A Pattern Language for Knowledge Discovery in a Semantic Web Context

Mehdi Adda, University of Quebec at Rimouski, Canada

ABSTRACT

Ontologies are used to represent data and share knowledge of a specific domain, and in recent years they tend to be used in many applications such as database integration, peer-to-peer systems, e-commerce, semantic web services, bioinformatics, or social networks. Feeding ontological domain knowledge into those applications has proven to increase flexibility, inter-operability and interpretability of data and knowledge. As more data is gathered/generated by those applications, it becomes important to analyze and transform it to meaningful information. One possibility is to use data mining techniques to extract patterns from those large amounts of data. One challenging general problem in mining ontological data is taking into account not only domain concepts, properties and instances, but also hierarchical structures of those concepts and properties. In this paper, the author researches the specific problem of extracting ontology-based sequential patterns.

Keywords: Data Mining, Ontological Data, Ontologies, Semantic Web, Sequential Patterns

I. INTRODUCTION

Web usage mining (WUM) is about the discovery of trends in user navigational behavior within a Web application (Mobasher, Jin, & Zhou, 2004; Pierrakos, Paliouras, Papatheodorou, & Spyropoulos, 2003; Eirinaki, Lampos, Vazirgiannis, & Varlamis, 2003). A WUM task is typically motivated by the need to understand the strategies adopted by the user (e.g., within an e-Learning context) or to personalize the content served by the application (e.g., in e-Commerce). The standard approach for WUM is association rule extraction from the application logs (Mobasher, Jin, & Zhou, 2004; Melville, Mooney, & Nagarajan, 2002; Jin & Mobasher, 2003).

With the emergence of the Semantic Web and the associated use of ontologies to describe the semantic content of web documents, the emphasis has been shifted on the ability of WUM engines to incorporate large amounts of domain knowledge (Eirinaki, Lampos, Vazirgiannis, & Varlamis, 2003; Dai & Mobasher, 2004; Liao, Chen, & Hsu, 2009). Although the topic of combining semantics and WUM has been largely covered in the literature, the specific impact of using an ontology in the mining process has not been clearly stated. In fact, most studies assume the domain ontology to be a mere conceptual hierarchy, thus reducing...
the mining problem to, at worst, a generalized sequential pattern extraction.

It is noteworthy that the full-blown ontologies that may underlay sophisticated Web information systems such as portals, virtual museums, Web e-learning and e-commerce environments, etc., have more structure than a mere conceptual is-a tree, in particular, they often comprise a collection of inter-concept relations, possibly organized into a separate is-a hierarchy. At the individual level, the relations materialize as inter-object links that belong to object descriptions and hence may be relevant for mining.

Such links induce higher-order dependencies between the content objects in a sequence, leading to a secondary graph structure (Bandopadhyay, Maulik, Holder, & Cook, 2005; Mineau, Moulin, & Sowa, 1993; Huan, Wang, Prins, & Yang, 2004). While the mining of graph-shaped data is now a relatively well understood topic, its enhancement to graphs with hierarchically organized labels on both vertices (objects) and edges (links) as provided by a full-blown ontology, is far less studied (Washio & Motoda, 2003). The goal of this study is therefore to define a mining discipline for this new pattern type.

The work presented in this paper is an enhancement and completeness of a framework previously proposed to mine this new category of patterns (Adda, Valtchev, Missaoui, & Djeraba, 2007). It is composed of a pair of languages (a data and a pattern one) provided with a generality relationship as well as a level-wise mining method, xPMiner, based on a set of specialization operators.

II. RELATED WORK

Semantic Web is designed to let users and machines describe resources, share that data in a distributed manner and enable interpretation and processing of the related data (Lee, Hendler, & Lassila, 2001; Lee, 2001). In recent years, ontologies are widely used to realize the data layer of Semantic Web. In this paper the specific problem of mining ontology-powered systems is addressed. One reason for focusing in the knowledge discovery aspect of a Semantic Web environment is that the concepts and relations of an ontology have an impact on the quality and quantity of the patterns that may be extracted from such graph-based data (Di-Jorio, Bringay, Fiot, Laurent, & Teisseire, 2008; Liao, Chen, & Hsu, 2009; Rajapaksha & Kodagoda, 2008).

Mining data in the context of ontology-powered systems involves different fields such as ontology engineering, knowledge discovery and pattern mining. Hereafter, the background knowledge related to this study on both pattern mining and ontologies is presented.

II.1. Pattern Mining

Given a universe of objects, or items, \( O \), a database \( D \) made of records combining items from \( O \), and a frequency, or support, threshold \( \sigma \), frequent pattern mining amounts to extract the family \( F_{\sigma} \) of item collections, or patterns, that are present in at least \( \sigma \) records. Two languages, the pattern language \( \Gamma \) and the data one \( \Delta \), and two binary relations underlay the problem: the generality between patterns \( \sqsubseteq \) and instantiation between a data record and a pattern \( \prec \). Generality follows instantiation as given a pattern \( f \in \Gamma \) and a super-pattern thereof \( \overline{f} (f \sqsubseteq \overline{f}) \), each record \( d \in \Delta \) instantiating \( d \prec \overline{f} \) instantiates \( \overline{f} \) as well.

In the simplest settings, both data records and patterns are sets of items (itemsets), i.e., \( \Delta = \Gamma = 2^{O} \), while \( \sqsubseteq \) and \( \prec \) boil down to set-theoretic inclusion. Hence the mining goal amounts to finding all the frequent subsets of a family of sets \( D \sqsubseteq \Delta \).

More elaborate record structures have been studied including sequences (Agrawal & Srikant, 1995; Di-Jorio, Bringay, Fiot, Laurent, & Teisseire, 2008) \( \Delta, \Gamma, = \omega^{O} \) and graphs (Inokuchi, Washio, & Motoda, 2000) with equally more complex generality and instan-
Related Content

Design and Operation of a Cell Phone-Based Community Hazard Information Sharing System
[www.igi-global.com/article/design-operation-cell-phone-based/65068?camid=4v1a](www.igi-global.com/article/design-operation-cell-phone-based/65068?camid=4v1a)

Ripple Effect in Web Applications
[www.igi-global.com/article/ripple-effect-web-applications/44919?camid=4v1a](www.igi-global.com/article/ripple-effect-web-applications/44919?camid=4v1a)
Virtual Machine Placement Strategy for Cloud Data Center
www.igi-global.com/chapter/virtual-machine-placement-strategy-for-cloud-data-center/140829?camid=4v1a

MCWDF: Micro Chunk Based Web Delivery Framework
www.igi-global.com/article/mcwdf/193006?camid=4v1a