Task Selection for Multi-Task Bayes Net Structure Learning

Alexandru Niculescu-Mizil, Rich Caruana
Department of Computer Science
Cornell University
Ithaca, NY 14853
{alexn,caruana}@cs.cornell.edu

Bayes Nets [1] provide a compact, intuitive description of the dependency structure of a domain by using a directed acyclic graph to encode probabilistic dependencies between variables. Acquiring the dependency graph from human experts, is often difficult and expensive, so significant research has focused on learning Bayes Nets from data. The learned dependency graph provides useful information about a problem and is often used as a data analysis tool. For example Friedman et al. used Bayes Nets learned from gene expression level data to discover regulatory interactions between genes for a species of yeast [2].

Most of the Bayes Net structure learning research has focused on learning the dependency graph for one problem in isolation. In many situations, however, data is available for multiple related problems. In these cases, inductive transfer [3, 4, 5] suggests that it may be possible to learn more accurate dependency graphs by *transferring* information between problems.

In a previous paper [6], we proposed a multi-task Bayes Net structure learning method that is able to find more accurate network structures for each task by taking advantage of the information conveyed by the other related tasks. Using synthetic problems created from the ALARM and the INSURANCE networks, we showed that multi-task structure learning performs very well, reducing the number of incorrect arcs in the learned structures by 20% - 57% and KL-Divergence by 6% - 26% when compared to learning the structures independently. In that paper however, we made the assumption that the user is able to identify a set of related tasks and provide it to the algorithm.

Here we tackle the task selection problem for multi-task structure learning. Task selection is an important, but usually hard, problem in inductive transfer: given a set of tasks, select a subset to use as relate tasks in a multi-task learner. The more related the tasks in the selected subset are, the more benefit multi-task learning will provide. The approach we take here is to first order all the tasks by a measure of their relatedness to a principal task, then select the first N to use as the related tasks, where N can either be specified by the user, or selected using an independent validation set.

Unlike other multi-task learning settings, where the notion of task relatedness is not well defined, in the multi-task structure learning setting there is a clear definition of relatedness: two tasks are related if they have similar structures. While computing the relatedness between two tasks this way is a chicken and egg problem (since the network structures are exactly what we are trying to learn), we can approximate how related two tasks are by using network structures learned from data instead of using the real network structures. This approximation is surprisingly effective. On synthetic problems, the average rank correlation between the task orderings obtained using this approximative measure of relatedness and the true task orderings is 0.96.

Besides synthetic problems, we have also applied our multi-task structure learning technique to a real problem in the ornithology domain. We generated multiple tasks by assigning the data collected in each Bird Conservation Region (BCR) to a separate task.(See Figure 2). The left graph in Figure 1 shows the mean log likelihood on an independent test set for one of the BCRs as a function of the number of related tasks that are selected by the task selection procedure. The graph shows that the performance first improves (up is good) until the four most related tasks are added to the set of extra tasks, then then it starts to decrease as less related tasks are added.

The right graph in Figure 1 shows the improvement in mean log likelihood that multi-task structure learning provides over single task structure learning on each of the eleven tasks. Multi-task structure learning yields better mean log likelihood than single task structure learning on ten of the eleven BCRs, with a significant improvement on the first six BCRs. Since the Bird Conservation Regions have been carefully designed to be ecologically distinct (thus have dissimilar characteristics) it is actually surprising that multi-task learning is able to provide such a significant improvement for some of the BCRs.

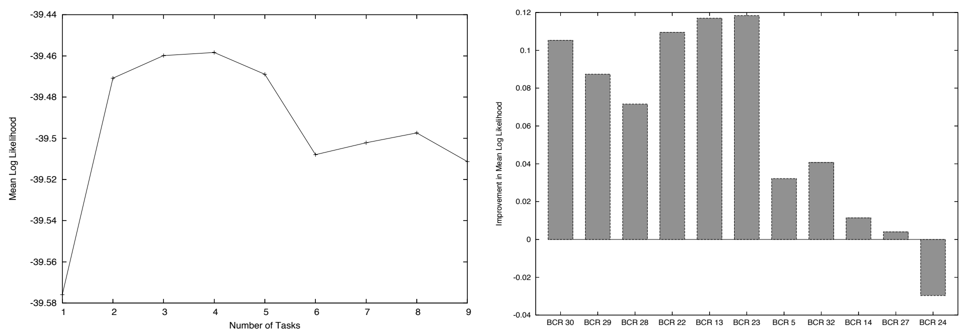

Figure 1: Results on the ornithology data-set.

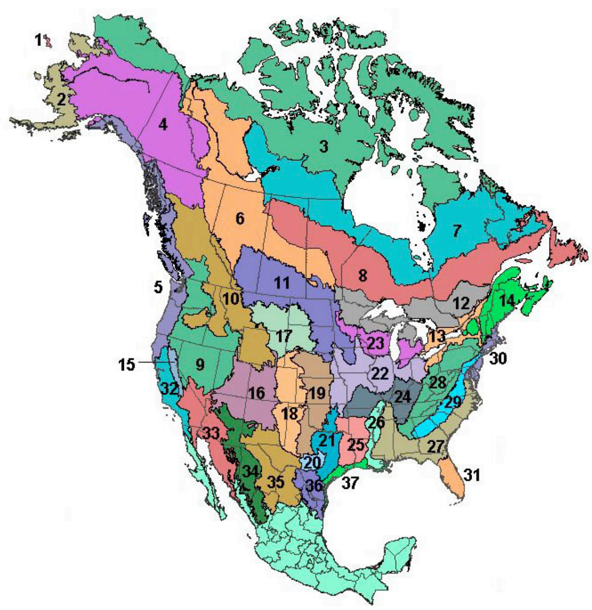

Figure 2: Bird Conservation Regions Map

References

[1] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.

[2] N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using bayesian networks to analyze expression data. *J. Comput. Biol.*, 7(3-4):601–620, 2000.

[3] Rich Caruana. Multitask learning. *Machine Learning*, 28(1):41–75, 1997.

[4] J. Baxter. A bayesian/information theoretic model of learning to learn via multiple task sampling. *Mach. Learn.*, 28(1):7–39, 1997.

[5] S. Thrun. Is learning the n-th thing any easier than learning the first? In *Advances in Neural Information Processing Systems*, 1996.

[6] Alexandru Niculescu-Mizil and Rich Caruana. Inductive transfer for bayesian network structure learning. In *Proc. 11th International Conf. on AI and Statistics*, 2007.

Topic: graphical models
Preference: oral