Universidad Central “Marta Abreu” de Las Villas

Facultad de Matemática, Física y Computación

Trabajo para optar por el Título Académico

Máster en Ciencia de la Computación

Empleo de relaciones de similaridad borrosa para el cálculo de pesos en algoritmos de aprendizaje

Autora:

Lenniet Coello Blanco

Tutores:

Dra. C Yaima Filiberto Cabrera

Dra. C Yailé Caballero Mota

Consultante:

Dr. C Rafael Bello Pérez

2015

Resumen

La Medida de Calidad de la Similaridad basada en la Teoría de los Conjuntos Aproximados constituye una alternativa eficaz para el análisis de conjuntos de datos ya que representa el grado de similaridad entre los objetos de un sistema de decisión heterogéneo. No obstante esta medida presenta la limitante de usar umbrales en la construcción de las relaciones de semejanza. Estos umbrales constituyen parámetros a ajustar, haciendo los métodos derivados muy sensible a pequeñas variaciones. En esta investigación se propone usar las relaciones borrosas de la Teoría de los Conjuntos Borrosos para eliminar los umbrales y flexibilizar la construcción de las relaciones de similaridad. Con esta modificación se redefine una nueva métrica denominada Medida de la Calidad de la Similaridad Borrosa. La nueva medida sustituye la relación de semejanza entre dos objetos por una relación borrosa binaria, que cuantifica la fortaleza de la relación en un rango de [0,1]. La relación borrosa está caracterizada por una función de pertenencia, que en este caso se define por una función de semejanza basada en la Teoría de los Conjuntos Aproximados Extendida. La nueva medida incluye el cálculo de los pesos de los rasgos; y se estudia su impacto al igual que su precursora para mejorar el desempeño de los k-Vecinos más Cercanos y el Multilayer Perceptron. Los resultados experimentales en bases de datos internacionales y aplicados en dos áreas del conocimiento: la Ingeniería Civil y el Rendimiento académico; muestran un comportamiento semejante en cuanto a precisión entre ambas medida.

Abstract

The Similarity Quality Measure based on Rough Sets Theory is an effective alternative for intelligent analysis of heterogeneous data sets. This measures represents the degree of similarity between objects of a mixed decision system. However this measure has the limitation of use thresholds building similarity relations. These thresholds are parameters to adjust the method, making it very sensitive to small variations. In this research it use fuzzy relations to eliminate the thresholds in the Measure Quality of Similarity and make it more flexible building the relationships of similarity. With this modification it redefines a new metric called Fuzzy Similarity Quality measure. The new measure replaces the relation of similarity between two objects by a binary fuzzy relation, which quantify the strength of the relationship in a range of [0.1]. The fuzzy relation is characterized by a membership function, which in this case is defined by a similarity function based on extended Rough Set theory. The new measure includes calculate features weights; and its impact as its predecessor studying for improve the performance of the k-Nearest Neighbors and Multilayer Pereceptron. Experimental results in International database and it applied in two areas of knowledge: Engineering Civil and academic performance; the result show a similar behavior between them for accuracy measure.

Tabla de contenido

INTRODUCCIÓN **1**

CAPÍTULO I. FUNDAMENTOS TEÓRICOS **6**

1.1. Análisis Inteligente de los Datos y Aprendizaje Automático 6

1.2. Teoría de los Conjuntos Aproximados 9

1.2.1. Teoría de los Conjuntos Aproximados Extendida *12*

1.3. Medida de Calidad de la Similaridad 14

1.4. Teoría de los Conjuntos Borrosos 16

1.5. Relaciones Borrosas 20

1.6. La Teoría Fuzzy- Rought Set y la Teoría Rough-Fuzzy Set 21

Conclusiones parciales 24

CAPÍTULO II. RELACIONES BORROSAS EN LA MEDIDA DE CALIDAD DE LA SIMILARIDAD **25**

2.1. Medida de Calidad de la Similaridad Borrosa 25

2.2. Método para cálculo de pesos para los rasgos basado en la medida de Medida de Calidad de la Similaridad Borrosa 26

2.3. Aplicación de la medida de Medida de Calidad de la Similaridad Borrosa en métodos del Aprendizaje Automático 32

2.3.1. Efecto del empleo de los pesos de los rasgos en el método k-vecinos más similares *32*

2.3.2. Efecto del empleo de los pesos de los rasgos en el método Perceptron multicapa *35*

2.4. Estudio experimental estadístico para validar la Medida de Calidad de la Similaridad Borrosa. 37

2.4.1 Validación experimental en conjuntos de datos internacionales *41*

Conlcusiones parciales 45

CAPÍTULO III. APLICACIONES DE LA MEDIDA DE CALIDAD DE LA SIMILARIDAD BORROSA **46**

3.1. Aplicaciones en la Ingeniería Civil *46*

3.2. Aplicaciones en el Rendimiento Académico *51*

Conclusiones parciales 56

CONCLUSIONES **57**

RECOMENDACIONES **58**

REFERENCIAS BIBLIOGRÁFICAS ..... **I**

Introducción

Durante las últimas décadas, el tamaño de los conjuntos de datos legibles por máquina se ha incrementado drásticamente y el problema de la explosión de datos se ha hecho evidente. Por otro lado, la evolución reciente de la informática ha proporcionado la infraestructura básica para el acceso rápido a grandes cantidades de datos en línea. Muchos de los métodos computacionales avanzados para extraer información de grandes cantidades de datos están empezando a madurar. Estos acontecimientos han creado una nueva gama de problemas y desafíos para los analistas, así como nuevas oportunidades de los sistemas inteligentes de análisis de datos. También han dado lugar a la aparición del campo de Análisis Inteligente de Datos (AID), una combinación de diversas disciplinas, incluyendo Artificial Inteligencia y Estadística, en particular.

Las técnicas de análisis estadístico, desarrolladas hace tiempo, permiten obtener cierta información útil, pero no inducir relaciones cualitativas generales, o leyes, previamente desconocidas; para esto se requieren técnicas de análisis inteligente. La Teoría de los Conjuntos Aproximados dentro del campo de la Inteligencia Artificial (IA) no convencional, ha abierto nuevas tendencias en el desarrollo de técnicas de análisis de datos, esencialmente para tratar con datos incompletos e inexactos.

La Teoría de los Conjuntos Aproximados (*Rough Set Theory*, RST) propuesta por Pawlak en 1982 brinda varias medidas para el análisis de los datos. La medida llamada calidad de la clasificación se emplea cuando se usa la RST para construir la función de evaluación; esta medida permite calcular la consistencia de un sistema de decisión. Su principal limitante es que solo se usa para sistemas de decisión donde el dominio de los rasgos es discreto.

En (Filiberto et al., 2010a) se propone una nueva medida para el caso de sistemas de decisión en el que el dominio de los rasgos, incluido el rasgo de decisión, no tiene que ser necesariamente discreto. Esta medida nombrada Medida de Calidad de la Similaridad representa el grado de similaridad entre los objetos de un sistema de decisión heterogéneo.

En (Filiberto et al., 2010b) se propuso además un método para construir relaciones de similaridad, a partir de la combinación de la metaheurística optimización basada en partículas (*Particule Swarm Optimization*, PSO) y la Medida de Calidad de la

Similaridad, la cual se utiliza como función de evaluación heurística. El proceso de construcción de la relación de similaridad incluye el cálculo de los pesos de los rasgos. Los autores realizaron un estudio sobre el impacto que tiene el método, llamado PSO+RST, empleando la relación de similaridad y los pesos de los rasgos para mejorar el desempeño de algunos métodos de aprendizaje, como son: método de los k-Vecinos más Cercanos y un Multilayer Perceptron.

En (Filiberto, Bello, Caballero, & Frias, 2013) se presenta una revisión de diferentes métodos para el aprendizaje automático que se han desarrollado usando la Medida de Calidad de la Similaridad, entre ellas los ya mencionados: *Multi-Layer Perceptron* (Filiberto et al., 2011), y el método k-NN (Filiberto, Bello, Caballero, & Larrua, 2010a) y (Filiberto, Bello, Caballero, & Larrua, 2010b). Además la medida se ha usado en la construcción de prototipos (Bello et al., 2013), y en el descubrimiento de reglas de clasificación (Filiberto, Bello, Caballero, & Frías, 2011).

No obstante la Medida de Calidad de la Similaridad presenta la limitante de usar umbrales en la construcción de las relaciones de similaridad entre los objetos del sistema de decisión. Estos umbrales constituyen parámetros a ajustar del método, y los parámetros son agravantes cuando se analiza cualquier algoritmo. La precisión del método se hace muy sensible a pequeñas variaciones en los umbrales, además estos valores son dependientes de la aplicación, por lo que se hace necesario un proceso de ajuste de los umbrales para maximizar el rendimiento del proceso de descubrimiento de conocimiento (Fernández et al., 2014). El empleo de los umbrales provoca que el PSO converja hacia óptimos locales afectando la estabilidad de este algoritmo. Por todo lo anteriormente planteado se hace necesario incorporar alguna técnica que permita manejar la imprecisión.

La Lógica Difusa es una de las herramientas para hacer los sistemas informáticos capaces de resolver problemas que involucran la imprecisión. En 1965, L. A. Zadeh introdujo el concepto de la Teoría de Conjuntos Difusos. La Teoría de Conjuntos Difusos o Borrosos es una extensión de la Teoría de Conjuntos Clásica (Zadeh, 1971). Una lógica que no es muy precisa se llama una Lógica Borrosa. La forma imprecisa de ver las cosas y manipularlas es mucho más poderosa que la forma precisa de mirarlas y manipularlas (Hussain, 2010).

Una relación es una descripción matemática de una situación en la que ciertos elementos de los conjuntos están relacionados entre sí de alguna manera. Las relaciones borrosas son conceptos importantes en la Teoría de los Conjuntos Borrosos y han sido ampliamente utilizados en muchos campos tales como el agrupamiento borroso, el control difuso y razonamiento bajo incertidumbre. También juegan un papel importante diagnóstico difuso y modelado difuso.

El tema de las relaciones difusas fue analizado por Zadeh en (Zadeh, 1971), en el que se propone un marco conceptual unificado para el tratamiento de las relaciones. Al permitir grados intermedios de relación, las relaciones difusas proporcionan mucha más libertad para expresar las preferencias humanas. Las Relaciones difusas generalizan el concepto de conjuntos difusos a universos multidimensionales e introducen la noción de grado de asociación entre los elementos de un mismo universo (Bodenhofer, 2000). Las Relaciones difusas generalizan el concepto de las relaciones de la misma forma que los conjuntos borrosos generalizan la idea fundamental de conjuntos.

Una relación borrosa binaria R entre dos conjuntos, X e Y, se define como un subconjunto de $X \times Y$. Denotado por $R(X, Y)$, esta relación se asocia con una función de pertenencia $\mu R(X,Y)$ que pertenece a {0; 1} para todos los (x, y) en $X \times Y$. Es decir, $\mu R(X,Y) = 1$ si $(x, y) \in R(X,Y)$ y $\mu R(X, Y) = 0$ si $(x, y) \notin R(X, Y)$. Zadeh define una relación borrosa R entre X e Y como un subconjunto borroso de $X \times Y$ y por una extensión permite que $\mu R(X,Y)$ asuma valores en el intervalo [0; 1]. El valor de $\mu R(X,Y)$ representa la fuerza de la relación entre x e y. (Yang & Shih, 2001)

La Teoría de los Conjuntos Borrosos como uno de los elementos centrales de la *softcomputing*, usa las relaciones borrosas para hacer métodos computacionales más tolerantes y flexibles a la imprecisión, especialmente en el caso de datos mezclados. (Verdegay, Yager, & Bonissone, 2008).

Tomando en consideración los estudios realizados y el análisis de las reflexiones anteriores, se plantea como **Problema Científico**: ¿Cómo flexibilizar la construcción de las relaciones de similaridad en la Medida de Calidad de la Similaridad?

Objetivo General: Mejorar la tolerancia y flexibilidad de la Medida de Calidad de la Similaridad utilizando la Teoría de los Conjuntos Borrosos.

Para lograr este objetivo se plantean los siguientes **Objetivos Específicos** de la investigación:

1. Diseñar un método para la construcción de relaciones de similaridad usando relaciones borrosas binarias a partir de la cual se pueda flexibilizar la la Medida de Calidad de la Similaridad.
2. Redefinir el método PSO+RST a partir de la combinación de la metaheurística PSO, la la Medida de Calidad de la Similaridad y las relaciones borrosas.
3. Mostrar la efectividad de los pesos que se calculan usando el nuevo método propuesto al aumentar la eficacia de dos técnicas de aprendizaje automático, el k-Vecinos más Cercanos y *Multilayer Perceptron*, que tienen entre sus parámetros los pesos de los rasgos.
4. Validar el nuevo método para la resolución de problemas sobre la Ingeniería Civil y el Rendimiento Académico.

Objeto de estudio: Medida de Calidad de la Similaridad y Teoría de los Conjuntos Borrosos.

Campo de acción: Construcción de relaciones de similaridad borrosa.

Hipótesis: Con la sustitución de relaciones de semejanza por relaciones borrosas binarias se logrará mayor flexibilidad de la Medida de Calidad de la Similaridad.

De los objetivos anteriores la tesis se estructura en tres capítulos. El Capítulo 1 explica detalladamente la Medida de Calidad de la Similaridad y el método de cálculo de pesos PSO+RST. Se hace énfasis en el problema del uso de umbrales en la construcción de las relaciones de similaridad. Además se describen los aspectos fundamentales de la Teoría de los Conjuntos Borrosos. También se realiza un análisis crítico referido al problema del pesado de rasgos en los métodos de aprendizajes k-Vecinos más Cercanos y *Multilayer Perceptron*.

En el Capítulo 2 se propone una nueva medida para construir relaciones de similaridad basada en la Teoría de los Conjuntos Borrosos y se redefine el algoritmo para el cálculo de pesos de los rasgos para problemas con datos mezclados, a partir de la combinación de la metaheurística PSO y la nueva medida propuesta. Se estudia el desempeño del método k-NN y MLP al emplear los pesos calculados por la nueva medida propuesta. Los resultados de la nueva medida son validados en este capítulo usando conjuntos de

datos internacionales y se realiza un análisis estadístico de los resultados. Por último en el Capítulo 3 se validan estadísticamente los resultados teóricos de la investigación en problemas reales provenientes de las áreas de la Docencia y la Ingeniería Civil. El documento culmina con las conclusiones, recomendaciones, referencias bibliográficas y los anexos.

Capítulo I. Fundamentos teóricos

En este capítulo se presenta un estudio acerca de los fundamentos teóricos del análisis inteligente de los datos y el aprendizaje automático. Luego se presentan los conceptos fundamentales de la Teoría de los Conjuntos Aproximados, y las medidas de inferencia que nos brinda esta teoría para el análisis inteligente de los datos. También se presentan los elementos fundamentales de la Teoría de los Conjuntos Borroso y su empleo para el manejo de información imprecisa, así como un análisis crítico de la Medida de Calidad de la Similaridad.

1.1. Análisis Inteligente de los Datos y Aprendizaje Automático

Desde hace más de dos décadas se vienen desarrollando y utilizando complejos algoritmos para la extracción de patrones útiles en grandes conjuntos de datos. Gran parte de esta información representa transacciones o situaciones que se han producido, siendo útil no sólo para explicar el pasado, sino para entender el presente y predecir la información futura. En muchas ocasiones, el método tradicional de convertir los datos en conocimiento consiste en un análisis e interpretación realizada de forma manual por especialistas en la materia estudiada. Esta forma de actuar es lenta, cara y altamente subjetiva. De hecho, la enorme cantidad de datos desborda la capacidad humana de comprenderlos y el análisis manual hace que las decisiones se tomen según la intuición de los especialistas.

A finales de la década de los 80, la creciente necesidad de automatizar todo este proceso inductivo abrió una línea de investigación para el análisis inteligente de datos. Al conjunto de métodos matemáticos y técnicas software para el análisis de datos y búsqueda de regularidades y tendencias, aplicados de forma iterativa e interactiva, se denomina técnicas de Minería de Datos o *Data Mining* (DM). Su tarea fundamental es la de encontrar modelos inteligibles a partir de los datos, y para que el proceso sea efectivo debe ser automático, generando patrones que ayuden en la toma de decisiones beneficiosas para la organización.

Para obtener conclusiones válidas y útiles al aplicar minería de datos, es necesario complementar este proceso con una adecuada preparación de los datos previa al proceso de minería y un análisis posterior de los resultados obtenidos. Se puede afirmar que el

proceso de minería de datos pertenece a un esquema más amplio, denominado extracción o descubrimiento de conocimiento en bases de datos (*Knowledge Discovery in Databases*, KDD). Como se plantea en (Sánchez, 2006) se puede intuir que el KDD es un campo multidisciplinario, donde las principales áreas contribuyentes son el Aprendizaje Automático, las Bases de datos y la Estadística.

El Aprendizaje Automático es el área de la IA que se ocupa de desarrollar técnicas capaces de aprender, es decir, extraer de forma automática conocimiento subyacente en la información. El aprendizaje automático junto con la estadística constituyen el corazón del Análisis Inteligente de los Datos (AID) (Sánchez, 2006).

El análisis inteligente de los datos surgió al igual que KDD dado el problema de la explosión de datos y la necesidad de analizarlos con inteligencia. En (Zeger and Liang, 1986) se plantea que el objetivo del AID no es crear más algoritmos de análisis de datos, sino para dar sentido a los datos, casi todos los cuales se generan por expertos en sus dominios.

A medida que más y más datos se recogen y se almacenan, diversas técnicas de análisis de datos se han desarrollado sobre la base de los trabajos de reconocimiento de patrones, estadística, la inteligencia artificial, aprendizaje automático, sistema de base de datos, internet/intranet y otros. Una tarea del AID incluye el conocimiento, descubrimiento, predicción y el proceso y/o modelado de sistemas.

En los sistemas basados en el conocimiento existen muchas experiencias en la aplicación de métodos del AID en diversas áreas como el marketing, medicina, economía y la agricultura. Herramientas y aplicaciones de AID han generado resultados positivos y han estimulado continuamente la exploración de nuevas áreas de aplicación, debido a los beneficios aportados por esta tecnología (Chandrasekaran, 1986).

El volumen y la rápida expansión de datos en tiempo real, como resultado de la explosión de la actividad de la web, elementos multimedia, dispositivos electrónicos, comercio y otros, ha contribuido a la demanda y provisión de métodos más sofisticados del AID (Peralta, 2004). En general la idea de analizar las grandes cantidades de datos con una buena descripción es a la vez atractiva e intuitiva, pero técnicamente es significativamente desafiante y difícil. Debe haber algunas estrategias que deben

aplicarse para un mejor uso de los datos recogidos de tales fuentes de grandes volúmenes de datos y complejos.

Un proceso típico de AID comienza con la identificación del problema en función del interés de un analista de datos. A continuación, todas las fuentes de información se identifican y un subconjunto de datos es generado a partir de los datos acumulados para la aplicación del AID (R. Wang & Strong, 1996).

Para garantizar la calidad, el conjunto de datos es pre-procesado para la eliminación de ruido, el tratamiento de la información ausente y la transformación a un formato adecuado. Luego se aplica entonces al conjunto de datos derivados una técnica de AID o una combinación de técnicas apropiadas para el tipo de conocimiento.

El conocimiento descubierto es entonces evaluado e interpretado, por lo general empleando algunas herramientas de post-procesamiento y técnicas de visualización para presentar la información al usuario. Por lo general, el conocimiento adquirido es un tipo de reglas de clasificación, reglas de asociación, relaciones funcionales, dependencias funcionales, reglas causales y/o *clusters*.

De acuerdo con los objetivos e intereses de un usuario final, tales como caracterizar el contenido del conjunto de datos en su conjunto o establecer vínculos entre los subgrupos de patrones en el conjunto de datos, un análisis de datos proceso puede tener tres tareas posibles: modelado predictivo, análisis de agrupación y asociación (Amant and Cohen, 1998).

Las técnicas de aprendizaje automático son uno de los más utilizados para el análisis de datos debido a su capacidad para identificar patrones a partir de los conceptos que se les enseñe.

Una fuente importante de dificultades para los métodos de análisis de datos es la calidad de los datos. Los datos pueden contener ruido, información incompleta y datos redundantes e inútiles. Datos ruidosos, corruptos e incompletos pueden desviar la búsqueda y hacer un análisis incorrecto aún con el método más preciso.

Los métodos de análisis de datos deben proporcionar un mecanismo adecuado para encontrar resultados precisos de datos ruidosos. También deben permitir la selección de los datos relevantes, y el aprendizaje con conocimiento incompleto.

En este sentido las técnicas de IA resultan ventajosas en el manejo de datos inexactos. En particular la Teoría de los Conjuntos Aproximados constituye una herramienta útil para el análisis de datos inteligente en datos que involucran imprecisión. La Teoría de los Conjuntos Aproximados es de vital importancia para la IA, las ciencias cognitivas, especialmente en las áreas de aprendizaje automático, la adquisición de conocimientos, análisis de decisiones, el descubrimiento de bases de datos, sistemas expertos, razonamiento inductivo y reconocimiento de patrones.

La principal ventaja de la Teoría de los Conjuntos Aproximados es el análisis de datos en bruto y que no necesita ninguna información preliminar o adicional acerca de la probabilidad de datos -como en estadísticas (Pawlak, 1984). Esta teoría se abordará en profundidad en el siguiente epígrafe.

1.2. Teoría de los Conjuntos Aproximados

La Teoría de los Conjuntos Aproximados fue propuesta por Pawlak en 1982. La filosofía de los conjuntos aproximados se basa en la suposición que cada objeto x del universo de discurso U tiene asociada cierta información (datos y conocimiento), expresados por medio de atributos que describen el objeto (Komorowski and Pawlak, 1999b). Entre las ventajas de la RST para el análisis de datos están: se basa en los datos originales y no necesita información externa, por lo que no es necesario realizar ninguna suposición acerca de los datos; y permite el análisis tanto de rasgos cualitativos como cuantitativos (Y. Filiberto, Bello, Caballero, & Larrua, 2011).

Principales definiciones de la Teoría de los Conjuntos Aproximados.

Definición 1.1. Sistema de Información y Sistema de Decisión

Sea un conjunto de atributos $A=\{a_1, a_2,\ldots,a_n\}$ y un conjunto U no vacío llamado universo de discurso o ejemplos (objetos, entidades, situaciones o estados) descritos usando los atributos en A; se denomina Sistema de Información al par (U,A) (Komorowski and Pawlak, 1999b). Si a cada elemento de U se le agrega un nuevo atributo d llamado atributo de decisión, indicando la decisión tomada en ese estado o situación, entonces se obtiene un Sistema de decisión $(U, A \cup \{d\}), donde\ d \in A$.

Definición 1.2. Función de Información

Se llama Función de Información a la función f: Ux$A \rightarrow V$, $V=\{v_1,v_2,\ldots,v_n\}$ tal que $f(x,a_i) \in v_j$ para cada $a_i \in A$, $x \in U$, donde v_i es el dominio asociado a cada atributo a_i (Komorowski and Pawlak, 1999b).

El atributo de decisión d induce una partición del universo U de objetos. Sean los conjuntos $Y_i=\{x \in U: x(d)=i\}$, entonces $\{Y_1,\ldots,Y_b\}$ es una colección de clases de equivalencias, llamadas clases de decisión, donde dos objetos pertenecen a la misma clase si ellos tienen el mismo valor para el atributo decisión.

Se dice que un atributo $a_i \in A$ separa o distingue un objeto x de otro y, y se escribe *Separa* (a_i, x, y), si y solo si se cumple:

$$f(x,\alpha_i) \neq f(y,\alpha_i) \tag{1.1}$$

Aunque de forma simplificada se define de la siguiente forma:

$$x(i) \neq y(i) \tag{1.2}$$

Donde $x(i)$ denota el valor del rasgo i en el objeto x.

La relación de separabilidad se basa en la comparación de los valores de un atributo, para lo cual se ha usado la igualdad (o desigualdad) estricta. Sin embargo, es posible usar una condición de comparación menos estricta definida de esta forma:

$$Separa(\alpha_i,x,y) \Leftrightarrow |x(i)-y(i)| > E \tag{1.3}$$

Definición 1.3. Relación de Inseparabilidad

A cada subconjunto de atributos B de A, $B \subseteq A$, está asociada una Relación Binaria de Inseparabilidad denotada por R, la cual es el conjunto de pares de objetos que son inseparables uno de otro por esa relación (Komorowski and Pawlak, 1999b). La relación de inseparabilidad binaria es definida en la expresión 1.4.

Una Relación de Inseparabilidad definida a partir de formar subconjuntos de elementos de U que tienen igual valor para un subconjunto de atributos B de A, $B \subseteq A$, es una Relación de Equivalencia.

$$R = \{(x,y) \in U \times U : x(i) = y(i) \forall \alpha_i \in B\} \tag{1.4}$$

Los conceptos básicos de RST son los conceptos de aproximación inferior y superior. Una definición clásica de aproximación inferior y superior fue originalmente introducida con referencia a una relación de inseparabilidad, la cual es una relación de equivalencia.

Sea $\mathrm{DS} = (\mathrm{U}, \mathrm{A} \cup \{\mathrm{d}\})$ un sistema de decisión, $B \subseteq A$ y $X \subseteq U$, B define una relación de equivalencia y el subconjunto X es un concepto en el universo U. El subconjunto X se

puede aproximar usando sólo la información contenida en *B* mediante la construcción de la aproximaciones *B*-inferior y *B*-superior, denotadas por B_*X y B^*X respectivamente, y definidas por las expresiones 1.5 y 1.6 respectivamente.

$$B^*X = \{x \in U: [x]_B \cap X \neq \emptyset\} \quad (1.5)$$

$$B_*X = \{x \in U: [x]_B \subseteq X\} \quad (1.6)$$

Donde $[x]_B$ denota la clase de *x* de acuerdo a la relación de inseparabilidad *B*. Los objetos en B_*X son con certeza miembros de *X*, mientras los objetos en B^*X son posiblemente miembros de *X*.

Medidas de inferencia clásicas de la Teoría de los Conjuntos Aproximados

La Teoría de los Conjuntos Aproximados ofrece algunas medidas para analizar los sistemas de información (Arco & Bello, 2006; Y. Caballero, 2007; Yaile Caballero, 2007; Caballero & Bello, 2006; Filiberto, 2012). A continuación se muestran las principales medidas de inferencia.

Precisión de la aproximación

Un conjunto aproximado *X* puede ser caracterizado numéricamente por el coeficiente llamado precisión de la aproximación, donde |*X*| denota la cardinalidad de *X , X* ≠ ∅.

$$\alpha(X) = \frac{R_*'(X)}{R'^*(X)} \quad (1.7)$$

Obviamente, 0 ≤α(X) ≤1. Si *α(X)=1*, *X* es duro (exacto), si *α(X) <1*, *X* es aproximado (vago, inexacto), siempre respecto al conjunto de atributos considerado (Skowron, 1993).

Calidad de la aproximación

El coeficiente de calidad de la aproximación expresa la proporción de objetos que pueden ser correctamente clasificados en la clase *X*.

$$\gamma(X) = \frac{R_*'(X)}{|X|} \quad (1.8)$$

Además, *0 ≤α(X) ≤ γ(X) ≤1*, y *γ(X) =0* si y solo si *α(X)=0*, mientras *γ(X)=1* si y solo si *α(X)=1* (Skowron 1999).

Función de pertenencia aproximada

La función de pertenencia aproximada cuantifica el grado de solapamiento relativo entre *R'(x)* (clase de similitud de x) y la clase a la cual el objeto *x* pertenece. Se define por la expresión 1.9.

$$\mu_x(x) = \frac{|X \cap R'(x)|}{|R'(x)|} \quad (1.9)$$

La función de pertenencia aproximada también puede ser interpretada como la probabilidad condicional de que el objeto *x* pertenezca al conjunto *X*.

Calidad de la clasificación

La inexactitud de las clasificaciones aproximadas es descrita por el coeficiente de calidad de la clasificación, dado por la expresión 1.10.

$$\Gamma(DS) = \frac{\sum_{i=1}^{l} |R'_*(X_i)|}{|U|} \quad (1.10)$$

Este coeficiente expresa la proporción de objetos que pueden estar correctamente clasificados en el sistema. Si el coeficiente es igual a 1, entonces el sistema de decisión es consistente, en otro caso es inconsistente (Skowron, 1993).

Esta última medida permite determinar la consistencia de un sistema de decisión y funciona como una métrica para valorar la calidad de un conjunto de datos. Su principal limitante es que solo se usa para sistemas de decisión donde el dominio de los rasgos es discreto.

1.2.1. Teoría de los Conjuntos Aproximados Extendida

Las definiciones previas de conjuntos aproximados se basan en una relación *R* que define como inseparables a pares de objetos que tienen igual valor para un subconjunto de rasgos *B*, es decir, los objetos *(x,y)* son inseparables según *B* si $\alpha_i(x) = \alpha_i(y)$, para todo *i* en *B*, donde $\alpha_i(x)$ denota el valor del rasgo *i* en el objeto *x*. Definida de esta forma, *R* es una relación de equivalencia.

En ocasiones cuando el dominio de los rasgos en *B* no es discreto, una relación de inseparabilidad definida de esta forma no es aplicable. La relación de equivalencia es muy estricta en caso de dominios continuos pues ligeras diferencia entre los valores de los objetos para un rasgo pueden no ser significativas al analizar su inseparabilidad; por ejemplo, una temperatura de 37,8 grados puede ser considerada igual a otra de 37,9

grados. Esto es especialmente importante en el caso de rasgos numéricos en los cuales pequeños errores en la medición de los mismos pueden generar estas diferencias.

En estos casos es necesario emplear otros tipos de relaciones de inseparabilidad entre los objetos del universo *U*. Para ello se define una relación de semejanza en lugar de una relación de equivalencia, remplazando la relación de equivalencia por una relación binaria más débil. De esta forma se obtiene una extensión del enfoque clásico de la RST. Esta extensión se logra a través el concepto de inseparabilidad entre objetos de modo que se agrupen en la misma clase los objetos similares, no idénticos, según una relación de semejanza *R* (Filiberto et al., 2011b). Las relaciones de semejanza no inducen una partición del universo *U*, sino generan clases de semejanza para cualquier objeto $x \in U$.

La clase de semejanza de *x*, de acuerdo a la relación de semejanza *R* se denota por *R(x)* y se define como $R(x) = \{y \in U: y\ R\ X\}$. Esta se lee como "el conjunto de objetos del universo *U* que son similares al objeto *x* de acuerdo a la relación *R*". Un ejemplo es el caso de las relaciones de tolerancia (*tolerance relation*), donde la relación $R \subseteq XxU$ es reflexiva *(xRx)* para cualquier $x \in U$ y simétrica ($xRy \Rightarrow yRx$) para cualquier par $x, y \in U$.

Mientras que las relaciones de equivalencias inducen una partición del universo, las relaciones de semejanza inducen un cubrimiento del universo. El cubrimiento del universo es una familia de subconjuntos no vacíos cuya unión es el propio universo. Una partición de *U* a su vez es un cubrimiento, de modo que el concepto de cubrimiento es una extensión del concepto de partición.

Distintos investigadores han desarrollado estudios sobre la RST en el caso de espacios de aproximación que son cubrimientos, algunos de ellos orientados a construir una generalización de la RST basada en cubrimientos. En estos casos es importante encontrar una adecuada relación de semejanza para el sistema de decisión dado.

Resumiendo el orden actuación para trabajar con la RST. En el caso de datos continuos se puede elegir entre dos alternativas, discretizar los datos o usar una extensión de la teoría. En el primer caso, el sistema de información original se transforma en otro donde es aplicable el enfoque clásico. En la otra alternativa se emplea una relación de semejanza, definida como, $xRy \leftrightarrow F(x, y) \geq \mathcal{E}, x, y \in U$, $F(x, y)$ es una función de semejanza, y $\mathcal{E}$ es un umbral (Filiberto et al., 2011b). Estos umbrales son limitantes cuando se usa la relación de semejanza en algún método, como ejemplo tenemos la

Medida de Calidad de la Similaridad que se abordará en el siguiente epígrafe. Eliminar estos umbrales de la relación de semejanza sin afectar la precisión de los métodos derivados es el objetivo principal de la presente investigación.

1.3. Medida de Calidad de la Similaridad

En (Filiberto et al., 2010a) se propone una nueva medida para el caso de sistemas de decisión en el que el dominio de los rasgos, incluido el rasgo de decisión, no tiene que ser necesariamente discreto. Esta medida nombrada Medida de Calidad de la Similaridad representa el grado de similaridad entre los objetos de un sistema de decisión heterogéneo y constituye una herramienta importante para el análisis inteligente de los datos.

Sea un sistema de decisión DS donde *A* denota el conjunto de rasgos de condición y *d* el rasgo de decisión, y el dominio de los rasgos en *A, {d}* pueden ser valores discretos o continuos. Sean las relaciones de similaridad *R1* y *R2* entre los objetos, definidas por las expresiones 1.11 y 1.12 para todo par de objetos *(x, y)* en *U x U*.

$$xR_1y \text{ si y solo si } F_1(x,y) \geq \mathcal{E}_1 \quad (1.11)$$

$$xR_2y \text{ si y solo si } F_2(x,y) \geq \mathcal{E}_2 \quad (1.12)$$

Donde *F1* y *F2* son funciones de semejanza para calcular el grado de similaridad entre pares de objetos en el universo *U*; *F1* considera los rasgos en *A* y *F2* calcula la similaridad entre dos objetos de acuerdo al valor del rasgo *d*; $\mathcal{E}_1$ y $\mathcal{E}_2$ son umbrales de semejanza que toman valores entre 0 y 1.

Usando las relaciones *R1* y R2 se definen los conjuntos de objetos *N1* y *N2* para cualquier objeto *x* en el universo *U* según las expresiones 1.13 y 1.14, donde *N1(x)* y *N2(x)* son los conjuntos de objetos similares a *x* de acuerdo a las relaciones *R1* y *R2* respectivamente:

$$N_1(x) = \{y \in U : xR_1y\} \quad (1.13)$$

$$N_2(x) = \{y \in U : xR_2y\} \quad (1.14)$$

El problema de encontrar una relación de semejanza adecuada para la RST extendida consiste en encontrar las relaciones *R1* y *R2* que logran la mayor semejanza posible entre los conjuntos *N1* y *N2*, denotada *N1(x)~N2(x)*, donde el símbolo ~ representa la mayor similaridad posible entre los conjuntos *N1(x)* y *N2(x)* dado los umbrales $\mathcal{E}_1$ y $\mathcal{E}_2$.

El grado de similaridad entre estos conjuntos se define por la expresión 1.15, la cual se basa en la medida o coeficiente de similaridad.

$$\varphi(x) = \frac{|N_1(x) \cap N_2(x)|}{0.5 * |N_1(x)| + 0.5 * |N_2(x)|} \quad (1.15)$$

A partir de la expresión 1.15 la Medida de Calidad de la Similaridad del sistema de decisión se define por la expresión 1.16.

$$\theta(DS) = \frac{\sum_{\forall x \in U} \varphi(x)}{|U|} \quad (1.16)$$

La medida $\theta(DS)$ (Filiberto et al., 2010a) representa el grado en el cual la similaridad entre los objetos, usando los rasgos en *A*, es equivalente a la similaridad que se obtiene de acuerdo al rasgo de decisión *d*. El problema es encontrar las relaciones *R1* y *R2* que maximizan la Medida de Calidad de la Similaridad, según la expresión 1.17.

$$max \rightarrow \frac{\sum_{\forall x \in U} \varphi(x)}{|U|} \quad (1.17)$$

En el caso de sistemas de decisión en los cuales el dominio del rasgo de decisión es discreto, como es el caso de los problemas de clasificación, la relación R_2 se define como $\mathbf{x}R_2\mathbf{y} \Leftrightarrow \mathbf{x}(d) = \mathbf{y}(d)$, donde *x(d)* es el valor del rasgo de decisión *d* para el objeto *x*.

La Medida de Calidad de la Similaridad se ha implementado con éxito en la construcción de prototipos (Bello et al., 2013), en el descubrimiento de reglas de clasificación (Filiberto, Bello, Caballero, & Frías, 2011) y en mejorar el comportamiento de técnicas de aprendizaje supervisado como el *Multi-Layer Perceptron* (Filiberto et al., 2011), y el método k-NN (Filiberto, Bello, Caballero, & Larrua, 2010a).

No obstante la Medida de Calidad de la Similaridad presenta la limitante de usar umbrales en la construcción de las relaciones de similaridad en las expresiones (1.11) y (1.12). Estos umbrales constituyen parámetros a ajustar del método, y los parámetros son agravantes cuando se analiza cualquier algoritmo. La precisión del método se hace muy sensible a pequeñas variaciones en los umbrales, además estos valores son dependientes de la aplicación, por lo que se hace necesario un proceso de ajuste de los umbrales para maximizar el rendimiento del proceso de descubrimiento de conocimiento (Fernández et al., 2014). Dadas las importantes aplicaciones de esta medida se decide

trabajar en la robustez y flexibilidad de esta medida por lo que se hace necesario incorporar alguna técnica que permita manejar la imprecisión. La Lógica Borrosa es una de las herramientas para hacer los sistemas informáticos capaces de resolver problemas que involucran la imprecisión. La Teoría de los Conjuntos Borrosos como uno de los elementos centrales de la *softcomputing*, usa las relaciones borrosas para hacer métodos computacionales más tolerantes y flexibles a la imprecisión, especialmente en el caso de datos mezclados. (Verdegay, Yager, & Bonissone, 2008). Los principales conceptos de la Teoría de los Conjuntos Borrosos se profundizarán en el siguiente epígrafe, así como las características de las relaciones borrosas binarias.

1.4. Teoría de los Conjuntos Borrosos

Los Conjuntos Difusos o Borrosos (CB) y la Lógica Difusa (LD) son potentes herramientas matemáticas para modelar sistemas con incertidumbre en la industria, la naturaleza y fenómenos cotidianos. Son facilitadores para el razonamiento de sentido común en la toma de decisiones en ausencia de información completa y precisa.

Sus funciones son significativas cuando se aplica a fenómenos complejos que no se describen fácilmente por métodos matemáticos tradicionales, especialmente cuando el objetivo es encontrar una buena solución aproximada. Un enfoque racional hacia la toma de decisiones debe tener en cuenta la subjetividad humana, en lugar de solo emplear medidas de probabilidad objetiva (Helvacioglu and Ozen, 2014).

Las CB y la LD han sido investigadas ampliamente derivando en exitosas aplicaciones, así los abordan diversos autores: (Bodenhofer, 2000, Degang and Deli, 2014, Degang et al., 2005, Helvacioglu and Ozen, 2014, Hussain, 2010, Ignjatovic et al., 2015, Kawahara et al., 1999, Li and Cui, 2015, Mitra and Pal, 2005, Ovchinnikov, 2002, Pedrycz, 1997, Qian et al., 2015, Ramakrishnan and Rao, 1992, RoyChowdhury and Shukla, 2003, Verdegay et al., 2008, Wang et al., 2013, Wang and Xue, 2014, Yang and Shih, 2001, Yao, 1998, Zadeh, 1971, Zeng et al., 2015, Zhao and Hu, 2015).

La Teoría de los Conjuntos Borrosos, permite representar de forma más adecuada el conocimiento humano, ya que permite que los fenómenos y observaciones tengan más de dos estados lógicos, como lo son en la lógica clásica que sólo tiene verdadero o falso,

o bien 0 ó 1. Una de las aplicaciones de la Lógica Borrosa es cuando se quiere construir Conjuntos Borrosos para ser usados en sistemas inteligentes (Ramírez & Laguna, 2012).
En un conjunto clásico se asigna el valor 0 ó 1 a cada elemento para indicar la pertenencia o no a dicho conjunto. Esta función puede generalizarse de forma que los valores asignados a los elementos del conjunto caigan en un rango particular, y con ello indicar el grado de pertenencia de los elementos al conjunto en cuestión (Ramírez and Laguna, 2012). A dicha función se le denomina "función de pertenencia" y el conjunto por ella definida "Conjunto Borroso". Esta función de pertenencia es definida por μA para un Conjunto Borroso A, determinado por el rango de los números entre [0,1], definido de la siguiente manera: $\mu A = X \rightarrow [0.1]$.
A diferencia de un conjunto clásico, los elementos pertenecen o no a él totalmente, por ejemplo, un número puede pertenecer o no al conjunto de números pares, pero no pertenecerá con un determinado grado al conjunto, como se hace en los Conjuntos Borrosos en donde hay grados de pertenencia en referencia a un universo local.
La minería de datos a través de grandes volúmenes de datos revela información útil en forma de nuevas relaciones, patrones o clusters, para la toma de decisiones por parte de un usuario. Los CB apoyan una búsqueda centrada en los datos, especificada en términos lingüísticos. Esto ayuda a evitar la búsqueda de patrones sin sentido o triviales en una base de datos, y a descubrir dependencias o relaciones entre los datos en formato cualitativo o semicualitativo (Mitra & Pal, 2005).
La Teoría de los Conjuntos Borrosos ha sido de gran importancia en el ámbito del reconocimiento de patrones, permitiendo:

- La representación de entradas lingüísticas.
- La estimación o representación de la información perdida en términos de valores de pertenencia.
- La representación de una membresía multiclase de patrones ambiguos, en la generación de reglas y en inferencias con términos lingüísticos.
- La extracción de regiones de imágenes mal definidas y describiendo las relaciones entre ellos como subconjuntos borrosos (Mitra y Pal, 2005).

Operaciones sobre CB

Originalmente la Teoría de los Conjuntos Borrosos se formuló en base a un conjunto de operadores también válidos para conjuntos clásicos. En la Teoría de los Conjuntos Borrosos se definen tres operaciones básicas:

Negación: $$\mu\neg A(x) = 1 - \mu A(x) \quad (1.18)$$

Unión: $$\mu A \cup B(x) = max[\mu A(x), \mu B(x)] \quad (1.19)$$

Intersección: $$\mu A \cap B(x) = min[\mu A(x), \mu B(x)] \quad (1.20)$$

Funciones de Pertenencia:

Cada CB, tiene asociada una función de pertenencia para sus elementos. Las funciones de pertenencia más típicas son la triangular 1.21, trapezoidal 1.22, gaussiana 1.23 y la campana 1.24.

Triangular $$triangle(x; a, b, c) = \begin{cases} 0, x \leq \alpha \\ \dfrac{x-a}{b-a}, a \leq x \leq b \\ \dfrac{c-a}{c-b}, b \leq x \leq c \\ 0, c \leq x \end{cases} \quad (1.21)$$

Trapezoide $$trapezoid(x; a, b, c, d) = \begin{cases} 0, x \leq \alpha \\ \dfrac{x-a}{b-a}, a \leq x \leq b \\ 1, b \leq x \leq c \\ \dfrac{d-x}{d-c}, c \leq x \leq d \\ 0, d \leq x \end{cases} \quad (1.22)$$

Gaussiana $$gaussian(x; c, \sigma) = e^{-1/2\left(\frac{x-c}{\sigma}\right)^2} \quad (1.23)$$

Campana $$bell(x; a, b, c) = \frac{1}{1 + \left|\frac{x-c}{a}\right|^{2b}} \quad (1.24)$$

Conceptos sobre CB

Existe un principio que permite la generalización de conceptos matemáticos de Conjuntos Clásicos a la Teoría de Conjuntos Borrosos, denominado "Principio de extensión". Además de las operaciones descritas anteriormente, existen varios conceptos sobre subconjuntos difusos:

Núcleo: El núcleo de un subconjunto borroso *A*, es el conjunto de los elementos $x \in X$, que pertenecen totalmente a *A*, es decir que su función de pertenencia es igual a1. El núcleo de un subconjunto borroso está dado por la expresión 1.25.

$$núcleo(A) = \{x \in X | \mu_A(x) = 1\} \quad (1.25)$$

Soporte: El soporte de un subconjunto borroso *A*, es el conjunto de los $x \in X$, que pertenecen en cierta medida a *A*, es decir que su función de pertenencia es mayor que 0. El soporte de un subconjunto borroso está dado por la expresión 1.26.

$$soporte(A) = \{x \in X | \mu_A(x) > 0\} \quad (1.26)$$

Inclusión: Sean *A* y *B* dos partes difusas o borrosas del conjunto *C*. Se dice que *A* está incluido en *B* si para todo $x \in C$, tenemos $\mu_A(x) \leq \mu_B(x)$, es decir, que los elementos de *A* siempre pertenecen en mayor medida a *B* que a *A*.

Familia: Partiendo del subconjunto borroso *A*, se puede definir la familia de los conjuntos clásicos A_t con *t* variando en [0,1]. $A_t = (x \in B / \mu_A(x) \geq t)$, el conocimiento de esta familia define totalmente a *A*.

Similaridad entre Conjuntos Borrosos

La similaridad entre dos conjuntos depende del punto de vista de los diferentes autores, lo cual hace difícil decir cuál definición es la mejor entre todas las medidas de similitud para dos Conjuntos Borrosos. En (Wang, 1997) se proponen dos nuevas definiciones de medida de similitud entre dos Conjuntos Borrosos. Estas definiciones también se extienden a definir las medidas de similitud entre dos elementos en los Conjuntos Borrosos.

Dos medidas de similaridad entre dos Conjuntos Borrosos *A* y *B* son definidas en (Wang, 1997) por la expresión 1.27 y 1.28.

$$S(A,B) = \left\{ \sum_{i=1}^{n} \left[\frac{m(\mu_A(X_i), \mu_B(X_i))}{M(\mu_A(X_i), \mu_B(X_i))} \right] \right\} / n \quad (1.27)$$

Donde $m(a, b)$ y $M(a, b)$ denotan los valores mínimos y máximos de *a* y *b*, respectivamente. Para evitar que el denominador pueda ser cero, el autor asume 0/0 = 1 en el empleo de la expresión 1.27.

$$S(A,B) = \frac{\sum_{i=1}^{n} [1 - |\mu_A(X_i) - \mu_B(X_i)|]}{n} \quad (1.28)$$

Si la similitud $S(A,B)=1$, entonces se cumple que $\mu_A(X_i)=\mu_A(X_i)$, para toda $X_i \in X$, y se dice que $A=B$. Para cualquier dos Conjuntos Borrosos cualesquiera *A* y *B*, $A=B$ no significa que el conjunto *A* es exactamente el "mismo" que el conjunto *B*. Definir del término "mismo" o "igual" entre conjuntos difusos es muy difícil, pues la información

tiene cierto grado de incertidumbre (Wang, 1997). La definición de medidas de similitud que satisfacen la propiedad *S(A,B)=1*, se consideran medidas más fiables.

1.5. Relaciones Borrosas

En (Zadeh, 1971) una relación borrosa binaria *R* se define como una colección borrosa de pares ordenados. Por lo tanto, si *X={x}* e *Y={y}* son colecciones de objetos denotados genéricamente por *x* e *y*, una relación borrosa de *X* a *Y* o, equivalentemente, una relación borrosa en *X∪Y*, es un subconjunto borroso de *X×Y*, caracterizado por una función de membresía (característica) *μR*. Esta función asocia a cada par *(x, y)* su "grado de pertenencia", en *R*. Para simplificar, en este trabajo supondremos que el rango de *μR* es el intervalo [0, l] y se referirá al número *μR(x, y)* como la fuerza de la relación entre *x* e *y*.

Principales definiciones de las Relaciones Borrosas

Definición 1.4. Dominio de una relación borrosa

El dominio de una relación borrosa *R* es denotada por *domR* y es un conjunto borroso definido por la expresión 1.29.

$$\mu_{dom\,R}(x) = max_y \mu_R(x, y), x \in X \tag{1.29}$$

Donde el máximo es tomado sobre todos los $y \in Y$.

Definición 1.5. Rango de una relación borrosa

El rango de una relación borrosa *R* es denotado por *ranR* y está definido por la expresión 1.30.

$$\mu_{ran\,R}(y) = max_x \mu_R(x, y), x \in X, y \in Y \tag{1.30}$$

Definición 1.6. Tamaño de una relación borrosa

El tamaño de una relación borrosa *R* es denotado por *h(R)* y está definido por la expresión 1.31.

$$h(R) = max_x max_y \mu_R(x, y), x \in X, y \in Y \tag{1.31}$$

Una relación borrosa es subnormal si *h(R)<1* y normal si *h(R)=1*.

Definición 1.7. Soporte de una relación borrosa

El soporte de una relación borrosa *R* es denotado por *S(R)* y se define como el subconjunto no borroso sobre el cual *μ(x,y)>0*.

Definición 1.8. Contención de una relación borrosa

La contención de una relación borrosa R en una relación borrosa Q, es denotada por $R \subset Q$ y es definida por $\mu_R \leq \mu_Q$, lo cual significa que $\mu_R(x,y) \leq \mu_Q(x,y)$ para todo (x,y) en $X \times Y$.

Definición 1.9. Unión de relaciones borrosas

La unión de las relaciones borrosas R y Q es denotado por $R+Q$ en lugar de $R \cup Q$ y está definida por la expresión 1.32.

$$\mu_{R+Q}(x,y) = \max(\mu_R(x,y), \mu_Q(x,y)), x \in X, y \in Y \quad (1.32)$$

Definición 1.10. Intersección de relaciones borrosas

La intersección de las relaciones borrosas R y Q es denotado por $R \cap Q$ y está definida por la expresión 1.33.

$$\mu_{R\cap Q}(x,y) = \min(\mu_R(x,y), \mu_Q(x,y)), x \in X, y \in Y \quad (1.33)$$

Definición 1.11. Producto de relaciones borrosas

El producto de las relaciones borrosas R y Q es denotado por RQ y está definido por $\mu_{RQ} = \mu_R \mu_Q$. Note que, si R, Q y T (T es alguna relación borrosa desde X a Y) entonces se cumple la expresión 1.34.

$$R(Q+T) = RQ + RT \quad (1.34)$$

Definición 1.12. Complemento de una relación borrosa

El complemento de la relación borrosa R es denotado por R' y es definido por $\mu_{R'} = 1 - \mu_R$.

Definición 1.13. Relaciones de similaridad en una relación borrosa

Una relación de similaridad S en la relación borrosa X, es una relación borrosa en X la cual cumple las propiedades enunciadas por las expresiones 1.35, 1.36 y 1.37.

a. Reflexiva $\mu_S(x,y) = 1$ Para toda x en dominio S (1.35)

b. Simétrica $\mu_S(x,y) = \mu_S(y,x)$ Para toda x, y en dominio S (1.36)

c. Transitiva $\mu_S(x,z) \geq max_y(\mu_S(y,x) min\, \mu_S(y,z))$ (1.37)

1.6. La Teoría Fuzzy- Rought Set y la Teoría Rough-Fuzzy Set

La Teoría de los Conjuntos Aproximados y la Teoría de los Conjuntos Borrosos son complementarias. Es natural combinar los dos modelos de incertidumbre (la vaguedad de los Conjuntos Borrosos y la aproximación de los conjuntos aproximados) para conseguir una precisión más exacta trabajando con información imperfecta.

La Teoría de los Conjuntos Aproximados y la Teoría de los Conjuntos Borrosos crean muchas posibilidades para la hibridación. El número de resultados reportados en esta dirección están aumentando continuamente; algunos ejemplos los podemos encontrar en (Komorowski and Pawlak, 1999a, Li and Cui, 2015, Mitra et al., 1999, Nguyen, 2011, Qian et al., 2015, Riza et al., 2014, Sun et al., 2014, Wua et al., 2003, Yao, 1998, Zeng et al., 2015, Zhao and Hu, 2015, Komorowski et al., 1996, Pawlak, 1984).

Resulta interesante establecer una comparación entre las definiciones de conjuntos clásicos, conjuntos difusos y los conjuntos aproximados. Los conjuntos clásicos es una noción primitiva y se define de manera intuitiva. Los Conjuntos Borrosos se definen mediante el empleo de la función de pertenencia difusa, lo que implica estructuras matemáticas avanzadas, número y funciones. Los conjuntos aproximados se definen por aproximaciones que requieren también conceptos matemáticos avanzados.

Los métodos de los Conjuntos Aproximados proporcionan descripciones aproximadas de conceptos y pueden ser empleados para construir una descripción aproximada de los conceptos difusos también. Esto es muy importante para la representación más comprimida de conceptos, reglas y patrones en KDD porque el uso de conceptos difusos puede describir estos elementos de una manera más compacta. Estas descripciones, por otra parte, pueden ser más adecuadas para la comunicación con el ser humano (Komorowski and Pawlak, 1999b).

Para tratar con los datos numéricos y discretos, o una mezcla de ambos, en conjuntos de datos, el "modelo aproximado borroso" se propuso por primera vez por Dübois y Prad, combinando los conjuntos aproximados y difusos. Las aproximaciones inferiores y superiores en los "aproximado borroso" intentan asignar una función de pertenencia a cada objeto del conjunto de datos.

Como Dübois y Prade definieron, si un Conjunto Borroso se acerca a una familia de conjuntos aproximados (*crips*) en el mismo universo, entonces las correspondientes aproximaciones inferior y superior se llaman "conjuntos borrosos aproximados"; y si un Conjunto Aproximado se acerca a una familia de conjuntos borrosos en el mismo universo, entonces aproximaciones inferior y superior correspondiente se llaman "conjuntos aproximados borrosos" (Zeng et al., 2015).

La función de pertenencia aproximada donde x se pueden utilizar para definir aproximaciones y la región límite de un conjunto, como se muestra a continuación

$$B_*(X) = \{x \in U: \mu_X^B(x) = 1\} \quad (1.38)$$

$$B^*(X) = \{x \in U: \mu_X^B(x) \geq 0\} \quad (1.39)$$

$$BN_B(X) = \{x \in U: 0 < \mu_X^B(x) < 1\} \quad (1.40)$$

La función de pertenencia aproximada tiene las siguientes propiedades (Pawlak, 1984).

$$\mu_X^B(x) = 1 \; si \; solo \; si \; x \epsilon B_*(X) \quad (1.41)$$

$$\mu_X^B(x) = 0 \; si \; solo \; si \; x \epsilon \; U - B^*(X) \quad (1.42)$$

$$0 < \mu_X^B(x) < 1 \; si \; solo \; si \; x \epsilon BN_B(X) \quad (1.43)$$

$$\mu_{U-X}^B(x) = 1 - \mu_X^B(x) \; para \; x \in U \quad (1.44)$$

$$\mu_{X \cup Y}^B(x) \geq \max\left(\mu_X^B(x), \mu_Y^B(x)\right) para \; x \in U \quad (1.45)$$

$$\mu_{X \cap Y}^B(x) \geq \min\left(\mu_X^B(x), \mu_Y^B(x)\right) para \; x \in U \quad (1.46)$$

Las propiedades anteriores muestran claramente la diferencia entre la pertenencia difusa y aproximada. En las propiedades particulares 1.45 y 1.46 se muestra que el número de miembros aproximados puede ser considerado como una generalización de función de pertenencia difusa.

Se puede ver fácilmente que existe una estrecha conexión entre la vaguedad y la incertidumbre. Como se ha mencionado anteriormente la vaguedad se relaciona con conjuntos o conceptos, donde la incertidumbre se relaciona con elementos de conjunto. El enfoque de conjunto aproximados muestra conexión entre estos dos conceptos.

Combinando los conjuntos aproximados y conjuntos difusos es posible obtener aproximaciones de conjuntos difusos, así como aproximaciones de conjuntos por medio de relación de similitud difusa. Este segundo caso llamado conjuntos aproximados borrosos (Pawlak, 1984, Komorowski and Pawlak, 1999a). Estos se definen por su función en el universo de membrecía de objetos U por:

$$\mu_{S^*(x)}(x) = sup_{\omega \in X} \mu_S(x, \omega \quad (1.47)$$

$$\mu_{S_*(x)}(x) = inf_{\omega \notin X}(1 \perp \mu_S(x, \omega)) \quad (1.48)$$

Donde S es una relación de relación de similitud difusa y x∈U. En este caso tenemos en cuenta las relaciones de similitud difusa en lugar de las relaciones de indiscernibilidad. En caso de relación de equivalencia se obtiene:

$$\mu_{S^*(x)}(x) = 1 \; si \; solo \; si \; x \epsilon S^*(X) \quad (1.49)$$

$$\mu_{S_*(x)}(x) = 1 \; si \; solo \; si \; x \epsilon S_*(X) \quad (1.50)$$

Donde $S^*(X)$ y $S_*(X)$ denota las aproximaciones inferior y superior de X con respecto a la relación indiscernibilidad S (Pawlak, 1984).

Como se mencionó anteriormente la Teoría de los Conjuntos Aproximados y la Teoría de los Conjuntos Borrosos brindan posibilidades para la hibridación. En este trabajo se presenta la idea de utilizar una función de semejanza basada en la teoría de los conjuntos aproximados en su enfoque extendido como función de pertenencia para caracterizar una relación binaria difusa. Existen otras relaciones interesantes de conjuntos aproximados con conjuntos difusos (Dubois and Prade, 1987, Dubois and Prade, 1990, Dubois and Prade, 1998).

Conclusiones parciales

En este capítulo se realizó un estudio acerca de la Teoría de los Conjuntos Aproximados para el análisis inteligente de los datos y de la Teoría de los Conjuntos Borroso para manejar información imprecisa, exponiendo de forma resumida sus conceptos básicos así como sus principales ventajas. Se realizó un análisis crítico de la Medida de Calidad de la Similaridad, arrojando como resultado una limitante al contener parámetros que ajustar en la definición de las relaciones de similaridad.

A partir de este diagnóstico se comprobó la necesidad de incorporar alguna técnica que haga menos sensible el uso de la medida, que aumente la interpretación del sistema a la vez que mantenga la imprecisión. En esta tesis se propone una redefinición de la Medida de Calidad de la Similaridad mediante la sustitución de relaciones de semejanza por relaciones binarias borrosas, con el objetivo de lograr un método más tolerable e igual de preciso. Esta nueva medida es abordará en el capítulo 2.

Capítulo II. Relaciones Borrosas en la Medida de Calidad de la Similaridad

En este capítulo se plantea la nueva medida denominada Medida de la Calidad de la Similaridad Borrosa. La nueva medida se emplea para mejorar los resultados de dos métodos de aprendizaje automático: k-Vecinos más Cercanos y Redes Neuronales Artificiales. También se presenta un estudio sobre los métodos estadísticos utilizados en la comparación de una nueva propuesta con varios algoritmos. Por último se muestra la validación experimental de los resultados.

2.1. Medida de Calidad de la Similaridad Borrosa

Las relaciones binarias borrosas son caracterizadas por una función de pertenencia que indica en un intervalo del [0,1] la fortaleza de la relación entre un par de objetos. En la medida de Calidad de la Similaridad Borrosa, la función de pertenencia se define por la función de semejanza basada en la Teoría de los Conjuntos Aproximados Extendida entre dos objetos.

En las expresiones (2.1) y (2.2) se refleja la sustitución de las relaciones de semejanza por relaciones borrosas binarias. En las expresiones de (1.11) y (1.12) de la Medida de Calidad de la Similaridad las relaciones de semejanza se establecían en dependencia de si las funciones F_1 y F_2 eran mayor o no que determinado umbral. El resultado de estas relaciones de semejanza establecían si existía- sí o no- una relación de semejanza entre los objetos *x* e *y*. En la Medida de Calidad de la Similaridad Borrosa siempre va a existir una relación entre los objetos *x* e *y*, esta relación asume valores en el intervalo [0; 1] que representa la fuerza de la relación entre *x* e *y*.

Para todo par de objetos *(x, y)* en *UxU* se cumplen las expresiones 2.1 y 2.2.

$$xR_1y = F_1(x,y) \quad (2.1)$$

$$xR_2y = F_2(x,y) \quad (2.2)$$

Donde *R1* y *R2* son relaciones difusas para cuantificar el nivel de semejanza entre los objetos *x* e *y* respecto a los rasgos de condición y el rasgo de decisión respectivamente. Las relaciones borrosas *R1* y R2 quedan definidos por las funciones de semejanza *F1* y *F2*, las cuales están definidas por las expresiones 2.3 y 2.4.

$$F_1(x,y) = \sum_{i=1}^{n} w_i * \partial_i (X_i, Y_i) \quad (2.3)$$

$$F_2(x,y) = \partial (d(X), d(Y)) \quad (2.4)$$

Donde

$$\partial_i(X(i), Y(i)) = \begin{cases} 1 - \dfrac{|X(i) - Y(i)|}{Max(\alpha_i) - Min(\alpha_i)} & si\ i\ es\ continuo \\ 1 & si\ i\ es\ discreto\ y\ X(i) = Y(i) \\ 0 & si\ i\ es\ discreto\ y\ X(i) \neq Y(i) \end{cases} \quad (2.5)$$

A partir de las relaciones borrosas *R1* y *R2* se definen los Conjuntos Borrosos *N1(x)* y *N2(x)* por las expresiones 2.6 y 2.7 respectivamente.

$$N_1(x)=\{ (y,\mu R_1(x,y)), \text{ para todo } y \in U \} \quad (2.6)$$

$$N_2(x)=\{ (y,\mu R_2(x,y)), \text{ para todo } y \in U \} \quad (2.7)$$

El grado de similaridad entre ambos conjuntos para un objeto *x* se calcula como la semejanza entre los Conjuntos Borrosos *N1(x)* y *N2(y)*. Esta semejanza se calcula mediante la expresión 2.8.

$$\varphi(x) = \frac{\sum_{i=1}^{n}[1 - |\mu R_1(x_i) - \mu R_2(x_i)|]}{n} \quad (2.8)$$

Usando la expresión 2.8 la Medida de Calidad de la Similaridad de un DS con un universo de N objetos se define por la expresión 2.9. Estas modificaciones permiten disminuir la cantidad de parámetros a considerar y mantener la eficacia del método.

$$\theta(DS) = \left\{ \frac{\sum_{i=1}^{n} \varphi(x)}{n} \right\} \quad (2.9)$$

2.2. Método para cálculo de pesos para los rasgos basado en la medida de Medida de Calidad de la Similaridad Borrosa

A partir de la Medida de Calidad de la Similaridad se generó un método para el cálculo de pesos hibridando la meta heurística Particle Swarm Optimization (PSO) como método de búsqueda para encontrar el conjunto de pesos que maximicen la Medida de Calidad de la Similaridad. Con este nuevo método llamado PSO+RST se obtuvieron excelentes resultados lo cuales aparecen reflejados en (Filiberto et al., 2010b).

Particle Swarm Optimization

La inteligencia de enjambre (SI), inspirada en el comportamiento biológico de las aves, es una innovadora técnica de optimización inteligente. Las técnicas SI se basan en el comportamiento colectivo de enjambres de abejas, banco de peces, y colonias de insectos, que mientras buscan alimento, se comunican entre sí y socializan en sus colonias.

La optimización basada en partículas (*Particle Swarm Optimization* PSO) es un algoritmo de inteligencia de enjambre basada en la metaheurística propuesta por Kennedy y Eberhart (Kennedy & Eberhart, 1995).

En PSO, a cada individuo del enjambre, llamado partícula, se comporta como un agente dentro de un entorno altamente descentralizado e inteligente. Cada partícula del enjambre contribuye al ambiente siguiendo reglas muy sencillas, mediante la cooperación y la comunicación con otras partículas del enjambre. Una conducta de colectividad global compleja surge en el enjambre. Este comportamiento global complejo se explota para resolver un problema de optimización complejo. La alta descentralización, la cooperación entre las partículas y aplicación sencilla hacen de PSO una técnica eficaz aplicable a los problemas de optimización.

PSO tiene tres componentes principales, los componentes sociales y cognoscitivos de las partículas, y la velocidad de las partículas. Cada partícula tiene una medida de calidad, así como una posición y velocidad en el espacio de la búsqueda, donde la posición determina el contenido de la posible solución. En el espacio de un problema donde puede haber más de una posible solución y se requiere buscar la solución óptima del problema, una partícula va a representar una solución individual al problema.

El aprendizaje de las partículas proviene de dos fuentes, una fuente es la propia experiencia de la partícula, llamada aprendizaje cognitivo y la otra fuente es el aprendizaje combinado de todo el enjambre, llamado aprendizaje social. El aprendizaje cognitivo está representado por la mejor experiencia del individuo (*pBest*) y el aprendizaje social está representado por la mejor experiencia del enjambre, el valor global (*gBest*) (Alam, Dobbie, Koh, Riddle, & Rehman, 2014). La solución *pBest* es la mejor solución que la partícula ha logrado nunca en su historia. El valor *gBest* es la mejor posición que el enjambre ha logrado nunca. El enjambre guía la partícula con el

parámetro *gBest*. Juntos el aprendizaje cognitivo y social se utilizan para calcular la velocidad de las partículas a su próxima posición.

Cuando es aplicado a los problemas de optimización, un algoritmo PSO típico empieza con la inicialización de varios parámetros. Una de las inicializaciones más importantes es seleccionar el enjambre inicial. El número de partículas en el enjambre depende de la complejidad del problema. Una opción inicial de soluciones se hace normalmente aleatoria. Sin embargo una suposición inicial que se propaga a las partículas de manera uniforme en el espacio de solución puede acelerar la aparición de una solución óptima.

Un número inicial típico de partículas para PSO en un enjambre va de 20 a 40, pero varía de aplicación a aplicación y de problema a problema. Las partículas comienzan a moverse de una posición a otra en busca de una mejor solución sobre la base del componente social y el cognitivo. El *pBest* del componente cognoscitivo para los problemas de minimización es calculado mediante la expresión 2.10.

$$pBest_i(t+1)\begin{Bmatrix} pBest_i(t) & if\ (X_i(t+1)) \geq f\big(pBest_i(t)\big) \\ X_i(t+1) & if\ (X_i(t+1)) < f\big(pBest_i(t)\big) \end{Bmatrix} \tag{2.10}$$

donde *Xi (t + 1)* es la nueva posición de la partícula, $pBest_i(t)$ es la actual mejor marca personal, y $pBest_i(t+1)$ es la nueva mejor posición personal de la partícula. El valor de *gBest* viene del aprendizaje social del enjambre que muestra el mejor ajuste que cualquier partícula del enjambre tiene jamás alcanzada. La expresión 2.11 se emplea para calcular el *gBest* para el mismo problema de minimización.

$$gBest(t) = argMin_{i=1}^{n}\{f\big(pBest_i(t)\big)\} \tag{2.11}$$

Donde *n* es el número total de partículas. Juntos *pBest* y *gBest* se combinan para definir la velocidad de la partícula que guía a las partículas hacia una mejor solución (Eberhart & Shi, 2000). La velocidad de una partícula se calcula mediante la expresión 2.12.

$$V_i(t+1) = w \times V_i(t) + q_1 r_1\ (pBest_i(t) - X_i(t)) + q_2 r_2\ (gBest_i(t) - X_i(t)) \tag{2.12}$$

Donde *Vi(t)* representa la velocidad de la corriente de la partícula i, *Vi (t + 1)* representa la nueva velocidad de la partícula alcanzará para pasar de la posición actual a la nueva posición. La gama de velocidades está delimitada entre *Vmax* y *Vmin*, donde *Vmax* es la velocidad máxima y *Vmin* es la velocidad mínima. Los parámetros q_1 y q_2 son constantes que reprsentan el peso de los componentes sociales y cognitivos, r_1 y r_2 son

números aleatorios que van de 0 a 1, y ω es la inercia de la partícula. La velocidad añadida a la posición actual ofrece la nueva posición de las partículas, dada por la expresión 2.13.

$$X_i(t+1) = X_i(t) + V_i(t+1) \tag{2.13}$$

El *pBest* y *gBest* afectan el movimiento de partículas desde la posición *Xi (t)* a *Xi (t + 1)*. Hay muchos factores que afectan a la convergencia y el rendimiento del algoritmo PSO. El tamaño enjambre, el ajuste de la velocidad, la posición de la partícula, la topología de los vecinos, y las actualizaciones síncronas o asíncronas. Además, los valores de la masa de inercia y los coeficientes de aceleración, la parte cognitiva y la parte social del modelo, respectivamente, pueden afectar significativamente la eficiencia y la fiabilidad del PSO.

En el algoritmo PSO, un control adecuado de la exploración global y explotación local es un asunto crucial. Shi y Eberhart introdujeron el concepto de masa de inercia ω de la versión original de PSO para equilibrar las búsquedas locales y globales durante el proceso de la evolución (Shi and Eberhart, 1998).

En general, una gran masa de inercia impuesta en las etapas tempranas de búsqueda produce que el espacio de búsqueda sea explorado a fondo. Si este peso de inercia es disminuido gradualmente, se logran soluciones más refinadas en la fase de la búsqueda final. El objetivo principal de esta modificación es evitar la convergencia prematura en las etapas tempranas de búsqueda y mejorar la convergencia a la solución óptima a nivel global. El concepto de "linealmente decreciente del peso inercia" aplicado a la optimización de enjambre de partículas (LPSO) se introdujo en (Shi and Eberhart, 1998) y (Shi and Eberhart, 1999) y está dada por la expresión 2.14.

$$\omega = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{iter_{max}} iter \tag{2.14}$$

Dónde el ω_{max} y el ω_{min} son los valores iniciales y finales de la masa de inercia, respectivamente, *iter* es el número de iteración actual e $iter_{max}$ es el número máximo de iteraciones permitidas. Por lo general, los parámetros ω_{max} y ω_{min} se establecen en 0,9 y 0,4, respectivamente. Por lo tanto, las partículas usan un peso mayor inercia durante la exploración inicial y utilizan los valores de peso bajos para posteriores iteraciones.

PSO constituye una de las técnicas mas usadas en las investigaciones actuales por su fácil implementación y sus pocos parametros que son facilmente ajustables. La segmentación y agrupamiento de imagenes fue unas de las primeras áreas de aplicación del PSO. Algunos de los otros dominios de procesamiento de imágenes basados en PSO incluyen la detección de bordes, la recuperación de la imagen, la eliminación de ruido, selección de características de imagen y clasificación de imágenes.

Esta técnica también ha sido muy empleada en las redes del sensores. Su aplicación abarca la localización de redes de sensores inalámbricos, optimización y control, la cobertura y enrutamiento de la red y la agrupación de nodos (Zhang et al., 2014). La robótica es otra área importante donde PSO se ha empleado con éxito para la colaboración del robot y la comunicación, el aprendizaje del robot y la toma de decisiones.

Algunas de las áreas dentro del KDD donde PSO ha sido implementado son: la detección de ruido, algoritmos de clasificación y reglas de asociación, selección de características basadas enjambre de partículas, minería de texto y sistemas de recomendación (Zhang, Ma, Wei, & Liang, 2014). Lo anteriormente expuesto demuestra la efectividad de la metaheristica PSO y la justificación de la selección de esta técnica para hibridar en un metodo para el pesado de rasgos, como se abordará en la próxima sección.

Los parámetros utilizados en la experimentación de nuestra investigación, fueron: Número de partículas 40, Número de iteraciones 100, q_1=2 y q_2=2 son constantes que representan el peso de los componentes sociales y cognitivos. La condición de parada es: cuando se alcancen 100 iteraciones o cuando en cinco iteraciones no mejore el valor fitness (maximizar medida calidad de la similaridad 2.9).

PSO+RST+FUZZY

En (Filiberto et al., 2010b) se propuso un método llamado PSO+RST, el cual permite construir relaciones de similaridad a partir de funciones de semejanza definidas como una suma pesada de la comparación a nivel de rasgo, constituyendo un nuevo procedimiento de aprendizaje de pesos cuya aplicación no está limitada a problemas de clasificación. En este caso, el objetivo del algoritmo de aprendizaje es encontrar el

conjunto de pesos W= $\{w_1, w_2,\ldots,w_n\}$, que maximiza la Medida de Calidad de la Similaridad.

El nombre de PSO+RST estuvo determinado porque el empleo de PSO para encontrar el conjunto de pesos. En procesos de selección de rasgos, PSO junto a los conjuntos aproximados ya se ha estudiado por otros autores reportando muy buenos resultados (Filiberto et al., 2010b). Además ha demostrado tener un buen desempeño en los problemas de optimización continua.

Teniendo en cuenta que la Medida de Calidad de la Similaridad es redefinida con el uso de las relaciones borrosas, se rediseña el método para pesado de rasgos, en esta nueva versión, llamado PSO+RST+FUZZY. En el PSO cada individuo está constituido por *n* componentes en dependencia de la cantidad de rasgos del problema a tratarse. El PSO se emplea como procedimiento de generación de subconjuntos; y la Medida de Calidad de la Similaridad Borrosa es empleada como función de evaluación de los subconjuntos.

A continuación se describe el funcionamiento del algoritmo PSO+RST propuesto en (Filiberto et al., 2010b)

Paso 1: Inicializar una población de partículas con posiciones y velocidades aleatorias en un espacio *D*-dimensional.

Paso 2: Para cada partícula, evaluar la Medida de Calidad de la Similaridad Borrosa empleando la expresión 2.15, en *D* Variables.

$$max \rightarrow \left\{ \frac{\sum_{\forall x \in U} \varphi(x)}{|U|} \right\} \qquad (2.15)$$

Paso 3: Comparar la Medida de Calidad de la Similaridad actual de cada partícula con la Medida de Calidad de la Similaridad de su mejor posición anterior *pBest*. Si el valor actual es mejor que el de *pBest*, entonces asignarle a *pBest* el valor actual, y *Pi =Xi*, es decir, la localización actual pasa a ser la mejor hasta el momento.

Paso 4: Identificar la partícula en la vecindad con mayor valor para Medida de Calidad de la Similaridad y asignar su índice a la variable *g* y asignar el mejor valor de la Medida de Calidad de la Similaridad a *m*.

Paso 5: Ajustar la velocidad y posición de la partícula según las ecuaciones 2.16 y 2.17 (para cada dimensión)

$$v_i(t+1) = \alpha * v_i(t) + U(0,\varphi 1)\big(pbest(t) - x_i(t)\big) + U(0,\varphi 2)\big(gbest(t) - x_i(t)\big) \quad (2.16)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (2.17)$$

Paso 6: Verificar si se cumple el criterio de parada (máximo número de iteraciones o si lleva cinco iteraciones sin mejorar la Medida de Calidad de la Similaridad *global (m)*), si no, ir al Paso2.

La única diferencia entre PSO+RST y PSO+RST+FUZZY, está en el Paso 2 del algoritmo, donde en lugar evaluar cada partícula con la Medida de Calidad de la Similaridad (expresión 1.16) se utiliza la Medida de Calidad de la Similaridad Borrosa (expresión 2.9). La validación estadística del PSO+RST+FUZZY se abordará en los epígrafes 2.4 y 2.4.1.

2.3. Aplicación de la medida de Medida de Calidad de la Similaridad Borrosa en métodos del Aprendizaje Automático

A los efectos de demostrar la efectividad de los pesos que se calculan por PSO+RST+FUZZY para construir la relación de similaridad se muestran dos casos de aplicación, que son dos técnicas de Aprendizaje Automático: el k-Vecinos más Cercanos y el *Multilayer Perceptron*. Ambas técnicas tienen entre sus parámetros el peso de los rasgos. En el k-Vecinos más Cercanos un correcto proceso de asignación de pesos a los rasgos permite una recuperación más eficiente de los casos más similares mientras que en el *Multilayer Perceptron* puede acelerar la convergencia hacia un óptimo global.

2.3.1. Efecto del empleo de los pesos de los rasgos en el método k-vecinos más similares

Una forma práctica y de fácil aplicación para predecir o clasificar un nuevo dato, basado en observaciones conocidas o pasadas, es la técnica del vecino más cercano. (Mora et al., 2008). En este epígrafe se presenta un estudio acerca del empleo del PSO+RST+FUZZY para mejorar el desempeño del método de aprendizaje k-Vecinos más Cercanos.

Las técnicas de k-Vecinos más Cercanos basan su criterio de aprendizaje en la hipótesis de que los miembros de una población suelen compartir propiedades y características con los individuos que los rodean, de modo que es posible obtener información

descriptiva de un individuo mediante la observación de sus vecinos más cercanos (Chavez, 2008). La idea básica sobre la que se fundamenta es que un nuevo caso se va a clasificar en la clase más frecuente a la que pertenecen sus *k*-Vecinos más Cercanos.

Este método utiliza funciones de distancia o similitud para generar predicciones a partir de ejemplos almacenados. Este algoritmo pertenece a la clase de algoritmo perezoso (*Lazy Learning Algoritms*), pues almacena el conjunto de entrenamiento y deja todo el procesamiento para la fase de clasificación.

Entre sus ventajas se encuentra que es muy simple e intuitivo. Su estrategia de clasificación resulta muy fácil de comprender, y su sencillez hace que sea uno de los clasificadores más populares dentro del Reconocimiento de Patrones y la Inteligencia Artificial. Entre sus desventajas está que asume, en su versión original, que los objetos están representados en espacios numéricos, por lo que no admite datos mixtos ni ausencia de información. Se han desarrollado versiones, como la Regla del Vecino Más Similar, que eliminan esta restricción (Villuendas et al., 2010).

En algunas ciencias como la Medicina, Geociencias, Sociología, etc., los prototipos generalmente están descritos por atributos numéricos y no numéricos (datos mixtos). En estos casos, la función de comparación no siempre cumple propiedades métricas. Por esta razón, es importante desarrollar clasificadores rápidos basados en la búsqueda de los k vecinos más similares (k-MSN) para datos mezclados que usen funciones de comparación no métricas (Hernández et al., 2010).

La regla k vecinos más similares (k-MSN) aunque no tiene las restricciones del k-NN y permite el trabajo con datos mixtos e incompletos, emplea funciones de similaridad que no son opuestas o inversas a una distancia, lo que mantiene un elevado costo computacional. También en la literatura se han reportado trabajos donde utilizan clasificadores k-NN con funciones de analogía que no cumplen las propiedades de las distancias, llamándole a estas funciones de diferente forma, usualmente como "distancias no métricas".

La forma de comparar dos objetos debe ser extraída de la modelación del problema, según el concepto de similaridad que utiliza el especialista en el dominio. Sin embargo, en muchos trabajos donde se utilizan repositorios estándar de bases de datos, esto no es posible. En estos casos se escoge entre un amplio repertorio de funciones, que ya han

sido utilizadas con anterioridad, o se crean funciones propias. Es importante destacar que la representación de los objetos impone restricciones importantes en cuanto a las funciones que se pueden aplicar. En (Villuendas et al., 2010) se brindan información sobre varias funciones de similaridad que admiten datos mixtos e incompletos.

El k-NN es altamente sensible a la definición de las funciones de distancia o similitud utilizadas y a las características de los rasgos que conforman cada uno de los ejemplos almacenados. Muchas variantes del k-NN se han propuesto reducir esta sensibilidad parametrizando la función de distancia o similitud con características de peso.

Los resultados presentados en (Duch and Grudzinski, 1999) muestran que un aspecto importante en los métodos basados en grados de similaridad, como el método k-NN, es el peso asignado a los rasgos, pues esto mejora significativamente el desempeño del método. También se muestra que el empleo de métodos de búsqueda para el cálculo de los pesos de los rasgos, obtiene mejores resultados que usar procedimientos de minimización de funciones.

El algoritmo k-NN para aproximar una función objetivo de valor discreto, es decir clasificación se define en la expresión 2.18 por (Mitchell, 1997).

$$f(X_q) \leftarrow argmax_{v \in V} \sum_{i=1}^{k} w_i * \delta(v, f(x_i)) \qquad (2.18)$$

El k-nn se adapta fácilmente a problemas de aproximación de funciones. Para lograr esto, tenemos que calcular el valor medio de los ejemplos de entrenamiento *k* más cercanos, en lugar de calcular su valor más común.

$$f(X_q) \leftarrow \frac{\sum_{i=1}^{k} W_i * f(X_i)}{\sum_{i=1}^{k} W_i} \qquad (2.19)$$

En la propuesta de esta investigación los valores de *Wi* son los pesos calculados por el método PSO+RST+FUZZY, con el objetivo de mejorar la recuperación de objetos similares. Para calcular la similaridad entre los objetos existen varias métricas, en este trabajo se emplea la planteada por la ecuación 2.20.

$$\delta(x, y) = \sum_{i=1}^{n} w_i * \partial_i (X_i, Y_i) \qquad (2.20)$$

De manera general, el aumento de los volúmenes de datos ha hecho que resulte muy complejo hallar los vecinos más cercanos a un objeto determinado. Para ello es

necesario realizar un gran número de comparaciones, proporcional al tamaño del conjunto de datos, y con ello un alto consumo de recursos y tiempo, para alcanzar el resultado (Villuendas et al., 2008). En el contexto de los clasificadores k-NN y k-MSN este problema ha sido abordado fundamentalmente a través de la selección de los rasgos. En (Filiberto et al., 2010a) se presenta un estudio acerca del mejoramiento del desempeño del método k-NN usando los pesos calculados por el método PSO+RST. Este método constituye una vía para mejorar el desempeño del método k-NN usado tanto para problemas de aproximación de funciones como de clasificación. El mismo permite encontrar el conjunto de pesos *W* usado en la función de semejanza que mejora la recuperación de objetos similares. En este trabajo se propone emplear conceptos tomados de la Teoría de los Conjuntos Borrosos a través del método PSO+RST+FUZZY para lograr una mayor eficacia del k-NN tanto como aproximador de funciones que como clasificador.

2.3.2. Efecto del empleo de los pesos de los rasgos en el método *Perceptron* multicapa

Las Redes Neuronales Artificiales (RNA) denotan un conjunto de modelos conexionistas inspirados en el comportamiento del cerebro humano. En particular, el *Multilayer Perceptron* (MLP) es la arquitectura más popular dentro de las RNAs, donde las neuronas se agrupan en capas y sólo hacia adelante existen conexiones. Esto proporciona una poderosa base de aprendizaje, con ventajas en la cartografía no lineal, en la tolerancia al ruido y se utiliza cada vez más en la minería de datos (MD) y Aprendizaje Automático (ML), debido a su buen comportamiento en términos de conocimiento predictivo (Rocha et al., 2005).

Este tipo de red utiliza un algoritmo de aprendizaje muy popular, el algoritmo *Back Propagation* (BP). Este algoritmo es una generalización de la regla LMS "*Least Mean Square*"; por lo que, también se basa en corregir el error (Hocenski et al., 2008).

El proceso de BP comprende básicamente dos pasos a través de las diferentes capas de la red, un paso adelante y un paso hacia atrás. En el pase hacia adelante, un patrón del vector de entrada se aplica a la capa de entrada; este efecto se propaga a través de las diferentes capas y, en consecuencia produce un vector de salida. Durante este proceso, los pesos sinápticos de la red son fijos y no cambian.

Durante el pase hacia atrás los pesos se cambian y se modifican de acuerdo a la regla de corrección de errores. La señal actual de salida se compara con la señal deseada, y los resultados de la señal de error se propagan en la dirección opuesta a través de la red mediante la modificación de los pesos. Al volver el vector de entrada hacia adelante, la respuesta estará más estrechamente relacionada con la salida deseada (Hocenski et al., 2008).

Cada red multicapa se define en términos de su arquitectura, sus funciones de activación, umbrales y pesos. Las dos últimas variables se van a utilizar en el momento de emplear un algoritmo de entrenamiento para que la red aprenda. En el entrenamiento, además de ajustar los pesos y umbrales, es necesario optimizar el número de neuronas que componen la red, ya que tienen una influencia directa en la velocidad de la red (Fu et al., 2010).

En problemas supervisados, los algoritmos de aprendizaje se basan en el error de salida. Este error es la diferencia entre la salida de la red neural y la salida deseada. Los pesos son ajustados por el algoritmo para minimizar el error de la salida de la red neuronal (Hocenski et al., 2008). El carácter esencial del algoritmo BP es el descenso de gradiente. El algoritmo de descenso de gradiente es estrictamente dependiente de la forma de la superficie de error y esta superficie puede tener algún mínimo local y multimodal. Esto da lugar a caer en algún mínimo local y apareciendo convergencia prematura (Fu et al., 2010).

El entrenamiento del BP es muy sensible a las condiciones iniciales. En términos generales, la elección del vector W_0 de pesos iniciales puede acelerar la convergencia del proceso de aprendizaje hacia un global o un mínimo local. Si esto sucede es que se encuentra dentro de la atracción basada en ese mínimo. Por el contrario, si W_0 inicia la búsqueda en una región relativamente plana de la superficie de error, se ralentiza la adaptación de los pesos de conexión (Adam et al., 2009).

Un MLP se compone de una capa de entrada, una capa de salida y una o más capas ocultas, pero se ha demostrado que para la mayoría de los problemas es suficiente con una sola capa oculta. El tamaño de la red depende del número de capas y el número de neuronas en la capa oculta. El número de unidades ocultas está directamente relacionada con las capacidades de la red, en este trabajo el número de neuronas en la capa oculta se

determina mediante *(i*j)/2*, donde *i* es la cantidad de neuronas de entrada y *j* las de salida.

Cada entrada de la red tiene un peso asociado *W*, que se modifica en el llamado proceso de aprendizaje. La capa de entrada es responsable de asignar pesos *Wij* a las entradas utilizando el método propuesto PSO+RST+FUZZY. Esta información se pasa a la capa oculta, y luego se transmite a la capa de salida, que es la responsable de producir la respuesta de la red.

Los valores iniciales de W_{ij}, donde W_{ij} es el peso de la conexión entre el nodo *j* de la capa oculta y el nodo *i* de la capa de entrada, se asignan de la siguiente manera: $W_{ij} = W_i$ para todo *j,* donde W_i es el peso calculado para el rasgo de entrada *i*, según el método PSO+RST+FUZZY descrito anteriormente. Los pesos para los enlaces entre los nodos de la capa oculta y los nodos en la capa de salida se generan aleatoriamente.

Para analizar el efecto del empleo de los pesos de los rasgos en el método Perceptron multicapa se realizaron tres experimentos. Estos experimentos se detallan en el próximo epígrafe.

2.4. Estudio experimental estadístico para validar la Medida de Calidad de la Similaridad Borrosa.

Durante los últimos años la comunidad de investigadores de Aprendizaje Automático se han percatado de la creciente necesidad de validar estadísticamente los resultados publicados. Este fenómeno puede atribuirse a la madures del área, el creciente número de aplicaciones de estos algoritmos a problemas del mundo real, la disponibilidad de programas de código abierto de aprendizaje automático que permiten de manera fácil crear nuevos algoritmos o modificar los ya existentes y compararlos entre ellos mismos (Demsar, 2006).

En un artículo de aprendizaje automático lo típico es seleccionar un juego de datos para probar, se corren los algoritmos y la calidad de los modelos resultantes se evalúa usando una medida apropiada usualmente la exactitud de la clasificación, sin hacer una verificación estadística de la hipótesis. Por suerte muchos investigadores que se enfrentan al problema de comparar dos clasificadores en un solo conjunto de datos han propuesto varias soluciones.

El interés en llevar a cabo una rigurosa metodología de comparación estadística entre algoritmos es cada vez más solicitado entre la comunidad científica. Por simplicidad y ausencia de condiciones iniciales en su aplicación, las técnicas estadísticas no paramétricas han demostrado ser una buena herramienta para comparar algoritmos. En la literatura especializada, podemos encontrar propuestas para utilizar tests estadísticos en el análisis de experimentos (Bartz-Beielstein, 2006).

La mayoría de los trabajos validan resultados a partir de estadísticos básicos, como medias y varianzas, junto con técnicas estadísticas paramétricas (Mori et al., 2005, Demsar, 2006). Sin embargo, es bien conocido que las técnicas estadísticas paramétricas requieren una serie de condiciones sobre la muestra de resultados que no siempre se cumplen.

En los problemas de clasificación supervisada para validar resultados se utilizan criterios tales como: porciento de clasificaciones correctas, diferentes medidas del error, el índice de Kappa, medida F, funcionales de calidad, error, etc. Una de las medidas más conocidas para evaluar la clasificación son los porcientos de clasificaciones correctas que se encuentran dentro de las técnicas de evaluación del rendimiento de clasificadores se eligió la validación cruzada. Esta técnica se incluye dentro de las inferencias probabilística no paramétrica.

A continuación se abordarán las técnicas de comparaciones más relevantes para la validación de resultados utilizando técnicas de estadística no paramétrica. En (Demsar, 2006, García and Herrera, 2008, Sheskin, 2003b) se plantea que cuando se desea realizar una comparación entre varios clasificadores se tienen dos alternativas, las pruebas de Anova o el test de Friedman. Anova es el método estadístico habitual pero requiere de las condiciones paramétricas que como se mencionó anteriormente no se cumplen. El Test de Friedman es la versión no-paramétrica del ANOVA, y al no cumplirse las condiciones impuestas por ANOVA se usa el Test de Friedman.

En estadística, la prueba de Friedman es una prueba no paramétrica desarrollada por el economista Milton Friedman. Esta prueba puede utilizarse en aquellas situaciones en las que se seleccionan n grupos de k elementos de forma que los elementos de cada grupo sean lo más parecidos posible entre sí, el método consiste en ordenar los datos por filas o bloques, reemplazándolos por su respectivo orden.

H0: No existen diferencias entre los grupos.

H1: Existen diferencias entre los grupos.

Este test trabaja asignando rankings *rij* a los resultados obtenidos por cada algoritmo *j* en cada problema *i*. Esto es, para cada problema, se asigna un ranking $1 \leq rij \leq k$, donde *k* es el número de algoritmos a comparar (García & Herrera, 2009). Estos rankings se asignan de forma ascendente, es decir, 1 al mejor resultado, 2 al segundo, etc., en caso de haber empates, se asignan rankings medios.

El Test de Friedman requiere el cálculo de los rankings medios de los algoritmos sobre los *n* problemas. Para ello se emplea la expresión 2.21.

$$R_j = \frac{\sum_{i=1}^{n} rij}{n} \tag{2.21}$$

La hipótesis nula que indica que todos los algoritmos se comportan similarmente, por lo que sus rankings *Rj* deben ser similares. Siguiendo esta hipótesis, el estadístico de Friedman se define por la expresión 2.22, y se distribuye de acuerdo a una distribución X^2con *k-1* grados de libertad.

$$X_r^2 = \frac{12}{nk\,(k-1)}\left[\sum_{j=1}^{k}\left(\sum R_j\right)2\right] - 3n(k+1) \tag{2.22}$$

Con los valores de rankings obtenidos con el Test de Friedman podemos establecer una ordenación clara entre algoritmos e incluso medir las diferencias entre ellos, pero no se puede concluir que una propuesta es mejor que otra a no ser que la hipótesis de comparación quede rechazada (García, Molina, Lozano, & Herrera, 2007).

Este estadístico fue mejorado a su vez por Iman y Davenport (Iman and Davenport, 1980), quienes mostraron que el estadístico de Friedman presenta un comportamiento demasiado conservativo. Para evitar este problema, propusieron otro estadístico más ajustado, el cual se refleja en la expresión 2.23 y se distribuye de acuerdo a una distribución F con *k-1* y *(k-1)(n-1)* grados de libertad.

$$F_{ID} = \frac{(n-1)X_r^2}{n\,(k-1) - X_r^2} \tag{2.23}$$

Si se rechaza la hipótesis nula, se puede proceder con un test *post-hoc* para encontrar diferencias a *posteriori* (García and Herrera, 2008).Para incrementar la potencia de los test múltiples existen un conjunto de métodos, llamados métodos secuenciales, o test de

post-hoc, que además sirven para evitar la redundancia en las conclusiones y organizar el orden en el que se realizan los contrastes. Los procedimientos que vamos a usar son: Holm (Holm, 1979); Holland-Copenhaver (Holland and Copenhaver, 1987); Hochberg(Hochberg, 1998); Hommel (Hommel, 1988)y Rom (Rom, 1990).

A través del test de Friedman ya se obtuvo el valor de ranking medio *Ri* para el algoritmo *i* sobre todas las funciones, el valor z de cada hipótesis se calcula mediante la expresión de la ecuación 2.24:

$$z = (\mathrm{R_i} - \mathrm{R_j})\sqrt{\mathrm{M(M+1)/6N}} \qquad (2.24)$$

El valor de z de cada hipótesis se puede utilizar para calcular la probabilidad de error (valor p) mediante la distribución normal (Sheskin, 2003a). El test de Friedman no identifica las diferencias existentes entre el mejor algoritmo encontrado (denominado como algoritmo de control) y el resto. Estos test se limitan a detectar la existencia o no de diferencias en todo el conjunto de resultados. Si existen diferencias, es necesario proceder con un procedimiento *post-hoc* de identificación de diferencias, basados en la distribución normal.

A continuación, presentamos los 4 procedimientos *post-hoc* más útiles para detectar diferencias a *posteriori*. Una vez obtenido el valor *z* empleando la expresión 2.24 es posible obtener a partir de él el *p*-valor no ajustado correspondiente.

Test de Holm

Es un procedimiento incremental que define diferentes niveles de rechazo por hipótesis (Holm, 1979). Sea *p1, ..., pm* los valores *p* ordenados, del más pequeño al mayor, y *Hi, ...,Hm* las correspondientes hipótesis, el procedimiento de rechaza *H1* hasta *H(i−1)* si *i* es el entero más pequeño tal que *pi > α/(m − i + 1)*.

Test de Holland- Copenhaver

Siguiendo el mismo esquema que Holm, el procedimiento de Holland-Copenhaver (Holland and Copenhaver, 1987) rechaza *H1* hasta *H(i−1)* si *i* es el entero más pequeño tal que $p_i > 1 - (1-\alpha)^{1/(m+i-1)}$.

Procedimiento de Hochberg

Hochberg (Hochberg, 1998) desarrolló un procedimiento decremental que consiste en utilizar la misma ordenación y niveles $\alpha_1^{'}. \ldots \alpha_m^{'}$ ajustados que Holm.

Si $p(m-1) < \alpha'_{(m-1)}$, entonces todas las hipótesis restantes son rechazadas, si no $H(m-1)$ es aceptada, y $p(m-2)$ es comparada con $\alpha'_{(m-2)}$; así sucesivamente. Este método es más potente que el procedimiento de Holm, pero requiere que los test estadísticos sean independientes.

Procedimiento de Hommel

Es un procedimiento decremental que también requiere la independencia de los tests estadísticos (Hommel, 1988). Determina los valores críticos ajustados en dos fases.

Sea $J = \{i' \in \{1, ...,m\} : p(m-i'+k) > k\alpha/i'; k = 1, ..., i'\}$. La primera fase del procedimiento utiliza los valores p obtenidos para determinar el número de miembros en J. La segunda fase obtiene el nivel de significancia de rechazo usando $\alpha' = \alpha/j'$, donde j' es el número más grande en J. Si J está vacío, todas las $Hi(i = 1, ...,m)$ son rechazadas. Si J no está vacío, Hi es rechazado siempre que $pi \leq \alpha/j'$. Este procedimiento es más potente que el de Hochberg.

Procedimiento de Rom

Rom en (Rom, 1990) desarrolló una modificación al procedimiento de Hochberg para incrementar su potencia. El incremento se consigue identificando los niveles de significancia apropiados que controlan los errores de Tipo I cuando los test estadísticos son independientes. Funciona exactamente igual que el procedimiento de Hochberg, pero los valores α^l se calculan mediante la expresión

$$\alpha'_{m-i+1} = \left[\sum_{j=1}^{i-1} \alpha^j - \sum_{j=1}^{i-2} \binom{i}{j} \alpha'^{(i-j)}_{(m-j)}\right] / i \tag{2.25}$$

en donde $\alpha'_m = \alpha$ y $\alpha'_{(m-1)} = \alpha/2$.

2.4.1 Validación experimental en conjuntos de datos internacionales

Para este estudio se emplearon conjuntos de datos reconocidos internacionalmente, descritos en el Anexo 1, provenientes del UCI Repository de Aprendizaje Automático (Asuncion & Newman, 2007). En la presente investigación se utiliza el test de Friedman y el Iman-Davenport para obtener los rankings de los algoritmos e identificar si existen diferencias significativas entre los mismos. Se emplea el test de Holm para probar secuencialmente las hipótesis ordenadas según su significancia e identificar las diferencias existentes entre el algoritmo de control y el resto. Este test tiene la ventaja

de no realizar ninguna suposición adicional sobre las hipótesis chequeadas y en (García and Herrera, 2008) se considera que el Test de Holm debe aparecer siempre en toda validación experimental.

Se obtuvieron los conjuntos de entrenamiento y control, tomando para el primero el 75% de los casos y para el segundo el 25%, de forma totalmente aleatoria. Siguiendo este principio de selección aleatoria se repitió el proceso diez veces y se obtuvieron para cada conjunto de datos diez conjuntos de entrenamiento y diez conjuntos de control, con el fin de aplicar validación cruzada (cross validation) para una mejor validación de los resultados (Filiberto, 2012).

El desempeño del método k-NN usando los pesos calculados por el método propuesto en esta investigación, llamado PSO+RST+FUZZY, se comparó con el resultado alcanzado por el método precedente PSO+RST y con los pesos calculados por otros métodos altamente referenciados como: el método del Gradiente Conjugado (KNN_{VSM}), el método Relief y Estándar (1/Cantidad-Rasgos). Estos métodos son los mismos empleados en la validación del PSO+RST en (Filiberto, 2012).

Para evaluar la precisión de los resultados alcanzados para el caso de aproximación de funciones se utilizó la medida PMD (promedio de las diferencias entre el valor deseado y el producido por el método)(Weber and Foix, 2007), definida por la expresión 2.25, donde: $\boldsymbol{x}_i(d)$ es el valor del rasgo de decisión para el objeto $\boldsymbol{x}_i$; q_i es el valor producido por el método y m es el número de objetos.

$$PMD = \frac{\sum_{i=1}^{m} |X_i(d) - q_i|}{m} \qquad (2.26)$$

Por último es importante resaltar que en (García and Herrera, 2008) se plantea que en la comparación de nuevas propuestas con varios algoritmos, el número de problemas (dominios) debe ser superior al doble de algoritmos como mínimo. Con menos problemas, es altamente probable que no se pueda rechazar ninguna hipótesis. Para cumplir con esta norma se decidió variar la cantidad de conjuntos de datos para cada experimento, así en el Experimento 1 y 2 se usan 10 conjuntos de datos para comprar 4 algoritmos. En el Experimento 3, 4 y 5 se usan 12 conjuntos de datos para comprar 5 algoritmos. Por último para el Experimento 6 se usan 14 conjuntos de datos para comprar 6 algoritmos.

Para analizar el efecto del empleo de los pesos de los rasgos en el método k-vecinos más similares se realizaron tres experimentos. Estos experimentos se detallan a continuación.

Experimento 1: Comparar los resultados del estimador de funciones k-NN, cuando se utilizan los diferentes métodos de cálculo de pesos: Estándar (1/Cantidad-Rasgos), KNNVSM, PSO+RST y el método de cálculo de pesos propuesto PSO+RST+FUZZY respecto al PMD.

Objetivos: Determinar si el método propuesto PSO+RST+FUZZY es significativamente inferior a los métodos Estándar, KNN_{VSM} y PSO+RST respecto al PMD.

Las pruebas de Friedman e Iman-Davemport arrojaron que existen diferencias significativas entre los resultados del Experimento 1. La prueba de Holm demostró que los resultados de las medidas cuando se calculan los pesos por el método PSO+RST+FUZZY son significativamente inferiores a los que se obtienen cuando se calculan los pesos mediante las otras medidas. En el Anexo 2 se muestran las tablas de las pruebas estadísticas relacionadas con el experimento 1.

Experimento 2: Comparar los resultados del estimador de funciones K-NN, cuando se utilizan los diferentes métodos de cálculo de pesos: Estándar (1/Cantidad-Rasgos), KNN_{VSM}, PSO+RST y el método de cálculo de pesos propuesto PSO+RST+FUZZY, respecto al coeficiente de correlación de Pearson, al coeficiente R2 y el error estándar.

Objetivos: Determinar si el método propuesto PSO+RST+FUZZY es significativamente superior a los métodos Estándar, KNN_{VSM}, PSO+RST respecto al coeficiente de correlación de Pearson, al coeficiente R^2 y significativamente superior con respecto al error estándar.

Las medidas de coeficiente de correlación de Pearson, coeficiente R^2 y error estándar se recomiendan en (Grau, 1994) para problemas de aproximación de funciones. En el caso de esta investigación se decidieron usar estas medidas por el interés que representa para el campo de la ingeniería civil, que es el área del problema real donde se aplicaron posteriormente estos métodos.

Experimento 3: Comparar los resultados de la exactitud general de la clasificación del k-NN, cuando se utilizan los diferentes métodos de cálculo de pesos: Estándar (1/Cantidad-Rasgos), KNN_{VSM}, Relief, PSO+RST y el método de cálculo de pesos propuesto PSO+RST+FUZZY

Objetivos: Determinar si el método propuesto PSO+RST+FUZZY es significativamente superior a los métodos Estándar, KNN_{VSM}, Relief y PSO+RST en cuanto a la exactitud general de la clasificación obtenida por el k-NN.

Las pruebas de Friedman e Iman-Davemport arrojaron que existen diferencias significativas entre los resultados del Experimento 3, donde el mejor ranking es obtenido por el método PSO+RST. La prueba de Holm, respecto a la exactitud general de la clasificación rechaza todas las hipótesis con p-valor ≤ 0.025: Estándar, KNN_{VSM} y Relief. Con respecto a PSO+RST+FUZZY vs PSO+RST la hipótesis nula no se rechaza, esto es equivalente a decir que no hay diferencias significativas entre sus comportamientos y por lo tanto se puede concluir que son igualmente eficaces. En el Anexo 4 se muestran las tablas de las pruebas estadísticas relacionadas con el experimento 3.

Experimento 4: Comparar los resultados del MLP usado en problemas de aproximación de funciones, cuando se utilizan los diferentes métodos de cálculo de pesos: Aleatorios (MLP-AL), Estándar (1/Cantidad-Rasgos), KNNVSM, PSO+RST y el método de cálculo de pesos propuesto PSO+RST+FUZZY, respecto al PMD.

Objetivos: Determinar si los valores de PMD son significativamente inferiores cuando se inicializan los pesos por el método PSO+RST+FUZZY que cuando se inicializan los pesos por los métodos Aleatorios (MLP-AL), Estándar, KNN_{VSM} y PSO+RST.

Las pruebas de Friedman e Iman-Davemport arrojaron que existen diferencias significativas entre los resultados del Experimento 4, donde el mejor ranking es obtenido por el método propuesto en esta investigación PSO+RST+FUZZY. La prueba de Holm, respecto a la exactitud general de la clasificación rechaza todas las hipótesis con p-valor ≤ 0.025: Estándar, KNN_{VSM} y Relief. Con respecto a PSO+RST+FUZZY vs PSO+RST la hipótesis nula no se rechaza, esto es equivalente a decir que no hay diferencias significativas entre sus comportamientos y por lo tanto se puede concluir que son igualmente eficaces. En el Anexo 5 se muestran las tablas de las pruebas estadísticas relacionadas con el experimento 4.

Experimento 5: Comparar los resultados del MLP usado en problemas de aproximación de funciones, cuando se inicializan los pesos por los diferentes métodos de cálculo de pesos: Aleatorios (MLP-AL), Estándar (1/Cantidad-Rasgos), KNN_{VSM},

PSO+RST y el método de cálculo de pesos propuesto PSO+RST+FUZZY, respecto al coeficiente de correlación de Pearson, al coeficiente R^2 y al error estándar.

Objetivos: Determinar si el método propuesto PSO+RST+FUZZY es significativamente superior a los métodos Aleatorios (MLP-AL), Estándar, KNN_{VSM} y PSO+RST respecto al coeficiente de correlación de Pearson, al coeficiente R^2 y al error estándar.

Experimento 6: Comparar los resultados de la exactitud de la clasificación general del MLP, cuando se utilizan los diferentes métodos de cálculo de pesos: Aleatorios (MLP-AL), Estándar (1/Cantidad-Rasgos), KNN_{VSM}, Relief, PSO+RST y el método de cálculo de pesos propuesto PSO+RST+FUZZY.

Objetivos: Determinar si los resultados de la exactitud general del MLP es significativamente superior cuando se inicializan los pesos a través del método propuesto PSO+RST+FUZZY que al emplear los métodos Aleatorios (MLP-AL), Estándar, KNN_{VSM}, Relief y PSO+RST.

Las pruebas de Friedman e Iman-Davemport arrojaron que existen diferencias significativas entre los resultados del Experimento 6. La prueba de Holm demostró que los resultados de las medidas cuando se calculan los pesos por el método PSO+RST+FUZZY son significativamente superiores a los que se obtienen cuando se calculan los pesos mediante los otros métodos. En el Anexo 7 se muestran las tablas de las pruebas estadísticas relacionadas con el experimento 6.

Conclusiones parciales

Utilizando los conceptos de la Teoría de los Conjuntos Borrosos y específicamente las relaciones borrosas binarias se redefinió la Medida de Calidad de la Similaridad, planteando una nueva métrica denominada Medida de Calidad de la Similaridad Borrosa. Esta modificación se reutilizó como medida calidad para evaluar las partículas en el también redefinido método para pesado de rasgos PSO+RST+FUZZY. Se demostró la eficacia de los pesos calculados por el PSO+RST+FUZZY para mejorar el desempeño de los métodos de aprendizaje k-NN y MLP, obteniendo resultados experimentales similares con el empleo del nuevo método en los 6 experimentos realizados.

Capítulo III. Aplicaciones de la Medida de Calidad de la Similaridad Borrosa

En el capítulo anterior se demostró con bases de datos internacionales, el buen desempeño que tienen la medida de la calidad de la similaridad borrosa con el método de pesado de rasgos PSO+RST+FUZZY. En este capítulo se aplica la medida en la solución de problemas reales relacionados con las ramas de la Ingeniería Civil y la Rendimiento Académico.

En estos dominios se tienen problemas con dominios continuos y de clasificación.

El procedimiento seguido para resolver los problemas en las ramas de la Ingeniería Civil y el Rendimiento Académico fue el siguiente:

1. Construir el sistema de decisión para el dominio de aplicación.
2. Calcular la importancia de los rasgos (usando PSO+RST+FUZZY).
3. Aplicar el método k-NN y el MLP para problemas de aproximación de funciones en aplicaciones de la Ingeniería Civil.
4. Aplicar el método k-NN y el MLP para clasificación en aplicaciones del Rendimiento Académico.

3.1. Aplicaciones en la Ingeniería Civil

Uno de los campos donde la computación está logrando grandes avances es la Ingeniería Civil y específicamente en la construcción compuesta. Esta ya está siendo empleada en la práctica constructiva a nivel internacional ((Filiberto et al., 2010b).

Los incendios causan anualmente elevadas pérdidas tanto en lo referente a vidas humanas como a intereses económicos. En Cuba las normas técnicas relacionadas al proyecto de estructuras en situación de incendio son obsoletas respecto a las utilizadas por los países con más avances en el tema. Por otra parte los modelos simplificados de cálculo permiten obtener el aumento de la temperatura de manera homogénea en cada elemento, por medio de procedimientos analíticos sencillos, pero pueden conducir a resultados anti-económicos. A lo anterior se añade que para la verificación a momento flector y esfuerzos cortantes de las secciones transversales compuestas no existen herramientas que puedan ser aplicadas de forma directa para el dimensionamiento en

situación de incendio. Además, debe tenerse en cuenta que el programa sueco SuperTempcalc de probada eficacia en el análisis térmico bidimensional de las secciones estructurales en situación de incendio, no contiene un módulo de cálculo para la determinación de los esfuerzos resistentes de las secciones transversales de vigas compuestas en esta situación.

Una construcción compuesta se caracteriza por la presencia de dos o más elementos que trabajen en conjunto. La exposición de estos materiales estructurales a altas temperaturas hace degenerar sus características físicas y químicas, causando reducción de la resistencia y la rigidez. Esto debe ser tomado en cuenta en el diseño de las estructuras en situación de incendio.

Para la verificación a momento flector y a cortante vertical de secciones transversales compuestas de hormigón y acero no existen herramientas que pueden ser aplicadas de forma directa para el dimensionamiento en situación de incendio. Se considera que el desenvolvimiento de tales herramientas será bastante útil y, para eso, es necesario considerar métodos avanzados para la obtención de los campos térmicos en las secciones transversales de las vigas, con vistas a un dimensionamiento más racional. Por lo anterior se enfrentó la tarea de desarrollar una herramienta computacional capaz de realizar el análisis estructural de secciones compuestas de acero y hormigón, denominada SCBEAM (por afinidad con los módulos estructurales SBEAM y CBEAM de SuperTempcalc), que es capaz de predecir a través de técnica de Inteligencia Artificial con efectividad y rapidez los momentos flectores resistentes, positivos y negativos, a partir de los resultados emitidos por el módulo de análisis térmico del programa de modelación térmica bidimensional SuperTempcalc. SCBEAM es completamente pertinente y necesario, con alcance nacional e internacional, favorece el desarrollo de ayudas de diseño no computacionales para la verificación a momento flector y cortante, en situación de incendio de las secciones transversales en las regiones de momentos positivos para vigas compuesta, consistentes en gráficos que permiten, de manera práctica y directa, determinar el tiempo de resistencia al fuego (TRF) asociado a esa verificación, todo lo que resulta novedoso a nivel internacional. Este software actualmente posee como técnica de aprendizaje automático un MLP estándar como método para la regresión.

En esta investigación se propone experimentar con los algoritmos MLP y k-NN con los pesos obtenidos mediante el método propuesto en este trabajo PSO+RST+FUZZY para validar si estas técnicas se pueden implementar en el SCBEAM para predecir la resistencia de las estructuras compuestas en caso de incendio. Se realizó el estudio experimental para el conjunto de datos. La variable de salida es la capacidad resistente última de un conector, determinada experimentalmente, *Qexp* con dominio real, por lo que se trata de un problema de aproximación de funciones. En las tablas 3.2 y 3.3 se muestran los resultados obtenidos con el algoritmo k-NN y en la tabla 3.1 se muestra la descripción del conjunto de datos de incendio.

Tabla 3.1. Descripción del conjunto de datos de incendio.

Atributos	**Tipo de dato**
hp	real
hw	real
bf	real
tala	real
talma	real
d	real
Aala	real
Aalma	real
A	real
alasup	real
alma	real
alainf	real
total	real
TRF	real
Clases	real
Cantidad de instancias: 117	

Tabla 3.2. Resultados del error PMD obtenido con el algoritmo k-NN usando los pesos obtenidos con los métodos (KNNVSM, Estándar, PSO+RST y PSO+RST+FUZZY), con k=1, k=2 y k=3, para el conjunto de datos incendio.

Conjunto de datos	Métodos	*k*	PMD
incendio	PSO+RST+FUZZY	*k*=1	**0.037**
		k=2	0.098
		k=3	0.15
	PSO+RST	*k*=1	0.04
		k=2	0.10
		k=3	0.15
	K-NN_{VSM}	*k*=1	0.49
		k=2	0.58
		k=3	0.55
	Estándar	*k*=1	0.55
		k=2	0.55
		k=3	0.54

Tabla 3.3. Resultados del coeficiente de correlación, coeficiente R^2 y error estándar obtenidos con el algoritmo k-NN usando los pesos obtenidos con los métodos (KNNVSM, Estándar, PSO+RST y PSO+RST+FUZZY), con k=1 para el conjunto de datos incendio.

Conjunto de datos	Métodos	Coeficiente de correlación	Coeficiente R^2	Error estándar
incendio	PSO+RST+FUZZY	0.955	0.912	0.115
	PSO+RST	0.955	0.912	0.115
	KNN_{VSM}	0.289	0.083	0.375
	Estándar	0.318	0.101	0.377

En las tablas 3.4 y 3.5 se muestran los resultados obtenidos con el algoritmo MLP usando los pesos obtenidos con los métodos (MLP-AL, KNN_{VSM}, Estándar, PSO+RST y PSO+RST+FUZZY) para el conjunto de datos incendio.

Tabla 3.4. Resultados del errores MAPE y PMD para el algoritmo MLP, usando los pesos obtenidos con los métodos (KNNVSM, Estándar, MLP-AL, PSO+RST y PSO+RST+FUZZY).

Conjuntos de datos	Métodos	MAPE (%)	PMD
incendio	PSO+RST+FUZZY	**6.17**	**0.015**
	PSO+RST	10.12	0.0248
	KNNVSM	28.75	0.0615
	Estándar	31.03	0.0602
	MLP-AL	8.11	0.0232

Tabla 3.5. Resultados del coeficiente de correlación, coeficiente R^2 y error estándar para el algoritmo MLP, usando los diferentes métodos de cálculo de peso (KNNVSM, Estándar, MLP-AL, PSO+RST y PSO+RST+FUZZY).

Conjunto de datos	Métodos	Coeficiente de correlación	Coeficiente R^2	Error estándar
Incendio	PSO+RST+FUZZY	0.998	0.997	0.02
	PSO+RST	0.996	0.993	0.03
	KNNVSM	0.981	0.962	0.076
	Estándar	0.981	0.963	0.075
	MLP-AL	0.996	0.992	0.033

Como se puede observar en las tablas 3.2 y 3.4, cuando se utiliza el método PSO+RST+FUZZY para obtener los pesos en los algoritmos k-NN y MLP para la base de incendio, los resultados son superiores a cuando se utilizan los restantes métodos de pesado.

Se concluye que ambos métodos tienen un buen desempeño en problemas de aproximación de funciones. A partir de estos resultados se decide que el algoritmo que se utilizará en el software SCBEAM para calcular la resistencia de estructuras compuestas en caso de incendio es el MLP con PSO+RST+FUZZY, pues es la combinación que logra mayor reducción del PMD.

3.2. Aplicaciones en el Rendimiento Académico

Investigadores cubanos han desarrollado diferentes estudios en los que utilizan técnicas matemáticas para la predicción de los resultados docentes (Luan, 2002). Dentro de los métodos más utilizados se encuentran los arboles de clasificación, redes neuronales, entre otras (Luan, 2002). Lo más usual en la aplicación de estas técnicas es para lograr predecir dos estados: "éxito" o "fracaso" (Heredia et al., 2012).

En la Universidad de Camagüey "Ignacio Agramonte y Loynaz", se creó la Facultad de Informática en el año 1999 y desde los primeros años de su creación, siempre ha sido una preocupante para los profesores, específicamente los que imparten asignaturas propias de la carrera, poder determinar que estudiantes tendrán problemas o no en su paso para cursar el segundo año de la carrera. Debido a la variedad de estudiantes que ingresan en esta carrera sus resultados docentes no son los mismos.

Los estudiantes provienen de diferentes municipios, tienen diferentes edades, y su fuente de ingreso es variada, donde la principal diferencia es la preparación docente que precede a la superior. Dentro de las características que se tienen en cuenta para realizar la predicción de los resultados de estos estudiantes se encuentran gran cantidad de datos que se llenan en la planilla de ingreso de los mismos y las evaluaciones de todas las asignaturas recibidas por ellos en primer año de la carrera. Todos estos datos fueron tomados de la Secretaría Docente de la Facultad de Informática de la Universidad de Camagüey "Ignacio Agramonte y Loynaz". A continuación se explican las características de las bases de casos creadas para realizar las predicciones.

Se realizaron 3 bases de conocimientos. Una primera con los datos de los alumnos al ingresar al primer año y nos permite predecir si pasaran el primer el año, bien, mal o regular. Una segunda base para la asignatura Introducción a la Inteligencia Artificial, la cual mantiene los rasgos de ingreso de los estudiantes e incluye a diferencia de la anterior las notas de las asignaturas de primer año. Esta segunda base nos permite predecir si pasaran la materia de Introducción a la Inteligencia Artificial con o sin problemas. Una tercera base igual que la anterior, nos permite predecir si pasaran a materia de Estructura de Datos con o sin problemas. Las instancias se corresponden con la información recopilada durante 10 años. Las materias Introducción a la Inteligencia

Artificial y Estructura de Datos fueron seleccionadas por ser las que permitían obtener bases balanceadas.

En la tabla 3.6 se muestra la descripción del conjunto de datos de docencia.

Tabla 3.6. Descripción del conjunto de datos de Primer Año.

Atributos	**Tipo de dato**
Municipio	symbol
Fuente de Ingreso	symbol
Origen Académico	symbol
Estado Civil	symbol
Organización Política	symbol
Sexo	symbol
Color de Piel	symbol
Tipo de Estudiante	symbol
Nivel Académico de los Padres	symbol
Tipo de Servicio Militar	symbol
Edad de ingreso	entero
Índice Académico	real
Lugar en el Escalafón	entero
Opción en la cual solicitó la carrera	entero
Clases	**Tipo de dato**
Bien	symbol
Mal	symbol
Regular	symbol
Cantidad de instancias: 222	

En la tabla 3.7 y 3.8 se muestra la descripción del conjunto de datos de Estructura de Datos e Inteligencia Artificial respectivamente.

Tabla 3.7. Descripción del conjunto de datos de Estructura de Datos

Atributos	**Tipo de dato**
Municipio	symbol
Fuente de Ingreso	symbol

Origen Académico	symbol
Estado Civil	symbol
Organización Política	symbol
Sexo	symbol
Color de Piel	symbol
Tipo de Estudiante	symbol
Nivel Académico de los Padres	symbol
Tipo de Servicio Militar	symbol
Edad de ingreso	entero
Pre-universitario	symbol
Índice Académico	real
Lugar en el Escalafón	entero
Opción en la cual solicitó la carrera	entero
Nota Asignatura _mat_uno	entero
Nota Asignatura _algebra	entero
Nota Asignatura _filosof	entero
Nota Asignatura _ftos_inf	entero
Nota Asignatura _intro_prog	entero
Nota Asignatura _efisica_uno	entero
Nota Asignatura _ingles_uno	entero
Nota Asignatura _historia	entero
Nota Asignatura _mat_dos	entero
Nota Asignatura _mat_disc	entero
Nota Asignatura _eco_pol_uno	entero
Nota Asignatura _ingles_dos	entero
Nota Asignatura _poo	entero
Nota Asignatura _efisca_dos	entero
Nota Asignatura intro_gest_sw	entero
Clases	**Tipo de dato**
Con Problema	symbol

Sin Problema	symbol
Cantidad de instancias: 231	

Tabla 3.8. Descripción del conjunto de datos de Inteligencia Artificial.

Atributos	Tipo de dato
Municipio	symbol
Fuente de Ingreso	symbol
Origen Académico	symbol
Estado Civil	symbol
Organización Política	symbol
Sexo	symbol
Color de Piel	symbol
Tipo de Estudiante	symbol
Nivel Académico de los Padres	symbol
Tipo de Servicio Militar	symbol
Edad de ingreso	entero
Pre-universitario	symbol
Índice Académico	real
Lugar en el Escalafón	entero
Opción en la cual solicitó la carrera	entero
Nota Asignatura _mat_uno	entero
Nota Asignatura _algebra	entero
Nota Asignatura _filosof	entero
Nota Asignatura _ftos_inf	entero
Nota Asignatura _intro_prog	entero
Nota Asignatura _efisica_uno	entero
Nota Asignatura _ingles_uno	entero
Nota Asignatura _historia	entero
Nota Asignatura _mat_dos	entero
Nota Asignatura _mat_disc	entero

Nota Asignatura _eco_pol_uno	entero
Nota Asignatura _ingles_dos	entero
Nota Asignatura _poo	entero
Nota Asignatura _efisca_dos	entero
Nota Asignatura intro_gest_sw	entero
Clases	**Tipo de dato**
Con Problema	symbol
Sin Problema	symbol
Cantidad de instancias: 231	

Para dar solución a esta problemática se utilizaron los algoritmos MLP y k-NN con los pesos obtenidos mediante el método propuesto en esta tesis PSO+RST+FUZZY. Se realizó el estudio experimental para los conjunto de datos de la problemática del rendimiento académico. En las tablas 3.9 y 3.10 se muestran los resultados obtenidos con ambos algoritmos.

Tabla 3.9. Resultados de la exactitud de la clasificación general, para el algoritmo k-NN para k=1 usando los diferentes métodos de cálculo de peso para el conjunto de datos de Inteligencia Artificial, Estructura de Datos y Primer Año.

Conjuntos de datos	Métodos	Precisión de la Clasificación
Inteligencia Artificial	PSO+RST+FUZZY	**83.9**
	PSO+RST	82.3
	KNNVSM	80.85
	Estándar	76.89
Estructura de Datos	PSO+RST+FUZZY	**76.8**
	PSO+RST	71.3
	KNNVSM	75.3
	Estándar	68.39
Primer Año	PSO+RST+FUZZY	68.4
	PSO+RST	65.5
	KNNVSM	68

	Estándar	**69**

Tabla 3.10. Resultados de la exactitud de la clasificación general, para el algoritmo MLP usando los diferentes métodos de cálculo de peso para el conjunto de datos de primer año.

Conjuntos de datos	Métodos	Precisión de la Clasificación
Primer Año	PSO+RST+FUZZY	**77.27**
	PSO+RST	64.98
	KNNVSM	60.41
	Estándar	53.39
	Relief	61.4
	MLP-AL	58.71

Como se puede observar en las tablas 3.9 y 3.10, cuando se utiliza el método de cálculo de pesos PSO+RST+FUZZY para obtener los pesos en los algoritmos k-NN y MLP para problemas de clasificación los resultados son superiores a cuando se utilizan los restantes. El método k-NN demostró ser eficiente para predecir si pasaran las materias de Inteligencia Artificial y Estructura de Datos con o sin problemas. El MLP por su parte demostró ser eficiente para predecir si pasaran el primer el año, bien, mal o regular.

Conclusiones parciales

En este capítulo se demuestra que la Medida de Calidad de la Similaridad Borrosa es también útil en la solución de problemas reales relacionados con las ramas de la Ingeniería Civil y el Rendimiento Académico. Los resultados alcanzados confirmaron que las modificaciones realizadas a los métodos de aprendizaje k-Vecinos más Cercanos y *Multilayer Perceptron* mejoran el desempeño de estos ya que los mejores resultados se obtuvieron al asignarle los pesos obtenidos por el método propuesto PSO+RST+FUZZY.

Conclusiones

A través de la Teoría de los Conjuntos Borrosos y las relaciones borrosas binarias se logró definir una medida, denominada Calidad de la Similaridad Borrosa. Esta medida cuantifica la similaridad de un sistema de decisión con datos mixtos y rasgo de decisión continuo. La medida puede ser utilizada como métrica para evaluar la calidad de un conjunto de datos.

El empleo de la Calidad de la Similaridad Borrosa como función de evaluación heurística de cada individuo proveniente de la meta heurística PSO posibilitó la redefinición de un método llamado PSO+RST+FUZZY para pesar los rasgos.

La efectividad de los pesos calculados quedó probada en los resultados arrojados por las modificaciones realizadas a los métodos de aprendizaje k- Vecinos más cercanos y *Multilayer Perceptron*, obteniendo PSO+RST+FUZZY un mejor valor de ranking respecto a otros algoritmos de pesado de rasgos.

La investigación realizada mostró que los métodos propuestos no solo tienen un buen desempeño en los conjuntos de datos internacionales, sino que también son efectivos en la solución de problemas reales relacionados con las ramas de la Ingeniería Civil y el Rendimiento Académico.

Recomendaciones

- Estudiar el comportamiento de la Medida de Calidad de la Similaridad Borrosa utilizando otras medidas de similaridad entre Conjuntos Borrosos propuestos en (Wang, 1997).
- Estudiar la Medida de Calidad de la Similaridad Borrosa en el contexto de las "Fuzzy Rough Set" y "Rough Fuzzy Set".
- Aplicar la Medida Calidad de la Similaridad Borrosa en la generación de reglas de clasificación y en la construcción de prototipos.
- Desarrollar un estudio comparativo entre el método propuesto PSO+RST+FUZZY con otras técnicas propuestas para realizar el cálculo de los pesos iniciales basadas en la metaheurística Particle Swarm Optimization.

Referencias Bibliográficas

1. ADAM, S., ALEXIOS, D. & VRAHATIS, M. 2009. Revisiting the Problem of Weight Initialization for Multi-Layer Perceptrons Trained with Back Propagation. *Heidelberg Springer,* ICONIP 5507**,** 308-315.
2. ALAM, S., DOBBIE, G., KOH, Y. S., RIDDLE, P. & REHMAN, S. U. 2014. Research on particle swarm optimization based clustering: A systematic review of literature and techniques. *Swarm and Evolutionary Computation,* 17 1–13.
3. AMANT, R. & COHEN, P. R. 1998. Interaction With a Mixed-Initiative System for Exploratory Data Analysis. *Knowledge-Based Systems,* 10**,** 265-273.
4. ARCO, L. & BELLO, R. 2006. On clustering validity measures and the Rough Set Theory. *5th Mexican International Conference on Artificial Intelligence.* IEEE Computer Society Press.
5. ASUNCION, A. & NEWMAN, D. 2007. UCI machine learning repository, A study of the behaviour of several methods for balancing machine learning training data. *SIGKDD Explorations,* 6**,** 2-29.
6. BARTZ-BEIELSTEIN, T. 2006. Experimental research in evolutionary computation: The new experimentalism. *Springer*.
7. BAZAN, J., SON, N. H., SKOWRON, A. & SZCZUKA, M. A View on Rough Set Concept Approximations. *In:* 26392003, L., ed. Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing. 9th International Conference, RSFDGRC2003, 2003 Chongqing, China.
8. BELLO, M., GARCÍA, M. M. & BELLO, R. 2013. A method for building prototypes in the nearest prototype approach based on similarity relations for problems of function approximation. *LNCS,* 7629**,** 39-50.
9. BODENHOFER, U. 2000. A similarity-based generalization of fuzzy orderings preserving the classical axioms. *Uncertainty and Fuzziness Knowledge-Based Systems,* 8**,** 593–610.
10. CABALLERO, Y. 2007a. *Acerca del Aprendizaje Automatizado, la Selección de Características y la Edición de conjuntos de Entrenamient.* Tesis de Doctorado, Universidad de Camagüey.

11. CABALLERO, Y. 2007b. *Aplicación de la Teoría de los Conjuntos Aproximados en el preprocesamiento de los conjuntos de entrenamiento para algoritmos de aprendizaje automatizado.* Universidad Central de Las Villas.
12. CABALLERO, Y. & BELLO, R. 2006. Two new feature selection algorithms with Rough Sets Theory. *IFIP 19th World Computer Congress.* Sydney, Australia.
13. CHANDRASEKARAN, B. 1986. Generic tasks in knowledge-based reasoning: High-level building blocks for expert systems design. *IEEE Expert,* 1**,** 23-30.
14. CHAVEZ, M. 2008. *Modelo de redes bayesianas en el estudio de secuencias genómicas y otros problemas biomédicas.* Doctor en Ciencias Técnicas, Universidad Central "Marta Abreu" de las Villas.
15. DEGANG, C. & DELI, Z. 2014. Structure of feature spaces related to fuzzy similarity relations as kernels. *Fuzzy Sets and Systems,* 237 90–95.
16. DEGANG, C., DELI, Z., JIASHANG, J., CONGXIN, W. & TSANG, E. C. C. 2005. Some notes on equivalent fuzzy sets and fuzzy subgroups. *Fuzzy Sets and Systems* 152**,** 403 – 409.
17. DEMSAR, J. 2006. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research,* 7**,** 1-30.
18. DUBOIS, D. & PRADE, H. 1987. Two fold fuzzy sets and rough sets- Some issues in knowledge representation. *Fuzzy Sets and Systems,* 23**,** 3-18.
19. DUBOIS, D. & PRADE, H. 1990. Rough fuzzy sets and fuzzy rough sets. *International J. General Systems,* 17**,** 191-209.
20. DUBOIS, D. & PRADE, H. 1998. Similarity versus preference in fuzzy set based logics. *International J. General Systems,* 17**,** 440-460.
21. DUCH, W. & GRUDZINSKI, K. 1999. Weighting and selection features. *Proceedings of the Workshop held in Ustron.* Poland: Intelligent Information System.
22. EBERHART, R. C. & SHI, Y. Comparing inertia weights and constriction factors in particle swarm optimization. 2000 Congress on evolutionary computation, 2000. IEEE, 84-88.

23. FERNÁNDEZ, Y., COELLO, L., FILIBERTO, Y., BELLO, R. & FALCON, R. 2014. Learning Similarity Measures from Data with Fuzzy Sets and Particle Swarms. *In:* IEEE (ed.) *11th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE).* Ciudad del Carmen, Campeche, Mexico: IEEE.
24. FILIBERTO, Y. 2012. *Métodos de aprendizaje para dominios con datos mezclados basados en la teoría de los conjuntos aproximados extendida.* Universidad Central de Las Villas.
25. FILIBERTO, Y., BELLO, R., CABALLERO, Y. & FRIAS, M. An analysis about the measure quality of similarity and its applications in machine learning. 4th International Workshop on Knowledge Discovery, Knowledge Management and Decision Support (EUREKA 2013), 2013 Mexico. 130–139.
26. FILIBERTO, Y., BELLO, R., CABALLERO, Y. & FRÍAS, M. 2011a. Algoritmo para el aprendizaje de reglas de clasificación basado en la teoría de los conjuntos aproximados extendida. *Revista DYNA,* 78**,** 62-70.
27. FILIBERTO, Y., BELLO, R., CABALLERO, Y. & LARRUA, R. A method to build similarity relations into extended Rough Set Theory. 10th International Conference on Intelligent Systems Design and Applications (ISDA2010), 2010a Cairo, Egipto.
28. FILIBERTO, Y., BELLO, R., CABALLERO, Y. & LARRUA, R. Using PSO and RST to Predict the Resistant Capacity of Connections in Composite Structures. NICSO, 2010b Berlin Heidelberg Springer, 359–370.
29. FILIBERTO, Y., BELLO, R., CABALLERO, Y. & LARRUA, R. 2011b. Una medida de la teoría de los conjuntos aproximados para sistemas de decisión con rasgos de dominio continuo. *Revista de la Facultad Ingeniería de la Universidad de Antioquia***,** 141-152.
30. FILIBERTO, Y., BELLO, R., CABALLERO, Y. & RAMOS, G. Improving the MLP Learning by Using a Method to Calculate the Initial Weights of the Network Based on the Quality of Similarity Measur. MICAI 2011, 2011 Berlin Heidelberg. Springer-Verlag, 351–362.

31. FU, X., ZHANG, S. & PANG, Z. 2010. A Resource Limited Immune Approach for Evolving Architecture and Weights of Multilayer Neural Network. *Heidelberg Springer,* ICSI 6145**,** 328–337
32. GARCÍA, S. & HERRERA, F. 2008. An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. *Journal of Machine Learning Research,* 9**,** 2677-2694.
33. GARCÍA, S. & HERRERA, F. 2009. Evolutionary under-sampling for classification with imbalanced data sets: Proposals and taxonomy. *Evolutionary Computation,* 17**,** 275-306.
34. GARCÍA, S., MOLINA, D., LOZANO, M. & HERRERA, F. 2007. Un estudio experimental sobre el uso de test no paramétricos para analizar el comportamiento de los algoritmos evolutivos en problemas de optimización. *Actas del Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados.* España.
35. GRAU, R. 1994. *Estadística aplicada con ayuda de paquetes de software,* Jalisco, México.
36. HELVACIOGLU, S. & OZEN, E. 2014. Fuzzy based failure modes and effect analysis for yacht system design. *Ocean Engineering* 79 131–141.
37. HEREDIA, J., RODRÍGUEZ, A. & VILALTA, J. 2012. Empleo de la regresión logística ordinal para la predicción del rendimiento académico. *Revista Investigacion Operacional,* 33**,** 252-267.
38. HERNÁNDEZ, S., MARTÍNEZ, J. & CARRASCO, J. 2010. Fast Most Similar Neighbor (MSN) classifiers for Mixed Data. *Computación y Sistemas,* 14**,** 73-85.
39. HOCENSKI, Z., ANTUNOVIC, M. & FILKO, D. 2008. Accelerated Gradient Learning Algorithm for Neural Network Weights Update. *Heidelberg Springer,* LINAI 5177**,** 49-56.
40. HOCHBERG, Y. 1998. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika,* 75**,** 811-818.
41. HOLM, S. 1979. A simple sequentially rejective multiple test procedure. *Journal of Statistics,* 6**,** 65-70.

42. HOLLAND, B. S. & COPENHAVER, M. D. 1987. An improved sequentially rejective Bonferroni test procedure. *Biometrics,* 43**,** 417-423.
43. HOMMEL, G. 1988. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika,* 75**,** 383-386.
44. HUSSAIN, M. 2010. *Fuzzy Relation.* Master of Science in Mathematical Modelling and Simulation, Institute of Technology School of Engineering.
45. IGNJATOVIC, J., CIRICA, M., ŠEŠELJA, B. & TEPAVCEVIC, A. 2015. Fuzzy relational inequalities and equations, fuzzy quasi-orders,closures and openings of fuzzy sets. *Fuzzy Sets and Systems,* 260 1–24.
46. IMAN, R. & DAVENPORT, J. 1980. Approximations of the critical region of the friedman statistic. *Communications in Statistics, Part A Theory Methods,* 9**,** 571-595.
47. KAWAHARA, Y., FURUSAWA, H. & MORI, M. 1999. Categorical representation theorems of fuzzy relations. *Information Sciences,* 119 235-251.
48. KENNEDY, J. & EBERHART, R. 1995. Particle swarm optimization. *IEEE International Conference on Neural Networks.* IEEE.
49. KOMOROWSKI, J. & PAWLAK, Z. 1999a. Rough Fuzzy Hybridization: A new trend in decision-making. *In:* SPRINGER (ed.) *Rough Sets: A tutorial.*
50. KOMOROWSKI, J. & PAWLAK, Z. 1999b. Rough Sets: A tutorial. *Rough Fuzzy Hybridization: A new trend in decision-making. Springer***,** 3-98.
51. KOMOROWSKI, J., PAWLAK, Z., POLKOWSKI, L. & SKOWRON, A. 1996. Rough Sets: A Tutorial. *Fundamenta Informaticae,* 17**,** 105-115.
52. LI, Z. & CUI, R. 2015. Similarity of fuzzy relations based on fuzzy topologies induced by fuzzy rough approximation operators. *Information Sciences* 305 219–233.
53. LUAN, J. 2002. Data Mining and Knowledge Management in Higher Education-Potential Applications. *Annual Forum for the Association for Institutional Research.* Toronto, Canadá.
54. MITCHELL, T. 1997. *Machine Learning*, McGraw-Hill Science.
55. MITRA, P., MITRA, S. & PAL, S. 1999. Modular Rough Fuzzy MLP: Evolutionary Design. *Heidelberg Springer,* RSFDGrC 1711**,** 128-137.

56. MITRA, S. & PAL, S. K. 2005. Fuzzy sets in pattern recognition and machine intelligence. *Fuzzy Sets and Systems,* 156 381 – 386.
57. MORA, J., MORALES, G. & BARRERA, R. 2008. Evaluación del clasificador basado en los k vecinos más cercanos para la localización de la zona en falla en los sistemas de potencia *Ingeniería e Investigación,* 28**,** 25-28.
58. MORI, N., TAKEDA, M. & MATSUMOTO, K. 2005. A statistical comparison study between Genetic Algorithms and bayesian Optimization Algorithms. *Progress of Theoretical Physics Supplement,* 157**,** 353-356.
59. NGUYEN, H. S. 2011. Rough sets and fuzzy sets in natural computing. *Theoretical Computer Science* 412 5816–5819.
60. OVCHINNIKOV, S. 2002. Numerical representation of transitive fuzzy relations. *Fuzzy Sets and Systems,* 126 225–232.
61. PAWLAK, Z. 1984. Rough sets. *International Journal of Computer and Information Sciences,* 11**,** 341-356.
62. PEDRYCZ, W. 1997. Fuzzy sets in pattern recognition: Accomplishments and challenges. *Fuzzy Sets and Systems* 90 171-176.
63. PERALTA, V. 2004. Evaluating and Enforcing Data Freshness in Data Information Systems. *In:* REPÚBLICA, U. D. L. (ed.). Uruguay.
64. QIAN, Y., WANG, Q., CHENG, H., LIANG, J. & DANG, C. 2015. Fuzzy-rough feature selection accelerator. *Fuzzy Sets and Systems* 258**,** 61-78.
65. RAMAKRISHNAN, R. & RAO, C. J. M. 1992. Representations of fuzzy relations with an application to group decision making. *Fuzzy Sets and Systems,* 49**,** 143-150.
66. RICH, E. 1988. *Inteligencia artificial*.
67. RIZA, L. S., JANUSZ, A., BERGMEIR, C., CORNELIS, C., HERRERA, F., S´LEZAK, D. & BENÍTEZ, J. M. 2014. Implementing algorithms of rough set theory and fuzzy rough set theory in the R package ‘‘RoughSets’’. *Information Sciences* 287 68–89.
68. ROCHA, M., CORTEZ, P. & NEVES, J. 2005. Simultaneous Evolution of Neural Network Topologies and Weights for Classification and Regression. *Springer, IWANN* 3512**,** 59-66.

69. ROM, D. M. 1990. A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika,* 77**,** 663-665.
70. ROYCHOWDHURY, P. & SHUKLA, K. K. 2003. Incorporating fuzzy concepts along with dynamic tunneling for fast and robust training of multilayer perceptrons. *Neurocomputing,* 50 319 – 340.
71. RUSELL, S. & NORVIG, P. 1995. *Artificial intelligence: a modern approach*.
72. SHESKIN, D. 2003b. *Handbook of parametric and nonparametric statistical procedures, chapman & hall*.
73. SHI, Y. & EBERHART, R. A modified particle swarm optimizer. IEEE World Congress Computational Intelligence, Evolutionary Computation, 1998. IEEE, 69-73.
74. SHI, Y. & EBERHART, R. Empirical study of particle swarm optimization. Congress on evolutionary computation 1999. IEEE, 1945-1950.
75. SKOWRON, A. 1993. Boolean reasoning for decision rules generation. In: Komorowski. *Lectures Notes in Artificial Intelligence Springer,* 689**,** 295-305.
76. SUN, B., MAB, W. & CHEN, D. 2014. Rough approximation of a fuzzy concept on a hybrid attribute information system and its uncertainty measure. *Information Sciences,* 284 60–80.
77. VERDEGAY, J. L., YAGER, R. R. & BONISSONE, P. P. 2008. On heuristics as a fundamental constituent of soft computing. *Fuzzy Sets and Systems* 159**,** 846-855.
78. VILLUENDAS, Y., GARCÍA, M. & RUIZ-SHULCLOPER, J. 2010. Selección combinada de rasgos y objetos par a el mejoramiento de clasificadores NN. *In:* AZUL, S. (ed.). Cuba: CENATAV.
79. VILLUENDAS, Y., RODRÍGUEZ, M., MEDINA, M., GARCÍA, M. & RUIZ-SHULCLOPER, J. 2008. Métodos para acelerar la búsquedade los k vecinos más cercanos utilizando distancias y disimilaridades. *In:* AZUL, S. (ed.). Cuba: CENATAV.
80. WANG 1997. New similarity measures on fuzzy sets and on elements. *Fuzzy Sets and Systems,* 85**,** 305-309.

81. WANG, R. & STRONG, D. 1996. Beyond accuracy: What data quality means to data consumers. *Journal on Management of Information Systems,* 12**,** 5-34.
82. WANG, X., CAO, X., WU, C. & CHEN, J. 2013. Indicators of fuzzy relations. *Fuzzy Sets and Systems,* 216 91 – 107.
83. WANG, X. & XUE, Y. 2014. Traces and property indicators of fuzzy relations. *Fuzzy Sets and Systems,* 246 78–90.
84. WILTAKER, J. 1990. Graphical Models in Applied Multivariate Statistical. *Probabilit & Statistics*. Wiley Series.
85. WUA, W., MI, J. & ZHANG, W. 2003. Generalized fuzzy rough sets. *Information Sciences,* 151 263–282.
86. YANG, M. S. & SHIH, H. M. 2001. Cluster analysis based on fuzzy relations. *Fuzzy Sets and Systems,* 120**,** 197–212.
87. YAO, Y. Y. 1998. A comparative study of fuzzy sets and rough sets. *Joun al of Information Sciences,* 109 227-242.
88. ZADEH, L. A. 1971. Similarity Relations and Fuzzy Orderings. *Information Sciences,* 3**,** 177-200.
89. ZEGER, S. L. & LIANG, K. 1986. Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics***,** 121-130.
90. ZENG, A., LI, T., LIU, D., ZHANG, J. & CHEN, H. 2015. A fuzzy rough set approach for incremental feature selection on hybrid information systems. *Fuzzy Sets and Systems,* 258 39–60.
91. ZHANG, W., MA, D., WEI, J.-J. & LIANG, H.-F. 2014. A parameter selection strategy for particle swarm optimization based on particle positions. *Expert Systems with Applications,* 41**,** 3576–3584.
92. ZHAO, X. R. & HU, B. Q. 2015. Fuzzy and interval-valued fuzzy decision-theoretic rough set approaches based on fuzzy probability measure. *Information Sciences* 298 534–554.

Anexos

Anexo 1. Descripción de los conjuntos de datos internacionales que se usaron en los experimentos

Tabla T1.1. Descripción de los conjuntos de datos internacionales que se usaron en los experimentos.

Conjuntos de datos	Número de instancias	Número de atributos	Tipo de dato del atributo decisión
baskball	96	4	numérica
bodyfat	252	14	numérica
detroit	13	13	numérica
diabetes_numeric	43	2	numérica
elusage	55	2	numérica
fishcatch	158	7	numérica
pollution	60	15	numérica
pwLinear	200	10	numérica
pyrim	7	27	numérica
sleep	51	7	numérica
vineyard	52	3	numérica
schlvote	37	5	numérica
Tae	151	5	nominal
Diabetes	768	7	nominal
Iris	150	3	nominal
Hepatitis	155	19	nominal
zoo	101	17	nominal
bridges-version1	107	12	nominal
Biomed	194	8	nominal
Schizo	104	14	nominal
soybean-small	47	35	nominal
Cars	392	7	nominal
heart-statlog	270	13	nominal
glass	214	9	nominal
liver-disorders	375	6	nominal
ecoli	336	7	nominal
new thyroid	215	5	nominal
wine	178	13	nominal
pima	768	8	nominal
balance-scale	625	4	nominal
haberman	306	3	nominal

Anexo 2. Resultados experimentales del efecto del uso de PSO+RST+FUZZY en el algoritmo k-NN, relacionados con el experimento 1

Tabla T2.1. Resultados del error PMD para el algoritmo k-NN para problemas de aproximación de funciones con k=3.

No.	Conjuntos de datos	PMD Mean absolute error			
		Estándar	KNN_{VSM}	PSO+RST	PSO+RST+FUZZY
1	baskball	0.08	0.08	0.07	0.07
2	detroit	36.71	36.71	32.52	30.8
3	elusage	9.02	8.94	8.32	7.782
4	pollution	37.51	42.71	36.38	35.08
5	pwLinear	1.96	2.2	1.7	2.15
6	pyrim	0.06	0.06	0.05	0.05
7	sleep	2.6	2.6	2.41	2.56
8	vineyard	1.95	1.95	1.8	1.66
9	bodyfat	2.56	2.56	1.06	1.26
10	fishcatch	57.02	57.02	40.02	38.52

Figura F2.1. Resultados del error PMD para el algoritmo k-NN para *k*=3, usando los diferentes métodos de cálculo de peso para problemas de aproximación de funciones.

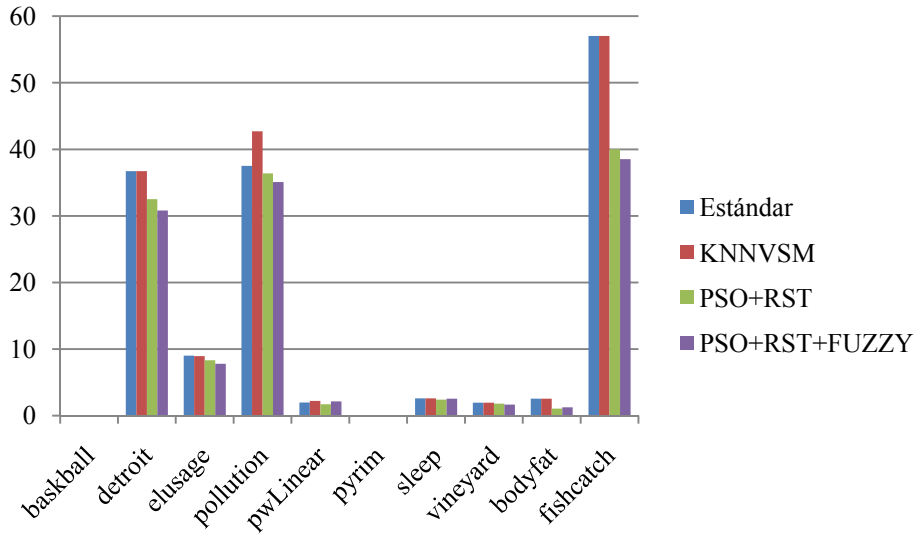

Tabla T2.2. Resultados de la prueba estadística de Friedman para la medida de error PMD para el método K-NN.

Algoritmos	Ranking
PSO+RST+FUZZY	1.5
PSO+RST	1.6
ESTÁNDAR	3.35
KNNVSM	3.55

Iman-Davemport (distribución de F con 3 y 27 grados de libertad): 23.967033

Valor de p computado por el test Iman-Davemport : 0.000000089925

Tabla T2.3. Prueba de Holm para α=0.05 para la medida de error PMD, tomando como control PSO+RST+FUZZY.

i	Algoritmos	$z=(R_0-R_i)/SE$	p	Holm	Hipótesis
3	KNNVSM	3.550704	0.000384	0.016667	Se rechaza
2	ESTÁNDAR	3.204294	0.001354	0.025	Se rechaza
1	PSO+RST	0.173205	0.86249	0.05	Se rechaza

El test de Holm rechaza todas las hipótesis con valor de $p \leq 0.05$.

Anexo 3. Resultados experimentales del efecto del uso de PSO+RST+FUZZY en el algoritmo k-NN, relacionados con el experimento 2

Tabla T3.1. Resultados del coeficiente de correlación para el algoritmo k-NN para k=3, usando los diferentes métodos de cálculo de peso para problemas de aproximación de funciones.

No.	Conjuntos de datos	Coeficiente-Correlación			
		PSO+RST +FUZZY	PSO+RST	KNN_{VSM}	Estándar
1	baskball	0.49	0.6	0.5	0.5
2	bodyfat	0.91	0.98	0.93	0.93
3	detroit	0.76	0.82	0.81	0.81
4	elusage	0.87	0.87	0.87	0.86
5	fishcatch	0.96	0.99	0.97	0.97
6	pollution	0.69	0.66	0.55	0.65
7	pwLinear	0.78	0.87	0.76	0.8
8	pyrim	0.65	0.76	0.72	0.73
9	sleep	0.73	0.77	0.74	0.74
10	vineyard	0.8	0.81	0.79	0.79

Tabla T3.2. Resultados del coeficiente R2 para el algoritmo k-NN para k=3, usando los diferentes métodos de cálculo de peso para problemas de aproximación de funciones.

No.	Conjuntos de datos	Coeficiente-R^2			
		PSO+RST +FUZZY	PSO+RST	KNN_{VSM}	Estándar
1	baskball	0.24	0.36	0.25	0.25
2	bodyfat	0.9	0.96	0.85	0.85
3	detroit	0.59	0.67	0.65	0.65
4	elusage	0.76	0.76	0.75	0.72
5	fishcatch	0.93	0.97	0.92	0.92

6	pollution	0.47	0.44	0.31	0.42
7	pwLinear	0.71	0.76	0.62	0.67
8	pyrim	0.42	0.57	0.52	0.53
9	sleep	0.53	0.6	0.55	0.55
10	vineyard	0.64	0.66	0.63	0.63

Tabla T3.3. Resultados del error estándar, para el algoritmo k-NN para k=3, usando los diferentes métodos de cálculo de peso para problemas de aproximación de funciones.

No.	**Conjuntos de datos**	**Error-Estándar**			
		PSO+RST +FUZZY	**PSO+RST**	**KNN_{VSM}**	**Estándar**
1	baskball	0.09	0.09	0.09	0.09
2	bodyfat	1.69	1.67	3.24	3.24
3	detroit	51.11	43.6	45.07	45.07
4	elusage	11.6	11.89	12.19	12.77
5	fishcatch	90.93	61.66	98.96	98.96
6	pollution	46.44	47.01	52.2	47.94
7	pwLinear	2.82	2.19	2.78	2.59
8	pyrim	0.1	0.08	0.09	0.09
9	sleep	3.33	3	3.18	3.18
10	vineyard	2.66	2.6	2.7	2.7

Anexo 4. Resultados experimentales del efecto del uso de PSO+RST+FUZZY en el algoritmo k-NN, relacionados con el experimento 3

Tabla T4.1. Resultados de la exactitud de la clasificación general, para el algoritmo k-NN con k=1, usando los diferentes métodos de cálculo de peso para problemas de clasificación.

No.	Conjuntos de de datos	***k*=1**				
		PSO+RST +FUZZY	**PSO+RST**	**KNN_{VSM}**	**Estándar**	**Relief**
1	tae	67.3	63.2	58.94	58.94	58.94
2	bridges_version1	62.4	52.3	42.4	32.2	45.2
3	diabetes	98.55	99.4	67.8	68.5	68.2
4	biomed	94.2	99.4	99.4	95.8	97.8
5	iris	94.5	95.8	94.4	94.4	94.4
6	zoo	95	71.4	69.4	69.4	71.4
7	schizo	60.3	75.7	67	48.5	53.5
8	soybean-small	100	100	100	100	100
9	cars	86.15	82.3	86.1	75.2	76
10	heart-statlog	74.4	77.4	77.2	77.5	75.7
11	liver-disorders	66.4	65.8	65.1	65.1	65.7
12	glass	83.51	82.3	78.4	72.1	73.9

Tabla T4.2. Resultados de la exactitud de la clasificación general, para el algoritmo k-NN con k=2, usando los diferentes métodos de cálculo de peso para problemas de clasificación.

No.	Conjuntos de de datos	k=2				
		PSO+RST +FUZZY	PSO+RST	KNN_{VSM}	Estándar	Relief
1	tae	61.4	55.9	51.66	51.66	51.66
2	bridges_version1	56.7	48.1	40.9	29.4	39.3
3	diabetes	98.55	99.4	67.8	68.5	68.2
4	biomed	94.2	99.4	99.4	95.8	97.8
5	iris	93.5	96.5	94.4	94.4	95.1
6	zoo	95.2	71.4	66.4	69.4	71.4
7	schizo	60.3	75.7	67	48.5	53.5
8	soybean-small	97.5	100	100	100	100
9	cars	82.6	79.2	82.7	75.3	75
10	heart-statlog	74.4	77.4	77.2	77.5	75.7
11	liver-disorders	66.4	65.8	65.1	65.1	65.7
12	glass	80.5	79.2	76.2	71.6	71.2

Tabla T4.3. Resultados de la exactitud de la clasificación general, para el algoritmo k-NN con k=3, usando los diferentes métodos de cálculo de peso para problemas de clasificación.

No.	Conjuntos de de datos	*k*=3				
		PSO+RST +FUZZY	PSO+RST	KNN_{VSM}	Estándar	Relief
1	tae	57.4	52	48.34	48.34	47.68
2	bridges_version1	52.5	48	35.2	32.4	37.9
3	diabetes	96.62	98.9	71.1	71.7	71.2
4	biomed	96.55	98.9	99.4	94.1	97.3
5	iris	95.2	96.5	93.7	94.4	95.1
6	zoo	93	70.4	64.4	64.4	70.4
7	schizo	56.7	72	72.7	55.3	51.8
8	soybean-small	97.5	100	100	97.5	100
9	cars	86.9	80.1	86.3	75.9	76.8
10	heart-statlog	81.2	82.8	81.6	80.1	80.2
11	liver-disorders	60.6	65.8	66.4	65.3	65.3
12	glass	81.25	77.9	76.6	69.7	69.1

Figura F4.1. Resultados de la exactitud de la clasificación general, para el algoritmo k-NN para k=3, usando los diferentes métodos de cálculo de peso para problemas clasificación.

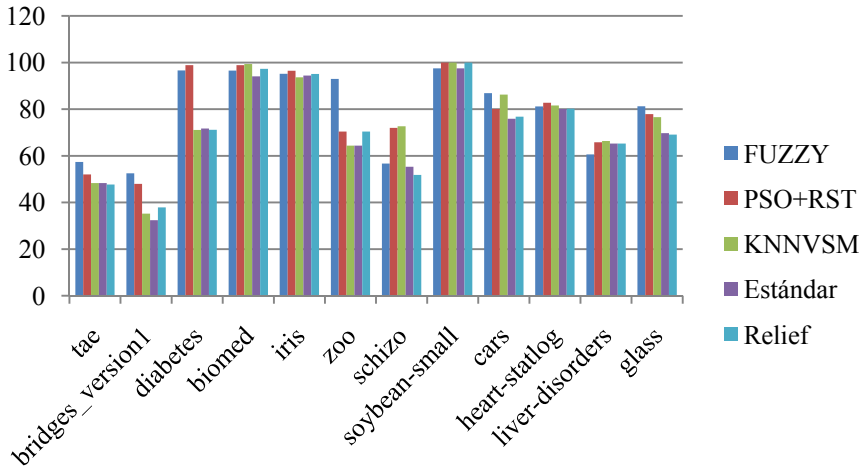

Tabla T4.4. Resultados de la prueba estadística de Friedman para la exactitud de la clasificación general con k-NN para *k*=3.

Algoritmos	Ranking
PSO+RST	1.875
PSO+RST+FUZZY	2.375
KNNVSM	2.8333
RELIEF	3.6667
ESTÁNDAR	4.25

Iman-Davemport (distribución de F con 4 y 44 grados de libertad): 6.435333

Valor de p computado por el test Iman-Davemport: 0.000362389097

Tabla T4.5. Prueba de Holm para α=0.05 para la exactitud de la clasificación general con k-NN para k=3, tomando como control PSO+RST.

i	Algoritmos	$z=(R_0-R_i)/SE$	p	Holm	Hipótesis
4	ESTÁNDAR	3.679334	0.000234	0.0125	Se rechaza
3	RELIEF	2.775638	0.005509	0.016667	Se rechaza
2	KNNVSM	1.484644	0.137638	0.025	Se rechaza
1	PSO+RST+FUZZY	0.774597	0.438578	0.05	No se rechaza

El test de Holm rechaza todas las hipótesis con valor de p ≤0.025.

Anexo 5. Resultados experimentales del efecto del uso de PSO+RST+FUZZY en el algoritmo MLP, para problemas de aproximación de funciones, relacionados con el experimento 4

Tabla T5.1. Resultados del error PMD para el algoritmo MLP para problemas de aproximación de funciones.

No.	Conjuntos de datos	PMD				
		PSO+RST +FUZZY	PSO+RST	KNN_{VSM}	Estándar	MLP-AL
1	baskball	0.073	0.08	0.08	0.08	0.08
2	bodyfat	0.52	0.52	0.54	0.54	2.13
3	diabetesnumeric	0.53	0.53	0.53	0.53	0.53
4	elusage	10.45	10.37	10.77	10.67	10.76
5	pollution	39.02	43.36	42.41	42.41	47.39
6	pwLinear	1.642	1.67	2.07	2.07	1.68
7	pyrim	0.072	0.08	0.09	0.09	0.08
8	sleep	2.71	2.89	3.17	3.17	3.28
9	vineyard	2.21	2.1	2.41	2.41	2.09
10	schlvote	3.96	3.32	9.92	9.92	3.32
11	detroit	34.99	25.65	32.78	32.78	33.06
12	fishcatch	33.03	33.99	47.93	47.93	38.6

Figura F5.1. Resultados del error PMD para el algoritmo MLP, usando los diferentes métodos de cálculo de peso para problemas de aproximación de funciones.

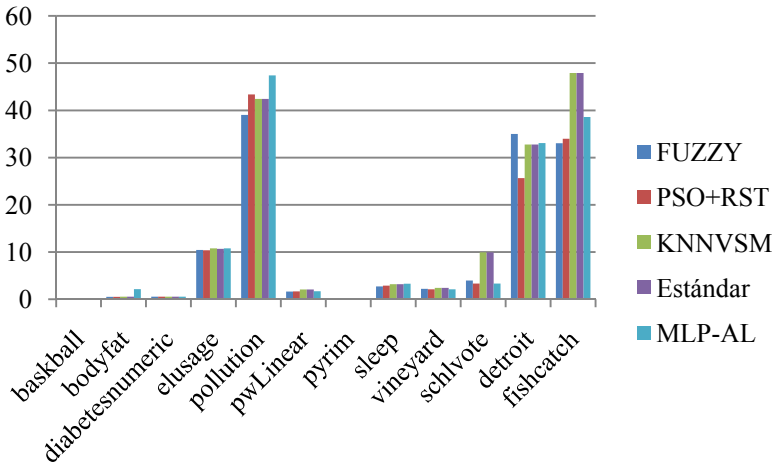

Tabla T5.2. Resultados de la prueba estadística de Friedman para la medida de error PMD para el MLP.

Algoritmos	Ranking
PSO+RST+FUZZY	1.9583
PSO+RST	2.1667
MLP-AL	3.375
ESTÁNDAR	3.6667
KNNVSM	3.8333

Iman-Davemport (distribución de F con 4 y 44 grados de libertad): 4.847924.

Valor de p computado por el test Iman-Davemport: 0.002508524983

Tabla T5.3. Prueba de Holm para α=0.025 para la medida de error PMD, tomando como control PSO+RST+FUZZY.

i	Algoritmos	$z=(R_0-R_i)/SE$	p	Holm	Hipótesis
4	KNNVSM	2.904738	0.003676	0.0125	Se rechaza
3	ESTÁNDAR	2.646539	0.008132	0.016667	Se rechaza
2	MLP-AL	2.194691	0.028186	0.025	Se rechaza
1	PSO+RST	0.322749	0.746886	0.05	No se rechaza

El test de Holm rechaza todas las hipótesis con valor de p ≤0.025.

Anexo 6. Resultados experimentales del efecto del uso de PSO+RST+FUZZY en el algoritmo MLP, relacionados con el experimento para problemas de aproximación de funciones, relacionados con el experimento 5

Tabla T6.1. Resultados del coeficiente de correlación para el algoritmo MLP, usando los diferentes métodos de cálculo de peso para problemas de aproximación de funciones.

No.	Conjuntos de datos	Coeficiente de correlación				
		PSO+RST FUZZY	PSO+RST	KNN_{VSM}	Estándar	MLP-AL
1	baskball	0.58	0.52	0.47	0.47	0.46
2	detroit	0.84	0.89	0.85	0.85	0.83
3	bodyfat	0.98	1	1	1	1
4	diabetes_numeric	0.49	0.51	0.48	0.48	0.48
5	elusage	0.86	0.84	0.81	0.81	0.83
6	pollution	0.78	0.71	0.64	0.64	0.55
7	pwLinear	0.89	0.9	0.84	0.84	0.89
8	pyrim	0.63	0.69	0.45	0.45	0.59
9	sleep	0.71	0.68	0.63	0.63	0.62
10	vineyard	0.8	0.8	0.67	0.67	0.8
11	schlvote	0.41	0.43	0.2	0.2	0.15
12	fishcatch	0.99	0.99	0.99	0.99	0.99

Tabla T6.2. Resultados del coeficiente R^2 para el algoritmo MLP, usando los diferentes métodos de cálculo de peso para problemas de aproximación de funciones.

No.	Conjuntos de datos	Coeficiente R^2				
		PSO+RST FUZZY	PSO+RST	KNN_{VSM}	Estándar	MLP-AL
1	baskball	0.33	0.27	0.21	0.21	0.21
2	detroit	0.75	0.79	0.73	0.73	0.69
3	bodyfat	0.97	0.97	0.97	0.97	0.97

4	diabetes_numeric	0.25	0.26	0.23	0.23	0.23
5	elusage	0.69	0.71	0.66	0.66	0.7
6	pollution	0.41	0.5	0.41	0.41	0.3
7	pwLinear	0.8	0.79	0.65	0.65	0.77
8	pyrim	0.45	0.48	0.2	0.2	0.35
9	sleep	0.51	0.47	0.4	0.4	0.38
10	vineyard	0.61	0.64	0.45	0.45	0.63
11	schlvote	0.04	0.18	0.04	0.04	0.02
12	fishcatch	0.96	0.98	0.96	0.96	0.98

Tabla T6.3. Resultados del error estándar, para el algoritmo MLP, usando los diferentes métodos de cálculo de peso para problemas de aproximación de funciones.

No.	**Conjuntos de datos**	**Error estándar**				
		PSO+RST FUZZY	**PSO+RST**	**KNN_{VSM}**	**Estándar**	**MLP-AL**
1	baskball	0.09	0.09	0.1	0.1	0.1
2	detroit	33.55	35.06	39.74	39.74	42.64
3	bodyfat	1.39	1.33	1.45	1.45	1.41
4	diabetes_numeric	0.67	0.63	0.64	0.64	0.64
5	elusage	13.42	13.05	14.1	14.07	13.36
6	pollution	49.27	44.35	48.35	48.35	52.46
7	pwLinear	1.99	2.03	2.64	2.64	2.02
8	pyrim	0.1	0.09	0.12	0.12	0.1
9	sleep	3.52	3.48	3.71	3.71	3.7
10	vineyard	2.83	2.65	3.29	3.29	2.69
11	schlvote	46.558	72.14	124.41	124.41	78.88
12	fishcatch	75.06	50.67	75.07	75.07	53.57

Anexo 7. Resultados experimentales del efecto del uso de PSO+RST+FUZZY en el algoritmo MLP, en cuanto a exactitud de la clasificación general, relacionados con el experimento 6

Tabla T7.1. Resultados de la exactitud de la clasificación general, para el algoritmo MLP usando los diferentes métodos de cálculo de peso para problemas de clasificación.

No.	Conjuntos de datos	PSO+RST+ FUZZY	PSO+RST	KNNVSM	Estándar	MLP-AL	RELIEF
1	tae	59.12	58.94	55.63	49.01	54.3	54.97
2	diabetes	76.43	74.8	74.22	76.69	75.39	74.74
3	iris	97.33	97.9	96.67	95.33	97.33	98
4	hepatitis	84.45	82.01	81.29	78.06	80	79.35
5	postoperative-patient-data	57.77	57.78	53.33	54.44	55.56	55.56
6	zoo	97.66	96.04	40.59	73.27	94.29	75.25
7	bridges-version1	68.03	71.43	41.9	41.9	69.52	60
8	biomed	92.98	90.7	82.99	83.51	86.08	83.51
9	schizo	69.54	68.27	62.5	63.46	65.38	63.46
10	soybean-small	100	100	76.6	78.72	100	74.47
11	cars	80.58	80.3	71.17	71.17	78.06	71.17
12	heart-statlog	83.33	81.85	80.37	80.37	78.15	80.37
13	haberman	74.17	74.7	73.52	75.47	73.52	73.51
14	liver-disorders	69.29	67.5	67.25	67.2	66.39	67.26

Figura F7.1. Resultados de la exactitud de la clasificación general, para el algoritmo MLP usando los diferentes métodos de cálculo de peso para problemas clasificación.

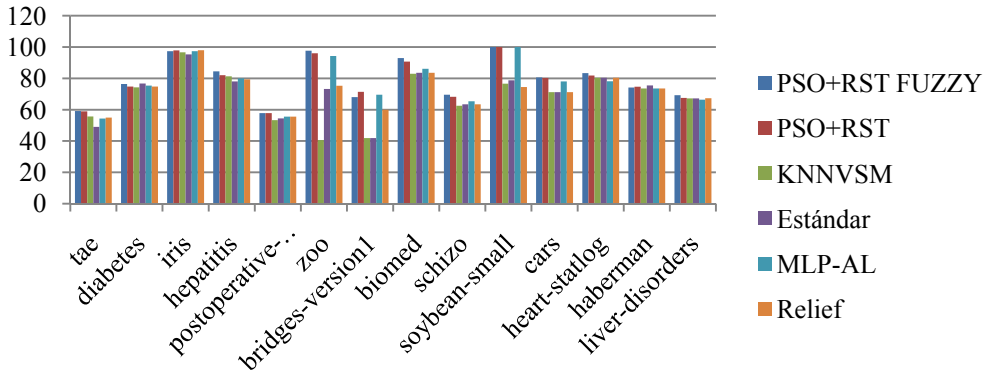

Tabla T7.2. Resultados de la prueba estadística de Friedman para la exactitud de la clasificación general con MLP.

Algoritmos	Ranking
PSO+RST+FUZZY	1.6786
PSO+RST	2
MLP-AL	3.6786
RELIEF	4.25
ESTÁNDAR	4.4643
KNNVSM	4.9286

Iman-Davemport (distribución de F con 5 y 65 grados de libertad): 14.189024.

Valor de p computado por el test Iman-Davemport: 0.000000002666

Tabla T7.3. Prueba de Holm para α=0.5 para la exactitud de la clasificación general con MLP, tomando como control PSO+RST.

i	Algoritmos	$z=(R_0-R_i)/SE$	p	Holm	Hipótesis
5	KNNVSM	4.596194	0.000004	0.01	Se rechaza
4	ESTÁNDAR	3.939595	0.000082	0.0125	Se rechaza
3	RELIEF	3.636549	0.000276	0.016667	Se rechaza
2	MLP-AL	2.828427	0.004678	0.025	Se rechaza
1	PSO+RST	0.454569	0.64942	0.05	Se rechaza

El test de Holm rechaza todas las hipótesis con valor de p ≤0.05.

Anexo 8. Resultados de la aplicación de PSO+RST+FUZZY para problemas de Rendimiento Académico e Ingeniería Civil

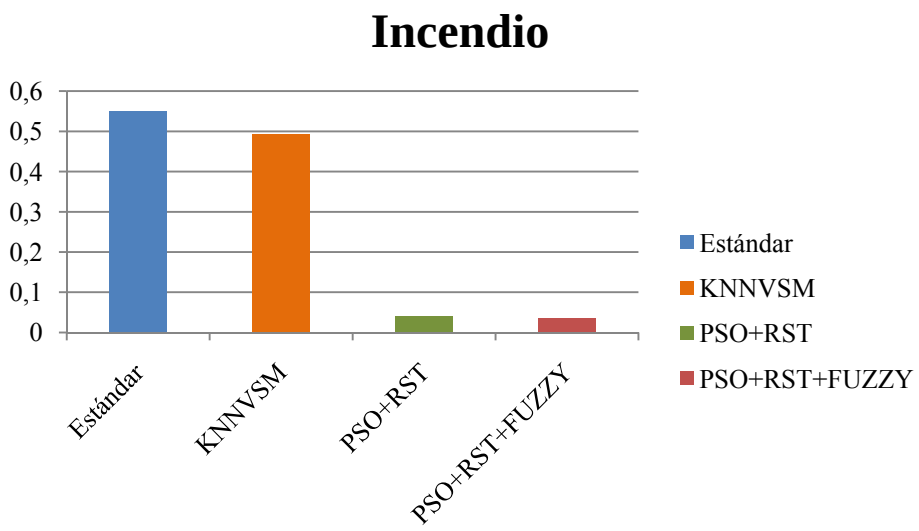

Figura F8.1. Resultados del error PMD para el algoritmo k-NN para k=1, usando los diferentes métodos de cálculo de peso para los conjuntos de incendio.

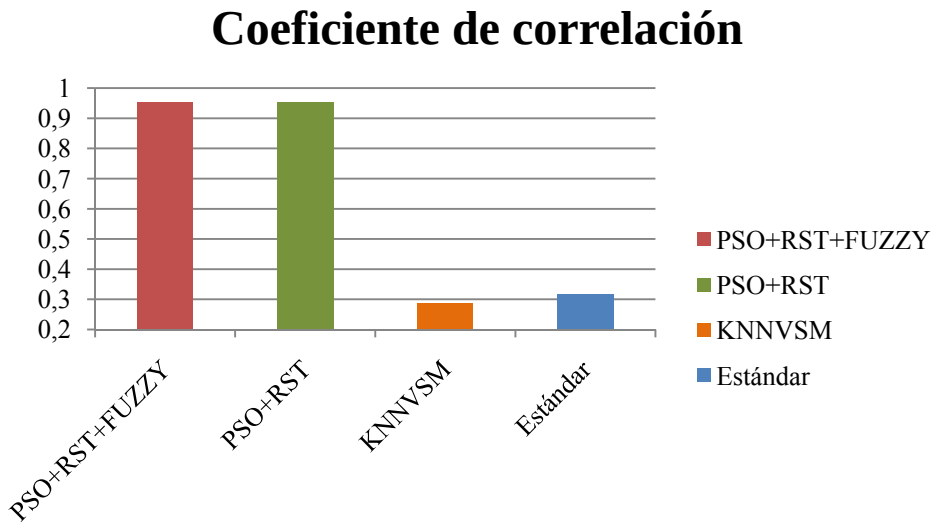

Figura F8.2. Resultados coeficiente de correlación para el algoritmo k-NN para k=3, usando los diferentes métodos de cálculo de peso para los conjuntos de incendio.

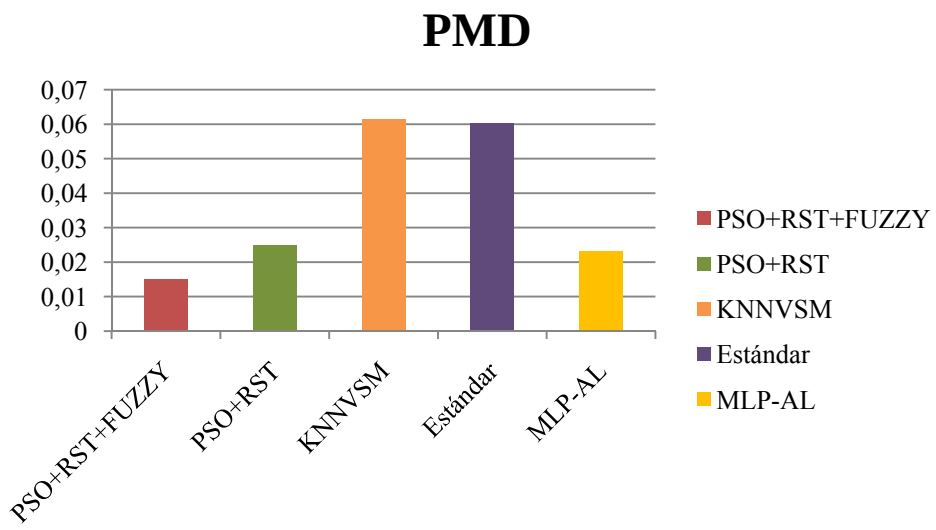

Figura F8.3. Resultados del error PMD para el algoritmo MLP, usando los diferentes métodos de cálculo de peso para los conjuntos de incendio.

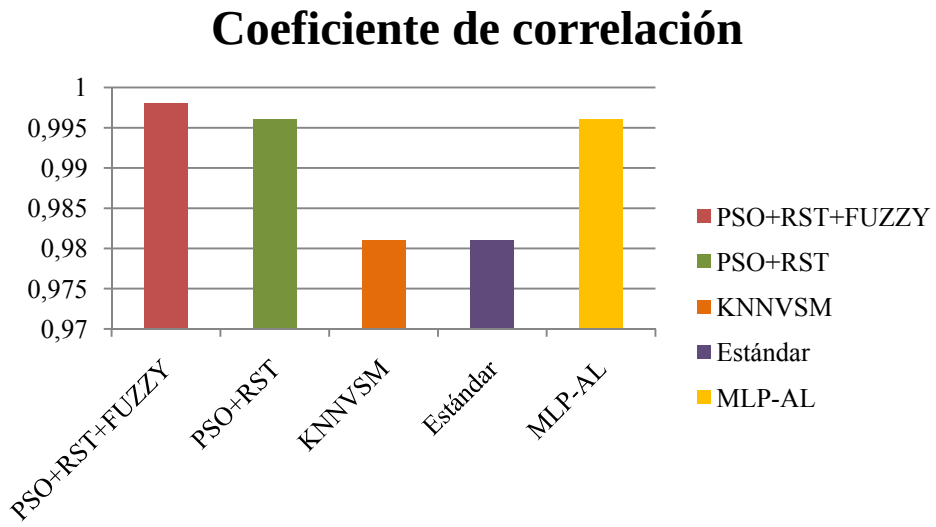

Figura F8.4. Resultados coeficiente de correlación para el algoritmo MLP usando los diferentes métodos de cálculo de peso para los conjuntos de incendio.