

# CONTRIBUCIONES AL APRENDIZAJE AUTOMATIZADO A TRAVÉS DE LA TEORÍA DE LOS CONJUNTOS APROXIMADOS EXTENDIDA

**Autoría principal:** Yailé Caballero Mota

**Otros autores:** Yaima Filiberto Cabrera, Rafael Bello Pérez, Rafael Larrúa Quevedo, Enislay Ramentol Martínez y Francisco Herrera Triguero

**Colaboradores:** 34

**Entidad ejecutora principal:** Universidad de Camagüey

**Otras entidades participantes:** Facultad de Matemática, Física y Computación, Universidad Central “Marta Abreu” de Las Villas (UCLV), Facultad de Construcciones, Universidad de Camagüey, Facultad de Informática, Universidad de Granada, España

**Dra. Yailé Caballero Mota** (29%). Director general del grupo investigador, participó en todas las líneas de trabajo.

**Dra. Yaima Filiberto Cabrera** (25%). Implementación del nuevo algoritmo de generación de reglas para problemas balanceados y modificaciones para problemas con desbalance en los datos. Sistema inteligente para la predicción en Ingeniería Civil. Sistema inteligente para la predicción de variables meteorológicas a mediano y largo plazos.

**Dr. Rafael Bello Pérez** (18%). Creación de métodos y algoritmos: Nueva medida denominada calidad de la similaridad de un sistema de decisión, nuevo método para construir relaciones de similaridad, nuevos algoritmos de generación de reglas.

**Dr. Rafael Larrúa Quevedo** (15%). Director del grupo investigador de Ingeniería Civil, participó en todas las líneas de trabajo relacionadas con esta área.

**MSc. Enislay Ramentol Martínez** (8%). Implementación de algoritmos para la edición de conjuntos de entrenamiento desbalanceados.

**Dr. Francisco Herrera Triguero** (5%). Diseño de nuevos algoritmos para la edición de conjuntos de entrenamiento desbalanceados.

**Autor para la correspondencia:**

Dra. Yailé Caballero Mota

Dirección postal: Avenida Ignacio Agramante # 140 % Avenida Camagüey y Aserradero. La Zambrana, Camagüey, Cuba. CP: 70300. Fax. 53-32-293799

## RESUMEN

Las investigaciones realizadas en el empleo de la Teoría de los Conjuntos Aproximados en el aprendizaje automático mostraron las limitaciones del enfoque clásico de esta teoría para el caso de dominios de aplicación con datos mezclados. Para superar estas limitaciones se han desarrollado investigaciones sobre los métodos de aprendizaje a través de la Teoría de los Conjuntos Aproximados Extendida (TCAE), en problemas de aproximación de funciones y en problemas de clasificación para datos balanceados y desbalanceados. El problema a resolver está dado por la necesidad de desarrollar la TCAE para facilitar el desarrollo de las técnicas de aprendizaje automático basadas en ella. Los objetivos del trabajo están dirigidos a obtener un método para la construcción de las relaciones de similitud necesarias para la aplicación de la TCAE en un dominio de aplicación dado, y a partir de ella mejorar las técnicas de aprendizaje automático o en la mejora de los existentes. Los principales resultados de la investigación son los siguientes:

- a) Medida denominada calidad de la similitud de un sistema de decisión.
- b) Método para construir relaciones de similitud, a partir de la combinación de la metaheurística optimización basada en partículas y la medida propuesta en esta investigación.
- c) Empleo de la relación de similitud y los pesos de los rasgos para mejorar el desempeño del método de los k-Vecinos más Cercanos; así como en la inicialización de la red neuronal Perceptron Multicapa a partir de los pesos de los rasgos.
- d) Algoritmos de edición de conjuntos de entrenamientos desbalanceados.
- e) Algoritmos para el descubrimiento de reglas de clasificación para problemas balanceados y desbalanceados.
- f) Aplicación de los resultados científicos obtenidos en la creación de sistemas inteligentes para la predicción en Ingeniería Civil y para la predicción de variables meteorológicas a mediano y largo plazos.

La producción científica asociada a estos resultados consiste en la publicación de 28 trabajos, de ellos: 14 en revistas y bases de datos referenciadas, así como la presentación de 14 ponencias en prestigiosos eventos científicos internacionales y el registro por CENDA de 6 productos de software. Además, forman parte de los resultados 24 tesis defendidas: 1 tesis de doctorado, 9 tesis de maestría y 14 trabajos de diplomas. Se han obtenido 14 Premios nacionales e internacionales. Se arriba a las conclusiones siguientes: los métodos propuestos han sido estudiados experimentalmente usando bases de datos internacionales; así como su aplicación para la solución de problemas reales en dos áreas del conocimiento: la Ingeniería Civil y la Meteorología. Se han obtenido resultados novedosos y relevantes desde el punto de vista teórico y práctico, lo cual se demuestra en la producción científica asociada y en los avales que se presentan en los anexos.

**Nota:** Los resultados que sustentan este trabajo son novedosos y en ningún caso coinciden con los presentados en el Premio de la Academia de Ciencias de Cuba 2009 “Métodos para el procesamiento de los conjuntos de entrenamiento en el aprendizaje automatizado basados en la Teoría de los Conjuntos Aproximados”, aunque corresponden a la misma línea de investigación. Los aportes merecedores del Premio anterior radicaban en algoritmos y métodos para el preprocesamiento de los datos, previo a la clasificación supervisada, a través de la Teoría de los Conjuntos Aproximados clásica. Los algoritmos y métodos que se proponen en el presente trabajo se basan en la Teoría de los Conjuntos Aproximados Extendida; además, los algoritmos de edición de conjuntos de entrenamiento que se proponen son para datos desbalanceados. La aplicación, en el 2009 fue en el pronóstico de variables meteorológicas a corto plazo. En el presente año las aplicaciones son en Ingeniería Civil y en el pronóstico meteorológico a mediano y largo plazos.

## COMUNICACIÓN CORTA

Se presentan los resultados obtenidos al aplicar la Teoría de los Conjuntos Aproximados Extendida en el aprendizaje automático, particularmente en problemas de aproximación de funciones y en problemas de clasificación para datos balanceados y desbalanceados; así como datos mezclados. El aporte científico principal del trabajo científico desarrollado está dado por los aportes en el desarrollo de esta teoría y las nuevas técnicas de aprendizaje automático y las propuestas de mejoras de otras existentes. A partir de estos resultados se han desarrollado sistemas inteligentes para resolver problemas reales. Los métodos propuestos han sido estudiados experimentalmente usando bases de datos internacionales; así como su aplicación en dos áreas del conocimiento: la Ingeniería Civil y la Meteorología.

### Introducción

La Teoría de los Conjuntos Aproximados (RST) ha sido utilizada como base para construir nuevas técnicas de aprendizaje automático; como ejemplo de esos resultados están los alcanzados por diferentes autores y los presentados en [1]. Sin embargo, la RST clásica está definida para problemas con rasgos de dominio discreto, de modo que los métodos para el descubrimiento de conocimiento derivados de ella se ven afectados por la presencia de rasgos con dominio continuo o datos mezclados (existencia simultánea en el dominio de aplicación de variables con dominio discreto y continuo), requiriendo usualmente un proceso de discretización de los rasgos continuos.

Dado el hecho de que en muchos de los problemas que surgen en la realidad aparecen datos mezclados, esto resulta una limitante importante. Para enfrentar esta problemática en la RST se han desarrollado nuevos conceptos y métodos basados en relaciones de similitud en lugar de relaciones de equivalencia, esta se conoce como RST extendida. Ello da lugar a la necesidad de encontrar las

relaciones de similitud apropiadas para cada dominio de aplicación, convirtiéndose en un problema relevante a resolver para aplicar los métodos de la RST extendida.

El objetivo fundamental de la investigación es diseñar métodos que faciliten la construcción de relaciones de similitud y permitan mejorar el desempeño de métodos de descubrimiento de conocimiento para sistemas de decisión con datos mezclados, incluyendo el caso en que el rasgo de decisión tiene dominio continuo.

### **Descripción de los resultados alcanzados**

Todos los resultados de esta investigación incluidos en esta propuesta de premio están definidos para el caso de problemas con datos mezclados, y se basan en el empleo de los conceptos de la RST formulados usando relaciones de similitud. Los aportes principales que se describen en esta sección son una nueva medida denominada calidad de la similitud de un sistema de decisión; y un nuevo método para construir relaciones de similitud basada en ella. A partir de estos resultados se proponen mejoras al desempeño del método de los k- Vecinos más Cercanos y para el entrenamiento de la red neuronal Perceptron Multicapa; así como nuevos algoritmos de edición de conjuntos de entrenamientos desbalanceados; nuevos algoritmos de generación de reglas de clasificación tanto para problemas balanceados como con desbalance en los datos; y la aplicación de estos nuevos métodos en la implementación de sistemas inteligentes para la predicción en Ingeniería Civil y para la predicción de variables meteorológicas a mediano y largo plazos. Seguidamente se describen brevemente estos resultados.

### **Nueva medida denominada calidad de la similitud de un sistema de decisión**

La definición de esta medida parte de dos aspectos básicos sobre el universo de los objetos: la granularidad de la información y el principio del razonamiento basado en casos. La medida propuesta representa el grado en el cual la similitud entre los objetos de acuerdo a los rasgos de condición es equivalente a la similitud que se obtiene de acuerdo al rasgo de decisión [2]; es decir, mide la similitud entre la granulación según los rasgos de condición y la granulación según los rasgos de decisión.

### **Nuevo método para construir relaciones de similitud**

El método propuesto, denominado PSO+RST, permite construir relaciones de similitud 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 [3]. En este caso, el objetivo del algoritmo de aprendizaje es encontrar el conjunto de pesos que maximiza la medida calidad de la similitud. La búsqueda del conjunto de pesos se realiza mediante una metaheurística; el nombre de PSO+RST estuvo determinado porque la primera metaheurística empleada fue PSO (Particle Swarm Optimization), aunque luego se han obtenidos buenos resultados con otras metaheurísticas [4].

### **Mejora del desempeño del método de los k-Vecinos más Cercanos.**

La idea básica del método k-NN es que valores de entradas similares tienen salidas similares. Un aspecto importante en los métodos basados en grados de similaridad es la forma de recuperación de los casos similares, en lo cual el uso de funciones de similaridad definidas como sumas pesadas mejoran significativamente el desempeño del método. La investigación realizada muestra como el empleo de los pesos calculados según el método PSO+RST produce una mejora en el desempeño del k-NN, tanto para problemas de aproximación de funciones como de clasificación; los resultados obtenidos fueron comparados con los alcanzados usando otros métodos como el del Gradiente Conjugado y el método Relief y se demostró la superioridad de los resultados al emplearse el método aquí propuesto [5] y [6].

### **Método para la inicialización de la red neuronal Perceptron Multicapa.**

Las Redes Neuronales Artificiales como el Perceptron Multicapa (MLP) son modelos apropiados para la solución de problemas no lineales. La construcción de redes neuronales es un problema de optimización no lineal en la que el objetivo es encontrar un conjunto de pesos que minimizan cierta función de costo. Esta función de costo puede ser caracterizada por un gran número de mínimos locales en las proximidades del mínimo global. La topología de la red y los pesos iniciales juegan un papel muy importante. Una pregunta clave para una red neuronal es la siguiente: ¿Cómo se calcula el conjunto de pesos de la red neuronal? Esto incluye un algoritmo de inicialización y un algoritmo de entrenamiento que ajusta los valores de los pesos. Los pesos pueden ser inicializados de forma aleatoria, pero es importante tener en cuenta que la eficacia y eficiencia del entrenamiento dependerá en gran medida el valor de estos pesos. Se ha demostrado que una de las principales razones de la lenta convergencia y la incapacidad para lograr generalización en el entrenamiento de un MLP es la falta de una correcta inicialización de los pesos que se ajustan. Sofisticados procedimientos de aprendizaje no son todavía capaces de compensar los malos valores iniciales de pesos, mientras que una buena inicialización llevaría a una convergencia o capacidad de generalización rápida.

En esta investigación se propone una nueva alternativa para inicializar los pesos de una red multicapas. En este caso, 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. Los experimentos realizados muestran el buen desempeño de la red Perceptron Multicapa al inicializar los pesos a través de la nueva alternativa que se propone [7].

### **Nuevos algoritmos de edición de conjuntos de entrenamientos desbalanceados.**

El aprendizaje a partir de conjuntos de entrenamientos no balanceados es uno de los grandes desafíos que enfrenta la comunidad científica del aprendizaje automatizado, debido al mal desempeño de los algoritmos frente a conjuntos de

este tipo. La ocurrencia de sucesos poco frecuentes provoca la desproporción de la representatividad de las clases. Los algoritmos de clasificación obtienen alta precisión con los elementos de la clase más representada (mayoritaria), ocurriendo lo contrario con los de la clase menos representada (minoritaria). En los conjuntos no balanceados el conocimiento original suele estar en la clase minoritaria, sin embargo muchos clasificadores suelen etiquetarlos como rarezas o ruido, enfocándose exclusivamente en las medidas globales. Se proponen nuevos algoritmos basados en la RST. El algoritmo **SMOTE-RSB\*** genera instancias sintéticas de la clase minoritaria usando el método SMOTE y la medida de calidad de clasificación de la RST [8]. El algoritmo **SMOTE-FRST** [9] genera ejemplos sintéticos usando SMOTE, y aplica un método de limpieza basado en la selección de prototipos; los estudios experimentales han mostrado su eficacia especialmente en presencia de ruido. El algoritmo **FRIPS** (Fuzzy Rough Imbalanced Prototype Selection) [10] y [11], es un método de selección de prototipos para conjuntos de datos no balanceados, basado en la RST y los conjuntos borrosos.

### **Nuevos algoritmos para el descubrimiento de reglas de clasificación para problemas balanceados y desbalanceados.**

Otro resultado alcanzado en esta investigación es un algoritmo para la inducción de reglas de clasificación: **IRBASIR** [12] y [13], el cual permite descubrir conocimiento a partir de sistemas de decisión con datos mezclados; lo cual hace que no requiera realizar ningún proceso de discretización, ni antes del aprendizaje como ID3 o LEM2 ni durante el aprendizaje como C4.5 o MODLEM (algoritmos clásicos en el descubrimiento de reglas de clasificación). El algoritmo induce reglas de la forma *Si P entonces Q*, pero en este caso la condición *P* tiene la forma  $\sum w_i * a_i() \geq \epsilon$ , donde  $w_i$  es el peso del rasgo  $i$ ,  $a_i()$  es la función de comparación para el rasgo  $i$  y  $\epsilon$  es un umbral; los pesos se calculan según el método PSO-RST. Este algoritmo busca el conjunto mínimo de reglas siguiendo una estrategia de cubrimiento secuencial, para lo cual construye clases de similaridad de los objetos del sistema de decisión. Se construyó el algoritmo **imbalance-IRBASIR** [14], el cual es una modificación de IRBASIR para el caso desbalanceado, a partir de la modificación de la medida calidad de la similaridad.

### **Aplicabilidad de la investigación en la solución de problemas reales**

Se mostró la efectividad de los métodos propuestos en la solución de problemas reales relacionados con las ramas de la Ingeniería Civil y la Meteorología. Esto incluye la construcción de un sistema inteligente para el pronóstico de la capacidad resistente en mampostería, predicción de la capacidad resistente de los Conectores y a la predicción de respuestas diversas de muros de mampostería confinada [15]; como Ayuda a los especialistas en Ingeniería Civil se desarrollaron los sistemas automatizados **PROCON v1.0**, **PROCON v2.0** y **PROCON v3.0**.

En el área de la Meteorología se solucionaron problemas relacionados con los pronósticos de variables meteorológicas a mediano y largo plazos de las variables Anomalías de las temperaturas mínimas, máximas y las precipitaciones. Los mejores resultados se obtuvieron cuando se usó el método propuesto de cálculo

de pesos PSO+RST con el estimador de funciones k-NN. En este sentido se desarrolló el sistema automatizado **PROVARMET 1.0**. Todos los softwares han sido registrados.

### **Conclusiones**

Se proponen resultados que facilitan el empleo de la de la Teoría de los Conjuntos Aproximados Extendida. Se define una medida, denominada calidad de la similaridad, que permite dar una valoración de la calidad de la información contenida en un sistema de decisión con datos mezclados; usando esta medida se propone un método para construir relaciones de similaridad. El uso de la relación de similaridad y los pesos de los rasgos calculados mediante el método propuesto en varias técnicas del aprendizaje automatizado permitió valorar que estos permiten mejorar el desempeño de estas técnicas. Se solucionaron problemas reales relacionados con las ramas de la Ingeniería Civil y la Meteorología.

### **Referencias**

[1] Caballero, Y et al. Métodos para el procesamiento de los conjuntos de entrenamiento en el aprendizaje automatizado basados en la Teoría de los conjuntos aproximados. Premio nacional de la Academia de Ciencia de Cuba 2009.

La producción científica asociada a estos resultados consiste en la publicación de **28** trabajos, de ellos: **14** en revistas y bases de datos referenciadas, así como la presentación de **14** ponencias en prestigiosos eventos científicos internacionales y el registro por CENDA de **6** productos de software. Además, forman parte de los resultados **24** tesis defendidas: **1** tesis de doctorado, **9** tesis de maestría y **14** trabajos de diplomas. Se han obtenido **14** Premios nacionales e internacionales como expresión del reconocimiento a los resultados obtenidos.

A continuación se detallan los resultados alcanzados. Se han obtenido las siguientes publicaciones científicas en forma de artículos, ponencias y capítulos o epígrafes de libros. **En el Anexo I aparecen los certificados que acreditan estas publicaciones.**

### **Artículos científicos en Revistas o Libros: Publicaciones en Revistas y bases de datos referenciadas:**

[1] Yumika Fernández, Rafael Bello, Yaima Filberto, Mabel Frias and Yaile Caballero. Effects of Using Reducts in the Performance of the Irbasir Algorithm. Revista DYNA. Indizada por: THOMSON REUTERS, SciELO Colombia, LATINDEX, ACTUALIDAD IBEROAMERICANA, GEOREF, Red ALyC, CHEMICAL ABSTRACTS, DOAJ (Directory of Open Access Journal). Revista de la Facultad de Minas de la Universidad Nacional de Colombia. Sience Citation Index Expanded (SciSearch), Journal Citation Reports/Science Edition. 2013 (*Para salir en la próxima edición*)

- [2] Nele Verbiest, Enislay Ramentol, Chris Cornelis, Francisco Herrera. Preprocessing Noisy Imbalanced Datasets using SMOTE enhanced with Fuzzy Rough Prototype Selection. *Sometido a la Revista Pattern Recognition Letters*. Elsevier. Indizada por: Science Citation Index, SCOPUS.
- [3] Yaima Filiberto, Rafael Bello, Yaile Caballero and Mabel Frias. An analysis about the measure quality of similarity and its applications in machine learning. Fourth International Workshop of Knowledge Discovery, Knowledge Management and Decision Making. EUREKA 2013. Published by Atlantis Press. 2013. Indizada por: SCOPUS.
- [4] E. Ramentol, Y. Caballero, R. Bello, F. Herrera. SMOTE-RSB\*: A Hybrid Preprocessing Approach based on Oversampling and Undersampling for High Imbalanced Data-Sets using SMOTE and Rough Sets Theory. *Revista Knowledge and Information Systems (KAIS)*. DOI: 10.1007/s10115-011-0465-6. ISSN: 0219-1377. E-ISSN: 0219-3116. Springer Verlag. Noviembre 2011. Volume 33, Issue 2, Pp: 245-265. Indizada por: Science Citation Index (SciSearch), SCOPUS. Factor de impacto: 2.25.
- [5] 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 measure. Part II, LNAI 7095, pp. 351-362. ISBN: 978-3-642-25329-4. DOI: 10.1007/978-3-642-25330-0\_31. Springer, Heidelberg (2011). Indizada por: Science Citation Index, SCOPUS.
- [6] Filiberto Y., Bello R., Caballero Y., Larrua, R. Una medida de la Teoría de los Conjuntos Aproximados para sistemas de decisión con rasgos de dominio continuo. *Revista de Ingeniería de la Universidad de Antioquia*. Indizada por: Science Citation Index Expanded (SCIE) del ISI. Journal Citation Report (JCR) Institute for Scientific Information (ISI). ISSN 0120-6230, Número 60, Pp. 141-152. Septiembre, 2011.
- [7] Filiberto Y., Bello R., Caballero Y., Frias M. Algoritmo para el aprendizaje de reglas de clasificación basado en la Teoría de los Conjuntos Aproximados Extendida. *Revista DYNA*, Vol. 68. Nro. 169. Pp: 62-70. ISSN 0012-7353. Indizada por: Thomson Reuters, SCIELO Colombia, LATINDEX, ACTUALIDAD IBEROAMERICANA, GEOREF, Red ALyC, CHEMICAL ABSTRACTS, DOAJ (Directory of Open Access Journal). *Revista de la Facultad de Minas de la Universidad Nacional de Colombia*. Science Citation Index Expanded (SciSearch), Journal Citation Reports/Science Edition. Octubre, 2011.
- [8] Enislay Ramentol, Nele Verbiest, Rafael Bello, Yailé Caballero, Chris Cornelis, Francisco Herrera. SMOTE-FRST: A new resampling method using Fuzzy Rough Set Theory. *Uncertainty Modeling in Knowledge Engineering and Decision Making*. Pp: 800-805. Indizada por: Science Citation Index, SCOPUS. Agosto, 2012.
- [9] Nele Verbiest, Enislay Ramentol, Chris Cornelis, Francisco Herrera. Improving SMOTE with Fuzzy Rough Prototype Selection to detect Noise in Imbalanced Classification data. Publicado en la serie *Advances in Artificial Intelligence of Springer-Verlag*. LNAI. Volume 7637, 2012, Pp: 169-178. Indizada por: Science Citation Index, SCOPUS. Noviembre, 2012.
- [10] Filiberto Y., Bello, R., Caballero Y., Larrua R. A method to build similarity relations into extended Rough Set Theory. *IEEE Catalog Number CFP 10384*

CDR, ISBN 978-1-4244-8135-4. DOI: 10.1109/ISDA.2010.6587091. Pp: 1314-1319. Indizada por: SCOPUS. 2010.

[11] Caballero, Y., Bello, R., Ramentol, E., Arco, L., García, M. Knowledge Discovery Using Rough Set Theory. Advances in Machine Learning I. Series: Studies in Computational Intelligence, Vol. 262, Pp. 367-383. Koronacki, J.; Ras, Z.W.; Wierzchon, S.T.; Kacprzyk, J. (Eds.) ISSN: 1860-949X. ISBN: 978-3-642-05176-0. Springer. Indizada por: Science Citation Index, SCOPUS. 2010

[12] Filiberto Y., Bello R., Caballero Y., Larrua R. Using PSO and RST to predict the resistant capacity of connections in composite structures. Studies in Computational Intelligence. Springer Berlin / Heidelberg. ISSN: 1860-949X (Print) 1860-9503 (Online). Volume 284/2010. ISBN: 978-3-642-12537-9. Pp 359-370. Indizada por: Science Citation Index, SCOPUS. 2010

[13] Caballero, Y. Bello, R., Arco, L., Márquez, Y., García, M. La Teoría de los Conjuntos Aproximados para el Descubrimiento de Conocimiento. Revista DYNA, Vol. 77, Edición 162, Pp 261-270. ISSN 0012-7353. Indizada por: Thomson Reuters, SCIELO Colombia, LATINDEX, ACTUALIDAD IBEROAMERICANA, GEOREF, Red ALyC, CHEMICAL ABSTRACTS, DOAJ (Directory of Open Access Journal). Revista de la Facultad de Minas de la Universidad Nacional de Colombia. Indizada por: Science Citation Index Expanded (SciSearch), Journal Citation Reports/Science Edition. Junio 2010

[14] Larrua R., Caballero Y., Filiberto Y., otros. Aplicación de la inteligencia artificial a la predicción de la capacidad resistente última de las conexiones en estructuras compuestas acero-hormigón. Revista de la Construcción. (Pontificia Universidad Católica de Chile). ISSN 0717-7925 versión impresa. Indizada por: Science Citation Index Expanded, PERIODICA, Chemical Abstracts, Actualidad Iberoamericana, RedALyC. Se encuentra en la Biblioteca Científica SCIELO Chile on-line. Vol 8 Nro 2. Pp 109-119. diciembre 2009.

#### **Otras publicaciones reconocidas y memorias de eventos prestigiosos:**

[15] Y. Caballero y colaboradores. Efecto de la selección de rasgos en el algoritmo para la generación de reglas de clasificación IRBASIR. RECPAT 2012. Camagüey. Noviembre, 2012. ISBN: 978-959-16-2065-1.

[16] Y. Caballero y colaboradores. Estudio de la selección de atributos relevantes aplicando algoritmos que combinan conjuntos aproximados e inteligencia colectiva. RECPAT 2012. Camagüey. Noviembre, 2012. ISBN: 978-959-16-2065-1.

[17] Y. Caballero y colaboradores. Uso de colonias de hormigas artificiales para el mejoramiento de clasificadores del vecino más cercano. RECPAT 2012. Camagüey. Noviembre, 2012. ISBN: 978-959-16-2065-1.

[18] Y. Caballero y colaboradores. Modelo inteligente para la predicción de variables meteorológicas a mediano y largo plazo. RECPAT 2012. Camagüey. Noviembre, 2012. ISBN: 978-959-16-2065-1.

[19] Filiberto Y., Caballero Y., Bello R., Frías M. Algoritmo de generación de reglas. Revista Cubana de Ciencias Informáticas. Vol. 5, Nro. 4. 2011.

[20] K. Marrero, E. Ramentol, Y. Caballero, R. Bello. Nuevo método de edición para conjuntos no balanceados usando operadores genéticos y k-NN. IX Congreso

de Reconocimiento de Patrones RECPAT2011. ISBN: 978-959-250-658-9. Noviembre 2011.

[21] Y. Filiberto, R. Bello, Y. Caballero, M. Frías. Algoritmo para el aprendizaje de reglas de clasificación basado en la Teoría de los Conjuntos Aproximados Extendida. IX Congreso de Reconocimiento de Patrones RECPAT2011. ISBN: 978-959-250-658-9. Noviembre 2011.

[22] Filiberto Y., Bello R., Caballero Y., Frias M. Método para el aprendizaje de reglas de decisión basada en la teoría de de los conjuntos aproximados. ISBN: 978-959-7213-01-7. Proceedings Conferencia Internacional de Ciencias Computacionales. XIV Convención y Feria Internacional Informática2011. Febrero, 2011.

[23] Larrua R., Olivera I., Filiberto Y., Bonilla J. Aplicación de la Inteligencia Artificial a la predicción de la capacidad resistente última de las conexiones en estructuras compuestas acero-hormigón. IV Jornada Internacional de la Ingeniería Civil en Cuba. Varadero, Cuba. Junio 2010.

[24] Filiberto Y., Bello R., Caballero Y., Larrua, R. A new measures in the Rough Set Theory for continuos decision systems. Memorias de Cuba-Flanders Workshop on Machine Learning and Knowledge Discovery CF-WML-KD2010. Santa Clara, Cuba. February 3 - 5, 2010.

[25] Ramentol E., Caballero Y., Bello, R., Moreno A., Herrera, F. SMOTE-RSB: A Hybrid Preprocessing Approach based on Oversampling an Undersampling por High Imbalanced Data-Sets using SMOTE and Rough Sets Theory. Memorias de Cuba-Flanders Workshop on Machine Learning and Knowledge Discovery CF-WML-KD2010. Santa Clara, Cuba. February 3 - 5, 2010.

[26] Filiberto Y., Bello R., Caballero Y., Larrua, R. Improving the k-NN method for the function approximation problem using a Rough Set. 9th International Conference on Operations Research. Colegio San Jerónimo de La Habana, Cuba. Feb 22-26 2010.

[27] Ramentol E., Caballero Y., Bello R., Herrera F., Sánchez Y. Nuevo algoritmo de edición de conjuntos de entrenamiento no balanceados hibridando operadores genéticos y la Teoría de los Conjuntos Aproximados. VII Congreso Nacional de Reconocimiento de Patrones, RECPAT2009. Santiago de Cuba, diciembre 2009. ISBN: 978-959-207-381-4.

[28] Filiberto Y., Bello R., Caballero Y. 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. VII Congreso Nacional de Reconocimiento de Patrones, RECPAT2009. Santiago de Cuba, diciembre 2009. ISBN: 978-959-207-381-4.

**AUTORIA DE SOFTWARE** Los modelos desarrollados han sido implementados en los siguientes productos de software. **En el Anexo II se muestran las respectivas certificaciones.**

1. Registro de Software número 2184-2012 del Centro Nacional de Derecho de Autor a favor de: PROCON v3.0. Pronóstico de la capacidad resistente de conectores en estructuras compuestas de hormigón-acero. 2012.

2. Registro de Software número 2191-2012 del Centro Nacional de Derecho de Autor a favor de: SAICCA Selección de Atributos con Inteligencia Colectiva y Conjuntos Aproximados. 2012.
3. Registro de Software número 2185-2012 del Centro Nacional de Derecho de Autor a favor de: R.R-Generator. Facilita un uso adecuado y eficiente del Sistema Reduct and Rule Generator para la selección de los atributos relevantes de una base de casos y su posterior. 2012.
4. Registro de Software número 2301-2011 del Centro Nacional de Derecho de Autor a favor de: PROCON v2.0. Cálculo del pronóstico de la capacidad resistente de conexiones en estructuras compuestas de hormigón y acero, con conectores diversos. 2011
5. Registro de Software número 1271-2011 del Centro Nacional de Derecho de Autor a favor de: PROCON v1.0. Cálculo del pronóstico de la capacidad resistente de conexiones en estructuras compuestas de hormigón y acero, con conectores diversos. 2011
6. Registro de Software número 2064-2010 del Centro Nacional de Derecho de Autor a favor de: "PROVARMET 1.0" Sistema automatizado para el pronóstico de variables meteorológicas de Camagüey y Las Tunas. 2010

## **TESIS DE DOCTORADO EN CIENCIAS TECNICAS, MAESTRIAS Y TRABAJOS DE DIPLOMA DEFENDIDAS CON ÉXITO A PARTIR DE LOS RESULTADOS CIENTIFICOS AQUÍ EXPUESTOS**

### **Tesis de Doctorado:**

[T1] Métodos de aprendizaje para dominios con datos mezclados basados en la Teoría de los Conjuntos Aproximados Extendida. Autora: Yaima Filiberto Cabrera, de la Universidad de Camagüey, Cuba. Enero, 2012. (Tesis premiada a nivel internacional en el Consorcio Internacional de Estudiantes de Doctorado, México). (Premio a la Mejor Tesis de Ciencias Técnicas defendida en el año 2012). (Premio al Doctor más joven defendido en el año 2012).

### **Tesis de Maestrías:**

[T2] Edición de conjuntos de entrenamiento no balanceados hibridando operadores genéticos y la Teoría de los Conjuntos Aproximados. 2010. DEA Softcomputing. Maestrante: Enislay Ramentol Martínez (Universidad de Camagüey-Universidad de Granada).

[T3] Predicción de variables meteorológicas a corto, mediano y largo plazos a través de técnicas de Inteligencia Artificial. 2011. Maestrante: Adán Montiel Valdés. (Maestría en Informática Aplicada. Universidad de Camagüey).

[T4] Pronóstico de la capacidad resistente de conexiones en estructuras compuestas de hormigón y acero, utilizando técnicas de Inteligencia Artificial. 2011. Maestrante: Yulién Martínez Conde. (Maestría en Informática Aplicada. Universidad de Camagüey).

[T5] Implementación de nuevo algoritmo de oversampling hibridando SMOTE y la Teoría de los Conjuntos Aproximados. 2011. Maestrante: Adrián Moreno (Maestría en Informática Aplicada. Universidad de Camagüey).

[T6] Edición de conjuntos de entrenamiento no balanceados, haciendo uso de operadores genéticos y la teoría de los conjuntos aproximados. 2011. Maestrante:

- Yanet Sánchez (Maestría en Informática Aplicada. Universidad de Camagüey).
- [T7] Pronóstico de la demanda en destinos turísticos con el uso de Redes Neuronales Artificiales; el caso de Santa Lucía de Cuba. 2011. Maestrante: Tomás Molinet Berenguer (Maestría en Gestión Turística. Universidad de Camagüey).
- [T8] Desarrollo de modelos inteligentes para la predicción de variables meteorológicas a mediano y largo plazos. 2011. Maestrante: Lisset Cabrera Leyva. (Maestría en Informática Aplicada. Universidad de Camagüey).
- [T9] Algoritmo de clasificación para conjuntos no balanceados. 2012. Maestrante: Karel Marrero Estévez. (Maestría en Informática Aplicada. Universidad de Camagüey).
- [T10] Sistema de apoyo a la edición de conjuntos desbalanceados. 2012. Maestrante: Mairelys Ramírez. (Maestría en Informática Aplicada. Universidad de Camagüey).

### **Trabajos de Diploma**

- [T11] Pronóstico de la capacidad resistente de conexiones en estructuras compuestas de hormigón y acero, utilizando técnicas de aprendizaje automatizado. 2008-2009. Mayte Guerra. Trabajo de Diploma en opción al título de Ingeniero Informático. Universidad de Camagüey.
- [T12] Aplicación de la inteligencia artificial al pronóstico de la capacidad resistente de conexiones en estructuras compuestas hormigón – acero. 2008-2009. Iliana Olivera. Trabajo de Diploma en opción al título de Ingeniero Civil. Universidad de Camagüey.
- [T13] Generación aleatoria de instancias para el pronóstico de conectores. 2009-2010. Iliana Varona Horta. Trabajo de Diploma en opción al título de Ingeniero Informático. Universidad de Camagüey.
- [T14] Edición de conjuntos de entrenamiento no balanceados haciendo uso de técnicas evolutivas y de recombinación de métodos de limpieza. 2009-2010. Karel Marrero. Trabajo de Diploma en opción al título de Ingeniero Informático. Universidad de Camagüey.
- [T15] Adaptación de algoritmos de selección de rasgos en sistemas de decisión continuos a la herramienta Weka. 2009-2010. Yakelín Cardoso. Trabajo de Diploma en opción al título de Ingeniero Informático. Universidad de Camagüey.
- [T16] Pronóstico de la capacidad resistente de conexiones en estructuras compuestas de hormigón y acero, utilizando las técnicas evolutivas de los algoritmos genéticos. 2009-2010. Dayexi Libera. Trabajo de Diploma en opción al título de Ingeniero Informático. Universidad de Camagüey.
- [T17] Algoritmo de generación de reglas basado en la Teoría de los Conjuntos Aproximados Extendida. 2010-2011. Mabel Frías Domínguez. Trabajo de Diploma en opción al título de Ingeniero Informático. Universidad de Camagüey.
- [T18] Aplicación de la inteligencia artificial al pronóstico de la respuesta de muros de mampostería confinada. 2010-2011. Lorena Chio. Trabajo de Diploma en opción al título de Ingeniero Civil. Universidad de Camagüey.
- [T19] Selección de atributos relevantes aplicando algoritmos que combinan conjuntos aproximados e inteligencia colectiva. 2011-2012. Yanela Rodríguez

Álvarez. Trabajo de Diploma en opción al título de Ingeniero Informático. Universidad de Camagüey.

[T20] Generación de reglas de clasificación utilizando selección de atributos. 2011-2012. Marcos Enrique Rivero Macías. Trabajo de Diploma en opción al título de Ingeniero Informático. Universidad de Camagüey.

[T21] Modificación del Algoritmo IRBASIR para la Generación de Reglas de Clasificación basado en la Teoría de los Conjuntos Aproximados Extendida. 2011-2012. Mario Nicolau Matsyka. Trabajo de Diploma en opción al título de Ingeniero Informático. Universidad de Camagüey.

[T22] Efecto del empleo de los pesos de los rasgos en el método MLP. 2011-2012. Marité Suárez Oliveros. Trabajo de Diploma en opción al título de Ingeniero Informático. Universidad de Camagüey.

[T23] Predicción de la capacidad resistente de conexiones tipo canal en estructuras compuestas hormigón-acero. 2011-2012. J. Espino. Trabajo de Diploma en opción al título de Ingeniero Civil. Universidad de Camagüey.

[T24] Predicción del comportamiento de muros de mampostería con la aplicación de inteligencia artificial. 2011-2012. I. Albojaire, M. Magariño. Trabajo de Diploma en opción al título de Ingeniero Civil. Universidad de Camagüey.

### **PREMIOS Y RECONOCIMIENTOS RECIBIDOS En el Anexo III aparecen las respectivas certificaciones.**

1. **Premio al trabajo tutorado:** PROCON v4.0 Pronóstico de la capacidad resistente en Construcciones. Fórum Científico Nacional de Estudiantes Universitarios de Ciencias Técnicas 2013.

2. **Premio a la Mejor Tesis Doctoral** defendida en el año 2012 en Ciencias Técnicas. Título: Métodos de aprendizaje para dominios con datos mezclados basados en la Teoría de los Conjuntos Aproximados Extendida.

3. **Premio a Doctor más joven** defendido en el año 2012.

4. **Premio del Rector al Mérito Científico Técnico 2012**, en la categoría: al Colectivo de investigación más destacado en el trabajo de investigación y en la promoción de procesos innovativos en el 2012.

5. **Premio al Mérito Científico Técnico 2012**, en la categoría: al resultado que refleje el avance científico de mayor trascendencia y originalidad durante el año 2012.

6. **Premio Nacional Colectivo Forjadores del Futuro** otorgado por las BTJ al Colectivo de Investigación de Inteligencia Artificial de la Universidad de Camagüey por los aportes significativos al desarrollo científico durante el 2011.

7. **Premio del Rector al Mérito Científico Técnico 2011**, en la categoría: al Colectivo de investigación más destacado en el trabajo de investigación y en la promoción de procesos innovativos en el 2011.

8. **Premio al Mérito Científico Técnico** en la categoría: al resultado que refleje el avance científico de mayor trascendencia y originalidad durante el año 2011.

9. **Premio al Mérito Científico Técnico 2011**, en la categoría: Al resultado ya aplicado de mayor aporte al desarrollo social.

10. **Premio Internacional a Tesis de Doctorado** Improving the MLP learning by using a method to calculate the initial weights of the network based on the quality of similarity measure. Publa, México en el consorcio de Estudiantes de Doctorado en el marco del X Mexican International Conference on Artificial Intelligence (MICAI 2011). Métodos de aprendizaje para dominios con datos mezclados basados en la Teoría de los Conjuntos Aproximados Extendida. Autora: Yaima Filiberto Cabrera, de la Universidad de Camagüey, Cuba. Tutores: Dr. Rafael Bello y Dra. Yailé Caballero.

11. **Premio Destacado al trabajo tutorado:** Pronóstico de la capacidad resistente en estructuras compuestas. XIX Fórum Científico Nacional de Estudiantes Universitarios de Ciencias Técnicas 2011.

12. **Premio CITMA provincial:** La predicción de la respuesta de estructuras diversas con la hibridación de técnicas inteligentes y modelos bioinspirados. 2011

13. **Premio CITMA provincial:** La predicción de la capacidad resistente de las conexiones en estructuras compuestas acero-hormigón con aplicación de Inteligencia Artificial. 2010

14. **Premio del Rector al Mérito Científico Técnico 2009**, en la categoría: al Colectivo de investigación más destacado en el trabajo de investigación y en la promoción de procesos innovativos en el 2009.