Enislay Ramentol Martinez

◦ Premio Extraordinario de Doctorado 2013-14 (Intenierías y Arquitectura)

NUEVOS METODOS DE EDICION DE CONJUNTOS DE ENTRENAMIENTO NO BALANCEADOS USANDO LA TEORIA DE LOS CONJUNTOS APROXIMADOS

Resumen

La creación de herramientas informáticas que “aprendan” a partir de la “experiencia” y sean capaces de diagnosticar, recomendar, predecir un hecho del que no existe precedente; desde hace algunos años atrapa la atención de la comunidad científica del área de la computación. La necesidad de disponer de algoritmos que permitan la generación de modelos a partir de los datos disponibles, ha propiciado el desarrollo vertiginoso de la minería de datos, y dentro de ésta del aprendizaje automático. El aprendizaje a partir de datos no balanceados se sitúa como uno de los 10 grandes desafíos que enfrenta el aprendizaje automático, debido a que cada vez son más las aplicaciones donde se presenta esta problemática.

Uno de los enfoques usados en el desarrollo de técnicas para del descubrimiento de conocimiento, es el basado en la Teoría de los Conjuntos Aproximados (TCA). La Teoría TCA se ha convertido en una herramienta relevante en el análisis inteligente de datos, la misma se utiliza tanto para el desarrollo de algoritmos de preprocesamiento como para el descubrimiento de conocimiento. La combinación de los conjuntos aproximados y los borrosos (fuzzy rough set y rough fuzzy sets) ha significado un nuevo paso en esta dirección. La tesis “Nuevos Métodos de Edición de Conjuntos de Entrenamiento no Balanceados usando la Teoría de los Conjuntos Aproximados” ha permitido la creación de algoritmos (utilizando los conceptos de la TCA y su hibridación difusa) que logran mejorar el desempeño de los clasificadores de manera significativa cuando éstos se enfrentan a datos donde la representatividad de los conceptos a clasificar está desbalanceada.

Algunas de las aportaciones más importantes derivadas de esta Tesis doctoral

Los principales resultados de la tesis están divididos en 3 grupos:

1. Algoritmos de preprocesamiento. Se proponen 3 nuevos algoritmos de preprocesamiento: SMOTE-RSB*, SMOTE-FRST y SMOTE-FRST-2T.
2. Algoritmo de clasificación. Se propone el algoritmo IFROWANN, un robusto clasificador que no necesita preprocesamiento, el mismo está avalado por una extensa experimentación realizada sobre 102 bases de datos internacionales.
3. Aplicación a un problema de la ingeniería. Se aplica el algoritmo SMOTE-FRST-2T a la predicción del diagnóstico de la necesidad de mantenimiento de los interruptores de alta potencia en la provincia de Camagüey, Cuba. Se desarrolla un software que a partir de la captura de datos de los interruptores en funcionamiento, emite señales de alarmas cuando un interruptor necesita mantenimiento.

Estos resultados son el producto de la colaboración entre investigadores de la Universidad de Granada (España), las Universidades cubanas de Camagüey y Central de Las Villas (Cuba) y la Universidad de Ghent, Bélgica. Los principales resultados a nivel teóricos están avalados por las publicaciones logradas en revistas como “Knowledge and Information Systems”, “Engineering Applications of Artificial Intelligence” y “IEEE Transaction on Fuzzy Systems”, esta última rankeada como número 1 en el área de Inteligencia Artificial.

Fuente: https://escuelaposgrado.ugr.es/doctorado/escuelas/edcti/pages/premios_extraordinarios/2013_14/ramentol_martinez_enislay/index