

Título: ALGORITMOS PARA LA CLASIFICACIÓN MULTINSTANCIA

Nombre: Sánchez Tarragó, Dánel

Universidad: Universidad de Granada

Departamento: CIENCIAS DE LA COMPUTACION E INTELIGENCIA ARITIFICIAL

Fecha de lectura: 25/06/2014

Programa de doctorado: Programa Oficial de Doctorado en Tecnologías de la Información y la Comunicación

Dirección:

- > **Director:** FRANCISCO HERRERA TRIGUERO
- > **Codirector:** Rafael Esteban Bello Pérez
- > **Codirector:** Chris Cornelis

Tribunal:

- > **presidente:** Antonio González Muñoz
- > **secretario:** JORGE CASILLAS BARRANQUERO
- > **vocal:** SEBASTIAN EMILIO VENTURA SOTO
- > **vocal:** Alberto Fernández Hilario
- > **vocal:** CÉSAR HERVÁS MARTÍNEZ

Descriptor:

- > INTELIGENCIA ARTIFICIAL

El fichero de tesis ya ha sido incorporado al sistema

- > <http://0-hera.ugr.es.adrastea.ugr.es/tesisugr/24077677.pdf>

Localización: DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN E INTELIGENCIA ARTIFICIAL

Resumen: Dentro del campo de estudio del aprendizaje automático, la clasificación multinstancias tiene como objetivo construir, a partir de un conjunto de ejemplos, un modelo matemático que permita clasificar objetos descritos por múltiples vectores de atributos. En la tesis se realiza una profunda revisión bibliográfica que contribuye a organizar la literatura dispersa y sistematizar el conocimiento existente sobre clasificación multinstancias, y se proponen dos soluciones para sendos problemas identificados en el campo de la clasificación multinstancias.

Uno de los problemas identificados es el problema de las clases desbalanceadas, el cual ocurre cuando hay mucha diferencia en el tamaño de las clases y provoca que el modelo de aprendizaje inducido a partir de los datos no represente adecuadamente el concepto que se pretende aprender y consecuentemente incurra en muchos errores de clasificación. En la tesis se propone un nuevo método general basado en técnicas de muestreo para enfrentar este problema en la clasificación multinstancias. El algoritmo propuesto, bautizado como MISMOTE, consiste a groso modo en modificar la distribución de los datos de entrenamiento, insertando

ejemplos sintéticos de objetos multinstancias en la clase minoritaria, para equilibrar el balance de las clases. Las pruebas experimentales validadas por métodos estadísticos mostraron que la solución propuesta mejora significativamente la calidad de la clasificación y es competitiva con otras soluciones existentes en la literatura.

El otro problema identificado es que aún no han sido adaptados al enfoque del aprendizaje multinstancias algunos algoritmos de clasificación tradicionales orientados específicamente a determinados dominios de aplicación. En la tesis se propone un nuevo algoritmo de clasificación multinstancias que permite aplicar el clasificador de Rocchio en el escenario multinstancias. El clasificador de Rocchio es muy usado para la recuperación y clasificación de documentos textuales en aplicaciones on-line. El algoritmo propuesto, bautizado como MIRocchio, implementa una nueva hipótesis multinstancias introducida en la tesis, la cual está basada en proporción de instancias positivas. La validez del algoritmo se comprueba experimentalmente en el problema de la recomendación de páginas web índices, siendo capaz de mejorar el desempeño de los clasificadores multinstancias del estado del arte en este dominio de aplicación.