

Título: MODELOS DE APRENDIZAJE BASADOS EN PROGRAMACIÓN GENÉTICA PARA CLASIFICACIÓN MULTI-ETIQUETA

Nombre: AVILA JIMENEZ, JOSE LUIS

Universidad: Universidad de Córdoba

Departamento: Informática y análisis numérico

Fecha de lectura: 13/06/2013

Programa de doctorado: INGENIERÍA Y TECNOLOGÍA

Dirección:

> **Director:** SEBASTIAN EMILIO VENTURA SOTO

> **Director:** EVA LUCRECIA GIBAJA GALINDO

Tribunal:

> **presidente:** José Ángel Olivas Varela

> **secretario:** AMELIA ZAFRA GÓMEZ

> **vocal:** MARÍA JOSÉ DEL JESÚS DÍAZ

Descriptor:

> PANIFICACION

El fichero de tesis ya ha sido incorporado al sistema

> <http://hdl.handle.net/10396/10427>

Resumen: RESUMEN DE LA TESIS DOCTORAL DE D./D^a

El resumen de la tesis para la base de datos Teseo debe ser una presentación de la tesis y tener la extensión suficiente para que quede explicado el argumento de la tesis doctoral. El formato debe facilitar la lectura y comprensión del texto a los usuarios que accedan a Teseo, debiendo diferenciarse las siguientes partes de la tesis:

1. introducción o motivación de la tesis

El problema de clasificación consiste en asociar una serie de etiquetas a una serie de ejemplos o patrones. En la clasificación clásica a cada patrón de entrenamiento solamente se le puede asociar una sola etiqueta de un conjunto de etiquetas. Por tanto se consideran que los conjuntos de clases objetivo en los que se agruparán los patrones son por definición conjuntos disjuntos.

En el caso de la clasificación multi-etiqueta, los conjuntos objetivos no son disjuntos, pudiendo haber patrones a los que se les asocie más de una etiqueta. Por tanto, los ejemplos se asocian a un conjunto de etiquetas y el resultado puede tomar varios valores dentro del conjunto de etiquetas[1].

El objetivo que se plantea en esta tesis es el desarrollo de una serie de modelos de programación genética para resolver problemas de clasificación multi-etiqueta.

2. contenido de la investigación

En la presente tesis se han propuesto dos modelos de programación genética para solventar problemas de clasificación multi-etiqueta.

Una primera aproximación al problema ha consistido en el uso de funciones discriminantes para determinar si un patrón pertenece o no a cada una de las etiquetas del problema, y se ha desarrollado en el algoritmo GEP-MLC [2].

Una segunda aproximación más elaborada en el aspecto de la comprensibilidad, así como en el manejo de conjuntos de datos categóricos es utilizar reglas de clasificación. Dentro de las reglas de clasificación se han desarrollado dos modelos, el denominado GC-1 en el que las reglas solamente tienen asociada una etiqueta en el consecuente, y el modelo GC-k, que comparte rasgos similares con el anterior pero en el que las reglas pueden estar asociadas a más de una etiqueta en su consecuente[3].

3. conclusión

Todos los algoritmos han sido evaluados en un marco experimental adecuado, utilizando numerosos conjuntos de datos y comparando su rendimiento frente a otros métodos ya publicados de contrastada calidad. Los resultados obtenidos, que han sido verificados mediante la aplicación de test estadísticos no paramétricos [4] demuestran los beneficios de utilizar los algoritmos propuestos.

4. bibliografía

- [1] G. Tsoumakas, I. Katakis, and I. Vlahavas, *Data Mining and Knowledge Discovery Handbook*, ch. Mining Multi-label Data(2009). Springer.
- [2] Ávila-Jiménez, J. L., Gibaja, E. L., Zafra, A. and Ventura, S. (2011). A niching algorithm to learn discriminant functions with multi-label patterns. *Journal of multiple-valued logic and soft computing*, 1-10.
- [3] Ávila-Jiménez, J. L., Gibaja, E. L. and Ventura, S (2010). Evolving Multi-label Classification Rules with Gene Expression Programming: A Preliminary Study. En *Proceedings of th 5th International Conference on Hybrid Artificial Intelligence Systems (HAIS'10)*, páginas 9-16.
- [4] Demsar, J. (2006). Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research* , 7, 1-30.