

Scalable Statistical Relational Learning for NLP

Instructors: William Yang Wang and William W. Cohen

Prerequisites: No prior knowledge of statistical relational learning is required.

Abstract:

Statistical Relational Learning (SRL) is an interdisciplinary research area that combines first-order logic and machine learning methods for probabilistic inference. Although many Natural Language Processing (NLP) tasks (including text classification, semantic parsing, information extraction, coreference resolution, and sentiment analysis) can be formulated as inference in a first-order logic, most probabilistic first-order logics are not efficient enough to be used for large-scale versions of these tasks. In this tutorial, we provide a gentle introduction to the theoretical foundation of probabilistic logics, as well as their applications in NLP. We describe recent advances in designing scalable probabilistic logics, with a special focus on ProPPR. Finally, we provide a hands-on demo about scalable probabilistic logic programming for solving practical NLP problems.

Outline:

- **Part 1: Foundations and Applications of Probabilistic First-Order Logic**
- We will provide a brief review of some first-order learning systems that have been developed in the past: Markov Logic Networks (Richardson and Domingos, 2006), Stochastic Logic Programs (Muggleton, 1996). In this part, we introduce the semantics of the above languages with their inference (and learning) approaches. We analyze and discuss the core ideas behind of such language. We show various applications of probabilistic logics in NLP.
- **Part 2: Scalable Probabilistic Logics: A Case Study of ProPPR.**
- We will focus on the efficiency issue, and introduce recent advances of scalable probabilistic logics, including lifted inference techniques (Van den Broeck and Suciu, 2014) and probabilistic soft logic (Bach et al., 2015). In particular, we will take CMU's ProPPR (Wang et al., 2013) as a case study. We describe the main contributions of ProPPR: including its approximate personalized PageRank inference scheme, parallel stochastic gradient descent learning method, and its flexibility in theory engineering. We then introduce the structure learning methods in ProPPR (Wang et al., CIKM 2014), including a structured regularization method as an alternative to predicate invention (Wang et al., IJCAI 2015). We will also cover our latest attempt of learning first-order logic formula embeddings, and discuss its relationship to (and possible connections between) even newer approaches to modeling knowledge bases, relationships, and inference using deep learning methods. To conclude this part, we show an interesting application of ProPPR (Wang et al., ACL-IJCNLP 2015): a joint information extraction and knowledge reasoning engine.
- **Part 3: Demos and Practical Applications.**

- We switch from the theoretical presentations to an interactive demonstration session: we aim at providing a hands-on lab session to transfer the theories of scalable probabilistic logics into practices. More specifically, we will provide a demo of several applications on synthetic and real-world datasets. Participants are encouraged to check out our repository on Github (<https://github.com/TeamCohen/ProPPR>) and bring laptops to the tutorial. The list of demo examples to be considered are text categorization, entity resolution, knowledge base completion (Wang et al., MLJ 2015), dependency parsing (Wang et al., EMNLP 2014), structure learning, and joint information extraction & reasoning.

About the Instructors:

William Yang Wang: Carnegie Mellon University, yww@cs.cmu.edu, <http://www.cs.cmu.edu/~yww/>

William Wang is a final-year PhD student at the School of Computer Science, Carnegie Mellon University. He works with William Cohen on designing scalable learning and inference algorithms for statistical relational learning, knowledge reasoning, and information extraction. He has published about 30 papers at leading conferences and journals including ACL, EMNLP, and NAACL. He has received best paper awards (or nominations) at ASRU, CIKM, and EMNLP, a best reviewer award at NAACL 2015, the Richard King Mellon Presidential Fellowship in 2011, and he is a Facebook Fellowship finalist. He is an alumnus of Columbia University, and a former research scientist intern at Yahoo! Labs, Microsoft Research Redmond, and University of Southern California.

William W. Cohen: Carnegie Mellon University, wcohen@cs.cmu.edu, <http://www.cs.cmu.edu/~wcohen/>

William Cohen is a professor of machine learning at Carnegie Mellon University. Dr. Cohen's research interests include information integration and machine learning, particularly information extraction, text categorization and learning from large datasets. He has a long-standing interest in statistical relational learning and learning models, or learning from data, that display non-trivial structure. He holds seven patents related to learning, discovery, information retrieval, and data integration, and is the author of more than 200 publications. He was a past president of International Machine Learning Society. He is a AAAI fellow, and was a winner of SIGMOD Test of Time Award and SIGIR Test of Time Award.

References:

Richardson, Matthew, and Pedro Domingos. "Markov logic networks." *Machine learning* 62.1-2 (2006): 107-136.

Muggleton, Stephen. "Stochastic logic programs." *Advances in inductive logic programming* 32 (1996): 254-264.

Van den Broeck, G., and Dan Suciu. "Lifted probabilistic inference in relational models." *UAI tutorials* (2014).

Bach, Stephen H., et al. "Hinge-loss Markov random fields and probabilistic soft logic." *arXiv preprint arXiv:1505.04406* (2015).

Wang, William Yang, Kathryn Mazaitis, and William W. Cohen. "Programming with personalized pagerank: a locally groundable first-order probabilistic logic." *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM, 2013.

Wang, William Yang, Kathryn Mazaitis, and William W. Cohen. "Structure learning via parameter learning." *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 2014.

Wang, William Yang, Kathryn Mazaitis, and William W. Cohen. "A soft version of predicate invention based on structured sparsity." *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI 2015)*, Buenos Aires, Argentina. 2015.

Wang, William Yang, and William W. Cohen. "Joint information extraction and reasoning: A scalable statistical relational learning approach." *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and The 7th International Joint Conference of the Asian Federation of Natural Language Processing (ACL-IJCNLP 2015)*, Beijing, China. 2015.

Wang, William Yang, et al. "Efficient inference and learning in a large knowledge base." *Machine Learning* 100.1 (2015): 101-126.

William Yang Wang, Lingpeng Kong, Kathryn Mazaitis, and William W. Cohen, "Dependency Parsing for Weibo: An Efficient Probabilistic Logic Programming Approach", in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, short paper, Doha, Qatar, Oct. 25-29, 2014, ACL.