

Pattern-Based Transformation Approach to Relational Domain Learning Using Dynamic Aggregation for Relational Attributes.

Abstrak

Due to the widespread use of relational databases (MySQL, Oracle, DB2, MsSQL), most data are stored as multiple tables in what can be a very large database. As a result, more efficient algorithms for mining data from multirelational domain need to be implemented. Inductive Logic programming (ILP) techniques are useful for analyzing data in multi-relational databases. Unfortunately, even though not complex in structure, such business data are often large and contain highly non-determinate components, making them difficult for ILP learners geared towards structurally complex tasks. In this paper, we build a novel transformation-based approach to relational domain learning and describe the transformation process implemented through relational aggregation based on pattern distance. In this paper, we present the prototype of "Dynamic Aggregation of Relational Attributes" (hence called DARA) that is capable of mapping one-to-many relationship into one-to-one relationship, while preventing loss of information, in handling classification task in relational domains. We experimentally show these results in a multi-relational domain that show higher percentage of correctly classified instances and illustrate set of rules extracted using our approach.

Fig. 1. Financial Dataset Schema Example (PKDD CUP 99) Keywords— Data Pre-processing, Relational Data