Scalable Large-Margin Structured Learning:
Theory and Algorithms

Liang Huang   **Kai Zhao**

The City University of New York

{liang.huang.sh, kzhao.hf, lemaoliu}@gmail.com

1 Motivations

Much of NLP tries to map structured input (sentences) to some form of structured output (tag sequences, parse trees, semantic graphs, or translated/paraphrased/compressed sentences). Thus structured prediction and its learning algorithm are of central importance to us NLP researchers. However, when applying machine learning to structured domains, we often face scalability issues for two reasons:

1. Even the fastest exact search algorithms for most NLP problems (such as parsing and translation) is too slow for repeated use on the training data, but approximate search (such as beam search) unfortunately breaks down the nice theoretical properties (such as convergence) of existing machine learning algorithms.

2. Even with inexact search, the scale of the training data in NLP still makes pure online learning (such as perceptron and MIRA) too slow on a single CPU.

This tutorial reviews recent advances that address these two challenges. In particular, we will cover principled machine learning methods that are designed to work under vastly inexact search, and parallelization algorithms that speed up learning on multiple CPUs. We will also extend structured learning to the latent variable setting, where in many NLP applications such as translation and semantic parsing the gold-standard derivation is hidden.

2 Contents

1. Overview of Structured Learning

    (a) key challenge 1: search efficiency

    (b) key challenge 2: interactions between search and learning

2. Structured Perceptron

    (a) the basic algorithm

    (b) convergence proof – a purely geometric approach   **(updated in 2015)**

    (c) voted and averaged perceptrons, and efficient implementation tricks

    (d) applications in tagging, parsing, etc.

    (e) inseparability and generalization bounds   **(new in 2015)**

3. Structured Perceptron under Inexact Search

    (a) convergence theory breaks under inexact search

    (b) early update

    (c) violation-fixing perceptron

    (d) applications in tagging, parsing, etc.

   —coffee break—

4. Large-Margin Structured Learning with Latent Variables

    (a) examples: machine translation, semantic parsing, transliteration

    (b) separability condition and convergence proof   **(updated in 2015)**

    (c) latent-variable perceptron under inexact search

    (d) applications in machine translation

5. Parallelizing Large-Margin Structured Learning

    (a) iterative parameter mixing (IPM)

    (b) minibatch perceptron and MIRA

3   Instructor Biographies

Liang Huang is an Assistant Professor at the City University of New York (CUNY). He received his Ph.D. in 2008 from Penn and has worked as a Research Scientist at Google and a Research Assistant Professor at USC/ISI. His work is mainly on the theoretical aspects (algorithms and formalisms) of computational linguistics, as well as theory and algorithms of structured learning. He has received a Best Paper Award at ACL 2008, several best paper nominations (ACL 2007, EMNLP 2008, and ACL 2010), two Google Faculty Research Awards (2010 and 2013), and a University Graduate Teaching Prize at Penn (2005). He has given three tutorials at COLING 2008, NAACL 2009 and ACL 2014.

Kai Zhao is a Ph.D. candidate at the City University of New York (CUNY), working with Liang Huang. He received his B.S. from the University of Science and Technology in China (USTC). He has published on structured prediction, online learning, machine translation, and parsing algorithms. He was a summer intern with IBM TJ Watson Research Center in 2013, Microsoft Research Redmond in 2014, and Google Research NYC in 2015.