

Learning and Inference in Structured Prediction Models

AAAI 2016 Tutorial Syllabus

Kai-Wei Chang¹, Gourab Kundu², Dan Roth³, Vivek Srikumar⁴

¹ Microsoft Research, MA

² IBM Research, NY

³ Department of Computer Science, University of Illinois, Urbana, IL

⁴ School of Computing, University of Utah, Salt Lake City, UT

Goal of the Tutorial

This tutorial targets AI researchers who are interested in designing and applying structured prediction models to problems with interdependent output variables. The tutorial will introduce the problem of structured prediction exemplifying it in multiple AI problems, and then cover recent developments in efficient inference and learning methods in discriminative structured models and outline further research directions in this topic.

Tutorial Description

Many prediction problems required structured decisions. That is, the goal is to assign values to multiple interdependent variables. The relationships between the output variables could represent a sequence, a set of clusters, or in the general case, a graph. When solving these problems, it is important to make consistent decisions that take the interdependencies among output variables into account. Such problems are often referred to as structured prediction problems. In past decades, multiple structured prediction models have been proposed and studied and success has been demonstrated in a range of applications, including natural language processing, information extraction, computer vision and computational biology. However, the high computational cost often limits both models' expressive power and the size of the data that can be handled. Therefore, designing efficient inference and learning algorithms for these models is a key challenge for structured prediction.

In this tutorial, we will focus on recent developments in discriminative structured prediction models such as Structured SVMs and Structured Perceptron. Beyond introducing the algorithmic approaches in this domain, we will discuss ideas that result in significant improvements both in the learning and in the inference stages of these algorithms. In particular, we will discuss the use of caching techniques to reuse computations and methods for decomposing complex structures, along with learning procedures that make use of it to simplify the learning stage. We will also present a recently proposed formulation that captures similarities between structured labels by using distributed representation. Participants will learn about existing trends in learning and the inference for the structured prediction models, recent tools developed in this area, and how they can be applied to AI applications.

Tutorial Outline

This tutorial will systematically introduce the theoretical and the practical aspects of structured prediction models. We begin by motivating the need for solving structured prediction problems. Then, we provide an overview of existing models, and discuss recent advances in inference and learning methods for structured prediction. On the inference side, we focus on integer linear programming formulations of the inference, discuss its generality, and discuss the recently introduced amortized inference method that allows re-using computations from previous inference instances to speed up inference on new instances. On the learning side, we focus on an efficient learning algorithm that is based on dual coordinate descent (DCD). We then combine the learning and inference ideas and show how amortized inference techniques can be incorporated in the learning stage to reduce training time. Finally, we introduce an advanced structured learning formulation, where labels are represented using dense vectors. This distributed representation allows semantically similar labels to share parameters, resulting in better performance. We conclude the tutorial by demonstrating an open-source library, IllinoisSL, that includes many of the above-mentioned techniques, and facilitates applying structured prediction models to the participants' own problems easily. The tutorial is designed to take 3.5 hours. The detailed outline is given below.

Motivation [20 min]

We will present several AI applications with structured outputs to motivate the need for structured prediction models.

Discriminative Structured Prediction Models [40 min]

First, we will present Constrained Conditional Models (CCM) [3], a framework that is used to model interdependencies between output variables using constraints and features. We then discuss how to formulate an inference problem as an Integer Linear Programming [7, 8]. We will also describe several paradigms to learn the parameters of a CCM. Necessary background about structured prediction will be provided in this section.

Amortized Inference for Structured Prediction Models [30 min]

We will describe a recently developed technique, amortized inference, for speeding up inference for structured prediction models by caching previous inference samples [9]. We will also discuss how to further improve the amortized inference techniques by incorporating a dual decomposition approach which decomposes the output structure and makes use of Lagrangian relaxation methods [6].

Efficient Learning for Structured Prediction Models [30 min]

We will first describe several structured learning approaches such as Structured SVMs [11, 12] and Structured Perceptron [5]. We then describe efficient learning algorithms for Structured SVMs based on a dual coordinate descent method [4, 1]. Finally, we will present methods that make use of amortized inference techniques [2].

Distributed Representation for Structured Prediction [30 min]

We will present a recently proposed structured learning formulation, *Distro* [10], which models meaning of labels using real valued vectors. We will also describe inference and learning algorithms in this model.

Structured Prediction Software [30 min]

We will introduce IllinoisSL – a Java based discriminative structured learning library. We will use the problem of part-of-speech tagging as a running example and demonstrate how to implement a sequential tagging model using the library.

Conclusion and Future Research Directions [30 min]

Structured prediction models are widely used in AI. Therefore, designing efficient learning and inference algorithms for them could have a great impact. We will conclude the tutorial by presenting some challenges and potential research topics in designing and applying structured prediction models.

Prerequisite Knowledge

No specific background knowledge is assumed of the audience.

Tutorial Instructors

Kai-Wei Chang

Microsoft Research New England, MA
One Memorial Drive, Cambridge, MA 02142
Email: kw@kwchang.net

Kai-Wei Chang is a post-doctoral researcher at Microsoft Research. He will be joining the Department of Computer Science at the University of Virginia as Assistant Professor in the Fall 2016. His research interests are in designing practical machine learning techniques for large and complex data, and applying them to real-world applications. He has been working on various topics in Machine Learning and Natural Language Processing, including large-scale classification, structured learning, co-reference resolution, and relation extraction. He

has been also involved in developing machine learning packages such as LIBLINEAR, Vowpal Wabbit, and Illinois-SL. He was awarded the KDD Best Paper Award in 2010 and won the Yahoo! Key Scientific Challenges Award in 2011.

Gourab Kundu

IBM Research

1101 Route 134 Kitchawan Rd, Yorktown Heights, NY 10598

Email: gkundu@us.ibm.com

Gourab Kundu is a research staff member at IBM research. He is broadly interested in all aspects of machine learning and natural language processing. He has publications in top tier machine learning and natural language processing conferences along with a best student paper in CoNLL 2011.

Dan Roth

Department of Computer Science, University of Illinois at Urbana-Champaign

201 N Goodwin Ave, MC-258, Urbana, IL 61801

Email: danr@illinois.edu

Phone: +1-(217)-244-7068

Dan Roth is a Professor in the Department of Computer Science and the Beckman Institute at the University of Illinois at Urbana-Champaign and a University of Illinois Scholar. Roth is a Fellow of the AAAS, ACM, AAAI and ACL, for his contributions to Machine Learning and to Natural Language Processing. He has published broadly in machine learning, natural language processing, knowledge representation and reasoning, and learning theory, and has developed advanced machine learning based tools for natural language applications that are being used widely by the research community and commercially. Roth is the Editor-in-Chief of the Journal of Artificial Intelligence Research (JAIR) and has served on the editorial board of several of the major journals in his research areas. He was the program chair of AAAI'11, ACL'03 and CoNLL'02 and serves regularly as an area chair and senior program committee member in the major conferences in his research areas. Prof. Roth received his B.A Summa cum laude in Mathematics from the Technion, Israel, and his Ph.D in Computer Science from Harvard University in 1995.

Vivek Srikumar

School of Computing, University of Utah
50 Central Campus Dr, Salt Lake City, UT 84112
Email: svivek@cs.utah.edu

Vivek Srikumar is an assistant professor in the School of Computing at the University of Utah. His research interests are in the areas of machine learning and natural learning processing in the context of structured learning and prediction. His research has primarily been driven by questions arising from the need to learn structured representations of text using little or indirect supervision and to scale inference to large problems. His work has been published in various AI, NLP and machine learning venues and recently received the best paper award at EMNLP. Previously, he obtained his Ph.D. from the University of Illinois at Urbana-Champaign in 2013 and was a post-doctoral scholar at Stanford University.

References

- [1] K.-W. Chang, V. Srikumar, and D. Roth. Multi-core structural SVM training. In *ECML*, 2013.
- [2] K.-W. Chang, S. Upadhyay, G. Kundu, and D. Roth. Structural learning with amortized inference. In *AAAI*, 2015.
- [3] M. Chang, L. Ratinov, and D. Roth. Structured learning with constrained conditional models. *Machine Learning*, 2012.
- [4] M. Chang and W. Yih. Dual coordinate descent algorithms for efficient large margin structural learning. *Transactions of the Association for Computational Linguistics*, 2013.
- [5] M. Collins. Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms. In *EMNLP*, 2002.
- [6] G. Kundu, V. Srikumar, and D. Roth. Margin-based decomposed amortized inference. In *ACL*, 2013.
- [7] D. Roth and W. Yih. A linear programming formulation for global inference in natural language tasks. In Hwee Tou Ng and Ellen Riloff, editors, *CoNLL*, 2004.
- [8] D. Roth and W. Yih. Global inference for entity and relation identification via a linear programming formulation. In Lise Getoor and Ben Taskar, editors, *Introduction to Statistical Relational Learning*, 2007.
- [9] V. Srikumar, G. Kundu, and D. Roth. On amortizing inference cost for structured prediction. In *EMNLP*, 2012.

- [10] V. Srikumar and C. D Manning. Learning distributed representations for structured output prediction. In *NIPS*, pages 3266–3274. 2014.
- [11] B. Taskar, V. Chatalbashev, D. Koller, and C. Guestrin. Learning structured prediction models: a large margin approach. In *ICML*, 2005.
- [12] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research*, 2005.