

Una medida de la teoría de los conjuntos aproximados para sistemas de decisión con rasgos de dominio continuo

A measure in the rough set theory to decision systems with continuo features

Yaima Filiberto¹, Rafael Bello², Yailé Caballero^{*1}, Rafael Larrua³

¹ Departamento de Computación. Universidad de Camagüey. Carretera Circunvalación Norte km 5 ½. Camagüey, Cuba.

² Departamento de Ciencia de la Computación. Universidad Central de Las Villas. Carretera a Camajuani km 5 Santa Clara, CP 54838, Cuba.

³ Centro de Estudio de Estructuras. Universidad de Camagüey. Carretera Circunvalación Norte km 5 ½. Camagüey, Cuba.

(Recibido el 23 de agosto de 2010. Aceptado el 20 de junio de 2011)

Resumen

La Teoría de los conjuntos aproximados ofrece diversas medidas y técnicas para el análisis de datos, especialmente en el caso de problemas donde el rasgo de decisión tiene dominio discreto. En este artículo se propone una medida y métodos basados en ella que permiten extender la aplicabilidad de esta teoría para el caso de problemas donde los rasgos tienen dominio continuo, especialmente el rasgo de decisión. Los estudios experimentales realizados muestran su efectividad.

----- *Palabras clave:* Selección de rasgos, aproximación de funciones, teoría de los conjuntos aproximados

Abstract

The Rough set theory provides several measures and techniques to data analysis, especially in the case of problems in which the decision feature has a discrete domain. In this paper a measure and methods are proposed which allow extending this theory to the case of problems in which the features have a continuo domain, especially the decision feature. The experimental studies show their affectivity.

----- *Keywords:* Feature selection, function approximation, rough set theory

* Autor de correspondencia: teléfono: + 53 + 3 + 229 37 99, correo electrónico: yailec@yahoo.com. (Y. Caballero)

Introducción

La teoría de Los Conjuntos Aproximados (Rough Set Theory, RST) fue propuesta por Pawlak en 1982 [1]. La filosofía de los conjuntos aproximados se basa en la suposición que alguna información es asociada con cada objeto del universo de discurso [2]. Entre las ventajas de la RST para el análisis de datos está que esta solo se basa en los datos originales y no necesita cualquier información externa; no es necesaria ninguna suposición acerca de los datos; así, como que sirve para analizar tanto rasgos cualitativos como cuantitativos [3, 4].

El uso de reductos (conjunto mínimo de atributos que preserva la partición del universo) en la selección de rasgos ha sido estudiado por diversos autores como [5-13]. La selección de rasgos es útil en diferentes tareas computacionales, por ejemplo, en el proceso de aprendizaje automatizado. Una apropiada representación del espacio de aprendizaje por la selección de atributos relevantes es un asunto crucial en el aprendizaje [14-17]. Por lo general, no todos los rasgos que describen el ejemplo son relevantes en el proceso de clasificación y algunos de ellos son irrelevantes. Rasgos irrelevantes aumentan la complejidad de proceso de aprendizaje y disminuyen la calidad del conocimiento inducido [18-20].

La selección de rasgos es un caso particular de un problema más general de selección de subconjuntos en el cual se maximiza algún criterio adoptado. Los métodos de selección de rasgos buscan entre los subconjuntos de rasgos y tratan de encontrar el mejor subconjunto entre los $2^N - 1$ subconjuntos candidatos según alguna medida de evaluación, donde la N denota el número de rasgos. Cada estado representa un subconjunto de rasgos posibles en el espacio de búsqueda.

Todos los métodos de selección de rasgos contienen dos componentes importantes: una función de evaluación y un algoritmo de generación de subconjuntos. La función de evaluación da una medida de calidad (bondad) de un subconjunto producido por algún procedimiento de generación. Un subconjunto siempre es óptimo con res-

pecto a una determinada función de evaluación. Hay varias categorías de funciones de evaluación, tales como medidas de distancia, las medidas de información (por ejemplo Entropía), las medidas de dependencia, las medidas de consistencia, y la medida por ciento del error del clasificador [20]. El algoritmo de generación de subconjunto se basa en un método de búsqueda. Las estrategias de búsqueda son importantes porque el proceso de selección de rasgos puede consumir mucho tiempo y una búsqueda exhaustiva del subconjunto “óptimo” no es práctico, incluso para los problemas de tamaño moderado [21].

Usualmente, 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. Otras medidas útiles en los procesos de selección de rasgos definidas en RST son una para computar la dependencia entre rasgos y otra para computar el peso de los rasgos. Estas medidas muestran la aplicabilidad de la RST clásica en el análisis de datos; pero, todas ellas se definen en el caso de sistemas de decisión donde el dominio de los rasgos es discreto.

Con el fin de considerar los rasgos de predicción continuos se han desarrollado algunas extensiones de la RST clásica (llamada RST extendida), como la RST basada en las relaciones de similitud. Pocas alternativas consideran el caso de rasgos de decisión con dominio continuo [2, 22].

En este artículo 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. Se analiza el uso de esta medida para construir la relación de semejanza en la RST extendida; el método propuesto también permite para calcular el peso de los rasgos.

En este artículo se presentan los conceptos y medidas fundamentales de la RST, se analiza la extensión de la misma para el caso de rasgos con dominios continuos, y se muestra cómo el empleo de relaciones de similitud en la RST lleva a la necesidad de calcular los pesos de los rasgos

involucrados en las funciones de similaridad. Se introduce además una nueva medida. En Se muestra el empleo de esta medida en un método para el cálculo de los pesos de los rasgos de un sistema de decisión continuo para el problema de aproximación de funciones, realizándose un estudio experimental del método.

La teoría de los conjuntos aproximados: fundamentos

En la Teoría de los Conjuntos Aproximados la información es representada por una tabla donde cada fila representa un objeto y cada columna representa un rasgo. Esta tabla se llamada Sistema de Información; más formalmente, es un par $S = (U, A)$, donde U es un conjunto finito no vacío de objetos llamado Universo y A es un conjunto finito no vacío de atributos. Un Sistema de Decisión es cualquier sistema de información de la forma $DS = (U, A \cup \{d\})$, donde $d \notin A$ es el atributo de decisión. Los componentes principales de la RST son el Sistema de Información (o el Sistema de Decisión) y una relación de inseparabilidad (indiscernibility relation). 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 [1, 23].

Sea $DS = (U, A \cup \{d\})$ un sistema de decisión y $B \subseteq A$ y $X \subseteq U$. B define una relación de equivalencia y el subconjunto X es un concepto en el universo U . 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 $B_*X = \{x \in U : [x]_B \subseteq X\}$ y $B^*X = \{x \in U : [x]_B \cap X \neq \emptyset\}$ respectivamente.

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 . Basada en las aproximaciones la RST ofrece la medida calidad de la clasificación, definida por la

expresión (1). Sea una partición $Y = \{Y_1, \dots, Y_n\}$ de U de acuerdo a los valores del rasgo de decisión d (clases), donde los subconjuntos Y_p son llamados clase de decisión.

$$\gamma_B(Y) = \frac{\sum_{i=1}^n |B_*Y_i|}{|U|} \tag{1}$$

El coeficiente $\gamma_B(Y)$ expresa el porciento de objetos que pueden ser clasificados correctamente en las clases Y_1, \dots, Y_n usando los rasgos en B solamente. Esta última medida puede ser usada para evaluar la consistencia de un sistema de decisión y mejorar la selección de rasgos. Un sistema de decisión es inconsistente si objetos inseparables pertenecen a clases de decisión diferentes. En el caso del problema de selección de rasgos, esta medida puede ser usada para evaluar la calidad de un subconjunto de rasgos, [5-13]. Un reducto es un conjunto minimal de atributos $B \subseteq A$ tal que $IND(B) = IND(A)$, es decir ambos generan la misma partición del universo U , lo cual es equivalente a $\gamma_B(Y) = \gamma_A(Y)$.

Pero si el dominio del rasgo de decisión no es discreto, no se puede realizar una partición del universo en clases de decisión, de modo que no es posible usar la medida calidad de la clasificación para medir la consistencia del sistema ni para realizar la selección de rasgos basado en el concepto de reducto. La medida que se propone en las secciones siguientes es una alternativa de solución para esta restricción.

La teoría de los conjuntos aproximados: extensiones

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 $a_i(x) = a_i(y)$, para todo rasgo i en B , donde $a_i(x)$ denota el valor del rasgo i en el objeto x . Definida de esta forma, R es una relación de equivalencia.

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, al medir la temperatura corporal de dos personas. 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. Es necesario usar otros tipos de relaciones de inseparabilidad entre los objetos del universo U . Emplear una relación de semejanza en lugar de una relación de equivalencia es más adecuado en estos casos. Reemplazando la relación de equivalencia por una relación binaria más débil, se obtiene una extensión del enfoque clásico de la RST. Algunas extensiones se estudian en [23-29].

Esto se logra extendiendo 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 . 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 $R(x) = \{y \in U : yRx\}$. 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) [5-30], donde la relación $R \subseteq X \times U$ 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. Un cubrimiento del universo U es una familia de subconjuntos no vacíos cuya unión es igual al universo. Una partición de U 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, tales como [29, 31-33]; algunos de ellos orientados a construir una generalización de la RST basada en cubrimientos. Por eso, es muy interesante e importante encontrar una adecuada relación de semejanza para un sistema de decisión dado.

Un ejemplo de una extensión de la RST basada en relaciones de semejanza fue presentada por R. Slowinski y D. Vanderpooten [24]. Ellos definen las aproximaciones inferior y superior de un conjunto X de la forma siguiente. Sean $x \subseteq U$ y R una relación binaria y reflexiva sobre U . Entonces la aproximación inferior y superior se definen como $R_*(X) = \{x \in U : R^{-1}(x) \subseteq X\}$ y $R^*(X) = \{x \in U : R^{-1}(x) \cap X \neq \emptyset\}$, donde R^{-1} denota la relación inversa de R .

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 [34, 35]. En la otra alternativa se emplea una relación de semejanza, definida como $xRy \Leftrightarrow F(x,y) \geq \varepsilon$, $x, y \in U$, $F(x,y)$ es una función de semejanza (la cual podría ser definida como en la expresión 8), y ε es un umbral (threshold). La inseparabilidad entre dos objetos depende del conjunto de pesos w_p , como se analiza más adelante. Esta es una de las posibles aplicaciones de la nueva medida que se propone en este artículo, calcular el mejor conjunto de pesos para construir la relación de semejanza.

Medida calidad de la similaridad

Sea un sistema de decisión $DS = (U, A \cup \{d\})$ donde el dominio de los rasgos en $A \cup \{d\}$ pueden ser valores discretos o continuos, y relaciones de semejanza $R1$ y $R2$ entre los objetos definidas de la forma siguiente. Para todo par de objetos (x, y) en U :

$$xR1y \text{ si y solo si } F1(x,y) \geq e1 \quad (2)$$

$$xR2y \text{ si y solo si } F2(x,y) \geq e2 \quad (3)$$

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 ; $e1$ y $e2$ son umbrales de semejanza.

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 (4) y (5), donde $N1(x)$ y $N2(x)$ son los conjuntos de objetos similares a x se acuerdo a las relaciones $R1$ y $R2$ respectivamente:

$$N1(x) = \{y \in U : xR1y\} \quad (4)$$

$$N2(x) = \{y \in U : xR2y\} \quad (5)$$

Entonces 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 $N2$, $N1(x) \approx N2(x)$, donde el símbolo (\approx) denota la mayor similaridad posible entre los conjuntos $N1(x)$ y $N2(x)$ dado los umbrales $e1$ y $e2$. El grado de similaridad entre estos conjuntos se define por la expresión (6):

$$\varphi(x) = \frac{|N1(x) \cap N2(x)|}{0.5 * |N1(x)| + 0.5 * |N2(x)|} \quad 0 \leq \varphi(x) \leq 1 \quad (6)$$

A partir de la expresión (6) la medida calidad de la similaridad del sistema de decisión

$DS = (U, A + \{d\})$ se define por la expresión (7):

$$\theta(DS) = \frac{\sum_{\forall x \in U} \varphi(x)}{|U|} \quad (7)$$

La medida $\theta(DS)$ 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 . Entonces el problema es encontrar las relaciones $R1$ y $R2$ que maximizan la medida calidad de la similaridad.

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 $R2$ se define como $xR2y \Leftrightarrow d(x) = d(y)$. En este caso, el problema es encontrar la relación $R1$ tal que las clases de semejanza de un objeto se corresponden con su clase de decisión. El costo computacional de calcular esta medida es $O(m^2 * (c1 + c2))$, donde $c1$ y $c2$ representan el costo computacional de evaluar las funciones $F1(x,y)$ y $F2(x,y)$; lo cual es ligeramente superior al costo de calcular la calidad de la clasificación según (1), el cual es $O(k * l * m^2)$, considerando que el costo de calcular la aproximación inferior de un conjunto es $O(l * m^2)$ de acuerdo a [Bel 98], donde l es la cantidad de rasgos, m la cantidad de objetos del universo y k la cantidad de clases

Un método para calcular los pesos de los rasgos usando la medida calidad de la similaridad

La asignación de pesos a los rasgos puede ser vista como una generalización del problema de selección de rasgos. La noción de selección de rasgos ha evolucionado en el tiempo, según un enfoque clásico el resultado de un selector de rasgos es binario indicando si el rasgo debe ser considerado o no. Este enfoque se puede generalizar mediante la introducción de un concepto de “grado de relevancia”, llamado también peso o importancia del rasgo, usualmente definido en el intervalo $[0, 1]$.

La RST clásica ofrece una medida para calcular la relevancia de un rasgo en un sistema de decisión $(U, A \cup \{d\})$. Primero se define una medida $\sigma(A, d)$ para calcular el grado de dependencia del rasgo de decisión d respecto al conjunto de rasgos A . Si $\sigma(A, d) = 1$ entonces se dice que d depende completamente de A ; si $\sigma(A, d) = 0$ entonces el rasgo de decisión es totalmente independiente de A , en otro caso d depende de A en grado k ($0 \leq k \leq 1$). Cuando el conjunto A se modifica, este grado de dependencia también se puede modificar; por eso, la diferencia numérica entre

$\sigma(A,d)$ y $\sigma(A-\{a\}, d)$ se puede interpretar como una medida del grado de influencia que tiene el atributo a en $\sigma(A, d)$.

A continuación se propone una alternativa para el caso de la importancia de los rasgos para el caso de la RST extendida, la cual es particularmente útil en el caso de que el dominio del rasgo decisión sea continuo. El cálculo del peso de los rasgos permite construir la relación de semejanza RI , definida por la expresión (2), y considerando la función FI como una medida de semejanza que considera tres componentes principales: (i) Una medida de similaridad local usada para comparar los valores de un rasgo (llamada función de comparación del rasgo); (ii) El conjunto de pesos que representa la importancia relativa de cada atributo; (iii) Una medida de similaridad global responsable de computar el grado de similaridad entre dos objetos basada en las medidas de similaridad locales y el conjunto de pesos (llamada función de semejanza, y que se define por la expresión (8)).

Una expresión clásica para calcular el grado de similaridad entre dos objetos (X,Y) de un universo es:

$$F(X, Y) = \sum_{i=1}^n w_i * \delta_i(X_i, Y_i) \quad (8)$$

donde: n es la cantidad de rasgos; w_i es el peso del rasgo i ; X_i, Y_i son los valores del rasgo i en los objetos X, Y respectivamente; δ_i es la función de comparación para el rasgo i .

Ejemplos de funciones de comparación de rasgos son la expresiones (9) y (10), la primera más usada en el caso de rasgos con dominio discreto y la segunda para rasgos de valores reales:

$$\delta(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases} \quad (9)$$

$$\delta(x, y) = \begin{cases} 1 & \text{if } |x - y| \leq \epsilon \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

Definido de esta forma, el problema de construir la relación de semejanza RI definida por la expresión (2), a partir de la función de semejanza FI definida por la expresión (8), consiste en encontrar el conjunto de pesos $W = \{w_1, w_2, \dots, w_n\}$, donde n es la cantidad de rasgos en el conjunto A . El método que se propone para encontrar W , busca el conjunto de pesos que maximiza la expresión (7); en este caso, la relación de semejanza $R2$ definida por la expresión (3) utiliza como función de semejanza la expresión $F2(X, Y) = \partial(d(X), d(Y))$, donde ∂ es una función de comparación para el rasgo de decisión d . Como se planteó antes, en casos de problemas de clasificación donde el dominio del rasgo de decisión d es discreto basta con usar la expresión (9) como función de comparación de rasgos al construir $F2$.

Para encontrar el conjunto W que maximiza la expresión (7) se propone usar un método de búsqueda heurística tal como las meta heurísticas Optimización basada en partículas (Particle Swarm Optimization, PSO) [36], o Algoritmos genéticos (Genetics Algorithms, GA), en cuyo caso se han hecho estudios particulares para considerar rasgos continuos [37, 38]. En este artículo el estudio experimental se realizó usando PSO, pues esta metaheurística ha mostrado un buen desempeño en los problemas de optimización continua [39, 40]. Además, el empleo de PSO en procesos de selección de rasgos junto a los conjuntos aproximados ya se ha estudiado por otros autores con buenos resultados como se muestra en [9, 10].

En este caso las partículas, que representan el vector W , tienen n componentes (una por cada rasgo en A). La calidad de las partículas se calcula usando la expresión (7). Al final del proceso de búsqueda desarrollado por el método PSO, la mejor partícula constituye el vector de pesos W a usar en la función de semejanza FI (definida por la expresión 8), y con esta queda construida la relación de semejanza RI (definida por la expresión 2).

El método propuesto constituye un nuevo procedimiento de aprendizaje de pesos cuya

aplicación no está limitada a problemas de clasificación. Ejemplos de algoritmos de aprendizaje de pesos para problemas diferentes a la clasificación son los basados en ordenamiento (ranking, score) de casos en el contexto del razonamiento basado en casos (Case-based Reasoning), en los cuales se retroalimenta al algoritmo con el orden deseado de los casos [41, 42]; en este caso, el objetivo del algoritmo de aprendizaje es optimizar una medida de similaridad de una manera tal que el orden dado a los casos recuperados sea igual al orden indicado por un tutor. Por otra parte, Hall presentó un algoritmo de selección de rasgos el cual se puede aplicar a problemas con rasgo de decisión con dominio continuo aplicando adecuadas medidas de correlación [43, 44].

Estudio del método para el problema de aproximación de funciones

En los problemas de clasificación el dominio del rasgo de decisión es discreto, de modo que un clasificador se puede considerar como una función cuya imagen es un conjunto de valores discretos. Mientras en el problema de aproximación de funciones es diferente, pues la imagen de la función es un conjunto continuo. Aunque la RST se ha desarrollado fundamentalmente para problemas de clasificación, también se han desarrollado extensiones de la misma para la aproximación de funciones, como el enfoque denominado “rough real functions” [2]. En [22], los autores discuten el problema de aproximación de funciones usando los conjuntos aproximados.

Dado el sistema de decisión $DS = (U, A \cup \{d\})$ donde el dominio del rasgo d son los números reales se desea construir la función F tal que $F(O) : U \rightarrow [a, b]$ donde $[a, b] \subseteq \mathfrak{R}$. En este caso, usualmente los rasgos en A también tienen como dominio los números reales.

En el estudio que se presenta aquí, como aproximador de funciones se utiliza el método K-Nearest Neighbours (k-NN). La idea básica el método k-NN es que valores de entradas similares tienen salidas similares [45]. Las salidas de un

objeto son computadas promediando las salidas de los vecinos más similares a él. Sea el objeto X_h y el conjunto de objetos más similares denotado por $N(X_h)$, la salida de X_h , denotada por d_h , se calcula a partir de $N(X_h)$ usando la expresión (11):

$$d_h = \frac{\sum_{X_j \in N(X_h)} d_j}{k} \quad (11)$$

Donde k es la cardinalidad del conjunto $N(X_h)$. Para construir el conjunto $N(X_h)$, es necesario usar una función que permita calcular el grado de similaridad entre dos objetos, en este caso se usa la función definida por la expresión (8). Los resultados presentados en [46] 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; allí también se muestra que el empleo de métodos de búsqueda para el cálculo de los pesos de los rasgos, obtienen mejores resultados que usar procedimientos de minimización de funciones.

De modo que el método propuesto en este artículo para el cálculo de pesos, también constituye una vía para mejorar el desempeño del método K-NN usado como aproximador de funciones, pues el mismo permite encontrar el conjunto de pesos W usado en la función de semejanza que mejora la recuperación de objetos similares. La importancia de construir la medida de similaridad más adecuada en los sistemas basados en casos, particularmente para resolver problemas de aproximación de funciones, se estudia en [42]. Los resultados experimentales que se presentan seguidamente muestran este desempeño.

Se utilizan conjuntamente el método K-NN y el método propuesto en la sección 5 para encontrar el conjunto de pesos de los rasgos, al cual se denomina PSO+RST, para resolver el problema de aproximación de funciones en varios ejemplos de sistemas de decisión. El resultado alcanzado por el aproximador de funciones que usa el conjunto de pesos calculado basado en la medida calidad de la semejanza, el método PSO+RST,

se compara con otras dos alternativas para el cálculo de los pesos, una consiste en asignar el mismo valor de peso a todos los rasgos ($w = \frac{1}{|A|}$), llamada método Standard y la otra es el método de cálculo k-NN_{VSM} [47].

Para evaluar la precisión de los resultados alcanzados por ambos métodos se utilizaron tres medidas: (i) Porcentaje del error medio absoluto (Mean Absolute Percentage Error, MAPE), (ii) Raíz del error cuadrático medio (Root Mean Square Error, RMSE), y (iii) Promedio de las diferencias entre el valor deseado y el producido por el método (PMD). Estas medidas se definen por las expresiones (12)-(14):

$$MAPE = \frac{\sum_{i=1}^N \frac{|ai - yi|}{ai}}{N} * 100\% \quad (12)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N \frac{ai - yi^2}{ai}}{N}} * 100\% \quad (13)$$

$$PMD = \frac{\sum_{i=1}^N |ai - yi|}{N} \quad (14)$$

donde: ai es el valor de salida deseado; yi es el valor producido por el método y N es el número de objetos.

Los sistemas de decisión utilizados se describen en la tabla 1. En todos los casos son datos reales sobre variables meteorológicas provenientes de varias regiones de Cuba localizadas en las provincias de Camaguey y Las Tunas, y el objetivo es pronosticar el valor de la variable cuyo nombre aparece en el nombre del sistema de decisión.

Para todos los rasgos se utiliza como función de comparación de rasgos la expresión (15):

Tabla 1 Descripción de los sistemas de decisión utilizados

| Sistema de decisión | Número de objetos | Número de rasgos |
|-------------------------------|--------------------------|-------------------------|
| Temperatura máxima Camaguey | 1782 | 26 |
| Temperatura mínima Camaguey | 1011 | 26 |
| Temperatura máxima Florida | 1525 | 26 |
| Temperatura máxima Las Tunas | 1373 | 23 |
| Nubosidad Nuevitas | 470 | 34 |
| Fuerza del viento Camaguey | 358 | 23 |
| Dirección del viento Camaguey | 2086 | 29 |

$$\partial(X_i, Y_i) = \begin{cases} 1 - \frac{|X_i - Y_i|}{Max(N_i) - Min(N_i)} & \text{si } i \text{ es continuo} \\ 1 & \text{si } i \text{ es discreto y } X_i = Y_i \\ 0 & \text{si } i \text{ es discreto y } X_i \neq Y_i \end{cases} \quad (15)$$

Para el estudio se utiliza el método k-fold cross-validation, el cual divide el sistema de decisión original en k subsistemas de igual tamaño, y se usa uno como conjunto de prueba y el resto como conjunto de entrenamiento, calculándose la precisión del aproximador como el promedio obtenido en los k conjuntos de prueba; esta técnica elimina diferentes limitaciones de los procesos de aprendizaje [47]. El valor de k utilizado es 10, como se recomienda en [48].

Para el método k-NN se analizaron dos alternativas para k (3 y 5). Los resultados de las medidas MAPE, RMSE y PMD para cada alternativa se presentan en la tabla 2.

Seguidamente se aplicaron pruebas estadísticas no paramétricas para evaluar la comparación.

El Test de Wilcoxon fue usado para evaluar si las diferencias representaban mejoras estadísticamente significativas. El método de Monte Carlo fue aplicado con un intervalo de confianza de 99%, se consideraron los siguientes valores para el valor de p :

- i) significativo \rightarrow significación menor que 0,05 y mayor que 0,01;
 - ii) altamente significativo \rightarrow una significación menor que 0,01;
 - iii) poco significativo \rightarrow un resultado menor que 0,1 y mayor que 0,05;
- No significativo \rightarrow un resultado mayor que 0,1.

Tabla 2 Precisión obtenida por cada método de acuerdo a las medidas empleadas

| Sistema de decisión | K | PSO+RST | | | Standard | | | k-NN _{VSM} | | |
|-------------------------------|---|---------|-------|-------|----------|-------|-------|---------------------|-------|-------|
| | | MAPE | PMD | RMSE | MAPE | PMD | RMSE | MAPE | PMD | RMSE |
| Temperatura máxima Florida | 3 | 3,338 | 0,985 | 18,58 | 3,381 | 0,999 | 18,61 | 3,381 | 0,999 | 18,61 |
| | 5 | 3,206 | 0,947 | 18,12 | 3,338 | 0,983 | 18,82 | 3,338 | 0,983 | 18,82 |
| Nubosidad Nuevitas | 3 | 15,98 | 0,493 | 27,99 | 21,79 | 0,684 | 34,27 | 21,79 | 0,684 | 34,27 |
| | 5 | 15,56 | 0,487 | 26,32 | 22,08 | 0,690 | 34,78 | 22,08 | 0,690 | 34,78 |
| Temperatura máxima Camaguey | 3 | 3,310 | 0,970 | 5,095 | 3,402 | 0,998 | 5,163 | 3,342 | 0,985 | 5,137 |
| | 5 | 3,349 | 0,985 | 4,537 | 3,354 | 0,990 | 4,855 | 3,340 | 0,980 | 4,798 |
| Temperatura mínima Camaguey | 3 | 4,069 | 0,805 | 6,049 | 4,269 | 0,844 | 6,104 | 4,238 | 0,826 | 6,099 |
| | 5 | 4,020 | 0,793 | 6,079 | 4,164 | 0,821 | 6,088 | 4,109 | 0,805 | 6,084 |
| Temperatura máxima Las Tunas | 3 | 3,671 | 0,747 | 5,259 | 3,732 | 0,758 | 5,473 | 3,732 | 0,758 | 5,473 |
| | 5 | 3,558 | 0,721 | 5,144 | 3,702 | 0,750 | 5,413 | 3,702 | 0,750 | 5,413 |
| Fuerza del viento Camagüey | 3 | 15,72 | 6,691 | 21,64 | 16,86 | 7,036 | 24,46 | 16,55 | 7,006 | 24,01 |
| | 5 | 15,50 | 6,558 | 21,40 | 17,20 | 7,163 | 23,93 | 17,01 | 7,010 | 23,40 |
| Dirección del viento Camagüey | 3 | 21,61 | 16,92 | 52,18 | 36,35 | 30,30 | 89,46 | 36,35 | 30,30 | 89,46 |
| | 5 | 22,15 | 17,00 | 55,75 | 35,86 | 29,60 | 86,14 | 35,86 | 29,60 | 86,14 |

La tabla 3 muestra los resultados del Test Wilcoxon para cada sistema de decisión. Este análisis

estadístico sostiene que el método PSO+RST supera significativamente las otras alternativas.

Tabla 3 Resultados del Test Wilcoxon

| Sistema de decisión | Métodos | R+ | R- | p-value |
|-------------------------------|--------------------------------------|-----------|-----------|----------------|
| Temperatura máxima Florida | StandardK5 vs. PSO+RSTK5 | 626108 | 537467 | 0,009 |
| | k-NN _{VSM} K5 vs. PSO+RSTK5 | 613142 | 548477 | 0,008 |
| Nubosidad Nuevitas | StandardK5 vs. PSO+RSTK5 | 84286 | 18092 | 0,000 |
| | k-NN _{VSM} K5 vs. PSO+RSTK5 | 83253 | 20051 | 0,000 |
| Temperatura máxima Camagüey | StandardK3 vs. PSO+RSTK3 | 837292 | 751361 | 0,048 |
| | k-NN _{VSM} K3 vs. PSO+RSTK3 | 8127523 | 771631 | 0,047 |
| Temperatura mínima Camagüey | StandardK5 vs. PSO+RSTK5 | 274666 | 236900 | 0,042 |
| | k-NN _{VSM} K5 vs. PSO+RSTK5 | 257563 | 247658 | 0,040 |
| temperatura máxima Las Tunas | StandardK5 vs. PSO+RSTK5 | 501695 | 440183 | 0,036 |
| | k-NN _{VSM} K5 vs. PSO+RSTK5 | 497625 | 456372 | 0,034 |
| Fuerza del viento Camagüey | StandardK5 vs. PSO+RSTK5 | 36359 | 243671 | 0,001 |
| | k-NN _{VSM} K5 vs. PSO+RSTK5 | 34873 | 269832 | 0,000 |
| Dirección del viento Camagüey | StandardK3 vs. PSO+RSTK3 | 503602 | 652324 | 0,000 |
| | k-NN _{VSM} K3 vs. PSO+RSTK3 | 489726 | 678293 | 0,000 |

Conclusiones

La Teoría de los conjuntos aproximados ofrece diferentes técnicas para el análisis de los datos, en este contexto se propone en este artículo la medida calidad de la similaridad, la cual está diseñada para el caso de que los sistemas de decisión incluyen rasgos continuos, especialmente si el rasgo de decisión tiene dominio continuo. Su aplicabilidad fue analizada para los casos de construir relaciones de similaridad cuando se trabaja con extensiones de la RST y realizar el cálculo de los pesos de los rasgos en procesos de selección de rasgos.

Se analizó experimentalmente como el empleo del método para el cálculo de los pesos basado en la medida calidad de la similaridad permite mejorar el desempeño del método k-NN en el problema de aproximación de funciones.

Referencias

1. Z. Pawlak. "Rough Sets". *International journal of Computer and Information Sciences*. Vol. 11. 1982. pp. 341-356.
2. Z. Pawlak. "Rough sets, Rough function and Rough calculus". *Rough Fuzzy Hybridization: a new trend in decision-making*. Ed. Springer-Verlag. Berlin. 1999. pp. 235-246.

3. Z. B. Xu. "Inclusion degree: a perspective on measures for rough set data analysis." *Information Sciences*. Vol. 141. 2002. pp. 227-236.
4. F. E. Tay, L. Shen. "Economic and financial prediction using rough set model." *European Journal of Operational Research*. Vol. 141. 2002. pp. 641-659.
5. S. K. Pal, A. Skowron. *Rough Fuzzy Hybridization: A New Trend in Decision-Making*, Ed. Springer-Verlag. Singapore. 1999. pp. 389.
6. N. Zhong, J. Dong, S. Ohsuga, "Using Rough sets with heuristics for feature selection." *Journal of Intelligent Information System*. Vol. 16. 2001. pp. 199-214.
7. R. Jensen, S. Qiang. *Finding rough sets reducts with Ant colony optimization*. in UK Workshop on Computational Intelligence. Edinburgh (UK). 2003. pp. 15-22.
8. P. R. Bello, Y. Gomez, Y. Caballero, A. Nowe. "Using ACO and Rough set theory to feature selection". *WSEAS Transactions on Information Science and Applications*. Vol. 2. 2005. pp. 512-517.
9. X. Wang. "Rough set feature selection and rule induction for prediction of malignancy degree in brain glioma." *Computer methods and programs in biomedicine*. Vol. 83. 2006. pp. 147-156.
10. X. Wang. "Feature selection based on rough sets and particle swarm optimizatio". *Pattern Recognition Letters*. Vol. 28. 2007. pp. 459-471.
11. Y. Caballero, R. Bello. *Feature selection algorithms using Rough set theory*. in Seventh Int. Conf. on Intelligence Systems Design and Applications ISDA07. IEEE Computer Society. Rio de Janeiro. 2007. pp. 407-411
12. P. R. Bello, Y. Gomez, A. Puris, R. Falcon. *Feature selection through Dynamic Mesh Optimization*. Lecture Notes on Computer Science. 2008. pp. 348-355.
13. Y. Gomez, R. Bello, A. Puris. "Two step Swarm intelligence to solve the feature selection problem". *Journal of Universal Computer Science*. Vol. 14. 2008. pp. 2582-2596.
14. G. H. John. *Relevant features and the subset selection problem*. in Proceedings of the Eleventh International Conference on Machine Learning. Ed. Morgan Kaufmann. 1994. pp. 121-129.
15. I. Inza. "Feature subset selection by bayesian networks based optimization." *Artificial intelligence*. Vol. 123. 2000. pp. 157-184.
16. H. Xing, L. Xu. "Feature space theory –a mathematical foundation for data mining." *Knowledge-based systems*. Vol. 14. 2001. pp. 253-257.
17. J. Stefanowski. "An experimental evaluation of improving rule based classifiers with two approaches that change representations of learning examples." *Engineering Applications of Artificial Intelligence*. Vol.17. 2004. pp. 439-445.
18. P. Somol, P. Pudil. "Feature selection toolbox." *Pattern Recognition*. Vol. 35. 2002. pp. 2749-2759.
19. M. Dong, R. Kothari. "Feature subset selection using a new definition of classifiability." *Patter Recognition Letter*. Vol. 24. 2003. pp. 1215-1225.
20. M. Kudo, J. Sklansky. "Comparison of algorithms that select features for pattern classifiers." *Pattern Recognition*. Vol. 33. 2000. pp. 25-41.
21. H. Zhang, G. Sun. "Feature selection using tabu search method." *Pattern Recognition*. Vol. 35. 2002. pp. 710-711.
22. Z. Pawlak. *Approximating functions using rough sets*. in Proceedings of 2004 Annual Meeting of the North American Fuzzy Information Processing Society NAFIPS. 2004. pp. 785-790
23. Z. Pawlak, A. Skowron. "Rough sets: Some extensions." *Information Sciences*. Vol. 177. 2007. pp. 28-40.
24. R. Slowinski, D. Vanderpooten. "Similarity relation as a basis for rough approximations." *Advances in Machine Intelligence & Soft-Computing*. P. P. Wang, (editor) Ed. Duke University Press. Vol. 4. Durham. 1997. pp. 17-33.
25. Y. Yao. "Relational interpretations of neighborhood operators and rough set approximations operators". *Information Sciences*. Vol. 101. 1998. pp. 239-259.
26. R. Slowinski, D. Vanderpooten. "A generalized definition of rough approximations based on similarity". *IEEE Transactions on Data and Knowledge Engineering*. Vol. 12. 2000. pp. 331-336.
27. S. Greco. "Rough sets theory for multicriteria decision analysis." *European Journal of Operational Research*. Vol. 129. 2001. pp. 1-47.
28. A. Skowron, R. Swiniarski, P. Synak. "Approximation Spaces and Information Granulation". *Transactions on Rough Sets III. Lecture Notes in Computer Science*. J. F. Peters and A. Skowron, (editors). Ed. Springer-Verlag GmbH. 2005. pp. 175-189.

29. K. Qin, Y. Gao, Z. Pei. "On Covering Rough Sets." *Lectures Notes on Artificial Intelligence*. Vol 4481. 2007. pp. 34-41.
30. A. Skowron, J. Stepaniuk. "Tolerance approximation spaces." *Fundamenta Informaticae*. Vol. 27. 1996. pp. 245-253.
31. Y. Yao. "On generalizing Rough set theory." *Lectures Notes on Artificial Intelligence*. Vol. 2639. 2003. pp. 44-51.
32. W. Zhu, F. Wang. *Relations among three types of covering rough sets*. in IEEE GrC. Atlanta, USA. 2006. pp. 43-48.
33. D. Bianucci, G. Cattaneo, D. Ciucci. "Entropies and Co-entropies of coverings with applications to incomplete information systems." *Fundamenta Informaticae*. Vol. 75. 2007. pp. 77-105.
34. N. H. Son, N. S. Hoa. "Discretization of real value attributes for control problems." *Workshop on Robotics, Intelligent Control and Decision Support Systems, P/JWSTK*. Ed. Warsaw. Varsovia (Poland). 1999. pp. 47-52.
35. N. H. Son, N. S. Hoa. "An application of discretization methods in control." *Proc. Of the Workshop on Robotics, Intelligent Control and Decision Support Systems*. Polish-Japanese Institute of Information Technology. Ed. Warsaw. Poland. 1999. pp. 47-52.
36. J. Kennedy, R. C. Eberhart. *Particle swarm optimization*. in *In Proceedings of the 1995 IEEE International Conference on Neural Networks*. Piscataway. IEEE Service Center. New Jersey. 1995. pp. 1942-1948.
37. F. Herrera, M. Lozano, A. Sanchez. "A Taxonomy for the Crossover Operator for Real Coded Genetic Algorithms: An Experimental Study." *International Journal of Intelligent Systems*. Vol. 18. 2003. pp. 309-338.
38. M. García Martínez, F. Herrera. "Global and local real-coded genetic algorithms based on parent-centric crossover operators." *European Journal of Operational Research*. Vol. 185. 2008. pp. 1088-1113.
39. K. E. Parsopoulos, M. N. Vrahatis. "Recent approaches to global optimization problems through Particle Swarm Optimization." *Natural Computing*. Vol. 1. 2002. pp. 235-306.
40. M. Reyes Sierra, C. C. Coello. "Multi-objective Particle Swarm Optimizers: A survey of the State-of-the-Art." *International Journal of Computational Intelligence Research*. Vol. 2. 2006. pp. 287-308.
41. Z. Zhang, Q. Yang. *Dynamic refinement of feature weights using quantitative introspective learning*. in In Proceedings of the 16th International Joint Conference on Artificial Intelligence (*IJCAI99*). 1999. pp.57-71
42. A. Stahl. "Learning Feature Weights from Case Order Feedback." *Lecture Notes in Computer Science Springer*. Vol. 2080. 2001. pp. 502-516.
43. M. A. Hall. *Correlation-based feature selection for machine learning*, in *Department of Computer Science*. University of Waikato: Hamilton. New Zealand. 1998. pp. 145.
44. M. A. Hall. *Feature selection for discrete and numeric class machine learning*. University of Waikato: Hamilton. New Zealand. 1999.
45. R. L. López, E. Armengol. "Machine learning from examples: Inductive and Lazy methods." *Data & Knowledge Engineering*. Vol. 25. 1998. pp. 99-123.
46. W. Duch, K. Grudzinski. *Weighting and selection features*. *Intelligent Information System VIII*. in Proceedings of the Workshop held in Ustron. Ustron, Poland. 1999. pp.32-36
47. T. Mitchell. *Machine Learning*. Ed. McGraw Hill. Portland, OR, USA. 1997. pp. 414.
48. J. Demsar. "Statistical comparisons of classifiers over multiple data sets." *Journal of Machine Learning Research*. Vol. 7. 2006. pp. 1-30.