Learning and Inference in Natural Language: from Stand Alone Learning Tasks to Structured Representations

Dan Roth
Department of Computer Science
University of Illinois at Urbana-Champaign
Urbana, IL 61801-2302, USA
danr@uiuc.edu

I will describe some of our work on understanding the role of learning in supporting reasoning and other high-level cognitive tasks, with a focus on learning and inference tasks in natural language.

In recent years there has been a lot of work on stand along learning tasks in natural language processing. It has become clear, through work in learning theory and NLP that linear learning algorithms can be used to learn stand alone target concepts in this domain accurately and reliably [Rot99, Rot98]. Algorithmic properties, as well as their appropriateness given the sparsity characteristics of the domain are fairly well understood. The key issues in this area remain those of mapping the input domain into a more expressive feature space; a lot of work in this area is devoted to study these issues both theoretically and experimentally [CR03, KRS01, CR00].

Given that, a lot of recent work in natural language processing is devoted to higher level tasks, and this will be the focus of the talk.

Natural language decisions often involve assigning values to sets of variables where complex and expressive dependencies can influence, or even dictate, what assignments are possible. This is common in natural language tasks ranging from predicting pos tags of words in their context – governed by sequential constraints such as that no three consecutive words are verbs – to semantic parsing – governed by constraints such that certain verbs must have, somewhere in the sentence, three arguments of specific semantic types.

I will describe research on a framework that combines learning and inference for the problem of assigning globally optimal values to a set of variables with complex and expressive dependencies among them [PR01, RY, PRYZ04b, PRYZ04a].

The inference process of assigning globally optimal values to mutually dependent variables is formalized as an optimization problem and is solved as an integer linear programming (ILP) problem. Two general classes of training pro-

cesses are presented. In one, the inference process applied to derive a global assignment to the variables of interest is decoupled from the process of learning estimators to variables' values; in the second, dependencies among the variables are incorporated into the learning process, and directly induce estimators that yield a global assignment.

I will show how this framework generalizes existing approaches to the problem of learning structured representations, and discuss the advantages the two training paradigms have in different situations. Examples will be given in the context of semantic role labeling and of information extraction problems.

Both approaches suggest interesting questions that I will try to comments on, from the cognitive perspective. These pertain to the relation between the strictness of the constraints and the difficulty/easiness of the learning task, as well as the possibility of deriving good features for stand alone tasks as an outcome of global learning and inference tasks.

References

[CR00] C. Cumby and D. Roth. Relational representations that facilitate learning. In *Proc. of the International Conference on the Principles of Knowledge Representation and Reasoning*, pages 425–434, 2000.

[CR03] C. Cumby and D. Roth. On kernel methods for relational learning. In *Proc. of the International Conference on Machine Learning*, 2003.

[KRS01] R. Khardon, D. Roth, and R. Servedio. Efficiency versus convergence of boolean kernels for on-line learning algorithms. In *NIPS-14; The 2001 Conference on Advances in Neural Information Processing Systems*. MIT Press, 2001.

[PR01] V. Punyakanok and D. Roth. The use of classifiers in sequential inference. In *NIPS-13; The 2000 Conference on Advances in Neural Information Processing Systems*, pages 995–1001. MIT Press, 2001.

[PRYZ04a] V. Punyakanok, D. Roth, W. Yih, and D. Zimak. Learning via inference over structurally constrained output. In *NIPS-04 workshop on Structured Representations*, 2004.

[PRYZ04b] V. Punyakanok, D. Roth, W. Yih, and D. Zimak. Semantic role labeling via generalized inference over classifiers. In *COLING 2004, The 20th International Conference on Computational Linguistics*, pages 130–133, 2004.

[Rot98] D. Roth. Learning to resolve natural language ambiguities: A unified approach. In *Proc. of the American Association of Artificial Intelligence*, pages 806–813, 1998.

[Rot99] D. Roth. Learning in natural language. In *Proc. of the International Joint Conference on Artificial Intelligence*, pages 898–904, 1999.

[RY] D. Roth and W. Yih. A linear programming formulation for global inference in natural language tasks.