

## Una Metodología de KDD basada en Sistemas Difusos Genéticos Multiobjetivo para Modelado Causal en Marketing\*

Jorge Casillas, Óscar Delgado

Dpto. Ciencias de la Computación e I.A.  
Univ. de Granada  
18071 Granada  
casillas@decsai.ugr.es

Francisco J. Martínez-López

Dpto. Comercialización e Invest. Mercados  
Univ. de Granada  
18071 Granada  
fjmlopez@ugr.es

### Resumen

El trabajo presenta un problema novedoso basado en modelado causal en marketing y su resolución mediante descubrimiento de conocimiento, tal como se muestra con una aplicación real. Las características del problema (con datos inciertos y conocimiento experto disponible) y la optimización multiobjetivo propuesta hace a los sistemas difusos genéticos una buena herramienta para abordarlo.

### 1. Introducción

El trabajo presenta una metodología de *extracción de conocimiento en bases de datos* (KDD) a partir de modelos causales utilizados tradicionalmente en marketing. Se propone un método de *inducción descriptiva* [4] para descubrir reglas individuales que determinen patrones de especial interés en los datos. Para ello se consideran reglas difusas de asociación [2], pero con variables del antecedente y consecuente previamente fijadas. La extracción se realiza mediante sistemas difusos genéticos. Como es natural, llegado a este punto surgen dos cuestiones importantes: ¿por qué reglas difusas? y ¿por qué algoritmos genéticos (AGs)? Es decir, por qué usar estas herramientas de representación y aprendizaje frente a otras ampliamente usadas en KDD.

El uso de reglas difusas (en lugar de reglas intervalares, árboles de decisión, etc.) viene

justificado principalmente por el tipo de dato manejado (ver sección 3.1). En nuestro caso, cada variable se compone de un conjunto de parámetros (ítems) que aportan información parcial para describirla. Esto añade incertidumbre en los datos que puede tratarse adecuadamente con reglas difusas. Además, es posible expresar el conocimiento experto disponible en semánticas lingüísticas. Por último, las reglas difusas obtenidas presentan una gran legibilidad y ofrecen transiciones graduales de las relaciones, aspectos importantes en KDD.

Respecto al uso de AGs para inducir estas reglas difusas en lugar de otras técnicas de aprendizaje automático, se debe a los siguientes aspectos. Por un lado, dado que la calidad de las diferentes reglas difusas se valoran mediante objetivos contradictorios (tales como soporte y confianza), realizamos optimización multiobjetivo para tratarlos adecuadamente. Este es actualmente uno de los aspectos más prometedores y una de las señas de identidad de los AGs, donde se están destacando por su alto rendimiento frente a otras técnicas. Además, para conseguir mayor compacidad y, por tanto, interpretabilidad, consideramos una representación flexible de reglas difusas que puede manejarse cómodamente con AGs.

El trabajo se organiza de la siguiente forma. La sección 2 describe brevemente el problema abordado. La sección 3 introduce la metodología de KDD propuesta. La sección 4 muestra algunos resultados experimentales. Por último, la sección 5 presenta algunas conclusiones y trabajos futuros.

\* Trabajo soportado en parte por el Proyecto TIC2002-04036-C05-01 del MCyT y fondos FEDER