Chapter LIV
Ontologies Application to Knowledge Discovery Process in Databases

Héctor Oscar Nigro
Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

Sandra Elizabeth González Císaro
Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

INTRODUCTION

Nowadays one of the most important and challenging problems in Knowledge Discovery Process in Databases (KDD) or Data Mining is the definition of the prior knowledge; this can be originated either from the process or the domain. This contextual information may help select the appropriate information, features or techniques, decrease the space of hypothesis, represent the output in a more comprehensible way and improve the whole process.

Most part of this background knowledge is only present as implicit knowledge –in the analyst mind- or textual documentation. Therefore we need a conceptual model to help represent this knowledge. Advances in the field of Knowledge Engineering allow codifying previous knowledge under the ontological formalism. According to Gruber's (2002) ontology definition - explicit formal specifications of the terms in the domain and relations among them - we can represent the knowledge of the Knowledge Discovery process and knowledge about domain. Ontologies are used for communication (between machines and/or humans), automated reasoning, representation and reuse of knowledge. As a result, ontological foundation is a precondition for efficient automated usage of Knowledge Discovery information. As a result, we can perceive the relation between Ontologies and Data Mining in two ways:

- **From ontologies to data mining**, we are incorporating knowledge in the process through the use of ontologies, i.e. how the experts comprehend and carry out the analysis tasks. Representative applications are intelligent assistants for the discover

process (2005) interpretation and validation of mined knowledge, Ontologies for resource and service description and Knowledge Grids (Cannataro et al., 2007).

- **From data mining to ontologies**, we include domain knowledge in the input information or use the ontologies to represent the results. Therefore the analysis is done over these ontologies. The most distinctive applications are in Medicine, Biology and Spatial Data, such as Gene representation, Taxonomies, applications in Geosciences, medical applications and specially in evolving domains (Bogorny el al., 2006; Sidhu et al., 2006)

So far, the proposals or solutions that we find in KDD with ontologies are partial, i.e. they are centered on some of the steps of knowledge discovery. For instance Euler and Scholz (2004) present a metamodel of KDD preprocessing chains that contains an ontology describing conceptual domain knowledge. This metamodel is operational, yet abstract enough to allow the reuse of successful KDD applications in similar domains. Bernstein et al. (2005) propose an intelligent tool (IDA) based on a mining ontology. Brisson and Collar (2007) present the so-called KEOPS approach integrating expert knowledge all along the data mining process in a coherent and uniform manner. An ontology driven information system plays a central role in the approach.

The main goal of this paper is to present the issue of the ontologies application in KDD. As a result of our research, we will propose a general ontology-based model, which includes all discovery steps.

This paper is presented as follows: First, Background: main works in the field are introduced. Second, Main focus section is divided into: KDD Using Ontologies cycle in which we explain the knowledge process and propose a model, Domain Ontologies, Metadata Ontologies and Ontologies for Data Mining Process. Third: Future Trends, Conclusions, References and Key Terms.

BACKGROUND

This section describes the most recent research works in ontologies application to KDD. As you will appreciate, none of the research works alludes to the use of the ontologies in the whole process.

Singh et al. (2003) have developed a context aware data mining framework which provides accuracy and efficiency to data mining outcomes. Context factors were modeled using Ontological representation. Although the context aware framework proposed is generic in nature and can be applied to most of the fields. Hotho et al. (2003) have showed that using ontologies as filters in term selection prior to the application of a K-means clustering algorithm will increase the tightness and relative isolation of document clusters as a measure of improvement.

Pan and Shen (2005) have proposed architecture for knowledge discovery in evolving environments. The architecture creates a communication mechanism to incorporate known knowledge into the discovery process through ontology service facility.

Rennolls (2005) have developed an intelligent framework for Data Mining, Knowledge Discovery and Business Intelligence. The ontological framework will guide the user to the models from an expanded data mining toolkit, and the epistemological framework will assist the user to interpret and appraise the discovered relationships and patterns.

Li and Zhong (2006) have introduced an approach to automatically discover ontologies from data sets in order to build complete concept models for Web user information needs. They proposed a method to capture evolving patterns to refine discovered ontologies and established a process to assess relevance of patterns in an ontology by the dimensions of exhaustively and specificity.

Brezany et al. (2005) have addressed the issues of composing workflows with automated support developed on top of Semantic Web technolo-