

Incremental Structured Prediction Using a Global Learning and Beam-Search Framework

Yue Zhang[†], Meishan Zhang[‡], Ting Liu[‡]

[†]Singapore University of Technology and Design

yue_zhang@sutd.edu.sg

[‡]Research Center for Social Computing and Information Retrieval

Harbin Institute of Technology, China

{mszhang, tliu}@ir.hit.edu.cn

Abstract

This tutorial discusses a framework for incremental left-to-right structured prediction, which makes use of global discriminative learning and beam-search decoding. The method has been applied to a wide range of NLP tasks in recent years, and achieved competitive accuracies and efficiencies. We give an introduction to the algorithms and efficient implementations, and discuss their applications to a range of NLP tasks.

1 Introduction

This tutorial discusses a framework of online global discriminative learning and beam-search decoding for syntactic processing (Zhang and Clark, 2011b), which has recently been applied to a wide variety of natural language processing (NLP) tasks, including word segmentation (Zhang and Clark, 2007), dependency parsing (Zhang and Clark, 2008b; Huang and Sagae, 2010; Zhang and Nivre, 2011; Bohnet and Kuhn, 2012), context free grammar (CFG) parsing (Collins and Roark, 2004; Zhang and Clark, 2009; Zhu et al., 2013), combinational categorial grammar (CCG) parsing (Zhang and Clark, 2011a; Xu et al., 2014) and machine translation (Liu, 2013), achieving state-of-the-art accuracies and efficiencies. In addition, due to its high efficiencies, it has also been applied to a range of joint structural problems, such as joint segmentation and POS-tagging (Zhang and Clark, 2008a; Zhang and Clark, 2010), joint POS-tagging and dependency parsing (Hatori et al., 2011; Bohnet and Nivre, 2012), joint morphological analysis, POS-tagging and dependency parsing (Bohnet et al., 2013), and joint segmentation, POS-tagging and parsing (Zhang et al., 2013; Zhang et al., 2014).

In addition to the aforementioned tasks, the framework can be applied to all structural pre-

diction tasks for which the output can be constructed using an incremental process. The advantage of this framework is two-fold. First, beam-search enables highly efficient decoding, which typically has linear time complexity, depending on the incremental process. Second, free from DP-style constraints and Markov-style independence assumptions, the framework allows arbitrary features to be defined to capture structural patterns. In addition to feature advantages, the high accuracies of this framework are also enabled by direct interactions between learning and search (Daumé III and Marcu, 2005; Huang et al., 2012; Zhang and Nivre, 2012).

2 Tutorial Overview

In this tutorial, we make an introduction to the framework, illustrating how it can be applied to a range of NLP problems, giving theoretical discussions and demonstrating a software implementation. We start with a detailed introduction of the framework, describing the averaged perceptron algorithm (Collins, 2002) and its efficient implementation issues (Zhang and Clark, 2007), as well as beam-search and the early-update strategy (Collins and Roark, 2004). We then illustrate how the framework can be applied to NLP tasks, including word segmentation, joint segmentation & POS-tagging, labeled and unlabeled dependency parsing, joint POS-tagging and dependency parsing, CFG parsing, CCG parsing, and joint segmentation, POS-tagging and parsing. In each case, we illustrate how the task is turned into an incremental left-to-right output-building process, and how rich features are defined to give competitive accuracies. These examples can serve as guidance in applying the framework to other structural prediction tasks.

In the second part of the tutorial, we give some analysis on why the framework is effective. We discuss several alternative learning algorithms,

and compare beam-search with greedy search on dependency parsing. We show that accuracy benefits from interaction between learning and search. Finally, the tutorial concludes with an introduction to ZPar, an open source toolkit that provides optimized C++ implementations of all the above tasks.

3 Outline

- 1 Introduction (0.5 hours)
 - 1.1 An overview of the syntactic processing framework and its applications
 - 1.2 An introduction to the beam-search framework and comparison to dynamic programming
 - 1.3 Algorithm in details
 - 1.3.1 Online discriminative learning using the perceptron
 - 1.3.2 Beam-search decoding
 - 1.3.3 The integrated framework
- 2 Applications (1.25 hours)
 - 2.1 Overview
 - 2.2 Word segmentation
 - 2.3 Joint segmentation and POS-tagging
 - 2.4 Dependency parsing
 - 2.5 Context free grammar parsing
 - 2.6 Combinatory categorial grammar parsing
 - 2.7 Joint segmentation, POS-tagging and parsing
- 3 Analysis of the framework (0.75 hours)
 - 3.1 The influence of global learning
 - 3.2 The influence of beam-search
 - 3.3 Benefits from the combination
 - 3.4 Related discussions
- 4 The ZPar software tool (0.5 hours)

4 About the Presenters

Yue Zhang is an Assistant Professor at Singapore University of Technology and Design (SUTD). Before joining SUTD in 2012, he worked as a postdoctoral research associate at University of Cambridge. He received his PhD and MSc degrees from University of Oxford, and undergraduate degree from Tsinghua University, China. Dr Zhang's research interest includes natural language parsing, natural language generation, machine translation and machine learning.

Meishan Zhang is a fifth-year Phd candidate at Research Center for Social Computing and Information Retrieval, Harbin Institute of Technology, China (HIT-SCIR). His research interest includes Chinese morphological and syntactic parsing, semantic representation and parsing, joint modelling and machine learning.

Ting Liu is a professor at HIT-SCIR. His research interest includes social computing, information retrieval and natural language processing.

References

- Bernd Bohnet and Jonas Kuhn. 2012. The best of both worlds – a graph-based completion model for transition-based parsers. In *Proceedings of EACL*, pages 77–87, Avignon, France, April. Association for Computational Linguistics.
- Bernd Bohnet and Joakim Nivre. 2012. A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing. In *Proceedings of EMNLP*, pages 1455–1465, Jeju Island, Korea, July. Association for Computational Linguistics.
- Bernd Bohnet, Joakim Nivre, Igor Boguslavsky, Richard Farkas, Filip Ginter, and Jan Hajic. 2013. Joint morphological and syntactic analysis for richly inflected languages. *Transactions of the Association for Computational Linguistics*, 1:415–428.
- Michael Collins and Brian Roark. 2004. Incremental parsing with the perceptron algorithm. In *Proceedings of ACL 2004, Main Volume*, pages 111–118, Barcelona, Spain, July.
- Michael Collins. 2002. Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms. In *Proceedings of EMNLP*, pages 1–8. Association for Computational Linguistics, July.
- Hal Daumé III and Daniel Marcu. 2005. Learning as search optimization: Approximate large margin methods for structured prediction. In *International Conference on Machine Learning (ICML)*, Bonn, Germany.
- Jun Hatori, Takuya Matsuzaki, Yusuke Miyao, and Jun'ichi Tsujii. 2011. Incremental joint pos tagging and dependency parsing in chinese. In *Proceedings of IJCNLP*, pages 1216–1224, Chiang Mai, Thailand, November. Asian Federation of Natural Language Processing.
- Liang Huang and Kenji Sagae. 2010. Dynamic programming for linear-time incremental parsing. In *Proceedings of ACL 2010*, pages 1077–1086, Uppsala, Sweden, July. Association for Computational Linguistics.

- Liang Huang, Suphan Fayong, and Yang Guo. 2012. Structured perceptron with inexact search. In *Proceedings of NAACL 2012*, pages 142–151, Montréal, Canada, June. Association for Computational Linguistics.
- Yang Liu. 2013. A shift-reduce parsing algorithm for phrase-based string-to-dependency translation. In *Proceedings of the ACL*, pages 1–10, Sofia, Bulgaria, August. Association for Computational Linguistics.
- Wenduan Xu, Yue Zhang, and Stephen Clark. 2014. Shift-reduce ccg parsing with a dependency model. In *Proceedings of the ACL*.
- Yue Zhang and Stephen Clark. 2007. Chinese segmentation with a word-based perceptron algorithm. In *Proceedings of ACL 2007*, pages 840–847, Prague, Czech Republic, June. Association for Computational Linguistics.
- Yue Zhang and Stephen Clark. 2008a. Joint word segmentation and POS tagging using a single perceptron. In *Proceedings of ACL-08: HLT*, pages 888–896, Columbus, Ohio, June. Association for Computational Linguistics.
- Yue Zhang and Stephen Clark. 2008b. A tale of two parsers: Investigating and combining graph-based and transition-based dependency parsing. In *Proceedings of EMNLP*, pages 562–571, Honolulu, Hawaii, October. Association for Computational Linguistics.
- Yue Zhang and Stephen Clark. 2009. Transition-based parsing of the chinese treebank using a global discriminative model. In *Proceedings of IWPT'09*, pages 162–171, Paris, France, October. Association for Computational Linguistics.
- Yue Zhang and Stephen Clark. 2010. A fast decoder for joint word segmentation and POS-tagging using a single discriminative model. In *Proceedings of EMNLP 2010*, pages 843–852, Cambridge, MA, October. Association for Computational Linguistics.
- Yue Zhang and Stephen Clark. 2011a. Shift-reduce ccg parsing. In *Proceedings of ACL 2011*, pages 683–692, Portland, Oregon, USA, June. Association for Computational Linguistics.
- Yue Zhang and Stephen Clark. 2011b. Syntactic processing using the generalized perceptron and beam search. *Computational Linguistics*, 37(1):105–151.
- Yue Zhang and Joakim Nivre. 2011. Transition-based dependency parsing with rich non-local features. In *Proceedings of ACL 2011*, pages 188–193, Portland, Oregon, USA, June. Association for Computational Linguistics.
- Yue Zhang and Joakim Nivre. 2012. Analyzing the effect of global learning and beam-search on transition-based dependency parsing. In *Proceedings of COLING 2012: Posters*, pages 1391–1406, Mumbai, India, December. The COLING 2012 Organizing Committee.
- Meishan Zhang, Yue Zhang, Wanxiang Che, and Ting Liu. 2013. Chinese parsing exploiting characters. In *Proceedings of ACL 2013*.
- Meishan Zhang, Yue Zhang, Wanxiang Che, and Ting Liu. 2014. Character-level chinese dependency parsing. In *Proceedings of the ACL*.
- Muhua Zhu, Yue Zhang, Wenliang Chen, Min Zhang, and Jingbo Zhu. 2013. Fast and accurate shift-reduce constituent parsing. In *Proceedings of ACL 2013*.