Ontology has recently received considerable attention in the knowledge management community. This article discusses the needs of ontology development for data mining. Based on a domain analysis of knowledge representations in data mining, it proposes a generic structure of ontologies for data mining. Furthermore, this article specifies the unique ontology resources of the subdomain of innovative data mining with incomplete data. A project on an ontology-based data mining system for discovering knowledge from incomplete data demonstrates the effectiveness of ontology in knowledge management.

Keywords: data mining; incomplete data; knowledge management; ontology

INTRODUCTION

Knowledge management is one of the important topics in the database management field (D. Zhang & Zhao, 2006). There have been two important research themes related to knowledge management: data mining (Cunningham, Song, & Chen, 2006; Fayyad, Piatetsky-Shapiro, & Smyth, 1996; J. Wang, 2005) and ontology (Green & Rosemann, 2004; Kim, 2002). Data mining is the process of trawling through data to find interesting patterns (Hand, 1998). The process of data mining is a knowledge management process because it involves managing the data mining of experts’ knowledge (Brachman, Khabaza, Kloesgen, Piatetsky-Shapiro, & Simoudis, 1996). On the other hand, ontology is a science that studies explicit formal specifications of the resources and relations among them in the domain (Gruber, 1993). An ontology is a specification of a conceptualization (Gruber, 1995) and is intended for knowledge sharing among applications (Welty, 2003). In the past few years, the two themes have become well-recognized substrates for research into knowledge management (Li & Zhong, 2006; Nilakanta, Miller, & Zhu, 2006). Yet, potential benefits of joining the two themes have not been fully explored.

Presumably, the use of ontologies for data mining can be beneficial for knowledge management through sharing common understanding
of the context and techniques of data mining among data miners (Zaki & Pan, 2002). However, little research on the general structure of ontologies in data mining has been reported. In this article, we first discuss the key knowledge elements of data mining and propose a generic structure of ontology for the domain. We then place the emphasis on ontology development for data mining with incomplete data. Through a project on an ontology-based data mining system, we demonstrate the effectiveness of ontology in data mining.

ONTOGOY IN DATA MINING

Ontology

According to the resource description framework (RDF; World Wide Web Consortium [W3C], 2008), a primitive ontology is a triple containing a subject, an object, and a predicate (relationship; see Figure 1a). Its special form that represents the reciprocal relation of resources (dual subject and object) is shown in Figure 1b. A large ontology for an entire domain is a composition of a set of primitive ontologies.

Resources are knowledge elements including data, procedures, rules, ideas that guide actions, and decisions (Alter, 1996; Beckman, 1999; Tobin, 1996; van der Spek & Spijkervet, 1997). Given the complexity of knowledge structures in general, a resource itself can be represented by an ontology. In data mining, an ontology is a network of all these resources that shows the coordination of data mining actions for the data miner to achieve a certain goal.

Categories of Resources of Data Mining

There has been moderate literature on ontologies associated with data mining (e.g., Bernstein, Provost, & Hill 2005; Kim, 2002; Li & Zhong, 2006). However, few research reports have provided generic structures of ontologies for data mining. Our motivation for this article is to propose generic resource categories for the domain of data mining based on the premise that a taxonomy of formalized generic resource categories can help people to better understand and share the ontologies (Welty, 2003), and to demonstrate the usefulness of the proposed taxonomy.

Task

A task of data mining is to discover meaningful patterns and properties of the data for the data miner. A task can be formally described as a hierarchical structure of its subtasks.

Data

Data are the key resource in data mining. Definitions of the data items and metadata of

Figure 1. Primitive ontology

![Figure 1. Primitive ontology](image-url)
Related Content

Performance Comparison of Static vs. Dynamic Two Phase Locking Protocols
[www.irma-international.org/article/performance-comparison-static-dynamic-two/51102/](www.irma-international.org/article/performance-comparison-static-dynamic-two/51102/)

A Graph-Based Approach for Semantic Process Model Discovery
[www.irma-international.org/chapter/graph-based-approach-semantic-process/58622/](www.irma-international.org/chapter/graph-based-approach-semantic-process/58622/)

Behavioral Aspects of Data Production and Their Impact on Data Quality
[www.irma-international.org/article/behavioral-aspects-data-production-their/51119/](www.irma-international.org/article/behavioral-aspects-data-production-their/51119/)

Simple and Incremental Nearest-Neighbor Search in Spatio-Temporal Databases
[www.irma-international.org/chapter/simple-incremental-nearest-neighbor-search/29665/](www.irma-international.org/chapter/simple-incremental-nearest-neighbor-search/29665/)

Adaptive XML-to-Relational Storage Strategies
[www.irma-international.org/chapter/adaptive-xml-relational-storage-strategies/20771/](www.irma-international.org/chapter/adaptive-xml-relational-storage-strategies/20771/)