Ontology-Based Knowledge Model for Multi-View KDD Process

EL Moukhtar Zemmouri, Ecole Nationale Supérieure d’Arts et Métiers, Morocco

Hicham Behja, Ecole Nationale Supérieure d’Arts et Métiers, Morocco, and INRIA Sophia Antipolis, France

Abdelaziz Marzak, Universite Hassan II Mohammedia - Casablanca, Morocco

Brigitte Trousse, INRIA Sophia Antipolis, France

ABSTRACT

Knowledge Discovery in Databases (KDD) is a highly complex, iterative and interactive process that involves several types of knowledge and expertise. In this paper the authors propose to support users of a multi-view analysis (a KDD process held by several experts who analyze the same data with different viewpoints). Their objective is to enhance both the reusability of the process and coordination between users. To do so, they propose a formalization of viewpoint in KDD and a Knowledge Model that structures domain knowledge involved in a multi-view analysis. The authors’ formalization, using OWL ontologies, of viewpoint notion is based on CRISP-DM standard through the identification of a set of generic criteria that characterize a viewpoint in KDD.

Keywords: Cross-Industry Standard Process for Data Mining (CRISP-DM), Domain Knowledge, Knowledge Discovery in Databases (KDD) Process, Ontology, Viewpoint, Web Ontology Language (OWL)

INTRODUCTION

Knowledge Discovery in Databases (KDD) is a highly complex, iterative and interactive process, with a goal-driven and domain dependent nature (Fayyad et al., 1996). It involves three main steps (data preprocessing, data mining and post-processing) with many decisions made by the analyst (Figure 1). The complexity of KDD is mainly due to the nature of the analyzed data (distributed, incomplete, heterogeneous, etc.) and the nature of the process itself (since the KDD is by definition interactive and iterative).

Given this complexity of KDD, the analyst faces two major challenges. On the one hand, he must manipulate prior domain knowledge to better understand the data and the business objective. On the other hand, he must be able to choose, configure, compose and execute tools and methods from various fields (e.g., machine

DOI: 10.4018/jmcmc.2012070102
learning, statistics, artificial intelligence, databases) to achieve goals. The first challenge involves analyzed domain knowledge, while the second involves the analyst domain knowledge (Figure 1).

A multi-view KDD process is usually held by one or more experts who consequently manipulate several types of knowledge and know-how. They will have different objectives and preferences, different competences, and different visions of analyzed data, KDD methods and functions. In brief, they have different viewpoints. In this context, the KDD process will be guided by the analyst’s viewpoint (Behja et al., 2005) and several types of knowledge and expertise are incorporated.

Figure 2 shows an example of a multi-view analysis of data from an e-learning system (mainly: log files, database, and courses material). These data can be analyzed by different actors of the system (learners, teachers, administrator, marketing …). The objective of a teacher (e.g., evaluation of a course) is not the same as the administrator’s one (e.g., ensuring system reliability). Attributes used for evaluating a course are different from those used for studying the reliability. Similarly, chosen data mining methods, techniques and tools will be different, and the interpretation of data mining results depends on the analyst’s viewpoint. Therefore, it is fundamental to take into account the viewpoint of each analyst and incorporate the two types of domain knowledge in the KDD process.

In this paper we propose to assist the users of a multi-view KDD process. Our objective is to enhance both the reusability of the process and coordination between its different users. We propose a formalization of viewpoint notion in KDD following a knowledge engineering approach: eliciting, structuring, and formalizing information and knowledge involved in a multi-view analysis (Schreiber et al., 2000). Elicitation will be based on CRISP-DM standard (Chapman et al., 1999) to identify a set of generic criteria that characterize a viewpoint in KDD. Knowledge involved in a multi-view analysis will be structured as a knowledge model containing four hierarchical sub-models: domain model, task and method model, viewpoint model and viewpoint organizational model. The viewpoint sub-model will be formalized using
11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage: 
www.igi-global.com/article/ontology-based-knowledge-model-multi/69531?camid=4v1

www.igi-global.com/e-resources/library-recommendation/?id=2

Related Content

Quality of Experience Models for Multimedia Streaming
www.igi-global.com/article/quality-experience-models-multimedia-streaming/47328?camid=4v1a

Mobile Speech Recognition
www.igi-global.com/chapter/mobile-speech-recognition/26736?camid=4v1a

Headache App: Usability Assessment and Criterion Validity
www.igi-global.com/article/headache-app/205676?camid=4v1a