Towards Semantic Annotation Supported by Dependency Linguistics and ILP

Jan Dědek and Peter Vojtáš

Department of Software Engineering, Charles University,
Prague, Czech Republic
{dedek,vojtas}@ksi.mff.cuni.cz

Abstract. The abstract should summarize the contents of the paper and should contain at least 70 and at most 150 words. It should be written using the *abstract* environment.

Keywords: Semantic Annotation, Dependency Linguistics, Inductive Logic Programming, Information Extraction, Machine Learning

1 Introduction

a. The Semantic Web and Information Extraction b. Information Extraction and Machine Learning Problem of feature selection, deep linguistic analysis is not used very often or has to be at lo level. c. Linguistics: Phrase structure and dependency structure d. Similarity of RDF and dependency linguistics e. Our contributions and benefits of using the combination of tools

2 Related work

a. ILP users b. Deep parsing users c. Information extraction - GATE d. Semantic annotation - GATE

3 Exploited methods - ILP and linguistics

a. PDT and TectoMT b. Stanford dependencies c. Czech language d. ILP ILP is capable to find complex and meaningful rules that cover the intended information. As we do not have large amount of training data, there is no problem with excessive time demands during learning and the application of the learned rules is simple and quick.

4 Implementation

a. ILP Wrapper Help of "Machine Learning PR" which is old and almost obsolete but easy to extend. b. Learning / application c. ILP serialization, annotation / tree node intersection d. Root - subtree, robustness e. Semantic interpretation f. How to download

5   Evaluation

a. Dataset b. Paum classifier as a baseline c. Cross validation, at the end all documents are processed with such learned model that the particular document was not used in corresponding learning procedure. So all the documents are processed as unseen during learning. d. Results e. Examples of learned rules f. Experience with human-designed rules.

6   Conclusion and future work

a. Application? - accident seriousness ranking, environment protection, economy b. From our experiments can be seen that ILP is capable to find complex and meaningful rules that cover the intended information. But in terms of the performance measures the results are not better than those from propositional learner.

References