Research Interests

Brigham S. Anderson

Overview
My recent research has had the major themes of *fast* inference and *active* inference in graphical models. Inference in graphical models is a fascinating problem with broad scope, since 1) most machine learning tasks can be posed as inference on probabilistic models, and 2) most probabilistic models can be described in graphical terms.

Fast inference is clearly useful for real-world applications. The promise of active inference can be expressed intuitively as introducing *curiosity* into machine learning. This allows for closed-loop learning, in which the machine is more autonomous, and learns faster because it can actively seek data to improve its understanding of the world.

Active Inference in Graphical Models
Imagine you have a large graphical model of a human patient. There are some nodes that are "targets" (such as Has_Cancer and Life_Expectancy) and some nodes that are "queries" (such as Is_Smoker and Biopsy_Result.) The goal is to choose which query node to observe in order to learn the most about the target nodes. Medical settings and disease diagnosis are obvious applications, but recommender systems, sensor management, sensitivity analysis, and active learning [2, 3] all use similar computations.

A common technique is to compute the mutual information between each query and target. However, checking all pairs like this incurs a quadratic cost. In [3], I show how to perform a similar all-pairs computation with cost linear in network size, requiring only two passes through the network. The loss function that allows this is a probabilistic form of expected misclassification risk, which is decomposable and allows for a message-passing algorithm. The algorithm demonstrates several orders of magnitude speedups without decrease in accuracy. In the cost-sensitive domains examined, it achieves superior accuracy.

Active Inference and Learning in HMMs
Hidden Markov Models (HMMs) model sequential data in many fields such as speech, text, and video processing, but scarcity of labelled data often hampers learning. In [2], I introduced objective functions for active learning for three fundamental HMM problems: model learning, state estimation, and path estimation. In addition, I described a new set of efficient algorithms for finding optimal greedy queries using these objective functions. The algorithms are fast, i.e., linear in the number of time steps to select the optimal query. Empirical results show that the algorithms can significantly reduce the need for labelled training data.

Although current active learning algorithms can select examples to be labelled, it is not clear how to select sequences of examples to be *completely annotated*, i.e., have each individual state labelled, as in POS tagging. My work on information gain-based algorithms in [6] solves this problem. The algorithm can extract the most useful subsequence from a stream of data. Surprisingly, this computation can be performed in time linear in the number of timesteps by using dynamic programming. These results are general enough to apply to any time series model having the Markov property.

Active Learning Frameworks
Active learning currently suffers from difficulty in targetting its learning to a particular *test* set. The Query By Committee techniques, for instance, actively improve their model of the training set, but have no principled way to take the characteristics of the test set into account. This is significant when, for instance, your training set consists of the entire web, but you are only interested in learning about "resumes" or "universities".

I am developing my work in [3] as a new active learning framework that can achieve this efficiently in a provably correct manner. The framework can be expressed as a probabilistic form of Query by Committee which uses the message-passing algorithm of [3]. This allows the learner to perform extremely focused active learning (it may be the case that the test set consists of a single unlabelled example.)

Fast Inference in Bayesian Models
Another project of interest is fast inference in graphical models using mixture model approximations. The basic idea is to approximate the Bayesian network with a mixture model having tens to hundreds of thousands of hidden states. In addition to making inference trivially easy and fast, the approach also allows inference on hybrid Bayesian networks (those containing Gaussian or Poisson nodes, for instance.) Learning of this large number of hidden states can be achieved using a heirarchical clustering of the hidden states for fast indexing and pruning.

Stochastic Optimization
Another form of active inference is experiment design, which attempts to maximize the amount of information learned about the location and magnitude of an optimum. The field of optimization is well developed, but when noise is included in the problem, almost all numerical solutions do extremely poorly. Machine learning holds significant promise, as it is very good at modeling noisy functions and dealing with uncertain decisions.

In [4], I demonstrated a novel nonparametric version of [7] that is robust to noise and discontinuities in the objective function. I have applied both algorithms in various industrial settings, including 3M chemical engineering research, Caterpillar engine optimization, and astronomical modeling [1, 5]. I have also developed an Excel plug-in version of the algorithm.

References

[1] B. S. Anderson. *Nonparametric Optimization and Galactic Morphology.* PhD thesis, Carnegie Mellon University, 2003.

[2] B. S. Anderson and A. W. Moore. Active learning for hidden markov models: Objective functions and algorithms. In *International Conference on Machine Learning*, 2005.

[3] B. S. Anderson and A. W. Moore. Fast Information Value for Graphicl Models. In *Advances in Neural Information Processing Systems.* Morgan Kaufmann, 2005.

[4] B. S. Anderson, A. W. Moore, and D. Cohn. A Nonparametric Approach to Noisy and Costly Optimization. In *International Conference on Machine Learning*, 2000.

[5] B. S. Anderson, A. W. Moore, A. Connolly, and B. Nichol. Eigengalaxies for Fast Galaxy Morphology. In *Knowledge Discovery and Data Mining.* AAAI Press, 2004.

[6] B. S. Anderson, S. Siddiqqi, and A. W. Moore. Sequence Selection for Active Learning. Submitted to International Conference on Machine Learning, 2006.

[7] A. W. Moore, J. Schneider, J. Boyan, and M. S. Lee. Q2: A memory-based active learning algorithm for Blackbox Noisy Optimization. In J. Shavlik, editor, *Proceedings of the Fifteenth International Conference on Machine Learning*, pages 386–394. Morgan Kaufmann, 1998.