Boosted Estimation of Probability Functions

Eric Bauer

Problem: Probability functions are difficult to learn from data, especially when the data features are highly interdependent. Further, even in cases where independencies exist between data features, the task of discovering and verifying such independencies is expensive computationally. Yet without exploiting some decomposition of a probability function, even performing basic inferential queries becomes quite difficult. How can we decompose and robustly estimate highly interdependent probability functions?

Impact: Techniques for probabilistic reasoning are used widely in medical domains, automated software support systems, text categorization problems, speech recognition tasks, robotics, and other areas. However, the applicability of the techniques is often tied to the quality of the decomposition used. Many of the domains in which probabilistic reasoning techniques have been successfully applied have a low degree of feature interdependence or fairly simple causal relationships. In general, such properties do not necessarily hold. By finding more flexible decompositions, the applicability of probabilistic reasoning would be extended greatly. Further, by finding more robust estimation techniques for such decompositions, we can learn better probability functions and improve the quality of solutions based on such functions.

State of the Art: Recent work in probabilistic reasoning has focused on using graphical models such as Bayesian networks or hidden Markov models to represent dependence relationships between data features [4]. Such decompositions break up a large probability function into a collection of local probability models, which significantly reduces the computation needed to learn and to use a probability function [3]. Moreover, graphical representations allow models to exhibit context-specific independence, altering the dependence structure as the quantity of known data changes [1].

Approach: One simple and well-known probabilistic decomposition is the mixture model: a weighted sum of a number of simpler probabilistic models. Traditionally, such a decomposition has been used when multiple independent processes are known to be at work. But mixtures are not limited to such problems; as the size of the mixture grows, even extremely simple probability functions can be mixed to produce arbitrarily complex probability functions and dependence relationships. However, one would prefer not to have to estimate all the mixture components simultaneously for such a large mixture.

Much recent work in the field of learning classifiers has focused on an algorithm for generating and combining weak hypotheses. This class of algorithms, known as boosting algorithms, produces a strong additive model that does not overgeneralize and that does not estimate all the weak hypotheses simultaneously [2]. My current work concentrates on constructing a boosting algorithm for the mixture-model estimation task and studying its properties.

Future Work: Beyond exploring new applications of this type of estimation, it might be possible to use a boosted mixture model to determine the underlying dependence structure of a given problem. One method for doing this would be to first construct a rough estimate using the mixture model, then use this rough estimate to help answer dependence queries issued by a structure learner for a Bayesian network. A second and more challenging direction lies in using the behavior of boosting methods in this new scope to clarify why boosting is capable of learning without overgeneralization.

Selected References:

[1] Craig Boutilier, Nir Friedman, Moises Goldszmidt, and Daphne Koller. Context-specific independence in Bayesian networks. In Eric Horvitz and Finn Jensen, editors, *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence (UAI-96)*, pages 115–123, San Francisco, August 1–4 1996. Morgan Kaufmann Publishers.

[2] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.

[3] David Heckerman. A tutorial on learning bayesian networks. Technical Report MSR-TR-95-06, Microsoft Research, March 1995.

[4] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1991. (Revised 2nd Edition).