

A Unified Framework for Structured Prediction: From Theory to Practice

Wei Lu

Singapore University of Technology and Design
8 Somapah Road, Singapore 487372
luwei@sutd.edu.sg

Structured prediction is one of the most important topics in various fields, including machine learning, computer vision, natural language processing (NLP) and bioinformatics. In this tutorial, we present a novel framework that unifies various structured prediction models.

The hidden Markov model (HMM) and the probabilistic context-free grammars (PCFGs) are two classic generative models used for predicting outputs with linear-chain and tree structures, respectively. As HMM's discriminative counterpart, the linear-chain conditional random fields (CRFs) (Lafferty et al., 2001) model was later proposed. Such a model was shown to yield good performance on standard NLP tasks such as information extraction. Several extensions to such a model were then proposed afterward, including the semi-Markov CRFs (Sarawagi and Cohen, 2004), tree CRFs (Cohn and Blunsom, 2005), as well as discriminative parsing models and their latent variable variants (Petrov and Klein, 2007). On the other hand, utilizing a slightly different loss function, one could arrive at the structured support vector machines (Tsochantaridis et al., 2004) and its latent variable variant (Yu and Joachims, 2009) as well. Furthermore, new models that integrate neural networks and graphical models, such as neural CRFs (Do et al., 2010) were also proposed.

In this tutorial, we will be discussing how such a wide spectrum of existing structured prediction models can all be implemented under a unified framework¹ that involves some basic building blocks. Based on such a framework, we show how some seemingly complicated structured prediction models such as a semantic parsing model (Lu et al., 2008; Lu, 2014) can be implemented conveniently and quickly. Furthermore, we also show that the framework can be used to solve certain structured prediction problems that otherwise cannot be easily handled by conventional structured

prediction models. Specifically, we show how to use such a framework to construct models that are capable of predicting non-conventional structures, such as overlapping structures (Lu and Roth, 2015; Muis and Lu, 2016a). We will also discuss how to make use of the framework to build other related models such as topic models and highlight its potential applications in some recent popular tasks (e.g., AMR parsing (Flanigan et al., 2014)).

This tutorial consists of the following 3 main sections.²

Foundations of structured prediction models

Duration: 45 minutes

In this section, we introduce the basics of structured prediction models. We will review all the above-mentioned structured prediction models. We then provide a global picture that shows the underlying connections between different models.

Unified framework for structured prediction

Duration: 45 minutes

In this section, we formally introduce the framework that allows all such different structured prediction models to be unified in an elegant manner. We start with defining the basic building blocks required for constructing a structured prediction model. Next, we discuss how to make use of such building blocks for constructing different types of models.

Practical guide on model implementation

Duration: 60 minutes

In this section, we present a practical guide on how to implement seemingly very different types of structured prediction models using our unified framework. We will show

¹<http://statnlp.org/statnlp-framework>

²The material associated with this tutorial will be available at <http://statnlp.org/tutorials/>.

how to conveniently implement several structured prediction models within our framework using real code examples.

The framework has been extensively used by our research group for developing various structured prediction models, including models for information extraction (Lu and Roth, 2015; Muis and Lu, 2016a; Jie et al., 2017), noun phrase chunking (Muis and Lu, 2016b), semantic parsing (Lu, 2015; Susanto and Lu, 2017), and sentiment analysis (Li and Lu, 2017). It is our hope that this tutorial will be helpful for many natural language processing researchers who are interested in designing their own structured prediction models rapidly. We also hope this tutorial allows researchers to strengthen their understandings on the connections between various structured prediction models, and that the open release of the framework will bring value to the NLP research community and enhance its overall productivity.

About the Instructor

Wei Lu is an Assistant Professor at the Singapore University of Technology and Design (SUTD), directing the StatNLP research group (<http://statnlp.org/>). He received his Ph.D. from the National University of Singapore (NUS) in 2009. He visited CSAIL, Massachusetts Institute of Technology (MIT) in 2007-2008, and worked as a postdoctoral research associate at the University of Illinois at Urbana-Champaign in 2011-2013. His research interests include developing mathematical models and machine learning algorithms for solving natural language processing problems. He is particularly interested in semantic processing (in a broad sense). His papers appeared at venues such as ACL, EMNLP, NAACL, AACL, and CIKM. He served as a program committee member for conferences such as ACL, EMNLP, NAACL, EACL, AACL, IJCAI and NIPS, and is currently a member of the standing reviewer team for TACL. He served as an area co-chair for ACL 2016 and received the best paper award at EMNLP 2011.

References

Trevor Cohn and Philip Blunsom. 2005. Semantic role labelling with tree conditional random fields. In *Proc. of CONLL*, pages 169–172.

Trinh Do, Thierry Arti, et al. 2010. Neural conditional random fields. In *Proc. of AISTATS*, pages 177–184.

Jeffrey Flanigan, Sam Thomson, Jaime G Carbonell, Chris Dyer, and Noah A Smith. 2014. A discriminative graph-based parser for the abstract meaning representation. In *Proc. of ACL*, pages 1426–1436.

Zhanming Jie, Aldrian Obaja Muis, and Wei Lu. 2017. Efficient dependency-guided named entity recognition. In *Proc. of AAAI*.

John Lafferty, Andrew McCallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. of ICML*, pages 282–289.

Hao Li and Wei Lu. 2017. Learning latent sentiment scopes for entity-level sentiment analysis. In *Proc. of AAAI*.

Wei Lu and Dan Roth. 2015. Joint mention extraction and classification with mention hypergraphs. In *Proc. of EMNLP*.

Wei Lu, Hwee Tou Ng, Wee Sun Lee, and Luke S Zettlemoyer. 2008. A generative model for parsing natural language to meaning representations. In *Proc. of EMNLP*, pages 783–792.

Wei Lu. 2014. Semantic parsing with relaxed hybrid trees. In *Proc. of EMNLP*, pages 1308–1318.

Wei Lu. 2015. Constrained semantic forests for improved discriminative semantic parsing. In *Proc. of ACL-IJCNLP*, pages 737–742.

Aldrian Obaja Muis and Wei Lu. 2016a. Learning to recognize discontinuous entities. In *Proc. of EMNLP*, pages 75–84.

Aldrian Obaja Muis and Wei Lu. 2016b. Weak semi-markov crfs for noun phrase chunking in informal text. In *Proc. of NAACL*, pages 714–719.

Slav Petrov and Dan Klein. 2007. Discriminative log-linear grammars with latent variables. In *Proc. of NIPS*, pages 1153–1160.

Sunita Sarawagi and William W Cohen. 2004. Semi-markov conditional random fields for information extraction. In *Proc. of NIPS*, pages 1185–1192.

Raymond Hendy Susanto and Wei Lu. 2017. Semantic parsing with neural hybrid trees. In *Proc. of AAAI*.

Ioannis Tsochantaris, Thomas Hofmann, Thorsten Joachims, and Yasemin Altun. 2004. Support vector machine learning for interdependent and structured output spaces. In *Proc. of ICML*, page 104. ACM.

Chun-Nam John Yu and Thorsten Joachims. 2009. Learning structural svms with latent variables. In *Proc. of ICML*, pages 1169–1176. ACM.