Chapter 17
Database Marketing Process Supported by Ontologies: An Oil Company Distribution Network Case Study

Filipe Mota Pinto
Polytechnic Institute of Leiria, Portugal

ABSTRACT

The dramatic explosion of data and the growing number of different data sources are exposing researchers to a new challenge - how to acquire, maintain, and share knowledge from large databases in the context of rapidly applied and evolving research. This paper describes research on an ontological approach for leveraging the semantic content of ontologies to improve knowledge discovery in databases. We analyze how ontologies and knowledge discovery process may interoperate and present our efforts to bridge the two fields, knowledge discovery in databases and ontology learning for successful database usage projects.

INTRODUCTION

In artificial intelligence, ontology is defined as a specification of a conceptualization (Gruber, 1993). Ontology specifies at a higher level, the classes of concepts that are relevant to the domain and the relations that exist between these classes. Indeed, ontology captures the intrinsic conceptual structure of a domain. For any given domain, its ontology forms the heart of the knowledge representation.

In spite of ontology-engineering tools development and maturity, ontology integration in knowledge discovery projects remains almost unrelated.

DOI: 10.4018/978-1-4666-1598-4.ch017

Knowledge Discovery in Databases (KDD) process is comprised of different phases, such as data selection, preparation, transformation or modeling. Each one of these phases in the life cycle might benefit from an ontology-driven approach which leverages the semantic power of ontologies in order to fully improve the entire process (Gottgtroy et al., 2004).

Our challenge is to combine ontological engineering and KDD process in order to improve it. One of the promising interests in use of ontologies in KDD assistance is their use for guiding the process. This research objective seems to be much more realistic now that semantic web advances have given rise to common standards and technologies for expressing and sharing ontologies (Bernstein et al., 2005).

The three main operations of KDD can take advantage of domain knowledge embedded in ontologies such as: At the data understanding and data preparation phases, ontologies can facilitate the integration of heterogeneous data and guide the selection of relevant data to be mined, regarding domain objectives; During the modeling phase, domain knowledge allows the specification of constraints (e.g., parameters settings) for guiding data mining algorithms by, (e.g. narrowing the search space); finally, to the interpretation and evaluation phase, domain knowledge helps experts to visualize and validate extracted units.

KDD process is usually performed by experts who use their own knowledge for selecting the most relevant data in order to achieve domain objectives (Gottgtroy et al., 2003). Here we explore how the one ontology and its associated knowledge base can assist the expert at KDD process. Therefore, this document describes a research on a new approach to leveraging the semantic content of ontologies to improve KDD.

This paper is organized as follows: after this introductory part we present related background concepts. Then, we present related work on this area following the presentation and discussion of ontological assistance. The main contribution is presented in terms of a system prototype description and also system operation sections. Finally we draw some conclusions and address further research based on this research to future KDD data environment projects.

BACKGROUND

Knowledge Discovery in Databases

Knowledge discovery in databases (KDD) is the result of an exploratory process in order to achieve domain defined objectives involving the application of various algorithmic procedures for manipulating data, building models from data, and manipulating the models. The Data Mining phase deserves more attention from the research community: processes comprise multiple algorithmic components, which interact in nontrivial ways.

We consider tools that will help data analysts to navigate the space of KDD processes systematically, and more effectively. In particular, this paper focuses on a subset of stages of the KDD —those stages for which there are multiple algorithm components that can apply.

For most of this paper, we consider a prototypical KDD process template, similar to the one represented in Figure 1.The sequence of KDD phases is not strict. Moving back and forth between different phases is always required. It depends on the outcome of each phase, which one, or which particular task of a phase has to be performed next.

We focus our attention on the three main macro components of KDD life cycle: data understanding (data selection); data pre processing (all related data preparation and transformation activities), and modeling (data mining and the application of induction algorithms) We have chosen this set of components because, individually, they are relatively well understood—and they can be applied to a wide variety of benchmark data sets.