State-of-the-Art Kernels for Natural Language Processing

Alessandro Moschitti
Department of Computer Science and Information Engineering
University of Trento
Via Sommarive 5, 38123 Povo (TN), Italy
moschitti@disi.unitn.it

Introduction

In recent years, machine learning (ML) has been used more and more to solve complex tasks in different disciplines, ranging from Data Mining to Information Retrieval or Natural Language Processing (NLP). These tasks often require the processing of structured input, e.g., the ability to extract salient features from syntactic/semantic structures is critical to many NLP systems. Mapping such structured data into explicit feature vectors for ML algorithms requires large expertise, intuition and deep knowledge about the target linguistic phenomena. Kernel Methods (KM) are powerful ML tools (see e.g., (Shawe-Taylor and Cristianini, 2004)), which can alleviate the data representation problem. They substitute feature-based similarities with similarity functions, i.e., kernels, directly defined between training/test instances, e.g., syntactic trees. Hence feature vectors are not needed any longer. Additionally, kernel engineering, i.e., the composition or adaptation of several prototype kernels, facilitates the design of effective similarities required for new tasks, e.g., (Moschitti, 2004; Moschitti, 2008).

Tutorial Content

The tutorial aims at addressing the problems above: firstly, it will introduce essential and simplified theory of Support Vector Machines and KM with the only aim of motivating practical procedures and interpreting the results. Secondly, it will simply describe the current best practices for designing applications based on effective kernels. For this purpose, it will survey state-of-the-art kernels for diverse NLP applications, reconciling the different approaches with a uniform and global notation/theory. Such survey will benefit from practical expertise acquired from directly working on many natural language applications, ranging from Text Categorization to Syntactic/Semantic Parsing. Moreover, practical demonstrations using SVM-Light-TK toolkit will nicely support the application-oriented perspective of the tutorial. The latter will lead NLP researchers with heterogeneous background to the acquisition of the KM know-how, which can be used to design any target NLP application.

Finally, the tutorial will propose interesting new best practices, e.g., some recent methods for large-scale learning with structural kernels (Severyn and Moschitti, 2011), structural lexical similarities (Croce et al., 2011) and reverse kernel engineering (Pighin and Moschitti, 2009).

References

Danilo Croce, Alessandro Moschitti, and Roberto Basili. 2011. Structured Lexical Similarity via Convolution Kernels on Dependency Trees. In *Proc. of EMNLP*.

Alessandro Moschitti. 2004. A Study on Convolution Kernels for Shallow Semantic Parsing. In *Proceedings of ACL*.

Alessandro Moschitti. 2008. Kernel Methods, Syntax and Semantics for Relational Text Categorization. In *Proceedings of CIKM*.

Daniele Pighin and Alessandro Moschitti. 2009. Efficient Linearization of Tree Kernel Functions. In *Proceedings of CoNLL*.

Aliaksei Severyn and Alessandro Moschitti. 2011. Fast Support Vector Machines for Structural Kernels. In *ECML*.

John Shawe-Taylor and Nello Cristianini. 2004. *Kernel Methods for Pattern Analysis*. Cambridge Univ. Press.

Tutorial Abstracts of ACL 2012, page 2,
Jeju, Republic of Korea, 8 July 2012.