Query-Specific Learning and Inference for Probabilistic Graphical Models

Tuesday, June 14, 2011
Gates Hillman Complex 4405
5:00 p.m.

Abstract

Probabilistic graphical models (PGMs) have become the approach of choice for representing and reasoning with high-dimensional probability distributions. However, for most models capable of accurately representing real-life distributions, inference is fundamentally intractable. Optimally balancing the expressive power and inference complexity of the models, as well as designing better approximate inference algorithms, remain important open problems.

This thesis contributes algorithms for learning and approximate inference in probabilistic graphical models that improve on the state of the art by emphasizing the computational aspects of inference over the representational properties of the models. Our contributions fall into two categories: learning accurate models where exact inference is tractable and speeding up approximate inference.

In a generative setting, we propose a polynomial time algorithm for learning the structure of tractable graphical models based on a novel efficient upper bound for conditional mutual information. We provide PAC learnability and graceful degradation guarantees for the resulting structures. Ours is the first efficient algorithm to provide this class of guarantees.

In a discriminative setting, we propose an approach for learning tractable models that tailors the structure of the model to the particular value of evidence that becomes known at test time. By avoiding a commitment to a single tractable structure during learning, the representation power of the models can be expanded without sacrificing efficient exact inference and parameter learning. The result is the same accuracy and an order of magnitude speedup over the state of the art.

Finally, for applications where the intractable model structure is a given and approximate inference is needed, we propose a principled way to speed up convergence of belief propagation by focusing the computation on the query variables and away from the variables that are of no direct interest to the user, resulting in significant speedups over the state of the art.