

6. Automatic Semantic Role Labeling

Scott Wen-tau Yih and Kristina Toutanova, Microsoft Research

The goal of semantic role labeling is to map sentences to domain-independent semantic representations, which abstract away from syntactic structure and are important for deep NLP tasks such as question answering, textual entailment, and complex information extraction. Semantic role labeling has recently received significant interest in the natural language processing community. In this tutorial, we will first describe the problem and history of semantic role labeling, and introduce existing corpora and other related tasks. Next, we will provide a detailed survey of state-of-the-art machine learning approaches to building a semantic role labeling system. Finally, we will conclude the tutorial by discussing directions for improving semantic role labeling systems and their application to other natural language problems.

6.1 Tutorial Outline

1. Introduction
 - What is semantic role labeling?
 - Why is SRL important?
 - Existing corpora: FrameNet & PropBank
 - Corpora in development
 - Relation to other tasks
2. Survey of Existing SRL Systems
 - History of the development of automatic SRL systems
 - Pioneering Work
 - Basic architecture of a generic SRL system
 - Major components
 - Machine learning technologies
 - CoNLL-04 and CoNLL-05 shared tasks on SRL
 - Details of several CoNLL-05 systems
 - Overall comparisons of CoNLL-05 systems
3. Analysis of Systems and Future Directions
 - Error Analysis
 - Influence of parser errors
 - Per argument performance
 - Directions for improving SRL
4. Applications
 - Information Extraction
 - Textual Entailment
 - Machine Translation

6.2 Target Audience

The main target audience is NLP students and researchers who are interested in learning about semantic role labeling, but have not followed all developments in the field. Additionally, researchers already working on semantic role labeling should profit from a global view and summary of relevant work. The tutorial will also be valuable for researchers working in the related areas of information extraction and spoken language understanding.

Scott Wen-tau Yih received his PhD in Computer Science from the University of Illinois at Urbana-Champaign in 2005 and is currently a Post-Doc Researcher in the Machine Learning and Applied Statistics group at Microsoft Research. His research focuses on different problems in natural language processing and machine learning, such as information extraction and semantic parsing. Scott has published several papers on semantic role labeling in CoNLL-04&05, COLING-04 and IJCAI-05. The SRL system he built at UIUC was the best system in the CoNLL-05 shared task.

Kristina Toutanova obtained her PhD in Computer Science from Stanford University in 2005 and joined Microsoft Research as a Researcher in the Natural Language Processing group. Her areas of expertise include semantic role labeling, syntactic parsing, machine learning, and machine translation. Kristina has published two papers on semantic role labeling in CoNLL-05 and ACL-05. The SRL system she built at Stanford was the runner-up system in the CoNLL-05 shared task.