Semantic Role Labeling Tutorial

Diego Marcheggiani, Michael Roth, Ivan Titov, Benjamin Van Durme

Description

This tutorial describes semantic role labelling (SRL), the task of mapping text to shallow semantic representations of eventualities and their participants. The tutorial introduces the SRL task and discusses recent research directions related to the task. The audience of this tutorial will learn about the linguistic background and motivation for semantic roles, and also about a range of computational models for this task, from early approaches to the current state-of-the-art. We will further discuss recently proposed variations to the traditional SRL task, including topics such as semantic proto-role labeling.

We also cover techniques for reducing required annotation effort, such as methods exploiting unlabeled corpora (semi-supervised and unsupervised techniques), model adaptation across languages and domains, and methods for crowdsourcing semantic role annotation (e.g., question-answer driven SRL). Methods based on different machine learning paradigms, including neural networks, generative Bayesian models, graph-based algorithms and bootstrapping style techniques.

Beyond sentence-level SRL, we discuss work that involves semantic roles in discourse. In particular, we cover data sets and models related to the task of identifying implicit roles and linking them to discourse antecedents. We introduce different approaches to this task from the literature, including models based on coreference resolution, centering, and selectional preferences. We also review how new insights gained through them can be useful for the traditional SRL task.

Organizers

Diego Marcheggiani (University of Amsterdam)   `marcheggiani@uva.nl`
Institute for Logic, Language and Computation, 1098XG Amsterdam, Netherlands
Diego Marcheggiani is a postdoctoral researcher at the University of Amsterdam. He graduated with a Ph.D. in Computer Science from the University of Venice and during this period he worked at the ISTI-CNR in Italy as a researcher. His research focus ranges from relation extraction to semantic role labeling and frame-semantic parsing. He is interested in supervised and unsupervised learning approaches in the scope of tensor factorization models and neural networks.

Michael Roth (UIUC / Saarland University)   mroth@coli.uni-sb.de
Department of Language Science and Technology, 66123 Saarbruecken, Germany
Michael Roth is a postdoctoral researcher and DFG research fellow at Saarland University and University of Illinois at Urbana-Champaign, respectively. He graduated with a Ph.D. in Computational Linguistics from Heidelberg University in 2013. His research focus lies on computational models of language that can facilitate automatic text understanding beyond the sentence level. Recent work includes neural-network based approaches to semantic role labeling and discourse-level frame-semantic parsing. His models are the current state-of-the-art on the CoNLL-2009 and FrameNet 1.5 data sets.

Ivan Titov (University of Amsterdam)   titov@uva.nl
Institute for Logic, Language and Computation, 1098XG Amsterdam, Netherlands
Ivan Titov is an Associate Professor at the University of Amsterdam. He is the recipient of an ERC Starting Grant, a personal Vidi Grant from the Dutch NSF (NWO) and a Google Focused Research Award. Ivan is an action editor for the Journal of Machine Learning Research (JMLR) and Transactions of ACL (TACL), as well as an editorial board member of the Journal of Artificial Intelligence Research (JAIR). His interests are in probabilistic modeling of language, primarily in semantics and syntax as well as in multilingual NLP and semi-supervised learning for NLP.

Benjamin Van Durme (Johns Hopkins University)   vandurme@cs.jhu.edu
Department of Computer Science, 21218 MD Baltimore, USA
Benjamin Van Durme is an Assistant Professor at the Johns Hopkins University in Computer Science, with a courtesy appointment in Cognitive Science, and the lead of the Natural Language Understanding group at the Human Language Technology Center of Excellence. (HLTCOE). His research is broadly focused on discovering and extracting knowledge from language, exploring topics such as low resource, multilingual information extraction; scalable, streaming algorithms for processing large collections; and semantic analysis at various levels of complexity.