

# T8: Predicting Structures in NLP: Constrained Conditional Models and Integer Linear Programming NLP

---

Dan Goldwasser, Vivek Srikumar, Dan Roth

## ABSTRACT

Making decisions in natural language processing problems often involves assigning values to sets of interdependent variables where the expressive dependency structure can influence, or even dictate what assignments are possible. This setting includes a broad range of structured prediction problems such as semantic role labeling, named entity and relation recognition, co-reference resolution, dependency parsing and semantic parsing. The setting is also appropriate for cases that may require making global decisions that involve multiple components, possibly pre-designed or pre-learned, as in summarization, paraphrasing, textual entailment and question answering. In all these cases, it is natural to formulate the decision problem as a constrained optimization problem, with an objective function that is composed of learned models, subject to domain or problem specific constraints.

Constrained Conditional Models (CCM) formulation of NLP problems (also known as: Integer Linear Programming for NLP) is a learning and inference framework that augments the learning of conditional (probabilistic or discriminative) models with declarative constraints (written, for example, using a first-order representation). The key advantage of the CCM formulation is its support for making decisions in an expressive output space while maintaining modularity and tractability of training and inference. In most applications of this framework in NLP, following [Roth & Yih, CoNLL'04], integer linear programming (ILP) has been used as the inference framework, although other algorithms can be used.

This framework has attracted much attention within the NLP community over the last few years, with multiple papers in all the recent major conferences. Formulating structured prediction as a constrained optimization problem over the output of learned models has several advantages. It allows the incorporation of problem specific global constraints using a first order language “ thus freeing the developer from (much of the) low level feature engineering “ and guarantees exact inference. Importantly, it provides also the freedom of decoupling model generation (learning) from the constrained inference stage, often simplifying the learning stage as well as the engineering aspect of building an NLP system, while improving the quality of the

solutions. These advantages and the availability of off-the-shelf solvers have led to a large variety of NLP tasks being formulated within it, including semantic role labeling, syntactic parsing, co-reference resolution, summarization, transliteration and joint information extraction.

The goal of this tutorial is to introduce the framework of Constrained Conditional Models to the broader ACL community, motivate it as a generic framework for structured inference in global NLP decision problems, present some of the key theoretical and practical issues involved in using CCMs and survey some of the existing applications of it as a way to promote further development of the framework and additional applications. The tutorial will be useful for senior and junior researchers who are interested in structured prediction and global decision problems in NLP, providing a concise overview of recent perspectives and research results.

## **OUTLINE**

After briefly motivating and introducing the general framework, the main part of the tutorial is a methodological presentation of some of the key computational issues studied within CCMs that we will present by looking at case studies published in the NLP literature. In the last part of the tutorial, we will discuss engineering issues that arise in using CCMs.

### **1. Motivation**

We will begin by introducing structured prediction with various NLP examples. We will motivate the framework of Constrained Conditional Models using examples from sequential inference, sentence compression and semantic role labeling. [30 min]

### **2. Examples of Existing CCM Applications**

We will present several applications that use CCMs – including co-reference resolution, sentence compression and information extraction and use these to explain several of the key advantages the framework offers. In this context, we will discuss several ways in which expressive constraints can be introduced into an application. [30 min]

### **3. Modeling CCMs: Inference methods and Constraints**

We will present and discuss several possibilities for modeling inference in CCMs. We will discuss ways to model problems as structured prediction problems and

the use of hard and soft constraints to represent prior knowledge. We will look at ways in which Constrained Conditional Models can be used to augment probabilistic models with declarative knowledge based constraints and how these support expressive global decisions. We will also mention various possibilities for performing the inference, from commercial Integer Linear Programming packages to search techniques to Lagrangian relaxation approximation methods. [30 min]

#### **4. Training Paradigms**

The objective function used by CCMs can be decomposed and learned in several ways, ranging from “standard” structured learning, i.e., a complete joint training of the model along with the constraints, to a complete decoupling between the learning and the inference stage. We will present the advantages and disadvantages offered by different training paradigms and provide theoretical and experimental understanding. In this part, we will also compare CCMs to other approaches studied in the literature. [30 min]

#### **5. Beyond Supervised Learning**

In the standard supervised setting, we require a corpus that is annotated with the structures of interest, which is expensive and often impractical. In this part of the tutorial, we will show how to go beyond the supervised setting using CCMs, which provide ways to use declarative constraints to guide supervised and semi-supervised training in the presence of partial and indirect supervision. We will also present recent advances in training structured predictors indirectly using a data for a companion binary task. We will review several successful applications of these methods in diverse tasks such as information extraction and textual entailment. [30 min]

#### **6. Developing CCMs Applications**

We will present a “cookbook approach” for developing applications within the CCM framework and discuss templates for possible applications. [30 min]

## **BIOS**

### **Dan Goldwasser**

Computer Science Department, University of Illinois at Urbana-Champaign, IL, 61801

Email: goldwas1--AT--illinois.edu

Phone: +(217) 333-2584

Dan Goldwasser is a PhD candidate in the Department of Computer Science at the University of Illinois at Urbana-Champaign. He is the recipient of the 2010 C.L. and Jane Liu Award for research promise. He published several papers in natural language processing, machine learning and semantic interpretation. His research work studies the role that an external context, such as the real world or a simulated world, plays in semantic interpretation and learning protocols.

### **Vivek Srikumar**

Computer Science Department, University of Illinois at Urbana-Champaign, IL. 61801

Email: vsrikum2--AT--illinois.edu

Phone: +(217) 333-2584

Vivek Srikumar is a Ph.D. candidate at the University of Illinois, Urbana-Champaign. He has worked on Machine Learning in the context of Natural Language Processing and has published papers in several conferences. His research deals with applying structure learning and prediction formalized as Constrained Conditional Models to semantic role labeling and its extensions.

### **Dan Roth**

Computer Science Department, University of Illinois at Urbana-Champaign, IL, 61801

Email: danr--AT--cs.uiuc.edu

Phone: +(217) 244-7068

Dan Roth is a Professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign and the Beckman Institute of Advanced Science and Technology (UIUC) and a University of Illinois Scholar. He is a fellow of AACL and the ACM. Roth has published broadly in machine learning, natural language processing, knowledge representation and reasoning and received several paper, teaching and research awards. He has developed several machine learning based natural language processing systems that are widely used in the computational linguistics community and in industry and has presented invited talks and tutorials in several major conferences. Dan Roth has written the first paper on formulating global NLP decisions as ILP problems with his student Scott Yih, presented in CoNLL'04, and since then has worked on further developing Constrained Conditional Models, on learning and inference issues within this framework and on applying it to several NLP problems, including Semantic Role Labeling, Information and Relation Extraction and Transliteration.