June 2011

Dan Roth
Research Contributions

- **Research Summary (somewhat outdated) can be found in http://L2R.cs.uiuc.edu/Rsummary.pdf**
- **Current Research Projects can be found in http://cogcomp.cs.illinois.edu/page/projects**

My research focuses on the computational foundations of intelligent behavior. We develop theories and systems pertaining to intelligent behavior using a unified methodology – at the heart of which is the idea that learning has a central role in intelligence. My work centers around the study of machine learning and inference methods to facilitate **natural language understanding**. In doing that I have pursued several interrelated lines of work that span multiple aspects of this problem - from fundamental questions in **learning and inference** and how they interact, to the study of a range of natural language processing (NLP) problems, including context sensitive disambiguation problems, semantic role labeling, named entities recognition, co-reference, question answering and textual entailment, to large scale Natural Language Processing and Information Extraction system development - resulting in a number of software packages and NLP tools that are now widely used by the community. Below is a brief summary of some of my key research contributions.

Global Inference in Natural Language Processing: Most decisions made while interpreting natural language involve assigning values to multiples interrelated variables. We have introduced a constrained optimization framework, Constrained Conditional Models, which is mostly known in the NLP community as **Integer Learning Programming (ILP) for NLP** to support incorporating declarative knowledge into statistical models for making global decisions [Roth& Yih, 2004, 2007]. Along with it we have proposed a Constrained Driven Learning framework that allows learning simple models while making decisions with more expressive ones, by biasing the learning of the model towards solutions that better satisfy our expectations but cannot be expressed in the model family [Chang et. al 2007, 2009]. This line of work has been extremely influential and has been followed up by dozens of papers in the NLP community, including multiple best papers awards in top conferences on this topic.

Joint Learning and Inference: My work on integrating Learning and Inference goes back to my Ph.D. thesis, where I have developed the *Learning to Reason (L2R)* framework, a theoretical framework that exhibits the advantage of studying learning and reasoning jointly via a common knowledge representation [Khardon, Roth, 94,97]. Over the last few years we have developed an integrated learning and inference approach in the context of NLP, studying **Learning and Inference over structured and constrained output** [Roth, Yih 04,05,07, Punyakanok et. eal 05, 08] which views inference as a **Constrained Optimization problem**. Along with it we developed theoretical understanding for when to jointly learn and when to decouple learning from joint inference. This inference framework also facilitated recent developments of a new joint learning protocol that enables learning structured representation with minimal indirect supervision [Chang et. al 10].

Machine Learning in Natural Language Processing: We have addressed several key problems in natural language from a unified point of view and developed a theoretical understanding of several fundamental issues and an experimental paradigm that builds on this view. Our key contributions in this area are three fold: **(1)** Our theoretical work in this area has contributed to developing better understanding of the relations between probabilistic models of classification and discriminative models and we have explained the success of generative models in this area via the learning theoretical notion of Structural Risk Minimization [Roth'99] . **(2)** Our work has established the ubiquity of linear classifiers and has shown that popular models, including naïve Bayes and Hidden Markov Models, are linear models and can be studied and understood this way [Roth'98, Roth 99]. These studies have had a significant impact on NLP research, which goes beyond explaining the generalization and robustness properties of widely used methods and, in particular, they provide insight into possible extensions of these methods to structure learning. **(3)** Along with studying theories, models and algorithms, we have developed a number of mature tools that were made available to the research community and have been downloaded and used by thousands of researchers, in classes and in industry. In addition to our basic machine learning package, SNoW, and a feature extraction language, FEX, we have made available a collection of state-of-the art NLP software packages. Our on-line demos were also made available and are being frequently used, including in NLP and Computational Linguistics classes. Demos, tools, software packages and data are all available from **http://cogcomp.cs.illinois.edu/.** Of particular importance here is our work on semantic analysis of sentences (Semantic Role Labeling, we have developed the best tool available in this area), our state of the art named entity recognition tool (which is the best broad coverage NER tool available) and our work on inference models for textual entailment.

Probabilistic Inference: Over the years we have made several key contributions in Probabilistic Inference. **(1)** We have proved a now classical result showing that exact inference in Bayesian network is #P-complete, and studied approximate inference, showing that in many cases even approximate inference, in a well-defined sense, is computationally intractable. **(2)** We were the first to develop general purpose and exact Lifted, first-order probabilistic inference algorithm. This algorithm takes first-order logic description of a Markov network, and, without propositionalzing it (that is, it does not depend on the vocabulary size), performs exact inference [Braz et. al, 2005]. This algorithm has already had significant impact on the probabilistic inference community and is bound to change the way probabilistic inference is done. **(3)** Given the intractability of general purpose inference with respect to standard representations of probability distributions (Bayes Nets and Markov Nets) we have proposed a new, Multi-Linear Representation (MLR) of discrete distributions; we have shown that it can be used to concisely represent classes of distributions which have exponential size in other commonly used representations, while supporting probabilistic inference in time linear in the size of the representation. Moreover, we presented algorithms for learning bounded-size distributions represented using MLR [Roth & Samdani'09].

Language Acquisition and Psycholinguistics: Through collaboration with psycholinguists we have started to study over the last few years possible mechanisms through which children acquire language. Using our state-of-the-art Semantic Role Labeling approach to mimic language-learning processes in children, we are able to control what knowledge sources are available to the learner. With this setup we are able to reproduce findings from child learning experiments, demonstrating the efficacy of psycholinguistic theories in a controlled learning environment contributing to the structure mapping account of the Syntactic Bootstrapping theory [Connor et. al 2011, 2010].

Understanding Natural Language in Context: Understanding natural language, reasoning, and acting in the world have been long recognized as fundamental phenomena of intelligence. Over the last few years we started to develop an understanding for how having a model of the world — concepts, objects and relations — contributes to understanding the meaning of natural language utterances about this world. Our key contribution to date is a new learning paradigm, Response based Learning, that facilitates learning from "natural instructions", thus allowing the teacher to communicate the relevant domain expertise to the learner without necessarily knowing anything about the internal representations used in the learning process.[Goldwasser, Roth 2011, Clarke et. al 2010].

Trustworthiness of Information: Much work in NLP has focused on determining what a document means, but we also must know whether or not to believe it. This has become even more important in the era of information overload and ease of publishing and has significant societal impact. We have begun a line of work in this area and our key contribution so far has been the introduction of a framework for incorporating prior knowledge into fact-finding algorithms, expressing both general "common-sense" reasoning and specific facts already known to the user.