Integer Linear Programming in NLP - Constrained Conditional Models

Ming-Wei Chang, Nicholas Rizzolo, Dan Roth
University of Illinois at Urbana-Champaign

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. Structured learning problems such as semantic role labeling provide one such example, but the setting is broader and includes a range of problems such as name entity and relation recognition and co-reference resolution. The setting is also appropriate for cases that may require a solution to make use of multiple (possible pre-designed or pre-learned components) as in summarization, 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 (aka Integer Linear Programming formulation of NLP problems) 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) as a way to support 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) was used as the inference framework, although other algorithms can be used for that purpose.

This framework, with and without Integer Linear Programming as its inference engine, has recently attracted much attention within the NLP community, with multiple papers in all the recent major conferences, and a related workshop in NAACL'09. Formulating problems as constrained optimization problems over the output of learned models has several advantages. It allows one to focus on the modeling of problems by providing the opportunity to incorporate problem specific global constraints using a first order language – thus frees the developer from (much of the) low level feature engineering – and it guarantees exact inference. It provides also the freedom of decoupling the stage of model generation (learning) from that of the constrained inference stage, often resulting in simplifying the learning stage as well as the engineering problem 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 natural language processing tasks being formulated within framework, including semantic role labeling, syntactic parsing, coreference resolution, summarization, transliteration and joint information extraction.

The goal of this tutorial is to introduce the framework of Constrained Conditional Models (CCMs) to the broader ACL community, motivate it as a generic framework for learning and 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 thus be useful for many of the senior and junior researchers that have interest in global decision problems in NLP, providing a concise overview of recent perspectives and research results.

Tutorial Outline

After shortly 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 and present some tool that facilitate developing CCM models.

1. Motivation and Task Definition [30 min]

We will motivate the framework of Constrained Conditional Models and exemplify it using the example of Semantic Role Labeling.

2. Examples of Existing Applications [30 min]

We will present in details several applications that made use of CCMs – including coreference resolution, sentence compression and information extraction and use these to explain several of the key advantages the framework offers. We will discuss in this context several ways in which constraints can be introduced to an application.

3. Training Paradigms [30 min]

The objective function used by CCMs can be decomposed and learned in several ways, ranging from 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 discuss comparison to other approaches studied in the literature.

4. Inference methods and Constraints [30 min]

We will present and discuss several possibilities for modeling inference in CCMs, from Integer Linear Programming to search techniques. We will also discuss the use of hard constraints and soft constraints and present ways for modeling constraints.

5. Introducing background knowledge via CCMs [30 min]

We will look at ways in which Constrained Conditional Models (CCMs)can be used to augment probabilistic models with declarative constraints in order to support decisions

in an expressive output space, and how declarative constraints can be used to aid supervised and semi-supervised training.

6. Developing CCMs Applications [30 min]

We present a modeling language that facilitates developing applications within the CCM framework and present some "templates" for possible applications.

Tutorial Instructors

Ming-Wei Chang

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

Email: mchang21@uiuc.edu

Ming-Wei Chang is a Phd candidate in University of Illinois at Urbana-Champaign.

He has done work on Machine Learning in Natural Language Processing and Information Extraction and has published a number of papers in several international conferences including "Learning and Inference with Constraints" (AAAI'08), "Guiding Semi-Supervision with Constraint-Driven Learning" (ACL'07) and "Unsupervised Constraint Driven Learning For Transliteration Discovery. (NAACL'09). He co-presented a tutorial on CCMs in EACL'09.

Nicholas Rizzolo

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

Email: ratinov2@uiuc.edu

Nicholas Rizollo is a Phd candidate in University of Illinois at Urbana-Champaign.

He has done work on Machine Learning in Natural Language Processing and is the principal developer of Learning Based Java (LBJ) a modeling language for Constrained Conditional Models. He has published a number of papers on these topics, including "Learning and Inference with Constraints" (AAAI'08) and "Modeling Discriminative Global Inference" (ICSC'07)

Dan Roth

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

Phone: +(217) 244-7068; Email: danr@cs.uiuc.edu

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 Willett Faculty Scholar of the College of Engineering. He has published broadly in machine learning, natural language processing, knowledge representation and reasoning and received several best paper and research awards. He has developed several machine learning based natural language processing systems including an award winning semantic parser, and has presented invited talks in several international conferences, and several tutorials on machine learning for NLP. Dan Roth has written the first paper on Constrained Conditional Models along with his student Scott Yih, presented in CoNLL'04, and since then has worked on learning and inference issue within this framework as well as on applying it for several NLP problems, including Semantic Role Labeling, Information Extraction and Transliteration. He has presented several invited talks that have addresses aspect of this model.

Bibliography

Dan Roth and Wen-tau Yih. A Linear Programming Formulation for Global Inference in Natural Language Tasks. In Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004), pages 1-8, 2004.

Vasin Punyakanok, Dan Roth, Wen-tau Yih, and Dav Zimak. Semantic Role Labeling Via Integer Linear Programming Inference. In Proceedings of the International Conference on Computational Linguistics (COLING-2004), pages 1346-1352, 2004.)

Tomacz Marciniak and Michael Strube. Beyond the Pipeline: Discrete Optimization in NLP. In Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005), pages 136-145, 2005.

Tzong-Han Tsai, Chia-Wei Wu, Yu-Chun Lin, and Wen-Lian Hsu. Exploiting Full Parsing Information to Label Semantic Roles Using an Ensemble of ME and SVM via Integer Linear Programming. In Proceedings of the Ninth Conference on Computational Natural Language Learning: Shared Task (CoNLL-2005) Shared Task, pages 233-236, 2005.

Vasin Punyakanok, Dan Roth and Wen-tau Yih. The Necessity of Syntactic Parsing for Semantic Role Labeling. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-2005), pages 1117-1123, 2005.

Vasin Punyakanok, Dan Roth, Wen-tau Yih and Dav Zimak. Learning and Inference over Constrained Output. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-2005), pages 1124-1129, 2005.

Dan Roth and Wen-tau Yih. Integer Linear Programming Inference for Conditional Random Fields. In Proceedings of the International Conference on Machine Learning (ICML-2005), pages 737-744, 2005.

Regina Barzilay and Mirella Lapata. Aggregation via Set Partitioning for Natural Language Generation. In Proceedings of the Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL-2006), pages 359-366, 2006.

James Clarke and Mirella Lapata. Constraint-Based Sentence Compression: An Integer Programming Approach. In Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions (ACL-2006), pages 144-151, 2006.

Sebastian Riedel and James Clarke. Incremental Integer Linear Programming for Non-projective Dependency Parsing. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP-2006), pages 129-137, 2006.

Philip Bramsen, Pawan Deshpande, Yoong Keok Lee, and Regina Barzilay. Inducing Temporal Graphs. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP-2006), 189-198, 2006.

Yejin Choi, Eric Breck, and Claire Cardie. Joint Extraction of Entities and Relations for Opinion Recognition. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP-2006), 431-439, 2006.

Manfred Klenner. Grammatical Role Labeling with Integer Linear Programming. In Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, Conference Companion (EACL-2006), pages 187-190, 2006.

Pascal Denis and Jason Baldridge. Joint Determination of Anaphoricity and Coreference Resolution using Integer Programming. In Proceedings of the Annual Meeting of the North American Chapter of the Association for Computational Linguistics - Human Language Technology Conference (NAACL-HLT-2007), pages 236-243, 2007.

James Clarke and Mirella Lapata. Modelling Compression with Discourse Constraints. In Proceedings of the Conference on Empirical Methods in Natural Language Processing and on Computational Natural Language Learning (EMNLP-CoNLL-2007), pages 1-11, 2007.

Manfred Klenner. Enforcing Consistency on Coreference Sets. In Recent Advances in Natural Language Processing (RANLP), pages 323-328, 2007

Dan Roth and Wen-tau Yih. Global Inference for Entity and Relation Identification via a Linear Programming Formulation. Introduction to Statistical Relational Learning, 2007.

K. Ganchev, João Graça and B. Taskar. Expectation Maximization and Posterior Constraints, Neural Information Processing Systems Conference (NIPS), Vancouver, BC, December 2007.

James Clarke and Mirella Lapata. Global Inference for Sentence Compression: An Integer Linear Programming Approach. Journal of Artificial Intelligence Research (JAIR), 31, pages 399-429, 2008.

Vasin Punyakanok, Dan Roth and Wen-tau Yih. The Importance of Syntactic Parsing and Inference in Semantic Role Labeling. Computational Linguistics 34(2), pages 257-287, 2008.

Jenny Rose Finkel and Christopher D. Manning. Enforcing Transitivity in Coreference Resolution. In Proceedings of the Annual Meeting of the Association for Computational Linguistics - Human Language Technology Conference, Short Papers (ACL-HLT-2008), pages 45-48, 2008.

K. Ganchev, João Graça and B. Taskar. Better Alignments = Better Translations?, Association for Computational Linguistics (ACL), Columbus, Ohio, June 2008.

Hal Daumé. Cross-Task Knowledge-Constrained Self Training In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP-2008).

Dan Goldwasser and Dan Roth. Transliteration as Constrained Optimization. In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP-2008), pages 353-362, 2008.