Semantic Matching using Kernel Methods

Alessandro Moschitti
Qatar Computing Research Institute, Doha, Qatar
amoschitti@qf.org.qa

ABSTRACT

Semantic matching (SM) for textual information can be informally defined as the task of effectively modeling text matching using representations more complex than those based on simple and independent set of surface forms of words or stems (typically indicated as bag-of-words). In this perspective, matching named entities (NEs) implies that the associated model can both overcomes mismatch between different representations of the same entities, e.g., *George H. W. Bush* vs. *George Bush*, and carry out entity disambiguation to avoid incorrect matches between different but similar entities, e.g., the entity above with his son *George W. Bush*. This means that both the context and structure of NEs must be taken into account in the IR model.

SM becomes even more complex when attempting to match the shared semantics between two larger pieces of text, e.g., phrases or clauses, as there is currently no theory indicating how words should be semantically composed for deriving the meaning of text.

The complexity above has traditionally led to define IR models based on bag-of-word representations in the vector space model (VSM), where (i) the necessary structure is minimally taken into account by considering n-grams or phrases; and (ii) the matching coverage is increased by projecting text in latent semantic spaces or alternatively by applying query expansion. Such methods introduce a considerable amount of noise, which negatively balances the benefit of achieving better coverage in most cases, thus producing no IR system improvement.

In the last decade, a new class of semantic matching approaches based on the so-called Kernel Methods (KMs) for structured data (see e.g., [4]) have been proposed. KMs also adopt scalar products (which, in this context, take the names of kernel functions) in VSM. However, KMs introduce two new important aspects:

- the scalar product is implicitly computed using smart techniques, which enable the use of huge feature spaces, e.g., all possible skip n-grams; and
- KMs are typically applied within supervised algorithms, e.g., SVMs, which, exploiting training data, can filter out irrelevant features and noise.

In this talk, we will briefly introduce and summarize, the latest results on kernel methods for semantic matching by focusing on structural kernels. These can be applied to match syntactic and/or semantic representations of text shaped as trees. Several variants are available: the Syntactic Tree Kernels (STK), [2], the String Kernels (SK) [5] and the Partial Tree Kernels (PTK) [4]. Most interestingly, we will present tree kernels exploiting SM between words contained in a text structure, i.e., the Syntactic Semantic Tree Kernels (SSTK) [1] and the Smoothed Partial Tree Kernels (SPTK) [3]. These extend STK and PTK by allowing for soft matching (i.e., via similarity computation) between nodes associated with different but related labels, e.g., synonyms. The node similarity can be derived from manually annotated resources, e.g., WordNet or Wikipedia, as well as using corpus-based clustering approaches, e.g., latent semantic analysis (LSA). An example of the use of such kernels for question classification in the question answering domain will illustrate the potentials of their structural similarity approach.

1. REFERENCES

[1] S. Bloehdorn and A. Moschitti. Combined syntactic and semantic kernels for text classification. In *Proceedings of ECIR 2007*, Rome, Italy, 2007.

[2] M. Collins. Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pages 1–8, University of Pennsylvania, United States, 2002.

[3] D. Croce, A. Moschitti, and R. Basili. Structured Lexical Similarity via Convolution Kernels on Dependency Trees. In *Proceedings of EMNLP 2011*, 2011.

[4] A. Moschitti. Efficient convolution kernels for dependency and constituent syntactic trees. In *Proceedings of ECML*, 2006.

[5] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis.* Cambridge University Press, 2004.

Bio

Alessandro Moschitti is a Senior Research Scientist at the Qatar Computing Research Institute (QCRI) and a tenured professor at the CS Department of the University of Trento, Italy. He obtained his PhD in CS from the University of Rome in 2003. He has been the only non-US faculty member to participate in the IBM Watson Jeopardy! challenge. He has significant expertise in both theoretical and applied machine learning for NLP, IR and Data Mining. He has devised innovative kernels for advanced syntactic/semantic processing with support vector and other kernel-based machines. He is an author or co-author of about 190 scientific articles in the area above. He has been an area chair for the semantics track at ACL and IJCNLP conferences and for machine learning at ACL and ECML. Additionally, he has been PC chair of other important conferences and workshops for the ML and ACL communities. Currently, he is the General Chair of EMNLP 2014 and he is on the editorial board of JAIR, JNLE and JoDS. He has received three IBM Faculty Awards, one Google Faculty Award and three best paper awards.