I am interested in statistical machine learning, especially statistical relational learning and learning models with efficient inference, with the goal of making machine learning more powerful, more practical, and easier to use.

Machine learning is one of the most important applications of computers today, but, like the Internet in the mid-90's, its full potential is far from realized. Most approaches to date only learn from sets of independent examples, where each example is a simple vector of feature values. However, most real-world problems are interdependent and highly structured. Consider the task of labeling web pages: in addition to the words on the page, the links to other web pages are highly informative, as are the labels of the linked pages. Flattening such a problem into vectors throws away key information. The ever increasing availability and variety of data leads to more and more structured information that is poorly handled by flat approaches. My work has led to more powerful representations and more efficient algorithms for statistical relational learning, helping the field move to richer solutions for these rich problems.

When moving beyond simple classification, one of the greatest hurdles for statistical machine learning is the problem of inference. For most models of moderate size and complexity, exact inference is intractable, while approximate inference may be prohibitively slow or inaccurate. My work has shown that by taking inference cost into account while learning, you can learn accurate models that still guarantee efficient, exact inference. I hope to extend some of these results to the relational setting, where large, complex models lead to particularly difficult inference problems.

My overall approach is to find general solutions to important, real-world problems by bridging theory and practice. I believe in thorough empirical evaluation, along with releasing code and data whenever possible so that the experiments are reproducible.

Statistical Relational Learning

Most real-world problems are both complex and uncertain. Rich relational structure is traditionally handled with first-order logic; uncertain data (such as that generated by stochastic processes, corrupted by noise, or incompletely understood) is typically modeled using probability. The goal of statistical relational learning is to combine first-order logic and probability, providing a more powerful and general framework for machine learning. Instead of reducing complex problems to vectors, a statistical relational approach allows us to work directly with the natural problem structure, often yielding better results with less work.

I am interested in developing general-purpose statistical relational representations and algorithms. While it may always be possible to create specially engineered solutions that outperform generic methods, we are likely to see the biggest gains from improving the performance and reliability of our best "generic" methods. A good analogy is microprocessors: while specialized chips still have their purpose, generic microprocessors (x86, ARM, etc.) offer an effective and low-cost solution for many problems. For machine learning and artificial intelligence in general, the best general-purpose solution to date is Markov logic, a language for statistical relational representation, learning, and reasoning. In its simplest form, Markov logic just attaches a weight to each formula in a first-order knowledge base. Given a set of constants for a domain, the sum of the weights of satisfied formulas determines the relative probability of a possible world. Weights and formulas can be learned or revised from data using ideas from convex optimization and inductive logic programming. Already, Markov logic has been used to build state-of-the-art systems for computational biology, natural language processing, entity resolution, and more.

My work to date has improved and extended Markov logic in two key ways. First, I developed a multi-layer generalization of Markov logic called recursive random fields (RRFs). RRFs resolve a key inconsistency in the Markov logic representation, represent many probability distributions exponentially more compactly than an MLN, and allow for new, more flexible approaches to learning. One can view a logical knowledge base as a single formula tree, in which each leaf is a variable and each interior node is a conjunction or disjunction, or an existential or universal quantifier. Markov logic "softens" this knowledge base by replacing the top-level universal quantifiers and conjunctions with a noisy AND, but all lower levels remain purely logical. Recursive random fields make every level of the formula tree probabilistic, creating softened disjunctions and existentials. This is useful for representing "m-of-n" or "m-of-all" concepts, e.g., a certain diagnosis is likely if at least four of ten symptoms are present, or we might say that popularity depends on having at least 20 friends. Furthermore, since weights are attached to every part of the network structure, we can effectively revise the structure through weight learning. This is similar to backpropagation neural networks, which are capable of representing and learning many different functions starting from a simple structure and simple learning procedure.

Recursive random fields have many nice theoretic properties, but under current methods, learning is very sensitive to the initial weights and inference is significantly slower than for MLNs. So I put my work on RRFs aside to focus on making MLNs more practical, in hopes of eventually applying those ideas to RRFs. At the time, the weight learning algorithms for Markov logic were rather crude and frequently failed on hard problems. Other researchers had been forced to simplify their models or set weights via various hacks. I adapted several convex optimization methods to cope with the particular challenges of Markov logic. The algorithms I developed are now the standard approach to Markov logic weight learning.

Learning for Efficient Inference

One of the important remaining challenges for Markov logic, and in fact, for graphical models in general, is efficient inference. Exact inference computes the most accurate probabilities, but is often intractable since probabilistic inference is #P-complete. Approximation algorithms range from Markov chain Monte Carlo (MCMC) sampling to message passing, each with its distinct set of advantages and disadvantages.

I have worked to address the problem of efficient inference for the case of propositional graphical models, with the hope of eventually extending these results to Markov logic or a similar representation. Rather than trying to solve the hard inference problem directly, I developed techniques to avoid it by only learning models that allow for efficient inference. Through experiments on 50 datasets, I showed that naive Bayes mixture models, one of the simplest and most efficient models for probability estimation, can be competitive with state-of-the-art Bayesian networks. For the cases where Bayesian networks remain a more accurate representation, we can bias the learning algorithm to prefer networks with cheaper inference. By combining learning and inference, I was able to learn models that appeared very complex (treewidth ¿ 100) but still allowed for very efficient inference (¡ 100ms). In general, inference cost is exponential in treewidth. I got around that by using arithmetic circuits, which can exploit context-specific independence and other forms of structure in the probability distribution. My methods are more effective and more efficient than related work that only restricts treewidth.

Future Work

Much remains to be done to make machine learning more powerful and more practical. I am currently looking into inference by approximate compilation: building an arithmetic circuit by dynamically introducing context-specific independence or other forms of structure. Arithmetic circuits have never been used before for approximate inference, but could offer much better results by adapting the approximation to the complexity of the problem and the available resources. Markov logic networks offer a particularly exciting application of dynamic inference methods, since their structures are determined by highly structured sets of formulas.

For the future, I plan to continue exploring rich, statistical relational representations and efficient algorithms. People who want to apply machine learning should be able to express their knowledge and data naturally, learn a good model automatically, and use the model to make predictions effficiently. Machine learning has come a long way towards this goal for the subproblem of classification using vector data. Statistical relational learning brings greater challenges, but also greater opportunities to use our data to solve problems in bioinformatics, natural language processing, social network analysis, user interface design, and more. One of the things that excites me about machine learning is its power to solve such a wide range of problems in different fields. But to have the greatest impact, machine learning must be more flexible and require less knowledge of machine learning. That is what I want to help accomplish.

Other Research

The work on spam filtering I conducted while interning at Microsoft Research is already well-cited, and led to invited talks at Oregon State University and the University of Cagliari, Italy. In this work, I looked at the adversarial problem of defeating a spam filter by adjusting the content of an email message. I developed simple but novel ways to reverse engineer spam filters by sending test messages, as well as a way to defeat spam filters with no prior knowledge. These practical attacks offer insights into what make spam filters weak, as well as methods for evaluating spam filter vulnerability. I also formulated the adversarial classifier reverse engineering (ACRE) learning problem so that the hardness of attacking different classifiers can be formally analyzed.

In addition to spam filtering, I have worked on evaluating collaborative filtering algorithms and developing effective methods for desktop activity recognition. These applications of machine learning have great potential to reduce information overload, a quickly growing problem.