Chapter 3

AL-QUIN: An Onto-Relational Learning System for Semantic Web Mining

Francesca A. Lisi
Università degli Studi di Bari “Aldo Moro”, Italy

ABSTRACT

Onto-Relational Learning is an extension of Relational Learning aimed at accounting for ontologies in a clear, well-founded and elegant manner. The system AL-QUIN supports a variant of the frequent pattern discovery task by following the Onto-Relational Learning approach. It takes taxonomic ontologies into account during the discovery process and produces descriptions of a given relational database at multiple granularity levels. The functionalities of the system are illustrated by means of examples taken from a Semantic Web Mining case study concerning the analysis of relational data extracted from the on-line CIA World Fact Book.

INTRODUCTION

Semantic Web Mining is an application area which aims at combining the two areas of Semantic Web and Web Mining from a twofold perspective (Stumme, Hotho, & Berendt, 2006). On one hand, the new semantic structures in the Web can be exploited to improve the results of Web Mining. On the other hand, the results of Web Mining can be used for building the Semantic Web. Most work in Semantic Web Mining simply extends previous work to the new application context, e.g., Maedche and Staab (2000) apply a well-known algorithm for association rule mining to discover conceptual relations from text. Indeed, we argue that Semantic Web Mining can be considered as
Data Mining (DM) for/from the Semantic Web. Current DM systems could serve the purpose of Semantic Web Mining if they were more compliant with, e.g., the standards of representation for ontologies in the Semantic Web and/or interoperable with well-established tools for Ontological Engineering (Gómez-Pérez, Fernández-López, & Corcho, 2004), e.g., Protégé (Knublauch, Ferguson, Noy, & Musen, 2004), that support these standards. In particular, those Knowledge Representation (KR) formalisms known as Description Logics (DLs) which collectively form a family of decidable fragments of First Order Logic (FOL) (Baader, Calvanese, McGuinness, Nardi, & Patel-Schneider, 2007) have been the starting point for the W3C recommended Ontology Web Language (OWL) (Horrocks, Patel-Schneider, & van Harmelen, 2003).

Algorithms for Relational Learning try to avoid the limitation of classical Machine Learning (ML) in describing the relationships between more than one object by using a richer representation language, such as (fragments of) FOL (De Raedt, 2008). This logic-based approach to Relational Learning is also known under the name of Inductive Logic Programming (ILP) (Nienhuys-Cheng & de Wolf, 1997). ILP has been historically concerned with concept learning from examples and background knowledge within the representation framework of Horn Clausal Logic (HCL) (underlying the Logic Programming paradigm) and with the aim of prediction, e.g., the system FOIL learns classification rules from examples expressed in the form of relations (Quinlan, 1990). More recently the ILP community has developed an interest in applications for relational databases, thus giving rise to the field of Relational Data Mining (Džeroski & Lavrač, 2001). Systems like Warmr (Dehaspe & Toivonen, 1999) adopt Datalog (Ceri, Gottlob, & Tanca, 1990) (a logical language for relational databases which corresponds to a function-free fragment of HCL) as representation language and support tasks of description rather than prediction in challenging applications, such as trying to understand the structure of proteins and designing pharmaceuticals. Though ILP is widely recognized as a major approach to Relational Learning, ILP as such cannot face the challenges of the Semantic Web, notably the availability of ontologies to deal with. This inadequacy of Relational Learning is due to the following KR issue: DLs are incomparable with HCL as regards the expressive power (Borgida, 1996) and the semantics (Rosati, 2005). Yet, since DLs and HCL can be combined according to some limited forms of hybridization (Rosati, 2005), the adoption of hybrid KR systems such as $\mathcal{AL}$-log (Donini, Lenzerini, Nardi, & Schaerf, 1998) can help overcome the current difficulties in accommodating ontologies in ILP (Lisi & Esposito, 2009). We claim that this paves the way to an extension of Relational Learning, called Onto-Relational Learning, to account for ontologies in a clear, well-founded and elegant manner.

In this paper we present an Onto-Relational Learning system, named $\mathcal{AL}$-QuIn, that adopts ILP as a methodological apparatus and $\mathcal{AL}$-log as a KR formalism. The system solves a variant of the popular DM problem of frequent pattern discovery which takes taxonomic ontologies into account during the discovery process and produces descriptions of a given relational database at multiple granularity levels. The functionalities of the system are illustrated by means of examples taken from a Semantic Web Mining case study concerning the analysis of relational data extracted from the on-line CIA World Fact Book.

The paper is structured as follows. First we shall provide a motivation for Onto-Relational Learning by showing the limits of systems like Warmr in dealing with prior knowledge in the form of taxonomic ontologies, and the necessary background information to readers not acquainted with reasoning and learning in $\mathcal{AL}$-log. Then we shall present the techniques and the algorithms employed in $\mathcal{AL}$-QuIn and describe how $\mathcal{AL}$-log
Related Content

Role of Vocabularies for Semantic Interoperability in Enabling the Linked Open Data Publishing
Ahsan Morshed (2013). *Teaching Cases Collection* (pp. 84-104).
[www.igi-global.com/chapter/role-vocabularies-semantic-interoperability-enabling/77201?camid=4v1a](www.igi-global.com/chapter/role-vocabularies-semantic-interoperability-enabling/77201?camid=4v1a)

Online Semantic Knowledge Management for Product Design Based on Product Engineering Ontologies
[www.igi-global.com/article/online-semantic-knowledge-management-product/63644?camid=4v1a](www.igi-global.com/article/online-semantic-knowledge-management-product/63644?camid=4v1a)

Representing Knowledge
[www.igi-global.com/chapter/representing-knowledge/44923?camid=4v1a](www.igi-global.com/chapter/representing-knowledge/44923?camid=4v1a)

Semantics for the Semantic Web: The Implicit, the Formal and the Powerful
[www.igi-global.com/article/semantics-semantic-web/2802?camid=4v1a](www.igi-global.com/article/semantics-semantic-web/2802?camid=4v1a)