

QÜESTIÓ, vol. 25, 3, p. 479-498, 2001

La minería de datos, entre la estadística y la inteligencia artificial

Data mining, between statistics and artificial intelligence

Tomàs Aluja

Departamento de Estadística e Investigación Operativa. Universitat Politècnica de Catalunya (UPC). E-mail: tomas.aluja@upc.es

En la pasada década hemos asistido a la irrupción de un nuevo concepto en el mundo empresarial: el «data mining» (minería de datos). Algunas empresas han implementado unidades de minería de datos estrechamente vinculados a la dirección de la empresa y en los foros empresariales las sesiones dedicadas a la minería de datos han sido las protagonistas. La minería de datos se presenta como una disciplina nueva, ligada a la Inteligencia Artificial y diferenciada de la Estadística. Por otro lado, en el mundo estadístico más académico, la minería de datos ha sido considerada en su inicio como una moda más, aparecida después de los sistemas expertos, conocida desde hacía tiempo bajo el nombre de «data fishing».

¿Es esto realmente así? En este artículo abordaremos las raíces estadísticas de la minería de datos, los problemas que trata, haremos una panorámica sobre el alcance actual de la minería de datos, presentaremos un ejemplo de su aplicación en el mundo de la audiencia de televisión y, por último, daremos una visión de futuro.

In the last decade a new concept had raised in the entrepreneurial side: data mining. Some companies have created data mining units directly linked to the CRM direction and in the professional forums data mining sessions have gained appeal. Data mining has appeared as a new discipline linked to Machine Learning, Artificial Intelligence and Data Bases, clearly differentiated from Statistics. On the other side, on the well-established statistics academia, data mining has been seen as the last fashion of a bad-known trend of data fishing and data dredging.

Is it really so? In this paper we will focus on the statistical roots of data mining, we will try to make an overview of the actual scope of data mining, we will present an application in the TV audience measure and we give some insights for the next future.

Palabras clave: Data mining, análisis de datos, modelización, inteligencia artificial, KDD, redes neuronales, árboles de decisión

Keywords: Data mining, data analysis, modelling, artificial intelligence, KDD, neural networks, decision trees

Clasificación AMS (MSC 2000): 62-07, 68T10, 62P30
