

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS INDUSTRIALES Y DE TELECOMUNICACIÓN

Titulación :

INGENIERO EN INFORMÁTICA

Título del proyecto:

“Estudio del efecto de las medidas de similaridad
intervalares difusas en sistemas de clasificación”

Alumno: David Jurío Pérez

Tutor: Jose Antonio Sanz

Pamplona, Mayo de 2010

INDICE

1. INTRODUCCIÓN.....	- 6 -
2. TEORÍA DE CONJUNTOS DIFUSOS	- 8 -
2.1 CONJUNTOS DIFUSOS	- 8 -
2.1.1. Conceptos básicos	- 8 -
2.1.2. Conectivos de los conjuntos difusos.....	- 11 -
2.1.3. Generalización de conectivos	- 12 -
2.1.4. Propiedades estructurales	- 16 -
2.2 CONJUNTOS DIFUSOS INTERVALO-VALORADOS	- 17 -
2.2.1. Características de los IVFS	- 17 -
2.2.2. Conectivos de los IVFS	- 19 -
2.3. MEDIDAS DE SIMILARIDAD	- 19 -
2.3.1. Similaridad de Mitchell	- 20 -
2.3.2. Similaridad de Gorzalczany.....	- 22 -
2.3.3. Similaridad de Bustince.....	- 22 -
2.3.4. Similaridad de Zeng y Li	- 23 -
2.3.5. Similaridad de Mendel	- 24 -
3. SISTEMAS DE CLASIFICACIÓN	- 26 -
3.1. Árboles de decisión	- 26 -
3.2. Redes neuronales artificiales	- 27 -
3.2.1 Neuronas Biológicas y artificiales.....	- 28 -
3.2.2. El aprendizaje de las redes neuronales artificiales:	- 31 -
3.3. SVM	- 31 -
3.4. Sistemas de Clasificación Basados en Reglas Difusas.....	- 32 -
3.4.1 Determinación del Método de Razonamiento Difuso:	- 35 -
3.4.4 Métodos para crear la Base de Conocimiento:	- 37 -
3.4.4.1 Algoritmo de Chi:.....	- 39 -
3.4.5.1 Método de Ishibuchi para el cálculo del grado de certeza:	- 41 -
4. SCBRD CON IVFS Y MEDIDAS DE SIMILARIDAD	- 42 -
4.1. Etiquetas Lingüísticas representadas mediante IVFS.....	- 42 -
4.2. Método de razonamiento difuso	- 43 -
4.2. Entradas del Sistema.....	- 44 -
4.2.1. Conjuntos Singleton	- 45 -
4.2.2. Conjuntos Difusos	- 45 -
4.2.2. Conjuntos IVFS	- 46 -

5. ESTUDIO EXPERIMENTAL	- 48 -
5.1 Configuración del experimento	- 48 -
5.2 Experimentos	- 49 -
5.2.1 Conjuntos Singleton como entrada.....	- 50 -
5.2.1.1 Resultados de Pruebas con Regla Ganadora y Producto	- 53 -
5.2.1.2 Resultados de Pruebas con Suma Aditiva y Producto.....	- 56 -
5.2.1.3 Resultados de Pruebas con Regla Ganadora y Mínimo.....	- 59 -
5.2.1.4 Resultados de Pruebas con Suma Aditiva y Mínimo	- 62 -
5.2.2 Conjuntos Difusos como entrada.....	- 64 -
5.2.2.1 Resultados de Pruebas con Regla Ganadora y Producto	- 65 -
5.2.2.2 Resultados de Pruebas con Suma Aditiva y Producto.....	- 67 -
5.2.3 Entrada con IVFS	- 70 -
5.2.3.1 Resultados de Pruebas con Regla Ganadora y Producto	- 70 -
5.2.3.2 Resultados de pruebas con Suma Aditiva y Producto	- 73 -
6. CONCLUSIONES.....	- 76 -
REFERENCIAS	- 80 -

INDICE DE FIGURAS

Ilustración 1. Ejemplo Representación de FS con representación triangular	- 9 -
Ilustración 2. Representación de FS con representación con función gaussiana.....	- 10 -
Ilustración 3. Representación de FS con representación de función exponencial.....	- 10 -
Ilustración 4. Conectivo complementación, punto de equilibrio.....	- 12 -
Ilustración 5. Ejemplo de IVFS	- 18 -
Ilustración 6. Creación de IVFS a partir de FS	- 18 -
Ilustración 7. Árbol de decisión para determinar recomendación o no de cirugía ocular - 27 -	
Ilustración 8. Representación de una neurona artificial	- 28 -
Ilustración 9. Ejemplo red neuronal	- 30 -
Ilustración 10. Esquema de un SCBRD.....	- 33 -
Ilustración 11. Ejemplo de IVFS representando la Altura.....	- 43 -
Ilustración 12. IVFS de tipo Singleton	- 45 -
Ilustración 13. Ejemplo Conjunto difuso.....	- 46 -
Ilustración 14. Ejemplo conjunto difuso IVFS.....	- 47 -

1. INTRODUCCIÓN

Uno de los tópicos en los que más se está investigando dentro de la Minería de Datos son los Sistemas de Clasificación. Estos sistemas son muy útiles hoy en día ya que extraen el conocimiento de grandes bases de datos de forma automática. El tamaño de las BD puede ser muy grande y por ello la tarea de extracción del conocimiento por parte del experto puede ser inviable.

El problema de clasificación, desde el punto de vista del aprendizaje supervisado, consiste en el establecimiento de una regla de decisión que permita determinar la clase de un nuevo objeto en una de las clases existentes y conocidas, $C_j \in C = \{1, \dots, M\}$, siendo M el número de clases. Cada uno de los objetos $e \in E$, a los que también se denomina ejemplos, se describe mediante un conjunto de observaciones $X(e) = (e_1, \dots, e_N)$ y a cada una de estas observaciones se le denomina variable, atributo o característica. El diseño de un *Clasificador o Sistema de Clasificación* se puede ver como la búsqueda de una correspondencia

$$D: X(E) \rightarrow C,$$

optimal en el sentido de un cierto criterio $\delta(D)$ que nos determine la bondad del Clasificador. Normalmente el objeto final, es el diseño de Sistema de Clasificación que asigne clases a cualquier punto del espacio de atributos con el mínimo error posible.

Dentro de la Minería de Datos existen varias técnicas que proponen una solución al problema, entre ellas:

- **Redes neuronales:** Se entrena el sistema con ejemplos de los que sabemos a qué clase pertenecen y automáticamente la red se ajusta de forma que sea capaz de clasificar nuevos ejemplos.
- **Sistemas expertos:** Éstos a partir de una base de reglas (obtenida en el entrenamiento) y un mecanismo de razonamiento difuso establecerán cual es la regla más adecuada y por tanto la clase a la que pertenece.

En este proyecto se van a emplear sistemas expertos de clasificación basados en reglas difusas (SCBRD) que es una herramienta muy útil para tratar problemas de clasificación. Estos están muy extendidos porque tienen la capacidad de construir un

modelo lingüístico interpretable por los usuarios y además permiten incorporar la información proveniente de un experto e información proveniente de modelos matemáticos o medidas empíricas. La importancia de los SCBRD se puede observar en la gran cantidad de las aplicaciones reales que hacen uso de ellos.

Estos sistemas emplean lógica difusa. La idea que subyace dentro de la misma es la de modelar la información procedente del mundo real, que por su propia naturaleza imprecisa y ambigua, no permite su total entendimiento mediante la lógica tradicional. La lógica difusa se utiliza cuando la complejidad del proceso en cuestión es muy alta y no existen modelos matemáticos precisos.

En este proyecto emplearemos el algoritmo de aprendizaje de reglas de Chi et al. Modificaremos la base de conocimiento que genere para incluir conjuntos intervalo-valorados difusos para manejar la incertidumbre asociada al proceso de definición de las funciones de pertenencia. El objetivo del proyecto es analizar la influencia de diferentes medidas de similaridad intervalares en el rendimiento del sistema.

En resumen, el presente proyecto trata de aplicar los conocimientos sobre la teoría de conjuntos difusos, conjuntos intervalo-valorados difusos y teoría de sistemas inteligentes, adquiridos en la asignatura “Inteligencia Artificial e Ingeniería del Conocimiento” durante el 4º curso de Ingeniería informática y asentados en la asignatura “Sistemas de Inteligencia Artificial” del 5º curso de la misma titulación.

En la memoria quedará reflejado el estudio realizado, además de capítulos teóricos acerca de conceptos que se utilizan en el proyecto. Así los capítulos serán: el segundo capítulo, es un capítulo teórico sobre lógica difusa, explicando aspectos como conjuntos difusos, conjuntos difusos intervalo valorados y medidas de similaridad; el tercer capítulo, será un recorrido teórico sobre todos los sistemas de clasificación existentes centrándose en los *sistemas de clasificación basados en reglas difusas*; en el cuarto, se describirá el modelo utilizado en el proyecto; en el quinto se mostrarán los experimentos realizados y los resultados obtenidos; y en el sexto se presentan las conclusiones.

2. TEORÍA DE CONJUNTOS DIFUSOS

La teoría sobre conjuntos difusos surgió como extensión a la teoría de conjuntos clásicos, por la necesidad de expresar incertidumbre en la función de pertenencia de un conjunto. Esta necesidad aparece al intentar representar la realidad en nuestro sistema, ya que en la realidad la definición de algo no siempre está delimitada de una forma precisa, y sobre todo el lenguaje humano describe de una forma vaga las cosas. Con vago, queremos decir por ejemplo que lo que significa bajo para unos no es lo mismo que para otros. Esta representación se puede realizar de diferentes formas.

En este capítulo veremos dos de las representaciones que se utilizarán en este proyecto: conjuntos difusos y conjuntos difusos intervalo valorados (IVFS). Además veremos en profundidad las diferentes medidas de similaridad que se utilizarán. Estas medidas nos dan un valor que indica la semejanza entre dos conjuntos difusos.

2.1 CONJUNTOS DIFUSOS

2.1.1. Conceptos básicos

Definición:

La lógica difusa es una extensión de la teoría de conjuntos clásica. Dado un conjunto X referencial finito y no vacío. Un conjunto difuso es:

$$A = \{(x, \mu_A(x)) \mid x \in X\} \text{ donde } \mu_A: X \rightarrow [0,1],$$

donde $\mu_A(x)$ es el grado de pertenencia del elemento x al conjunto A . $FS(X)$ representa todos los conjuntos difusos de X .

Variable Lingüística: Es una variable cuyos estados vienen representados por conjuntos difusos. Un ejemplo de variable lingüística es la altura, donde los conjuntos difusos son bajo, mediano o alto.

Las funciones de pertenencia se pueden modelar de muchas formas, algunas de las funciones utilizadas para modelar conjuntos difusos son: triangular, gaussiana, exponencial. En la Ilustración 1, puede verse como serían las funciones de pertenencia triangulares con las que modelamos las etiquetas lingüísticas que componen la variable lingüística. Por ejemplo la función de pertenencia que modela la etiqueta lingüística bajo sería la siguiente:

$$\mu_{BAJO}(X) = \begin{cases} \frac{1}{1,55-1,40}(x - 1,40) & \text{si } x \in [1,40,1,55] \\ \frac{1}{1,70-1,55}(x - 1,70) & \text{si } x \in [1,55,1,70] \\ 0 & \text{en cualquier otro caso} \end{cases}$$

En la Ilustración 1 podemos ver las funciones de pertenencia que modelan las etiquetas lingüísticas bajo, medio y alto es el siguiente:

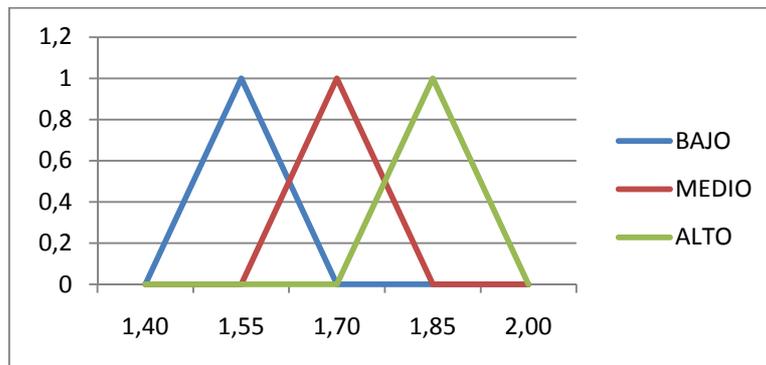


Ilustración 1. Ejemplo Representación de FS con representación triangular

En la Ilustración 2 se muestra como sería la representación de la variable lingüística altura mediante funciones de pertenencia gaussianas. El valor que toma la función de pertenencia para cada uno de los conjuntos se calcula mediante la expresión siguiente:

$$\mu_{bajo}(x) = e^{\left(\frac{-1(x-b)^2}{2c}\right)},$$

siendo b el punto donde la función toma valor 1 c la dispersión de los datos.

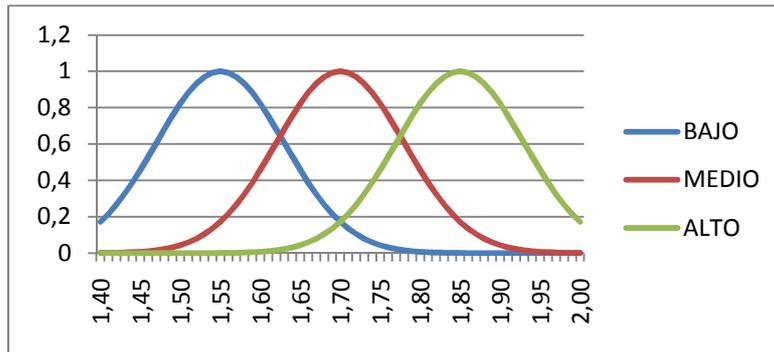


Ilustración 2. Representación de FS con representación con función gaussiana

Existen otras representaciones como las exponenciales, que vienen representadas por ejemplo con una fórmula de este tipo:

$$\mu_{bajo}(x) = e^{-|5(x-1)|}.$$

La Ilustración 3 muestra la representación de las etiquetas lingüísticas: bajo, medio y alto; según serían utilizando la fórmula mostrada.

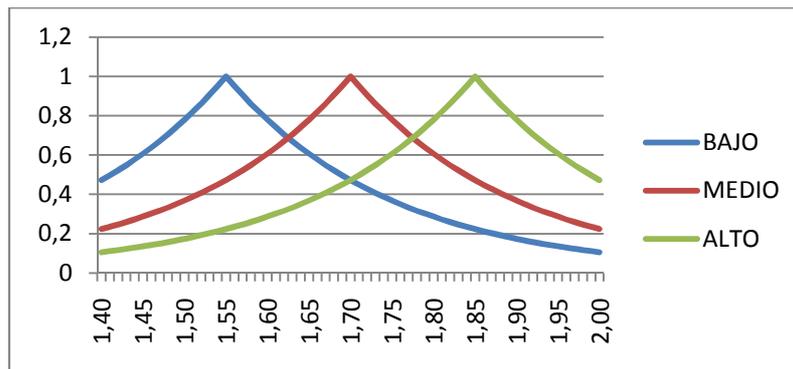


Ilustración 3. Representación de FS con representación de función exponencial

Dado un conjunto difuso A y $\alpha \in [0,1]$, llamamos α – **corte** de A a un conjunto clásico (conjunto cuyas funciones de pertenencia toman valores entre 0 y 1) que esté formado por los elementos que tienen pertenencia al conjunto $A \geq \alpha$. Esto se denota

$${}^{\alpha}A = \{x | \mu_A(x) \geq \alpha\}$$

Se llama **conjuntos de nivel** al conjunto formado por sus α – **cortes**.

Se denomina **soporte de un conjunto A** al “cero-corte fuerte” de A.

$$Sop(A) = {}^{0+}A = \{x | \mu_A(x) > \alpha\}$$

La **altura de un conjunto** es la mayor de las pertenencias a dicho conjunto $h(A) = \text{Sup}\{\mu_A(x)\}$.

Un conjunto se dice que es **normal**, si la altura es igual a 1, en caso contrario decimos que es un conjunto **subnormal**.

Llamamos **cardinal** de un conjunto A. $\text{Cardinal}(A) = \sum_i \mu_A(x_i)$, es la suma de la función de pertenencia.

2.1.2. Conectivos de los conjuntos difusos

Los conectivos se van a explicar para un conjunto X referencial $FS(X)$, si $A = \{(x, \mu_A(x)) | x \in X\}$

1. **Complementario** de A: $A_c = \overline{A} = A' = A_n$ será:

$$A_c = \{(x_j, \mu_{A_c}(x)) | x \in X\} \text{ tal que } \mu_{A_c}(x) = 1 - \mu_A(x) \forall x \in X.$$

Para este conectivo existe lo que se llama **punto de equilibrio** en el que se cumple:

$$\mu_A(x) = \mu_{A_c}(x).$$

En la Ilustración 6 aparecen representado el complementario de un conjunto difuso. Además las aspas rojas indican los puntos de equilibrio.

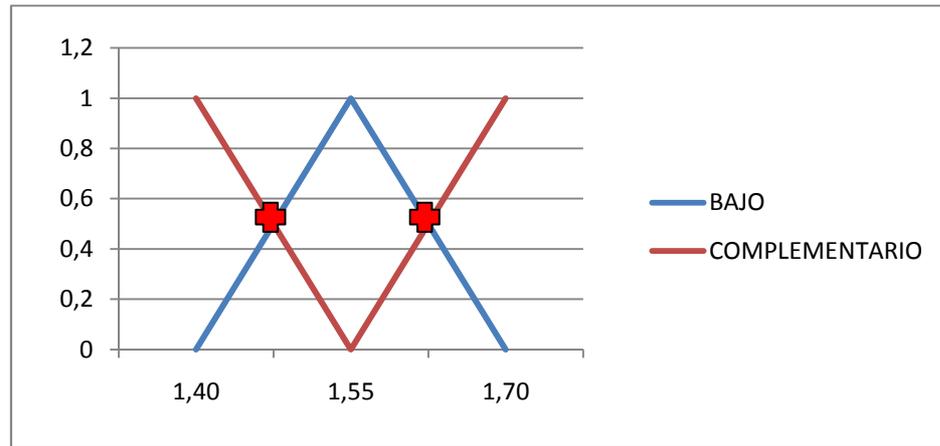


Ilustración 4. Conectivo complementación, punto de equilibrio

2. **Unión** de dos conjuntos $A, B \in FS(X)$. Se expresa de la siguiente forma:

$$A \cup B = \{(x, \mu_{A \cup B}(x)) | x \in X\} \text{ tal que } \mu_{A \cup B}(x) = \vee(\mu_A(x), \mu_B(x))$$

3. **Intersección** de dos conjuntos $A, B \in FS(X)$. Se expresa de la siguiente forma:

$$A \cap B = \{(x, \mu_{A \cap B}(x)) | x \in X\} \text{ tal que } \mu_{A \cap B}(x) = \wedge(\mu_A(x), \mu_B(x))$$

2.1.3. Generalización de conectivos

1. Complementario

Definición:

Una negación es una función $n: [0,1] \rightarrow [0,1]$ tal que cumple:

- 1) $n(0)=1$ y $n(1)=0$
- 2) Es decreciente: si $x \leq y$ entonces $n(x) \geq n(y)$

Una negación es **estricta** si:

- 3) es continua.
- 4) Es estrictamente decreciente: si $x < y$ entonces $n(x) > n(y)$

Una negación estricta es **involutiva** si:

- 5) $n(n(x))=x \forall x \in [0,1]$

Si se cumplen las propiedades 2 y 5 se cumplen las demás

Si una negación es continua, existe un único elemento $e \in (0,1)$ al que se le llama **punto de equilibrio** tal que $n(e)=e$. Este punto tiene la propiedad de ser invariante para la negación. Este punto nunca tomará los valores 0 o 1.

Existe una función de negación ideal: $n(x)=1-x$. Para la cual $n(0)=1$ y $n(1)=0$. Es estricta y es involutiva porque $n(n(x))=1-n(x)=1-(1-x)=x$. Además para $n(x)=1-x$ el punto de equilibrio es 0,5 ya que $n(0,5)=1-0,5=0,5$.

Definición:

Una función $\varphi: [a, b] \rightarrow [a, b]$ es un **automorfismo** en el intervalo $[a,b]$ si:

- Es continua
- Es estrictamente creciente
- $\varphi(a) = a, \varphi(b) = b$

Normalmente, será en $[0,1]$ con $\varphi(0) = 0, \varphi(1) = 1$

Teorema:

Una negación $n: [0,1] \rightarrow [0,1]$ es fuerte si y sólo si existe un automorfismo en el intervalo $[0,1]$ tal que:

$$n(x) = \varphi^{-1}(1 - \varphi(x)).$$

Ejemplo de negación es: $n(x)=1-x$. Para la cual $n(0)=1$ y $n(1)=0$.

Entonces, para este caso, el complemento viene dado por la siguiente expresión:

$$\text{Sea } A_c = \{(x, N(\mu_A(x)) = (1 - n(\mu_A(x))) \mid x \in X\}$$

2. Intersección.

Definición:

Una función $T: [0,1]^2 \rightarrow [0,1]$ se llama t-norma (intersección), si cumple:

1. $T(1,x)=x \forall x \in [0,1]$
2. $T(x,y)=T(y,x)$ (conmutativa) $\forall x, y \in [0,1]$
3. Crecimiento en ambos argumentos.

$$x \leq x' \Rightarrow T(x, y) \leq T(x', y)$$

$$y \leq y' \Rightarrow T(x, y) \leq T(x', y')$$

4. Asociatividad.

$$T(x, T(y, z)) = T(T(x, y), z) \quad \forall x, y, z \in [0,1]$$

La única t-norma que cumple la idempotencia $\wedge(x, x) = x$ es el mínimo. Además cualquier t-norma T es siempre menor igual que el mínimo $T \leq \wedge$.

$$T(A, B) = \{(x, \mu_{T(A, B)}(x) = T(\mu_A(x), \mu_B(x)) \mid x \in X\}$$

Teorema:

Una t-norma continua satisface $T(x, n(x))=0$ si y sólo si existe un automorfismo φ del intervalo unidad tal que:

$$T(x, y) = \varphi^{-1}(1 - \varphi(x)) \forall x \in [0,1] \text{ con } c(x) \leq \varphi^{-1}(1 - \varphi(x)) \forall x \in [0,1]$$

Teorema:

Una t-norma T continua es estricta si y sólo si existe un automorfismo φ tal que:

$$T(x, y) = \varphi^{-1}(\varphi(x) * \varphi(y))$$

Dados dos conjuntos difusos $A, B \in FS(X)$.

$$A = \{(x, \mu_A(x)) \mid x \in X\}$$

$$B = \{(x, \mu_B(x)) \mid x \in X\}$$

Entonces, la intersección viene dada por la siguiente expresión:

$$A \cap B = T(A, B) = \{(x, \mu_{A \cap B}(x) = \wedge (\mu_A(x), \mu_B(x)) \mid x \in X\}$$

3. Unión

Definición:

Una función $S: [0,1]^2 \rightarrow [0,1]$ se llama t-conorma si cumple:

1. $S(x,0)=x \forall x \in [0,1]$
2. $S(x,y)=S(y,x)$ (conmutativa) $\forall x, y \in [0,1]$
3. Crecimiento en ambos argumentos.

$$x \leq x' \Rightarrow S(x, y) \leq S(x', y)$$

$$y \leq y' \Rightarrow S(x, y) \leq S(x, y')$$

4. Asociatividad.

$$S(x, S(y, z)) = S(S(x, y), z) \quad \forall x, y, z \in [0,1]$$

La única t-conorma que cumple la idempotencia $\vee(x, x) = x$ es el máximo. Además cualquier t-conorma S es siempre mayor igual que el máximo $S(x, y) \leq \vee(x, y)$.

Cualquier t-norma T es siempre $T(x, y) \leq \wedge(x, y) \leq \text{Medias} \leq \vee(x, y) \leq S(x, y)$

Teorema:

Sea S una t-conorma continua satisface $S(x, n(x)) = \wedge$ (“tercio exclusivo”) si y sólo si existe un automorfismo φ del intervalo unidad tal que $S(x, y) = \varphi^{-1}(\wedge(\varphi(x) + \varphi(y), 1))$ con $n(x) \geq \varphi^{-1}(1 - \varphi(x))$

Dados dos conjuntos difusos $A, B \in FS(X)$.

$$A = \{(x, \mu_A(x)) \mid x \in X\}$$

$$B = \{(x, \mu_B(x)) \mid x \in X\}$$

La unión viene dada por la siguiente expresión:

$$A \cup B = T(A, B) = \{(x, \mu_{A \cup B}(x) = \vee (\mu_A(x), \mu_B(x)) \mid x \in X\}$$

2.1.4. Propiedades estructurales

Teorema:

Sea $FS(X)$. En estas condiciones: $\{FS(X), \cup, \cap\}$ es un retículo. Por ello cumple las siguientes propiedades:

- Idempotencia $\begin{cases} A \cup A = A \\ A \cap A = A \end{cases}$
- Conmutativa: $\begin{cases} A \cup B = B \cup A \\ A \cap B = B \cap A \end{cases}$
- Asociativa: $\begin{cases} A \cup (B \cap C) = (A \cup B) \cap C \\ A \cap (B \cup C) = (A \cap B) \cup C \end{cases}$
- Absorción $\begin{cases} A \cup (A \cap B) = A \\ A \cap (A \cup B) = A \end{cases}$

Teorema:

Sea un retículo $\{F(X), \cup, \cap\}$, dada la relación siguiente

$$A \leq B \text{ si } A \cap B = A, (A \cup B) = B,$$

que es reflexiva, antisimétrica y transitiva, decimos que tenemos un orden parcial.

Dados dos conjuntos $A, B \in FS(X)$. $A \subseteq B$ si y solo si $\mu_A(x) \leq \mu_B(x) \forall x \in X$.

$A=B$ si y solo si $A \subseteq B$ y $B \subseteq A$ y si y solo si $\mu_A(x) = \mu_B(x) \forall x$.

2.2 CONJUNTOS DIFUSOS INTERVALO-VALORADOS

2.2.1. Características de los IVFS

Los conjuntos difusos intervalo-valorados, aparecen como extensión a los conjuntos difusos vistos en la sección 2.1. En los conjuntos intervalo-valorados la función de pertenencia asigna a cada elemento del referencial x un subintervalo del intervalo cerrado $[0,1]$.

Concretamente para un conjunto X referencial finito y no vacío, un conjunto IVFS A sobre X es una expresión de la siguiente forma:

$$A = \{(x, A(x)) | x \in X\}$$

donde

$$A(x) = [\underline{A}(x), \bar{A}(x)] \in L([0,1])$$

siendo

$$L([0,1]) = \{x = [\underline{x}, \bar{x}] | (\underline{x}, \bar{x}) \in [0,1]^2 \text{ y } \underline{x} \leq \bar{x}\}$$

Obviamente, $A(x) = [\underline{A}(x), \bar{A}(x)]$ es el grado pertenencia en el punto $x \in X$.

La función de pertenencia $\forall x \in X$ viene dada por $\mu_A(x) = [\underline{A}(x), \bar{A}(x)] = [\underline{A}(x), \bar{A}(x)]$

Ejemplo:

$$X = \{x_1, x_2, x_3\}$$
$$X = \{(x_1, [0,0.2]), (x_2, [0.5,0.7]), (x_3, [0,0.9])\}$$

En la Ilustración 7 se puede observar como es un IVFS que representa la función de pertenencia al conjunto Bajo mediante un IVFS.

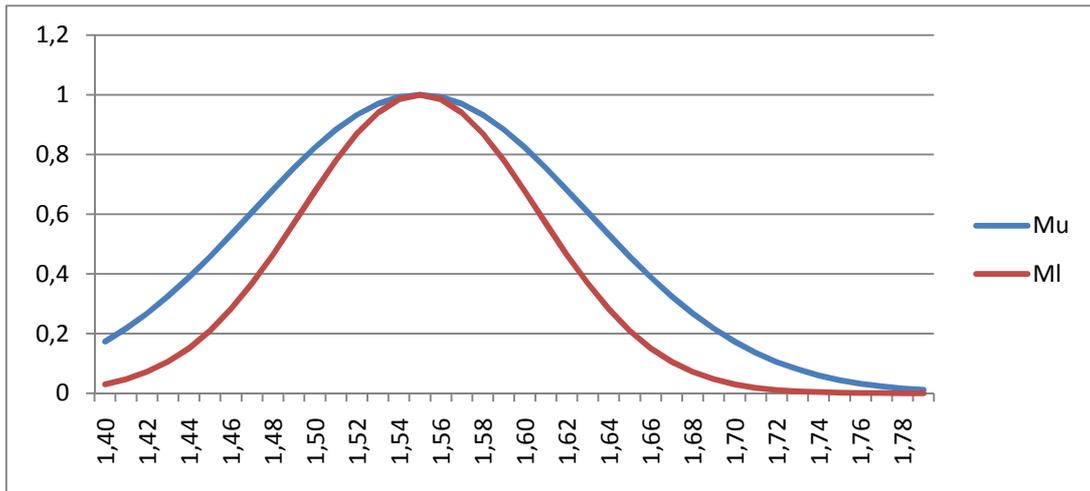


Ilustración 5. Ejemplo de IVFS Mu es $\bar{A}(x)$ y MI es $\underline{A}(x)$

Existen muchos métodos para crear IVFS a partir de conjuntos difusos. Uno de ellos, dado por Tizhoosh, para cada valor de x se le asocia un intervalo partiendo de la función de pertenencia difusa. Sea A perteneciente al FS(x), Tizhoosh constituye A' perteneciente al IVFS(x) de la siguiente forma:

- $\mu_{A'}(x) = [\underline{A}(x), \bar{A}(x)]$, donde
 - $\underline{A}(x) = (\mu_A(x))^{1/\alpha}$, $\alpha \geq 1$
 - $\bar{A}(x) = (\mu_A(x))^\alpha$, $\alpha \geq 1$

En la Ilustración 8 se puede observar el IVFS creado según la construcción exterior con $\alpha = 2$ y Mu es $\bar{A}(x)$ y MI es $\underline{A}(x)$.

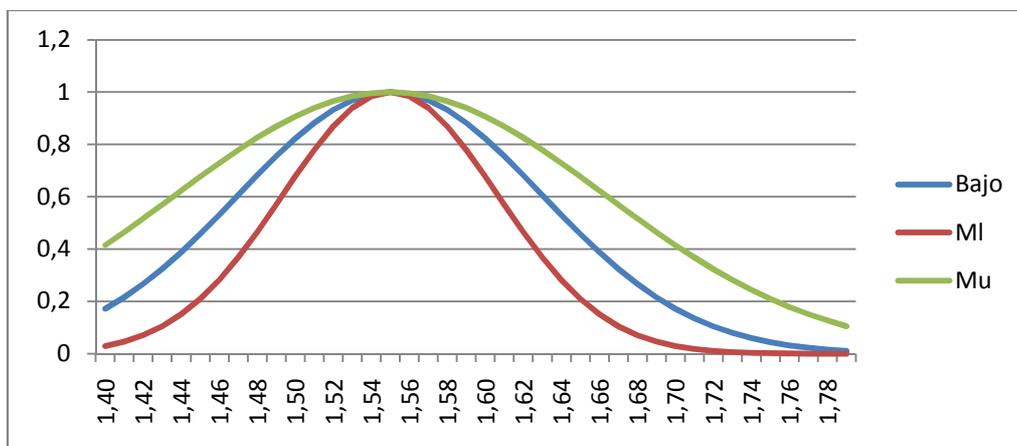


Ilustración 6. Creación de IVFS a partir de FS Mu es $\bar{A}(x)$ y MI es $\underline{A}(x)$

2.2.2. Conectivos de los IVFS

A continuación se describe el funcionamiento de los diferentes conectivos aplicados a los IVFS:

1. **Complementario.** Dado un conjunto difuso intervalo-valorado

$$A \in IVFS(x) | A = \{(x, M_A(x) = [\underline{A}(x), \overline{A}(x)] | x \in X\}.$$

Entonces el complementario de éste será el siguiente conjunto:

$$A_c = \overline{A} = A^c = \{(x, M_{A_c}(x) = [1 - \underline{A}(x), 1 - \overline{A}(x)] | x \in X\}$$

2. **Unión.** Dados dos conjuntos $A, B \in IVFS(x)$. La unión viene dada por la siguiente expresión:

$$A \cup B = \{(x, M_{A \cup B}(x)) = [\vee(\underline{A}(x), \underline{B}(x)), \vee(\overline{A}(x), \overline{B}(x))] | x \in X\}$$

Ambos conjuntos vienen definidos por las siguientes expresiones:

$$A = \{(x, A(x)) = [\underline{A}(x), \overline{A}(x)] | x \in X\}$$

$$B = \{(x, B(x)) = [\underline{B}(x), \overline{B}(x)] | x \in X\}$$

3. **Intersección.** Sean los conjuntos $A, B \in IVFS(x)$, dados en el punto anterior, la intersección de ambos se calculará de la siguiente manera:

$$A \cap B = \{(x, M_{A \cap B}(x)) = [\wedge(\underline{A}(x), \underline{B}(x)), \wedge(\overline{A}(x), \overline{B}(x))] | x \in X\}$$

2.3. MEDIDAS DE SIMILARIDAD

La literatura acerca de las medidas de similaridad entre conjuntos difusos es bastante extensa. De acuerdo a algunos autores hay más de 50 expresiones para determinar la similaridad.

Siguiendo el artículo “A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets” nos centraremos ahora en describir las medidas de similitud entre IVFS, que en este artículo se mencionan, ya que posteriormente serán las que utilizaremos en nuestro estudio.

La medida de similitud expresa la semejanza entre dos conjuntos IVFS. De esta manera, una función de similitud $SM: IVFS(x) * IVFS(x) \rightarrow [0,1]$ cumple:

1. $SM(A,B)=SM(B,A)$
2. $SM(A,A_c)=0$ si A es un conjunto clásico (petenencias con valor 0 ó 1)
3. $SM(A,B)=1$ si: $A=B$ o
4. Si $A \leq A' \leq B' \leq B$ entonces $SM(A,B) \leq SM(A',B')$
5. $SM(A,B)=SM(A_c,B_c)$

Las medidas de similitud que vamos a emplear en este proyecto son las de: Mitchell, Gorzalczany, Bustince, Zeng y Li ,y Mendel. Todas ellas aparecen en el artículo anteriormente mencionado.

2.3.1. Similitud de Mitchell

La primera similitud que definimos es la de H.B. Mitchell que define una medida de similitud para conjuntos difuso tipo 2. A continuación se muestra como se realiza el cálculo cuando \tilde{A} y \tilde{B} son de tipo 2.

Los pasos a realizar son los siguientes:

1. Discretizar la variable del universo del discurso, X, en L puntos, de forma que se usen en \tilde{A} y \tilde{B} .
2. Encontrar M conjuntos difusos tipo uno embebidos para el conjunto difuso tipo 2, \tilde{A} , de forma que para cada $m=1,2,\dots,M$ tenemos la función de pertenencia:

$$\mu_{A_e^m}(x_l) = r_m(x_l) \times [\bar{\mu}_{\tilde{A}}(x_l) - \underline{\mu}_{\tilde{A}}(x_l)] + \underline{\mu}_{\tilde{A}}(x_l) \quad l = 1,2, \dots, L$$

Donde $r_m(x_1)$ es un número aleatorio elegido dentro del intervalo $[0,1]$, y $\underline{\mu}_{\tilde{A}}(x_1)$ y $\bar{\mu}_{\tilde{A}}(x_1)$ forman el intervalo de pertenencia del elemento al conjunto en ese punto.

- De forma similar, se calculan N conjuntos difusos tipo uno embebidos para el conjunto difuso tipo 2, \tilde{B} , de forma que para cada $n=1,2,\dots,N$ tenemos:

$$\mu_{B_e^n}(x_1) = r_n(x_1) \times [\bar{\mu}_{\tilde{B}}(x_1) - \underline{\mu}_{\tilde{B}}(x_1)] + \underline{\mu}_{\tilde{B}}(x_1) \quad l = 1, 2, \dots, L$$

- Se computa la similaridad para un conjunto difuso tipo 2 como la media de similaridades de conjuntos difusos tipo 1, S_{mn} , que son calculadas para cada uno de las MN combinaciones de los conjuntos difusos de tipo 1, es decir:

$$S_M(\tilde{A}, \tilde{B}) = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N S_{mn} \text{ donde } S_{mn} = s(A_e^m, A_e^n)$$

donde los conjuntos con subíndice “e” son los embebidos.

En nuestro caso S_{mn} se calcula mediante $d_r(A, B)$, la medida de similaridad para conjuntos difusos dada por Minkowski en la expresión siguiente:

$$d_r(A, B) = \left(\sum_{i=1}^n |\mu_A(x_i) - \mu_B(x_i)|^r \right)^{\frac{1}{r}}, r \geq 1$$

Por tanto, a $S(A, B)$ deberemos asignarle el valor normalizado dado según la expresión:

$$S_N(A, B) = 1 - d_r(A, B)/n$$

Algunos de los problemas que presenta este algoritmo son:

- Generalmente la similaridad entre dos conjuntos \tilde{A} y \tilde{B} no es nunca 1 incluso en el caso de que sean exactamente idénticos. Esto es debido a la variable aleatoria introducida en la generación de los conjuntos embebidos de tipo 1.
- Otro problema es que debido a la aleatoriedad, los resultados pueden variar de experimento en experimento, y aunque con M y N muy grandes se produce una convergencia estocástica, el coste computacional también crece enormemente.

2.3.2. Similaridad de Gorzalczany

M.Gorzalczany propone una medida de compatibilidad para un conjunto difuso intervalo-valorado. La medida resultante es un intervalo y se trata de una medida que no es simétrica. Esto quiere decir que para dos IVFS \tilde{A}, \tilde{B} $SM(A, B) \neq SM(B, A)$. Así Gorzalczany define el grado de compatibilidad $S_G(\tilde{A}, \tilde{B})$ como:

$$S_G(\tilde{A}, \tilde{B}) = [S_{\min}, S_{\max}]$$

donde,

$$S_{\min} = \min \left(\frac{\max_{x \in X} \{ \min(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}}(x)) \}}{\max_{x \in X} \underline{\mu}_{\tilde{A}}(x)}, \frac{\max_{x \in X} \{ \min(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}}(x)) \}}{\max_{x \in X} \overline{\mu}_{\tilde{A}}(x)} \right)$$

$$S_{\max} = \max \left(\frac{\max_{x \in X} \{ \min(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}}(x)) \}}{\max_{x \in X} \underline{\mu}_{\tilde{A}}(x)}, \frac{\max_{x \in X} \{ \min(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}}(x)) \}}{\max_{x \in X} \overline{\mu}_{\tilde{A}}(x)} \right)$$

2.3.3. Similaridad de Bustince

H.Bustince también propone una medida de similaridad para conjuntos difusos intervalo-valorados. Cabe destacar que como en el caso de la similaridad definida por M.Gorzalczany, la similaridad entre dos IVFS va a ser un intervalo. En este caso satisface las siguientes propiedades:

1. $S_B(\tilde{A}, \tilde{B}) = S_B(\tilde{B}, \tilde{A})$
2. $S_B(\tilde{A}, c(\tilde{A})) = 0$
3. $S_B(\tilde{A}, \tilde{A}) = [1, 1]$
4. Si $\tilde{A} \leq \tilde{B} \leq \tilde{C}$, entonces $S_B(\tilde{B}, \tilde{A}) \geq S_B(\tilde{C}, \tilde{A})$ y $S_B(\tilde{B}, \tilde{C}) \geq S_B(\tilde{A}, \tilde{C})$
5. Si A y B son conjuntos difusos tipo 1 entonces $S_B(A, B) \in [0, 1]$ es un número.

Por tanto la similaridad viene dada por:

$$S_B(\tilde{A}, \tilde{B}) = [S_L(\tilde{A}, \tilde{B}), S_U(\tilde{B}, \tilde{A})]$$

donde,

$$S_L(\tilde{A}, \tilde{B}) = r_L(\tilde{A}, \tilde{B}) \star r_L(\tilde{B}, \tilde{A})$$

$$S_U(\tilde{A}, \tilde{B}) = r_U(\tilde{A}, \tilde{B}) \star r_U(\tilde{B}, \tilde{A})$$

\star puede ser cualquier t-norma y $[r_L(\tilde{A}, \tilde{B}), r_U(\tilde{A}, \tilde{B})]$ es un indicador del grado de inclusión de intervalos de \tilde{A} en \tilde{B} . Ambos se calculan de la siguiente forma:

$$r_L(\tilde{A}, \tilde{B}) = \min_{x \in X} \left\{ 1, \min \left(1 - \underline{\mu}_{\tilde{A}}(x) + \underline{\mu}_{\tilde{B}}(x), 1 - \bar{\mu}_{\tilde{A}}(x) + \bar{\mu}_{\tilde{B}}(x) \right) \right\}$$

$$r_U(\tilde{A}, \tilde{B}) = \min_{x \in X} \left\{ 1, \max \left(1 - \underline{\mu}_{\tilde{A}}(x) + \underline{\mu}_{\tilde{B}}(x), 1 - \bar{\mu}_{\tilde{A}}(x) + \bar{\mu}_{\tilde{B}}(x) \right) \right\}$$

2.3.4. Similaridad de Zeng y Li

La medida de similaridad dada por Zeng y Li es para conjuntos difusos intervalo-valorados. Se trata de una función real $S_Z: IVFS(\tilde{A}) \times IVFS(\tilde{B}) \rightarrow [0,1]$ como una medida de Similaridad para conjuntos difusos intervalo-valorados, si S_Z cumple las siguientes propiedades:

1. $S_Z(\tilde{A}, \tilde{B}) = S_Z(\tilde{B}, \tilde{A})$
2. $S_Z(\tilde{A}, c(\tilde{A})) = 0$
3. $S_Z(\tilde{A}, \tilde{B}) = 1 \Leftrightarrow \tilde{A} = \tilde{B}$
4. Si $A \leq B \leq C$ entonces $S_Z(\tilde{A}, \tilde{C}) \leq S_Z(\tilde{A}, \tilde{B})$ y $S_Z(\tilde{A}, \tilde{C}) \leq S_Z(\tilde{B}, \tilde{C})$

Así proponen las siguientes medidas de Similaridad para conjuntos difusos intervalo-valorados si los universos del discurso de \tilde{A} y \tilde{B} son discretos

$$S_Z(\tilde{A}, \tilde{B}) = 1 - \frac{1}{2n} \sum_{i=1}^n (|\underline{\mu}_{\tilde{A}}(x_i) - \underline{\mu}_{\tilde{B}}(x_i)| + |\bar{\mu}_{\tilde{A}}(x_i) - \bar{\mu}_{\tilde{B}}(x_i)|)$$

2.3.5. Similaridad de Mendel

Según J.M.Mendel, cuando se mide la similaridad entre dos conjuntos IVFS \tilde{A} y \tilde{B} es necesario comparar sus formas y también su proximidad. Por ello, J.M.Mendel propone medir la Similaridad mediante un vector de similaridad, el cuál tendrá dos componentes, por ejemplo:

$$S_v(\tilde{A}, \tilde{B}) = (S_1(\tilde{A}, \tilde{B}), S_2(\tilde{A}, \tilde{B}))^T$$

Donde $S_1(\tilde{A}, \tilde{B}) \in [0,1]$ es la medida de Similaridad para la forma y $S_2(\tilde{A}, \tilde{B}) \in [0,1]$ es la medida de proximidad entre \tilde{A} y \tilde{B} .

Debido a que la proximidad entre \tilde{A} y \tilde{B} se calcula en $S_2(\tilde{A}, \tilde{B})$, en el cálculo de S_1 consideraremos que ambos conjuntos están alineados, así sus formas pueden ser comparadas. Los centroides de \tilde{A} y \tilde{B} los denotaremos como $C_{\tilde{A}} = [c_l(\tilde{A}), c_r(\tilde{A})]$ y $C_{\tilde{B}} = [c_l(\tilde{B}), c_r(\tilde{B})]$, respectivamente y los centros de $C_{\tilde{A}}$ y $C_{\tilde{B}}$ como:

$$C(\tilde{A}) = [c_l(\tilde{A}) + c_r(\tilde{A})]/2$$

$$C(\tilde{B}) = [c_l(\tilde{B}) + c_r(\tilde{B})]/2$$

Un método de alineamiento es mover un conjunto o ambos de forma que $c(\tilde{A})$ y $c(\tilde{B})$ coincidan. Esto no afectará al valor de S_1 , que puede aparecer como un valor o como un intervalo.

Un cálculo bastante intuitivo para S_1 es definirlo como una extensión de la Similaridad de Jaccard de la Similaridad para conjuntos difusos para conjuntos difusos intervalo-valorados utilizando la siguiente expresión:

.

$$S_1(\tilde{A}, \tilde{B}) = \frac{AC[FOU(\tilde{A} \cap \tilde{B}')] }{AC[FOU(\tilde{A} \cup \tilde{B}')]}$$

$$= \frac{\text{card}(\underline{\mu}_{\tilde{A}}(x) \cap \underline{\mu}_{\tilde{B}'}(x)) + \text{card}(\underline{\mu}_{\tilde{A}}(x) \cap \underline{\mu}_{\tilde{B}'}(x))}{\text{card}(\underline{\mu}_{\tilde{A}}(x) \cup \underline{\mu}_{\tilde{B}'}(x)) + \text{card}(\underline{\mu}_{\tilde{A}}(x) \cup \underline{\mu}_{\tilde{B}'}(x))}$$

$S_1(\tilde{A}, \tilde{B})$ tendrá las siguientes propiedades:

- a) $0 \leq S_1(\tilde{A}, \tilde{B}) \leq 1$
- b) $S_1(\tilde{A}, \tilde{B}) = 1 \Leftrightarrow \tilde{A} = \tilde{B}$

$S_2(\tilde{A}, \tilde{B})$ mide la proximidad de \tilde{A} y \tilde{B} , se define como:

$$S_2(\tilde{A}, \tilde{B}) \equiv h(d(\tilde{A}, \tilde{B}))$$

donde $d(\tilde{A}, \tilde{B}) = |c(\tilde{A}) - c(\tilde{B})|$ es la distancia Euclídea entre centros de centroides de \tilde{A} y \tilde{B} , y h es una función que satisface las siguientes propiedades:

1. $\lim_{x \rightarrow \infty} h(x) = 0$
2. $h(x)=1$ si y sólo si $x=0$
3. $h(x)$ es monótonamente decreciente con el crecimiento de x

La medida $S_v(\tilde{A}, \tilde{B})$ permite cuantificar separadamente las características de forma y proximidad. El propósito de esta medida es encontrar el conjunto que más se parece a otro, por lo tanto nos tiene que permitir crear un ranking. Por esta razón se necesita crear un escalar con el que comparar.

Dicho valor se obtendrá calculando el producto de las similaridades de forma y distancia. Es decir:

$$S_v(\tilde{A}, \tilde{B}) = S_1(\tilde{A}, \tilde{B}) \times S_2(\tilde{A}, \tilde{B})$$

3. SISTEMAS DE CLASIFICACIÓN

La minería de datos es el proceso de extracción de conocimiento de grandes bases de datos que nos permiten extraer modelos que nos ayudarán a interpretar correctamente la información que manejamos. Uno de los tópicos en los que más se está investigando dentro de la MD son los sistemas de clasificación.

El problema de clasificación, desde el punto de vista del aprendizaje supervisado, consiste en el establecimiento de una regla de decisión que permita determinar la clase de un nuevo objeto en una de las clases existentes y conocidas, $C_j \in C = \{1, \dots, M\}$, siendo M el número de clases. Cada uno de los objetos $e \in E$, a los que también se denomina ejemplos, se describe mediante un conjunto de observaciones $X(e) = (e_1, \dots, e_N)$ y a cada una de estas observaciones se le denomina variable, atributo o característica. El diseño de un *Clasificador o Sistema de Clasificación* se puede ver como la búsqueda de una correspondencia

$$D: X(E) \rightarrow C$$

optimal en el sentido de un cierto criterio $\delta(D)$ que nos determine la bondad del Clasificador. Normalmente el objeto final, es el diseño de Sistema de Clasificación que asigne clases a cualquier punto del espacio de atributos con el mínimo error posible.

Algunos de los sistemas de inteligencia artificial que permiten la clasificación son: los árboles de decisión, las redes neuronales, las máquinas de vectores de soporte o los sistemas de clasificación basados en reglas difusas. En los siguientes apartados describiremos brevemente algunos de estos sistemas.

3.1. Árboles de decisión

De todos los métodos de aprendizaje, los sistemas de aprendizaje basados en árboles de decisión son quizás el método más fácil de utilizar y de entender. Un árbol de decisión es un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final a tomar se puede determinar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas. Los árboles de decisión se

3. Sistemas de Clasificación

utilizan desde hace siglos y son especialmente apropiados para expresar procedimientos médicos, legales, comerciales, estratégicos, matemáticos lógicos, etc.

Una de las grandes ventajas de los árboles de decisión es que, en su forma más general, las opciones posibles a partir de una determinada condición son excluyentes. Esto permite analizar una situación y, siguiendo el árbol de decisión apropiadamente, llegar a una sola acción o decisión

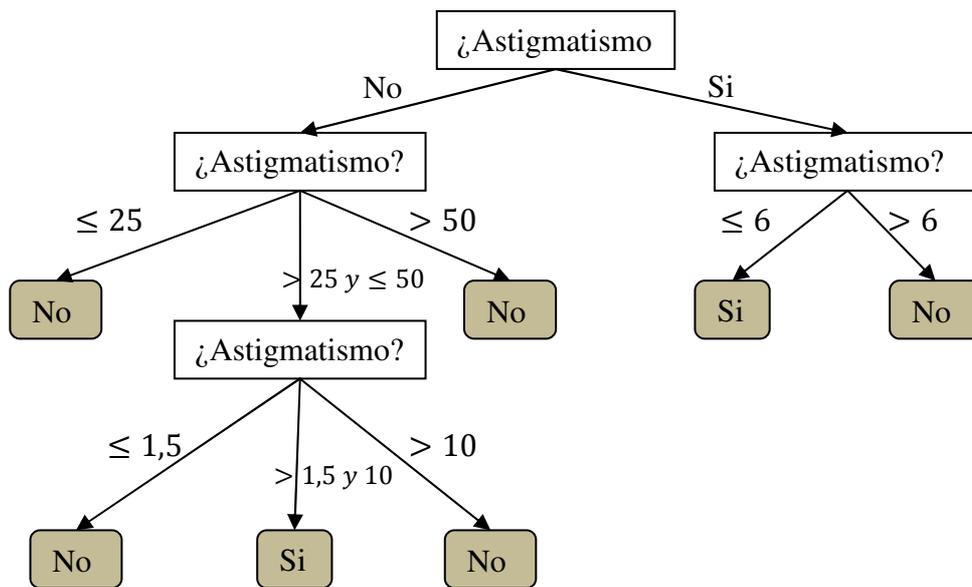


Ilustración 7. Árbol de decisión para determinar recomendación o no de cirugía ocular

3.2. Redes neuronales artificiales

Las redes neuronales artificiales (RNA) son sistemas conexionistas dentro del campo de la Inteligencia Artificial, las cuales, dependiendo del tipo de arquitectura neuronal, pueden tener diferentes aplicaciones. Pueden utilizarse para el reconocimiento de patrones, la comprensión de información y la reducción de la dimensionalidad, el agrupamiento, la clasificación, la visualización, etc.

Las redes neuronales artificiales son un método de aprendizaje cuya finalidad inicial es la de emular los procesadores biológicos de información. Las RNA parten de la presunción que la capacidad humana de procesar información se debe a la naturaleza biológica de nuestro cerebro.

3.2.1 Neuronas Biológicas y artificiales

Una neurona biológica recibe la información a través de la sinapsis de sus dendritas. Cada sinapsis representa la unión de un axón de una neurona con una dendrita de otra neurona. Una transmisión electro-química tiene lugar en la sinapsis, la cual permite a la información ser transmitida desde una neurona a la próxima. La información es entonces transmitida a lo largo de las dendritas hasta que alcanza el cuerpo de la célula. Allí tiene lugar el sumatorio de los impulsos eléctricos que lo alcanzan y se aplica según algún tipo de función de activación a éste. La neurona se activará si el resultado es superior a un determinado límite o umbral. Esto significa que enviará una señal (en forma de una onda ionización) a lo largo de su axón con la finalidad de comunicarse con otras neuronas. Ésta es la manera en la que la información pasa de una parte de la red de neuronas a otra. Las sinapsis tienen diferente rendimiento al transcurrir el tiempo de vida de la neurona.

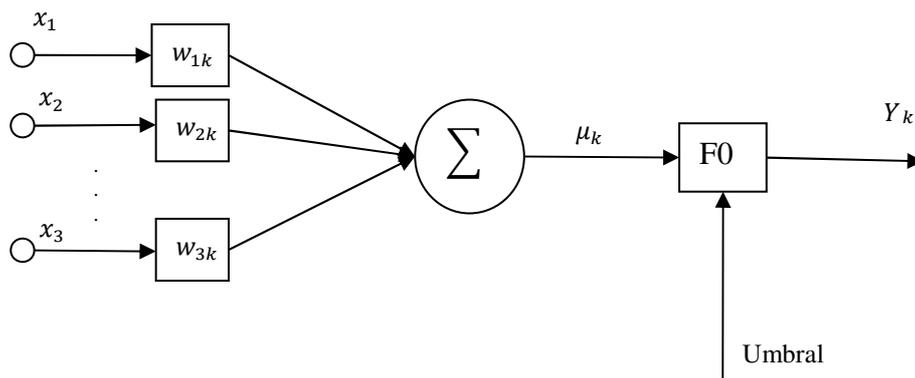


Ilustración 8. Representación de una neurona artificial

Normalmente se modela una neurona biológica como aparece en la Ilustración 8. Esta figura incluye una entrada externa adicional, denominada polarización o “bias”

3. Sistemas de Clasificación

cuya finalidad es la de poder aumentar o disminuir el umbral de excitación de la neurona dependiendo de si es un valor positivo o negativo, respectivamente.

Las entradas se representan por el vector de entrada, x , y el rendimiento de las sinapsis se modela mediante un vector de pesos, w . entonces el valor de salida de esta neurona viene dado por:

$$y = f\left(\sum_i w_i x_i\right) = f(w, x) = f(w^T x)$$

donde f es la función de activación.

Cuando tenemos una red de neuronas, las salidas de unas se conectan con las entradas de otras. Si el peso entre dos neuronas conectadas es positivo, el efecto producido es de excitación. Por el contrario, si es negativo, este efecto es de inhibición.

Por tanto, podemos ver que una única neurona es una unidad de procesamiento muy simple. Se considera que el potencial de las redes neuronales artificiales proviene de la capacidad que proporciona el empleo de muchas de estas unidades simples y robustas al actuar en paralelo. Se puede considerar que estamos intentando simular lo que ocurre en el cerebro humano, aunque es imposible ya que los seres humanos poseen varios cientos de billones de neuronas. En la figura que aparece a continuación se puede observar la representación de una red neuronal artificial.

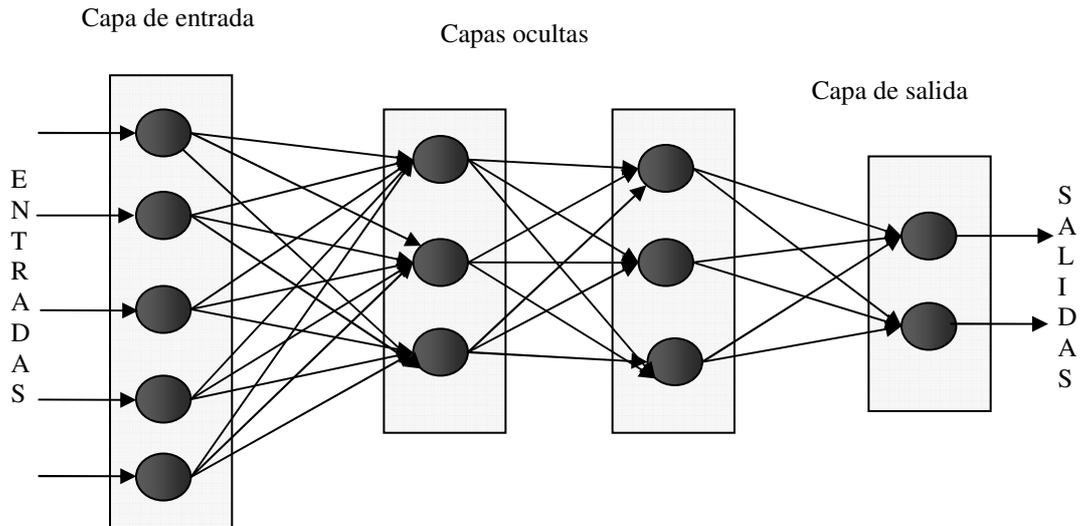


Ilustración 9. Ejemplo red neuronal

En la Ilustración 9 se observa un conjunto de entradas (el vector de entrada, x) accediendo a la red desde el lado izquierdo y se propaga a través de la red hasta que la activación alcanza la capa de salida. Las capas intermedias son conocidas como capas ocultas ya no son accesibles desde fuera de la red.

Existen dos modos de trabajo en un RNA:

- modo de transferencia de activación: cuando la activación es transmitida por toda la red. Éste es el funcionamiento o de aplicación y está asociado a la operación de propagación hacia adelante.
- modo de aprendizaje: cuando la red se organiza normalmente a partir de la transferencia de activación más reciente.

A la hora de aplicar las redes neuronales a un problema de clasificación, éstas se usan en modo aprendizaje.

3.2.2. El aprendizaje de las redes neuronales artificiales:

Las redes neuronales no necesitan volver a ser programadas al cambiar de entorno. Esto no quiere decir que sus comportamientos no cambien con la finalidad de adaptarse al nuevo entorno. Los cambios en los resultados son debido al cambio de los pesos de las conexiones entre neuronas de una capa y otra capa dando lugar a lo que denominamos aprendizaje. Estos cambios se producen para modelar los cambios en el rendimiento de la sinapsis de las redes neuronales reales.

Para problemas de clasificación podrían utilizarse redes neuronales artificiales de una o varias capas, esto dependerá de la complejidad del problema, para realizar una clasificación.

Por último decir, que existen diferentes tipos de redes neuronales artificiales dependiendo de la arquitectura que se use para construirlas. Entre ellas están las multicapa, las recurrentes, de base radial, etc.

3.3. SVM

Los fundamentos teóricos de las máquinas de vectores soporte (SVM, del inglés Support Vector Machines) se encuentran en los trabajos de Vapnik y otros autores sobre la teoría del aprendizaje estadístico, desarrollados a finales de los años 70 y durante los 80. El modelo SVM, tal como se entiende actualmente, fue presentado en la conferencia COLT (COmputational Learning Theory) del año 1992 por el mismo Vapnik, junto con Boser y Guyon [BOS92] y descrito con más detalle posteriormente en [CVAP95; Vapnik 1998], posibilitando el salto de la formulación teórica a su aplicación práctica en problemas reales de reconocimiento de formas. A partir de ese momento el interés por este modelo de aprendizaje no ha parado de crecer hasta el presente. Como consecuencia, tanto su desarrollo teórico (introduciendo conexiones con otros modelos de aprendizaje) como el campo de las aplicaciones reales han experimentado un avance muy importante, hasta el punto que, hoy en día, las SVM constituyen un referente completamente establecido para las disciplinas del aprendizaje automático y de la minería de datos. Los campos donde las SVM han sido aplicadas con éxito incluyen

entre otros, la visión por computador, la bioinformática, la recuperación de información, el procesamiento del lenguaje natural y el análisis de series temporales.

Las máquinas de vectores soporte mapean el espacio la entrada original en características del espacio de varias dimensiones por medio de ciertas funciones núcleo (evitando el cálculo del producto interno de dos vectores). En una nueva característica del espacio, el hiperplano óptimo de separación con máximo margen es determinado para minimizar un máximo riesgo esperado en vez del riesgo empírico. Se utiliza el algoritmo de entrenamiento SMO para obtener el SVM clasificador básico.

3.4. Sistemas de Clasificación Basados en Reglas Difusas.

Los Sistemas de Clasificación Basados en Reglas Difusas (SCBRD) son una herramienta muy útil para tratar problemas de clasificación. Éstos son muy empleados por su capacidad de construir un modelo lingüístico interpretable por los usuarios y que permita incorporar la información proveniente de: la experiencia de un experto, y la información proveniente de modelos matemáticos o medidas empíricas. La importancia de los SCBRD se puede observar en la cantidad de aplicaciones reales que tienen. Algunas de estas son: la detección de intrusiones, aplicaciones médicas, predicción de ventas de periódicos, entre otras.

Un SCBRD está compuesto por una Base de Conocimiento (BC) y un Método de Razonamiento Difuso (MRD) que, utilizando la información de BC, determina una clase para cualquier patrón de datos admisible que llegue al sistema. Esta es la estructura que refleja la Ilustración 10:

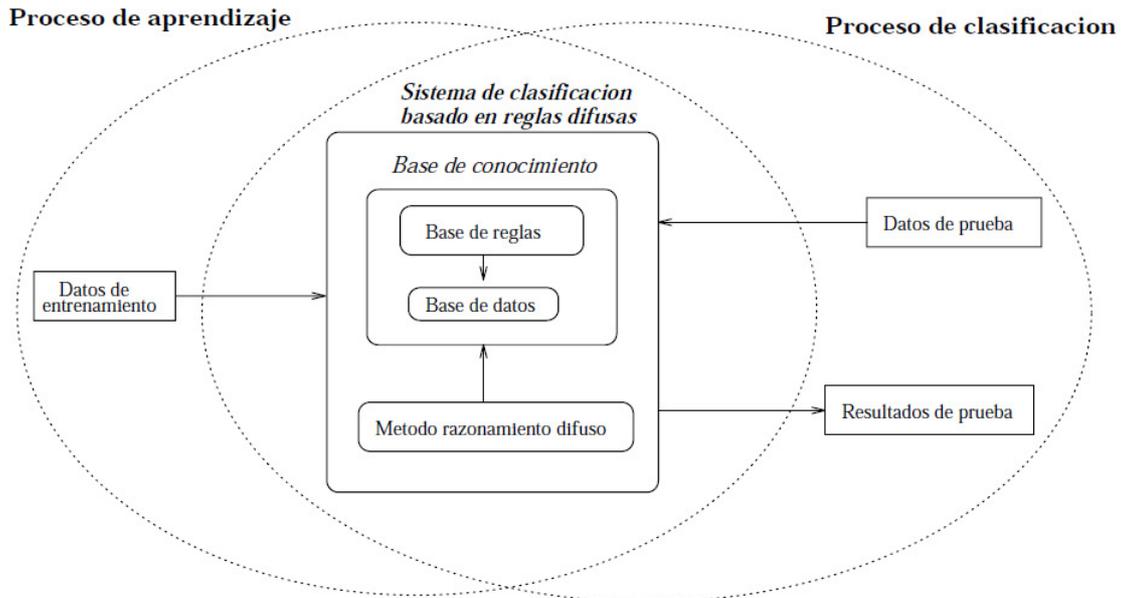


Ilustración 10. Esquema de un SCBRD

La BC está formada por dos componentes:

- La Base de Datos (BD), que contiene la definición de los conjuntos difusos asociados a los términos lingüísticos utilizados en la Base de Reglas.
- La Base de Reglas (BR), formada por un conjunto de reglas de clasificación $R = \{R_1, R_2, \dots, R_L\}$, donde L es el número de reglas, de uno de los tipos siguientes utilizados en la literatura especializada para SCBRDs.

La Base de Reglas difusas almacena el conjunto de reglas del sistema, estas reglas pueden ser de los siguientes tipos:

a) **R_K : Si X_1 es A_1^K y ... y X_N es A_N^K entonces Y es C_j**

donde X_1, \dots, X_N son las variables asociadas a los diferentes atributos del sistema de clasificación A_1^K, \dots, A_N^K son las etiquetas lingüísticas utilizadas para discretizar los dominios continuos de las variables, e Y es la variable que indica la clase C_j a la cual pertenece el patrón.

b) **Reglas difusas con una clase y un grado de certeza asociado a la clasificación para esa clase en el consecuente**

R_K : Si X_1 es A_1^K y ... y X_N es A_N^K entonces Y es C_j con grado r^k donde r^k es el grado de certeza asociado a la clasificación en la clase C_j para los ejemplos pertenecientes al subespacio difuso delimitado por el antecedente de la regla. Este tipo de regla son empleadas por Ishibuchi y otros métodos.

c) Reglas difusas con grados de certeza asociados a cada una de las clases en el consecuente

$$R_K: \text{Si } X_1 \text{ es } A_1^K \text{ y ... y } X_N \text{ es } A_N^K \text{ entonces } (r_1^k, \dots, r_M^k)$$

donde r_j^k es el grado de fuerza de la regla R_k para predecir la clase C_j para un ejemplo perteneciente a la región difusa representada por el antecedente de la regla.

Este tipo de reglas se utilizan en los SCBRDs propuestos por Mandal y otros.

El último modelo de regla extiende a los dos anteriores empleando distintos valores para los parámetros del vector (r_1^k, \dots, r_M^k) . Así, con:

$$r_v^k = 1, r_j^k = 0 \quad j \neq v, j = 1, \dots, M$$

tenemos una regla difusa tipo (a), y con

$$r_v^k = r^k, r_j^k = 0 \quad j \neq v, j = 1, \dots, M$$

una regla tipo (b).

En algunas propuestas de SCBRDs se utilizan reglas con consecuente difuso, es decir, reglas difusas en las que el consecuente es una etiqueta lingüística definida a través de un conjunto difuso. Este tipo de reglas, en un Sistema de Clasificación, requieren un proceso de defuzzificación adicional que permita la obtención de un valor nítido de clasificación y no serán consideradas en esta memoria.

La necesidad de interpretar los resultados, exigida a un Sistema de Clasificación en la mayoría de los casos, hace que el enfoque adoptado para su diseño sea con frecuencia el de un SCBRD descriptivo como el explicado en esta sección y utilizado en esta memoria. En él, las variables consideradas para describir las observaciones del sistema se consideran variables lingüísticas y se asocia un conjunto de etiquetas a cada una de ellas, almacenándose la semántica de estas últimas en la BD. En un SCBRD aproximativo las variables se consideran variables difusas y, en cada regla, cada una de

3. Sistemas de Clasificación

las variables tomará como valor un conjunto difuso distinto por lo que no es necesaria una BD en la BC, del sistema, que queda reducida de esta forma a una BR formada por un conjunto de reglas difusas aproximativas.

El diseño de un SCBRD mediante un proceso de aprendizaje inductivo supervisado comienza con un conjunto de ejemplos correctamente clasificados y tiene como objetivo final la obtención de un SCBRD que determine, con el mínimo error posible, la etiqueta clase para ejemplos no clasificados. Una vez diseñado el SCBRD, se determina su rendimiento sobre datos de prueba para tener una estimación del error real del SCBRD. Este proceso se describe en la figura 13 del comienzo de este punto.

El análisis de los componentes y del funcionamiento de un SCBRD nos permite determinar que el proceso de aprendizaje puede implicar cinco tareas complementarias entre si:

- La selección de las variables más informativas para el problema de clasificación a resolver.
- La determinación del MRD a utilizar en el proceso de clasificación de nuevos ejemplos.
- La generación de un conjunto de reglas.
- La selección de un subconjunto de reglas con la mejor cooperación y la menor redundancia.
- El establecimiento y ajuste de las particiones difusas para los dominios de las variables del problema.

La definición automática de cualquiera de estas tareas se puede tratar como un problema de optimización o búsqueda.

3.4.1 Determinación del Método de Razonamiento Difuso:

El MRD en un SCBRD determina la forma de utilización de la información contenida en un conjunto de reglas que describen la relación entre las variables del

sistema en el proceso de predicción de la clase a la que pertenece un nuevo patrón. Es, por tanto, uno de los aspectos más importantes en un SCBRD y más determinantes de su rendimiento e interpretabilidad.

La formulación de un MRD debe permitir aprovechar al máximo la potencia del Razonamiento Aproximado propia de los Sistemas Difusos, que permite obtener un resultado incluso cuando no se tenga compatibilidad exacta entre el patrón y los antecedentes de las reglas.

Para Sistemas de Control Difuso en los que el consecuente de las reglas es un conjunto difuso, se han formulado distintas versiones de la regla composicional de inferencia de Zadeh [Zad73] que permite obtener la salida de una regla difusa [MZ82, DP92] y distintos operadores de defuzzificación que combinan las salidas de todas las reglas disparadas para obtener una respuesta global del sistema.

En los SCBRDs en los que el consecuente no es un conjunto difuso, el MRD más utilizado es el denominado de compatibilidad máxima o de regla ganadora. Kuncheva define en [Kun96] de forma general el funcionamiento de este MRD: Para cada regla difusa y para cada clase, se combina el grado de compatibilidad del ejemplo con su antecedente con el grado de certeza de la clasificación en esa clase. La salida del SCBRD es, en base al principio de la regla ganadora, la clase para la que se alcanza valor máximo de combinación entre compatibilidad y certeza para una de las reglas. En caso de empate, se puede seleccionar aleatoriamente una de las clases para las cuales se obtiene ese valor máximo o determinar cómo salida la clase para la cual se han disparado más reglas (considerando que una regla se dispara en la clasificación de un ejemplo, si éste tiene grado de compatibilidad con antecedente de la regla mayor que cero).

A diferencia de los empleados habitualmente en control difuso, este MRD utiliza la información que proporciona una sola regla. Se han desarrollado otras propuestas que tratan de combinar la información aportada por las reglas disparadas:

- El MRD denominado de combinación aditiva o de compatibilidad máxima acumulativa, basado en la votación máxima, que acumula para cada clase la

3. Sistemas de Clasificación

combinación del grado de compatibilidad con el grado de certeza aportado por cada regla, y clasifica con la clase para la que se obtiene el valor máximo.

- El MRD basado en la defuzzificación del centroide que para una BR con L reglas difusas tipo “a”, determina la salida S para un ejemplo e de acuerdo a la siguiente expresión:

$$S = \frac{\sum_{k=1}^L R^k(e) \cdot C_k}{\sum_{k=1}^L R^k(e)},$$

donde C_k es la clase indicada en el consecuente de la regla k y $R^k(e)$ es el grado de compatibilidad del ejemplo con el antecedente de la regla k (o lo que es lo mismo, el grado de pertenencia del ejemplo a la región difusa determinada por el antecedente de la regla).

Una vez obtenida esta salida S se clasifica con la clase

$$C_j = (\text{int})(S + 0.5),$$

donde (int) denota el valor entero más pequeño.

Este tipo de MRD sólo trabaja bien con problemas de clasificación en dos clases ya necesita un gran número de reglas difusas para dar resultados satisfactorios en cualquier otro caso, por lo que no es muy utilizado.

Como se ha mencionado, el MRD incluido en un SCBRD determina su rendimiento e interpretabilidad, por lo que nuevas propuestas de MRDs que integren la información de todas las reglas o seleccionen parte de las mismas en un proceso de clasificación pueden llevar a la obtención de mejores SCBRDs.

3.4.4 Métodos para crear la Base de Conocimiento:

Un proceso de generación de la BC de un SCBRD descriptivo debe obtener un conjunto de reglas difusas que representen, de la forma más precisa posible, las

relaciones entre las variables del problema de clasificación y permitan la clasificación correcta de nuevos ejemplos que se presenten al sistema.

En este problema uno de los primeros aspectos a determinar es la partición difusa utilizada para cada una de las variables. La obtención del conjunto de etiquetas lingüísticas (y conjuntos difusos asociados) óptimo para cada variable es un problema difícil que se ha afrontado de diferentes formas en distintos métodos:

- Estableciendo una partición difusa fija para cada variable en base a conocimiento experto o a una división equidistante del dominio con un tipo de función de pertenencia preestablecida o determinada tras una experimentación. Este enfoque se sigue en la extensión a problemas de clasificación del proceso de generación de Wang y Mendel [CWY95,CYP96], en el proceso de extracción propuesto por Pal y Mandal en [PM92] o en el método genético desarrollado por Yuan y Zhuang[YZ96].
- Utilizando técnicas de agrupamiento que determinan grupos de datos con determinadas relaciones entre las características y construyendo los conjuntos difusos mediante proyecciones. Habitualmente esto determina conjuntos difusos para cada una de las reglas, por lo que se obtienen SCBRDs aproximativos.
- Integrando el proceso de búsqueda de la partición difusa más adecuada en el proceso de aprendizaje, como hacen Hong y Li en el método descrito en [HL96].
- Utilizando distintas particiones difusas para cada variable, que se diferencia no sólo en el número de términos utilizados sino también en la forma de la función de pertenencia y, posteriormente, seleccionando las reglas con mejor comportamiento en el sistema para obtener así reglas difusas con distinta granularidad en la partición, como proponen Ishibuchi y otros.

Los métodos de extracción de BCs difusas se diferencian no sólo en la forma de obtención de la partición difusa, sino también en el enfoque utilizado para la generación de reglas. Así tenemos esto métodos:

- Métodos iterativos que ó exploran el conjunto de ejemplos de entrenamiento, como el propuesto por Chi, ó exploran el espacio de patrones y determinan la

3. Sistemas de Clasificación

mejor regla para cada subregión del espacio determinada por la partición difusa, como lo hace Ishibuchi.

- Métodos basados en Redes Neuronales que permiten obtener BCs descriptivas o aproximadas.
- Métodos basados en técnicas de agrupamiento.
- Métodos basados en la derivación de reglas difusas a partir de ID3.
- Métodos basados en Algoritmos Genéticos.

A continuación pasaremos a analizar más en profundidad el método de Chi para extraer la base de reglas. Existen otros como el de Ishibuchi o Hong y Li, pero como utilizaremos el de Chi para la realización del proyecto, solamente profundizaremos en el análisis de éste.

3.4.4.1 Algoritmo de Chi:

Es una extensión, de un método de generación de reglas difusas para problemas de control definido por Wang y Mendel, hecha por Chi para problemas de clasificación. A continuación se describe la forma en que se utiliza.

Consideremos un conjunto de datos de entrenamiento con la siguiente estructura:

$$e^1 = (e_1^1, \dots, e_N^1, s^1),$$

$$e^2 = (e_1^2, \dots, e_N^2, s^2),$$

.

.

.

$$e^p = (e_1^p, \dots, e_N^p, s^p)$$

,siendo s^h la etiqueta de la clase a la que pertenece el ejemplo e^h y e_N^p la variable lingüística N del ejemplo, p es el número de datos de entrenamiento y N el número de características que se tienen en cuenta de cada ejemplo.

Para generar un conjunto de reglas a partir de este conjunto de entrenamiento que describan las relaciones entre las variables y determinen una correspondencia entre el espacio de características y el conjunto de clases $C = \{C_1, \dots, C_M\}$, donde M es el número de clases, se siguen los siguientes pasos:

1. Fuzzificar el espacio de características. Una vez determinado el dominio de variación de cada atributo o característica x_i , se divide en Z_i regiones ($i=1, \dots, N$) y se define un conjunto difuso para cada región difusa. Cada conjunto difuso estará modelado por funciones triangulares en este proyecto.
2. Generar una regla difusa para cada ejemplo $e^h = (e_1^h, \dots, e_N^h, s^h)$. Para ello hay que hacer:
 - a. Calcular los grados de pertenencia del ejemplo e^h a las distintas regiones difusas. Por ejemplo, si estamos hablando de características de una persona, que mide 1,75 y pesa 80 kg. Tendremos que determinar cuál es el grado de pertenencia a las diferentes particiones de alto (bajo, medio, alto) y a las particiones de peso (delgado, medio, obeso).
 - b. Asignar el ejemplo e^h a la región difusa a la que tiene mayor grado de pertenencia. Basándonos en el ejemplo anterior le asignaríamos la etiqueta lingüística que mayor grado de pertenencia tiene, por ejemplo alto, obeso.
3. Generar una regla difusa para el ejemplo, cuyo antecedente está determinado por la región difusa seleccionada y con la etiqueta de clase del ejemplo en el consecuente. Ejemplo: si x_1 es alto y x_2 es obeso entonces es la clase hombre.
4. Cuando tenemos el mismo antecedente y distinta clase, para decidir que clase es la que elegimos, calculamos el grado de certeza, esto lo haremos según el método de Ishibuchi.

3.4.5.1 Método de Ishibuchi para el cálculo del grado de certeza:

En nuestras reglas, hemos definido el consecuente de la regla y el grado de certeza de cada regla, calculado a partir de una serie de patrones de entrenamiento. Ishibuchi define un método para asignar el grado de confianza a cada regla.

Si consideramos m patrones etiquetados $x_p = (x_{p1}, \dots, x_{pn}), p = 1, 2, \dots, m$ donde x_{pi} es el i -ésimo atributo que puede tomar valores $i=1, 2, \dots, n$. Cada regla difusa será de la siguiente forma:

$$R_j: \text{Si } x_1 \text{ es } A_{j1} \text{ y } \dots \text{ y } x_n \text{ es } A_{jn} \text{ entonces Clase} = C_j \text{ con } RW_j$$

donde R_j es la etiqueta de la j -ésima regla, $x = (x_1, \dots, x_n)$ es un vector del patrón de dimensión n , A_{ji} es el conjunto difuso del antecedente representando un término lingüístico, C_j es la etiqueta de la clase, y RW_j es el peso de la regla. Específicamente, el peso de la regla se computa usando un factor de penalización de sus siglas en inglés (PCF) definido como:

$$PCF_j = \frac{\sum_{x_p \in \text{Class } C_j} \mu_{A_j}(x_p) - \sum_{x_p \notin \text{Class } C_j} \mu_{A_j}(x_p)}{\sum_{p=1}^m \mu_{A_j}(x_p)}$$

4. SCBRD CON IVFS Y MEDIDAS DE SIMILARIDAD

En los capítulos anteriores hemos descrito la teoría sobre lógica difusa, distintos sistemas de clasificación utilizados en la actualidad y en este capítulo nos vamos a centrar en describir y profundizar más en el modelo planteado para las pruebas a realizar en el proyecto.

Hemos hablado de las ventajas de estos sistemas, que son la interpretabilidad del modelo de salida. Sin embargo, una desventaja de estos sistemas es la incertidumbre inherente a la definición de las funciones de pertenencia que pueden estar definidas homogéneamente en el espacio de entrada o por conocimiento experto. Este hecho sugiere representar el grado de pertenencia de los objetos a los conjuntos por la medio de intervalos. Esto es, emplear Conjuntos Difusos Intervalo-Valorados (CDIV) para caracterizar las etiquetas lingüísticas que componen los atributos del problema.

El objetivo de este proyecto es estudiar la eficacia en la clasificación de un sistema de clasificación basado en reglas difusas. Se harán pruebas sobre los mismos datos para cada una de las medidas de similaridad y se compararán los resultados. Para poder aplicar las medidas de similaridad, representaremos las entradas del sistema como conjuntos difusos tipo 2 o IVFS. En el sistema, el grado de emparejamiento se calcula mediante medidas de similaridad entre el ejemplo y el antecedente.

Analizaremos los resultados de estas pruebas utilizando un algoritmo muy conocido para la extracción de reglas difusas de un conjunto de datos, el algoritmo de CHI et al. Se ha elegido el algoritmo de CHI et al porque es uno de los más utilizados.

4.1. Etiquetas Lingüísticas representadas mediante IVFS

A partir de una Base de Conocimiento, que podemos generar mediante cualquier método ad-hoc, incluimos los IVFS para cada etiqueta lingüística. Para construir los IVFS consideramos el conjunto difuso existente en la BC como el extremo inferior,

4. SCBRD con IVFS y Medidas de Similitud

además añadimos el extremo superior centrado en el máximo del extremo inferior y con una amplitud un 50% mayor que la del extremo inferior siendo simétrico a ambos lados.

En la Ilustración 11 podemos ver el ejemplo de una variable lingüística representada por 3 etiquetas (IVFS) en el estado inicial. Las líneas solidas representan el extremo inferior (\underline{A}_j) y las líneas a puntos representan el extremo superior (\overline{A}_j).

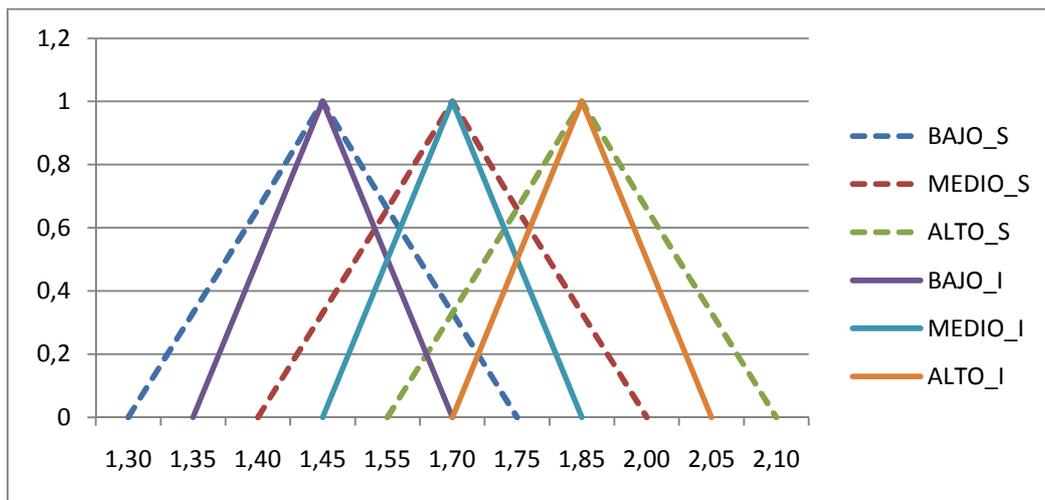


Ilustración 11. Ejemplo de IVFS representando la Altura

4.2. Método de razonamiento difuso

La modificación de la representación de las etiquetas difusas, implica que debemos adaptar el método de razonamiento difuso utilizado para la clasificación de nuevos patrones. El modelo general de razonamiento difuso para clasificación, presentado en la sección 3, será modificado de la siguiente manera. Si tenemos un nuevo patrón $x_p = (x_{p1}, \dots, x_{pn})$ haremos lo siguiente:

1. Para calcular el grado de emparejamiento entre ejemplo y antecedente, utilizamos las medidas de similitud vistas en el capítulo 3. Estas medidas operan con los valores de los extremos superiores e inferiores de los IVFS y nos devuelven un valor de similitud. Algunas de ellas devuelven un intervalo, formado por el valor máximo y el mínimo que toma la similitud. En este caso

para poder trabajar con un único valor, calcularemos el punto medio de ambos valores. El cálculo del grado de emparejamiento se hará de la siguiente forma:

$$T(SM(x_{p1}, A_1), \dots, SM(x_{pn}, A_n))$$

donde la SM la calculamos mediante una de las similaridades vistas en el capítulo 2, y T es una T-norma como el mínimo o el producto.

2. Como **grado de asociación** con cada una de las reglas, se toma la medida del grado de emparejamiento con cada antecedente de las reglas por el peso de la regla. El peso de la regla vendrá denotado como RW y su cálculo se realiza siguiendo el método que aparece en el capítulo 3 en el apartado del método de Ishibuchi para el cálculo del grado de certeza.

En este punto ya podemos aplicar el resto del MRD como en el mostrado en el Capítulo 2.

4.2. Entradas del Sistema

Hasta ahora nos hemos dedicado a explicar la parte del modelo “estática”, con este término nos referimos a la parte del modelo que no se modificará una vez se hayan generado las reglas y el sistema sólo tenga que clasificar, es decir, a las base de reglas, a la base de datos y al MRD. Sabemos como serán las reglas de la base de conocimiento y como se representarán las etiquetas de las variables lingüísticas de las reglas. Ahora nos queda ver como se representarán los datos que entran al sistema para poder calcular sobre ellos la similaridad con IVFS.

Cuando queremos clasificar un nuevo objeto, este tiene una serie de valores para cada una de las variables. Para poder trabajar en nuestro sistema con estos objetos y poder comparar las distintas características con las de las reglas de nuestra base de reglas, deberemos construir un IVFS. Nos vamos a centrar en la representación de cada vector de los ejemplos mediante 3 tipos de conjuntos, que detallaremos a continuación.

4.2.1. Conjuntos Singleton

Este tipo de conjuntos se construyen a partir del valor para cada una de las variables de los patrones de entrada. El conjunto Singleton es un conjunto IVFS especial, que toma valor 1 tanto en la función de pertenencia inferior como en la superior en el punto del universo del discurso que coincide con el valor del patrón para esa variable lingüística y toma valor cero en el resto del universo del discurso. En la Ilustración 12, viene detallado como se construyen este tipo de conjuntos.

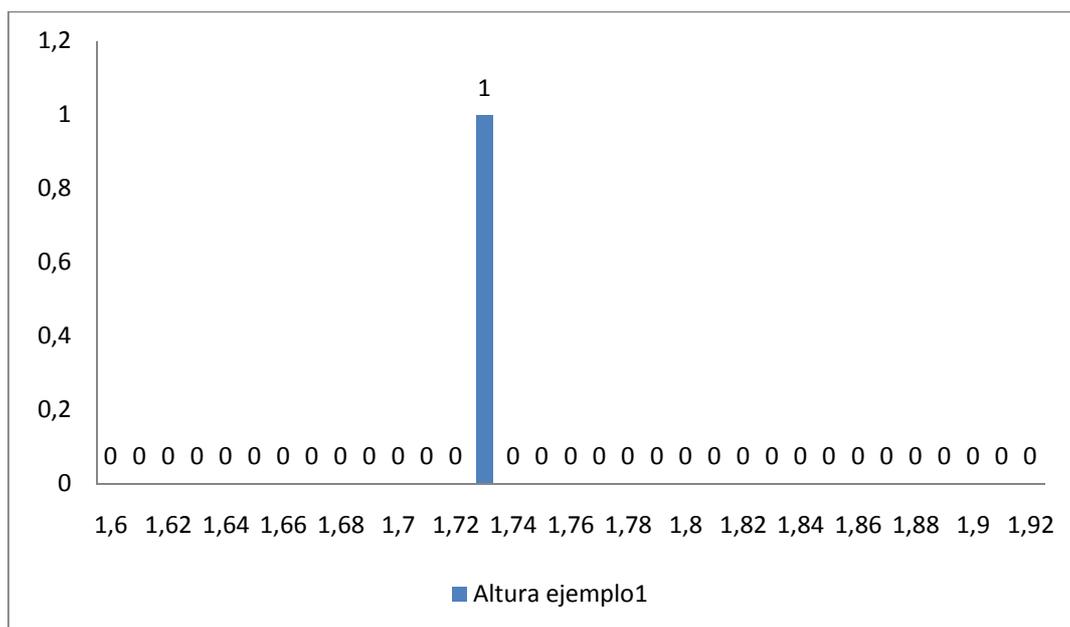


Ilustración 12. IVFS de tipo Singleton

El ejemplo de la Ilustración anterior se muestra un ejemplo en el que se representa mediante un IVFS el valor de la variable lingüística altura para un ejemplo de entrada. En él se ve como el ejemplo de entrada tiene una altura de 1,73. En dicho punto el conjunto toma valor 1, para cualquier otro valor toma valor 0.

4.2.2. Conjuntos Difusos

Los conjuntos difusos se construyen como si fuese un conjunto IVFS con la misma función de pertenencia para el extremo superior e inferior. Las funciones de pertenencia tomarán valor 1 en el punto del universo del discurso que coincide con el valor del

patrón para esa variable lingüística. En el resto del universo del discurso tomará distintos valores dependiendo de qué función de pertenencia se elija. En nuestro caso de las funciones que aparecía en el tema 2 se eligió una función de pertenencia de tipo Gaussiana, cuya media será la del valor del ejemplo y desviación estándar 2. En la Ilustración 13, se puede ver cómo será ésta.

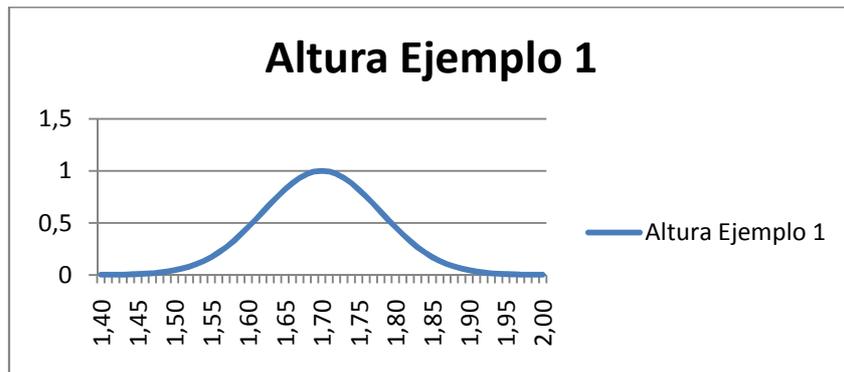


Ilustración 13. Ejemplo Conjunto difuso

En la Ilustración 13 se puede ver como la altura del ejemplo 1, la cual tiene un valor de 1,70, toma valor 1 en ese punto y luego sigue una distribución de tipo Gaussiano en el resto del universo del discurso. De esta forma representamos una cierta incertidumbre en la altura que encajaría este ejemplo como clase.

4.2.2. Conjuntos IVFS

Como ya hemos comentado anteriormente, este tipo de conjuntos vienen definidos por una función de pertenencia inferior y otra superior. Para nuestras pruebas, hemos seleccionado para ambas funciones de pertenencia funciones de tipo gaussiano. Lo único que diferenciará a una de la otra es la desviación estándar que en el caso de la función superior será mayor. Ambas funciones estarán normalizadas a 1, para que cumpla la condición de que el valor medio (valor que toma el ejemplo para esa variable) tome valor 1. En la Ilustración 14 se muestra la representación de estas dos funciones de pertenencia. Donde,

$$FP_{INF} = N(\text{ejemplo}, ds1)$$

$$FP_{SUP} = N(\text{ejemplo}, ds2)$$

$$\text{con } ds2 > ds1$$

4. SCBRD con IVFS y Medidas de Similitud

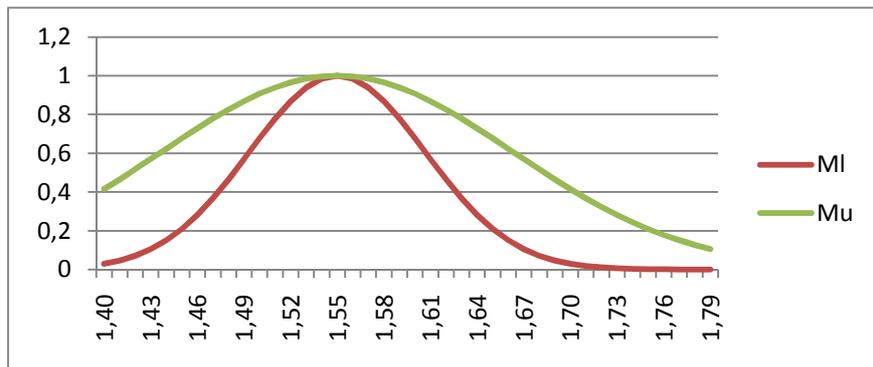


Ilustración 14. Ejemplo conjunto difuso IVFS

5. ESTUDIO EXPERIMENTAL

Como se mencionó en la descripción del modelo, el modelo se construía para realizar una serie de pruebas con el fin de comprobar el rendimiento del sistema de clasificación basado en reglas difusas, aplicando medidas de similaridad durante la clasificación. Además, compararemos los resultados con el SCBRD de Chi y el de Chi con IVFS para el caso singleton.

Antes de la realización de las pruebas procederemos a configurar el entorno donde se realizarán las pruebas, refiriéndonos con esto a los datos sobre los que se realizarán las pruebas, el tipo de conjuntos de entrada que se utilizarán, el tipo de inferencia y la T-norma a utilizar en el cálculo del grado de emparejamiento, etc. En las secciones siguientes, se detallarán todos estos aspectos.

5.1 Configuración del experimento

Para las pruebas se seleccionan 23 data-sets del repositorio UCI. En la Tabla 1 aparece la siguiente información de cada data-set: el número de ejemplos (#Eje), el número de atributos (#Atr) y el número de clases (#Clas).

Para llevar a cabo los distintos experimentos se utiliza el modelo un 5-folder cross-validation. Ésto significa que para cada data-set se crearán 5 particiones aleatorias. Una de ellas con el 20% de datos del data-set, y 4 particiones más con el 80% de los datos. Las 4 particiones últimas se utilizarán para entrenar, y el resto para probar. Para cada data-set se harán 5 pruebas, utilizando cada vez para testear una partición distinta y el resto como entrenamiento. De esta forma el resultado de cada data-set, será la media del resultado de las 5 particiones.

Este entorno de pruebas experimentales se compartirá para los experimentos de clasificación con los diferentes tipos de conjuntos usados como entradas del sistema (Singleton, Difuso, IVFS), para cada uno de los tipos de conjunto de entrada, se harán pruebas con las diferentes medidas de similaridad (Mitchell, Bustince, Gorzalczany,

4. SCBRD con IVFS y Medidas de Similitud

Zeng-Li y Mendel). Para poder extraer diferencias significativas entre los resultados obtenidos por los diferentes métodos a estudio, es necesario un análisis estadístico. Para ello utilizaremos test-no paramétricos, que son robustos para la comparación de clasificadores. En este estudio empírico aplicaremos la comparación entre dos algoritmos usando el test de Wilcoxon [SHE06] con rangos con signo, estableciendo en 0.05 el nivel de confianza (α) en todos los casos.

Tabla 1. Características de los data-sets

Data-set	#Eje.	#Atr.	#Clas.
Balance	625	4	3
Bupa	345	6	2
Cleveland	297	13	5
Ecoli	336	7	8
Glass	214	9	6
Haberman	306	3	2
Iris	150	4	3
Magic	1902	10	2
New-thyroid	215	5	3
Page-Blocks	548	10	5
Penbased	1099	16	10
Pima	768	8	2
Ring	768	8	2
Vehicle	846	18	4
Wine	178	13	3
Wisconsin	683	9	2
Hayes-Roth	160	5	3
Letter	20000	16	26
Satimage	6435	36	7
Segment	2310	19	7
Shuttle	15	6	7
Twonorm	7400	21	2
Yeast	1484	8	10

5.2 Experimentos

En este capítulo realizaremos los experimentos con cada uno de los tipos de conjuntos de entrada. En el subapartado 5.2.1 veremos los resultados con los conjuntos tipo singleton, en el 5.2.2 se mostrarán los resultados para conjuntos difusos normales y en la subsección 5.2.3 se verán los de los IVFS.

5.2.1 Conjuntos Singleton como entrada

En esta subsección se muestran los resultados de los experimentos considerando las entradas al sistema como conjuntos singleton. En el capítulo 4.2, vimos como se construían este tipo de conjuntos difusos. Si recordamos como eran, veremos que eran bastante especiales, ya que tomaban valor 1 en punto del universo del discurso que coincidía con el valor del patrón para esa variable, en el resto de universos del discurso valía cero. Para poder calcular la similaridad con respecto a conjuntos IVFS, se creaba un singleton con valor de la función de pertenencia inferior igual a la función de pertenencia superior, la cual toma los valores anteriormente descritos.

El hecho de que la entrada clásica al sistema sea la representada por conjuntos singleton lleva a un estudio más pormenorizado de cómo funcionan con cada una de las similaridades que se desean analizar. El hecho es que al estudiarlo en detalle, se descubre que tres de ellos (Gorzalczany, Bustince y Zeng-Li) arrojan el mismo valor de similaridad. Por ello, se decide hacer el estudio con 3 de las similaridades que se mencionaron, concretamente con las de Bustince, Mitchell y Mendel. Para el resto de tipos de entrada se harán los experimentos con cada una de ellas.

En primer lugar vamos a ver el estudio que se hizo con datos sencillos para llegar a la conclusión de que con conjuntos de tipo singleton se obtienen los mismos resultados con las medidas de similaridad de Gorzaczany, Bustince y Zeng-Li.

El método de Mendel no se tiene en cuenta para la realización del estudio en el que se analiza que medidas de similaridad son diferentes. Esto se debe a que además de la forma del conjunto, se tiene en cuenta la distancia entre ambos siendo este un nuevo factor que el resto no tienen en cuenta. Como queda demostrado en el artículo de J.M Mendel, “*A vector similarity meadure for linguistic approximation: Interval type-2 and type-1 fuzzy sets*”, donde compara esta medida de similaridad con las anteriormente mencionadas.

Por lo tanto, descartando esta primer medida, nos queda comparar las medidas de similaridad de Gorzalczany, Bustince, Mitchell y Zeng and Li. Para ello haremos un cálculo sencillo mediante el ejemplo formado por los 2 conjuntos siguientes:

4. SCBRD con IVFS y Medidas de Similitud

$$A = \{x_1, [0.8, 0.9]\}$$

$$B = \{x_1, [1, 1]\}$$

Gorzalczany:

Como esta medida de similitud no es simétrica, se hace el cálculo en un sentido y en otro.

$$S_G(\tilde{A}, \tilde{B}) \left[\min \left(\frac{\max_{x \in X} \{\min(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}}(x))\}}{\max_{x \in X} \underline{\mu}_{\tilde{A}}(x)}, \frac{\max \{\min(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}}(x))\}}{\max_{x \in X} \overline{\mu}_{\tilde{A}}(x)} \right), \max \left(\frac{\max_{x \in X} \{\min(\underline{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{B}}(x))\}}{\max_{x \in X} \underline{\mu}_{\tilde{A}}(x)}, \frac{\max \{\min(\overline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{B}}(x))\}}{\max_{x \in X} \overline{\mu}_{\tilde{A}}(x)} \right) \right]$$

$$S_G(\tilde{A}, \tilde{B}) = \left[\min \left(\frac{0.8}{0.8}, \frac{0.9}{0.9} \right), \max \left(\frac{0.8}{0.8}, \frac{0.9}{0.9} \right) \right] = [1, 1] \Rightarrow \text{La media es 1}$$

$$S_G(\tilde{B}, \tilde{A}) = \left[\min \left(\frac{0.8}{1}, \frac{0.9}{1} \right), \max \left(\frac{0.8}{1}, \frac{0.9}{1} \right) \right] = [0.8, 0.9] \Rightarrow \text{La media es 0,85}$$

Bustince:

Esta medida es simétrica, así que se hará solamente el cálculo en un sentido. Siguiendo los siguientes pasos.

$$i. S_B(\tilde{A}, \tilde{B}) = [S_L(\tilde{A}, \tilde{B}), S_U(\tilde{B}, \tilde{A})]$$

$$ii. S_L(\tilde{A}, \tilde{B}) = r_L(\tilde{A}, \tilde{B}) * r_L(\tilde{B}, \tilde{A})$$

$$S_U(\tilde{A}, \tilde{B}) = r_U(\tilde{A}, \tilde{B}) * r_U(\tilde{B}, \tilde{A})$$

$$iii. r_L(\tilde{A}, \tilde{B}) = \min_{x \in X} \left\{ 1, \min \left(1 - \underline{\mu}_{\tilde{A}}(x) + \underline{\mu}_{\tilde{B}}(x), 1 - \overline{\mu}_{\tilde{A}}(x) + \overline{\mu}_{\tilde{B}}(x) \right) \right\}$$

$$r_U(\tilde{A}, \tilde{B}) = \min_{x \in X} \left\{ 1, \max \left(1 - \underline{\mu}_{\tilde{A}}(x) + \underline{\mu}_{\tilde{B}}(x), 1 - \overline{\mu}_{\tilde{A}}(x) + \overline{\mu}_{\tilde{B}}(x) \right) \right\}$$

Con los datos del ejemplo, son los cálculos siguientes:

$$\begin{cases} r_L(\tilde{A}, \tilde{B}) = \min_{x \in X} \{1, \min(1 - 0.8 + 1, 1 - 0.9 + 1)\} = 1 \\ r_U(\tilde{A}, \tilde{B}) = \min_{x \in X} \{1, \max(1 - 0.8 + 1, 1 - 0.9 + 1)\} = 1 \end{cases}$$

$$\begin{cases} r_L(\tilde{B}, \tilde{A}) = \min_{x \in X} \{1, \min(1 - 1 + 0.8, 1 - 1 + 0.9)\} = 0.8 \\ r_U(\tilde{B}, \tilde{A}) = \min_{x \in X} \{1, \max(1 - 1 + 0.8, 1 - 1 + 0.9)\} = 0.9 \end{cases}$$

$$S_L(\tilde{A}, \tilde{B}) = 1 * 0.8 = 0.8$$

$$S_U(\tilde{A}, \tilde{B}) = 1 * 0,9 = 0,9$$

$$S_B(\tilde{A}, \tilde{B}) = [0,8,0,9] \Rightarrow \text{media } 0,85$$

Mitchell:

El cálculo de la similaridad de Mitchell, se hace a partir de la siguiente fórmula:

$$\mu_{A_e^m}(x_l) = r_m(x_l) \times [\bar{\mu}_{\tilde{A}}(x_l), \underline{\mu}_{\tilde{A}}(x_l)] \underline{\mu}_{\tilde{A}}(x_l) \quad l = 1, 2, \dots, L$$

Con esta fórmula, la cual se repite m veces, con distintos valores de r aleatorios, y haciendo el sumatorio de todos los valores obtenidos y dividiendo entre LxM, se obtiene la similaridad. Viendo que r es un número aleatorio, es imposible que la similaridad de siempre el mismo valor exacto. Por lo tanto esta medida será una de las que deberemos incluir en el estudio.

Zeng-Li:

Realizando los cálculos para el ejemplo. Partimos de la fórmula siguiente:

$$S_Z(\tilde{A}, \tilde{B}) = 1 - \frac{1}{2n} \sum_{i=1}^n (|\underline{\mu}_{\tilde{A}}(x_i) - \underline{\mu}_{\tilde{B}}(x_i)| + |\bar{\mu}_{\tilde{A}}(x_i) - \bar{\mu}_{\tilde{B}}(x_i)|)$$

$$S_Z(\tilde{A}, \tilde{B}) = 1 - \frac{1}{2n} \sum_{i=1}^n |1 - 0,8| + |1 - 0,9| = 1 - \frac{1}{2}(0,3) = 1 - 0,15 = 0,85$$

Como resultado del estudio, vemos que las medidas de Zeng-Li, Gorzalczany y Bustince coinciden en los resultados, por lo tanto con hacer las pruebas para una de ellas, conoceremos el resultado de los otros métodos. Para este cometido se elige la similaridad de Bustince. Debido a este razonamiento, se harán las pruebas para los algoritmos de Mitchell, Bustince y Mendel.

4. SCBRD con IVFS y Medidas de Similitud

El estudio estar  formado por 4 pruebas que son todos las posibles combinaciones entre los dos MRD considerados (Regla Ganadora (WR) y Suma Aditiva (AC)) y las 2 T-normas consideradas para calcular el grado de emparejamiento (producto y m nimo):

- WR + prod
- WR + min
- AC + prod
- AC + min

5.2.1.1 Resultados de Pruebas con Regla Ganadora y Producto

En esta subsecci n se muestran los resultados de las pruebas realizadas con las 3 medidas de similitud, con el m todo de Chi original y con el m todo de Chi con IVFS, utilizando en el MRD de la regla ganadora y el producto para el c lculo del grado de emparejamiento.

En la tabla 2, se muestran los resultados:

Tabla 2. Tabla de resultados del método Chi et al en entrenamiento (tr) y test (tst). Mediante el algoritmo de Chi clásico (Chi) y el algoritmo de Chi con IVFS (chi3-ivfs) y mediante las distintas similitudes (Mitchell, Bustince, Mendel). Para RG y producto

	chi3		chi3-ivfs		MITCHELL		BUSTINCE		MENDEL	
	tr	tst	tr	tst	tr	tst	tr	tst	tr	tst
balance	91,62	89,92	91,10	90,24	91,66	90,40	91,66	90,40	87,54	86,56
bupa	60,73	57,68	59,71	58,26	60,65	57,68	60,51	57,97	60,95	57,97
cleveland	92,22	36,01	90,02	53,21	91,04	52,20	90,87	51,19	90,95	50,52
ecoli	76,18	72,64	69,98	67,58	75,80	71,15	75,65	70,85	83,79	78,89
glass	66,28	57,95	55,94	51,87	63,11	57,93	60,76	56,99	66,39	63,09
haberman	74,57	72,88	73,83	73,53	74,00	72,55	73,91	72,88	74,00	72,88
iris	92,94	92,67	93,61	92,00	94,29	92,67	94,29	92,00	94,62	94,00
magic	75,98	74,87	73,31	72,66	76,10	74,97	76,15	74,92	75,73	74,60
newthyroid	86,32	84,65	81,75	80,93	87,37	86,98	87,02	86,98	89,01	88,84
pageblocks	92,73	91,42	91,50	90,88	92,64	91,42	92,45	91,42	91,95	91,06
penbased	98,66	94,27	95,31	92,27	98,41	94,64	98,41	94,73	97,72	93,91
pima	75,45	72,53	69,71	67,97	75,19	72,78	75,25	72,52	74,76	72,66
ring	59,53	52,70	51,17	50,41	58,48	52,16	58,58	52,16	0,00	0,00
vehicle	65,85	60,88	54,87	51,55	65,11	61,00	64,93	60,76	63,45	59,34
wine	98,73	92,67	98,02	94,90	98,44	94,90	98,44	94,33	98,30	94,94
wisconsin	98,17	90,49	97,65	96,05	98,09	95,46	98,06	95,46	97,91	95,76
Hayes-Roth	82,21	58,40	83,36	72,05	79,93	65,95	80,68	64,47	83,55	67,55
letter	82,21	58,40	30,94	28,95	31,04	29,65	31,17	29,85	31,13	29,30
satimage	48,11	48,06	46,51	46,65	49,28	49,15	48,03	47,90	0,00	0,00
segment	87,05	85,93	82,23	81,99	86,83	86,67	87,14	86,71	87,18	86,45
shuttle	80,26	80,23	78,85	78,80	80,26	80,23	80,26	80,23	80,26	80,23
twonorm	88,83	85,54	81,25	77,84	89,75	86,49	89,78	85,81	87,95	84,05
yeast	29,61	28,91	37,63	36,99	34,36	33,90	34,85	33,83	34,33	32,82
average	78,44	71,29	73,40	69,90	76,17	71,78	76,04	71,49	71,80	67,63

4. SCBRD con IVFS y Medidas de Similitud

A partir de los resultados de la tabla anterior, se realiza el estudio estadístico mediante el test de Wilcoxon comparando de esta manera los resultados que se obtienen con el método de Chi antiguo (tabla 3) y con el algoritmo de Chi modificado para IVFS (tabla 4), con la introducción de las diferentes medidas de similitud.

Tabla 3. Test de Wilcoxon para comparar el algoritmo de Chi original con cada uno de los métodos que utilizan la similitud con conjuntos de entrada singleton. R+ corresponde a la suma de rangos de los métodos con similitud y R- al método de Chi.

Comparación	R ⁺	R ⁻	Hipótesis ($\alpha=0.05$)	p-value
Mitchell vs. Chi	145	45	Rechazado para Mitchell	0.044
Bustince vs. Chi	137	73	Aceptado	0.232
Mendel vs. Chi	132	99	Aceptado	0,566

Como se observa en la tabla 3, el único que no alcanza el valor de aceptación marcado es el método de Mitchell, y aunque todos ellos en suma de rangos superan a los resultados obtenidos mediante el método de Chi, solamente los resultados de Mitchell demuestran estadísticamente que ambos métodos son distintos. La medida de similitud que peor funciona es la de Mendel, lo cual se refleja tanto en la suma de rangos como en el p-value que es el más elevado.

A continuación, en la tabla 4, se muestran los resultados del test estadístico de Wilcoxon al comparar cada uno de los métodos con las distintas similitudes con el algoritmo de Chi con IVFS.

Tabla 4. Test de Wilcoxon para comparar el algoritmo de Chi con IVFS con cada uno de los métodos que utilizan la similitud con conjuntos de entrada singleton. R+ corresponde a la suma de rangos de los métodos con similitud y R- al método de Chi con IVFS.

Comparación	R ⁺	R ⁻	Hipótesis ($\alpha=0.05$)	p-value
Mitchell vs. Chi-IVFS	197	56	Rechazado para Mitchell	0.022
Bustince vs. Chi-IVFS	191	62	Rechazado para Bustince	0.036
Mendel vs. Chi-IVFS	167	109	Aceptado	0,378

Como se puede observar en la tabla 4, todos los métodos que utilizan medidas de similaridad para el cálculo del grado de emparejamiento, superan al método de Chi con IVFS. Esto se puede ver en la suma de rangos, donde los métodos que utilizan medidas de similaridad son muy superiores al método de CHI con IVFS. A pesar de esto, los únicos métodos que no superan el valor de aceptación del p-value que nos habíamos marcado, son el de Mitchell y el de Bustince. En consecuencia, solo podemos decir que estadísticamente estos dos métodos son diferentes del método de Chi y por consiguiente mejores estadísticamente. Por lo tanto podríamos decir que los resultados obtenidos son mejores al compararlo frente al método de Chi con IVFS que con respecto al Chi original.

5.2.1.2 Resultados de Pruebas con Suma Aditiva y Producto

En esta sección se muestran los resultados de las pruebas realizadas en el modelo antes descrito utilizando en el MRD, la inferencia por medio de la suma aditiva y el producto.

En la tabla siguiente, se muestran los resultados con las distintas medidas de similaridad y los distintos algoritmos:

Tabla 5. Tabla de resultados del método Chi et al en entrenamiento (tr) y test (tst). Mediante el algoritmo de Chi clásico (Chi) y el algoritmo de Chi con IVFS (chi3-ivfs) y mediante las distintas similitudes (Mitchell, Bustince, Mendel) para el cálculo del grado de emparejamiento, con suma aditiva y producto

Data-sets	chi3		chi3-ivfs		MITCHELL		BUSTINCE		MENDEL	
	tr	tst	tr	tst	tr	tst	tr	tst	tr	tst
balance	91,14	89,76	90,78	90,08	90,46	90,08	90,30	89,92	88,58	88,16
bupa	59,71	58,26	59,64	57,97	59,64	58,55	59,71	58,84	60,44	58,26
cleveland	90,36	49,54	89,26	53,88	89,86	52,53	89,86	52,86	89,77	51,85
ecoli	71,48	69,34	69,75	66,99	77,37	74,71	77,22	74,41	83,64	80,07
glass	58,17	53,73	55,47	51,87	63,58	58,86	61,00	57,93	66,27	60,76
haberman	73,83	73,53	73,83	73,53	73,83	73,53	73,83	73,53	73,83	73,20
iris	93,61	92,67	93,45	93,33	94,12	94,67	93,61	94,67	91,43	90,67
magic	73,81	73,50	73,41	72,82	76,71	75,87	76,60	75,92	75,65	74,50
newthyroid	83,04	81,40	81,52	80,93	87,02	86,05	87,37	86,51	88,54	87,91
pageblocks	91,63	91,06	91,36	90,88	92,18	91,06	92,09	91,06	92,00	90,88
penbased	96,00	92,82	94,88	91,73	97,04	94,36	97,06	94,36	97,50	94,64
pima	70,69	68,88	69,06	67,71	73,43	71,74	73,43	71,74	73,92	73,31
ring	52,25	50,81	51,03	50,41	56,45	52,30	56,14	51,89	0,00	0,00
vehicle	56,85	54,49	54,42	51,67	60,55	57,81	60,94	57,45	59,84	55,80
wine	98,16	93,79	97,74	95,46	97,88	94,90	98,02	94,90	98,30	96,60
wisconsin	97,62	95,17	97,62	95,90	97,80	96,20	97,80	96,20	97,76	95,46
Hayes-Roth	81,07	53,85	80,88	66,67	78,78	69,69	79,36	68,89	83,36	74,27
letter	82,21	58,40	29,68	27,70	30,02	28,65	30,09	28,65	30,54	28,80
satimage	55,94	55,84	54,38	54,59	55,20	54,75	55,51	54,75	0,00	0,00
segment	87,44	86,71	80,34	80,13	86,93	86,45	87,01	86,32	87,56	87,10
shuttle	80,26	80,23	80,26	80,23	80,35	80,32	80,26	80,23	80,26	80,23
twonorm	96,75	95,68	88,76	88,24	96,79	95,81	96,89	96,08	96,01	94,86
yeast	29,59	28,91	37,68	37,20	36,79	35,99	36,99	36,46	36,50	35,38
average	77,03	71,67	73,70	70,43	76,21	72,82	76,13	72,76	71,81	68,38

A partir de los resultados de la tabla 5, se realiza el estudio estadístico mediante el test de Wilcoxon comparando de esta manera los resultados que se obtienen con el método de Chi y Chi con IVFS, con los métodos que utilizan las diferentes medidas de similitud.

Primero se comparan todos los métodos que utilizan medidas de similitud con el método de Chi original. Los resultados del test de Wilcoxon se pueden observar en la tabla 6.

Tabla 6. Test de Wilcoxon para comparar el algoritmo de Chi original con cada uno de los métodos que utilizan la similitud con conjuntos de entrada singleton. R+ corresponde a la suma de rangos de los métodos con similitud y R- al método de Chi.

Comparación	R⁺	R⁻	Hipótesis ($\alpha=0.05$)	p-value
Mitchell vs. Chi	200	31	Rechazado para Mitchell	0.003
Bustince vs. Chi	181	29	Rechazado para Bustince	0.05
Mendel vs. Chi	144	87	Aceptado	0,322

Como se puede observar en los resultados obtenidos, todos los métodos que utilizan medidas de similitud para calcular el grado de emparejamiento, obtienen mejores resultados que el método de Chi original. En cambio solo el método de Bustince y Mitchell no superan el valor 0.05 en el p-value, siendo el mejor resultado, el obtenido por el método de Mitchell. Por lo tanto solo podemos decir de estos dos métodos que son mejores estadísticamente.

A continuación, en la tabla 7 se puede observar los resultados del test al comparar el método de Chi con IVFS con cada uno de los métodos con medidas de similitud para calcular el grado de emparejamiento.

5. Estudio Experimental

Tabla 7. Test de Wilcoxon para comparar el algoritmo de Chi con IVFS con cada uno de los métodos que utilizan la similaridad con conjuntos de entrada singleton. R+ corresponde a la suma de rangos de los métodos con similaridad y R- al método de Chi con IVFS

Comparación	R ⁺	R ⁻	Hipótesis ($\alpha=0.05$)	p-value
Mitchell vs. Chi-IVFS	208	23	Rechazada para Mitchell	0.001
Bustince vs. Chi-IVFS	209.5	21.5	Rechazada para Bustince	0.001
Mendel vs. Chi-IVFS	151	80	Aceptada	0,217

En la tabla 7 vemos que los resultados para los métodos que utilizan medidas de similaridad son mucho mejores. En la suma de rangos ganan todos al método de Chi con IVFS. En cambio el valor de aceptación no es alcanzado por el método de Mitchell y de Bustince, siendo para ambos de 0,001. El método de Mendel supera este valor. Por lo tanto no podemos concluir que este último sea estadísticamente mejor, ya que no podemos decir que estadísticamente Mendel y Chi con IVFS sean distintos métodos.

5.2.1.3 Resultados de Pruebas con Regla Ganadora y Mínimo

En esta subsección se muestran los resultados de las pruebas realizadas con las 3 medidas de similaridad utilizando en el MRD de la regla ganadora y el mínimo para el cálculo del grado de emparejamiento. Estas pruebas se realizan solamente con 16 de los 23 data –sets anteriores: balance, bupa, cleveland, ecoli, glass, haberman, iris, magic, newthyroid, pageblocks, penbased, pima, ring, vehicle, wine, wisconsin.

En la tabla 8, se muestran los resultados:

Tabla 8. Tabla de resultados del método Chi et al en entrenamiento (tr) y test (tst). Mediante el algoritmo de Chi clásico (Chi) y el algoritmo de Chi con IVFS (chi3-ivfs) y mediante las distintas similitudes (Mitchell, Bustince, Mendel) para el cálculo del grado de emparejamiento, con regla ganadora y mínimo

	chi3		chi3-ivfs		MITCHELL		BUSTINCE		MENDEL	
	tr	tst	tr	tst	tr	tst	tr	tst	tr	tst
balance	91,62	89,92	91,10	90,24	91,58	90,40	91,18	90,40	90,42	89,60
bupa	60,73	57,68	59,71	58,26	59,27	58,26	59,71	58,26	59,93	58,26
cleveland	92,22	36,01	90,02	53,21	93,64	52,54	90,19	51,54	89,94	51,88
ecoli	76,18	72,64	69,98	67,58	76,40	68,46	71,03	68,46	72,45	69,64
glass	66,28	57,95	55,94	51,87	66,47	53,27	57,11	52,80	58,53	53,27
haberman	74,57	72,88	73,83	73,53	74,07	73,21	73,83	73,53	73,83	73,53
iris	92,94	92,67	93,61	92,00	92,44	92,67	93,45	92,00	93,61	93,33
magic	75,98	74,87	73,31	72,66	75,86	73,40	73,75	73,45	73,67	73,34
newthyroid	86,32	84,65	81,75	80,93	88,89	81,40	83,04	81,40	83,51	82,33
pageblocks	92,73	91,42	91,50	90,88	92,45	91,06	91,59	91,06	91,54	90,88
penbased	98,66	94,27	95,31	92,27	98,86	92,64	96,04	92,64	95,79	92,64
pima	75,45	72,53	69,71	67,97	75,04	69,01	70,72	69,01	70,59	69,27
ring	59,53	52,70	51,17	50,41	58,04	50,95	52,55	50,95	40,91	40,41
vehicle	65,85	60,88	54,87	51,55	65,33	54,61	56,85	54,37	56,65	54,02
wine	98,73	92,67	98,02	94,90	99,29	94,90	98,16	94,90	97,88	94,35
wisconsin	98,17	90,49	97,65	96,05	97,61	95,90	97,65	95,90	97,62	96,05
average	81,62	74,64	77,97	74,02	78,62	74,54	78,55	74,42	77,93	73,92

5. Estudio Experimental

A partir de los resultados de la tabla 8, se realiza el estudio estadístico mediante el test de Wilcoxon comparando de esta manera los resultados que se obtienen con el método de Chi y Chi con IVFS, con los métodos que utilizan las diferentes medidas de similitud.

En la tabla 9 vemos los resultados del test estadístico al comparar los métodos que utilizan medidas de similitud para calcular el grado de emparejamiento, con el método de Chi original. En esta tabla se puede ver que aunque todos los métodos que utilizan medidas de similitud empeoran con la suma de rangos con respecto al método de Chi, aunque no podemos decir que el método de Chi sea estadísticamente mejor, ya que el p-value es superior 0,05.

Tabla 9. Test de Wilcoxon para comparar el algoritmo de Chi con cada uno de los métodos que utilizan la similitud con conjuntos de entrada singleton. R+ corresponde a la suma de rangos de los métodos con similitud y R- al método de Chi

Comparación	R ⁺	R ⁻	Hipótesis ($\alpha=0.05$)	p-value
Mitchell vs. Chi	44	76	Aceptada	0.363
Bustince vs. Chi	48	88	Aceptada	0.301
Mendel vs. Chi	49	87	Aceptada	0,326

En la tabla 10 vemos los resultados del test estadístico de los métodos con medidas de similitud comparados con el método de Chi con IVFS. Aquí se puede ver como por suma de rangos el método de Chi es superior a todos los métodos que utilizan medidas de similitud y estadísticamente mejor que los métodos de Mitchell y Bustince, ya que el valor de su p-value es menor que 0.05 y por lo tanto nos indica que ambos métodos son distintos.

Tabla 10. Test de Wilcoxon para comparar el algoritmo de Chi con IVFS con cada uno de los métodos que utilizan la similitud con conjuntos de entrada singleton. R+ corresponde a la suma de rangos de los métodos con similitud y R- al método de Chi con IVFS

Comparación	R ⁺	R ⁻	Hipótesis ($\alpha=0.05$)	p-value
Mitchell vs. Chi-IVFS	13.5	91.5	Rechazada para Chi-IVFS	0.014
Bustince vs. Chi-IVFS	12	66	Rechazada para Chi-IVFS	0.034
Mendel vs. Chi-IVFS	23.5	54.5	Aceptada	0,224

5.2.1.4 Resultados de Pruebas con Suma Aditiva y Mínimo

En esta subsección se muestran los resultados de las pruebas realizadas con las 3 medidas de similaridad utilizando en el MRD la suma aditiva y el mínimo para el cálculo del grado de emparejamiento. Estas pruebas se realizan solamente con 16 de los 23 data –sets que aparecieron al comienzo de la sección, concretamente: balance, bupa, cleveland, ecoli, glass, haberman, iris, magic, newthyroid, pageblocks, penbased, pima, ring, vehicle, wine, wisconsin.

A partir de los resultados de la tabla 12, se realiza el estudio estadístico mediante el test de Wilcoxon comparando de esta manera los resultados que se obtienen con el método de Chi y Chi con IVFS, con los métodos que utilizan las diferentes medidas de similaridad.

En la tabla 11 vemos los resultados del test estadístico al comparar los métodos que utilizan medidas de similaridad para calcular el grado de emparejamiento, con el método de Chi original. Aquí se puede ver como por suma de rangos el método de Chi es superior a todos los métodos que utilizan medidas de similaridad y es mejor estadísticamente, ya que el valor de su p-value es menor que 0.05, por lo que se sabe a ciencia cierta que ambos son métodos distintos.

Tabla 11. Test de Wilcoxon para comparar el algoritmo de Chi con cada uno de los métodos que utilizan la similaridad con conjuntos de entrada singleton. R+ corresponde a la suma de rangos de los métodos con similaridad y R- al método de Chi

Comparación	R ⁺	R ⁻	Hipótesis ($\alpha=0.05$)	p-value
Mitchell vs. Chi	30	90	Aceptada	0.088
Bustince vs. Chi	26	94	Aceptada	0.053
Mendel vs. Chi	29	91	Aceptada	0,078

Tabla 12. Tabla de resultados del método Chi et al en entrenamiento (tr) y test (tst). Mediante el algoritmo de Chi clásico (Chi) y el algoritmo de Chi con IVFS (chi3-ivfs) y mediante las distintas similitudes (Mitchell, Bustince, Mendel). Con suma aditiva y mínimo

	chi3		chi3-ivfs		MITCHELL		BUSTINCE		MENDEL	
	tr	tst	tr	tst	tr	tst	tr	tst	tr	tst
balance	90,66	89,92	90,66	89,92	88,46	88,00	88,94	88,80	89,06	88,80
bupa	59,64	58,26	59,42	57,97	59,05	57,97	59,05	57,97	59,05	57,68
cleveland	89,43	50,21	88,33	53,54	80,73	54,54	80,64	53,19	82,93	52,53
ecoli	69,98	67,28	66,91	65,22	65,65	63,40	65,80	63,70	68,93	65,19
glass	55,23	50,94	54,29	50,94	30,20	26,15	31,02	26,62	34,43	31,30
haberman	73,83	73,53	73,83	73,53	73,75	73,53	73,75	73,53	73,75	73,53
iris	91,09	92,67	88,74	88,00	71,60	71,33	70,08	68,67	73,28	72,00
magic	73,50	73,34	73,33	73,24	75,35	75,50	75,43	75,34	75,50	75,60
newthyroid	81,87	80,93	81,17	80,47	82,46	81,86	82,22	81,40	82,57	81,40
pageblocks	91,59	90,88	91,17	90,88	90,54	90,52	90,54	90,15	90,76	90,51
penbased	95,15	91,73	93,72	90,55	85,14	83,00	85,30	83,36	87,17	84,73
pima	69,32	67,71	68,80	67,58	65,76	65,36	65,76	65,23	66,03	65,23
ring	53,47	52,97	57,23	57,30	49,51	49,59	49,51	49,59	49,51	49,59
vehicle	53,65	50,84	53,68	50,84	38,09	37,83	37,29	36,65	37,64	37,71
wine	97,74	93,79	97,45	95,46	95,33	94,33	94,48	94,35	95,05	95,48
wisconsin	97,65	95,03	97,43	95,90	97,29	96,05	97,18	96,05	97,32	96,05
average	77,74	73,75	77,26	73,83	71,81	69,31	71,69	69,04	72,69	69,83

En la tabla 13 vemos los resultados del test estadístico de los métodos con medidas de similaridad comparados con el método de Chi con IVFS. Aquí se puede ver como por suma de rangos el método de Chi con IVFS es superior a todos los métodos que utilizan medidas de similaridad y es mejor estadísticamente, ya que el valor de su p-value es menor que el valor de aceptación.

Tabla 13. Test de Wilcoxon para comparar el algoritmo de Chi con IVFS con cada uno de los métodos que utilizan la similaridad con conjuntos de entrada singleton. R+ corresponde a la suma de rangos de los métodos con similaridad y R- al método de Chi con IVFS

Comparación	R ⁺	R ⁻	Hipótesis ($\alpha=0.05$)	p-value
Mitchell vs. Chi-IVFS	18	87	Rechazado para Chi-IVFS	0.03
Bustince vs. Chi-IVFS	13	92	Rechazado para Chi-IVFS	0.013
Mendel vs. Chi-IVFS	20	100	Rechazado para Chi-IVFS	0,023

Para concluir con el primer tipo de conjuntos de entrada al sistema, diremos que utilizando el producto vemos que los nuevos métodos mejoran en general al método de Chi original y al Chi con IVFS, pero con el mínimo éstos últimos funcionan mucho mejor. Por esta razón se decide que se van a ejecutar para el resto de tipos de conjuntos de entrada solamente pruebas con el producto.

5.2.2 Conjuntos Difusos como entrada

En esta subsección se muestran los resultados de los experimentos para entradas al sistema de conjuntos difusos.

En los siguientes apartados pasaremos a analizar resultados obtenidos tras aplicar el modelo clasificador a los 16 data-sets: balance, bupa, cleveland, ecoli, glass, haberman, iris, magic, newthyroid, pageblocks, penbased, pima, ring, vehicle, wine, wisconsin. El estudio estará formado por 2 pruebas que son todas las posibles combinaciones entre los dos MRD considerados (Regla Ganadora (WR) y Suma Aditiva (AC)) y la T-norma producto para calcular el grado de emparejamiento:

- WR + prod
- AC + prod

Aquí no se harán pruebas con la T-norma mínimo para calcular el grado de emparejamiento, porque para los conjuntos de entrada singleton dio muy mal resultado. Se harán pruebas para el modelo definido en el capítulo cuatro con cada una de las 5 medidas de similaridad vistas. Ahora los test estadísticos se harán con respecto al mejor de ellos, ya que no podemos comparar con Chi y Chi-IVFS ya que no están implementados para conjuntos difusos.

5.2.2.1 Resultados de Pruebas con Regla Ganadora y Producto

En esta subsección se muestran los resultados de las pruebas realizadas en los modelos antes descritos para el MRD con regla Ganadora y producto para el cálculo de la semejanza a cada regla.

En la tabla 14, se muestran los resultados:

Tabla 14. Tabla de resultados para conjuntos difusos mediante distintas medidas de similaridad (Mitchell, Bustince, Mendel, Gorzalczany y Zeng-Li)

data-sets	MITCHELL		BUSTINCE		MENDEL		GORZALCZANY		ZENG-LI	
	tr	tst	tr	tst	tst	tr	Tr	tst	tr	tst
balance	84,61	84,32	89,90	88,48	88,22	87,68	90,42	88,96	90,46	90,24
bupa	58,25	57,97	61,38	57,97	60,36	58,55	59,78	58,84	35,78	35,36
cleveland	59,10	54,21	93,24	47,81	90,70	54,21	84,71	55,24	90,11	51,19
ecoli	48,84	48,23	80,73	77,10	81,63	77,40	73,79	70,83	62,89	59,80
glass	15,51	15,87	67,45	60,30	64,98	61,23	32,32	27,54	37,29	34,88
haberman	63,96	64,50	75,31	73,53	73,75	73,86	73,75	73,53	59,08	58,78
iris	51,43	52,00	96,13	96,67	94,29	92,67	89,08	86,67	62,86	62,00
magic	39,95	39,58	73,79	72,92	75,94	75,24	77,18	76,50	51,89	51,88
newthyroid	19,06	19,07	87,25	86,51	87,95	86,98	86,55	86,05	52,87	50,70
pageblocks	90,17	90,15	92,77	91,60	91,72	90,88	91,04	90,70	53,87	53,80
penbased	50,42	49,18	98,84	94,91	97,63	94,45	95,09	92,55	19,45	18,91
pima	65,73	65,62	76,62	74,48	74,14	73,44	71,47	69,40	71,93	70,83
ring	49,51	49,59	55,91	56,35	64,74	60,27	50,93	51,08	50,15	49,86
vehicle	32,91	31,31	66,14	61,00	61,88	58,52	56,80	54,49	0,00	0,00
wine	69,58	66,87	98,16	92,65	98,59	95,46	97,60	94,90	32,67	33,14
wisconsin	47,42	47,40	98,35	96,05	97,84	96,19	97,95	96,49	52,59	52,11
average	52,90	52,24	82,00	76,77	81,52	77,31	76,78	73,36	51,49	48,34

5. Estudio Experimental

A partir de los resultados de la tabla anterior, se realiza el estudio estadístico mediante el test de Wilcoxon, comparando de esta manera los resultados que se obtienen con el método que da mejores resultados de los que utilizan medidas de similaridad, (en este caso el método con la medida de similaridad de Mendel), con cada uno de los métodos de que utilizan las diferentes medidas de similaridad para conjuntos difusos.

En la tabla 15 vemos los resultados del test estadístico al comparar los métodos que utilizan medidas de similaridad para calcular el grado de emparejamiento. Todos ellos se comparan con el que a priori parece el mejor. Mirando la tabla 14 se selecciona el método de Mendel que es el que mejores resultados medios arroja.

En esta tabla 15 se puede ver como con la similaridad de Mendel los resultados obtenidos son mejores que los obtenidos con el resto de métodos. De hecho se demuestra estadísticamente que es superior a los métodos de Mitchell, Zeng y Li, y Gorzalczany, ya que el p-value es inferior a 0,05 y por lo tanto ambos métodos son distintos. En cambio, aunque la suma de rangos para el método de Mendel es superior que para el de Bustince su p-value es 0.535, con lo que no podemos decir que sea estadísticamente superior.

Tabla 15. Test de Wilcoxon para comparar el método que utiliza la medida de similaridad de Mendel con los que utilizan el resto de medidas de similaridad para conjuntos difusos como entrada. R- corresponde a la suma de rangos del método con similaridad de Mendel y R+ al resto de métodos

Comparación	R ⁺	R ⁻	Hipótesis ($\alpha=0.05$)	p-value
Mitchell vs. Mendel	0	120	Rechazada para Mendel	0.001
Bustince vs. Mendel	56	80	Aceptada	0.535
Gorzalczany vs. Mendel	29	107	Rechazada para Mendel	0,044
Zeng-Li vs. Mendel	1	135	Rechazada para Mendel	0,001

5.2.2.2 Resultados de Pruebas con Suma Aditiva y Producto

En esta sección se muestran los resultados de las pruebas realizadas en el modelo descrito, utilizando en el MRD la suma aditiva y como T-norma el producto.

En la tabla 16, se muestran los resultados:

Tabla 16. Tabla de resultados para conjuntos difusos del método utilizando para calcular el grado de emparejamiento las medidas de similaridad de Mitchell, Bustince, Mendel, Gorzalczany y Zeng –Li

	MITCHELL		BUSTINCE		MENDEL		GORZALCZANY		ZENG-LI	
	tr	tst	tr	tst	tr	tst	Tr	tst	tr	tst
balance	82,00	81,76	90,62	89,28	88,62	88,16	90,38	90,08	90,30	89,92
bupa	58,11	57,97	61,16	58,26	60,22	58,26	59,27	58,55	35,78	35,36
cleveland	53,76	53,88	93,24	48,14	90,36	54,88	82,16	56,58	89,86	52,86
ecoli	47,35	48,23	80,81	77,40	81,70	77,99	72,52	70,25	63,27	60,98
glass	13,63	13,54	67,57	60,30	65,22	60,30	25,03	22,89	38,24	36,28
haberman	73,42	73,53	75,31	73,53	73,75	73,53	73,67	73,53	59,08	58,78
iris	46,55	46,67	96,64	96,67	93,45	93,33	88,07	86,67	62,86	62,00
magic	35,14	35,12	73,94	73,24	76,00	75,24	78,19	78,07	51,89	51,88
newthyroid	16,37	16,28	87,49	86,51	87,84	87,44	86,55	86,05	52,75	50,70
pageblocks	89,76	89,78	92,77	91,60	91,72	90,88	90,72	90,33	53,87	53,80
penbased	37,97	37,64	98,84	95,00	97,50	94,36	94,06	91,73	19,45	18,91
pima	65,21	65,10	76,65	74,61	74,01	73,31	70,20	69,01	71,77	70,70
ring	49,51	49,59	56,01	56,89	65,82	61,62	49,85	50,14	50,15	49,86
vehicle	29,98	29,31	66,03	61,12	61,20	57,69	53,92	52,02	0,00	0,00
ine	63,77	61,32	98,16	92,65	98,30	95,46	97,17	95,46	32,67	33,14
wisconsin	35,06	34,99	98,31	96,05	97,80	96,34	97,80	96,93	52,08	51,96
average	49,85	49,67	82,10	76,95	81,47	77,42	75,60	73,02	51,50	48,57

5. Estudio Experimental

A partir de los resultados de la tabla 16, se realiza el estudio estadístico mediante el test de Wilcoxon comparando de esta manera los resultados que se obtienen con las distintas medidas de similaridad para calcular el grado de emparejamiento, con el mejor de ellos. En la tabla 17 se puede observar los resultados del test estadístico.

En ella se puede observar que el método de Mendel es superior a todos ellos según la suma de rangos, pero solamente en el caso de las medidas de similaridad de Mitchell y Zeng-Li podemos decir que sea estadísticamente mejor, ya que en estos casos el p-value es inferior que el valor que habíamos fijado y por tanto sabemos que son métodos diferentes. En los casos de Bustince y Gorzalczany, no hay tanta diferencia en la suma de rangos y el p-value supera el valor que nos habíamos marcado, por lo tanto no podemos decir que sean métodos distintos. El mayor p-value es con el método de Bustince, lo cual indica que después de Mendel, la de Bustince es la medida de similaridad que mejor funciona.

Tabla 17. Test de Wilcoxon para comparar el método que utiliza la medida de similaridad de Mendel con los que utilizan las demás medidas de similaridad para conjuntos difusos como entrada. R- corresponde a la suma de rangos de los métodos con similaridad de Mendel y R+ al resto de métodos.

Comparación	R⁺	R⁻	Hipótesis ($\alpha=0.05$)	p-value
Mitchell vs. Mendel	0	120	Rechazada para Mendel	0.001
Bustince vs. Mendel	41	50	Aceptada	0.753
Gorzalczany vs. Mendel	23	82	Aceptada	0,064
Zeng-Li vs. Mendel	1	135	Rechazada para Mendel	0,001

5.2.3 Entrada con IVFS

En esta subsección se muestran los resultados de los experimentos para entradas al sistema de conjuntos IVFS.

En los siguientes apartados pasaremos a analizar resultados obtenidos tras aplicar el modelo clasificador a los 16 data-sets: balance, bupa, cleveland, ecoli, glass, haberman, iris, magic, newthyroid, pageblocks, penbased, pima, ring, vehicle, wine, wisconsin. El estudio estará formado por 2 pruebas que son todas las posibles combinaciones entre los dos MRD considerados (Regla Ganadora (WR) y Suma Aditiva (AC)) y la T-norma producto para calcular el grado de emparejamiento:

- WR + prod
- AC + prod

Aquí no se harán pruebas con la T-norma mínimo para calcular el grado de emparejamiento, porque para los conjuntos de entrada singleton dio muy mal resultado. Se harán prueba para el modelo definido en el capítulo cuatro con cada una de las 5 medidas de similaridad vistas. Ahora los test estadísticos se harán con respecto al mejor de ellos, ya que no podemos comparar con Chi y Chi-IVFS ya que no están implementados para conjuntos difusos.

5.2.3.1 Resultados de Pruebas con Regla Ganadora y Producto

En esta sección se muestran los resultados de las pruebas realizadas en el modelo antes descrito utilizando en el MRD el método de la regla ganadora y el producto para el cálculo de la semejanza a cada regla.

En la tabla 18, se muestran los resultados:

Tabla 18. Tabla de resultados para conjuntos IVFS del método Chi et al en entrenamiento (tr) y test (tst). Utilizando las distintas similitudes (Mitchell, Bustince, Mendel, Gorzalczany y Zeng-Li) para el cálculo del grado de emparejamiento.

Data-sets	MITCHELL		BUSTINCE		MENDEL		GORZALCZANY		ZENG-LI	
	tr	tst	tr	tst	tr	tst	Tr	tst	tr	tst
balance	80,68	79,84	88,10	86,08	87,62	86,40	91,62	90,56	91,66	90,40
bupa	56,65	55,07	62,47	57,97	60,73	57,97	59,49	57,10	47,35	44,93
cleveland	79,88	53,53	93,24	48,48	91,38	52,87	90,36	54,88	90,87	51,19
ecoli	45,93	43,46	80,81	76,80	82,15	77,70	76,70	72,03	62,67	58,31
glass	19,27	18,67	67,22	60,30	66,28	64,02	46,30	42,52	36,47	34,88
haberman	63,96	64,50	75,22	73,20	74,00	73,20	73,83	73,53	59,32	58,12
iris	64,54	64,00	95,46	93,33	95,97	94,00	95,29	93,33	68,57	66,00
magic	54,77	53,95	73,01	72,13	75,80	74,45	77,34	76,39	64,86	64,88
newthyroid	28,65	29,30	86,67	85,58	88,19	86,98	88,30	86,98	52,51	50,70
pageblocks	91,68	91,24	92,77	91,78	91,91	91,06	92,50	91,60	53,87	53,80
penbased	95,40	92,09	98,77	94,09	98,00	94,27	98,25	94,64	19,64	19,00
pima	66,35	66,16	76,95	74,48	74,83	73,18	74,44	72,66	73,07	71,22
ring	49,51	49,59	54,86	55,68	60,20	53,65	71,81	71,76	47,75	48,92
vehicle	46,26	45,40	66,14	60,41	64,43	60,88	60,91	57,22	0,00	0,00
wine	98,02	94,90	98,30	93,21	98,73	95,48	98,73	94,33	19,53	20,19
wisconsin	97,40	96,63	98,39	95,76	97,95	95,90	97,73	96,19	52,70	51,82
average	64,93	62,40	81,77	76,21	81,76	77,00	80,85	76,61	52,55	49,02

A partir de los resultados de la tabla 18, se realiza el estudio estadístico mediante el test de Wilcoxon comparando de esta manera los resultados que se obtienen con las distintas medidas de similaridad para calcular el grado de emparejamiento, con la mejor de las medidas. En la tabla 19 se puede observar los resultados del test estadístico.

Tabla 19. Test de Wilcoxon para comparar el método que utiliza la medida de similaridad de Mendel con los que utilizan las demás medidas de similaridad para conjuntos IVFS como entrada. R+ corresponde a la suma de rangos de los métodos con similaridad distinta de Mendel y R- al método con similaridad de Mendel. Utilizando regla ganadora para el cálculo del grado de emparejamiento.

Comparación	R ⁺	R ⁻	Hipótesis ($\alpha=0.05$)	p-value
Mitchell vs. Mendel	8	128	Rechazada para Mendel	0.002
Bustince vs. Mendel	24	81	Aceptada	0.074
Gorzalczany vs. Mendel	56	64	Aceptada	0.820
Zeng-Li vs. Mendel	3	133	Rechazada para Mendel	0.001

Como se puede observar en la tabla 19, el método de Mendel que a priori parecía el mejor, se ve como comparándolo con los otros métodos es mejor según el valor dado por la suma de rangos, ya que supera a todos ellos. En cambio solo podemos decir que estadísticamente es superior al método de Mitchell y Zeng-Li, ya que en el resto de casos el p-value es superior a 0,005 que es lo que habíamos marcado como umbral para aceptar la hipótesis de que ambos métodos son iguales.

Además del test estadístico se puede extraer que después de Mendel la medida de similaridad que mejor funciona para este modelo con IVFS como entrada es la de Gorzalczany, ya que es la que obtiene valores más igualados en la suma de rangos y la que obtiene un p-value mayor. Bustince también funciona bastante bien y en el caso Zeng-Li y Mitchell la medida de similaridad de Mendel es muy superior.

5.2.3.2 Resultados de pruebas con Suma Aditiva y Producto

En esta sección se muestran los resultados de las pruebas realizadas en el modelo antes descrito, utilizando como MRD que utiliza suma aditiva y producto.

A partir de los resultados de la tabla 20, se realiza el estudio estadístico mediante el test de Wilcoxon comparando de esta manera los resultados que se obtienen con las distintas medidas de similaridad para calcular el grado de emparejamiento, con el mejor de ellos. En la tabla 21 se puede observar los resultados del test estadístico.

En los resultados de la tabla 21 podemos ver como con la similaridad de Mendel se obtienen los mejores resultados. De hecho funciona mejor que todas las demás similaridades, aunque solamente podemos decir que es estadísticamente superior a la de Mitchell, Zeng-Li y Gorzalczany, ya que en la comparación con la de Bustince se obtienen peores resultados según la suma de rangos, pero el p-value que se obtiene es 0,069. Al ser superior al valor que se había fijado como valor de aceptación, no podemos decir que sean diferentes métodos y por tanto estadísticamente superior.

Es curioso que en la tabla 20, la media de los resultados de la medida de similaridad de Gorzalczany, es mejor que la media de los resultados de la medida de similaridad de Bustince. En cambio al hacer el test estadístico y compararlos ambos con la similaridad de Mendel, el que mejor resultados ofrece es el de Bustince.

En la tabla 20, se muestran los resultados:

Tabla 20. Tabla de resultados para conjuntos IVFS del método Chi et al en entrenamiento (tr) y test (tst). Con las distintas medidas de similaridad (Mitchell, Bustince, Mendel, Gorzalczany y Zeng-Li) combinación del producto e inferencia de suma aditiva.

data-sets	MITCHELL		BUSTINCE		MENDEL		GORZALCZANY		ZENG-LI	
	tr	tst	tr	tst	Tr	Tst	Tr	tst	tr	tst
balance	78,92	78,56	90,26	89,92	88,54	88,00	90,46	86,24	90,30	89,92
bupa	58,11	57,97	59,35	58,84	60,22	58,26	60,44	60,29	35,78	35,36
cleveland	53,76	53,88	83,43	55,90	89,94	55,20	92,80	51,33	89,86	52,86
ecoli	47,12	48,23	73,79	71,14	82,45	79,17	81,27	71,18	63,27	60,98
glass	13,63	13,54	28,79	28,97	64,63	61,23	62,59	51,16	38,24	36,28
haberman	73,42	73,53	73,75	73,53	73,83	73,53	74,40	72,58	59,08	58,78
iris	46,55	46,67	88,57	87,33	93,11	93,33	94,96	96,67	62,86	62,00
magic	35,14	35,12	78,03	77,76	75,89	75,03	74,80	73,23	51,89	51,88
newthyroid	16,37	16,28	86,32	85,58	87,84	87,44	88,07	83,72	52,75	50,70
pageblocks	89,76	89,78	90,76	90,33	91,72	90,88	92,13	92,18	53,87	53,80
penbased	34,81	34,73	94,61	92,36	97,50	94,36	98,04	94,64	19,45	18,91
pima	65,21	65,10	70,46	69,66	73,92	73,18	75,40	72,08	71,77	70,70
ring	49,51	49,59	51,51	51,49	66,23	62,03	54,75	54,73	50,15	49,86
vehicle	28,83	28,01	55,08	53,20	61,14	57,10	63,91	54,12	0,00	0,00
wine	60,80	59,62	97,03	95,46	98,30	96,03	98,87	95,00	19,53	20,19
wisconsin	35,06	34,99	97,80	96,78	97,80	96,19	97,87	95,91	52,08	51,96
average	49,19	49,10	76,22	73,64	81,44	77,56	81,30	75,32	50,68	47,76

5. Estudio Experimental

Tabla 21. Test de Wilcoxon para comparar el método que utiliza la medida de similaridad de Mendel con los que utilizan las demás medidas para conjuntos ivfs como entrada. R^+ corresponde a la suma de rangos de los métodos con similaridad distinta de Mendel y R^- al método con similaridad de Mendel. Utilizando regla ganadora para el cálculo del grado de emparejamiento.

Comparación	R^+	R^-	Hipótesis ($\alpha=0.05$)	p-value
Mitchell vs. Mendel	0	120	Rechazada para Mendel	0.001
Bustince vs. Mendel	28	92	Aceptada	0.069
Gorzalczany vs. Mendel	27.5	108.5	Rechazada para Mendel	0.036
Zeng-Li vs. Mendel	1	135	Rechazada para Mendel	0.001

6. CONCLUSIONES

Como ya hemos comentado varias veces en esta memoria, el propósito del proyecto era comparar el rendimiento del modelo propuesto en el artículo “Improving the Performance of Fuzzy Rule-Based Classification Systems with Interval-Valued Fuzzy Sets and Genetic Amplitude Tuning”, utilizando el MRD y el de Chi (implementación hecha en el artículo) y utilizando cada una de las 5 medidas de similaridad para comparar cada etiqueta lingüística con el valor del nuevo patrón para una variable dada. Es decir, ver si con las medidas de similaridad se mejoraban los resultados obtenidos anteriormente con el algoritmo de Chi y Chi con IVFS.

En los primeros capítulos se explicó la base teórica que sustenta las medidas de similaridad mencionadas. Después se presentó el modelo sobre el que se iba a probar el rendimiento de estas y finalmente en el capítulo 5 se presentaron los resultados de estos. En este capítulo se analiza de forma global los resultados obtenidos.

Tal y como se dividieron los experimentos según el tipo de entrada: Singleton, difuso o IVFS, se muestra a continuación la comparación de los resultados obtenidos en el capítulo 5, para cada uno de ellos, comparando primero según las distintas t-normas utilizadas, y después entre los diferentes métodos de inferencia. Para ello se utilizarán los mejores resultados de cada método de inferencia.

En las tablas siguientes, se pueden ver las conclusiones según los mejores resultados (obtenidos del método de Wilcoxon) para ambos métodos de inferencia, para entradas de tipo singleton, difuso y IVFS.

Para entradas de tipo singleton, los mejores resultados se obtienen mediante la t-norma producto para el cálculo del grado de emparejamiento. Esta t-norma funciona mejor que el mínimo comparando tanto con el método de Chi original, como con el método de Chi con IVFS. También en ambos casos se mejora utilizando como método de inferencia la suma aditiva en lugar de la regla ganadora. Entre todas medidas de similaridad estudiadas, los mejores resultados obtenidos para el caso de conjuntos de tipo singleton son para la medida de similaridad de Mitchell. Este método mejora

6. Conclusiones

sustancialmente los resultados obtenidos con los métodos de Chi original y Chi con IVFS, como se puede ver en las tablas de la 3 a la 13 del capítulo anterior, donde se muestran los resultados de los test estadísticos de las pruebas realizadas para este tipo de conjuntos.

Ahora pasamos a analizar los resultados obtenidos para conjuntos difusos y ver los resultados que se obtuvieron para los métodos de inferencia, suma aditiva y regla ganadora, al comparar Mendel con Bustince, que son los dos métodos más igualados. Con estos datos se crea la tabla 26, donde se puede ver que con suma aditiva los resultados se igualan más entre los dos métodos. Siendo algo superior el de Mendel, aunque como se cumple la hipótesis de que ambos son iguales, no podemos decir que sea estadísticamente superior.

Tabla 22. Comparativa del Método de Mendel con Bustince. Donde R- es la suma de rangos para Mendel y R+ la suma de rangos para Bustince.

Comparación	R ⁺	R ⁻	Hipótesis ($\alpha=0.05$)	p-value
Regla Ganadora	56	80	Aceptado	0.535
Suma Aditiva	41	50	Aceptado	0.753

A continuación analizamos los resultados obtenidos para conjuntos IVFS y ver los resultados que se obtuvieron para los métodos de inferencia, suma aditiva y regla ganadora, al comparar Mendel con Bustince. Se comparan los resultados obtenidos con estas dos medidas de similaridad porque son los dos más igualados. Con estos datos se crea la tabla 27. En ella donde se puede ver que en ambos casos los resultados son mejores para Mendel y es algo mejor para suma aditiva que para la regla ganadora.

Comparando las tablas 26 y 27, podemos decir que la similaridad de Bustince funciona ligeramente mejor con una entrada de difusos que con una entrada de conjuntos IVFS.

Tabla 23. Comparativa del Método de Mendel con Bustince. Donde R^- es la suma de rangos para Mendel y R^+ la suma de rangos para Bustince.

Comparación	R^+	R^-	Hipótesis ($\alpha=0.05$)	p-value
Regla Ganadora	24	81	Aceptado	0.074
Suma Aditiva	28	92	Aceptado	0.069

Por último, cabe mencionar que como resultado del proyecto ha sido aceptado en el congreso ISCAMI que se celebrará los días 21 y 22 de Mayo en Bratislava, el trabajo “*Using Similarity Measures in Fuzzy Rule-Based Classification Systems with Interval-Valued Fuzzy Sets*”.

REFERENCIAS

[BOS92] Boser, B.; Guyon, I.; Vapnik, V. "A Training Algorithm for Optimal Margin Classifiers". In Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, COLT, 1992.

[CVAP95] Cortes, C.; Vapnik, V. "Support Vector Networks". Machine Learning, 20(3):273-297, 1995.

[ECJM96] Enrique Castillo, José Manuel Gutiérrez, y Ali S. Hadi Sistemas Expertos y Modelos de Redes Probabilísticas. 1996

[JJC04] José Hernández Orallo, M^a José Ramírez Quintana, César Ferri Ramírez. Introducción a la minería de datos. 2004

[MZ82] Mizumoto M. y Zimmermann H.-J (1982) Comparison of fuzzy reasoning methods. Fuzzy Sets and Systems

[CWY95] Chi Z., Wu J., y Yan H. "Handwritten numeral recognition using self-organizing maps and fuzzy rules. Pattern Recognition". 1995

[CYP96] Chi Z., Yan H., y Pham T. "Fuzzy algorithms with applications to image processing and pattern recognition". 1996

[DP92] Dubois D. y Prade H. "Upper and lower images of a fuzzy set induced by a fuzzy relation: Applications of fuzzy inference and diagnosis". 1992

[HL96] Hong T. P. y Lee C. Y. "Induction of fuzzy rules and membership functions from training examples". 1996.

[Kun96] kuncheva L.I. "On the equivalence between fuzzy and statistical classifiers. International Journal of Uncertainty, Fuzzyiness and Knowledge-Based". 1996

7. Referencias

[PM92] Pal S. K. y Mnadal D. P. “Linguistic recognition system base don approximate reasoning”. 1992

[SHE06] D. Sheskin, Handbook of parametric and nonparametric statistical. procedures, (Chapman & Hall/CRC, second edition). 2006

[STL84] Stevens, L. (1984), Artificial Intelligence. The Search for the Perfect Machine. Hayden Book Company, Hasbrouck Heights, N.J.1984

[VAP98] Vapnik, V. Stitistical Learning Theory. Wiley, 1998.

[YZ96] Yuan Y. y Zhuang H. “A genetic algorithm for generating fuzzy classification rules”. 1996

[Zad73] Zadeh, L.A.(1973). Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on Systems, Man, and Cybernetics.