedents of the rules included in the FRBS.

In (69), the accuracy-interpretability trade-off is considered in the context of imbalanced classification problems. Usually, the accuracy of a classifier is measured as the percentage of correct classification, but this objective might not be suitable for problems characterized by highly imbalanced distributions of patterns. In this proposal, authors applied the well-known NSGA-II to provide a set of binary FRBCs with a

good trade-off between complexity and accuracy. In this case, complexity is computed as the sum of the conditions in the antecedents of the classifier rules, whereas accuracy is expressed in terms of two objectives, sensitivity and specificity. These express how well the system classifies patterns belonging to the positive and the negative class, respectively.

1.3.2.3 Approaches to simultaneous learning of KB components

KB learning of linguistic FRBSs aims to learn the DB and RB concurrently. This approach tackles a very large search space, which is also difficult for EAs to handle. Some approaches have been proposed to learn concurrently the overall RB and DB.

In (55) the authors proposed a method for feature selection and DB learning, to obtain FRBCs composed of a compact set of comprehensible fuzzy rules with high classification ability. The DB learning involves both the number of labels for each variable (granularity) and the form of each fuzzy membership function. A non-linear scaling function is used to adapt the fuzzy partition contexts for the corresponding granularity. This approach uses an MOEA to evolve the DB and considers a simple generation method to derive the RB. The MOEA has two goals: to improve the accuracy, by minimizing the classification error percentage over the training dataset, and to obtain a compact and interpretable KB, by penalizing fuzzy classifiers with large numbers of selected features and high granularity. The second objective is expressed by the product of the number of selected variables and their averaged granularity.

In (9) the authors proposed a technique to concurrently perform the RB identification and the DB learning of fuzzy models for regression problems. Two MOEAs are exploited to generate a set of linguistic FRBSs with different trade-offs between accuracy and interpretability. The proposed approach can learn RBs and membership function parameters of the associated linguistic labels, therefore the search space increases considerably. To manage the size of the search space, the linguistic two-tuple representation model (99) is included, which uses a reduced number of parameters to perform the symbolic translation of labels. The first MOEA is (2+2)M-PAES and it is compared with the well-known NSGA-II. Two objectives are considered: the MSE and the number of antecedents activated in each rule.

The same (2+2)M-PAES is exploited in (23) to generate linguistic FRBSs for regression problems, with different trade-offs between complexity and accuracy. The presented approach aims to learn the RB and the granularity of the uniform partitions defined by the input and output variables concurrently. Consequently, the concepts of

virtual and concrete RBs are introduced: the former is defined by uniformly partitioning each linguistic variable with a fixed maximum number of fuzzy sets. The latter takes into account, for each variable, the number of fuzzy sets determined by the specific partition granularity of that variable. RBs and membership function parameters are defined by the virtual partitions and, whenever a fitness evaluation is required, they are mapped to the concrete partitions. Two objectives are considered: the accuracy of the FRBSs, measured as the MSE, and their complexity, computed as the number of propositions used in the antecedent of the rules contained in the concrete RB.

This work is extended in (24), in which the same MOEA is used to concurrently learn not only the RB and partition granularity, but also membership function parameters. The same approach is presented in (26), where a partition integrity index is proposed as a third objective. This index measures to what extent a partition is different from an initially interpretable one. Furthermore, in (25) a novel interpretability index is proposed, which combines RB complexity with DB integrity.

In (41) a specific MOEA, called Pitt-DNF, is proposed to obtain FRBSs for regression problems. The Pittsburgh approach is chosen, therefore each chromosome encodes a complete set of fuzzy rules. Antecedents of rules are represented in Disjunctive Normal Form (DNF), i.e. each input variable can take an OR-ded combination of several linguistic terms as a value and the different input linguistic variables are combined by an AND operator. Nevertheless, the authors wrongly call Conjunctive Normal Form these kinds of fuzzy rules. This representation provides a high degree of compactness and improves the interpretability of fuzzy models, but the combination of the Pittsburgh approach with DNF-type fuzzy rules causes some problems to generate the rules themselves. The proposed learning algorithm, based on NSGA-II, has been developed to avoid the generation of DNF-type fuzzy rule sets with these problems and it gives a set of solutions with different trade-offs between complexity, computed as the number of DNF rules, and accuracy, measured by the MSE. One crossover operator and two mutation operators were specifically designed to take into account the particular representation of fuzzy rules, thus avoiding inconsistency, redundancy, over-generality and incompleteness in fuzzy rules.

In (149) an MOEA is proposed to learn the granularities of fuzzy partitions, tune the membership function parameters and learn the fuzzy rules of a linguistic FRBS for regression problems. A two-step evolutionary approach is applied: the fuzzy models are initialized using a method that combines the benefits of an ad-hoc RB generation algorithm and decision-tree algorithms, with the aim of reducing the search space. The initial population is then optimized by an MOEA that reduces the number of rules, rule

conditions, membership functions and input variables. The MOEA is based on NSGA-II and the original genetic operators are replaced with new ones that take into account dynamic constraints to ensure the transparency of fuzzy partitions. Two objectives are optimized: accuracy, expressed as the MSE, and complexity, computed as the total rule length (number of active rule conditions).

In (12), the authors propose a two stage approach to obtain linguistic KBs in classification problems based on the multi-objective fuzzy rule selection presented in (115), by also including a lateral tuning (8) within the same process and by considering the same three objectives: to maximize the number of correctly classified training patterns, to minimize the number of fuzzy rules and to minimize the total number of antecedent conditions. The first stage determines appropriate granularities for the DB and a set of candidate rules. The second stage performs multi-objective rule selection and tuning, based on using NSGA-II to obtain the final RB and the appropriate DB parameters.

A recent proposal can be found in (10), where the authors focus on the scalability issue of linguistic FRBSs in 17 regression problems. The first stage uses an improved MOEA (based on SPEA2) to perform an embedded genetic DB learning including feature selection, granularities and the reduced lateral displacement of fuzzy partitions in order to control the dataset dimensionality and obtain a reduced KB. For each DB definition an ad-hoc RB is derived by adding a cropping mechanism to avoid large RBs and to reduce the required computation time. Two minimization objectives are used: MSE and number of rules. Finally, a post-processing stage for fine tuning and rule selection is applied to the obtained KBs using the same objectives. A speeded-up version of a previous MOEA, namely Exploration-Exploitation based SPEA2 ($SPEA2_{E/E}$), is presented by including a new approach to fast fitness estimation which only uses a small percentage of the training data. Since this mechanism is proposed for any kind of EA, authors also include it in the first stage in order to address the problem of large datasets (many-instance datasets).

In (20), Alonso et al. propose embedding the High Interpretable Linguistic Knowledge (HILK) heuristic method (22) in a three-objective evolutionary algorithm, with the aim of getting a good accuracy-interpretability trade-off when building FRBCs. The well-known NSGA-II algorithm is employed, using two point crossover and Thrift's mutation (159). Three criteria are optimized: accuracy, by maximizing the right classification rate; readability, by minimizing the total rule length; comprehensibility, by minimizing the average number of rules fired at the same time (Average Fired Rules - AVR). Each chromosome includes a number of genes equal to the number of input variables and each gene represents the number of linguistic terms defined for the related

input. Recently, this proposal has been extended in (33) by considering a novel comprehensibility index called the Logical View Index (LVI), which estimates how much an RB satisfies logical properties. In this novel version, the AVR is substituted by the LVI as a better FRBCs comprehensibility measure. Finally, in (34), both LVI and AVR indexes are considered. The proposed evolutionary framework is used to set up two independent experimental sessions with two objectives: classification rate vs. AFR and classification rate vs. LVI. The study aims to find possible relationships between AFR and LVI, showing that the AFR minimization implies the LVI minimization, while the opposite is not verified.

1.3.3   Approaches that deal with TSK FRBSs

TSK fuzzy models provide a powerful tool for modeling complex non-linear systems, as multiple sub-models (typically linear models) are combined to describe the global behavior of the system. The resulting model is often more difficult to interpret, and few works can be found on this topic.

In (161) a technique based on a hierarchical MOEA (66), derived from MOGA (80), is proposed to construct TSK fuzzy models (157) from data, considering both their accuracy and interpretability. The initial model is generated through a two-step procedure: a fuzzy clustering method is used to preprocess the training data and to construct the rule antecedents, then the Recursive Least Square (RLS) method is applied to determine the consequent rule. Finally, the hierarchical MOEA is exploited to obtain the optimized fuzzy models, for regression problems. A hierarchical chromosome formulation is used, so that it can perform the simultaneous optimization of rule antecedents and number of rules, whereas consequents are obtained with the RLS method. A two-level hierarchical structure is used: control genes and parameter genes. Considering that there are two types of genes in the chromosome, a multi-point crossover is applied for control genes, whereas for the parameter genes which are represented in real numbers, BLX-α crossover is applied. During the optimization an interpretability-driven RB simplification is applied, to reduce the search space. Five objectives are optimized: the MSE for accuracy, the total number of fuzzy sets and the number of fuzzy rules for compactness, a purposely-defined aggregate index for both completeness and distinguishability, and finally an appropriate equation for non-redundancy.

In (164), a novel coevolutionary algorithm (143) is proposed to improve the performance of TSK fuzzy systems in regression problems. This algorithm is called the Pareto Multi-Objective Cooperative Coevolutionary Algorithm (PMOCCA). The

fuzzy system is decomposed into two species: antecedents of fuzzy rules and parameters of fuzzy sets. To obtain a good initial fuzzy system, a modified fuzzy clustering algorithm is used. Afterwards, the PMOCCA and some interpretability-driven simplification techniques are used to evolve the initial fuzzy system with three objectives: accuracy of the system, the number of fuzzy rules and the number of antecedents in each fuzzy rule.

The problem of the trade-off between accuracy and complexity in TSK fuzzy systems is also faced in (93), in which a specific version of NSGA-II is proposed to determine a Pareto-optimal set of fuzzy models for regression problems. In particular, two competing objectives are addressed: the accuracy, measured by the normalized RMSE, and the complexity, expressed by the number of fuzzy rules. The specialization of the algorithm is obtained first by using several heuristics to obtain a good initialization of the population, and second by designing crossover and mutation operators specific to the problem.

In (92) a Multi-Objective Neuro-Evolutionary Algorithm (MONEA) is proposed to obtain a parameter estimation of TSK fuzzy models for regression problems. Neural network based techniques and ad-hoc techniques for interpretability improvement are included in the MOEA to increase the efficacy of the algorithm: the fuzzy model is defined by a radial basis function neural network (47, 71). The number of neurons in the hidden layer of the neural network is equal to the number of rules in the fuzzy model and the firing strength of the ith neuron in the hidden layer matches the firing strength of the ith rule. The neurons in the output layer perform the computation for the function described in the consequents of the fuzzy model. The MONEA considers four objectives: accuracy, computed as the MSE, transparency, for which the similarity among distinct fuzzy sets is considered, and compactness, expressed by the number of rules and the number of antecedents in the fuzzy model.

Another proposal can be found in (94), in which the authors used a hybrid technique to optimize the structure of TSK fuzzy systems for regression problems. Firstly, a backpropagation algorithm is applied to optimize the membership function parameters and the parameters of fuzzy rules. In a second phase, NSGA-II is used to perform a fine tuning of parameters and to select the optimal number of fuzzy rules. The algorithm considers two objectives: the system's accuracy, computed as the MSE, and complexity, defined by the number of active fuzzy rules in the RB.

In (50), a regression problem named the ocean color inverse problem is approached by using the (2+2)M-PAES to optimize TSK FRBSs. The evolutionary optimization roughly identifies the structure of the fuzzy models, then a tuning process is performed: TSK FRBSs are implemented as an artificial neural network and by training the neural

network, the parameters of the fuzzy model are adjusted. The result is a set of fuzzy models with different trade-offs between accuracy and complexity.

A recent contribution has been presented in (52), in which the authors first analyze the time complexity for both the generation and the evaluation of TSK FRBSs. Since the identification of the rule consequent parameters is one of the most time-consuming phases in TSK FRBS generation, a simple and effective technique is proposed for speeding it up. Then, this approach is included in the optimization process of the structure of TSK fuzzy systems for regression problems. (2+2)M-PAES is applied and one-point crossover and three appropriately defined mutation operators are used. Two objectives are optimized: the MSE as a measure of accuracy and the total number of conditions different from *don't care* as a measure of complexity.

1.3.4 A summary of MOEFSs designed to generate FRBSs with different accuracy-interpretability trade-offs

In order to give an overview of the contributions described so far, Table 1.1 presents a summary of the works dealing with the accuracy-interpretability trade-off of FRBSs. Papers are grouped considering the components of the KB that are optimized and within each group they appear in chronological order. For each paper the type of FRBS approach is shown, together with the kinds of rules. The number and type of the objectives are reported together with the name of the MOEA, its generation type and the kind of proposal (novel, general use or based on a previous MOEA). The repetition of the objective type means the presence of two different optimized measures for the same objective. In the last column, the type of application problem is briefly described.

Except for some earlier works, the greater part of the approaches use a second-generation MOEA (i.e. MOEA with elitism) to tackle the accuracy-interpretability trade-off: in fact, the introduction of the concept of elitism is essential for the convergence of the algorithms. Moreover, the concept of interpretability becomes more complex and complete over the years: earlier contributions considered interpretability only in terms of complexity, whereas more recently, semantic interpretability has been studied in depth and included in the optimization process.

Looking at the FRBS approach, it is evident that Mamdani FRBSs are used more than TSK ones, probably due to the intrinsic interpretability of the Mamdani model. Finally, we can remark that earlier contributions scarcely considered the problem of learning the whole KB, which is progressively considered more often in the latter contributions.

Table 1.1: Summary of the proposals on MOEFSs designed to generate FRBSs with good Accuracy-Interpretability trade-offs.

	A-I trade-off			FRBS approach		Objectives		MOEA			Problem type
	Authors	Ref.	Year	Rules	Type	#Obj.	Type	Name	Gen.	Type	
DB TUNING (INCLUDING RULE SET LEARNING) — MEMBERSHIP FUNCTIONS TUNING	Wang et al.	(161)	2005	TSK	LING. ∗	5	A+C+C+S+S	MOHGA	1st	I ∘	REG.
	Alcalá et al.	(11)	2007	MAM.	LING.	2	A+C	$SPEA2_{ACC}$	2nd	I ●	REG.
	Gonzalez et al.	(93)	2007	TSK	SCAT.	2	A+C	NoN.	2nd	I †	REG.
	Gomez et al.	(92)	2007	TSK	SCAT.	4	A+C+C+S	MONEA	2nd	N	REG.
	Pulkkinen et al.	(148)	2008	MAM.	LING.	3	A+C+C	NSGA-II	2nd	G	CLAS.
	Pulkkinen et al.	(147)	2008	MAM.	LING.	3	A+A+S	NSGA-II	2nd	G	CLAS.
	Guenounou et al.	(94)	2009	TSK	LING. ∗	2	A+C	NSGA-II	2nd	G	REG.
	Gacto et al.	(82)	2009	MAM.	LING.	2	A+C	VARIOUS	2nd	G	REG.
	Botta et al.	(32)	2009	MAM.	LING.	2	A+S	NSGA-II	2nd	G	REG.
	Gacto et al.	(83)	2010	MAM.	LING.	3	A+C+S	SPEA2-SI	2nd	I ●	REG.
I.P. TUNING	Marquez et al.	(135)	2009	MAM.	LING.	2	A+C	VARIOUS	2nd	I †●	REG.
	Marquez et al.	(136)	2010	MAM.	LING.	3	A+C+S	NoN.	2nd	I †	REG.
KB LEARNING — LEARN. BY RULE SEL.	Ishibuchi et al.	(104, 105)	1997,1998	MAM.	LING.	2	A+C	NoN.	1st	N	CLAS.
	Ishibuchi et al.	(107)	2001	MAM.	LING.	3	A+C+C	GBML	1st	N	CLAS.
	Ishibuchi et al.	(115)	2004	MAM.	LING.	3	A+C+C	MOGLS	1st	N	CLAS.
	Alcalá et al.	(12)	2010	MAM.	LING.	3	A+C+C	NSGA-II	2nd	G	CLAS.
	Ishibuchi et al.	(110)	2006	MAM.	LING.	2	A+C	NSGA-II	2nd	G	CLAS.
	Ishibuchi et al.	(111)	2007	MAM.	LING.	3	A+C+C	GBML	2nd	I †	CLAS.
RB LEARNING	Setzkorn et al.	(152)	2005	MAM.	LING.	3	A+C+C	NoN.	2nd	I ●	CLAS.
	Cococcioni et al.	(51)	2007	MAM.	LING.	2	A+C	(2+2)M-PAES	2nd	I ★	REG.
	Xing et al.	(164)	2007	TSK	LING. ∗	2	A+C	PMOCCA	2nd	N	REG.,Ts.
	Ducange et al.	(69)	2010	MAM.	LING.	3	A+A+C	NSGA-II	2nd	G	IMB. CLAS.
SIMULTANEOUS LEARNING OF KB COMPONENTS	Cordón et al.	(55)	2003	MAM.	LING.	2	A+C	NoN.	1st	N	CLAS.
	Cococcioni et al.	(50)	2008	TSK	SCAT.	2	A+C	(2+2)M-PAES	2nd	I ★	REG.
	Alcalá et al.	(9)	2009	MAM.	LING.	2	A+C	(2+2)M-PAES	2nd	I ★	REG.
	Antonelli et al.	(23)	2009	MAM.	LING.	2	A+C	(2+2)M-PAES	2nd	I ★	REG.
	Antonelli et al.	(24)	2009	MAM.	LING.	2	A+C	(2+2)M-PAES	2nd	I ★	REG.
	Casillas et al.	(41)	2009	DNF-RULES	LING.	2	A+C	NoN.	2nd	I †	REG.
	Pulkkinen et al.	(149)	2010	MAM.	LING.	2	A+C	NoN.	2nd	I †	REG.
	Alonso et al.	(20)	2010	MAM.	LING.	3	A+C+S	NSGA-II	2nd	g	CLAS.
	Cannone et al.	(33)	2011	MAM.	LING.	3	A+C+S	NSGA-II	2nd	g	CLAS.
	Cannone et al.	(34)	2011	MAM.	LING.	2	A+S AND A+S	NSGA-II	2nd	g	CLAS.
	Cococcioni et al.	(52)	2011	TSK.	SCAT.	2	A+C	(2+2)M-PAES	2nd	I ★	REG.
	Antonelli et al.	(26)	2011	MAM.	LING.	3	A+C+S	(2+2)M-PAES	2nd	I ★	REG.
	Antonelli et al.	(25)	2011	MAM.	LING.	2	A+(C+S)	(2+2)M-PAES	2nd	I ★	REG.
	Alcalá et al.	(10)	2011	MAM.	LING.	2	A+C	NoN.	2nd	I †	REG.

I.P.=Inference Parameters, MAM.=Mamdani, TSK=Takagi-Sugeno-Kang, LING.=Linguistic, SCAT.=Scatter, ∗In the antecedent;
A=Accuracy, C=Complexity, S=Semantic aspects;
NoN.=No name, N=Novel algorithm, I=Improved version, G=General use;
CLAS.=Classification, REG.=Regression, Ts.=Time Series, IMB.=Imbalanced;
†NSGA-II based, ★PAES based, ∘MOGA based, ●SPEA2 based.

1.4 MOEFSs designed for multi-objective control problems

FLCs are one of the most common applications of fuzzy logic. An FLC includes a set of linguistic control rules related by the dual concepts of fuzzy implication and the compositional rule of inference (132). No fixed process for designing a fuzzy controller exists and the appropriate fuzzy parameters have to be chosen on the basis of an experimental study of the control objective. To overcome this difficulty, the application of EAs was proposed for the design of FLCs (7, 85, 100). Two problems arise during this process: the first issue concerns how to establish the structure of the controller; secondly, the numerical values of the controller's parameters have to be chosen.

Many contributions can be found in the literature on the use of EAs to obtain the optimal design of FLCs, both for tuning and learning tasks. Most of them take into account only one performance objective. The first multi-objective approaches were carried out by combining several performance objectives into a single one, by using a weighting approach. Afterwards, more objectives were included in the optimization process with the aim of considering not only different performance measures, but also characteristics such as time constraints, robustness and stability requirements, comprehensibility and the compactness of the obtained controller. EAs have been used either for off-line or on-line design of FLCs, although in the latter case the computation time is sometimes a critical issue. Further information on EAs applied to FLCs can be found in (117).

In the following we will analyze the existing works on the application of MOEFSs to Fuzzy Control, considering both categories presented in section 1.2.2. They are controller parameters' identification and the learning of controller structure. Unless otherwise specified, the contributions use a Mamdani-type FLC.

1.4.1 Controller parameters' identification

The first approach aims to modify the parameters that affect the controller's performance once an initial design of the FLC is established. Tuned parameters can be the scaling factors for each variable, the shape of fuzzy sets representing the meaning of linguistic values and the selected IF-THEN rules. This approach permits the reduction of the computational load required, since the search space is smaller than the one considered when learning all the components together. Nevertheless, since the parameters

and structure of fuzzy models are strictly related, the obtained solutions are affected by the initial system definition.

One of the first works on the use of a first-generation MOEA for the optimization of an FLC is presented in (2). A Mamdani-type fuzzy system is designed for the vibration control of a civil engineering structure in seismic zones. Consideration of the building performance includes both the safety and the comfort level of the user. The former issue is achieved by minimizing the peak displacement, the latter one by minimizing the peak acceleration. The trade-off between the two objectives is handled using a two-branch tournament GA that provides a set of Pareto optimal solutions and optimizes the parameters of the input and output membership functions. Each membership function is represented by a generalized bell shaped function defined by three values. One-point crossover is employed and the mutation is performed on a bit-by-bit basis, with a certain probability.

A similar approach is undertaken in (3), where a hybrid control system (using active and passive control strategies) is proposed for the structural vibration control of buildings. A tuned mass damper and an active mass driver are used as respective the passive and active control components of the hybrid control system. To control the active mass driver, an FLC is used and the two-branch tournament GA is applied to the optimization of the parameters of the input and output membership functions. In (4) a further objective is added. A three-branch tournament GA is used this time, in which the minimization of peak displacement, acceleration and rotation of the structure about its vertical axis are considered as the three objective functions.

In (5), the same approach is used for the optimization of an FLC that drives an active tuned mass damper towards the response control of wind-excited tall buildings. Furthermore, in (6), the authors improve the proposal presented in (3) by adding an active control system to the hybrid control system. The overall system is driven by an FLC, whose parameters are optimized by means of the two-branch tournament GA, presented in the previous works.

Further works use a first-generation MOEA to tune the parameters of the membership functions of an FLC. In (49) a hierarchical MOGA-based approach is used to tune fuzzy scheduling controllers for a gas turbine engine. The engine should satisfy nine large-signal performance criteria (e.g. steady-state accuracy, transient accuracy, disturbance rejection, stability, stall margin, structural integrity, engine degradation, etc). Once an initial suitable fuzzy scheduling controller is designed, parameters of membership functions and scaling factors are tuned to meet the former criteria.

In (119), an MOGA-based approach is presented to tune an FLC for a solid oxide fuel cell power plant. The obtained model achieves fast transient responses and has

1. MOEFSS: CURRENT STATUS AND FURTHER DIRECTIONS

very low total harmonic distortion in output current steady-state operation. To improve the fuzzy structure of the controller, a tuning process adapts the parameters of membership functions and scaling factors. Fuzzy sets are defined by the center points of normalized, triangular membership functions. Objectives are described by a system of equations that represent the harmonics to be minimized.

With regard to the use of second-generation MOEAs, in (124), the authors investigate the use of smart base-isolation strategies to reduce structural damage caused by severe loads. A friction pendulum system and a magnetorheological damper are employed in a smart base-isolation system and an FLC is used to modulate the magnetorheological damper. The classic NSGA-II is used to optimize parameters of membership functions and to find appropriate fuzzy rules for the FLC. Gaussian membership functions are used for all input and output variables of the FLC. The shapes of Gaussian membership functions are defined by two parameters and are coded into the chromosome with a real-valued representation. The optimization process aims to minimize root mean squared structural acceleration and base drift.

This problem is tackled again in (125), in which a novel control technique is proposed, by utilizing a hierarchical structure of FLCs. The structure consists of two lower-level controllers and a higher-level supervisory controller. Lower-layer controllers are optimized by NSGA-II, considering four objectives: reduction of peak superstructure acceleration, peak isolation system deformation, RMSE of superstructure acceleration and RMSE of isolation system deformation. Gaussian membership functions are used for all input and output variables of the FLC, as in the previous contribution.

In (154) an FLC is designed to manage two magnetorheological dampers for the mitigation of seismic loads. NSGA-II with Controlled Elitism is used for the optimization of FLC parameters. Fuzzy sets of input and output variables are represented by Gaussian membership functions, which are described by two parameters. These parameters are coded in the chromosome by means of floating point values. The overall optimization process aims to maximize four objective functions: peak interstory drift, peak acceleration, RMSE of interstory drift and RMSE of acceleration.

In (139), the authors present a multi-objective evolutionary process for tuning the fuzzy membership functions of a fuzzy visual system for autonomous robots. This fuzzy visual system is based on a hierarchical structure that includes three different linguistic FRBCs. The combined action of these classifiers allows robots to detect the presence of doors in the images captured by their cameras. The DB of the whole fuzzy visual system is coded in a single chromosome, which comprises the four parameters defining each trapezoidal-shaped membership function. Blend crossover (BLX-α) (73)

28

and random mutation are considered as genetic operators, whereas the conflicting objectives to be optimized are the true positive and false positive detection rates. Different single (a generational GA and CHC) and multi-objective (SPEA, SPEA2, NSGA-II) evolutionary algorithms are considered and compared, with NSGA-II obtaining the best performance.

In (60) the parameters of an adaptable hierarchical TSK fuzzy controller for blinds are optimized by NSGA-II, considering two objectives: energy consumption and thermal comfort. The fuzzy sets are represented by triangular membership functions, whose parameters are optimized. The performances of the FLC are tested by means of a software for dynamic simulation of indoor climate and energy.

In (70) an MOEA based on SPEA2 is developed to optimize the parameter of an FLC that aims to improve the water quality of a sewage treatment plant. The FLC uses ten parameters for its operation and each chromosome codifies a set of parameters. Depending on these parameters, the controller decides when to activate a blower in the aeration tank, in order to keep the water clean. Water quality is based on different criteria, therefore the optimization process tries to minimize the concentrations of three chemical compounds.

In (85), the authors proposed a tuning process combined with a rule selection process, to improve the performance of FLCs for the control of heating, ventilating and air conditioning (HVAC) systems, including several performance criteria such as energy performance, stability and indoor comfort requirements. The technique is based on $SPEA2_{E/E}$ and aims to obtain a more accurate controller by forcing the removal of unnecessary rules and biasing the search through those solutions that satisfy the performance objective to a higher degree. Two objectives are considered: maximizing the performance, expressed by aggregating five quality measures, and minimizing the complexity, computed as the number of rules obtained.

1.4.2 Learning of controller structure

Learning of controller structures is used for the generation of an FLC in situations where a reasonable set of rules is not immediately apparent. These kinds of approaches are able to take into account the synergy between the RB and DB, but they involve a heavier computational burden due to the increase in the search space.

One of the first works in this branch is (29). An FLC for a non-linear missile autopilot is designed using NSGA. Both the membership functions' distribution and the RB of the FLC are determined. The design process minimizes four objectives: the steady state error, the overshoot, the settling and the rising time.

In (156) a specific MOEA is presented for the on-line design of the structure of a fuzzy speed controller for a DC motor motion control platform. The optimization involves three objectives to be minimized: the current tracking error, the velocity tracking error and the power consumption of the system.

A medical application is presented in (77), where an MOEA is used to design FLCs to adjust the amount of drug dosage necessary to achieve the required neuromuscular blockade level in patients during surgery. The evolutionary approach is based on SPEA2 and considers two goals: the optimization of the amount of the drug required and the minimization of the complexity of the obtained FLC, so that the undertaken control decision can be explained in natural language.

Beyond the works presented above, there are some contributions that use a hybrid approach of fuzzy systems, neural networks and GAs, in order to automatically construct a controller. For example, in (48) an intelligent combustion controller is designed to handle an incineration process, by integrating different soft computing approaches. The proposed methodology applies three techniques simultaneously: a representative state function is modeled using a GA and a neural network. Then, this model is used as surrogate of the plant and a specific first-generation MOEA is applied to obtain a set of FLCs, represented by TSK-type control rules. Finally, the control RB is improved by a tuning process. In this specific application, two goals are considered: effluent quality and heat recovery.

In (150) a gain scheduling adaptive control scheme for nonlinear plants is presented. The controller is based on fuzzy systems, neural networks and GAs. A fuzzy PI controller is optimally designed using a specific MOEA to satisfy three objectives: minimizing overshoot time, minimizing settling time and smoothing output response. Then, the backpropagation algorithm is applied to design a neural gain scheduler with the aim of tuning the optimal parameters of the fuzzy PI controller at some operating points.

1.4.3 A summary of MOEFSs designed for multi-objective control problems

All contributions on MOEFSs designed for Fuzzy Control are grouped in Table 1.2. Papers are divided based on the aspects of the controller that are considered by the optimization process. A description of this type of table is given for Table 1.1 in Section 1.3.4. In almost all cases the objectives express a performance measure, therefore the objective type does not appear in this table. Due to the various application fields

of FLCs, the last column contains a brief description of the application framework. Within each group, papers are sorted in chronological order.

Table 1.2: Summary of the proposals on MOEFSs for multi-objective fuzzy control problems.

		Fuzzy Control			FRBS approach			MOEA			
		Authors	Ref.	Year	Rules	Type	#Obj.	Name	Gen.	Type	Application Framework
Controller parameters' identification		Ahlawat et al.	(2)	2001	Mam.	Ling.	2	NoN.	1st	I ⋆	Building vibration
		Ahlawat et al.	(3, 5)	2002,2004	Mam.	Ling.	2	NoN.	1st	I ⋆	Building vibration
		Ahlawat et al.	(4)	2002	Mam.	Ling.	3	NoN.	1st	I ⋆	Building vibration
		Chipperfield et al.	(49)	2002	Mam.	Ling.	9	NoN.	1st	N	Gas turbine engine
		Ahlawat et al.	(6)	2004	Mam.	Ling.	2	NoN.	1st	I ⋆	Building vibration
		Jurado et al.	(119)	2005	Mam.	Ling.	16	NoN.	1st	I ∘	Solid oxide fuel cell
		Kim et al.	(124)	2006	Mam.	Scat.	2	NSGA-II	2nd	G	Base-isolation system
		Kim et al.	(125)	2007	Mam.	Scat.	4	NSGA-II	2nd	G	Base-isolation system
		Shook et al.	(154)	2008	Mam.	Ling.	4	NSGA-II CE	2nd	I †	Seismic loads mitigation
		Muñoz et al.	(139)	2008	Mam.	Ling.	2	various	2nd	G	Fuzzy visual syst. for robots
		Daum et al.	(60)	2010	TSK	Scat.	2	NSGA-II	2nd	G	HVAC systems
		Ebner et al.	(70)	2010	‡	‡	3	NoN.	2nd	I •	Water treatment
		Gacto et al.	(85)	2010	Mam.	Ling.	2	$SPEA2_{E/E}$	2nd	I •	HVAC systems
Learning FLC struct.		Blumel et al.	(29)	2001	Mam.	Ling.	4	NSGA	1st	N	Missile autopilot
		Chen et al.	(48)	2002	TSK	Ling. ∗	2	NoN.	1st	N	Incineration process
		Stewart et al.	(156)	2004	Mam.	Ling.	3	NoN.	1st	N	DC motor motion ctrl.
		Serra et al.	(150)	2006	Mam.	Ling.	3	NoN.	2nd	N	Nonlinear plants
		Fazendeiro et al.	(77)	2007	Mam.	Ling.	2	NoN.	2nd	I •	Drug dosage for surgeries

Mam.=Mamdani, TSK=Takagi-Sugeno-Kang, Ling.=Linguistic, Scat.=Scatter, ∗In the antecedent, ‡Patented FLC, not available information;
A=Accuracy, C=Complexity, S=Semantic aspects;
NoN.=No name, N=Novel algorithm, I=Improved version, G=General use;
⋆2-branch tournament GA, ∘MOGA based, †NSGA-II based, •SPEA2 based.

In most cases, the proposal deals with the post-processing of FLC parameters, since it is the simplest approach and requires a reduced search space. Earlier works consider first-generation algorithms and only very recently have the best known second-generation MOEAs been applied. Finally, in almost all papers a Mamdani-type FRBS is used.

1.5 MOEFSs designed for fuzzy association rule mining

The knowledge extracted by the mining process can be represented in several ways, for example, using association rules. A general association rule is defined as an implication $X \Rightarrow Y$, where both X and Y are defined as sets of attributes. This implication is

interpreted as follows: "for a specified fraction of the existing transactions, a particular value of attribute set X determines the value of attribute set Y as another particular value under a certain confidence", where a transaction consists of a set of items I.

Two classic concepts are involved in association rules: support, that is the percentage of transactions that contains both X and Y, and confidence, that is, the ratio between the support of $X \cup Y$ and the support of X. Thus, the problem of association rule mining (167) consists of finding all association rules that satisfy user-specified minimum support and confidence. Early works used Boolean association rules, which consider only whether an item is present in a transaction or not, without evaluating its quantity. To take into account this aspect, fuzzy association rules (65) were introduced.

In the following, we describe those contributions that apply MOEFSs to fuzzy association rule mining. Then, a brief summary of the existing works is provided.

1.5.1 Description of the existent contributions

Fuzzy association rule extraction can be performed using MOEAs, as they obtain good results when dealing with problems involving several measures that could be contradictory to some degree. Moreover, they could also include interpretability concepts, since fuzzy association rules can explain the associations they represent.

For example, in (123) a specific Pareto-based multi-objective evolutionary approach is presented for mining optimized fuzzy association rules. Two different coding schemes are proposed: the first one tries to determine the appropriate fuzzy sets in a pre-specified rule, also called certain rule. In such cases, each individual represents the base values of membership functions of a quantitative attribute in the DB. The second coding scheme tries to find both rules and their appropriate fuzzy sets. In both approaches three objectives are maximized: support, confidence and comprehensibility of fuzzy association rules, where the last one is expressed by a measure related to the number of attributes in a rule.

A fuzzy data mining approach is presented in (46) for the Single-minimum-Support Fuzzy-Mining problem. An MOGA-based algorithm is proposed to extract both membership functions and association rules from quantitative transactions. The algorithm tries to maximize two objectives. The first is the suitability of membership functions, through a combination of coverage and overlap factors. This measure is used to reduce the membership functions that are redundant or too separate. The second objective is the total number of large 1-itemsets in a given set of minimum support values. Since a larger number of 1-itemsets will usually result in a larger number of all of the itemsets

with a higher probability, this implies more interesting association rules. Thus, this metric expresses the interestingness of a rule.

The earlier proposals in fuzzy association rule mining assumed that the number of fuzzy sets is pre-specified. In (18) an automated clustering method is proposed, which aims to automatically cluster values of a quantitative attribute, in order to obtain a large number of large itemsets in less time. The method uses an MOEA based on SPEA and the optimization process considers two objectives. The first is to maximize the number of large itemsets with respect to a given minimum support value, since a large itemset potentially leads to the discovery of some interesting fuzzy association rules. The second objective is to minimize the time required to find all large itemsets in a given database. In (158) a technique for mining optimized fuzzy association rules is proposed, to detect intrusions in a network. The proposed framework aims to concurrently identify fuzzy attributes and to define the membership functions by exploiting clustering techniques. Afterwards, MOGA (80) is applied to generate and optimize fuzzy association rules of different orders. The optimization process tries to maximize two objectives: confidence, which represents the strength of a rule, and support, which in this case identifies the generality of a rule.

A particular approach focused on predictive induction is presented in (42), in which an MOEA is used to derive fuzzy association rules from uncertain data for consumer behavior modeling. Rules are codified with DNF-type fuzzy rules. The proposed framework considers data collection, data mining and finally knowledge interpretation. During the mining process, an evolutionary scheme based on NSGA-II is applied and three objectives are minimized. The accuracy is expressed by the approximation error, the complexity is represented by the number of DNF-type fuzzy rules. This second objective does not completely assess the interpretability of the fuzzy system, since the internal structure of each DNF-type fuzzy rule is not considered. Thus, a third objective is added that measures the number of equivalent Mamdani-type fuzzy rules for each DNF-type fuzzy rule.

Beyond predictive and descriptive induction, there are mixed techniques that combine the characteristics of both types of induction. An example is Subgroup Discovery (128), which aims to extract descriptive knowledge from data that concerns a property of interest. Subgroup Discovery is a form of supervised inductive learning or subgroup description, in which the algorithm analyzes a set of data in order to find interesting subgroups, given a property of interest chosen by the user. The induction of rules that describe subgroups can be considered a multi-objective problem, since a Subgroup Discovery rule can be evaluated by means of different quality measures.

An application of an MOEA to Subgroup Discovery can be found in (38). The algorithm, called the Non-dominated Multi-objective Evolutionary algorithm for Extracting Fuzzy rules in Subgroup Discovery (NMEEF-SD), is based on the well-known NSGA-II and aims to extract novel and interpretable fuzzy rules that describe subgroups. In NMEEF-SD, the quality measures considered as objectives in the evolutionary process can be selected, making it possible to study the combinations of measures that provide better results. Three quality measures are available: support, confidence and unusualness, i.e. the weighted relative accuracy of a rule. These last measures attempt to obtain a good trade-off between the generality, interest and precision.

1.5.2 A summary of MOEFSs designed for fuzzy association rule mining

Table 1.3 contains all contributions that deal with MOEFSs designed for mining fuzzy association rules, presented in chronological order. As with Tables 1.1 and 1.2, a description of this type of table is given in Section 1.3.4, but the column describing the FRBS approach is no longer necessary. The remaining fields assume the meanings previously explained.

In most cases, the classical measures of data mining, support and confidence, are used as objectives. The application of MOEAs to extract fuzzy association rules is quite recent, beginning in 2006. Therefore, the majority of works exploit a second-generation MOEA.

Table 1.3: Summary of the proposals on MOEFSs for mining fuzzy association rules.

Fuzzy association rule mining					Objectives	MOEA		
Authors	Ref.	Year	#Obj.	Description		Name	Gen.	Type
Kaya et al.	(123)	2006	3	↑Sup. + ↑Con. + ↓Att.		NoN.	2nd	N
Alhajj et al.	(18)	2008	2	↑LI + ↓Tim.		NoN.	2nd	I •
Chen et al.	(46)	2008	2	↑L1I + ↑Sui.		NoN.	1st	I ∘
Thilgam et al.	(158)	2008	2	↑Sup. + ↑Con.		MOGA	1st	G
Casillas et al.	(42)	2009	3	↓Err. + ↓DNF-FR + ↓MAM-FR		NoN.	2nd	I †
Carmona et al.	(38) ∗	2010	3	↑Sup. + ↑FCon. + ↑Unu.		NMEEF-SD	2nd	I †

∗Applied for Subgroup Discovery;
Con.=Confidence, Sup.=Support, Tim.=Time, Err.=Error, LI=#Large itemsets, L1I=#Large 1-itemsets, Att.=#Attributes, Sui.=Suitability, DNF-FR=#DNF-type Fuzzy Rules, MAM-FR=#Equivalent Mamdani-type Fuzzy Rules, Unu.=Unusualness, FCon.=Fuzzy confidence, ↑Maximize, ↓Minimize, NoN.=No name, N=Novel algorithm, I=Improved version, G=General use; †NSGA-II based, ⋆PAES based, ∘MOGA based, •SPEA based.

1.6 Open problems and new trends in MOEFSs

In this section, some current trends in the field of MOEFSs will be presented and some recent contributions related to them will be described. In addition, some issues will be highlighted in order to focus researchers' attention on new problems that arise when using MOEFSs in real-world applications.

One important issue concerns the fact that MOEAs have not been specifically designed for MOEFSs, in which a chromosome represents parts of an FRBS and consequently assumes a complex structure that can even comprise a combination of binary, integer and real coding. Moreover, MOEFSs have to take into account test errors, which are not usually present in Evolutionary Multi-Objective Optimization benchmarks. Due to this fact, existent MOEAs may not be suited to optimize FRBS structures, thus producing sub-optimal solutions.

Considering this issue and the current state-of-the-art of MOEFSs described in the previous sections, we try to highlight some problems related to MOEFSs that should be investigated. The following subjects will be stated as open problems and briefly described:

a) performance evaluation of MOEFSs;

b) reliable interpretability measures;

c) objective dimensionality;

d) scalability issues;

e) application to imbalanced datasets;

f) automatic selection of the most suitable solution;

g) integration of decision maker's preferences;

h) design MOEFSs to generate type-2 fuzzy systems.

1.6.1 Performance evaluation of MOEFS approaches

Comparing different multi-objective optimization techniques is a difficult task, since the optimization result is a set of non-dominated solutions rather than a single solution. Researchers generally agree on considering two informal criteria to assess the quality of a solution set: the distance of the approximated points from the true Pareto

front should be minimized and solutions should be equally distribute along the front. Additionally, the extent of the obtained non-dominated front should be maximized.

In the literature, several performance measures have already been proposed to consider these criteria and to evaluate the search capacity of algorithms. The following measures are widely used: attainment surfaces, hypervolume, epsilon dominance (81, 170, 171), etc. A drawback of these measures is that the quality difference between the obtained FRBSs remains unclear. Moreover, the Pareto front approximation is generated with respect to the training data, whereas the performance of the algorithm should be evaluated with respect to test data by applying a statistical analysis.

A novel attempt based on the ideas in (82) to obtain representative mean values has been proposed in (9) to compare different multi-objective approaches: for each dataset and for each trial of an algorithm (considering cross validation), the approximated Pareto front is generated and three representative points are extracted (the best in the first objective, the median considering both objectives and the best in the second objective). Afterwards, for each dataset, the mean values and the standard deviations of some measures (first objective or training accuracy, second objective or complexity and test accuracy) are computed for each representative point over all the trials, and a non-parametric statistical test is applied locally for each measure at each representative point. In this way, authors were able to statistically compare the different algorithms by analyzing the performance of the obtained FRBSs when looking for the desired properties in the Pareto front extremes and in the mid point (equilibrium point).

This approach has been extended in (83) and applied to problems with more than two objectives. To make a statistical comparison of the different interesting points possible, the authors project the obtained Pareto fronts on the planes generated by considering pairs of objectives (in this case, accuracy-complexity and accuracy-semantic planes). In this way, they can analyze the non-dominated solutions by considering the said interesting points for each pair of objectives.

This technique presents some problems when the Pareto fronts generated by different algorithms reside in distant zones of the objective space, as it is not applicable in these cases. Therefore, a previous graphical representation of the averaged Pareto fronts is necessary to determine whether this technique is suitable or not. In cases where the obtained Pareto fronts are located in different parts of the objective space, it could be determined which representative points are comparable for each dataset by considering this graphical representation, constituting a first attempt to assess the quality difference between fronts.

1.6.2   Reliable interpretability measures

In Section 1.3 we have explained how the definition of interpretability heavily affects the comprehensibility of an FRBS and researchers are still looking for reliable and widely accepted interpretability measures. Some proposals attempt to define new indexes to consider multiple interpretability measures (19, 25, 83). This problem is mainly related to contributions of the first category and it is discussed deeply in (84), where a taxonomy is proposed to organize the different measures or constraints that are used in the literature to assess interpretability in linguistic FRBSs. A taxonomy with four quadrants is presented: complexity and semantic interpretability are taken into account at the level of RB or at the level of fuzzy partitions. Since the interpretability of linguistic FRBSs is still an open problem, the review tries to organize the different measures proposed so far, in order to help researchers to determine the most appropriate measures according to the part of the KB in which they want to maintain or improve interpretability.

This research (84) highlighted that there is not a single comprehensive global measure to quantify the interpretability of linguistic models, thus it would be necessary to consider appropriate measures from all the quadrants. It is necessary to establish a way to combine these measures globally. To this end, the different measures might be optimized as different objectives within a multi-objective framework, by also taking accuracy into account. However, the real problem resides in the choice of common and widely accepted measures for each of the quadrants, which is still an open problem for the useful application of MOEFSs that aim to discover the accuracy-interpretability trade-off of FRBSs.

1.6.3   Objective dimensionality

MOEAs usually work very well for two or three objective problems, whereas their search capacity worsens as the number of objectives increases. Problems with four or more objectives are often called many-objective problems (114).

These kinds of problems can be handled by different approaches:

- integrating many aspects into few objectives;

- selecting few aspects as objectives;

- using all the objectives.

The first approach aims to combine several objectives into a single one, using weights or appropriate aggregation operators. This method presents the common problems of a single-objective approach: the aggregation method and weights have to be chosen carefully, since they greatly influence the performance of the optimization process. However, it represents an effective way of handling many objectives when some of them are related and can be properly combined.

The second approach is achieved by reducing the dimensionality in the objective space, since not all the objectives may be necessary. If there is a certain number of non-conflicting objectives, these objectives must be considered redundant. On the other hand, in some cases there are some objectives (conflicting or not) that could be removed without significantly losing the problem information, in which case only the statistically significant conflicting objectives should be considered.

The third method is the most complex one, as when applying a classic MOEA to a many-objective problem, several problems arise. When the number of objectives increases, almost all solutions in a population become non-dominated, therefore the search capacity of MOEAs based on the Pareto-dominance concept is heavily affected. The number of solutions required to approximate the entire Pareto front increases exponentially with the number of objectives. This happens because in many-objective problems the Pareto front is represented by a hyper-surface in the objective space. The decision making process becomes harder, since the final solution is chosen from among a wider number of multi-objective solutions.

To overcome these problems, researchers found that the low selection pressure could be tackled by inducing a preference ordering over the points in the non-dominated set. The approaches based on preference ordering include relaxing the concept of Pareto-dominance, controlling the dominance area, modifying the rank definition, substituting the distance metric, etc. These approaches seem promising, but they still need further investigation.

1.6.4 Scalability issues

In recent years, having to deal with large or high dimensional datasets has become more common (97, 118). Large datasets include many instances, while high dimensional datasets refer to datasets with a large number of features. These kinds of datasets provide some difficulties: the size of large datasets affects the fitness function computation, thus increasing the computational time, whereas high dimensional datasets increase the search space. Moreover, in most of the cases, the wider the search space, the greater the number of generated rules. Resulting models can be very complex,

with interpretability heavily affected. This problem is particularly evident in the works belonging to the first and third groups of the taxonomy.

In the case of large datasets, these problems can be tackled by reducing the training set, i.e. removing irrelevant training instances prior to the learning process. The choice of the subset is a crucial task, since it has to describe the whole training set without the loss of information. When dealing with high dimensional datasets, it is also possible to perform a feature selection process that determines the most relevant variables before or during the learning process. Finally, the interpretability issue can be tackled by reducing the rule set through a post processing approach.

Large and high dimensional datasets increasingly occur in real-word problems, but until now there have been few works that attempt to approach them through the multi-objective evolutionary optimization of fuzzy systems, therefore this is still an interesting investigation field. A recent example can be found in (10), which proposes an MOEA for obtaining linguistic Mamdani compact models in 17 regression problems, including up to 80 variables and up to 40.000 example data. A variable selection mechanism is applied to ensure a fast convergence in the presence of a high number of variables. To handle problems with a high number of examples, an error estimation of the obtained models is computed by using a reduced subset of the training patterns within a new mechanism for fitness estimation which is applicable to any EA.

1.6.5 Imbalanced datasets

Problems with imbalanced datasets appear mainly when dealing with classification tasks (45). Usually, the accuracy of a classifier is evaluated according to the percentage of correct classification, which should be maximized by the optimization process. This measure is inappropriate when the application domain is characterized by a highly imbalanced distribution of samples, since positive cases compose just a small fraction of the available data used to train the classifier. In some cases the cost of misclassification is different between the positive and the negative classes. Thus, the obtained classifier presents a high predictive accuracy over the majority class and poor predictive accuracy over the minority class. Furthermore, the minority class examples can be considered as noise and completely ignored.

Two approaches can be followed to reduce or avoid bias toward the majority class.

- At data level: pre-processing mechanisms can be applied to patterns, to prevent imbalance. These solutions include different forms of re-sampling, i.e. oversampling, undersampling, and variations on or combinations of the previous techniques.

- At algorithmic level: solutions are mainly based on cost-sensitive approaches, by using metrics that take into account the misclassification costs of each class.

With regard to MOEFSs, imbalanced datasets could be handled in the application of FRBCs. A first approach in this sense can be found in (69), in which the performance of binary FRBCs are analyzed, considering an application domain characterized by highly imbalanced distributions of examples. To assess FRBCs' performance, two objectives are maximized: sensitivity and specificity. Sensitivity corresponds to the true positive rate, specificity to the complement of the false positive rate. These two metrics describe the system's ability to correctly classify patterns belonging to both the positive and the negative classes. The sum of the conditions in the antecedents of rules in the classifier is added as a third objective, in order to decrease the complexity. After the optimization process, the Receiver Operating Characteristic (ROC) curve analysis is used to compare the obtained binary classifiers and to select a set of potentially optimal classifiers.

Since these kinds of datasets are increasingly used in several fields, such as security systems, medicine, telecommunication systems, information retrieval tasks, etc, they are receiving increasing attention from researchers.

1.6.6 Automatic selection of the most suitable solution

The strength of MOEAs resides in their ability to approximate a wide part of the Pareto front, thus providing multiple solutions with different trade-offs between objectives. However, in many application fields, only a single solution is required. The problem of automatically choosing a single solution for a specific purpose has not been discussed in the studies presented so far.

Focusing on a set of obtained FRBSs (and on a single FRBS) represents a way to ease the choice of an appropriate single solution. However, this kind of visualization is a difficult task when the number of objectives increases, since it is impossible to show all the non-dominated solutions in many-dimensional visualization spaces.

The obtained FRBSs also present the problem of overfitting since they are evaluated according to test data (generalization ability). Since MOEFSs are expected to obtain a large set of FRBSs, the choice of a single solution should consider FRBSs with good generalization abilities. However, this is not an easy task since it has to be included in the learning process, so it is only possible to take into account the results of the training set, while the test set remains unused.

An approach to determine the most suitable FRBS from a given Pareto front in terms of its generalization ability has been proposed in (109). In this contribution the

authors propose a technique using a double cross-validation to evaluate the generalization ability of the obtained models. Double cross-validation has a nested structure of two cross-validation loops. The inner loop is used to determine the best complexity of fuzzy rule-based systems with the highest generalization ability for the training data in each run in the outer loop. That is, the inner loop plays the role of validation data. The determined best complexity is used to choose the final FRBS in each run in the outer loop.

1.6.7 Integration of decision maker's preferences

In a multi-objective optimization problem, exploring the whole search space can be unnecessary if the final goal is to find only those solutions that satisfy some requirements specified by the decision maker. A good strategy may be to direct the search process towards the areas of the Pareto front that better reflect the decision maker's preferences, by integrating these preferences into the optimization process. In this way, the search space is reduced and the efficiency of the search process is significantly increased. The incorporation of decision maker's preferences is an interesting research issue which has not yet been well explored in the literature.

In (30) the problem of Multi Criteria Decision Making (MCDM) is considered as the conjunction of three components: the search of the possible solutions, a preference trade-off process to select a single solution and an interactive visualization process to embed the decision-maker in the solution refinement and selection loop. Authors introduce a requirement framework to compare most MCDM problems, their solutions and analyze their performances.

A second example is presented in (140), where user preferences are incorporated into a rule selection process of FRBSs for pattern classification problems. Due to the difficulty in choosing an objective interpretability measure, multiple interpretability criteria are combined into a single *preference function*, which is used as one of the objective functions during the optimization process. Moreover, the preference function can be changed interactively by the user, through the modification of the priority of each interpretability criterion.

Another possibility to indirectly consider user's preferences is to concentrate the search on the most significant objectives. Usually, when dealing with MOEFSs, the objectives used present different difficulty levels. In this way, objectives that are easy to achieve, such as the complexity of the obtained models, bias the search, leading to sub-optimal models (overly simple models presenting inappropriate accuracies when using complexity measures). However, the user is not only interested in obtaining

simple models but also accurate ones. Some approaches concentrating the search on the accuracy objective as a way to obtain the most accurate models can be found in (11, 82, 85).

1.6.8 Design MOEFSs to generate type-2 fuzzy systems

At the end of the 90s, a new class of fuzzy system was presented (120), in which the antecedent or consequent membership functions were type-2 fuzzy sets. The concept of a type-2 fuzzy set is introduced by Zadeh (166) as a generalization of the concept of an ordinary fuzzy set, also referred to as type-1 fuzzy set. A type-2 fuzzy set incorporates uncertainty about the membership function into fuzzy set theory since its membership function is three-dimensional, where the third dimension is the value of the membership function at each point on its two-dimensional domain. If there is no uncertainty, a type-2 fuzzy set is reduced to a type-1 fuzzy set. Such sets are useful when it is difficult to determine an exact membership function for a fuzzy set.

As in the case of type-1 fuzzy systems, the hybridization of type-2 fuzzy systems and GAs was proposed in (144), in order to automatically design type-2 fuzzy systems, following which several contributions have been published, in which GAs, and in general EAs, are used to obtain type-2 fuzzy systems, mainly in control applications (43, 137, 160).

Despite this, as far as we know, no proposals have yet been presented to combine MOEAs with type-2 fuzzy systems, therefore this may be a new and promising research field.

1.7 Conclusion

In this Chapter a two-level taxonomy has been presented to categorize the considerable number of contributions presented on the application of MOEFSs. In the first level, the contributions are divided depending on the multi-objective nature of the problem tackled, while in the second level they are divided according to the type of GFS used.

In the first level, the most prolific category includes works on the application of MOEFSs to the trade-off between interpretability and accuracy. Therefore, many complex variations of existing MOEAs have been proposed in order to obtain better performances. The second category gathers works that deal with the application of MOEFSs to multi-objective fuzzy control problems, in which many contributions focus on first-generation algorithms, probably due to the fact that they could be efficiently applied in control problems, in spite of their simplicity. However, it should be remembered

that the introduction of the elitism concept in second-generation MOEAs is a theoretical requirement to assure convergence. Only recently, MOEFSs have been applied to extract fuzzy knowledge from databases, therefore the third category includes few contributions. Moreover, nor are there well-described measures that consider fuzziness in association rules.

Finally, several current trends and open problems have been highlighted, in order to draw the attention of the research community to their importance, since they are either unsolved or have still not been addressed. Among these problems, we focused on the scalability issues, i.e. the study of the accuracy-interpretability trade-off when dealing with high dimensional and large scale datasets. In particular, an analysis of the influence of granularity has been carried out, since it has been demonstrated that appropriate granularities contribute to the generation of more accurate models (59). Moreover, in some cases a granularity learning process can indirectly realize a feature selection process, in fact, when the granularity assigned to a certain variable is equal to one, the variable itself can be considered useless to the scope of the problem. This approach is particularly useful when dealing with high-dimensional problems.

PART I
Granularity Learning

Chapter 2

Improving a Fuzzy Association Rule-Based Classification Model by Granularity Learning based on Heuristic Measures over Multiple Granularities

2.1 Introduction

A FRBS is composed of two main parts: the inference system, which implements the fuzzy inference process needed to obtain an output when an input is given, and the Knowledge Base (KB), which represent the knowledge about the problem to be solved and it consists of the Rule Base (RB) and the Data Base (DB). When designing a FRBS, both components need to be specified and the precision of the FRBS is directly related to them. Focusing in the KB specification, many approach have been presented to automatically learn the RB from numerical data representing the system behavior, but there is not a similar effort for specifying the DB, although its design it also a critical task. Most of the RB learning methods assume the existence of a previously designed DB. In particular, one of the problems that arise is how to determine the number of fuzzy sets associated to each variable in the DB, i.e. the *granularities*, which can affect both accuracy and complexity of FRBCSs (59).

The easiest strategy consists in fixing a single a-priori granularity and creating uniform fuzzy partitions for all the variables (105, 113). Despite its simplicity, this approach can not be the most appropriate since it does not consider at all the available knowledge of a problem and usually induces the generation of a high number of fuzzy

rules, thus increasing the complexity of models. The choice of suitable granularities is crucial since it contributes to the generation of more precise models (59). A further approach consists in the use of multiple granularities (115), which is useful to reduce the number of rules in the obtained models, although it causes a loss of semantic interpretability.

For this reason, in (13) the authors proposed a mechanism to identify appropriate single granularities while performing a multi-objective evolutionary fuzzy rule selection process, based on the proposal presented in (115). The framework includes four steps: a) first, a heuristic procedure is used to create a pre-specified number of promising fuzzy rules; b) then, for each attribute a single granularity is learnt, considering the frequency of used partitions and the importance of the rules extracted in the previous step; c) next, these granularities are used to extract again a pre-specified number of fuzzy rules; d) finally, a multi-objective evolutionary algorithm is used to perform the rule selection process. Nevertheless, this method presents some limitations when dealing with high dimensional problems, due to the increase of the search space and of the time required by the fitness evaluation in the MOEA when it is working on the large amount of extracted fuzzy rules.

However, this would be possible considering more appropriate rule generation approaches dealing with high dimensional problems as the Fuzzy Association Rule-Based Classification model for High Dimensional problem (FARC-HD) proposed in (15). This algorithm currently belongs to the state-of-the-art of classification algorithms, since it has been demonstrated to outperform in accuracy some of the most widespread classification algorithms. In particular, the results presented in (15) show that it performs better than three other GFSs (FH-GBML, 2SLAVE, SGERD), two approaches to obtain fuzzy associative classifiers (LAFAR and CFAR) and five classic approaches for associative classification (C4.5, CBA, CBA2, CMAR, CPAR).

In the following, a method is proposed that combines the single granularity specification mechanism presented in (13) with a new multi-objective version of a fuzzy associative classification algorithm based on FARC-HD (15), especially designed to continue dealing with high-dimensional problems. We name this method MO-FARCG. The aim is to avoid the use of multiple granularities but still reducing the complexity of the obtained classifiers while maintaining high generalization ability, by considering both objectives within a multi-objective evolutionary framework.

The remaining part of this chapter is arranged as follows: Section 2.2 introduces some preliminary concepts about FRBSs for classification (FRBC) and fuzzy association rules. Section 2.3 describes in detail each stage of the proposed approach. Section

2.4 presents the experimental setup and discusses the obtained results. Finally, in Section 2.6, some concluding remarks are pointed out.

2.2 Preliminaries

In this section, FRBCs are briefly introduced. Then, fuzzy association rules are described and their application to classification problems is explained.

2.2.1 Fuzzy rule-based classifiers

Let us consider a set of m patterns $\mathbf{x}_p = (x_{p1}, ..., x_{pn})$, $p = 1, 2, ..., m$ to be classified and a set of M classes to be assigned. Considering an n-dimensional pattern space, x_{pi} is the attribute value of the p-th pattern for the i-th attribute ($i = 1, ..., n$). A classification problem is to assign a label to each pattern, in a way that it is consistent with some observed data we know about the problem (training data).

In this case, fuzzy rules of the following type are used:

$$R_q : \text{ If } x_1 \text{ is } A_{q1} \text{ and ... and } x_n \text{ is } A_{qn}$$
$$\text{then Class } C_q \text{ with } RW_q, \tag{2.1}$$

where R_q is the label of the q-th fuzzy rule, $\mathbf{x} = (x_1, ..., x_n)$ is an n-dimensional pattern vector, A_{qi} is an antecedent fuzzy set ($i = 1, ..., n$), C_q is a class label, and RW_q is the rule weight. The antecedents fuzzy sets of R_q are denoted as a fuzzy vector $\mathbf{A}_q = (A_{q1}, A_{q2}, ..., A_{qn})$. Each new pattern is classied as the class with the maximum total strength of the vote.

The performance of FRBCs is highly affected by the rule weight RW_q associated with each fuzzy rule (106). The rule weight can be defined in different ways and many mechanisms have been proposed in the literature. For example, in (116), the authors present several heuristic methods that can be used to specify the weight of fuzzy rules. In this study, the most common one has been chosen, i.e. the fuzzy confidence value or certainty factor (CF) (56).

The certainty factor is widely used for fuzzy classification as it just affects the strength of each fuzzy IF-THEN rule in the classification phase, without changing the positions of the antecedent fuzzy sets (106).

49

2.2.2 Fuzzy association rules for classification

Association rules are used to represent and identify dependences between items in a database (96, 167). They are expressions of the type $A \rightarrow B$, where A and B are sets of items, and $A \cap B = \oslash$. This means that if all the items in A exist in a transaction, then all the items in B with a high probability are also in the transaction, and A and B should have no common items (1).

Several studies have been presented in which association rules are extracted from databases with binary or discrete values, but many real-world problems include also quantitative data. Therefore, the new trend in this field is to design mining algorithms that can handle different types of data.

In this case, the fuzzy set theory represents an useful tool, due to its affinity with the human reasoning (108). Fuzzy association rules can be used to describe associations between data in a more effective way than simple association rules do. In fact, the linguistic representation makes them more understandable for human experts and fuzzy sets avoid sharp boundaries when partitioning the domain of an attribute. Recently, the mining of fuzzy association rules from quantitative data has been investigated in several studies (14, 65, 123).

An example of fuzzy association rule is a rule of the type:

$$A \; is \; Middle \rightarrow B \; is \; High. \tag{2.2}$$

where A and B are the attributes present in the database and *Middle* and *High* are linguistic terms associated to these variables.

The interestingness of a fuzzy association rule is commonly evaluated by two measures, namely support and confidence, which can be defined as follows:

$$Support(A \rightarrow B) = \frac{\sum_{x_p \in T} \mu_{AB}(x_p)}{|N|} \tag{2.3}$$

$$Confidence(A \rightarrow B) = \frac{\sum_{x_p \in T} \mu_{AB}(x_p)}{\sum_{x_p \in T} \mu_A(x_p)} \tag{2.4}$$

where $|N|$ is the number of transactions that appear in the database T, $\mu_A(x_p)$ is the matching degree of the transaction x_p with the antecedent of the rule, and $\mu_{AB}(x_p)$ is the matching degree of the transaction x_p with the antecedent and consequent of the rule.

In recent years, fuzzy association rules have been investigated to be used as classification rules (102, 103, 142). A fuzzy association rule can be used as a classification

rule if its consequent part includes only one class label ($C = C_1, ..., C_j, ..., C_S$). This type of rule is called *fuzzy associative classification rule* and assumes the following form:

$$A \rightarrow C_j \tag{2.5}$$

As common fuzzy association rules, it can be evaluated in terms of support and confidence:

$$Support(A \rightarrow C_j) = \frac{\sum_{x_p \in ClassC_j} \mu_A(x_p)}{|N|} \tag{2.6}$$

$$Confidence(A \rightarrow C_j) = \frac{\sum_{x_p \in ClassC_j} \mu_A(x_p)}{\sum_{x_p \in T} \mu_A(x_p)} \tag{2.7}$$

2.3 Multi-objective fuzzy association rule-based classifier with granularity learning (MO-FARCG)

In this section, the integration between the granularity learning and a fuzzy associative classification algorithm will be described. In this case, the framework is organized into three stages:

1) Setting stage: Learning the Appropriate Granularities. First, for each class a fixed pre-specified number of rules with multiple granularities are obtained, according to well-known data mining rule evaluation measures (1). Then, a single granularity for each attribute is chosen, depending on the frequency of extracted rules and some quality measures.

2) Learning stage: Extraction of Candidate Fuzzy Association Rules. All the possible frequent fuzzy itemsets are listed in a search tree and then fuzzy association rules are generated. Finally, rules are evaluated and sorted following a criterion and only the best rules are maintained in order to reduce the number of candidate rules.

3) Post-processing stage: Multi-Objective Evolutionary Fuzzy Rule Selection and MFs Tuning. The best cooperative rules are selected and tuned by using a Multi-Objective Evolutionary Algorithm based on the Strength Pareto Evolutionary Algorithm 2 (SPEA2) (169), exploiting the positive synergy of both techniques within the same process.

2.3.1 Setting stage: Learning the appropriate granularities

In this stage an initial set of rules is extracted, following the multiple granularity approach presented in (13). Then, a single granularity for each attribute is chosen on the basis of some information included in the extracted rules. These granularities will be used in the next stage to obtain a new initial set of fuzzy rules.

2.3.1.1 Extracting rules with multiple granularities

At the beginning, a multiple granularity is used for all attributes: fourteen fuzzy sets are considered, distributed into four fuzzy partitions (see Figure 2.1). An additional fuzzy set is also used, to represent a *don't care* condition (the domain interval $[0, 1]$), then the overall number of possible fuzzy rules is 15^n.

This number of rules is too large to be considered for the generation of the initial candidate rule set, therefore only rules with a small number of antecedent conditions are selected. This number is fixed on 3 for datasets with less than 30 attributes, and 2 for datasets with a number of attributes equal to or bigger than 30.

The heuristic procedure presented in (112) is used to determine the rule weight CF_q and the consequent class C_q for each fuzzy rule R_q. For the antecedent part $\mathbf{A}_q$, the confidence of each class is calculated as

$$c(\mathbf{A}_q \Rightarrow \text{Class } h) = \frac{\displaystyle\sum_{\mathbf{x}_p \in \text{Class } h} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}{\displaystyle\sum_{p=1}^{m} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}, \quad h = 1, \ ..., \ M. \tag{2.8}$$

Then, the consequent class C_q is specified according to the class with the maximum
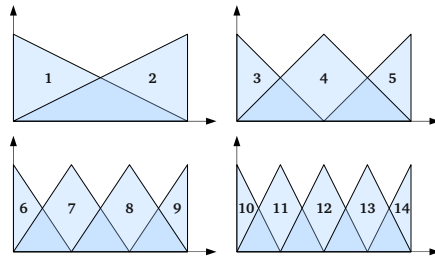
Figure 2.1: The fourteen antecedent fuzzy sets considered for each variable.

confidence:

$$c(\mathbf{A}_q \Rightarrow \text{Class } C_q) = \max_{h=1, 2, ..., M} \{c(\mathbf{A}_q \Rightarrow \text{Class } h)\}. \tag{2.9}$$

As previously said, the performance of FRBCs is highly influenced by the rule weight. In this proposal, the following definition of CF_q is used since it has led to good results (116):

$$CF_q = c(\mathbf{A}_q \Rightarrow \text{Class } C_q) - \sum_{\substack{h=1 \\ h \neq C_q}}^{M} c(\mathbf{A}_q \Rightarrow \text{Class } h). \tag{2.10}$$

In order to maintain only useful rules, a fuzzy rule R_q is not selected as a candidate rule if its confidence is smaller than 0.5.

The heuristic technique described before generates a large number of short fuzzy rules as candidate rules, including not interesting rules. To decrease this number and to select the most useful rules, a preventive rule reduction is performed. To this end, rules are evaluated and sorted according to the product $p(R_q) = s(R_q) \cdot c(R_q)$, where $c(R_q)$ is the confidence and $s(R_q)$ is the support, that is the percentage of samples covered by R_q. Finally, for each class, the best 300 rules are chosen.

2.3.1.2 Specifying a single granularity for each attribute

The pool of extracted fuzzy rules contains information that can be used to choose a single granularity for each attribute. This can be done by considering how many times a granularity appears in the extracted rules (weighted by the corresponding rule importance). This approach has been already used in (13) to establish a single granularity.

In particular, we use the following specification:

$$Gr(i) = \operatorname*{argmax}_{g=2,...,5} \left\{ \sum_{Gran(\mathbf{A}_{qi})=g} \text{Imp}(\mathbf{R}_q) \right\}, \tag{2.11}$$

where $Gran(\mathbf{A}_{qi})$ is the granularity of the partition that include the fuzzy set used in attribute i of rule R_q and $\text{Imp}(R_q)$ is a criterion to measure the importance of the rule. Frequency, confidence, weight, support and product of confidence and support are some of the criteria that can be used.

In order to promote more general rules, two different approaches were investigated in the original proposal (13): 1-ALL approach and 1-2-3 approach. Both of them give priority to those rules which include a single antecedent condition in the attribute considered. The only difference is when all the rules presents more than one condition: in 1-ALL approach, the granularity is obtained by considering all the rules including

the attribute itself, without taking into account the number of conditions, whereas, in 1-2-3 approach, the priority is given to shorter rules.

Here, we focus on the product and confidence criteria, combined with 1-ALL approach, since these combinations provided better results. At the end of this stage, a single granularity is obtained for each attribute.

2.3.2 Learning stage: Extraction of candidate fuzzy association rules

In this stage, the attributes' granularities learnt in the previous phase are used to extract from the original data a new set of fuzzy association rules. Afterwards, this set is reduced by applying a prescreening procedure. These two steps are as follows.

2.3.2.1 Rule extraction

For each class, all the possible fuzzy itemsets are constructed, using a search tree. The root level of each tree (level 0) is generated as an empty set. All attributes are assumed to have an order, which is the order of appearance in the training data, and all the one-itemsets constitute the first level of the search tree, following their order (level 1). The further level (level 2) for an attribute *A* is constructed by considering all the two-itemsets that combine the one-itemset of attribute *A* with all the one-itemsets for the other attributes in the order. The same procedure is used to construct the following levels of the tree. No repeated itemsets appear in the tree.

Moreover, each itemset is evaluated with respect to a minimum support and a minimum confidence: an itemset with a support higher than the minimum support is a frequent itemset and an itemset with a confidence higher than the confidence's threshold has reached the quality level demanded by the user. Therefore, if the support of an *n*-itemset in a node *A* is lower than the minimum support, the node is not extended anymore. At the same way, if the classification rule associated to an item set has a confidence higher than the minimum confidence, the correspondent node does not need to be extended further.

The procedure above lists all the frequent fuzzy itemsets, which are used to generate the candidate fuzzy association rules. Each rule will contain a frequent itemset in the antecedent and the corresponding class in the consequent and for all the classes this process is repeated. The number of frequent fuzzy itemsets extracted depends directly on the minimum support, which is defined for each class using the distributions of the classes over the data set. This stage generates a large number of rules, which can be

hardly handled by human users. To simplify the understanding of the model, the depth of the trees ($Depth_{max}$), and so the length of the fuzzy rules, is limited to a fixed value.

2.3.2.2 Rule prescreening

The rule extraction process generates a large number of rules, which can cause a problem of rule redundancy. To decrease this number by selecting only the best rules, a subgroup discovery technique is used, in particular the pattern weighting scheme described in (122).

Each pattern is associated to a weight $w(i) = \frac{1}{i+1}$, where i stores how many times the pattern has been covered by a rule. Initially, all the weights assume the same value $w(0) = 1$.

For each class, the algorithm selects the best rule, then the weights related to the patterns covered by this rule are decreased. In this way, the patterns that are still uncovered will have a greater possibility of being covered in the following iterations. When the i counter reach a threshold k_t, the correspondent pattern is deleted.

The remaining rules are sorted again and the procedure is repeated until either all patterns have been deleted, or there is no rule left in the rule base.

To evaluate the quality of fuzzy rules, we use a modification of the wWRAcc' measure described in [33]. The wWRAcc' measure has been modified to enable the handling of fuzzy rules. The new measure is defined as follows:

$$wWRAcc''(A \rightarrow C_j) = \frac{n''(A \cdot C_j)}{n'(C_j)} \cdot \left(\frac{n''(A \cdot C_j)}{n''(A)} - \frac{n(C_j)}{N}\right) \qquad (2.12)$$

where $n''(A)$ is the sum of the products of the weights of all covered patterns by their matching degrees with the antecedent part of the rule, $n''(A \cdot C_j)$ is the sum of the products of the weights of all correctly covered patterns by their matching degrees with the antecedent part of the rules, $n(C_j)$ is the number of patterns of class C_j and $n'(C_j)$ is the sum of the weights of patterns of class C_j.

2.3.3 Post-processing stage: Multi-objective evolutionary fuzzy rule selection and membership functions tuning

In the final step, a modification of the well-known SPEA2 is applied as post-processing algorithm to the KB generated by the previous stage. A similar version of this algorithm was already used for regression problems in (83), in which three objectives are considered instead of two. The SPEA2 was preferred to the well-known NSGA-II

since in (11) the approaches based on SPEA2 have been demonstrated to be more appropriate when a tuning of MFs is performed.

This SPEA2 modification performs a fuzzy rule selection together with a tuning of the MFs, aiming to improve the model accuracy as the first objective and to reduce the model complexity as the second objective. In the next sections, the main components of this algorithm are described, and then, the specific characteristics and its main steps are presented.

2.3.3.1 Objectives

Each chromosome is associated with a bi-dimensional vector, whose elements express the fulfillment degree of the following two objectives, respectively:

- classification error minimization: it is represented by the complement of the number of the classification rate;

- complexity minimization: it is represented by the number of selected rules;

To compute the classification error, the following function has been used:

$$Fitness(C) = 1 - \frac{\#Hits}{N} \qquad (2.13)$$

where $\#Hits$ is the number of patterns correctly classified and N is the total number of patterns.

2.3.3.2 Coding scheme and initial gene pool

A double coding scheme for both *rule selection* (C_S) and *tuning* (C_T) is used: $C^p = C_S^p C_T^p$, where C^p is the chromosome representing the individual p. The $C_S^p = (c_{S1}, \ldots, c_{Sm})$ part is represented by a binary-coded string with m genes, where m is the number of initial rules. Each gene contains a values of "1" if the correspondent rule is selected, "0" otherwise. The C_T^p part uses a real coding scheme to codify the three definition parameters of the triangular MFs, where m^i is the number of labels in the database for each of the n variables.

$C_T^p = C_1 C_2 \ldots C_n$
$C_i = (a_1^i, b_1^i, c_1^i, \ldots, a_{m_i}^i, b_{m_i}^i, c_{m_i}^i)\ i = 1, \ldots, m$

The first individual of the first population codifies the KB obtained by the previous step. The remaining individuals of the first population are generated randomly, with each value within the corresponding variation intervals.

2.3.3.3 Crossover and mutation

An intelligent crossover and mutation operators are described in this section. Each offspring is obtained in the following way.

- First, the C_T part of the offspring is obtained by applying blend crossover (BLX)-0.5 (74) to the C_T part of the parents.

- Then, the binary part C_S is generated depending on the C_T parts of parents and offspring: for each gene in the C_S part, the following steps are performed.

    - Each gene of the C_T part which represents the corresponding MFs of the rule, is considered for both parents and offspring. The MFs of these three rules are extracted.

    - Between the offspring rule and each parent rule, euclidean normalized distances are computed by considering the center points of the MFs involved in these rules. The differences between each pair of centers are normalized by the amplitudes of their respective variation intervals.

    - The parent's rule closer to the offspring's rule is selected and its value is duplicated in the C_S part of the offspring.

This process is repeated until each gene in the C_S part of the offspring is obtained. In each step four offspring are generated, although after applying mutation only the two best offspring are maintained. This type of crossover prevents the recovery of a bad rule already discarded, while permits the recovery of a rule that can be still considered good due to its MFs configuration.

The crossover operator performs a better exploration in the C_S part, therefore the mutation operator does not need to add rules. It simply changes randomly a gene value in the C_T part and sets to zero a random gene in the C_S part, with probability P_m.

The application of these operators brings some advantages: the crossover between individuals with very different rules allows the algorithm to explore different parts of the search space, while the mutation promotes rule extraction, since it is used to remove unnecessary rules.

2.3.3.4 Modifications of the classic SPEA2

Some changes have been introduced to the original selection mechanism of SPEA2, to improve the algorithm's search ability.

- A mechanism to prevent incest has been included, based on the concepts of CHC presented in (72). This avoid premature convergence in the real coding (C_T) part, which has a greater influence on the algorithm convergence and represents a wider search space than the binary coding part (C_S). In the CHC approach, parents are crossed only if their Hamming distance divided by 4 exceeds a threshold. To follow this approach, the real coding scheme needs to be converted in a binary one, thus each gene is transformed using a gray code with a fixed number of bits per gene (BGene). The threshold value is initially set to $L = (\#CT \times BGene)/4$, where $\#CT$ is the number of genes in the C_T part of the chromosome. This value is decreased by 1 at each generation of the algorithm, therefore in further generations closer solutions can be crossed.

- A restart operator has been introduced to renew the external population when we detect that all the crossover are allowed. Actually, to prevent premature convergence, the first restart is applied if 50% of crossovers are detected at any generation (the required ratio can be defined as $\%_{required} = 0.5$). Each time the restart is performed, the required ratio is updated ad follows: $\%_{required} = (1 + \%_{required})/2$.

  The external population after the restart includes the individuals with the best value in each objective, and the remaining individuals are initialized as follows: the C_S part is copied from the most accurate individual, while the values in the C_T part are generated randomly. In this way, the most accurate and interpretable solutions obtained so far are preserved.

  Some constraints to the application of restart have been introduced: a) a new restart cannot be applied if the most accurate solution has not been improved; b) the restart is not applied at the end, when the approximation of the Pareto front is well formed and needs to be preserved; c) restart is disabled if the midpoint of the total evaluations number is reached and it has been never applied before.

- A mechanism to promote the most accurate solutions has been introduced. At each stage of the algorithm, between restarting points, the number of solutions in the external population ($\bar{P}_{t+1}$) that can be used to constitute the mating pool is reduced progressively and the most accurate solutions are preferred. To this end, solutions are sorted according their accuracy and the number of eligible solutions is reduced progressively from 100% at the beginning to 50% at the end of each stage. This mechanism is disabled in the last evaluations (when restart is disabled too), in order to obtain a wide and well-formed Pareto front.

2.4   Experimental framework

To evaluate the performance of the proposed approach with respect to the original FARC-HD, we have considered 24 real-world datasets (Table 2.1). Only datasets with continuous attributes have been considered, since the two first steps of the method have not been designed to handle nominal data. Moreover, in case of instances presenting missing values, they have been removed from the datasets (Cleveland). All datasets can be downloaded from the Knowledge Extraction based on Evolutionary Learning (KEEL)-dataset repository (`http://sci2s.ugr.es/keel/datasets.php`).

To carry the different experiments out, a ten-fold cross-validation model is considered: each dataset is randomly split into ten folds, each containing 10% of the patterns of the dataset. Then, a single fold is used for testing and the remaining for training. The cross-validation process is repeated ten times, with each fold used exactly once for testing. For each of the ten partitions, three trials of the algorithm are executed and finally the results are averaged out over 30 runs.

The proposed method is called MO-FARCG and it has been compared with the original approach FARC-HD (15). Two versions of MO-FARCG have been considered, using product (MO-FARC-prod) and confidence (MO-FARC-conf) criteria, respectively. Due to the multi-objective nature of the SPEA2 included in MO-FARCG, the average of the most accurate solution from all the Pareto fronts is considered for the comparison.

Table 2.2 summarizes the parameters used for the methods' analysis. The maximum number of antecedents allowed for a fuzzy rule is restricted to 3 (*short* fuzzy rules), in order to encourage the generation of simpler models. While this was reported by the authors that the original FARC-HD did not present any change over 15000 evaluations, the proposed approach needs 20000 evaluations to reach the convergence. For this reason we fix 15000 for FARC-HD and 20000 for MO-FARCG, since no changes are obtained for the original approach over the specified number of evaluations. The parameter k_t is the threshold beyond which the covered patterns are completely eliminated.

Statistical analysis (88, 89) was adopted to evaluate the results, and in particular non-parametric tests, following the recommendations presented in (67), where a set of simple and robust non-parametric tests for statistical comparisons of classifiers has been described.

The Wilcoxons signed-ranks test (153, 162) for pair-wise comparison was used, with a confidence level of $\alpha = 0.05$ in all cases. A wider description of this test and a software to perform it can be found on the web site available at: `http://sci2s.ugr.es/sicidm/`.

Table 2.1: Datasets considered in the study.

Name	Attributes (R/I/N)	Patterns	Classes
Appendicitis	7 (7/0/0)	106	2
Cleveland	13 (13/0/0)	297(303)	5
Ecoli	7 (7/0/0)	336	8
Glass	9 (9/0/0)	214	7
Heart	13 (1/12/0)	270	2
Iris	4 (4/0/0)	150	3
Magic	10 (10/0/0)	19020	2
Monks	6 (0/6/0)	432	7
Movement Libras	90 (90/0/0)	360	15
Page-blocks	10 (4/6/0)	5472	5
Penbased	16 (0/16/0)	10992	10
Phoneme	5 (5/0/0)	5404	2
Pima	8 (8/0/0)	768	2
Ringnorm	20 (20/0/0)	7400	2
Satimage	36 (0/36/0)	6435	6
Sonar	60 (60/0/0)	208	2
Spambase	57 (57/0/0)	4597	2
Spectfheart	44 (0/44/0)	267	2
Texture	40 (40/0/0)	5500	11
Twonorm	20 (20/0/0)	7400	2
Vowel	13 (10/3/0)	990	11
Wdbc	30 (30/0/0)	569	2
Wine	13 (13/0/0)	178	3
Yeast	8 (8/0/0)	1484	10

2.5 Experimental results

This section shows the results of the experiments described in the previous section. Table 2.3 shows the average number of rules/conditions (**#R/#C**) and classification percentages in training (**Tra**) and test (**Tst**) of the most accurate solution from each of the obtained Pareto fronts, for the two versions of MO-FARCG, and of the best solution for FARC-HD. The overall mean values for each method are highlighted in the last row.

Table 2.2: Parameters of the methods considered for comparison.

Method	Parameters
FARC-HD	$Minsup = 0.05$, $Maxconf = 0.85$, $Depth_{max} = 3$, $k_t = 2$, $Pop = 50$, $Evaluations = 15000$ $BITSGENE = 30$
MO-FARCG-prod	$Minsup = 0.05$, $Maxconf = 0.85$, $Depth_{max} = 3$, $k_t = 2$, $Pop = 50$, $Evaluations = 20000$ $BITSGENE = 30$
MO-FARCG-conf	$Minsup = 0.05$, $Maxconf = 0.85$, $Depth_{max} = 3$, $k_t = 2$, $Pop = 50$, $Evaluations = 20000$ $BITSGENE = 30$

Table 2.3: Results referred to the most accurate solutions obtained by applying MO-FARC-prod and MO-FARC-conf, respectively, and to the solution obtained by applying FARC-HD.

DATASETS	MO-FARCG-prod				MO-FARCG-conf				FARC-HD			
	#RULE	#COND	TRA	TST	#RULE	#COND	TRA	TST	#RULE	#COND	TRA	TST
Appendicitis	4.73	1.83	93.43	85.79	4.80	1.68	93.26	86.09	6.8	1.8	93.82	84.18
Cleveland	21.23	2.85	79.03	56.22	24.53	2.87	81.93	56.45	61.3	2.9	88.18	55.24
Ecoli	7.07	2.75	64.41	62.80	7.07	2.75	64.41	62.80	33.8	2.4	92.33	82.19
Glass	10.23	2.35	76.38	66.56	11.20	2.38	75.89	68.55	22.7	2.5	81.1	70.24
Heart	14.40	2.60	93.00	81.48	15.70	2.58	93.53	82.22	27	2.6	93.91	84.44
Iris	4.77	1.13	98.20	95.56	4.77	1.15	98.17	95.56	4	1.1	98.59	96
Magic	4.23	2.24	81.15	80.80	10.77	1.86	83.58	82.99	43.3	2.5	85.36	84.51
Monk2	13.53	1.92	100.00	99.47	12.43	1.65	100.00	99.92	14.2	2	99.92	99.77
Movement libras	46.33	2.94	94.31	71.67	46.40	2.94	94.72	72.87	83.1	2.9	95.52	76.67
Pageblocks	5.77	2.17	94.40	94.30	6.17	2.05	94.36	94.23	19.1	2.3	95.62	95.01
Penbased	55.57	2.91	95.38	94.25	55.70	2.92	95.07	93.89	152.8	2.8	97.04	96.04
Phoneme	3.87	2.35	80.09	79.13	8.00	2.08	81.79	80.53	17.8	2.2	83.52	82.14
Pima	4.23	1.75	80.04	74.93	8.77	2.35	82.17	76.01	22.7	2.4	82.9	75.66
Ring	8.27	1.50	84.82	83.72	14.07	1.65	92.54	91.52	24	1.9	95.13	94.08
Satimage	30.00	2.68	88.88	86.77	32.63	2.67	89.21	87.04	76.1	2.7	88.68	87.32
Sonar	4.73	2.74	88.98	79.09	6.50	2.69	92.70	78.81	18	2.3	98.77	80.19
Spambase	2.83	2.75	77.78	77.42	2.93	2.76	74.76	74.11	29.8	2.4	92.37	91.93
Spectfheart	3.50	1.00	79.40	79.42	3.50	1.00	79.40	79.42	12.9	2.8	91.43	79.83
Texture	21.20	2.75	93.59	92.37	21.67	2.71	94.45	93.08	54.5	2.7	93.71	92.89
Twonorm	14.87	2.96	97.06	96.06	14.87	2.96	97.06	96.06	60.9	2.6	96.64	95.28
Vowel	27.50	2.77	50.06	46.16	34.97	2.75	59.60	54.38	72.3	2.9	80.48	71.82
Wdbc	5.23	1.44	96.69	94.56	5.37	1.54	97.27	95.09	10.4	1.7	98.57	95.25
Wine	8.83	1.87	99.67	95.53	8.43	1.80	99.90	95.88	8.7	1.6	99.94	94.35
Yeast	12.17	2.53	58.69	56.34	15.77	2.53	58.24	56.11	35.2	2.6	63.81	58.5
MEANS	**13.96**	**2.28**	**85.23**	**80.43**	**15.71**	**2.26**	**86.42**	**81.40**	**37.98**	**2.36**	**91.14**	**84.31**

A first comparison has been drawn between the two different approaches of MO-FARCG: the Wilcoxon's signed-ranks test has been applied to establish if the two versions are statistically equivalent (null-hypothesis). Table 2.4 shows that the Wilcoxon test applied on the test classification percentage of the most accurate solutions rejects the null hypothesis, since $p - value \leq /alpha$. Therefore, the two approaches are not statistically equivalent and the version MO-FARCG-conf is to be preferred.

For this reason, this version of MO-FARCG has been chosen to be compared with

Table 2.4: Comparison on test accuracy between MO-FARCG-prod and MO-FARCG-conf

Comparison (test accuracy)	**p value**	**Hypothesis**
MO-FARCG-prod vs MO-FARCG-conf	0.0084	Rejected

the FARC-HD and the Wilcoxon's signed-ranks test has been applied again on the test classification percentage of the most accurate solutions. The null-hypothesis is rejected, since $p - value \leq /alpha$ (see Table 2.5), hence the two methods are not statistically equivalents.

A further comparison has been drawn between the two algorithm with respect to the average number of rules of the most accurate solutions. Once more the null-hypothesis is rejected with $p - value = 1.35E - 005$ and the two method are not statistically equivalent (Table 2.5).

Table 2.5: Comparison on test accuracy and complexity between FARC-HD and MO-FARCG-conf

Comparison (test accuracy)	**p value**	**Hypothesis**
FARC-HD vs MO-FARCG-conf	0.00754	Rejected

Comparison (complexity)	**p value**	**Hypothesis**
FARC-HD vs MO-FARCG-conf	1.35E-005	Rejected

By looking at the results reported in Table 2.3, we can state that MO-FARCG-conf is outperformed by the original FARC-HD regarding the test accuracy, whereas the opposite is true when considering the complexity of the obtained systems. Nevertheless, against less than 3% loss on test accuracy, the complexity is reduced by more than 50%. This makes the models easier to understand but maintaining their accuracy at a similar level or even better for some datasets.

Ideally, when a multi-objective approach and a single-objective approach are applied to the same task, the solution set obtained by the former approach includes the solution obtained by the latter one. In the present study we cannot expect a similar result since the starting conditions are different for the multi-objective and the single-objective algorithms. In fact, FARC-HD uses an initial database generated by considering equidistributed granularities, while the initial database in MO-FARCG is constructed by considering the most promising granularities obtained in the initial steps of the algorithm. Therefore, the loss of accuracy and the complexity reduction are

provoked by the initial granularities choice rather than by the application of the multi-objective algorithm.

2.6 Conclusion

In this chapter, we have investigated how the granularity learning affects the performance of a FRBCs. To this aim, a method has been proposed, named MO-FARCG, by extending the FARC-HD algorithm proposed in (15). Two preliminary steps have been added: first a set of rules with multiple granularities is extracted and then a single granularity for each attribute is specified, depending on some measures performed on the extracted rules. Moreover, the single-objective genetic algorithm included in the original FARC-HD algorithm has been extended with a MOEA (that is a modification of SPEA2), in order to consider both accuracy and complexity of the obtained models.

According to the results obtained by performing experiments over 24 real-world datasets, we can conclude that the method proposed to learn attributes' granularities produces models with a slightly decreased accuracy, which is balanced by a considerable reduction of models' complexity.

Chapter 3

A Multi-Objective Evolutionary Method for Learning Granularities Based on Fuzzy Discretization to Improve the Accuracy-Complexity Trade-Off of Fuzzy Rule-Based Classification Systems: D-MOFARC

3.1 Introduction

In the previous chapter a method has been proposed to learn suitable granularities, and its influence over the accuracy-complexity trade-off of the obtained FRBSs has been investigated. The experimental results have highlighted that the proposed granularity learning process leads to models in which the complexity is reduced, but the accuracy is slightly decreased.

The ideal goal would be to concurrently improve both objective or, at least, to improve the accuracy while maintaining the complexity to an acceptable level. This issue is addressed in the present chapter, in which a granularity learning process is proposed to generate granularities that will be used to define the initial fuzzy partitions of the DB. To this end, we present a fuzzy discretization algorithm which extracts suitable granularities from data. This mechanism has been integrated within a MOEFS which evolves the initial KB. Although the main objective is the accuracy, the MOEA has been exploited as a tool to mainly improve precision while decreasing the complexity of the models (82).

3. GRANULARITY LEARNING WITH FUZZY DISCRETIZATION

In particular, the multi-objective evolutionary method, namely Multi-Objective Fuzzy Association Rule-Based Classification Model with granularity learning based on Discretization (D-MOFARC), comprehends the following steps:

- Step 1. A Fuzzy Discretization Algorithm is designed, in order to learn automatically suitable granularities for each variable and to generate the correspondent fuzzy partitions. This approach is based on the concept of discretization (44, 76), which represents the process of transforming the range of values of a continuous attribute in a set of intervals and assigning a discrete value to each interval. We extend this approach to the case of fuzzy partitions, considering attributes interdependencies. Therefore, after obtaining a set of intervals for each variable, a fuzzy set is assigned to each interval, instead of a discrete value, to obtain the fuzzy labels associated to each variable. The process has been integrated with a tree-based generation mechanism that considers attribute partitioning interdependencies.

- Step 2. An initial RB associated to the previous fuzzy partitions is created by extracting candidate fuzzy association rules. To this aim, the first two steps of the Fuzzy Association Rule-Based Classification Model for High-Dimensional Problems (FARC-HD) method proposed in (15) have been used. Since the extracted rules do not use all the labels generated in the initial step, the initial fuzzy partitions are finally refined by removing the unused labels.

- Step 3. A new specific MOEA is designed to concurrently perform the tuning of MFs in the DB and the selection of rules in the RB. This algorithm is a modified version of the Strength Pareto Evolutionary Algorithm 2 (SPEA2)(169), and aims to improve the accuracy while maintaining the complexity of the initial model, at the same time. As said, the multi-objective approach has been preferred since it has been demonstrated to be useful in generating more precise models, while the use of the number of rules as second objective helps in limiting their complexity (82).

This chapter is arranged as follows: Section 3.2 introduces some preliminary concepts about discretization methods in general and the CAIM discretization algorithm (131) in particular. Section 3.3 illustrates the characteristics of the fuzzy discretization approach. In Section 3.4 the features of the proposed method are introduced and described in detail. Section 3.5 illustrates the experimental framework, in which the experimental setup is described and the obtained results are presented and discussed. Finally, in section 3.6, some concluding remarks are pointed out.

3.2 Preliminaries: Discretization approaches and the CAIM algorithm

A possible approach for learning granularities is represented by discretization algorithms, which transform continuous attribute's values into a set of intervals and assign to each interval a numerical and discrete value. The aim is to minimize the number of discrete intervals while maximizing the interdependency between class labels and discrete attribute values, to prevent an excessive information loss during the discretization. A discretization scheme D on a continuous attribute A divides the continuous domain of A into n discrete intervals and can be represented as follows:

$$D : \{[d_0, d_1], (d_1, d_2], \cdots, (d_{n-1}, d_n]\}, \tag{3.1}$$

where d_0 and d_n are the minimum and the maximum values of the attribute A, respectively, and the values d_i identify the *cut points* for the discretization D.

The process consists of two steps: first, the number of intervals for each attribute are found and then the boundaries of the intervals are determined. Usually the first task is not performed automatically and the number of intervals must be specified by the user. Discretization algorithms can be grouped into two categories:

- Unsupervised (or class-blind) algorithms generate intervals without considering the class labels of each pattern.

- Supervised algorithms discretize attributes by considering the interdependence between class labels and the attribute values.

We will focus on this latter type. In (131) a discretization algorithm is proposed, that automatically determines the number of intervals for data partitioning and concurrently finds the boundaries of each interval. The method is named CAIM, from the name of the criterion optimized to measure the dependency between the class C and the discretization D for the attribute A. The criterion is defined for a given attribute by means of a matrix, called quanta matrix. Table 3.1 shows the quanta matrix for a generic attribute A, where S is the number of classes, q_{ir} is the total number of patterns that belong to the $i - th$ class and have the value of the attribute A in the interval $(d_{r-1}, d_r]$, M_{i+} is the total number of patterns belonging to the $i - th$ class, M_{+r} is the total number of values of the attribute A that are in the interval $(d_{r-1}, d_r]$, for $i = 1, \cdots, S$ and $r = 1, \cdots, n$ and M is the total number of patterns.

3. GRANULARITY LEARNING WITH FUZZY DISCRETIZATION

Table 3.1: Quanta Matrix for attribute A and discretization scheme D

Class	Interval					Class Total
	$[d_0, d_1]$	$\cdots$	$(d_{r-1}, d_r]$	$\cdots$	$(d_{n-1}, d_n]$	
C_1	q_{11}	$\cdots$	q_{1r}	$\cdots$	q_{1n}	M_{1+}
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
C_i	q_{i1}	$\cdots$	q_{ir}	$\cdots$	q_{in}	M_{i+}
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
C_S	q_{S1}	$\cdots$	q_{Sr}	$\cdots$	q_{Sn}	M_{S+}
Interval Total	M_{+1}	$\cdots$	M_{+r}	$\cdots$	M_{+n}	M

The CAIM criterion is expressed as follows:

$$CAIM(C, D|A) = \frac{\sum_{r=1}^{n} \dfrac{max_r^2}{M_{+r}}}{n} \tag{3.2}$$

where n is the number of intervals, max_r is the maximum value among all q_{ir} values and $i = 1, \cdots, S$.

The space of all the possible discretization schemes is large and the search of the optimal scheme cannot be optimally solved using exact algorithms due to the time requirement. Thus, the CAIM algorithm searches for an approximation of the optimal scheme by finding local maximum values of the CAIM criterion. This approach has demonstrated to be computationally not expensive and to lead to good approximations of the optimal discretization scheme. This algorithm is executed in two steps:

- For each continuous attribute A, a set of candidate interval boundaries B (i.e. the candidate *cut points*) and an initial discretization scheme are created: first, the minimum (d_0) and maximum (d_n) values of the attribute's domain are found and added to the set of candidate cut points. Then, all the distinct values of A in the dataset are considered and for each adjacent pair of values, a midpoint is calculated. All the midpoints are added to the set of candidate cut-points B. The algorithm starts by considering the a discretization scheme ($D : [d_0, d_n]$) that covers all the possible values of a continuous attribute and with the criterion *GlobalCAIM* = 0.

- These new cut-points are added to the initial discretization according to the maximization of the CAIM criterion. From all candidate cut-points, the algorithm chooses the point that gives the highest value of the CAIM criterion, then this point is added to the discretization if *CAIM* > *GlobalCAIM* or if there are less than S intervals in the discretization. In fact, it is assumed that each discretized

attribute needs at least the number of discretization intervals to be equal to the number of classes, since this guarantees that the discretized attribute can improve the subsequent classification. After the addition of a cut-point to the discretization, the *GlobalCAIM* criterion is updated and assigned with the *CAIM* value of the added cut-point.

A detailed description of the algorithm can be found in the original proposal (131). The CAIM algorithm generates separately a single discretization *D* for each attribute. However, when considering FRBSs, the attributes are combined together to form the rules and an appropriate discretization should take into account this aspect. Therefore, in the following section we propose a new discretization approach that try to exploit the interdependence among attributes to generate fuzzy discretizations, i.e. the initial fuzzy partitions.

3.3 Data base extraction: A proposal for fuzzy discretization on interdependent attribute scenarios

In order to extract an initial set of fuzzy rules from a dataset, it is necessary to build an associated DB and in particular to choose a granularity for each variable (13, 59). We already highlighted how a fixed by hand single granularity approach may not be the best choice, since it leads to models with a bad trade-off between accuracy and complexity.

To this end, we present a fuzzy discretization method to generate automatically a promising set of labels for each attribute. The proposed approach actually does not generate a single fuzzy partition, but instead a small set of promising fuzzy partitions for each attribute, in order to provide a wider number of labels for the rule extraction process. The number of partitions is fixed to a maximum of 3 for each attribute.

This method includes some concepts derived from the CAIM discretization algorithm, to generate the initial discretizations that will be subsequently *fuzzified* to obtain the associated fuzzy partitions. We do not use directly the CAIM algorithm since it does not consider a possible dependence among attributes in the generation process.

The proposed technique, named Fuzzy Discretization Algorithm, uses trees to take into account relationships among attributes. For each attribute a tree is constructed: first, the considered attribute is set as root node, then, children nodes are added by choosing from the remaining attributes and promoting the ones that present a good combination with the parent node.

The Fuzzy Discretization Algorithm comprehends the following three steps:

- Ordering the attributes according to a fitness, based on the CAIM criterion.

- Generating a set of discretizations for each attribute in the dataset.

- Transforming each discretization in a fuzzy partition to obtain the DB.

3.3.1   Step 1: Ordering the attributes

When dealing with classification datasets, attributes are supposed to be more or less important for the classification task. To establish an order among them, the CAIM algorithm is initially applied for each attribute and the obtained discretizations are evaluated and ordered with respect to the CAIM criterion (3.2). In this way, the CAIM algorithm itself determines the order of the attributes and the trees' construction will follow this order.

3.3.2   Step 2: Generating the discretizations

The previous step generates a CAIM discretization for each attribute. This discretization is discarded, since the CAIM algorithm usually promotes discretizations with a small number of cut points, especially when a dataset presents few classes. On the contrary, the first discretization for each attribute is generated by considering equidistributed intervals, fixing the number of cut points to a high value in order to promote better granularities. In our case, the number of cut points for the initial discretization has been fixed to 4, to generate at least 5 fuzzy labels for each attribute.

The first discretization is used to generate a tree for each attribute, in a recursive manner. Let's consider a parent node attribute A and its corresponding first discretization D_A. The input patterns are sorted with respect to the values of attribute A and they are grouped according to the intervals $i = 1, \cdots, I$ defined by the discretization D_A, where I is the number of intervals of discretization D_A. Let's now consider the group of patterns corresponding to an interval i. The CAIM algorithm is applied on these patterns, considering all the attributes $T_j \neq A$ ($j = 1, \cdots, S$) and for each of them a discretization is obtained. These discretizations are evaluated according to the CAIM criterion and the attribute corresponding to the discretization that maximize it is chosen as child node B for the interval i in the tree.

This process is repeated for each interval in the discretization D_A. After generating all the children of a node (i.e. a child for each interval) and their corresponding discretizations, the entire step is repeated recursively for each child (Figure 3.1) and it stops when one of the following conditions occurs:

- only two patterns of the dataset fall in an interval.

- the depth of the tree reaches $Depth_{max}$, fixed to 3.
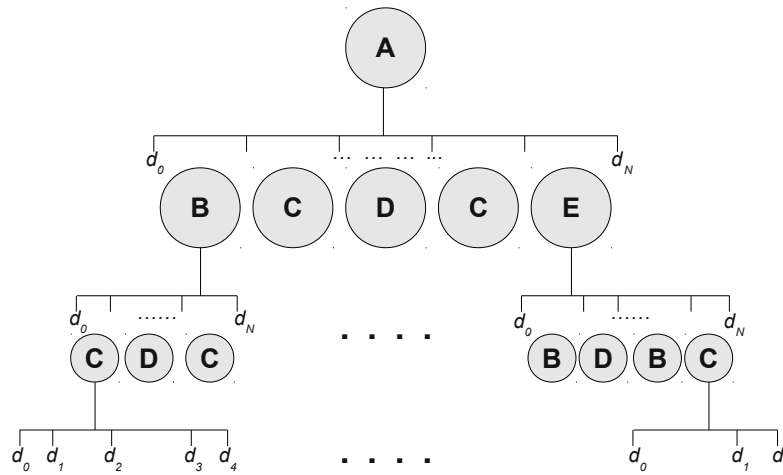
Figure 3.1: An example of generated tree for an attribute.

After this step, a tree for each attribute has been built and for each node a discretization has been generated. However, for each attribute a maximum of three discretizations are stored, in particular the first encountered in the trees' generation process, that are supposed to be the best ones.

3.3.3 Step 3: Generating the data base

In this step, the discretizations corresponding to each attributes are transformed into fuzzy partitions, by applying the concept of strong fuzzy partition (58). For each interval a triangular fuzzy set is generated: the centroid c_i is set in correspondence with the middle point between the boundaries of the interval $(d_{i-1}, d_i]$, while the left and right parameters β_l and β_r of the labels between two adjacent centroids c_i and c_{i+1} are set in correspondence with the adjacent centroids themselves (Figure 3.2).
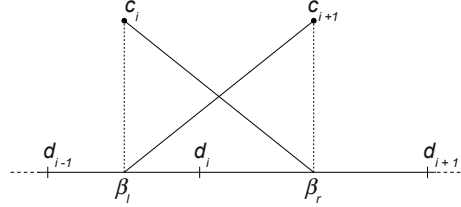
Figure 3.2: Fuzzification of a discretization according to the extended fuzzy CAIM criterion.

3.4 Multi-objective fuzzy association rule-based classification algorithm with granularity learning based on discretization (D-MOFARC)

The discretization approach presented so far has been integrated within a new Multi-Objective Evolutionary method, which evolves an initial KB by performing concurrently a rule selection process and a tuning process. The aim is to improve the precision of the initial fuzzy model through the tuning of the MFs, while maintaining or decreasing the complexity by means of the selection of rules. This proposed method can be summarized in three stages:

- Data base extraction. The Fuzzy Discretization Algorithm described in section 3.3 is used to construct a set of initial fuzzy partitions.

- Rule base extraction. An initial RB associated to the previous fuzzy partitions is created by extracting candidate fuzzy association rules. To this aim, the first two steps of the FARC-HD method proposed in (15) have been used. Since some of the labels belonging to the initial fuzzy partitions are never used in any of the extracted rules, the initial partitions are refined by removing the unused labels and an initial DB is generated.

- Data base tuning and rule selection. A new specifically designed MOEA is proposed, which concurrently performs the MFs tuning of the DB and the selection of rules from the RB. This algorithm is a modified version of the Strength Pareto Evolutionary Algorithm 2 (SPEA2)(169), and aims to improve the accuracy while reducing the complexity of the initial model at the same time.

Hereinafter, the stage 2 and 3 of the method are described in detail.

3.4.1 Rule base extraction

The Fuzzy Discretization Algorithm generates a maximum of three fuzzy partitions for each attribute, which are used to derive the initial RB. To this aim, the first two steps of the approach presented in (15) have been applied. This method uses fuzzy association rules to codify the information extracted from a dataset, that are an extension of association rules, used to represent dependencies between itemsets in a database (96, 167). Fuzzy association rules can consider not only binary or discrete values, but also quantitative values, and can be used as classification rules if their consequent is expressed with a class label. The generation of the RB is performed in two consecutive steps. In the following we briefly describe these two steps (for a more detailed explanation see (15)).

3.4.1.1 Step 1: Rule extraction

In this first step, a set candidate association rules is extracted: for each class a search tree is built, in order to list all the possible frequent itemsets of a class. itemsets are constructed, using a search tree. The root level of each tree (level 0) is generated as an empty set and all the one-itemsets constitute the first level of the search tree (level 1). The further level (level 2) for an attribute A is constructed by considering all the two-itemsets that combine the one-itemset of attribute A with all the one-itemsets for the other attributes. The same procedure is used to construct the following levels of the tree. No repeated itemsets appear in the tree. Once all the frequent itemsets have been listed, a candidate fuzzy association rule is constructed for each itemset, by setting the itemset itself in the antecedent and the corresponding class in the consequent. The maximum depth of the trees is fixed to three, in order to generate rules with few antecedent conditions, thus keeping the model simple.

3.4.1.2 Step 2: Rule prescreening

The rule extraction process generates a large number of rules, which can cause a problem of rule redundancy. To decrease this number by selecting only the best rules, a subgroup discovery technique is used, in particular the pattern weighting scheme described in (122). Each pattern is associated to a weight $w(i) = \frac{1}{i+1}$, where i stores how many times the pattern has been covered by a rule. Initially, all the weights assume the same value $w(0) = 1$. For each class, the algorithm selects the best rule, then the weights related to the patterns covered by this rule are decreased. In this way, the patterns that are still uncovered will have a greater possibility of being covered in the

73

following iterations. When the i counter reach a threshold k_t, the correspondent pattern is deleted. The remaining rules are sorted again and the procedure is repeated until either all patterns have been deleted, or there is no rule left in the rule base. To evaluate the quality of fuzzy rules, a modification of the wWRAcc' measure described in (122) has been used. The wWRAcc' measure has been modified in order to handle fuzzy rules. The new measure is defined as follows:

$$wWRAcc''(A \rightarrow C_j) = \frac{n''(A \cdot C_j)}{n'(C_j)} \cdot (\frac{n''(A \cdot C_j)}{n''(A)} - \frac{n(C_j)}{N}) \tag{3.3}$$

where $n''(A)$ is the sum of the products of the weights of all covered patterns by their matching degrees with the antecedent part of the rule, $n''(A \cdot C_j)$ is the sum of the products of the weights of all correctly covered patterns by their matching degrees with the antecedent part of the rules, $n(C_j)$ is the number of patterns of class C_j and $n'(C_j)$ is the sum of the weights of patterns of class C_j.

Since the rule prescreening process can remove a large number of rules, after this step it is possible that some of the labels that belong to the original fuzzy partitions do not appear anymore in any of the remaining rules. Therefore, the initial DB needs to be refined by removing all the unnecessary labels, thus contributing to decrease the complexity of fuzzy partitions.

3.4.2 Evolutionary multi-objective data base tuning and rule selection

In this step, the knowledge base previously obtained is improved by concurrently applying a tuning of the DB parameters and a rule selection process. A modification of the SPEA2 algorithm(169) is designed, with the aim improving the accuracy-complexity trade-off. In the following subsections the main features of this algorithm are presented. They are:

- Objectives.

- Coding scheme and initial gene pool.

- Crossover and mutation operators.

A further subsection is included at the end, which describes the main modifications introduced with respect to the classic SPEA2 algorithm.

3.4.2.1 Objectives

Each chromosome is associated with a bi-dimensional vector, whose elements express the fulfillment degree of the following two objectives, respectively:

- Classification error minimization. It is represented by the complement of the number of the classification rate, i.e. the error rate. To compute this classification error, the following function has been used:

$$Fitness(C) = 1 - \frac{\#Hits}{N} \tag{3.4}$$

  where $\#Hits$ is the number of patterns correctly classified and N is the total number of patterns.

- Complexity minimization. It is represented by the number of selected rules.

3.4.2.2 Coding scheme and initial gene pool

A double coding scheme for both *rule selection* (C_S) and *tuning* (C_T) is used: $C^p = C_S^p C_T^p$, where C^p is the chromosome representing the individual p. The $C_S^p = (c_{S1}, \ldots, c_{Sm})$ part is represented by a binary-coded string with m genes, where m is the number of initial rules. Each gene contains a value of "1" if the corresponding rule is selected, "0" otherwise. The C_T^p part uses a real coding scheme and codifies three definition parameters for each triangular MF.

$C_T^p = C_1 C_2 \ldots C_n$

$C_i = (a_1^i, b_1^i, c_1^i, \ldots, a_{m_i}^i, b_{m_i}^i, c_{m_i}^i)$ $i = 1, \ldots, n$ where m_i is the number of labels in the database for each of the n variables.

Each triangular MF is represented as $MF_j = (a_j, b_j, c_j)$, where $j = (1, \ldots, k)$ and k is the granularity of the considered fuzzy partition. Each MF parameters can assume the values within the following variation intervals:

$$
\begin{aligned}
[I_{a_j}^l, I_{a_j}^r] &= [a_j - (b_j - a_j)/2, \ \ a_j + (b_j - a_j)/2] \\
[I_{b_j}^l, I_{b_j}^r] &= [b_j - (b_j - a_j)/2, \ \ b_j + (c_j - b_j)/2] \\
[I_{c_j}^l, I_{c_j}^r] &= [c_j - (c_j - b_j)/2, \ \ c_j + (c_j - b_j)/2]
\end{aligned} \tag{3.5}
$$

The first individual of the first population codifies the KB obtained by the previous step. The remaining individuals of the first population are generated randomly, with each value within the corresponding variation intervals.

3.4.2.3 Crossover and mutation

Crossover and mutation operators have been specifically designed. Each offspring is obtained in the following way.

- First, the C_T part of the offspring is obtained by applying blend crossover (BLX)-0.5 (74) to the C_T part of the parents.

- Then, the binary part C_S is generated depending on the C_T parts of parents and offspring. For each gene in the C_S part, the following steps are performed:

  - Each gene of the C_T part which represents the corresponding MFs of the rule, is considered for both parents and offspring. The MFs of these three rules are extracted.

  - Between the offspring rule and each parent rule, euclidean normalized distances are computed by considering the center points of the MFs involved in these rules. The differences between each pair of centers are normalized by the amplitudes of their respective variation intervals.

  - The parent's rule closer to the offspring's rule is selected and its value is duplicated in the C_S part of the offspring.

This process is repeated until each gene in the C_S part of the offspring is obtained. In each step four offspring are generated, although after applying mutation only the two best offspring are maintained. This type of crossover prevents the recovery of a bad rule already discarded, while permits the recovery of a rule that can be still considered good due to its MFs configuration.

The crossover operator performs a better exploration in the C_S part, therefore the mutation operator does not need to add rules. It simply changes randomly a gene value in the C_T part and sets to zero a random gene in the C_S part, with probability P_m.

The application of these operators brings some advantages: the crossover between individuals with very different rules allows the algorithm to explore different parts of the search space, while the mutation promotes rule extraction, since it is used to remove unnecessary rules.

3.4.2.4 Modifications of the classic SPEA2 algorithm

Some changes have been introduced to the original selection mechanism of SPEA2, to improve the algorithm's search ability.

- A mechanism to prevent incest has been included, based on the concepts of CHC presented in (72). This avoid premature convergence in the real coding (C_T) part, which has a greater influence on the algorithm convergence and represents a wider search space than the binary coding part (C_S). In the CHC approach, parents are crossed only if their Hamming distance divided by 4 exceeds a threshold. To follow this approach, the real coding scheme needs to be converted in a binary one, thus each gene is transformed using a gray code with a fixed number of bits per gene (BGene). The threshold value is initially set to $L = (\#CT \times BGene)/4$, where $\#CT$ is the number of genes in the C_T part of the chromosome. This value is decreased by 1 at each generation of the algorithm, therefore in further generations closer solutions can be crossed.

- A restart operator has been introduced to renew the external population when we detect that all the crossover are allowed. Actually, to prevent premature convergence, the first restart is applied if 50% of crossovers are detected at any generation (the required ratio can be defined as $\%_{required} = 0.5$). Each time the restart is performed, the required ratio is updated ad follows: $\%_{required} = (1 + \%_{required})/2$.

  The external population after the restart includes the individuals with the best value in each objective, and the remaining individuals are initialized as follows: the C_S part is copied from the most accurate individual, while the values in the C_T part are generated randomly. In this way, the most accurate and interpretable solutions obtained so far are preserved.

  Some constraints to the application of restart have been introduced: a) a new restart cannot be applied if the most accurate solution has not been improved; b) the restart is not applied at the end, when the approximation of the Pareto front is well formed and needs to be preserved; c) restart is disabled if the midpoint of the total evaluations number is reached and it has been never applied before.

- A mechanism to promote the most accurate solutions has been introduced. At each stage of the algorithm, between restarting points, the number of solutions in the external population ($\bar{P}_{t+1}$) that can be used to constitute the mating pool is reduced progressively and the most accurate solutions are preferred. To this end, solutions are sorted according their accuracy and the number of eligible solutions is reduced progressively from 100% at the beginning to 50% at the end of each stage. This mechanism is disabled in the last evaluations (when restart is disabled too), in order to obtain a wide and well-formed Pareto front.

3.5 Experimental framework

The method proposed and described in the previous sections has been evaluated by comparing its results with the results obtained by applying a recent and well-performing GFS for classification problems, named FARC-HD (15). As we said in Section 2.1, this algorithm has been demonstrated to generate good fuzzy rule-based models, in particular when dealing with high dimensional problems, outperforming in accuracy some of the most widespread state-of-the-art classification algorithms.

3.5.1 Experimental setup

The experiments have been performed by considering 35 real-world datasets, whose characteristics are described in Table 3.2. The number of instances (**#Inst**), the number of attributes (**#Attr**) (numerical and nominal attributes **Num/Nom** are highlighted, respectively) and the number of classes (**#Cls**) of each dataset are shown. The web link to the Knowledge Extraction based on Evolutionary Learning(KEEL)-data set repository is also reported (16), from which the datasets can be downloaded. The instances that presented missing values have been removed from the datasets (in particular, from cleveland, crx, marketing datasets).

To carry the different experiments out, a ten-fold cross-validation model has been applied: each dataset has been randomly split into ten folds, each containing 10% of the patterns of the dataset. Then, a single fold has been used for testing and the remaining folds for training. The cross-validation process has been repeated ten times, with each fold used exactly once for testing. For each of the ten partitions, three trials of the algorithm have been executed and finally the results have been averaged out over 30 runs. Due to the multi-objective nature of the evolutionary algorithm included in the D-MOFARC method, the average of the most accurate solution from all the Pareto fronts, which is our main objective, has been considered for the comparison with the single-objective FARC-HD approach. To evaluate the results, statistical analysis has been adopted (88, 89), in particular non-parametric tests, following the recommendations presented in (67), where a set of simple and robust non-parametric tests for statistical comparisons of fuzzy rule-based classifiers has been described. The Wilcoxons signed-ranks test (153, 162) for pair-wise comparison was used, considering a confidence level of $\alpha = 0.05$. A wider description of this test and a software to perform it can be found on the web site available at: `http://sci2s.ugr.es/sicidm/`.

Table 3.2: List of the datasets used in the study.

Dataset	#Inst	#Attr(Num/Nom)	#Cls
iris	150	4 (4/0)	3
tae	151	5 (5/0)	3
hepatitis	155	19 (19/0)	2
wine	178	13 (13/0)	3
automobile	205	25 (15/10)	6
glass	214	9 (9/0)	7
newthyroid	215	5 (5/0)	3
heart	270	13 (13/0)	2
cleveland	297(303)	13 (13/0)	5
haberman	306	3 (3/0)	2
ecoli	336	7 (7/0)	8
bupa	345	6 (6/0)	2
balance	625	4 (4/0)	3
crx	653(690)	15 (6/9)	2
australian	690	14 (8/6)	2
wisconsin	699	9 (9/0)	2
pima	768	8 (8/0)	2
vehicle	846	18 (18/0)	4
german	1000	20 (7/13)	2
contraceptive	1473	9 (9/0)	3
titanic	2201	3 (3/0)	2
segment	2310	19 (19/0)	7
spambase	4597	57 (57/0)	2
banana	5300	2 (2/0)	2
phoneme	5404	5 (5/0)	2
page-blocks	5472	10 (10/0)	5
texture	5500	40 (40/0)	11
optdigits	5620	64 (64/0)	10
satimage	6435	36 (36/0)	7
thyroid	7200	21 (21/0)	3
ring	7400	20 (20/0)	2
twonorm	7400	20 (20/0)	2
coil2000	9822	85 (85/0)	2
penbased	10992	16 (16/0)	10
magic	19020	10 (10/0)	2

http://sci2s.ugr.es/keel/datasets.php

3.5.2 Results and analysis

This section shows the results of the experiments described in the previous section. Table 3.3 summarizes the average number of rules/conditions (**#R/#C**) and classification

percentages in training (**Tra**) and test (**Tst**) of the most accurate solution from each of the obtained Pareto fronts, for the D-MOFARC approach and of the best solution for FARC-HD (15). The overall mean values for each method are highlighted in the last row.

The two methods have been compared by applying the Wilcoxon's signed-ranks test, in order to understand if they are statistically equivalent (null-hypothesis). When considering the accuracy, the Wilcoxon's test is based on computing the differences between the average errors on the test set, whereas when considering the complexity the test is computed by taking into account the average number of rules that are obtained by a pair of algorithms. A normalized difference $DIFF$ has been adopted when considering the number of rules, which is defined as

$$DIFF = \frac{MeanRules(x) - MeanRules(RA)}{MeanRules(x)} \qquad (3.6)$$

where $MeanRules(x)$ represents the number of rules obtained on average by the x algorithm (D-MOFARC) and RA is the reference algorithm (FARC-HD). This difference expresses the improvement in percentage with respect to the reference algorithm.

Table 3.4 shows the statistics obtained by applying the Wilcoxon's signed-rank test on the accuracy achieved on the test set and on the number of rules, comparing the results obtained by applying the D-MOFARC approach with the results achieved by using the FARC-HD algorithm. The ranks R^+ and R^- and the p-value are shown and a further column is added to highlight if the null-hypothesis is rejected or not, with a significance level $\alpha = 0.05$.

When considering the accuracy on test, the $p\ value > \alpha$, therefore the test rejects the null-hypothesis of equivalence. The R^+ and R^- values highlight that the results of the proposed method outperform the results obtained by the FARC-HD algorithm. A further comparison has been drawn between the two methods with respect to the average number of rules of the most accurate solutions. In this case, the null-hypothesis is not rejected but the two approaches could be considered statistically different with a significance level of 0.065, i.e., a 93.5% confidence. Nevertheless, by looking at the average values presented in Table 3.3 we can notice that the average number of rules obtained by applying the D-MOFARC method is slightly lower than the average value obtained by applying the FARC-HD algorithm.

Figure 3.3 presents a representative example of the rule sets obtained by the D-MOFARC (3.3-a) and the FARC-HD (3.3-b) approaches, respectively, when considering the third fold of the bupa dataset. As well, Figure 3.4 represents the rule sets

Table 3.3: Comparison of the average results obtained by applying the D-MOFARC method and the FARC-HD algorithm.

10 folds	D-MOFARC				FARC-HD			
Dataset	**#R**	**#C**	**Tra**	**Tst**	**#R**	**#C**	**Tra**	**Tst**
iris	5.6	1.1	98.1	**96.0**	4.4	1.1	98.6	95.3
tae	20.2	2.7	82.1	**59.3**	19.9	2.3	71.7	59.0
hepatitis	11.4	1.9	100.0	**90.0**	10.4	2.1	99.4	88.7
wine	8.6	1.8	100.0	**95.8**	8.3	1.5	100.0	95.5
automobile	38.9	2.7	99.3	**81.0**	34.1	2.7	96.6	76.8
glass	27.4	2.6	95.2	**70.6**	18.2	2.4	79.0	69.0
newthyroid	9.5	1.7	99.8	**95.5**	9.6	1.8	99.2	94.4
heart	18.7	2.7	94.4	**84.4**	27.8	2.6	93.1	83.7
cleveland	45.6	2.9	90.9	52.9	42.1	2.8	82.2	**58.3**
haberman	9.2	2.0	81.7	69.4	5.7	1.2	79.2	**73.5**
ecoli	26.2	2.5	94.0	**82.7**	32.2	2.4	91.6	81.2
bupa	7.7	2.6	82.8	**70.1**	10.6	1.9	78.2	66.4
balance	20.1	2.3	89.4	85.6	18.8	1.0	92.2	**91.2**
crx	22.6	2.6	91.6	**84.9**	24.4	2.4	91.0	84.7
australian	19.8	2.7	91.3	86.0	25.9	2.5	90.8	**86.4**
wisconsin	9.0	1.4	98.6	**96.8**	13.6	1.1	98.3	96.2
pima	10.4	2.5	82.3	75.5	20.2	2.3	82.3	**76.2**
vehicle	22.4	2.8	84.5	**70.6**	31.6	2.5	77.2	68.0
german	55.4	2.8	86.8	**72.4**	63.3	2.7	83.4	71.7
contraceptive	49.0	2.9	63.6	53.6	68.5	2.6	61.2	**54.1**
titanic	10.4	1.8	78.9	78.7	4.1	1.3	79.1	**78.8**
segment	26.2	2.5	98.0	**96.6**	41.1	2.7	94.8	93.3
spambase	24.3	2.2	91.7	90.5	30.5	2.3	92.4	**91.6**
banana	8.7	2.0	90.3	**89.0**	12.9	1.5	86.0	85.5
phoneme	9.3	2.4	84.8	**83.5**	17.2	2.2	83.9	82.4
page-blocks	21.5	2.4	97.8	**97.0**	18.4	2.3	95.5	95.0
texture	60.5	2.8	97.5	**95.2**	50.9	2.7	93.7	93.0
optdigits	135.4	3.0	97.0	93.4	204.0	3.0	97.0	**93.5**
satimage	56.0	2.8	90.8	**87.5**	30.2	2.8	84.4	83.8
thyroid	5.9	2.4	99.3	**99.1**	4.9	1.8	94.3	94.1
ring	15.3	1.7	94.2	93.3	24.9	1.9	95.1	**94.0**
twonorm	10.2	2.7	94.5	93.1	60.4	2.6	96.6	**95.1**
coil2000	89.0	1.0	94.0	**94.0**	2.6	1.0	94.0	**94.0**
penbased	119.2	2.9	97.4	**96.2**	152.7	2.8	97.0	96.0
magic	32.2	2.6	86.3	**85.4**	43.8	2.5	85.4	84.8
Mean	**30.3**	2.4	**91.4**	**84.2**	33.9	**2.2**	89.0	83.6

Table 3.4: Wilcoxon's statistic on test accuracy and number of rules for D-MOFARC (R^+) vs FARC-HD (R^-), considering 37 datasets (10 folds).

Comparison (Tst Acc)	R^+	R^-	p-value	Hypothesis
D-MOFARC vs FARC-HD	414.5	180.5	0.04544	Rejected

Comparison (#Rules)	R^+	R^-	p-value	Hypothesis
D-MOFARC vs FARC-HD	428.0	202.0	0.0649	Not Rejected

obtained by the D-MOFARC (3.4-a) and the FARC-HD (3.4-b) approaches, respectively, when considering the third fold of the newthyroid dataset. In these specific cases, the figures highlight that the proposed method induces the generation of more compact rule sets and smaller granularities with respect to the FARC-HD algorithm.
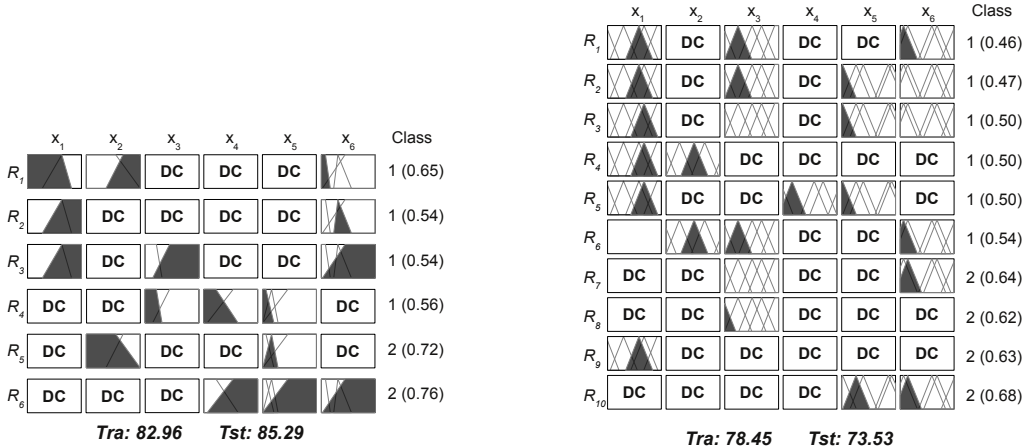
(a) Bupa rule set generated by D-MOFARC

(b) Bupa rule set generated by FARC-HD

Figure 3.3: Example rule sets obtained by the D-MOFARC(a) and the FARC-HD(b) approaches, respectively, when considering the third fold of the bupa dataset.

Although the results presented so far only consider the best solution obtained by the D-MOFARC method, it actually generates a set of Pareto-optimal solutions, each of them representing a trade-off between accuracy and complexity. Even though it was not our main purpose, it is possible for a decision maker to choose the solution that better satisfies the required trade-off for a certain problem. As an example, some Pareto-optimal sets are represented in Figure 3.5, in particular for heart, bupa, pima and segment datasets, respectively. Each Pareto front has been drawn by choosing a representative fold of the considered dataset and reporting the number of rules in the
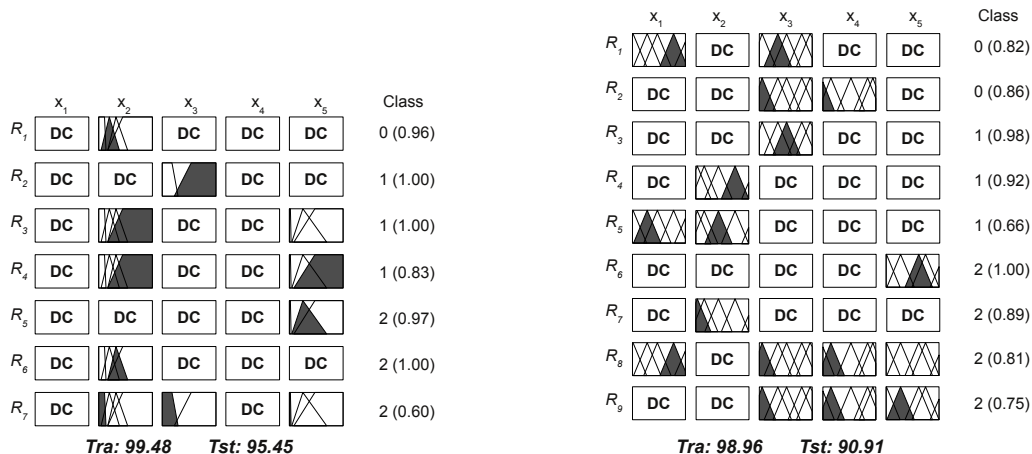
(a) Newthyroid rule set generated by D-MOFARC

(b) Newthyroid rule set generated by FARC-HD

Figure 3.4: Example rule sets obtained by the D-MOFARC(a) and the FARC-HD(b) approaches, respectively, when considering the third fold of the newthyroid dataset.

x-axis and the correct classification rate on training in the y-axis. The Pareto-optimal solutions are represented with dark square symbols in the graphs. In addition, the correspondent values for the correct classification rate on test are depicted with light square symbols. We have also depicted the correct classification rate on training (dark triangle symbols) and test (light triangle symbols) of the solution obtained by the application of the FARC-HD algorithm on the same folds. The observation of the Pareto fronts points out that in some cases there exist solutions with an accuracy level almost equivalent to the best solution's accuracy, while their complexity is considerably lower than the best solution's complexity.

3.6 Conclusion

In this chapter we have presented a multi-objective evolutionary method which performs concurrently a tuning process and a rule selection process on an initial KB of a FRBCS. In this method, a Fuzzy Discretization Algorithm has been integrated, in order to extract suitable granularities from data and generate the fuzzy partitions that constitute the initial DB. The associated RB has been generated by extracting a set of fuzzy association rules, according to the first two steps of the FARC-HD method presented in (15). The proposed MOEA is a modification of the SPEA2 algorithm and its aim is
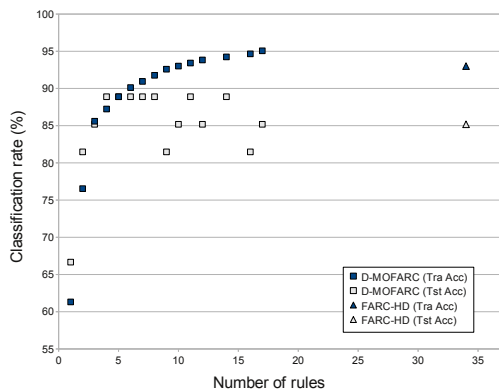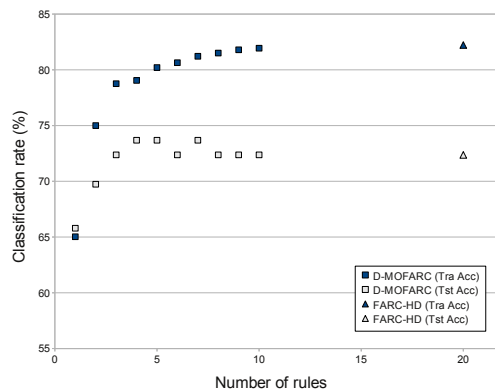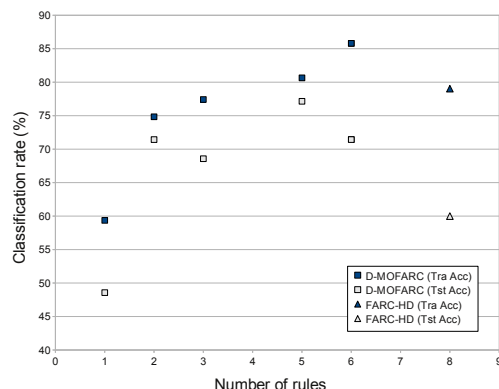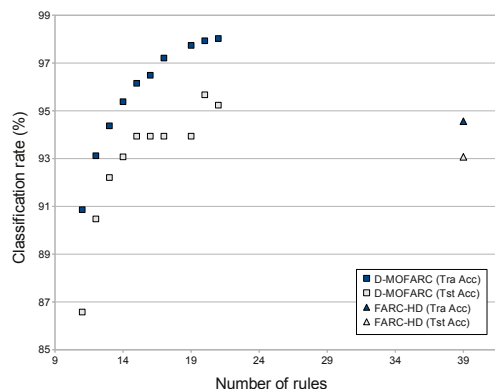
Figure 3.5: Pareto solutions obtained by the D-MOFARC method when considering representative folds of four datasets, namely heart(a), pima(b), bupa(c) and segment(d).

to generate compact and precise FRBCSs by considering concurrently two objectives: the correct classification rate and the number of rules of the obtained models.

The results obtained by comparing the proposed algorithm with a recent and well performing accuracy-driven GFS highlight that in this case our method clearly outperforms the FARC-HD when considering the precision of the obtained models, whereas less significant statistical difference has been observed when considering the complexity. Nevertheless, the mean results show that the number of rules is also reduced on average.

PART II
Instance Selection and GFSs

Chapter 4

A Study on the Application of Instance Selection Techniques with Genetic Fuzzy Rule-Based Classification Systems: Accuracy-Complexity Trade-Off

4.1 Introduction

The learning process of a GFS is strongly affected by the amount of instances used to generate the FRBS. A first problem is related to the computational time required by the fitness evaluation during the evolutionary process, since it is directly proportional to the number of instances. A second problem regards the complexity of the obtained models: in order to cover as much as possible instances of the dataset, the learning process tends to generate a high number of rules.

To reduce the amount of instances would speed up the learning process and possibly would lessen the complexity of the generated FRBSs. To this aim several approaches have been proposed in the literature. In particular, when considering medium and large datasets, the reduction can be obtained by applying techniques of Instance Selection (IS) (68, 86, 133, 163), which aim to extract a small representative subset of instances from the initial set, by removing superfluous instances. The subset should maintain all the information of the original set, so that it can be used to generate classification models with the same accuracy as models generated by using the original set.

4. APPLICATION OF IS TECHNIQUES WITH GFSS

4. APPLICATION OF IS TECHNIQUES WITH GFSS

IS techniques can be grouped into two categories, depending on the aim pursued after obtaining the reduced set:

- Prototype Selection (PS) methods (145): the reduced set is used by an instance-based classifier (for example K-NN) to classify new instances. Instance-based classifiers assume that unlabeled instances can be classified by relating them to the labeled instances, according to a certain similarity or distance function. The selected instances should provide the best trade-off between classification accuracy and reduction of the number of instances.

- Training Set Selection (TSS) methods (27, 37): the subset of instances is employed by a machine learning algorithm to build a predictive model (e.g. neural networks, FRBSs, decision trees, etc).

Several studies can be found in the literature for both PS and TSS. For example in (35), the authors perform IS by means of an evolutionary process. The quality of the reduced set is assessed by using the 1-NN classifier and a classification model constructed by the C4.5 algorithm. A comparison is carried out among evolutionary and non-evolutionary IS techniques, with respect to the classification accuracy and the instance reduction rate. This study has been subsequently extended in (36) and (37), where the concept of data stratification has been integrated in the framework with the aim of handling the scaling problem that appears when evaluating medium-large size datasets, and generating classification models with a good accuracy-interpretability trade-off.

A particular application of TSS is presented in (87). Here, the authors focus on classification problems in presence of imbalanced dataset. Data are re-balanced by undersampling the instances belonging to the majority class thorough a TSS method. TSS is integrated in an evolutionary algorithm and the quality of the reduced set is evaluated by generating a classification model with the well-known C4.5 algorithm.

A recent approach can be found in (27), where the authors investigate the use of TSS to reduce the set of instances required by a Multi-Objective Evolutionary Algorithm (MOEA) to generate FRBSs for regression problems. The TSS is integrated in a co-evolutionary framework: cyclically, a single-objective GA selects a subset of instances which are used by the MOEA for generating the FRBSs. The GA maximizes an index that measures the quality of the reduced set of instances.

In this chapter, we focus on the use of TSS techniques as pre-processing methods before applying a GA for generating Fuzzy Rule-Based Classification Systems (FR-BCSs). We aim to investigate if TSS techniques can help to reduce the complexity of

the generated FRBCSs, preserving or hopefully increasing their accuracy. A preliminary study discussed in (91), where a set of 20 small size datasets have been considered, has highlighted that a specific family of TSS methods is effective in achieving this objective.

This chapter extends the study in (91), in order to analyze the effects of instance selection preprocessing not only on small size datasets, but also on the kind of problems mainly addressed in this thesis, by considering medium-large size datasets, which frequently appear in real-world problems. For these datasets, the number of rules of the generated FRBCSs can be quite large and therefore their interpretability can be quite low, blurred by the complexity. We have considered 36 TSS techniques and 17 additional medium-large size datasets. The TSS techniques have been applied to each dataset and reduced datasets have been obtained. Then, the reduced datasets have been used to generate FRBCSs by exploiting a recently developed GFS, named Fuzzy Association Rule-based Classification model for High-Dimensional problems (FARC-HD) (15), which, as previously said, has been demonstrated to be efficient when working with high-dimensional datasets, i.e. datasets with a high number of variables. Since medium-large size datasets usually involve also a high number variables, FARC-HD results to be particularly suitable for these datasets. The goal is to understand if TSS techniques are able to decrease the number of instances in a dataset without losing the information needed for allowing FARC-HD to generate FRBCSs that achieve high classification rate despite a low complexity and a low computational time.

A further study has been performed by considering the combination of small and medium-large size datasets. The aim is to obtain more reliable results when applying statistical tests and to investigate if there exist TSS techniques that can be effectively used with datasets of any size.

Finally, an analysis of the computational time required by the application of TSS techniques and by the execution of the GFS on the reduced datasets is reported, in order to evaluate if the selected subsets lead to a reduction in the time required by the GFS to generate classification models.

This chapter is organized as follows: Section 4.2 contains a brief description of the IS process and IS methods in general. The methods used in this study are listed, according to the taxonomy proposed in (86). Section 4.3 describes the methodology used to carry out the experiments. Section 4.4 includes a brief overview of FARC-HD. In Section 4.5 the experimental framework is presented and the obtained results are examined and discussed. Finally, in section 3.6, some concluding remarks are stated.

4.2   Instance selection methods

Given a training set TR, the aim of an IS algorithm is to find a representative subset $S \subseteq TR$ of meaningful instances by removing superfluous instances (Figure 4.1). The resulting subset will be used to build a classifier.
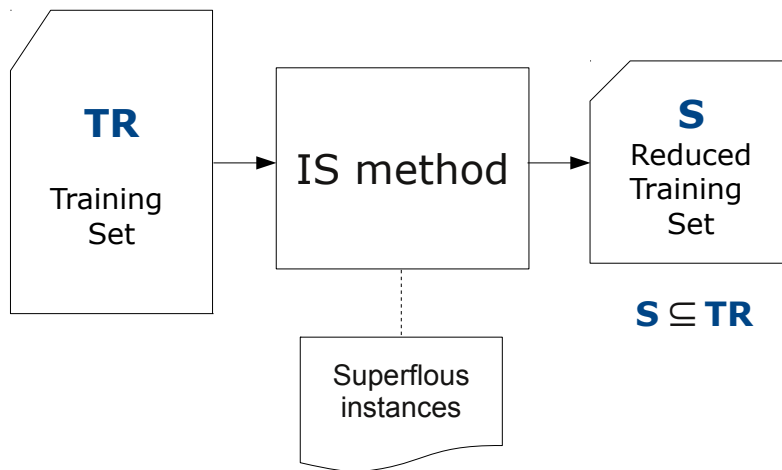
Figure 4.1: Instance Selection algorithm.

During the last years, more than fifty IS methods have been proposed in the literature and some reviews can be found in (28, 86, 90, 127, 133, 141, 163).

A comprehensive description of IS methods has been presented in a recent survey (86). Here, a taxonomy based on the main characteristics of the methods has been proposed, analyzing advantages and drawbacks of each of them. Hereinafter, the categories of IS methods are briefly described and then the methods selected for the current study are listed.

4.2.1   Classification of instance selection methods

IS methods have been grouped in different categories according to their common properties and to the description presented in (86):

- *Type of selection*: this characteristic is mainly influenced by the type of search strategy carried out by the IS algorithms, depending on the position of the instances to be retained with respect to the decision boundaries (border instances,

central instances or some other set of instances). The techniques are condensation, edition and hybrid. Condensation methods try to compute a consistent subset S by removing unnecessary instances that will not affect the classification accuracy on the training set. Edition methods aim to remove noisy instances, allowing the classifier to increase its accuracy. Hybrid methods search for a subset in which both noisy and unnecessary instances are concurrently eliminated.

- *Direction of search*: when searching for a subset S of instances from the training set TR, there are a variety of directions in which the search can proceed: incremental, decremental, batch, mixed and fixed. Incremental methods start with an empty subset S and add instances during the selection process according to some criterion; on the contrary decremental methods start with the whole training set ($S = TR$) and remove instances during the selection process. Batch methods examine all the instances before removing any of them and those instances that meet a removal criteria are eliminated at once. Mixed methods start with a subset S (that can be randomly chosen or obtained by an incremental or decremental process) and add or remove instances that meet a certain criterion. Finally, fixed methods act like mixed ones, but the number of instances to be added or removed is fixed in advance, thus the final number of selected instances is determined beforehand.

- *Evaluation of search*: according to the strategy used to add or remove instances in the subset S, the IS methods can be divided in wrapper and filter. Wrapper methods consider a selection criterion based on the accuracy obtained by a classifier. Usually, the instances that do not influence the accuracy classification are discarded. Filter methods use a selection function as selection criterion that is not based on a specific classifier.

- *Other properties*: there are other properties that can influence the results of an IS algorithm in combination with a given classifier. However these properties depend on the type of classifier used and therefore are not suitable for discriminating different types of instance selection techniques.

4.2.2   Instance selection methods used in this study

The characteristics explained above have been used in (86) to organize IS methods in a taxonomy, since they influence the behavior of the methods. According to this taxonomy, the methods chosen for the current study are shown in Figure 4.2.
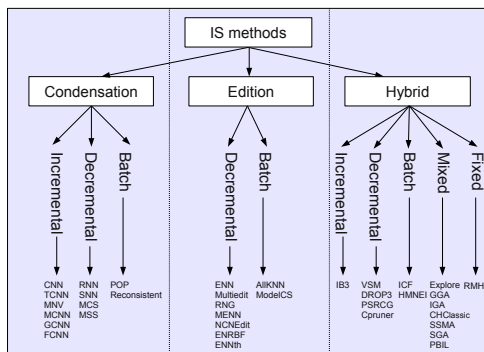
Figure 4.2: Taxonomy of Instance Selection methods.

This figure shows a classification hierarchy based on, first, type of selection and, then, direction of search. The selected methods are the most representative ones, according to the study described in (86). In Table 4.1, the methods are listed by reporting their complete and short names.

4.3 Experimental methodology for studying performances of IS methods as TSS techniques

In this section we give a description of the methodology used to assess the suitability of using IS methods together with a classification system (Figure 4.3) for TSS.

Let us consider a dataset DS consisting of a number N of instances. The instances are randomly split into a training set TR and a test set TS by using a K-fold cross-validation scheme. In particular, the instances are divided into K folds, each containing a percentage $1/K$ of the instances. Then, $K - 1$ folds are used as training set TR_i and the remaining fold is used as TS_i, where $i = 1, \cdots, K$. The analysis is performed on the TR_i, while the validation of the model is performed on the TS_i.

At each cross-validation round i, an IS method is applied to the TR_i, in order to extract a subset of instances S_i. The goal is to obtain a subset that does not include superfluous instances, is representative of TR_i, and preserves the original information contained in TR_i.

The extracted subset S_i is used by the GFS to generate a classifier, with the aim of obtaining a model with an accuracy $Acc(S_i)$ equal to or higher than the accuracy $Acc(TR_i)$ that would be obtained by using the complete training set TR_i to generate the classifier. The generalization abilities of the obtained classifier are evaluated on the

Table 4.1: Instance Selection methods used in the current study.

Complete name	Short name
All-KNN	AllKNN
C-Pruner	CPruner
CHC Evolutionary Algorithm	CHClassic
Condensed Nearest Neighbor	CNN
Decremental Reduction Optimization Procedure 3	DROP3
Edited Nearest Neighbor	ENN
Edited Nearest Neighbor Estimating	ENNTh
Edited Normalized Radial Basis Function	ENRBF
Explore	Explore
Fast Condensed Nearest Neighbor 1	FCNN
Generalized Condensed Nearest Neighbor	GCNN
Generational Genetic Algorithm	GGA
Hit Miss Network Edition Iterative	HMNEI
Instance Based 3	IB3
Intelligent Genetic Algorithm	IGA
Iterative Case Filtering	ICF
Minimal Consistent Set	MCS
Model Class Selection	ModelCS
Modified Condensed Nearest Neighbor	MCNN
Modified Edited Nearest Neighbor	MENN
Modified Selective Subset	MSS
MultiEdit	MultiEdit
Mutual Neighborhood Value	MNV
Nearest Centroid Neighbor Edition	NCNEdit
Patterns by Ordered Projections	POP
Population Based Incremental Learning	PBIL
Prototype Selection using Relative Certainty Gain	PSRCG
Random Mutation Hill Climbing	RMHC
Reconsistent	Reconsistent
Reduced Nearest Neighbor	RNN
Relative Neighborhood Graph Editing	RNG
Selective Nearest Neighbor	SNN
Steady-State Genetic Algorithm	SGA
Steady-State Memetic Algorithm	SSMA
Tomek Condensed Nearest Neighbor	TCNN
Variable Similarity Metric	VSM

test set TS_i. For each dataset, the results on training and test sets have been averaged over the K rounds.

To investigate the effectiveness of the IS techniques, the GFS has been also applied to the original TR_i (Figure 4.4). For each dataset, the results on the training and test sets have been averaged over the K rounds and they have been compared with the ones obtained by applying the IS methods.
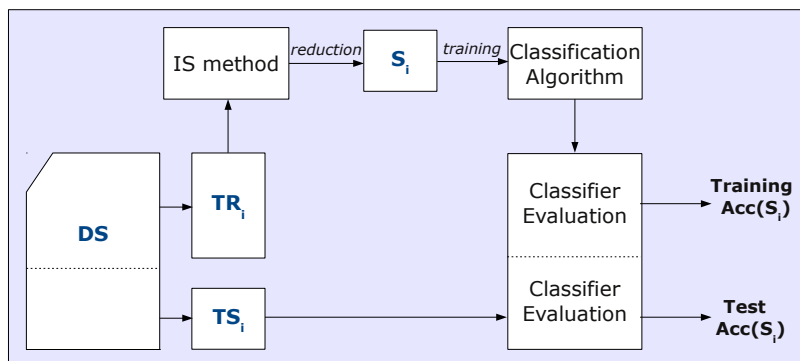
Figure 4.3: Application of IS methods to obtain a reduced set S_i that is consequently used to build a classifier.

4.4 Classification algorithm used in the study: FARC-HD

The aim of this study is to analyze the influence of TSS methods when combined with evolutionary classification algorithms. To this end, we choose a recently proposed classification algorithm, named FARC-HD (15), which has proved to obtain good performances and to be especially suitable for high-dimensional datasets. Hereinafter, a brief description of the algorithm is given.

FARC-HD uses fuzzy association rules to represent the information. These rules are an extension of the classical association rules, which are used to represent the dependencies between itemsets in a database (96, 167). Unlike classical association rules, fuzzy association rules can consider not only binary or discrete values, but also quantitative values. A fuzzy association rule can be represented as follows:

$$A \ is \ Middle \rightarrow B \ is \ High. \tag{4.1}$$
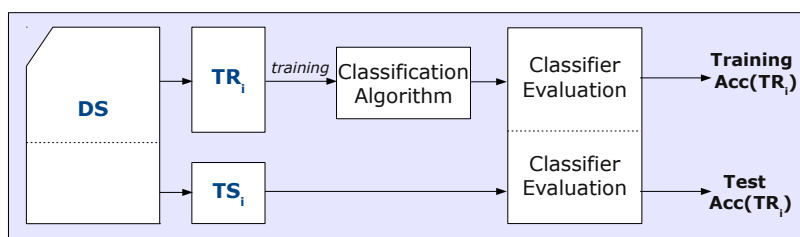
Figure 4.4: Using the original TR_i to construct a classifier.

Fuzzy association rules can be used for classification tasks if their antecedents contain fuzzy itemsets and their consequents contain only one class label. FARC-HD extracts a set of fuzzy association rules from the input data and uses them to build an FRBCS. The method consists of three steps, which will be briefly described.

- Extraction of fuzzy association rules for classification: A tree is built for each class, to list all the possible frequent itemsets, i.e. sets of items that frequently appear together in the dataset. After obtaining all the frequent itemsets, the rules are built by setting the frequent itemsets in the antecedent and the corresponding class in the consequent. The depth of the trees is fixed to $Depth_{max}$, in order to generate rules with few antecedent conditions and therefore to keep the model simple.

- Rule prescreening: this second step uses subgroup discovery to preliminarily reduce the rule set, by removing unnecessary rules. Each rule is evaluated using a modification of the wWRAcc' measure presented in (122), that correlates the quality of a rule with the number of input instances correctly covered and with their matching degrees with the antecedent part of the rule. After this prescreening process, only the most promising rules are maintained.

- Rule selection and Data Base tuning: the final step performs a rule selection process to further reduce the number of rules, by applying a genetic algorithm. Further, a tuning process of the membership functions is included in the genetic process, in order to obtain higher classification accuracy. The CHC genetic model (72) is used for the rule selection and tuning process, due to its ability in the exploration and the exploitation of the search space. The obtained final model is an FRBCS with low complexity and high accuracy.

4.5 Experimental framework

This section describes the experiments that have been performed to evaluate the effectiveness of the TSS methods. The results of each experiment are presented and analyzed to draw the conclusions. This section is divided into four parts:

- Experimental Set-Up.

- Results and analysis considering 17 medium-large datasets.

- Results and analysis considering 37 datasets (different sizes).

- Results and analysis of time performances.

4.5.1   Experimental set-up

In this study, 37 datasets have been considered, divided into small size and medium-large size datasets according to their number of instances. Datasets with more than 2000 instances are considered medium-large size datasets. Table 4.2 summarizes their characteristics, by reporting the number of instances (**#Inst**), the number of attributes (**#Attr**) (numerical and nominal attributes **Num/Nom** are highlighted, respectively) and the number of classes (**#Cls**) of each dataset. The web link to the Knowledge Extraction based on Evolutionary Learning(KEEL)-data set repository is also reported (16), from which the datasets can be downloaded. The instances that presented missing values have been removed from the datasets (in particular, from cleveland, crx, marketing datasets).

Table 4.2: List of the datasets used in the study, divided according to their number of instances.

Small DS	#Inst	#Attr(Num/Nom)	#Cls
australian	690	14 (8/6)	2
automobile	205	25 (15/10)	6
balance	625	4 (4/0)	3
bupa	345	6 (6/0)	2
cleveland	297(303)	13 (13/0)	5
contraceptive	1473	9 (9/0)	3
crx	653(690)	15 (6/9)	2
ecoli	336	7 (7/0)	8
german	1000	20 (7/13)	2
glass	214	9 (9/0)	7
haberman	306	3 (3/0)	2
heart	270	13 (13/0)	2
hepatitis	155	19 (19/0)	2
iris	150	4 (4/0)	3
newthyroid	215	5 (5/0)	3
pima	768	8 (8/0)	2
tae	151	5 (5/0)	3
vehicle	846	18 (18/0)	4
wine	178	13 (13/0)	3
wisconsin	699	9 (9/0)	2

Medium-Large DS	#Inst	#Attr(Num/Nom)	#Cls
abalone	4174	8 (7/1)	28
banana	5300	2 (2/0)	2
coil2000	9822	85 (85/0)	2
magic	19020	10 (10/0)	2
marketing	6876(8993)	13 (13/0)	9
optdigits	5620	64 (64/0)	10
page-blocks	5472	10 (10/0)	5
penbased	10992	16 (16/0)	10
phoneme	5404	5 (5/0)	2
ring	7400	20 (20/0)	2
satimage	6435	36 (36/0)	7
segment	2310	19 (19/0)	7
spambase	4597	57 (57/0)	2
texture	5500	40 (40 /0)	11
thyroid	7200	21 (21/0)	3
titanic	2201	3 (3/0)	2
twonorm	7400	20 (20/0)	2

http://sci2s.ugr.es/keel/datasets.php

A ten-fold cross validation scheme has been applied following the indications presented in Section 4.3 and for each round the method has been applied three times with different random seeds. Therefore, the results have been averaged over thirty executions for each dataset.

36 TSS techniques (Table 4.1) have been applied to each dataset and the reduced sets have been used by the FARC-HD algorithm to generate FRBCSs. For each TSS technique, the results have been averaged over the group of considered datasets. The FARC-HD algorithm has also been applied to the original datasets without using any TSS technique and again the results have been averaged over the group of considered datasets.

Thus, for each reduced dataset generated by a TSS technique and for the overall dataset, a single value is obtained for each metric. The following metrics have been computed:

- Instance reduction rate: percentage of instances that have been removed from the training set.

- Accuracy on the training set: percentage of training instances correctly classified.

- Accuracy on the test set: percentage of test instances correctly classified.

- Number of Rules: complexity of the obtained FRBCSs, computed as the average number of rules.

To analyze the results, non-parametric statistical tests have been considered, according to the recommendations made in (67). A wider description of these tests and a software for their application can be found on the web site available at: http://sci2s.ugr.es/sicidm/. In particular, the Wilcoxon's signed-rank test (162) has been applied to perform pairwise comparison between the results obtained by applying the FARC-HD algorithm to the original TS and to the reduced TSs generated by the different TSS techniques discussed in this chapter.

In order to ease the understanding of some parts of the following sections we included here the table of average results of the previous study (91) on small size datasets (Table 4.3), which will be referred in further sections.

4.5.2   Results and analysis considering 17 medium-large datasets

The average results obtained by applying the TSS techniques before FARC-HD to 17 medium-large datasets are presented in Table 4.4. We denote the sequences (specific TSS technique → FARC-HD) as name of the specific TSS technique followed by a hyphen and by FARC-HD. In the table, FARC-HD alone indicates the execution of FARC-HD on the original training set.

Table 4.3: Average results obtained by applying 36 TSS techniques to 20 small size data sets.

Inst Red Rate		Tra Acc		Tst Acc		#Rules	
Explore	0.980	FARC-HD	0.873	FARC-HD	0.785	Cpruner	2.530
CHC	0.980	POP	0.868	POP	0.778	CHC	3.390
SSMA	0.960	MSS	0.858	ModelCS	0.774	Explore	3.460
PBIL	0.950	ModelCS	0.853	MSS	0.768	SSMA	4.830
GGA	0.950	NCNEdit	0.830	NCNEdit	0.768	PBIL	6.240
SSGA	0.940	CNN	0.828	RNG	0.768	RNN	6.300
RNN	0.920	RNG	0.823	ENN	0.762	GGA	6.370
Cpruner	0.920	TCNN	0.823	AllKNN	0.751	SSGA	6.670
IGA	0.920	Reconsistent	0.823	Multiedit	0.750	ENRBF	7.430
MCNN	0.900	FCNN	0.822	ENNTh	0.749	IGA	8.200
RMHC	0.900	PSRCG	0.820	HMNEI	0.747	RMHC	8.320
DROP3	0.820	ENN	0.817	TCNN	0.745	MENN	8.630
ICF	0.750	MNV	0.812	CNN	0.741	MCNN	9.200
SNN	0.720	MCS	0.807	MENN	0.740	ENNTh	9.280
IB3	0.700	HMNEI	0.807	Reconsistent	0.740	Multiedit	11.130
FCNN	0.610	AllKNN	0.799	FCNN	0.735	AllKNN	12.070
MNV	0.610	IB3	0.796	PSRCG	0.734	DROP3	12.330
TCNN	0.600	GCNN	0.789	MCS	0.729	ICF	12.440
VSM	0.580	Multiedit	0.779	MNV	0.720	ENN	14.430
Reconsistent	0.580	ENNTh	0.776	GCNN	0.717	RNG	14.820
CNN	0.570	MENN	0.768	IB3	0.716	NCNEdit	15.070
MCS	0.570	ICF	0.763	RMHC	0.710	HMNEI	15.480
HMNEI	0.520	VSM	0.759	ICF	0.706	IB3	19.290
PSRCG	0.510	DROP3	0.739	ENRBF	0.692	SNN	20.510
GCNN	0.500	RMHC	0.736	GGA	0.686	ModelCS	21.160
MENN	0.470	ENRBF	0.713	DROP3	0.684	GCNN	21.240
ENNTh	0.460	SSGA	0.706	IGA	0.679	MNV	21.550
MSS	0.460	RNN	0.702	SSGA	0.677	TCNN	21.680
AllKNN	0.350	IGA	0.701	RNN	0.675	CNN	21.790
Multiedit	0.330	GGA	0.698	PBIL	0.673	FCNN	21.890
ENRBF	0.320	PBIL	0.694	VSM	0.655	VSM	21.890
ENN	0.250	SNN	0.685	Explore	0.647	MCS	22.070
RNG	0.250	SSMA	0.669	Cpruner	0.644	Reconsistent	22.220
NCNEdit	0.230	MCNN	0.668	SSMA	0.643	PSRCG	22.980
ModelCS	0.140	Explore	0.663	CHC	0.632	MSS	23.530
POP	0.100	Cpruner	0.653	MCNN	0.623	POP	23.600
FARC-HD	0.000	CHC	0.642	SNN	0.618	FARC-HD	24.480

Inst Red Rate, **Tra Acc**, **Tst Acc** and **#Rules** denote the instance reduction rate, the accuracies on the training and test sets, and the number of rules of the generated FRBCSs, respectively. Obviously, when FARC-HD is applied to the overall training set, the instance reduction rate is equal to 0. For each metric, the methods are sorted from the best to the worst. Table 4.4 shows that there exists an inverse proportionality between the complexity and the accuracy in most of the cases: the larger the number

Table 4.4: Average results obtained by applying 36 TSS techniques to 17 medium-large size data sets.

Inst Red Rate		Tra Acc		Tst Acc		#Rules	
CHClassic-FARC-HD	99.3	RNG-FARC-HD	83.7	RNG-FARC-HD	82.7	CHClassic-FARC-HD	8.8
Explore-FARC-HD	99.2	ModelCS-FARC-HD	83.3	ModelCS-FARC-HD	82.3	Explore-FARC-HD	10.7
MCNN-FARC-HD	98.7	POP-FARC-HD	83.2	FARC-HD	82.3	Cpruner-FARC-HD	14.0
SSMA-FARC-HD	98.5	FARC-HD	83.2	POP-FARC-HD	82.3	MCNN-FARC-HD	16.0
RNN-FARC-HD	95.9	NCNEdit-FARC-HD	82.7	AllKNN-FARC-HD	81.6	SSMA-FARC-HD	16.5
SGA-FARC-HD	94.8	AllKNN-FARC-HD	82.6	NCNEdit-FARC-HD	81.6	SNN-FARC-HD	17.0
GGA-FARC-HD	90.9	ENN-FARC-HD	82.5	ENN-FARC-HD	81.6	RNN-FARC-HD	18.6
DROP3-FARC-HD	90.6	ENNTh-FARC-HD	82.4	ENNTh-FARC-HD	81.5	IB3-FARC-HD	24.1
RMHC-FARC-HD	90.0	HMNEI-FARC-HD	82.2	IGA-FARC-HD	81.1	ENRBF-FARC-HD	24.4
PBIL-FARC-HD	89.8	IGA-FARC-HD	81.8	HMNEI-FARC-HD	81.0	SGA-FARC-HD	24.5
Cpruner-FARC-HD	89.3	Multiedit-FARC-HD	81.5	Multiedit-FARC-HD	80.5	DROP3-FARC-HD	24.6
SNN-FARC-HD	85.3	PBIL-FARC-HD	81.1	PBIL-FARC-HD	80.4	VSM-FARC-HD	24.7
ICF-FARC-HD	84.4	GGA-FARC-HD	80.7	GGA-FARC-HD	80.2	GGA-FARC-HD	25.6
VSM-FARC-HD	78.0	Reconsistent-FARC-HD	80.3	Reconsistent-FARC-HD	79.5	MCS-FARC-HD	26.1
IGA-FARC-HD	75.5	MENN-FARC-HD	79.7	MENN-FARC-HD	78.9	MNV-FARC-HD	26.1
IB3-FARC-HD	74.3	RMHC-FARC-HD	79.1	RMHC-FARC-HD	78.4	PBIL-FARC-HD	26.3
FCNN-FARC-HD	73.4	SGA-FARC-HD	78.3	SGA-FARC-HD	77.7	RMHC-FARC-HD	27.7
TCNN-FARC-HD	72.6	MSS-FARC-HD	78.0	MSS-FARC-HD	77.2	ICF-FARC-HD	28.5
CNN-FARC-HD	71.7	Cpruner-FARC-HD	76.1	Cpruner-FARC-HD	75.8	PSRCG-FARC-HD	29.5
PSRCG-FARC-HD	67.1	MCS-FARC-HD	75.5	MCS-FARC-HD	75.0	TCNN-FARC-HD	29.7
Reconsistent-FARC-HD	66.4	MNV-FARC-HD	75.5	MNV-FARC-HD	75.0	CNN-FARC-HD	29.9
MSS-FARC-HD	58.6	RNN-FARC-HD	74.7	RNN-FARC-HD	74.2	IGA-FARC-HD	30.5
GCNN-FARC-HD	57.7	IB3-FARC-HD	74.3	IB3-FARC-HD	73.6	FCNN-FARC-HD	30.5
MCS-FARC-HD	53.3	TCNN-FARC-HD	74.2	TCNN-FARC-HD	73.5	GCNN-FARC-HD	34.1
MNV-FARC-HD	53.3	CNN-FARC-HD	74.2	CNN-FARC-HD	73.5	MSS-FARC-HD	34.1
HMNEI-FARC-HD	51.4	FCNN-FARC-HD	74.1	FCNN-FARC-HD	73.4	HMNEI-FARC-HD	34.7
MENN-FARC-HD	31.4	PSRCG-FARC-HD	74.1	VSM-FARC-HD	73.3	Reconsistent-FARC-HD	37.2
ENRBF-FARC-HD	30.4	VSM-FARC-HD	73.9	PSRCG-FARC-HD	73.3	MENN-FARC-HD	37.3
ENNTh-FARC-HD	30.0	ICF-FARC-HD	73.3	ICF-FARC-HD	72.6	ENNTh-FARC-HD	38.3
Multiedit-FARC-HD	22.8	ENRBF-FARC-HD	71.7	ENRBF-FARC-HD	71.4	Multiedit-FARC-HD	41.5
AllKNN-FARC-HD	22.4	SSMA-FARC-HD	71.6	SSMA-FARC-HD	71.3	RNG-FARC-HD	43.9
ENN-FARC-HD	17.9	CHClassic-FARC-HD	71.2	CHClassic-FARC-HD	71.0	AllKNN-FARC-HD	43.9
NCNEdit-FARC-HD	16.4	DROP3-FARC-HD	69.4	Explore-FARC-HD	68.8	ENN-FARC-HD	44.0
RNG-FARC-HD	15.9	Explore-FARC-HD	69.2	DROP3-FARC-HD	68.7	NCNEdit-FARC-HD	44.1
POP-FARC-HD	8.7	GCNN-FARC-HD	66.8	GCNN-FARC-HD	66.4	FARC-HD	45.3
ModelCS-FARC-HD	8.6	MCNN-FARC-HD	64.0	MCNN-FARC-HD	63.6	ModelCS-FARC-HD	45.4
FARC-HD	0.0	SNN-FARC-HD	40.0	SNN-FARC-HD	39.4	POP-FARC-HD	45.6

of rules, the more accurate the obtained model. Nevertheless, the most accurate models are obtained by applying RNG-FARC-HD and ModelCS-FARC-HD, which also succeed in reducing the number of rules with respect to FARC-HD.

For the sake of brevity, among the 36 TSS methods, we will focus on the most promising ones, i.e. those methods that allow achieving a considerable reduction of the number of rules in the FRBCSs without penalizing their accuracy. To perform this analysis, we plot the last two columns of Table 4.4 in Figure 4.5. For the sake of clarity, the FARC-HD part of the name has not been reported. The x and y axes

represent the average values of the percentage of correct classification on the test set and of the number of rules of the obtained FRBCSs, respectively.

The methods represented with a gray circle belong to the set of the non-dominated solutions in the plane accuracy-complexity. A solution is non-dominated if none of its objective functions (accuracy and complexity in our case) can be improved without causing a worsening in the other objective function. Therefore, the methods that are non-dominated in the accuracy-complexity plane represent the best methods among the analyzed ones.

The non-dominated solutions, which are generated by using TSS selection methods of the same family, have been represented within a cloud to point out if methods of the same type appear in the same zone. We observe that the most accurate methods, namely RNG-FARC-HD and ENNTh-FARC-HD, belong to the *edition decremental* family, represented with a cloud filled with parallel vertical lines. Nevertheless, the reduction rate achieved by RNG-FARC-HD with respect to FARC-HD is modest.

A second group, represented within a uniform gray cloud, includes methods belonging to the *hybrid mixed* family. Almost all of them (except SGA-FARC-HD and CHClassic-FARC-HD) lead to a good reduction in the number of rules, while maintaining adequate accuracy on test ($\geq 80\%$) compared with the accuracy of FARC-HD (84.2%).

The last non-dominated method belongs to the *hybrid decremental* family (Cpruner-FARC-HD), and is represented within a cloud filled with parallel diagonal lines. It manages to reduce considerably the number of rules of the obtained models, but causing a severe detriment of the accuracy on the test set.

Among the methods belonging to the non-dominated set, the most promising ones have been selected by considering those that achieve an accuracy larger than 80%. The GGA-FARC-HD method has been excluded due to the fact that its complexity is almost equal to the complexity of PBIL-FARC-HD, but its accuracy is slightly lower. Four TSS methods have been finally selected to perform a statistical analysis on them. They are RNG-FARC-HD, ENNTh-FARC-HD, IGA-FARC-HD and PBIL-FARC-HD. The Wilcoxon's signed-rank test (162) has been applied to perform pairwise comparison both on accuracy and complexity, with a level of confidence $\alpha = 0.05$.

For each TSS technique, a single value per dataset, averaged over K rounds, has been used to perform the pairwise comparison. When considering the accuracy, the Wilcoxon's test is based on computing the differences between the average errors on the test set, whereas when considering the complexity the test is computed by taking
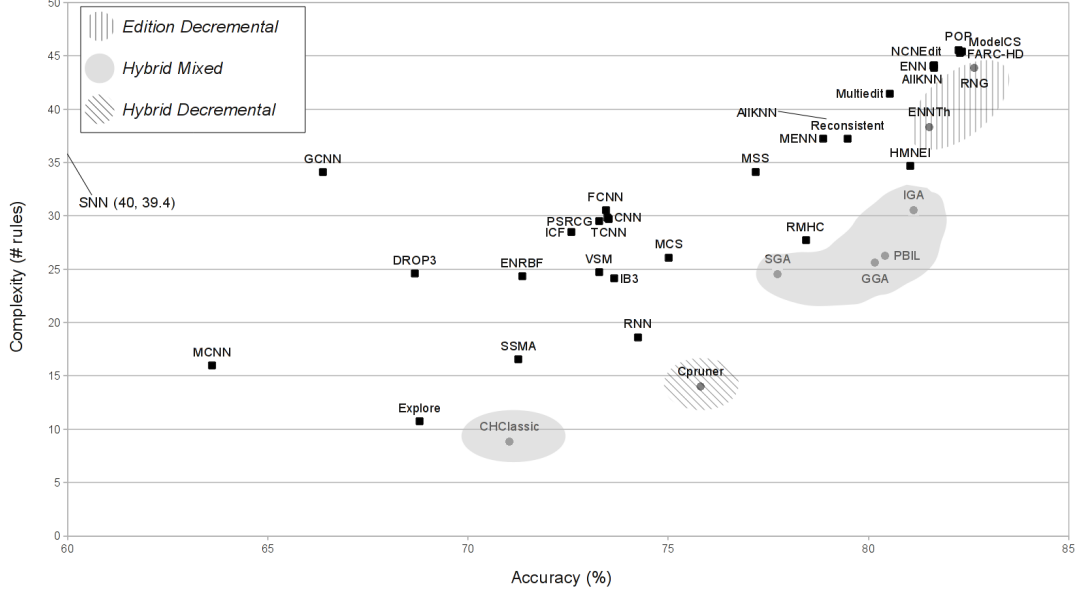
Figure 4.5: TSS methods in the accuracy-complexity plane

into account the average number of rules that are obtained by a pair of algorithms. We adopt a normalized difference $DIFF$ for the number of rules, which is defined as

$$DIFF = \frac{MeanRules(x) - MeanRules(RA)}{MeanRules(x)} \tag{4.2}$$

where $MeanRules(x)$ represents the number of rules obtained on average by the x algorithm and RA is the reference algorithm. This difference expresses the improvement in percentage with respect to the reference algorithm.

Table 4.5 shows the statistics obtained by applying the Wilcoxon's signed-rank test on the accuracy achieved on the test set, by comparing the results obtained by the FARC-HD algorithm with the ones obtained by the methods previously selected. In each row, the ranks R^+ and R^- and the p-value are shown. A further column is added to highlight if the Null-Hypothesis is rejected or not, with a significance level $\alpha = 0.05$.

The test rejects the Null-Hypothesis for two methods: IGA-FARC-HD and PBIL-FARC-HD. The R^- values corresponding to these methods are lower than the R^+ values corresponding to FARC-HD: we can conclude that these methods are statistically worse than FARC-HD in terms of accuracy on the test set. The Null-Hypothesis is not rejected in the case of the RNG-FARC-HD and ENN-FARC-HD methods. Thus, we can conclude that no significant difference is found with respect to the accuracy obtained on test set.

Table 4.5: Wilcoxon's statistics on the accuracy achieved on the test set obtained by comparing FARC-HD with the four most accurate non-dominated methods, considering 17 medium-large size datasets.

Comparison	R^+	R^-	p-value	Null-Hypothesis
FARC-HD vs RNG-FARC-HD	112.0	41.0	0.09838	Not Rejected
FARC-HD vs ENN-FARC-HD	86.5	49.5	0.32273	Not Rejected
FARC-HD vs IGA-FARC-HD	140.0	13.0	0.00134	Rejected
FARC-HD vs PBIL-FARC-HD	128.0	25.0	0.01286	Rejected

Table 4.6 shows the statistics obtained by applying the Wilcoxon's signed-rank test on the number of rules. The test rejects the Null-Hypothesis for two methods: IGA-FARC-HD and PBIL-FARC-HD. The R^+ and R^- values highlight that the rejected algorithms are statistically better than FARC-HD. Actually, in Table 3.3 we can observe that these two methods lead to models with an average number of rules considerably lower than the models obtained by applying FARC-HD directly to the original training set.

Table 4.6: Wilcoxon's statistics on number of rules obtained by comparing FARC-HD with the four most accurate non-dominated methods, considering 17 medium-large size datasets.

Comparison	R^+	R^-	p-value	Null-Hypothesis
RNG-FARC-HD vs FARC-HD	92.0	44.0	0.205204	Not Rejected
ENNTh-FARC-HD vs FARC-HD	78.0	75.0	0.924572	Not Rejected
IGA-FARC-HD vs FARC-HD	122.0	14.0	0.003356	Rejected
PBIL-FARC-HD vs FARC-HD	125.0	28.0	0.02016	Rejected

By observing all the results presented so far, we can draw some conclusions about the benefits of combining TSS techniques with a genetic FRBCS algorithm, when considering medium-large size datasets.

To this aim, we have computed the differences in percentage between the average results obtained by the four most accurate non dominated models and by the FARC-HD algorithm applied to the original training set, respectively. Table 4.7 shows these differences by considering as metrics the average accuracy on the test set (**TstAccDiff**) and the average number of rules (**RulesDiff**), respectively, taking the 17 medium large size datasets into account. The results highlight that, except for RNG-FARC-HD, the other analyzed TSS techniques cause a negligible loss (within a range of 3%) in accuracy. On the other hand, all the methods manage to decrease the complexity of the obtained models, although only IGA-FARC-HD and PBIL-FARC-HD achieve a significant reduction of the number of rules (-32.6% and -42%, respectively).

Table 4.7: Differences in percentage between the average values of accuracy on the test set (**TstAccDiff**) and of the number of rules (**RulesDiff**), obtained by the four most accurate non-dominated models and by FARC-HD, respectively, considering the 17 medium-large size datasets.

Method	TstAccDiff	RulesDiff
RNG-FARC-HD	0.4 %	-3.1 %
ENNTh-FARC-HD	-0.9 %	-15.4 %
IGA-FARC-HD	-1.4 %	-32.6 %
PBIL-FARC-HD	-2.3 %	-42.0 %

We can conclude that the TSS techniques belonging to the *edition decremental* family, i.e. RNG-FARC-HD and ENNTh-FARC-HD, are able to generate models with an accuracy almost equal to the one obtained by models generated by using the overall training set. Also, the reduction in complexity is not significant for RNG-FARC-HD, while it is more evident for ENNTh-FARC-HD.

On the other hand, two of the techniques belonging to the *hybrid mixed* family, i.e. IGA-FARC-HD and PBIL-FARC-HD, manage to reduce on average the number of rules by 30-40%, causing at the same time only a small loss in classification accuracy (1-2%).

These results are different from the ones obtained in the preliminary study discussed in (91), where a similar analysis has been performed by considering 20 small size datasets. Indeed, in (91), the *edition decremental* family has proved to be the most effective, since the reduction in complexity is about 50% on average while the accuracy on the test set is almost the same as for the FARC-HD applied to the overall training set. These techniques are listed in Table 4.8, which also shows the average accuracy obtained on the test set (**TstAcc**) in the 20 small size datasets, the difference in percentage of the accuracy achieved on the test set (**TstAccDiff**) with respect to FARC-HD applied to the original training set, the average number of rules (**#Rules**) and the difference in percentage of the number of rules (**RulesDiff**) with respect to the FRBCSs generated by FARC-HD applied to the original training set. On the contrary, *hybrid mixed* methods, which do not obtain good results when applied to small size datasets, seem to be more appropriate when dealing with medium-large size datasets.

4.5.3 Results and analysis considering 37 datasets (all dataset sizes)

The statistical analysis presented in the previous section considers 17 medium-large datasets. We have observed that the conclusions drawn by this analysis are different from the ones discussed in a preliminary work on the same subject, which however

Table 4.8: Average accuracy on the test set (**TstAcc**), average number of rules (**#Rules**) and difference in percentage of the average values of accuracy on the test set (**TstAccDiff**) and of the number of rules (**RulesDiff**) obtained by three models including edition decremental methods and by FARC-HD, respectively, considering 20 small size datasets.

Method	TstAcc	TstAccDiff	#Rules	RulesDiff
FARC-HD	78.51	0.0 %	24.5	0.0 %
NCNEdit-FARC-HD	76.79	-2.2 %	15.1	-38.4 %
RNG-FARC-HD	76.78	-2.2 %	14.8	-39.4 %
ENN-FARC-HD	76.25	-2.9 %	14.4	-41.0 %

had taken only small datasets into account. Thus, to investigate whether if there exists a TSS technique that is suitable for being used with datasets of any dimension, we combined the 20 small datasets analyzed in (91) (see this previous results in Table 4.3 in section 4.5.1) with the 17 medium-large datasets analyzed in this thesis and compared the best TSS techniques, namely ENN, RNG and NCNEdit, determined in (91) with the best TSS techniques, namely IGA and PBIL, identified from the previous analysis on medium-large size datasets performed in this chapter. The comparison is statistically validated by applying the Wilcoxon's signed rank test. The results are shown in Table 4.9. We observe that the Null-Hypothesis is rejected in all cases. We can conclude that none of the TSS techniques generates results that are statistically equivalent to the ones obtained when no TSS technique is applied.

Table 4.9: Wilcoxon's statistics on the accuracy achieved on the test set obtained by comparing FARC-HD with the selected methods, considering 37 datasets.

Comparison	R^+	R^-	p-value	Null-Hypothesis
FARC-HD vs ENN-FARC-HD	464	202	0.039300	Rejected
FARC-HD vs IGA-FARC-HD	683	20	5.398E-9	Rejected
FARC-HD vs NCNEdit-FARC-HD	490.5	212.5	0.035590	Rejected
FARC-HD vs PBIL-FARC-HD	657	46	2.776E-7	Rejected
FARC-HD vs RNG-FARC-HD	562	141	0.001059	Rejected

The Wilcoxon's signed rank test has been performed again by considering the normalized number of rules as metric and the results are shown in Table 4.10. In all cases the Null-Hypothesis is rejected, therefore the results obtained by applying the FARC-HD algorithm to the reduced sets of instances and to the original datasets are not statistically equivalent. The R^+/R^- values indicate that when a TSS technique is used, the obtained models are less complex.

Table 4.11 shows the average accuracy on the test set (**TstAcc**), the difference in percentage of the accuracy achieved on the test set (**TstAccDiff**) by the selected methods and by FARC-HD, respectively, the average number of rules (**#Rules**) and

Table 4.10: Wilcoxon's statistics on the number of rules obtained by comparing FARC-HD with the selected methods, considering 37 datasets.

Comparison	R^+	R^-	p-value	Null-Hypothesis
ENN-FARC-HD vs FARC-HD	583	120	2.69E-4	Rejected
IGA-FARC-HD vs FARC-HD	649	17	6.024E-9	Rejected
NCNEdit-FARC-HD vs FARC-HD	576	90	5.142E-5	Rejected
PBIL-FARC-HD vs FARC-HD	673	30	2.962E-8	Rejected
RNG-FARC-HD vs FARC-HD	587	79	1.936E-5	Rejected

the difference in percentage of the number of rules (**RulesDiff**). We observe that the application of IGA-FARC-HD and PBIL-FARC-HD methods reduces considerably the number of rules, but it also considerably worsens the accuracy performances. The other methods, i.e. ENN-FARC-HD, NCNEdit-FARC-HD and RNG-FARC-HD, lead to an average reduction of about 17% with respect to the number of rules, while the accuracy on the test set is only slightly decreased (-2% on average).

Table 4.11: Average accuracy on the test set (**TstAcc**), average number of rules (**#Rules**) and difference in percentage of the average accuracy on the test set (**TstAccDiff**) and of the average number of rules (**RulesDiff**), with respect to FARC-HD, considering 37 datasets.

Method	TstAcc	TstAccDiff	#Rules	RulesDiff
FARC-HD	80.25	0.0 %	34.05	0.0 %
RNG-FARC-HD	79.48	-1.0 %	28.17	-17.3 %
NCNEdit-FARC-HD	79.02	-1.5 %	28.42	-16.5 %
ENN-FARC-HD	78.71	-1.9 %	28.04	-17.6 %
IGA-FARC-HD	73.95	-7.8 %	18.45	-45.8 %
PBIL-FARC-HD	73.31	-8.6 %	15.45	-54.6 %

We can conclude that RNG, NCNEdit and ENN are the most promising TSS techniques to be used with GFS, since they induce the generation of less complex models without causing an excessive reduction in accuracy. On the other hand, the IGA and PBIL techniques lead to a considerable reduction of the complexity of the generated models despite a strong deterioration of the accuracy.

Probably, the best choice would be to apply different TSS techniques according to the size of the datasets. To this aim, Table 4.12 shows the average results obtained by applying RNG-FARC-HD for small size datasets, and IGA-FARC-HD and PBIL-FARC-HD for medium-large size datasets. These combinations lead to an improved balance between the reduction of the complexity of the generated FRBCSs and their accuracy, since they manage to reduce the average complexity by 38% on average, while the accuracy is only slightly decreased.

4.5.4 Results and analysis of time performances

In this section, the computational cost of using TSS techniques is analyzed. The time values reported in this analysis have been computed in the following way: for each dataset a TSS technique has been applied and the reduced training set has been obtained. Then, the FARC-HD algorithm has been applied to the reduced set of instances and the time required by its execution has been recorded.

This process has been repeated for each TSS technique and the results are shown in Figure 4.6 and in Figure 4.7, for small size datasets and medium-large size datasets, respectively. In the x-axis, the datasets are sorted in ascending order, according to the number of instances they include, while in the y-axis the time in seconds required to execute only the FARC-HD algorithm is reported.
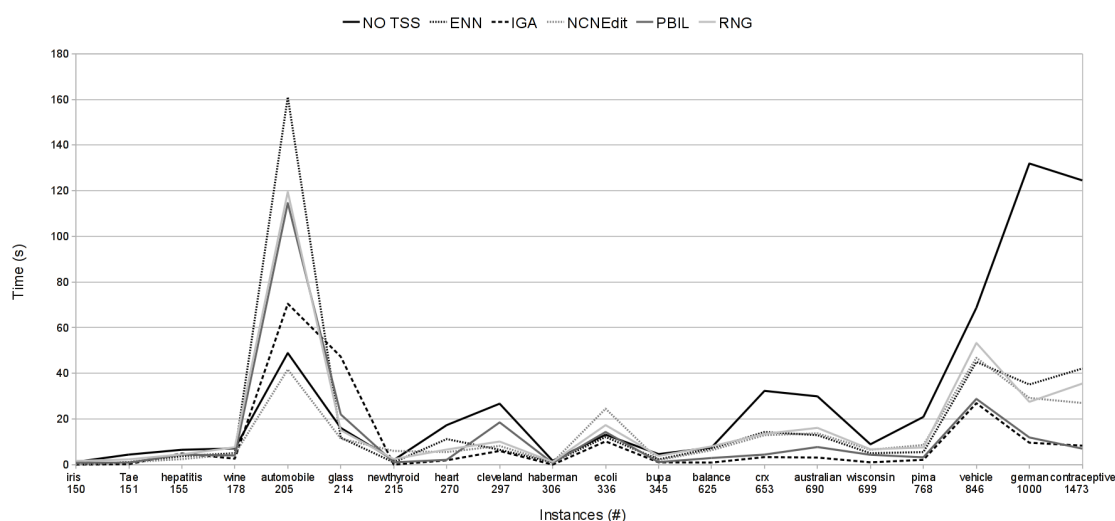
Figure 4.6: Time required by the FARC-HD algorithm to be performed for each small size dataset, without applying any TSS technique and after applying TSS techniques.

For each curve, a different TSS technique has been performed before applying the FARC-HD algorithm, except for the curve shown with a black solid line, which

Table 4.12: Average accuracy on the test set (**TstAcc**), average number of rules (**#Rules**) and difference in percentage of the average accuracy on the test set (**TstAccDiff**) and of the average number of rules (**RulesDiff**), with respect to FARC-HD, considering 37 datasets.

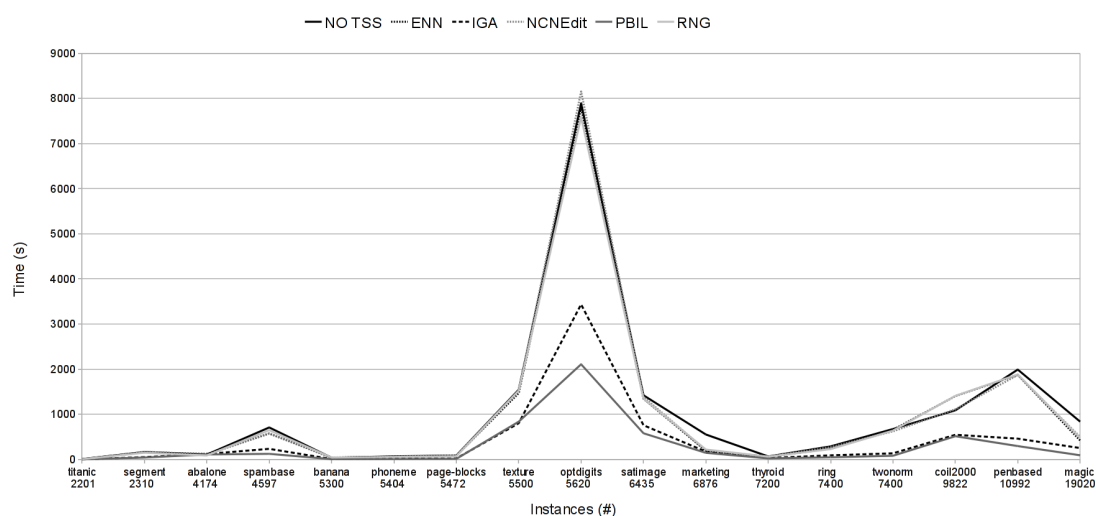| Method	TstAcc	TstAccDiff	#Rules	RulesDiff
FARC-HD	80.25	0.0 %	34.05	0.0 %
(RNG/IGA)-FARC-HD	78.78	-1.8 %	22.04	-35.3 %
(RNG/PBIL)-FARC-HD	78.46	-2.2 %	20.09	-41.0 %

Figure 4.7: Time required by the FARC-HD algorithm to be performed for each medium-large size dataset, without applying any TSS technique and after applying TSS techniques.

represents the time required to perform the FARC-HD algorithm without applying any TSS technique. In Figure 4.6, we notice that when the number of instances is smaller than 215 (new-thyroid dataset), the application of any TSS technique does not visibly reduce the time required by the FARC-HD algorithm to be performed. On the contrary, starting from this point in the graph, we can see that almost all the TSS techniques (excluding RNG and NCNEdit) manage to match or reduce the time required by the FARC-HD algorithm to be performed. In particular, starting from datasets with more than 345 instances (bupa dataset), the IGA (black dotted line) and PBIL (dark gray solid line) techniques are the most effective in reducing the time required by the FARC-HD algorithm. This conclusion is even more evident in Figure 4.7, in which the values of the curves corresponding to IGA and PBIL techniques are always equal to or smaller than the values of the other curves.

To better quantify the time reduction induced by each TSS technique on the execution of the FARC-HD algorithm, the values obtained for each curve have been averaged over all datasets and the results have been reported in the second column of Table 4.13. The table also shows in the first column the execution time of each TSS technique, averaged over all datasets, and in the third column the overall time required to perform each TSS technique and the FARC-HD algorithm. Finally, in the last column, the instance reduction rate for each TSS technique is reported. The first row includes the average time needed to execute the FARC-HD algorithm when no TSS technique is applied (the TSS time is 0 as well as the reduction rate).

Table 4.13: Execution times averaged over 37 datasets, for the TSS pre-processing and FARC-HD post-processing, respectively, overall execution times and instance reduction rate achieved by each TSS technique.

Method	TSS Time(s)	FARC-HD Time(s)	Overall Time(s)	Instance Reduction
NO TSS	0	489	489	0 %
ENN	2	456	458	-21.8 %
IGA	6730	198	6928	-84.2 %
NCNEdit	13	459	472	-20.2 %
PBIL	2995	142	3137	-92.5 %
RNG	161	453	614	-20.7 %
(RNG/IGA)	6711	202	6913	-48.1 %
(RNG/PBIL)	2985	145	3130	-54.7 %

As expected, the results highlight that the slower TSS techniques (IGA and PBIL) remove a larger number of instances and therefore the time required by the FARC-HD to be executed as a post-processing method is shorter. In particular, the PBIL method is the most effective, since its aggressive instance removal policy makes the execution of the FARC-HD algorithm three times faster than the execution of the FARC-HD algorithm when no TSS technique is applied. Nevertheless, taking into account that PBIL obtains too poor accuracy on the small size datasets and that the times for the RNG/PBIL combination are still very similar, we consider that such combination promotes the best equilibrium between performance and execution time of the FARC-HD algorithm.

On the other hand, if we consider the overall time needed to select a reduced set of instances with a TSS technique and to subsequently execute the FARC-HD algorithm on the reduced set, the results show that this time is frequently longer than the time needed to execute the FARC-HD algorithm on the original dataset, except when using ENN, NCNEdit and RNG techniques. In these cases, the overall computational time is almost equivalent to the time required by the execution of the FARC-HD algorithm, although the reduction obtained on the number of instances is modest (20%).

4.6   Conclusion

In this chapter we have presented an analysis of the influence of instance selection methods, used for training set selection, combined with a genetic fuzzy system for classification. The analysis was carried out by considering 36 methods and 37 datasets of different sizes in order to investigate if these methods are useful to decrease the

time of the evolutionary process and the complexity of the obtained models, although maintaining or improving their accuracy.

The analysis highlights that different IS techniques should be used depending on the dimension of the considered dataset. According to (91), we recommend to use techniques belonging to the *edition decremental* family, and in particular RNG when small size datasets are involved.

When dealing with medium-large size datasets, one of our main aims in this thesis, the IGA and PBIL techniques, belonging to the *hybrid mixed* family, have resulted to be the most suitable for generating models with a good trade-off between accuracy and complexity. Nevertheless, the execution of the PBIL technique is on average faster than the execution of the IGA technique, although the overall computational times required to perform both the TSS and the FARC-HD algorithm are longer than the time needed by the execution of the FARC-HD algorithm on the original training set. Therefore, the objective of using a TSS technique is not to reduce the overall computational time, but rather to improve the trade-off between accuracy and complexity, by reducing the number of rules while preserving most of the accuracy.

Chapter 5

Final Remarks

In the following, some overall conclusions are presented on the work developed in the thesis and on the results obtained. The publications associated with the thesis are listed and finally some ideas for possible future works are highlighted.

A   Summary and conclusions

This thesis focuses on the accuracy-complexity trade-off in MOEFSs, dealing with high dimensional and large scale classification problems. We illustrated two approaches to address this problem:

- The first strategy aims to learn the granularity of each variable appearing in a dataset, since it has been demonstrated that this approach produces a more appropriate DB, helping to improve the precision of the obtained models. The granularity learning has been combined with MOEAs, which are able to maintain or reduce the complexity during the evolutionary process thanks to their ability in managing two objectives concurrently. Moreover, they have been specifically designed to deal with dataset with a large number of variable and instances. To this end, two proposals have been presented: in the first one, we described a MOEFS that learns single granularities of the variables on the basis, of some heuristic measures computed on multiple granularities. This method achieves a reduction of the complexity, while the accuracy is slightly decreased. In the second proposal, we presented a MOEFS in which the granularity learning process is performed by using a fuzzy discretization algorithm. This method achieves an improvement of the accuracy, while the complexity is maintained at the same level or, in some cases, decreased.

- The second strategy addresses the problem through a pre-processing of the input data, by applying instance selection techniques in combination with a GFS. The study we performed highlighted that in the obtained models the complexity has been reduced considerably, without an appreciable reduction of the accuracy. A further analysis performed on the computational times evidenced that instance selection techniques are not useful to globally reduce the time required by the pre-processing plus the GFS application, since the total time increases.

B   Publications associated with this thesis

In the following, we present a list including the publications associated to this thesis:

- International journals:

  - M. Fazzolari, R. Alcala, Y. Nojima, H. Ishibuchi, F. Herrera, *A Review of the Application of Multiobjective Evolutionary Fuzzy Systems: Current Status and Further Directions*. IEEE Transactions on Fuzzy Systems, 21:1, 45-65 (2013).

  - M. Fazzolari, B. Giglio, R. Alcal, F. Marcelloni, F. Herrera, *A study on the application of instance selection techniques in genetic fuzzy rule-based classification systems: Accuracy-complexity trade-off*. Knowledge-Based Systems, in press, doi: 10.1016/j.knosys.2013.07.011 (2013).

  - M. Fazzolari, R. Alcala, F. Herrera, *A Multi-Objective Evolutionary Method for Learning Granularities Based on Fuzzy Discretization to Improve the Accuracy-Complexity Trade-Off of Fuzzy Rule-Based Classification Systems: D-MOFARC Algorithm*. Submitted to Applied Soft Computing.

- International conferences:

  - B. Giglio, F. Marcelloni, M. Fazzolari, R. Alcal, F. Herrera, *A Case Study on the Application of Instance Selection Techniques for Genetic Fuzzy Rule-Based Classifiers*. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2012), Brisbane (Australia), 920-927, 10-15 June 2012.

  - M. Fazzolari, R. Alcal, F. Herrera, *A Method for Learning Granularities to Improve the Accuracy-Complexity Trade-Off of Fuzzy Rule-Based Classifiers*. International Conference on Medical Imaging using Bio-Inspired and Soft Computing (MIBISOC 2013), Brussels (Belgium), (2013).

– M. Fazzolari, R. Alcala, Y. Nojima, H. Ishibuchi, F. Herrera, *Improving a Fuzzy Association Rule-Based Classification Model by Granularity Learning based on Heuristic Measures over Multiple Granularities*. IEEE International Workshop on Genetic and Evolutionary Fuzzy Systems (GEFS 2013), Singapore (Singapore), 44-51 (2013).

C   Future works

The conclusions presented so far pointed out that granularity learning and instance selection techniques represent valid tools to improve the balance between accuracy and complexity in FRBSs for high dimensional problems. Even so, there are other issues that should be considered. In the following we present some possible future investigation trends related with this problem.

- A possible research line can focus on the development of MOEFSs that consider not only complexity measures but also semantic interpretability measures. However, we should remember that there are still no measures commonly accepted by the scientific community. In some recent works new measures have been proposed to describe semantic interpretability (25, 83) and they have been applied to regression problems. A possibility would be to apply these measures to the case of MOEFSs for classification problems, with the aim of generating more interpretable DB from the semantic point of view.

- Regarding the instance selection, the study performed in this thesis pointed out that these techniques does not aim to reduce the total calculation time, but rather to improve the accuracy-complexity trade-off through the reduction of the rules of the obtained models. In this case, a possible improvement would be to specifically design instance selection techniques to be used in combination with GFSs and MOEFSs, in order to improve both objectives simultaneously.

- By analyzing the results presented in chapter 2 and 3, it is clear how the choice of appropriate granularities can influence the accuracy of FRBSs. One possible future research line can investigate the development of *wrapper* MOEFSs, in which the granularity learning process is performed within the evolutionary process based on the application of a fast rule induction process each time a DB needs to be evaluated. To this end, a fuzzy discretization algorithm can be used to initialize the chromosome that will be evolved by the MOEA.

- An alternative proposal to the use of instance selection techniques as pre-processing would be the use of instance reduction mechanisms integrated within the MOEA. A similar approach has been used in (27) for regression problems. Its advantage resides in the fact that the reduced set of examples is adapted during the evolutionary process, providing models with similar accuracy but reducing drastically the computational time required by the overall fitness evaluation.

References

[1] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo. Fast discovery of association rules. In *Advances in knowledge discovery and data mining*, pages 307–328. AAAI Press, 1996.

[2] A. S. Ahlawat and A. Ramaswamy. Multiobjective optimal structural vibration control using fuzzy logic control system. *J. Struct. Eng.*, 127(11):1330–1337, 2001.

[3] A. S. Ahlawat and A. Ramaswamy. Multi-objective optimal design of FLC driven hybrid mass damper for seismically excited structures. *Earthquake Eng. Struct. Dyn.*, 31(7):1459–1479, 2002.

[4] A. S. Ahlawat and A. Ramaswamy. Multiobjective optimal FLC driven hybrid mass damper system for torsionally coupled, seismically excited structures. *Earthquake Eng. Struct. Dyn.*, 31(12):2121–2139, 2002.

[5] A. S. Ahlawat and A. Ramaswamy. Multiobjective optimal fuzzy logic control system for response control of wind-excited tall buildings. *J. Eng. Mech.*, 130(4):524–530, 2004.

[6] A. S. Ahlawat and A. Ramaswamy. Multiobjective optimal fuzzy logic controller driven active and hybrid control systems for seismically excited nonlinear buildings. *J. Eng. Mech.*, 130(4):416–423, 2004.

[7] R. Alcalá, J. Alcalá-Fdez, M. J. Gacto, and F. Herrera. Improving fuzzy logic controllers obtained by experts: A case study in HVAC systems. *Appl. Intell.*, 31:15–30, 2009.

[8] 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 Trans. Fuzzy Syst.*, 15(4):616–635, 2007.

REFERENCES

[9] 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 Trans. Fuzzy Syst.*, 17(5):1106–1122, 2009.

[10] R. Alcalá, M. J. Gacto, and F. Herrera. A fast and scalable multi-objective genetic fuzzy system for linguistic fuzzy modeling in high-dimensional regression problems. *IEEE Trans. Fuzzy Syst.*, 19(4):666–681, 2011.

[11] 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 compact linguistic fuzzy rule-based systems. *Int. J. Uncertainty Fuzziness Knowledge Based Syst.*, 15(5):539–557, 2007.

[12] R. Alcalá, Y. Nojima, F. Herrera, and H. Ishibuchi. Multiobjective genetic fuzzy rule selection of single granularity-based fuzzy classification rules and its interaction with the lateral tuning of membership functions. *Soft Comput.*, 15(12):2303–2318, 2010.

[13] R. Alcalá, Y. Nojima, F. Herrera, and H. Ishibuchi. Multiobjective genetic fuzzy rule selection of single granularity-based fuzzy classification rules and its interaction with the lateral tuning of membership functions. *Soft Comput.*, 15:2303–2318, 2011.

[14] J. Alcalá-Fdez, R. Alcalá, M. J. Gacto, and F. Herrera. Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms. *Fuzzy Sets Syst.*, 160(7):905–921, 2009.

[15] J. Alcalá-Fdez, R. Alcalá, and F. Herrera. A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. *IEEE Trans. Fuzzy Syst.*, 19(5):857–872, 2011.

[16] J. Alcala-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. Garcia, L. Sanchez, and F. Herrera. Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *J. Multiple-Valued Logic Soft Comput.*, 17(2-3):255–287, 2011.

[17] J. Alcalá-Fdez, F. Herrera, F. Márquez, and A. Peregrín. Increasing fuzzy rules cooperation based on evolutionary adaptive inference systems: Research articles. *Int. J. Intell. Syst.*, 22:1035–1064, 2007.

[18] R. Alhajj and M. Kaya. Multi-objective genetic algorithms based automated clustering for fuzzy association rules mining. *J. Intell. Inf. Syst.*, 31:243–264, 2008.

[19] J. M. Alonso and L. Magdalena. HILK++: An interpretability-guided fuzzy modeling methodology for learning readable and comprehensible fuzzy rule-based classifiers. *Soft Comput.*, pages 1–22, 2010.

[20] J. M. Alonso, L. Magdalena, and O. Cordón. Embedding HILK in a three-objective evolutionary algorithm with the aim of modeling highly interpretable fuzzy rule-based classifiers. In *4th IEEE Int. Workshop Genet. Evol. Fuzzy Syst.*, pages 15–20, 2010.

[21] J. M. Alonso, L. Magdalena, and G. González-Rodríguez. Looking for a good fuzzy system interpretability index: An experimental approach. *Int. J.Approximate Reasoning*, 51(1):115–134, 2009.

[22] J. M. Alonso, L. Magdalena, and G. Serge. HILK: A new methodology for designing highly interpretable linguistic knowledge bases using the fuzzy logic formalism. *Int. J. Intell. Syst.*, 23:761–794, 2008.

[23] M. Antonelli, P. Ducange, B. Lazzerini, and F. Marcelloni. Learning concurrently partition granularities and rule bases of Mamdani fuzzy systems in a multi-objective evolutionary framework. *Int. J. Approximate Reasoning*, 50(7):1066–1080, 2009.

[24] M. Antonelli, P. Ducange, B. Lazzerini, and F. Marcelloni. Multi-objective evolutionary learning of granularity, membership function parameters and rules of Mamdani fuzzy systems. *Evol. Intell.*, 2(1-2):21–37, 2009.

[25] M. Antonelli, P. Ducange, B. Lazzerini, and F. Marcelloni. Learning concurrently data and rule bases of Mamdani fuzzy rule-based systems by exploiting a novel interpretability index. *Soft Comput.*, 15:1981–1998, 2011.

[26] 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 Comput.*, 15:2335–2354, 2011.

REFERENCES

[27] M. Antonelli, P. Ducange, and F. Marcelloni. Genetic training instance selection in multi-objective evolutionary fuzzy systems: A co-evolutionary approach. *IEEE Trans. Fuzzy Syst.*, 20(2):276–290, 2012.

[28] J. C. Bezdek and L. I. Kuncheva. Nearest prototype classifier designs: An experimental study. *Int. J. Intell. Syst.*, 16(12):1445–1473, 2001.

[29] A. L. Blumel, E. J. Hughes, and B. A. White. Multi-objective evolutionary design of fuzzy autopilot controller. In *Evolutionary multi-criterion optimization*, volume 1993, pages 668–680. Springer, 2001.

[30] P. Bonissone. Research issues in multi criteria decision making (MCDM): The impact of uncertainty in solution evaluation. In *Proc. 12th Int. Conf. Proc. Manag. Uncert. Knowledge-Based Syst.*, pages 1409–1416, 2008.

[31] L. Booker. *Intelligent behaviour as an adaption to the task environment*. PhD thesis, Univ. Michigan, 1982.

[32] 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 Comput.*, 13(5):437–449, 2009.

[33] R. Cannone, J. Alonso, and L. Magdalena. Multi-objective design of highly interpretable fuzzy rule-based classifiers with semantic cointension. In *5th IEEE Int. Workshop Genet. Evol. Fuzzy Syst.*, pages 1–8, 2011.

[34] R. Cannone, J. M. Alonso, and L. Magdalena. An empirical study on interpretability indexes through multi-objective evolutionary algorithms. In *Proc. Int. Conf. Fuzzy Logic. App.*, pages 131–138, 2011.

[35] J. Cano, F. Herrera, and M. Lozano. Using evolutionary algorithms as instance selection for data reduction in KDD: An experimental study. *IEEE Trans. Evol. Comput.*, 7(6):561–575, 2003.

[36] J. R. Cano, F. Herrera, and M. Lozano. Stratification for scaling up evolutionary prototype selection. *Pattern Recognition Letters*, 26(7):953–963, 2005.

[37] J. R. Cano, F. Herrera, and M. Lozano. Evolutionary stratified training set selection for extracting classification rules with trade off precision-interpretability. *Data Knowl. Eng.*, 60(1):90–108, 2007.

[38] C. J. Carmona, P. Gonzalez, M. J. del Jesus, and F. Herrera. NMEEF-SD: Non-dominated multiobjective evolutionary algorithm for extracting fuzzy rules in subgroup discovery. *IEEE Trans. Fuzzy Syst.*, 18(5):958–970, 2010.

[39] B. Carse, T. C. Fogarty, and A. Munro. Evolving fuzzy rule based controllers using genetic algorithms. *Fuzzy Sets Syst.*, 80(3):273–293, 1996.

[40] J. Casillas, F. Herrera, O. Cordón, and L. Magdalena. *Interpretability issues in fuzzy modeling*. Springer, 2003.

[41] J. Casillas, P. Martinez, and A. D. Benitez. Learning consistent, complete and compact sets of fuzzy rules in conjunctive normal form for regression problems. *Soft Comput.*, 13(5):451–465, 2009.

[42] J. Casillas and F. J. Martinez-Lopez. Mining uncertain data with multiobjective genetic fuzzy systems to be applied in consumer behaviour modelling. *Expert Syst. Appl.*, 36(2, Part 1):1645–1659, 2009.

[43] O. Castillo, P. Melin, A. Alanis, O. Montiel, and R. Sepulveda. Optimization of interval type-2 fuzzy logic controllers using evolutionary algorithms. *Soft Comput.*, 15(6):1145–1160, 2011.

[44] J. Catlett. On changing continuous attributes into ordered discrete attributes. In *Lecture notes in computer science*, volume 482, pages 164–178. Springer, 1991.

[45] N. V. Chawla, N. Japkowicz, and A. Kotcz. Editorial: Special issue on learning from imbalanced data sets. *SIGKDD Explor. Newsl.*, 6:1–6, 2004.

[46] C.-H. Chen, T.-P. Hong, V. S. Tseng, and L.-C. Chen. A multi-objective genetic-fuzzy mining algorithm. In *IEEE Int. Conf. Granular Comput.*, pages 115–120, 2008.

[47] T. Chen and H. Chen. Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks. *IEEE Trans. Neural Networks*, 6(4):904–910, 1995.

[48] W. C. Chen, N.-B. Chang, and J.-C. Chen. GA-based fuzzy neural controller design for municipal incinerators. *Fuzzy Sets Syst.*, 129:343–369, 2002.

[49] A. J. Chipperfield, B. Bica, and P. J. Fleming. Fuzzy scheduling control of a gas turbine aero-engine: A multiobjective approach. *IEEE Trans. Ind. Electron.*, 49(3):536–548, 2002.

REFERENCES

[50] M. Cococcioni, G. Corsini, B. Lazzerini, and F. Marcelloni. Solving the ocean color inverse problem by using evolutionary multi-objective optimization of neuro-fuzzy systems. *Int. J. Know. Based Intell. Eng. Syst.*, 12(5-6):339–355, 2008.

[51] M. Cococcioni, P. Ducange, B. Lazzerini, and F. Marcelloni. A Pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems. *Soft Comput.*, 11(11):1013–1031, 2007.

[52] M. Cococcioni, B. Lazzerini, and F. Marcelloni. On reducing computational overhead in multi-objective genetic Takagi-Sugeno fuzzy systems. *Appl. Soft Comput.*, 11(1):675–688, 2011.

[53] C. A. C. Coello, G. B. Lamont, and D. A. V. Veldhuizen. *Evolutionary algorithms for solving multi-objective problems.* Springer-Verlag, 2007.

[54] O. Cordón. A historical review of evolutionary learning methods for mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems. *Int. J. Approximate Reasoning*, 52(6):894–913, 2011.

[55] O. Cordón, M. J. del Jesús, J. Casillas, 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*, pages 79–99. Springer-Verlag, 2003.

[56] O. Cordón, M. J. del Jesus, and F. Herrera. A proposal on reasoning methods in fuzzy rule-based classification systems. *Int. J. Approximate Reasoning*, 20:21–45, 1999.

[57] O. Cordón and F. Herrera. A three-stage evolutionary process for learning descriptive and approximate fuzzy-logic-controller knowledge bases from examples. *Int. J. Approximate Reasoning*, 17(4):369–407, 1997.

[58] O. Cordón, F. Herrera, F. Hoffmann, and L. Magdalena. *Genetic fuzzy systems. Evolutionary tuning and learning of fuzzy knowledge bases.* World Scientific Publishing, 2001.

[59] O. Cordón, F. Herrera, and P. Villar. Analysis and guidelines to obtain a good uniform fuzzy partition granularity for fuzzy rule-based systems using simulated annealing. *Int. J. Approximate Reasoning*, 25(3):187–215, 2000.

[60] D. Daum and N. Morel. Identifying important state variables for a blind controller. *Build. Environ.*, 45(4):887–900, 2010.

[61] J. de Oliveira. Semantic constraints for membership function optimization. *IEEE Trans. Syst. Man Cybern. Part A: Syst. Humans*, 29(1):128 –138, 1999.

[62] K. Deb. *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, Inc., 2001.

[63] K. Deb and R. B. Agrawal. Simulated binary crossover for continuous search space. Technical Report 9, 1994.

[64] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.*, 6(2):182–197, 2002.

[65] M. Delgado, N. Marin, D. Sanchez, and M. A. Vila. Fuzzy association rules: General model and applications. *IEEE Trans. Fuzzy Syst.*, 11(2):214–225, 2003.

[66] M. R. Delgado, F. V. Zuben, and F. Gomide. Hierarchical genetic fuzzy systems. *Inf. Sci.*, 136(1-4):29–52, 2001.

[67] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.*, 7:1–30, 2006.

[68] J. Derrac, S. García, and F. Herrera. A survey on evolutionary instance selection and generation. *Int. J. of Applied Metaheuristic Computing*, 1(1):60–92, 2010.

[69] P. Ducange, B. Lazzerini, and F. Marcelloni. Multi-objective genetic fuzzy classifiers for imbalanced and cost-sensitive datasets. *Soft Comput.*, 14(7):713–728, 2010.

[70] M. Ebner, P. Stalph, M. Michel, and R. Benz. Evolutionary parameter optimization of a fuzzy controller which is used to control a sewage treatment plant. *Water Sci. Technol.*, 61:53–66, 2010.

[71] S. Elanayar V.T. and Y. Shin. Radial basis function neural network for approximation and estimation of nonlinear stochastic dynamic systems. *IEEE Trans. Neural Networks*, 5(4):594 –603, 1994.

REFERENCES

[72] L. J. Eshelman. The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. In *Workshop on the foundations of genetic algorithms*, pages 265–283, 1990.

[73] L. J. Eshelman and J. D. Schaffer. Real-coded genetic algorithms and interval-schemata. In *Foundation of Genetic Algorithms 2*, pages 187–202, 1993.

[74] L. J. Eshelman and J. D. Schaffer. Real-coded genetic algorithms and interval-schemata. In *Foundation of Genetic Algorithms 2*, pages 187–202, 1993.

[75] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, 17:37–54, 1996.

[76] U. M. Fayyad and K. B. Irani. On the handling of continuous-valued attributes in decision tree generation. *Machine Learning*, 8:87–102, 1992.

[77] P. Fazendeiro, J. V. Oliveira, and W. Pedrycz. A multiobjective design of a patient and anaesthetist-friendly neuromuscular blockade controller. *IEEE Trans. Biomed. Eng.*, 54(9):1667–1678, 2007.

[78] 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 Trans. Fuzzy Syst.*, 21(1):45–65, 2013.

[79] P. J. Fleming and R. C. Purshouse. Evolutionary algorithms in control systems engineering: A survey. *Control Eng. Pract.*, 10(11):1223–1241, 2002.

[80] C. M. Fonseca and P. J. Fleming. Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization. In *Proc. 5th Int. Conf. Genet. Algorithms*, pages 416–423, 1993.

[81] C. M. Fonseca and P. J. Fleming. On the performance assessment and comparison of stochastic multiobjective optimizers. In *Parallel problem solving from nature*, volume 1141, pages 584–593. Springer, 1996.

[82] 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 Comput.*, 13(5):419–436, 2009.

[83] M. J. Gacto, R. Alcalá, and F. Herrera. Integration of an index to preserve the semantic interpretability in the multiobjective evolutionary rule selection and tuning of linguistic fuzzy systems. *IEEE Trans. Fuzzy Syst.*, 18(3):515–531, 2010.

[84] M. J. Gacto, R. Alcalá, and F. Herrera. Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. *Inf. Sci.*, 181(20):4340–4360, 2011.

[85] M. J. Gacto, R. Alcalá, and F. Herrera. A multi-objective evolutionary algorithm for an effective tuning of fuzzy logic controllers in heating, ventilating and air conditioning systems. *Appl. Intell.*, 36:330–347, 2012.

[86] S. Garcia, J. Derrac, J. Cano, and F. Herrera. Prototype selection for nearest neighbor classification: taxonomy and empirical study. *IEEE Trans. Pattern Anal. and Mach. Intell.*, 34(3):417–435, 2012.

[87] S. García, A. Fernández, and F. Herrera. Enhancing the effectiveness and interpretability of decision tree and rule induction classifiers with evolutionary training set selection over imbalanced problems. *Appl. Soft. Comput.*, 9(4):1304–1314, 2009.

[88] 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 Comput.*, 13:959–977, 2009.

[89] S. García and F. Herrera. An extension on statistical comparisons of classifiers over multiple data sets for all pairwise comparisons. *J. Mach. Learn. Res.*, 9:2677–2694, 2008.

[90] N. García-Pedrajas. Evolutionary computation for training set selection. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 1(6):512–523, 2011.

[91] B. Giglio, F. Marcelloni, M. Fazzolari, R. Alcalá, and F. Herrera. A case study on the application of instance selection techniques for genetic fuzzy rule-based classifiers. In *IEEE Int. Conf. Fuzzy Syst.*, pages 920–927, 2012.

[92] A. F. Gómez-Skarmeta, F. Jiménez, and G. Sánchez. Improving interpretability in approximative fuzzy models via multiobjective evolutionary algorithms. *Int. J. Intell. Syst.*, 22(9):943–969, 2007.

REFERENCES

[93] J. González, I. Rojas, H. Pomares, L. J. Herrera, A. Guillón, J. M. Palomares, and F. Rojas. Improving the accuracy while preserving the interpretability of fuzzy function approximators by means of multi-objective evolutionary algorithms. *Int. J. Approximate Reasoning*, 44(1):32–44, 2007.

[94] O. Guenounou, A. Belmehdi, and B. Dahhou. Multi-objective optimization of TSK fuzzy models. *Expert Syst. Appl.*, 36(4):7416–7423, 2009.

[95] S. Guillaume. Designing fuzzy inference systems from data: An interpretability-oriented review. *IEEE Trans. Fuzzy Syst.*, 9(3):426–443, 2001.

[96] J. Han, M. Kamber, and J. Pei. *Data mining: Concepts and techniques*. Morgan Kaufmann, 2006.

[97] F. Herrera. Genetic fuzzy systems: Taxonomy, current research trends and prospects. *Evol. Intell.*, 1:27–46, 2008.

[98] F. Herrera, M. Lozano, and J. L. Verdegay. Tuning fuzzy logic controllers by genetic algorithms. *Int. J. Approximate Reasoning*, 12(3-4):299–315, 1995.

[99] F. Herrera and L. Martinez. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Trans. Fuzzy Syst.*, 8(6):746 –752, 2000.

[100] F. Hoffmann. Evolutionary algorithms for fuzzy control system design. *Proc. of IEEE*, 89(9):1318–1333, 2001.

[101] J. Holland and J. Reitman. Cognitive systems based on adaptive algorithms. In *Pattern directed inference systems*, pages 313–329. Academic Press, 1978.

[102] Y.-C. Hu. Determining membership functions and minimum fuzzy support in finding fuzzy association rules for classification problems. *Know. Based Syst.*, 19(1):57–66, 2006.

[103] Y.-C. Hu, R.-S. Chen, and G.-H. Tzeng. Finding fuzzy classification rules using data mining techniques. *Pattern Recogn. Lett.*, 24:509–519, 2003.

[104] H. Ishibuchi, T. Murata, and M. Gen. Performance evaluation of fuzzy rule-based classification systems obtained by multi-objective genetic algorithms. *Comput. Ind. Eng.*, 35(3-4):575–578, 1998.

[105] 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 Syst.*, 89(2):135–150, 1997.

[106] H. Ishibuchi and T. Nakashima. Effect of rule weights in fuzzy rule-based classification systems. *IEEE Trans. Fuzzy Syst.*, 9(4):506–515, 2001.

[107] H. Ishibuchi, T. Nakashima, and T. Murata. Three-objective genetics-based machine learning for linguistic rule extraction. *Inf. Sci.*, 136(1-4):109–133, 2001.

[108] H. Ishibuchi, T. Nakashima, and M. Nii. *Classification and modeling with linguistic information granules: Advanced approaches advanced approaches to linguistic data mining.* Springer, 2004.

[109] H. Ishibuchi, Y. Nakashima, and Y. Nojima. Double cross-validation for performance evaluation of multi-objective genetic fuzzy systems. In *5th IEEE Int. Workshop Genet. Evol. Fuzzy Syst.*, pages 31–38, 2011.

[110] H. Ishibuchi and Y. Nojima. Evolutionary multiobjective optimization for the design of fuzzy rule-based ensemble classifiers. *Int. J. Hybrid Intell. Syst.*, 3:129–145, 2006.

[111] H. Ishibuchi and Y. Nojima. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. *Int. J. Approximate Reasoning*, 44(1):4–31, 2007.

[112] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka. Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms. *Fuzzy Sets Syst.*, 65(2-3):237–253, 1994.

[113] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka. Selecting fuzzy IF-THEN rules for classification problems using genetic algorithms. *IEEE Trans. Fuzzy Syst.*, 3(3):260 –270, 1995.

[114] H. Ishibuchi, N. Tsukamoto, and Y. Nojima. Evolutionary many-objective optimization. In *3rd Int. Workshop Genet. Evolving Syst.*, pages 47–52, 2008.

[115] H. Ishibuchi and T. Yamamoto. Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. *Fuzzy Sets Syst.*, 141(1):59–88, 2004.

REFERENCES

[116] H. Ishibuchi and T. Yamamoto. Rule weight specification in fuzzy rule-based classification systems. *IEEE Trans. Fuzzy Syst.*, 13(4):428–435, 2005.

[117] M. Jamshidi, R. A. Krohling, L. dos S. Coelho, and P. J. Fleming. *Robust control systems with genetic algorithms*. Control series. CRC Press, 2003.

[118] Y. Jin. Fuzzy modeling of high-dimensional systems: Complexity reduction and interpretability improvement. *IEEE Trans. Fuzzy Syst.*, 8(2):212–221, 2000.

[119] F. Jurado and M. Valverde. Enhancing the electrical performance of a solid oxide fuel cell using multiobjective genetic algorithms. *Renewable Energy*, 30(6):881–902, 2005.

[120] N. Karnik, J. Mendel, and Q. Liang. Type-2 fuzzy logic systems. *IEEE Trans. Fuzzy Syst.*, 7(6):643–658, 1999.

[121] C. L. Karr and E. J. Gentry. Fuzzy control of PH using genetic algorithms. *IEEE Trans. Fuzzy Syst.*, 1(1):46, 1993.

[122] B. Kavšek, N. Lavrač, and V. Jovanoski. APRIORI-SD: Adapting association rule learning to subgroup discovery. In *Advances in Intelligent Data Analysis V*, volume 2810, pages 230–241. Springer, 2003.

[123] M. Kaya. Multi-objective genetic algorithm based approaches for mining optimized fuzzy association rules. *Soft Comput.*, 10:578–586, 2006.

[124] H.-S. Kim and P. N. Roschke. Fuzzy control of base-isolation system using multi-objective genetic algorithm. *Comput.-Aided Civ. Infrastruct. Eng.*, 21(6):436–449, 2006.

[125] H.-S. Kim and P. N. Roschke. GA-fuzzy control of smart base isolated benchmark building using supervisory control technique. *Adv. Eng. Software*, 38(7):453–465, 2007.

[126] J. Kim, Y. Moon, and B. P. Zeigler. Designing fuzzy net controllers using genetic algorithms. *IEEE Control Syst. Mag.*, 15(3):66–72, 1995.

[127] S.-W. Kim and B. J. Oommen. A brief taxonomy and ranking of creative prototype reduction schemes. *Pattern Anal. Appl.*, 6:232–244, 2003.

[128] W. Klösgen and J. M. Zytkow. Subgroup discovery. In *Handbook of data mining and knowledge discovery*, pages 354–367. Oxford University Press, 2002.

[129] J. D. Knowles and D. W. Corne. Approximating the nondominated front using the Pareto Archived Evolution Strategy. *Evol. Comput.*, 8(2):149–172, 2000.

[130] L. Kuncheva. *Fuzzy Classifier Design*. Springer, 2000.

[131] L. A. Kurgan and K. J. Cios. CAIM discretization algorithm. *IEEE Trans. on Knowl. and Data Eng.*, 16(2):145–153, 2004.

[132] C. C. Lee. Fuzzy logic in control systems: Fuzzy logic controller. I. *IEEE Trans. Syst. Man Cybern.*, 20(2):404–418, 1990.

[133] H. Liu and H. Motoda. On issues of instance selection. *Data Min. Knowl. Discov.*, 6(2):115–130, 2002.

[134] E. H. Mamdani. Application of fuzzy algorithms for control of simple dynamic plant. *Proc. of IEEE*, 121(12):1585–1588, 1974.

[135] A. A. Marquez, F. A. Marquez, and A. Peregrin. Rule base and inference system cooperative learning of Mamdani fuzzy systems with multiobjective genetic algorithms. In *Proc. Joint 2009 Int. Fuzzy Syst. Assoc. World Congr. 2009 Eur. Soc. Fuzzy Logic Technol. Conf.*, pages 1045–1050, 2009.

[136] A. A. Marquez, F. A. Marquez, and A. Peregrin. A multi-objective evolutionary algorithm with an interpretability improvement mechanism for linguistic fuzzy systems with adaptive defuzzification. In *IEEE Int. Conf. Fuzzy Syst.*, pages 1–7, 2010.

[137] P. Melin, D. Sanchez, and L. Cervantes. Hierarchical genetic algorithms for optimal type-2 fuzzy system design. In *Fuzzy Inf. Processing Soc. (NAFIPS), 2011 Annual Meeting of the North American*, pages 1–6, 2011.

[138] C. Mencar, G. Castellano, and A. M. Fanelli. Distinguishability quantification of fuzzy sets. *Inf. Sci.*, 177(1):130–149, 2007.

[139] R. Munoz-Salinas, E. Aguirre, O. Cordón, and M. Garcia-Silvente. Automatic tuning of a fuzzy visual system using evolutionary-algorithms: Single-objective versus multiobjective approaches. *IEEE Trans. Fuzzy Syst.*, 16(2):485–501, 2008.

[140] Y. Nojima and H. Ishibuchi. Incorporation of user preference into multi-objective genetic fuzzy rule selection for pattern classification problems. *Artif. Life Rob.*, 14:418–421, 2009.

REFERENCES

[141] J. Olvera-López, J. Carrasco-Ochoa, J. Martínez-Trinidad, and J. Kittler. A review of instance selection methods. *Artif. Intell. Review*, 34:133–143, 2010.

[142] F. P. Pach, A. Gyenesei, and J. Abonyi. Compact fuzzy association rule-based classifier. *Expert Syst. Appl.*, 34:2406–2416, 2008.

[143] J. Paredis. Coevolutionary computation. *Artif. Life*, 2(4):355–375, 1995.

[144] S. Park and H. Lee-Kwang. Designing fuzzy logic controllers by genetic algorithms considering their characteristics. In *Proc. 2000 Cong. Evol. Comput.*, volume 1, pages 683–690, 2000.

[145] E. Pekalska, R. P. Duin, and P. Paclík. Prototype selection for dissimilarity-based classifiers. *Pattern Recogn.*, 39(2):189–208, 2006.

[146] G. Piatetsky-Shapiro and W. J. Frawley. *Knowledge discovery in databases*. AAAI Press Copublications, 1991.

[147] P. Pulkkinen, J. Hytonen, and H. Kolvisto. Developing a bioaerosol detector using hybrid genetic fuzzy systems. *Eng. Appl. Artif. Intell.*, 21(8):1330–1346, 2008.

[148] P. Pulkkinen and H. Koivisto. Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms. *Int. J. Approximate Reasoning*, 48(2):526–543, 2008.

[149] P. Pulkkinen and H. Koivisto. A dynamically constrained multiobjective genetic fuzzy system for regression problems. *IEEE Trans. Fuzzy Syst.*, 18(1):161–177, 2010.

[150] G. L. O. Serra and C. P. Bottura. Multiobjective evolution based fuzzy PI controller design for nonlinear systems. *Eng. Appl. Artif. Intell.*, 19(2):157–167, 2006.

[151] M. Setnes, R. Babuska, and H. Verbruggen. Rule-based modeling: precision and transparency. *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, 28(1):165–169, 1998.

[152] C. Setzkorn and R. C. Paton. On the use of multi-objective evolutionary algorithms for the induction of fuzzy classification rule systems. *Biosyst.*, 81(2):101–112, 2005.

[153] D. Sheskin. *Handbook of parametric and nonparametric statistical procedures*. Chapman and Hall/CRC, V edition, 2011.

[154] D. A. Shook, P. N. Roschke, P.-Y. Lin, and C.-H. Loh. GA-optimized fuzzy logic control of a large-scale building for seismic loads. *Eng. Struct.*, 30(2):436–449, 2008.

[155] S. Smith. *A learning system based on genetic adaptive algorithms*. PhD thesis, Univ. Pittsburgh, 1980.

[156] P. Stewart, D. A. Stone, and P. J. Fleming. Design of robust fuzzy-logic control systems by multi-objective evolutionary methods with hardware in the loop. *Eng. Appl. Artif. Intell.*, 17(3):275–284, 2004.

[157] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man Cybern.*, 15(1):116–132, 1985.

[158] P. S. Thilagam and V. S. Ananthanarayana. Extraction and optimization of fuzzy association rules using multi-objective genetic algorithm. *Pattern Anal. Appl.*, 11(2):159–168, 2008.

[159] P. Thrift. Fuzzy logic synthesis with genetic algorithm. In *4th Int. Conf. Genet. Algorithms*, pages 509–513, 1991.

[160] C. Wagner and H. Hagras. A genetic algorithm based architecture for evolving type-2 fuzzy logic controllers for real world autonomous mobile robots. In *IEEE Fuzzy Syst. Conf. 2007*, pages 1–6, 2007.

[161] 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 Syst.*, 149(1):149–186, 2005.

[162] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

[163] D. R. Wilson and T. R. Martinez. Reduction techniques for instance-based learning algorithms. *Mach. Learn.*, 38:257–286, 2000.

[164] Z.-Y. Xing, Y. Zhang, Y.-L. Hou, and L.-M. Jia. On generating fuzzy systems based on Pareto multi-objective cooperative coevolutionary algorithm. *Int. J. Control Autom. Syst.*, 5(4):444–455, 2007.

REFERENCES

[165] L. Zadeh. Fuzzy sets. *Inf. Control*, 8:338–353, 1965.

[166] L. Zadeh. The concept of a linguistic variable and its application to approximate reasoning I. *Inf. Sci.*, 8(3):199–249, 1975.

[167] C. Zhang and S. Zhang. *Association rule mining: Models and algorithms*. Springer-Verlag, 2002.

[168] 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 Syst.*, 159(23):3091–3131, 2008.

[169] E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: improving the strength pareto evolutionary algorithm. Technical Report 103, 2001.

[170] E. Zitzler and L. Thiele. Multiobjective optimization using evolutionary algorithms - a comparative case study. In *Parallel Problem Solving from Nature*, volume 1498, pages 292–301. Springer, 1998.

[171] E. Zitzler, L. Thiele, M. Laumanns, C. Fonseca, and V. da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Trans. Evol. Comput.*, 7(2):117–132, 2003.