

UN MÉTODO PARA SUSTITUCIÓN COMPLETA DE VALORES DESCONOCIDOS

Jesús A. Aranda B, Martha Millán G.
Universidad del Valle
Grupo GEDI
Escuela de Ingeniería de Sistemas y Computación
Cali-Colombia
{jesarana,millan}@eisc.univalle.edu.co

7 de julio de 2001

Resumen

Una de las etapas más importantes dentro del proceso de descubrimiento de conocimiento en bases de datos (DCBD), es la limpieza de datos, ya que permite atenuar la suciedad del conjunto de datos sobre el cual se aplican técnicas de data mining. Uno de los aspectos de interés dentro de la limpieza de datos lo constituye la sustitución de valores desconocidos.

En este artículo se propone un método completo y cooperativo para la sustitución de valores desconocidos. Lo completo hace referencia a que todos los valores desconocidos presentes en la base de datos son sustituidos y lo cooperativo a que se utilizan tres criterios de sustitución que se complementan dependiendo del tipo de valor desconocido de que se trate.

Palabras claves: Descubrimiento de conocimiento, Valores desconocidos, Reglas de asociación, Limpieza de datos, Bases de datos, Minería de datos.

Abstract

One of the most important in the Knowledge Discovery process in Databases (KDD) is the cleaning of data to reduce noise in a data set upon which data mining techniques are applied. One of the most important issues in data cleaning is the substitution of unknown values.

In this paper a complete and cooperative method for the substitution of unknown values is proposed. Complete refers to the fact that all unknown values in a data set are substituted. Cooperativeness has to do with the fact that three different substitution criteria are used to treat unknown values.

Keywords: Knowledge Discovery, Unknown values, Association rules, Data Cleaning, Databases, Data Mining