Semantic Role Labeling

Martha Palmer[*], Ivan Titov[§], Shumin Wu[*]
[*]University of Colorado
[§]Saarland University
Martha.Palmer@colorado.edu
titovian,wushumin@gmail.com

1   Overview

This tutorial will describe semantic role labeling, the assignment of semantic roles to eventuality participants in an attempt to approximate a semantic representation of an utterance. The linguistic background and motivation for the definition of semantic roles will be presented, as well as the basic approach to semantic role annotation of large amounts of corpora. Recent extensions to this approach that encompass light verb constructions and predicative adjectives will be included, with reference to their impact on English, Arabic, Hindi and Chinese. Current proposed extensions such as Abstract Meaning Representations and richer event representations will also be touched on.

Details of machine learning approaches will be provided, beginning with fully supervised approaches that use the annotated corpora as training material. The importance of syntactic parse information and the contributions of different feature choices, including tree kernels, will be discussed, as well as the advantages and disadvantages of particular machine learning algorithms and approaches such as joint inference. Appropriate considerations for evaluation will be presented as well as successful uses of semantic role labeling in NLP applications.

We will also cover techniques for exploiting unlabeled corpora and transferring models across languages. These include methods, which project annotations across languages using parallel data, induce representations solely from unlabeled corpora (unsupervised methods) or exploit a combination of a small amount of human annotation and a large unlabeled corpus (semi-supervised techniques). We will discuss methods based on different machine learning paradigms, including generative Bayesian models, graph-based algorithms and bootstrapping style techniques.

2 Outline

I. Introduction, background and annotation

- Motivation – who did what to whom

- Linguistic Background

- Basic Annotation approach

- Recent extensions

- Language Specific issues with English, Arabic, Hindi and Chinese

- Semlink – Mapping between PropBank, VerbNet and FrameNet.

- The next step – Events and Abstract Meaning Representations

II. Supervised Machine Learning for SRL

- Identification and Classification

- Features (tree kernel, English vs. Chinese)

- Choice of ML method and feature combinations (kernel vs feature space)

- Joint Inference

- Impact of Parsing

- Evaluation

- Applications (including multi-lingual)

III. Semi-supervised and Unsupervised Approaches

- Cross-lingual annotation projection methods and direct transfer of SRL models across languages

- Semi-supervised learning methods

- Unsupervised induction

- Adding supervision and linguistic priors to unsupervised methods

3 Speaker Bios

Martha Palmer[1] is a Professor of Linguistics and Computer Science, and a Fellow of the Institute of Cognitive Science at the University of Colorado. Her current research is aimed at building domain-independent and language independent techniques for semantic interpretation based on linguistically annotated data, such as Proposition Banks. She has been the PI on NSF, NIH and DARPA projects for linguistic annotation (syntax, semantics and pragmatics) of English, Chinese, Korean, Arabic and Hindi. She has been a member of the Advisory Committee for the DARPA TIDES program, Chair of SIGLEX, Chair of SIGHAN, a past President of the Association for Computational Linguistics, and is a Co-Editor of JNLE and of LiLT and is on the CL Editorial Board. She received her Ph.D. in Artificial Intelligence from the University of Edinburgh in 1985.

Ivan Titov[2] joined the Saarland University as a junior faculty and head of a research group in November 2009, following a postdoc at the University of Illinois at Urbana-Champaign. He received his Ph.D. in Computer Science from the University of Geneva in 2008 and his master's degree in Applied Mathematics and Informatics from the St. Petersburg State Polytechnic University (Russia) in 2003. His research interests are in statistical natural language processing (models of syntax, semantics and sentiment) and machine learning (structured prediction methods, latent variable models, Bayesian methods).

Shumin Wu is a Computer Science PhD student (advised by Dr. Martha Palmer) at the University of Colorado. His current research is aimed at developing and applying semantic mapping (aligning and jointly inferring predicate-argument structures between languages) to Chinese dropped-pronoun recovery/alignment, automatic verb class induction, and other applications relevant to machine translation.

---

[1]http://verbs.colorado.edu/~mpalmer/
[2]http://people.mmci.uni-saarland.de/~titov/