

Optimización en Colonias de Hormigas y en Enjambres de Partículas para la selección atributos

Ant Colony Optimization and Particle Swarm Optimization for feature selection

Yanela Rodríguez Álvarez¹, Yailé Caballero Mota¹, Rafael Bello Pérez².

¹Universidad de Camagüey, Cuba, yanela.rodrique@reduc.edu.cu, yaile.caballero@reduc.edu.cu

²Universidad Central "Marta Abreu" de Las Villas, rbellop@uclv.edu.cu.

RESUMEN

La selección de atributos relevantes puede ser vista como uno de los problemas más importantes en el campo del aprendizaje automático. En esta investigación se propone un enfoque que integra funciones de Optimización en Colonias de Hormigas (ACO) y Optimización en Enjambres de Partículas (PSO) para la generación de subconjuntos de rasgos principales que describen los datos y como función de evaluación la medida calidad de la clasificación de la Teoría de los Conjuntos Aproximados (RST). El algoritmo propuesto, nombrado ACO-PSO-RST, está basado en la variante Sistema de Colonias de Hormigas, utiliza PSO para la inicialización del rastro de feromona y RST se utiliza como función heurística por la metaheurística ACO y para determinar si el subconjunto es un reducto. Para verificar la eficiencia del método propuesto se llevaron a cabo experimentos con bases de casos internacionales y se realizaron comparaciones con otros métodos. Además, este método se aplicó en el preprocesamiento de los datos para predecir, de forma automatizada, la calidad del agua para el riego. Los resultados demostraron que el enfoque propuesto provee una solución eficiente al problema de selección de rasgos.

Palabras clave: selección de rasgos; Optimización en Colonias de Hormigas; Optimización en Enjambres de Partículas; Teoría de los Conjuntos Aproximados.

ABSTRACT

Feature selection can be viewed as one of the most fundamental problems in the field of machine learning. In this investigation is proposing an approach that uses Ant Colony Optimization (ACO) and Particles Swarms Optimization (PSO) to generation of subsets of principal features that describe the data and Rough Set Theory (RST) offers the heuristic function to measure the quality of a single subset. The proposed algorithm, named ACO-PSO-RST, is based in Ant Colony System variant, used PSO for the initialization pheromone raster and RST is used as heuristic function for ACO metheuristic and to determinate if the subset is a reduct. To verify the efficiency of method proposed, experiments are carried out on some standard international datasets. In addition, this method is applied in the pre-processing of data to predict, automatically, the water quality for irrigation. The results demonstrated that this approach could provide efficient solution to find a minimal subset of the features.

Keywords: feature selection; Ant Colony Optimization; Particle Swarm Optimization; Rough Set Theory.

1. INTRODUCCIÓN

La selección de rasgos consiste en encontrar el subconjunto de atributos del conjunto de datos original que mejor describe los objetos del dominio; tiene como meta reducir la dimensionalidad del conjunto de rasgos a través de la selección del subconjunto de rasgos de mejor desempeño bajo algún criterio de clasificación [1]. Este proceso de selección se hace eliminando rasgos irrelevantes y redundantes [2, 3], proporcionando así una mejor representación de la información original reduciendo significativamente el costo computacional y contribuyendo a una mejor generalización del algoritmo de aprendizaje. Normalmente este

proceso está presente en las etapas previas de las principales tareas de la minería de datos, ya sean supervisadas o no [4].

Es un hecho que el comportamiento de los clasificadores mejora cuando se eliminan los atributos no relevantes y redundantes. La selección de los atributos relevantes se debe a la preferencia por los modelos más sencillos frente a los más complejos. Esta preferencia ha sido utilizada con bastante frecuencia en la ciencia moderna y tiene sus orígenes en el denominado Principio de la Cuchilla de Occam (*Occam's Razor*) [5]. La selección de atributos es un campo de investigación y desarrollo productivo desde los años setenta, donde confluyen distintas áreas como el reconocimiento de patrones [6-8], el aprendizaje automático [9-13] y la minería de datos [14, 15]. Las técnicas de selección de características se aplican en muchos entornos diferentes, como por ejemplo la clasificación de textos, recuperación de imagen [16] y bioinformática. Se hace constar, que el proceso de selección de atributos, además de preceder a la clasificación, suele estar presente en las etapas previas de las principales tareas de la minería de datos, ya sean supervisadas o no, como regresión, agrupamiento y reglas de asociación [17].

Partiendo de la premisa de que en el proceso de selección de atributos se escoge un subconjunto de atributos del conjunto original, este proceso pretende elegir atributos que sean relevantes para una aplicación y lograr el máximo rendimiento con el mínimo esfuerzo. Se debe tener en cuenta que los atributos irrelevantes y redundantes existentes en una base de datos pueden tener un efecto negativo en los algoritmos de clasificación: (1) Al tener más atributos, normalmente implica la necesidad de tener más instancias para garantizar la fiabilidad de los patrones obtenidos (variabilidad estadística entre patrones de diferente clase). Por consiguiente, el algoritmo de clasificación entrenado con todos los datos tardará más tiempo. (2) Los atributos irrelevantes y los redundantes, pueden confundir a los algoritmos de aprendizaje. Por lo que en general, el clasificador obtenido es menos exacto que otro que aprenda sobre datos relevantes. (3) Además, con la presencia de atributos redundantes o de irrelevantes, el clasificador obtenido será más complejo, dificultando el entendimiento de los resultados. (4) Asimismo, la reducción de características se podría tener en cuenta en futuras capturas de datos, reduciendo el coste de almacenamiento y tal vez el económico. Los procedimientos de selección de rasgos constan de dos componentes principales: la función de evaluación y el método de generación de subconjuntos (basado en un proceso de búsqueda). Existen diferentes enfoques y técnicas para seleccionar atributos relevantes, tales como las técnicas de Inteligencia Colectiva, y las basadas en la Teoría de los Conjuntos Aproximados.

La Teoría de Conjuntos Aproximados fue introducida por Z. Pawlak en 1982 [18, 19]. Se basa en aproximar cualquier concepto, un subconjunto duro del dominio como, por ejemplo, una clase en un problema de clasificación supervisada, por un par de conjuntos exactos, llamados aproximación inferior y aproximación superior del concepto. Con esta teoría es posible tratar tanto datos cuantitativos como cualitativos, y no se requiere eliminar las inconsistencias previas al análisis; respecto a la información de salida puede ser usada para determinar la relevancia de los atributos y generar las relaciones entre ellos [20-29]. La inconsistencia describe una situación en la cual hay dos o más valores en conflicto para ser asignados a una variable [30]. Sobre los Conjuntos Aproximados se han manifestado diversos autores, los cuales ven esta teoría como la mejor herramienta para modelar la incertidumbre cuando esta se manifiesta en forma de inconsistencia, y como una nueva dirección en el desarrollo de teorías sobre la información incompleta [31-33]. La principal ventaja que tiene el análisis de datos basado en RST es que para operar este no requiere parámetros adicionales además de los datos de entrada [34]. En este epígrafe se describirán los conceptos fundamentales de los Conjuntos Aproximados para el enfoque clásico.

Por otro lado, la Inteligencia Colectiva (también llamada inteligencia de enjambre) es un paradigma inteligente, distribuido e innovador para la solución de problemas de optimización que originalmente tomó su inspiración en ejemplos biológicos de enjambres. Dentro de esta familia de algoritmos existen dos bien representativos: la Optimización mediante Enjambres de Partículas (PSO) [35] incorpora el comportamiento de enjambre observado en bandadas de pájaros, cardúmenes de peces o enjambres de abejas, de las que surgió la idea. Y la Optimización mediante Colonias de Hormigas (ACO) [36, 37] que tiene que ver con

sistemas inteligentes artificiales, que son inspirados a partir de observar el comportamiento de las hormigas reales en la búsqueda de comida, los cuales son usados para resolver problemas de optimización discretos. Los algoritmos de ACO reproducen el comportamiento de las hormigas reales en una colonia artificial. Estos han sido aplicados a un gran número de problemas cuyas soluciones generan explosión combinatoria como el clásico del vendedor ambulante, problemas de ruteo en redes de telecomunicaciones, planificación de tareas, etcétera. En [38] se plantea el uso de estas técnicas para el cálculo de reductos debido a que las hormigas pueden descubrir las mejores combinaciones de atributos en la medida en que atraviesan el grafo.

2. METODOLOGÍA

Para evaluar la calidad del modelo propuesto, se llevó a cabo un extenso estudio experimental sobre una colección de 13 conjuntos de datos reconocidos internacionalmente, para algunos de los cuales ya existen resultados reportados, obtenidos con otros métodos. Los conjuntos de datos son provenientes del depósito de datos para aprendizaje automatizado disponibles en el sitio ftp de la Universidad de Irvine, California, del sitio personal de Jensen, y en el Repositorio de la UCI para el aprendizaje automático estos son: Audiology, Biomed, Bupa (Liver Disorders), Cleveland, Exactly, Heart-statlog, Ionosphere, Iris, New Thyroid, Sonar, Vehicle, Wine y Zoo. En los experimentos se compara el método propuesto con los principales algoritmos encontrados en la literatura que combinan los enfoques de inteligencia colectiva y conjuntos aproximados aplicados a la selección de rasgos, para demostrar que el enfoque propuesto es superior: AS-RST-FS, ACS-RST-FS, AFSBRSACO, RSFSACO y ACO-RST-FSP. El desempeño de los algoritmos se evalúa a través de las métricas tamaño del reducto obtenido y calidad de la clasificación, y las diferencias significativas encontradas entre los resultados están validadas con un análisis estadístico. Específicamente para el análisis estadístico de los resultados se utilizó la prueba de Friedman para detectar diferencias estadísticamente significativas entre un grupo de resultados. Se emplea además la prueba de Holm [39] con el fin de encontrar los algoritmos significativamente superiores.

3. RESULTADOS Y DISCUSIÓN

Conjuntos de datos

Las características de los 13 conjuntos de datos con los que se obtuvieron resultados significativos aparecen en la **Tabla 1**.

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

Bases de Casos	Cantidad de rasgos	Cantidad de instancias	Cantidad de clases	Ausencia de información
Audiology	69	226	24	Si
Biomed	8	194	2	No
Bupa	6	345	2	No
Cleveland	13	303	5	No
Exactly	13	777	2	No
Heart-statlog				
Ionosphere	34	351	2	No
Iris	4	150	3	No
New Thyroid	5	215	3	No
Sonar	60	208	2	No
Vehicle	18	846	4	No
Wine	13	178	3	No
Zoo	16	101	7	No

Algoritmos analizados en el estudio experimental

AS-RST-FS dado en [40], está basado en el algoritmo Sistema de hormigas (AS) desarrollado por Dorigo en su tesis doctoral en 1992 [41], el primer algoritmo de ACO. En AS-RST-FS, ACO es utilizada para

generar subconjuntos de rasgos empleando una aproximación de filtro basada en selección hacia adelante. RST ofrece la función heurística para medir la calidad de un subconjunto de rasgos.

ACS-RST-FS algoritmo dado en [40], aunque está basado en ACS, es similar al anterior. Sin embargo, para esta variante se utiliza la regla de transición probabilística de ACS. Además, es diferente la forma en la que se actualiza el valor de la feromona, ya que esta vez el valor de la feromona se actualiza de forma local y luego de forma global, o sea, cada vez que un nodo correspondiente a un rasgo sea adicionado a un subconjunto se actualizará el valor de la feromona y además para el mejor subconjunto generado en un ciclo.

AFSBRASCO (Algorithm for feature selection based on Rough Set and Ant Colony Optimization), por sus siglas en inglés, dado en [42] es un algoritmo híbrido en el que RST se utiliza para definir la importancia de los rasgos mediante las aproximaciones superior e inferior, las cuales son empleadas como heurísticas para guiar el proceso de selección de rasgos. ACO se utiliza para implementar el método de búsqueda, generando subconjuntos de rasgos que usan una aproximación de filtro basada en la selección hacia adelante. En el caso de este algoritmo las hormigas parten del *CORE* o núcleo, además la feromona está asociada a los arcos denotando la posibilidad de ir por ejemplo a un nodo j desde un nodo i .

RSFSACO dado en [43], este algoritmo la información heurística se calcula dinámicamente durante el proceso de construcción de las soluciones. La importancia de los rasgos se adopta como una información heurística en RSFSACO. La importancia de los rasgos se define por la entropía y la información mutua. En RSFSACO, para la construir una solución cada hormiga debe comenzar a partir del núcleo de rasgos. Lo siguiente la hormiga selecciona aleatoriamente un rasgo, luego selecciona el siguiente rasgo de aquellos que quedan sin seleccionar con una alta probabilidad. Después de que cada hormiga haya construido una solución, se deberá actualizar la feromona de cada arista Si el subconjunto óptimo es encontrado o las iteraciones alcanzan su máximo ciclo, entonces el algoritmo para y devuelve el mínimo reducto de rasgos encontrado. Si ninguna condición se cumple, entonces la feromona se actualiza, se crea un nuevo conjunto de hormigas y el proceso itera una vez más.

ACO-RST-FSP que se propone en [44] se clasifica como “filtro”, utiliza ACO como procedimiento de generación de subconjuntos y como función de evaluación de la calidad de los subconjuntos la medida calidad de la clasificación de la Teoría de los Conjuntos Aproximados. Durante la ejecución del algoritmo cada hormiga construye un subconjunto de rasgos hasta que este alcance un valor de la calidad de la clasificación igual al calculado para el conjunto de todos los rasgos, es decir, hasta formar un reducto.

Modelo propuesto

El algoritmo propuesto, nombrado ACO+PSO+RST para el cálculo de rectos, está basado en el algoritmo ACO-RST-FS [44], expuesto anteriormente. La novedad científica de este algoritmo radica en la forma de inicializar el rastro de feromona. Este valor inicial de la feromona para cada rasgo se asignaba de acuerdo al cálculo de un valor aleatorio $\tau_i(0)$, $i = 1, \dots, nf$. En este método se propone tomar como valores iniciales de la feromona el conjunto de pesos W que se obtiene como resultado de aplicar algoritmo PSO+RST dado en [45]. Este conjunto de pesos W representa la importancia relativa de cada atributo y garantiza que las hormigas partan de los rasgos más relevantes al construir una solución.

El conjunto de pesos de los rasgos obtenido con el método, PSO+RST, ha sido empleado para mejorar el desempeño de algunos métodos de aprendizaje, como son: método de los k-Vecinos más Cercanos usado tanto como aproximador de funciones que como clasificador [46, 47] y un Multilayer Perceptron [48, 49]. En el caso del método k-NN, PSO+RST permite encontrar el conjunto de pesos W usado en la función de semejanza que mejora la recuperación de objetos similares. En el método Perceptron multicapa, por otro lado, se utilizan los pesos calculados basados en el método PSO+RST como pesos iniciales de los enlaces entre la capa de entrada y la capa oculta.

Resultados experimentales

Primeramente, se comparan los algoritmos estudiados respecto a la longitud de los reductos obtenidos, los resultados de este experimento se muestran en la Tabla 2.

Tabla 2 Longitud de los reductos obtenidos por los diferentes métodos de cálculo de reductos

Bases de Casos	RSFSACO	AS-RST-FS	ACS-RST-FS	AFSBRSAO	ACS-RST-FSP	ACO-PSO-RST
Audiology	4	3	3	4	2	2
Biomed	2	2	2	3	2	2
Bupa	3	3	3	4	3	3
Cleveland	5	3	3	5	3	3
Exactly	11	10	10	11	10	10
Heart-statlog	4	3	3	5	3	3
Ionosphere	5	2	2	4	2	2
Iris	3	3	3	4	3	3
New Thyroid	2	2	2	3	2	2
Sonar	1	1	1	2	1	1
Vehicle	3	3	3	5	3	3
Wine	2	2	2	3	2	2
Zoo	7	5	5	7	5	5

Con los resultados mostrados en la tabla anterior se realiza un análisis estadístico. De acuerdo con las peculiaridades que presentan las poblaciones se utilizan pruebas de comparaciones no paramétricas, pues las muestras no siguen una distribución normal y esta relacionadas. Primero se aplica una prueba de Iman-Davenport como una variante de la prueba de Friedman, en caso de que existan diferencias significativas en el grupo se utilizar una variante de la prueba de Holm para comparaciones par a par.

El indicador que describe el tamaño de los reductos calculados presenta un valor de Iman-Davenport con el cual se obtiene 0.000001023009 como valor ajustado y como este es menor que 0.05 (margen de error permitido), se puede decir que existen diferencias significativas entre los algoritmos. En la Tabla 3 se muestran los valores de los rangos promedios de los métodos.

Tabla 3 Rango promedio obtenido por cada método en la prueba de Friedman

Algoritmo	Ranking
RSFSACO	4.1538
AS-RST-FS	2.8462
ACS-RST-FS	2.8462
AFSBRSAO	5.7692
ACS-RST-FSP	2.6923
ACO-PSO-RST	2.6923

En la Tabla 4 se muestra la comparación entre los algoritmos tomando como método de control ACO-PSO-RST, utilizando la prueba de Holm, se puede determinar si se acepta o rechaza la hipótesis de igualdad entre las muestras (si el valor p es menor o igual que el margen de error permitido, dado por la prueba de Holm, la hipótesis es rechazada, en caso contrario se acepta).

Tabla 4 P-valores ajustados obtenidos a través de la aplicación del método de control Holm.

i	Algoritmo	$z = (R_0 - R_i)/SE$	p	Holm	Hipótesis
5	AFSBRSAO	4.193139	0.000028	0.01	Rechazada
4	RSFSACO	1.991741	0.046399	0.0125	Aceptada
3	AS-RST-FS	0.209657	0.833935	0.016667	Aceptada

2	ACS-RST-FS	0.209657	0.833935	0.025	Aceptada
1	ACS-RST-FSP	0	1	0.05	Aceptada

Los resultados de la prueba de Holm permiten asegurar que existen diferencias significativas entre el método propuesto ACO-PSO-RST, en cuanto a la longitud de los reductos encontrados, respecto al método AFSBRSACO, no siendo así respecto al resto de los algoritmos estudiados entre los cuales no existen diferencias significativas.

Por otro lado, los métodos AS-RST-FS, ACS-RST-FS y AFSBRSACO tienen la desventaja de no tener en cuenta mantener o no afectar significativamente la calidad de la clasificación obtenida con el conjunto reducido de rasgos, es decir reducen el número de atributos a tener en cuenta sacrificando la calidad del aprendizaje con la pérdida de información relevante. En la Tabla 5 se muestra el comportamiento de la calidad de la clasificación obtenida con las bases preprocesadas por los diferentes algoritmos en estudio.

Tabla 5 Calidad de la clasificación utilizando los reductos obtenidos por los métodos estudiados.

Bases de Casos	Con todos los rasgos	RSFSACO	AS-RST-FS	ACS-RST-FS	AFSBRSACO	ACS-RST-FSP	ACO-PSO-RST
Audiology	1	1	0.0708	0.0708	0.0708	1	1
Biomed	1	1	1	1	1	1	1
Bupa	1	1	1	1	1	1	1
Cleveland	1	1	1	1	1	1	1
Exactly	1	1	1	1	1	1	1
Heart-statlog	1	1	1	1	1	1	1
Ionosphere	1	1	1	1	1	1	1
Iris	1	1	1	1	1	1	1
New Thyroid	1	1	1	1	1	1	1
Sonar	1	1	1	1	1	1	1
Vehicle	1	1	0.9976	0.9976	0.9976	1	1
Wine	1	1	1	1	1	1	1
Zoo	1	1	1	1	1	1	1

Una pregunta clave para un enfoque ACO es la siguiente: ¿Cómo se inicializa el rastro de feromona? Esto incluye un algoritmo de inicialización y un algoritmo de evaporación. Los valores de feromona pueden ser inicializados de forma aleatoria, pero es importante tener en cuenta que el rendimiento dependerá en gran medida del valor de esta feromona.

Se ha demostrado que una de las principales razones de la lenta convergencia y la incapacidad para lograr generalización en los algoritmos basados en ACO es la falta de una correcta inicialización de los valores de feromona que se ajustan. Sofisticados procedimientos de aprendizaje no son todavía capaces de compensar los malos valores iniciales de feromona, mientras que una buena inicialización llevaría a una convergencia y/o capacidad de generalización rápida. Si se logra situar el rastro de feromona lo más cercano posible a una buena solución, el espacio de búsqueda se puede reducir y el aprendizaje se hará más rápido.

Convencionalmente, la regla empleada para la inicialización de la feromona es el uso de pequeños valores aleatorios, por lo general los valores iniciales del rastro de feromona toman valores aleatorios distribuidos uniformemente en pequeños intervalos $[-1, 1]$. El motivo de esto es que los grandes valores absolutos de peso pueden causar que algunos nodos de la red sean muy activos o sean inactivos para todas las muestras de entrenamiento, y por lo tanto insensible a los procesos de aprendizaje. Sin embargo, los valores muy pequeños también pueden retardar el aprendizaje. Esta inicialización puede ser adecuada en muchos problemas sencillos. Sin embargo, en los problemas más difíciles se ha encontrado que las

inicializaciones más deterministas pueden mejorar la convergencia del aprendizaje drásticamente en comparación con la inicialización aleatoria.

Respecto a *RSFSACO* y *ACS-RST-FSP* el método propuesto *ACO-PSO-RST* tiene la ventaja adicional de utilizar el conjunto de pesos W que se obtiene como resultado de aplicar algoritmo *PSO+RST* como valores iniciales de feromona, mejorando la convergencia del aprendizaje y la capacidad de una rápida generalización.

Aporte practico al problema de pronóstico de la calidad del agua para el riego de los cultivos

El agua para el riego de los campos de cultivo supone el mayor porcentaje de consumo en el mundo, ascendiendo a 2 680 km³ anualmente, de allí se estima, que el 70% del agua dulce en la Tierra es empleada para esta actividad. Una de las mayores preocupaciones en la utilización del agua con este fin es el exceso de sales, pues esto conlleva a la degradación de estos, implicando una disminución de su capacidad productiva, pérdida de recursos, mayor necesidad de aportes agrícolas, descenso del valor y una reducción de la población tanto animal, como vegetal.

A nivel mundial existen distintos criterios de interpretación de la calidad de las aguas de riego, destacándose varios autores e instituciones. En Cuba, basada en el anteproyecto de la Norma Cubana de la Metodología de Evaluación de las Aguas de Riego según las propiedades de los suelos, existe la Propuesta de la Dirección Nacional de Suelos y Fertilizantes del Ministerio de la Agricultura de Cuba, que establece la siguiente clasificación: Aguas adecuadas sin ninguna restricción de uso, Aguas con restricciones de ligeras a moderadas en su uso, Aguas con serias restricciones en su uso o Aguas inadecuadas. Esta considera indicadores de calidad, tales como: conductividad eléctrica (CE), Sales Solubles Totales (SST), Relación de Adsorción del Sodio Ajustado (RAS ajustado), Salinidad Efectiva, entre otros. La presente investigación utiliza la norma anteriormente mencionada, por lo abarcadora que es, pues no solo tiene en cuenta las características físicas, químicas y biológicas del agua, sino también, las propiedades de los suelos y la tolerancia de los cultivos.

A partir de las características del agua, las propiedades de los suelos y la tolerancia de los cultivos es posible finalmente predecir la clasificación del agua para riego. Los expertos declararon que ellos normalmente utilizan 12 indicadores para juzgar la calidad del agua para el riego (Tabla 6). De esta manera, se tienen 12 atributos predictores y un atributo objetivo para cada caso, la base de conocimiento cuenta con 860 instancias. Este es un típico problema de clasificación supervisada, donde se quiere predecir la clase correspondiente a los valores de la calidad del agua para el riego, dado un nuevo objeto.

Tabla 6 Descripción de las variables de la base de casos de calidad del agua

Atributos	Tipo	Posible escala de valores
CE	Númerico	[0.05,8.43]
Tipo Riego	Nominal	{Aspersión,Superficial}
Sensibilidad cultivo	Nominal	{Semitolerante,Sensible,Tolerante}
Tipo Suelo	Nominal	{Arenoso_con_Alta_Velocidad_de_Infiltración,Arenoso_con_Baja_Velocidad_de_Infiltración, Latosolizado_con_Alta_Velocidad_de_Infiltración, Latosolizado_con_Baja_Velocidad_de_Infiltración, No_Calcáreos_con_Alta_Velocidad_de_Infiltración, No_Calcáreos_con_Baja_Velocidad_de_Infiltración, Calcáreos_con_Alta_Velocidad_de_Infiltración, Calcáreos_con_Baja_Velocidad_de_Infiltración, Montmorilloníticos }
SST	Númerico	[32.0,5395.2]
Salinidad efectiva	Númerico	[0.17,871.5]
Sodio	Númerico	[0.03,189.45]
RAS Ajustado	Númerico	[-0.87,31.58]
Carbonato y bicarbonato	Númerico	[0.05,569.9]
Cloro	Númerico	[0.08,78.1]

Mg sales disueltas	Numérico	[8.03,86.45]
Boro	Numérico	[0.0,6.78]
Calidad AGUA	Nominal	Adecuada, con restricciones ligeras, con serias restricciones, inadecuada

Teniendo en cuenta las características de los datos, es viable la aplicación del método propuesto en esta investigación con el objetivo de eliminar aquellos atributos irrelevantes que pueden influir de manera nociva en la posterior asertividad del clasificador a la hora de predecir los valores de la calidad del agua. Al aplicar el algoritmo se obtuvieron resultados favorables, pues se logra reducir significativamente el número de atributos de la base de casos sin afectar la calidad de la clasificación. La base de conocimiento preprocesada para predecir la calidad del agua consta de un atributo predictor el cual se corresponde con el indicador de concentración de Carbonato y bicarbonato en el agua y corrobora investigaciones realizadas por el Departamento de Agricultura de los Estados Unidos como se muestra en [50].

4. CONCLUSIONES

Como resultado de esta investigación se realizó un estudio detallado del comportamiento del algoritmo propuesto ACO-PSO-RST que combinando la Optimización basada en Colonias de Hormigas y la Teoría de Conjuntos Aproximados, proporciona a los investigadores otra alternativa que permite encontrar subconjuntos reducidos de atributos, capaces de representar la información necesaria en problemas de aprendizaje supervisado. Particularmente se encontraron buenas soluciones al aplicar este algoritmo en el preprocesamiento de los datos para optimizar el pronóstico de la calidad del agua para el riego. El método que se estudia en la investigación para el cálculo de reductos logra reducciones altamente significativas de la cantidad de atributos respecto al conjunto original de datos, mientras que la medida calidad de la clasificación por clases de la Teoría de los Conjuntos Aproximados no se vio afectada por los conjuntos de atributos reducidos. Se realizaron comparaciones con otros métodos de selección de rasgos y los resultados obtenidos demuestran la superioridad de la propuesta. Estos resultados están apoyados por las pruebas estadísticas no paramétricas realizadas. En la solución al problema del pronóstico automatizado de la calidad del agua para el riego, se realizó un preprocesamiento de los datos para seleccionar los atributos relevantes. Cualquiera de los reductos de longitud mínima facilita a los especialistas qué preguntas realizar acerca de variables originales con potencia suficiente para un buen pronóstico.

5. REFERENCIAS BIBLIOGRÁFICAS

1. Liu, H. and H. Motoda, *Computational Methods of Feature Selection*. 2007.
2. Bell, D. and H. Wang, *A formalism for relevance and its application in feature subset selection*. Machine Learning, 2000.
3. Blum, A. and P. Langley, *Selection of relevant features and examples in machine learning*. Artificial Intelligence, 1997.
4. Liu, H. and L. Yu, *Toward integrating feature selection algorithms for classification and clustering*, in *IEEE Trans. On knowledge and data engineering*. 2005. p. 1-12.
5. Gamberger, D. and N. Lavrac, *Conditions for ocam's razor applicability and noise elimination*, in *9th European Conf. on Machine Learning*. 1997.
6. Kohavi, R. and B. Frasca, *Useful feature subsets and rough set reducts*, in *3rd Int. Workshop on Rough Set and Soft Computing*. 1994.
7. Yang, J. and V. Honavar, *Feature extraction, construction and selection*, K.A. Publishers, Editor. 1998. p. 117-136.
8. Dudani, S., *The distance-weighted k-nearest-neighbor rule*, in *Man and Cybernetics*. 1975, IEEE Transactions on Systems.
9. Breiman, L., et al., *Classification and regresion trees*, in *Wadsworth Int. Group*. 1984: Belmont CA.
10. Kohavi, R. and G. John, *Wrappers for feature subset selection*. Artificial Intelligence, 1997.

11. Koller, D. and M. Sahami, *Toward optimal feature selection*, in *13th Int. Conf. on Machine Learning*. 1996: Bari, Italia. p. 284-292.
12. Kudo, M., et al., *Comparison of classifier specific feature selection algorithms*. 2000: p. 677-686.
13. Hall, M., *Correlation-based Feature Selection for Machine Learning*, in *Department of Computer Science*. 1999, University of Waikato: Hamilton, New Zealand.
14. Davies, S. and S. Russell, *Np-completeness of searches for smallest possible feature sets*, in *AAAI Fall Symposium on Relevance*. 1994: N. Orleans, LA. p. 37-39.
15. Liu, H. and L. Yu, *Feature selection for data mining*. 2002, Arizona State University: Temp, Arizona.
16. Siedlecki, W. and J. Sklansky, *On automatic feature selection*. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 1988. 2: p. 197-220.
17. Ruiz, D.R., *Selección de Atributos mediante proyecciones*, in *Departamento de Lenguajes y Sistemas Informáticos*. 2005, Universidad de Sevilla: Sevilla. p. 180.
18. Pawlak, Z., *Rough sets*. *International Journal of Information & Computer Sciences* 1982. 11: p. 341-356.
19. Pawlak, Z., *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishing, 1991.
20. Choubey, S.K., *A comparison of feature selection algorithms in the context of rough classifiers*, in *Fifth IEEE International Conference on Fuzzy Systems*. 1996.
21. Chouchoulas, A. and Q. Shen, *A rough set-based approach to text classification*. *Lectures Notes in Artificial Intelligence*, 1999. 1711: p. 118-127.
22. Greco, S. and M. Inuiguchi, *Rough Sets and Gradual Decision Rules. Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, in *9th International Conference RSFDGRC2003*. 2003: Chongqing, China.
23. Grzymala-Busse, J.W. and S. Siddhaye, *Rough set approaches to rule induction from incomplete data*, in *10th International Conference on Information Processing and Management of Uncertainty in Knowledge-Bases systems IPMU2004*. 2004: Perugia, Italy.
24. Miao, D. and L. Hou, *An Application of Rough Sets to Monk's Problems Solving. Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, in *9th International Conference, RSFDGRC2003*. 2003: Chongqing, China.
25. Midelfart, H. and J. Komorowski, *Learning rough set classifiers from gene expression and clinical data*. *Fundamenta Informaticae*, 2003. 53: p. 155-183.
26. Piñero, P. and L. Arco, *Two New Metrics for Feature Selection in Pattern Recognition*. *Lectures Notes in Computer Science* 2003. LNCS 2905: p. 488-497.
27. Sugihara, K. and H. Tanaka, *Rough Sets approach to information systems with interval decision values in evaluation problems*, in *The International Symposium on Fuzzy and Rough Sets ISFUROS2006*. 2006: Santa Clara, Cuba.
28. Tsumoto, S., *Automated extraction of hierarchical decision rules from clinical databases using rough set model*. *Expert systems with Applications*, 2003. 24: p. 189-197.
29. Zhao, Y. and H. Zhang, *Classification Using the Variable Precision Rough Set. Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, in *9th International Conference, RSFDGRC2003*. 2003: Chongqing, China.
30. Parsons, S., *Current approaches to handling imperfect information in data and knowledges bases*. *IEEE Transaction On knowledge and data enginnering*, 2006.

31. Grabowski, A., *Basic Properties of Rough Sets and Rough Membership Function*. Journal of Formalized Mathematics, 2003. 15.
32. Skowron, A., *New directions in Rough Sets*, in *7th International Workshop (RSFDGRC'99)*. 1999: Japan.
33. Skowron, A. and J.F. Peters, *Rough Sets: Trends and Challenges*, in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing 9th International Conference, RSFDGRC 2003*. 2003.
34. Düntsch, I. and G. Gediga, *Rough Set Data Analysis: A road to non-invasive knowledge discovery*. Methodos Publishers, 2000.
35. Kennedy, J. and R.C. Eberhart, *Particle swarm optimization*, in *IEEE International Conference on Neural Networks*. 1995: Piscataway, New York, USA.
36. Dorigo, M., M. Birattari, and T. Stutzle, *Ant colony optimization*. Computational Intelligence, 2006. 1(4): p. 28-39.
37. Dorigo, M. and T. Stutzle, *Ant Colony Optimization*, in *Cambridge Massachussetts*. 2004, MIT Press.
38. Jensen, R. and Q. Shen. *Finding Rough Set Reducts with Ant Colony Optimization*. in *UK Workshop on Computational Intelligence*. 2003.
39. Holm, S., *A simple sequentially rejective multiple test procedure*. *Journal of Statistics*, 1979. 6: p. 65-70.
40. Bello, R., et al., *A Model based on Ant Colony System and Rough Set Theory to Feature Selection*, in *Genetic and Evolutionary Computation Conference (GECCO05)*. 2005: Washington DC, USA.
41. Dorigo, M., *Optimization, Learning and Natural Algorithms*. 1992, Politecnico di Milano.
42. Ming, H., *Feature Selection Based on Ant Colony Optimization and Rough Set Theory*, in *International Symposium on Computer Science and Computational Technology*. 2008 Beijing University of Technology, Beijing, China.
43. Chen, Y., D. Miao, and R. Wang, *A rough set approach to feature selection based on ant colony optimization*. Pattern Recognition Letters, 2010. 31: p. 226-233.
44. Gómez, M.Y., *Algoritmos que combinan conjuntos aproximados y optimización basada en colonias de hormigas para la selección de rasgos. Extensión a múltiples fuentes de datos.*, in *Departamento de Ciencia de la Computación*. 2010, Universidad Central "Marta Abreu" de Las Villas: Santa Clara. p. 105.
45. Filiberto, Y. and R. Bello, *Un método para el cálculo de los pesos de los atributos en sistemas de decisión continuos basada en nueva medida de la teoría de los conjuntos aproximados.*, in *VII Congreso Nacional de Reconocimiento de Patrones, RECPAT2009*. 2009: Santiago de Cuba, Cuba.
46. Filiberto, Y., et al., *Using PSO and RST to predict the resistant capacity of connections in composite structures*, in *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*. 2010, Springer. p. 359-370.
47. Filiberto, Y., et al. *A method to build similarity relations into extended Rough Set Theory*. in *2010 10th International Conference on Intelligent Systems Design and Applications*. 2010. IEEE.
48. Coello, L., et al., *Improving the multilayer perceptron learning by using a method to calculate the initial weights with the similarity quality measure based on fuzzy sets and particle swarms*. Computación y Sistemas, 2015. 19(2): p. 309-320.
49. Cabrera, Y.F., et al. *Improving the MLP learning by using a method to calculate the initial weights of the network based on the quality of similarity measure*. in *Mexican International Conference on Artificial Intelligence*. 2011. Springer.
50. Duarte, O., *Calidad Química del Agua*, in *Tecnología de Tierras y Aguas I*. 2003 México.