Editorial

Inductive Reasoning and Machine Learning for the Semantic Web

Claudia d'Amato [a], Nicola Fanizzi [a], Marko Grobelnik [b], Agnieszka Lawrynowicz [c] and Vojtech Svatek [d]

[a] *Department of Computer Science – University of Bari, Italy*
E-mail: {claudia.damato,nicola.fanizzi}@uniba.it
[b] *Jozef Stefan Institute, Ljubljana, Slovenia*
E-mail: marko.grobelnik@ijs.si
[c] *Poznan University of Technology, Poland*
E-mail: agnieszka.lawrynowicz@cs.put.poznan.pl
[d] *University of Economics, Prague, Czech Republic*
E-mail: svatek@vse.cz

1. Introduction

The goal of the Semantic Web research is to extend the current World WideWeb into one where information is given semantically precise meaning. The Semantic Web is supposed to provide a medium to share such information and is often perceived to be a global, distributed database of meaningfully represented data. From the speed of the development of initiatives like Linked Open Data or community efforts to integrate various ontological resources like bio-ontologies, we may expect to have, in the near future, access to immense data resources that would by orders of magnitude escape the possibility of direct human analysis. Hence, the problem of mining from the Semantic Web currently becomes as important as it used to be that of data mining from local databases. Exploiting this global resource of data requires new kinds of approaches to data mining and data analysis that would be able to deal at the same time with its scale and with the complexity, expressiveness, and heterogeneity of the representation languages, leverage on availability of ontologies and explicit semantics of the resources, and account for novel assumptions (e.g., the "open world") that underlie reasoning services within the Semantic Web. The open, distributed and inherently incomplete nature of the Semantic Web environment poses problems for deductive approaches, traditionally employed to reason with logic-based ontological data. Hence, one may witness a recent trend in the Semantic Web community to propose complementary forms of reasoning, preferably more efficient and noise-tolerant.

This Special Issue of the Semantic Web Journal builds upon three consecutive editions of the International Workshop on Inductive Reasoning and Machine Learning for the Semantic Web (IRMLeS), held in 2009–2011 and devoted to the the topic area outlined above. Out of a pool of 22 papers accepted, five have been selected by votes of the IRMLeS workshop chairs and their authors invited to submit an extended version to the Special Issue. Eventually, three extended papers were submitted and passed through the review process. Huang et al. addressed learning from general graph data expressing facts about entities, such that missing relationships could be predicted. Novel use of (urn-model-based) statistical sampling was elaborated for this purpose. Experiments have been carried out both in lay domains (FOAF and DBpedia data) and in a specialized (biomedical) domain. Vidal et al. focused on prediction of a special kind of relationship – that between a document and ontology/vocabulary terms that characterize its content. Technically, the relationship is first approximated using ranking functions. The task can however be viewed as a pre-processing task for actual discovery of novel,

scientifically valid relationships among entities, which is analogous to the previous case. The life science domain was once more addressed. Bouza & Bernstein addressed the recommender systems domain, and for this purpose introduced classification-based similarity metrics for comparing user preferences and discussed the concept of partial preference similarity based on a machine learning model. Their method enables incorporation of background knowledge in the form of ontologies into recommendation systems in cases when a machine learning algorithm exploiting ontologies is used. Regarding the type of (semantic) data sources, the approaches by Huang et al. and Vidal et al. both operate on data published in the form of semantic graphs represented in RDF. Huang et al. argue that Statistical Relational Learning (SRL) is particularly well suited to operate on interlinked, sparse and incomplete data (such as LOD) in contrast to classical machine learning approaches that operate on attribute-value representations and to Inductive Logic Programming (ILP) approaches that tackle machine learning in close-to-deterministic domains. They also show how ontologi-

cal background knowledge can be integrated into their approach. Vidal et al. stressed that not only knowledge represented within the ontologies but also the knowledge encoded in annotations may together provide the basis for new discoveries. Bouza & Bernstein differ from the previous two in that they do not directly address Semantic Web data; however, they model user preference similarity taking advantage of a preference ontology. Their approach thus mainly features the opposite direction of interaction between data mining and semantics: using semantic models in order to guide the analysis of data collected from the real world.

Generally, the Special Issue appears representative of the recent advances in this interdisciplinary field, and will hopefully become one of the seeds of future research. We would like to express our thanks to all people who contributed to its completion, from the authors through reviewers to the SWJ editorial staff.

Claudia d'Amato, Agnieszka Lawrynowicz, Nicola Fanizzi, Marko Grobelnik, Vojtech Svatek, guest editors.