
Apéndice D

Acrónimos

ACLS [Analogue Concept Learning System]

Algoritmo de construcción de árboles de decisión derivado de ID3.

ARM [Association Rule Mining]

Uno de los problemas de KDD que más atención ha recibido en los últimos años. Generalmente se aplica a bases de datos de transacciones comerciales (p.ej. grandes almacenes) para caracterizar el comportamiento del consumidor.

ART [Association Rule Tree]

Modelo de clasificación híbrido, propuesto por Juan Carlos Cubero, que pretende aprovechar las mejores cualidades de las reglas de asociación (como modelo de clasificación parcial) y de los algoritmos de construcción de árboles de decisión.

BDE [Borland Database Engine]

CLI de Borland.

CART [Classification and Regression Trees]

Algoritmo de construcción de árboles de decisión.

CLI [Call-Level Interface]

Protocolo estándar de acceso a una base de datos desde un lenguaje de alto nivel.

CLS [Concept Learning System]

El patriarca de la familia de algoritmos TDIDT.

DDL [Data Definition Language]

Las órdenes de un lenguaje de definición de datos para una base de datos relacional incluyen, entre otras, aquéllas que nos permiten manipular tablas y vistas.

DM [Data Mining]

Ver KDD.

DML [Data Manipulation Language]

Las órdenes de un lenguaje de manipulación de datos para una base de datos relacional permiten manipular las tuplas contenidas en las tablas: insertar, borrar, actualizar o consultar.

ID3 [Iterative Dichotomizer 3]

Ideado por el investigador australiano J.R. Quinlan, es posiblemente el algoritmo más conocido de la familia TDIDT para la construcción de árboles de decisión.

ISODATA [Iterative Self-Organizing Data Analysis Techniques]

Algoritmo de clustering similar al algoritmo K-Means (su objetivo es optimizar la suma de los cuadrados de las distancias a los centroides de los clusters) aunque bastante más complejo.

JDBC [Java DataBase Connectivity]

CLI estándar para el lenguaje de programación Java.

KDD [Knowledge Discovery in Databases]

Extracción no trivial de información potencialmente útil a partir de un gran volumen de datos en el cual la información está implícita.

KDDMS [Knowledge and Data Discovery Management Systems]

Sistema de Data Mining de “segunda generación”. Su objetivo es ser a las aplicaciones de KDD lo que los DBMSs son hoy a las aplicaciones de gestión. Debe permitir consultas mucho más flexibles y complejas que las permitidas por SQL.

KM [K-Means]

Algoritmo de clustering que, dado el número de clusters deseado, trata de minimizar la suma de los cuadrados de las distancias a los centroides de los clusters.

MAQA [Mining Association among Quantitative Attributes]

Algoritmo de extracción de reglas de asociación que permite el manejo de atributos numéricos mediante su discretización en intervalos.

MDL [Minimum Description Length principle]

Principio establecido por Rissanen según el cual el mejor modelo para representar un conjunto de datos es el que minimiza la suma del número de bits requerido para describir los datos en términos del modelo y del número de bits necesario para describir el modelo en sí.

ML [Machine Learning]

Proceso por el cual un ordenador acrecienta su conocimiento y mejora su habilidad: “el aprendizaje denota cambios en el sistema que son adaptativos en el sentido de que permiten al sistema hacer la misma tarea o tareas a partir de la misma posición más eficiente y/o efectivamente la siguiente vez” [Simon].

ODBC [Open DataBase Connectivity]

CLI de Microsoft.

OLAP [OnLine Analytical Processing]

Herramientas usadas para el análisis interactivo de datos multidimensionales, mejores que SQL a la hora de realizar resúmenes e informes.

SLIQ [Supervised Learning in Quest]

Algoritmo de construcción de clasificadores basados en árboles de decisión desarrollado en el proyecto Quest (el proyecto de Data Mining del IBM Almaden Research Center).

SPRINT [Scalable PaRallelizable Induction of decision Trees]

Otro algoritmo de construcción de árboles de decisión ideado por los miembros del proyecto Quest del IBM Almaden Research Center.

SQL [Structured Query Language]

Lenguaje estándar de definición y manipulación de datos (DDL y DML) cuya primera versión comercial fue introducida en el mercado por Relational Software Inc. (hoy Oracle Corporation) en 1979. SQL se deriva del lenguaje SEQUEL (“*Structured English Query Language*”), desarrollado por IBM Corporation haciendo uso del modelo relacional de Codd. El último estándar de SQL publicado por ANSI e ISO se denomina SQL-92 (a veces, SQL2): ANSI X3.135–1992 e ISO/IEC 9075:1992.

TDIDT [Top-Down Induction of Decision Trees]

Familia de algoritmos para la construcción recursiva de árboles de decisión. Miembros destacados de esta familia son ID3, CART y C4.5.