Probabilistic Learning in Bayesian and Stochastic Neural Networks

William H. Hsu
Department of Computer Science and Beckman Institute
University of Illinois at Urbana-Champaign
bhsu@cs.uiuc.edu, **http://anncbt.ai.uiuc.edu**

Abstract

[1] The goal of this research is to integrate aspects of artificial neural networks (ANNs) with symbolic machine learning methods in a probabilistic reasoning framework. Improved understanding of the semantics of neural nets supports principled integration efforts between seminumerical (so-called "subsymbolic") and symbolic intelligent systems. My dissertation focuses on learning of spatiotemporal (ST) sequences. In recent work, I have investigated architectures for modeling of ST sequences, and dualities between Bayesian networks and ANNs that expose their probabilistic and information theoretic foundations. In addition, I am developing algorithms for automated construction of Bayesian networks (and hybrid models); metrics for comparison of Bayesian networks across architectures; and a quantitative theory of feature construction (in the spirit of the PAC formalism from computational learning theory) for this learning environment. (Haussler 1988) Such methods for pattern prediction will be useful for building advanced knowledge based systems, with diagnostic applications such as intelligent monitoring tools.

Statistical learning systems such as neural networks and genetic algorithms currently lack formal quantitative characterization compared to pure Bayesian models such as: classification and regression trees (CART); generalized maximum likelihood algorithms for modeling mixtures of experts, such as Expectation-Maximization (EM); and other substrates for pure random sampling and parameter estimation, such as Hidden Markov Models (HMMs). My experiments are focused on "multimodal" ST sequences, a category of patterns which gives rise to classification problems such as: interpretation of speech signals, prediction of protein secondary fold, and biomedical monitoring. Each of these applications currently demands specialized methods for acquiring and organizing knowledge, presenting difficulties for multimodal systems (which attempt to combine multiple diagnoses and justifications). Isolated successes are reported in the literature using various ANN and hierarchical (blackboard architecture and HMM-based) methods for this mixture modeling problem. (Jordan & Jacobs 1994)

My approach uses dualities, between "pure probabilistic" models of causation (especially Bayesian networks and their dynamic decision- and utility-theoretic extensions) and their stochastic ANN analogues, for quantitative analysis. For example, we may wish to evaluate Bayesian network quality in terms of ability to: explain a sequence of causally related events; evoke a policy or action; predict the next in a sequence of observations; or relate a sequence of observable actions to recognized goals. The best Bayesian networks for such diagnostic, planning, predictive, or plan recognitive systems vary greatly in architecture, and are difficult to compare objectively and syntactically, because of these topological differences. Similarly, when field test results and representative data samples are not always available, it is infeasible to empirically evaluate structural quality.

An alternative is to convert the networks of interest to a common ANN representation (e.g., a restriction of Discrete Hopfield Networks called *sparse, bipartite Boltzmann Machines*) and to analyze this "baseline" dual architecture (e.g., in terms of stochastic behavior as in simulated annealing). A recent paper explains this duality, describes its application and benefits, and presents: combinatorial analysis of dual networks, an algorithm for interconversion (change of network representation), and early analysis of predictive accuracy of simple dual networks. (Hsu 1997)

The long-term goal of this research is the development of a quantitative theory of **N***etwork* **E***fficiently* **R***epresentable* **F***unctions* (*NERF*s) as characterized by Russell and Norvig. This theory will help to redress the semantic deficiency of ANNs, compared to pure probabilistic formalisms with well-understood symbolic interpretations. These developments will, in turn, improve our ability to develop hybrid systems capable of acquiring, representing, and recovering explanations about causal relationships in ST sequences. An important statistical pattern recognition aspect of this capability is *causal discovery*, the detection of hidden influents. Finally, integration of Bayesian and stochastic neural networks in a common probabilistic framework will not only prescribe and help optimize both kinds of network architectures for ST sequence learning, but will produce guidelines for decomposition of ST sequences in the mixture modeling paradigm. (Jordan & Jacobs 1994)

References

Haussler, D. 1988. Quantifying inductive bias: AI learning algorithms and Valiant's learning framework. *Artficial Intelligence* 36:177–221.

Hsu, W. H. 1997. A position paper on statistical inference techniques which integrate neural network and bayesian network models. In *International Conference on Neural Networks (ICNN-97), to appear*.

Jordan, M. I., and Jacobs, R. A. 1994. Hierarchical mixtures of experts and the EM algorithm. *Neural Computation* 6:181–214.