Chapter 13
Inductive Classification of Semantically Annotated Resources through Reduced Coulomb Energy Networks

Nicola Fanizzi
Università degli studi di Bari, Italy

Claudia d’Amato
Università degli studi di Bari, Italy

Floriana Esposito
Università degli studi di Bari, Italy

ABSTRACT

The tasks of resource classification and retrieval from knowledge bases in the Semantic Web are the basis for a lot of important applications. In order to overcome the limitations of purely deductive approaches to deal with these tasks, inductive (instance-based) methods have been introduced as efficient and noise-tolerant alternatives. In this paper we propose an original method based on a non-parametric learning scheme: the Reduced Coulomb Energy (RCE) Network. The method requires a limited training effort but it turns out to be very effective during the classification phase. Casting retrieval as the problem of assessing the class-membership of individuals w.r.t. the query concepts, we propose an extension of a classification algorithm using RCE networks based on an entropic similarity measure for OWL. Experimentally we show that the performance of the resulting inductive classifier is comparable with the one of a standard reasoner and often more efficient than with other inductive approaches. Moreover, we show that new knowledge (not logically derivable) is induced and the likelihood of the answers may be provided.

DOI: 10.4018/978-1-60960-593-3.ch013
1 INTRODUCTION

The tasks of resource classification and retrieval from knowledge bases (KBs) in the Semantic Web (SW) are the basis for many important knowledge-intensive applications. However, the inherent incompleteness and accidental inconsistency of knowledge bases in the Semantic Web requires new different methods which are able to perform such tasks efficiently and effectively (although with some acceptable approximation). Instance-related tasks are generally tackled by means of logical approaches that try to cope with the problems mentioned above. This has given rise to alternative methods for approximate reasoning (Wache, Groot & Stuckenschmidt, 2005), (Hitzzler & Vrandecic, 2005), (Haase, van Harmelen, Huang, Stuckenschmidt & Halberstadt, 2005), (Möller, Haarslev & Wessel, 2006), (Huang & van Harmelen, 2008), (Tserendorj, Rudolph, Krötzsch & Hitzler, 2008), (Rudolph, Tserendorj & Hitzler, 2008). Inductive methods for approximate reasoning are known to be often quite efficient, scalable, and noise-tolerant.

Recently, first steps have been taken to apply classic machine learning techniques for building inductive classifiers for the complex representations, and related semantics, adopted in the context of the SW (Fanizzi, d’Amato & Esposito, 2008a), especially through non-parametric\(^1\) statistical methods (d’Amato, Fanizzi & Esposito, 2008), (Fanizzi, d’Amato & Esposito, 2008d). Instance-based inductive methods may help a knowledge engineer populate ontologies (Baader, Ganter, Sertkaya & Sattle, 2007). Some methods are also able to complete ontologies with probabilistic assertions derived exploiting the missing and sparse data in the ontologies (Rettinger, Nickles & Tresp, 2009). Further sophisticated approaches are able of dealing with uncertainty encoded in probabilistic ontologies through suitable forms of reasoning (Lukasiewicz, 2008).

In this paper we propose a novel method for inducing classifiers from ontological data that may naturally be employed as an alternative way for performing concept retrieval (Baader, Calvanese, McGuinness, Nardi, Patel-Schneider, 2003) and several other related applications. Even more so, like its predecessors mentioned above, the induced classifier is also able to determine a likelihood measure of the induced class-membership assertions which is important for approximate query answering and ranking. Some assertions could not be logically derived, but may be highly probable according to the inductive classifier; this may help to cope with the uncertainty caused by the inherent incompleteness of the KBs even in absence of an explicit probabilistic model.

Specifically, we propose to answer queries adopting an instance-based classifier, the Reduced Coulomb Energy (RCE) network (Duda, Hart & Stork, 2001), induced by a non-parametric learning method. The essentials of this learning scheme have been extended to be applied to the standard representations of the SW via semantic similarity measures for individual resources. As with other similarity-based methods, a retrieval procedure may seek for individuals belonging to query concepts, exploiting the analogy with other training instances, namely the classification of the nearest ones (w.r.t. the measure of choice). Differently from other lazy-learning approaches experimented in the past (d’Amato, Fanizzi & Esposito, 2008) which do not require training, yet more similarly to the non-parametric methods based on kernel machines (Bloehdorn & Sure, 2007), (Fanizzi, d’Amato & Esposito, 2008d), the new method is organized in two phases:

- In the training phase, an RCE network based on prototypical individuals (parameterized for each prototype) is trained to detect the membership of further individuals w.r.t. some query concept;
- The network is then exploited by the classifier, during the classification phase, to make a decision on the class-membership...