Leveraging Domain Knowledge to Learn Multiple Bayesian Network Structures

Diane Oyen
Department of Computer Science
University of New Mexico
doyen@cs.unm.edu

1 Problem and Motivation

A Bayesian network is a standard tool in statistical data mining that gives a compact representation of relationships among variables in data. The structure of the network, or edges in the graph, represent conditional dependencies between variables. Machine learning algorithms have been developed to find Bayesian networks that model the underlying structure of relationships among data variables. For example, the variables could be the activity in various regions of the brain. Modern neuro-imaging technology allows us to look for patterns among the relationships between the activity of brain regions. Neuro-scientists suspect that certain mental illnesses, such as schizophrenia, exhibit abnormal brain networks. To explore these differences in networks, a different Bayesian network should be learned for each sub-population of subjects. Learning several different, but related, networks is known as multitask learning. The data is partitioned into tasks, and for each task a Bayesian network is learned.

Simply partitioning the data and learning completely independent networks is problematic when there is not enough data in each task. Furthermore, we believe that the data is actually generated by similar, rather than completely independent, networks. In fact, we often have domain information about how tasks are related. For example, Subject A has a family history of schizophrenia, is asymptomatic, and is an alcohol addict, while Subject B has no family history of schizophrenia, is asymptomatic, and is an alcohol addict. These are more similar to each other than either is to Subject C, who has no family history of schizophrenia, has acute symptoms of schizophrenia, but is not an alcohol addict. I present a novel algorithm that learns multiple Bayesian networks for a collection of unsupervised machine learning tasks when limited data is available and a metric of the relatedness of tasks is given.

2 Background and Related Work

Multitask learning algorithms generally represent the similarity among tasks in one of three ways: all tasks are assumed to be equally similar [2, 5]; the similarity among tasks is estimated from the same data that will be used to train the model [7, 3]; or the similarity between tasks is provided by some other source such as task-specific domain information or an expert [1]. This third option has been explored recently for zero-data classification problems when no training data are available for certain tasks [4, 6]. They show that knowledge of the relationships among tasks can compensate for lack of training data.

3 Approach and Uniqueness

I extend this use of domain knowledge to the unsupervised problem of multitask Bayesian network structure learning. Learning a different network for each task allows for variability at the cost of fragmenting the data into small chunks for each task. Domain knowledge can provide a task-relatedness metric to facilitate the intelligent sharing of information across tasks to compensate for lack of data. Therefore, the inputs to my algorithm come in two forms. The first is training data associated with each task and the second is a metric describing the relatedness of all tasks. This task-relatedness metric is given by an undirected graph in which edges represent expected similarity between tasks.

My algorithm learns a network for each task by optimizing a weighted combination of the likelihood of the data and the similarity across neighboring networks. To do this, I introduce a new regularization penalty into the structure score while searching over the space of graphs through an iterative process. This penalty is a graph-distance metric between neighboring networks and is applied in addition to the common-practice complexity penalty for each individual network. The resulting networks are essentially smoothed toward having similar structures as those for similar tasks.

4 Results and Contributions

Without multitask structure learning, we would have to decide between learning networks independently for each task or pooling the data to learn a single network representing all tasks. On synthetic data, I compare my task-relatedness aware multitask (TRAM) algorithm against standard multitask learning (all tasks are assumed to be equally similar), pooling data, and learning independent networks. When the task-relatedness metric is correct, TRAM outperforms standard multitask learning. TRAM performs better than independent learning when there is little data. TRAM almost always performs better than pooling data. Therefore, I show that task-relatedness knowledge improves the performance of multitask learning when little data is present.

In the neuro-imaging application, observed task-specific activity data are random variables in the networks. A set of clinical variables with contextual information represents an entirely different type of knowledge about the subjects and is used to describe task-relatedness. Incorporating this information into the multitask representation allows the algorithm to learn Bayesian networks for a very large space of contexts with limited data from a limited number of tasks. To my knowledge, this is the first multitask network learning algorithm that can incorporate domain knowledge about the relatedness of tasks. The other, purely data-driven, models ignore this valuable information which is often available in real datasets.

References

[1] Bakker, B., and Heskes, T. Task clustering and gating for bayesian multitask learning. *J. Mach. Learn. Res. 4* (2003), 83–99.

[2] Caruana, R. Multitask learning. *Machine Learning 28*, 1 (July 1997), 41–75.

[3] Eaton, E., Desjardins, M., and Lane, T. Modeling transfer relationships between learning tasks for improved inductive transfer. In *ECML PKDD '08: Proceedings of the 2008 European Conference on Machine Learning and Knowledge Discovery in Databases - Part I* (Berlin, Heidelberg, 2008), Springer-Verlag, p. 317–332.

[4] Larochelle, H., Erhan, D., and Bengio, Y. Zero-data learning of new tasks. In *AAAI'08: Proceedings of the 23rd national conference on Artificial intelligence* (2008), AAAI Press, p. 646–651.

[5] Niculescu-Mizil, A., and Caruana, R. Inductive transfer for bayesian network structure learning. In *Eleventh International Conference on Artificial Intelligence and Statistics (AISTATS-07)* (2007).

[6] Palatucci, M., Pomerleau, D., Hinton, G., and Mitchell, T. Zero-shot learning with semantic output codes. In *Neural Information Processing Systems (NIPS)* (2009), p. 1410–1418.

[7] Thrun, S., and O'Sullivan, J. Discovering structure in multiple learning tasks: The TC algorithm. In *International Conference on Machine Learning* (1996), Morgan Kaufmann, p. 489–497.