Inferring Tasks for Improved Network Structure Discovery

Diane Oyen*
University of New Mexico
doyen@cs.unm.edu

Eric Eaton
Bryn Mawr College
eeaton@brynmawr.edu

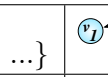
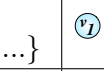
Terran Lane
University of New Mexico
terran@cs.unm.edu

Multitask network structure learning is an important problem in several scientific domains, such as, computational neuroscience and bioinformatics. Multitask learning algorithms have been shown to greatly improve the robustness of learned graphical models [4, 3, 5]. Intuitively and empirically, it is believed that the success of transfer is dependent upon the similarity among the tasks [1, 2].

Defining the tasks themselves remains a challenge in unsupervised multitask learning. Typically, the data are assumed to be a priori partitioned into tasks. The problem we address is how to define tasks within a dataset. This problem is not well studied from the machine learning perspective, but it becomes apparent when working with real data. For example, in group neuroimaging studies, we learn the functional brain networks for subjects from different populations, such as control subjects and patients with schizophrenia. We treat the data from each population as a task. Yet are these tasks appropriate or should subjects be divided into tasks based on their symptoms or based on their family history or based on their drug/alcohol use? Often, after seeing the output of the multitask network algorithm (from a priori task definitions), domain experts will revise their task definitions and re-run the algorithm. We give a framework to infer task definitions using both metadata and the learned networks so that the results learned by the algorithm produce high-scoring networks while partitioning data into tasks that reflect divisions in the metadata (see Figure 1).

Subject	Metadata = [**F**am hist, **S**ymp, **A**lc, ...]	⇒Task
A	$F_A = [1, 0, 1, \ldots]$	k=5
B	$F_B = [0, 0, 1, \ldots]$	k=3
C	$F_C = [0, 4, 1, \ldots]$	k=7
⋮	⋮	⋮

(a) Example metadata

Task	Data	⇒Network
k=5	$D_5 = \{\mathcal{X}_A =$  $, \mathcal{X}_D =$  $, ...\}$	
k=3	$D_3 = \{\mathcal{X}_B =$  $, \mathcal{X}_E =$  $, ...\}$	
k=7	$D_7 = \{\mathcal{X}_C =$  $, \mathcal{X}_F =$  $, ...\}$	
⋮	⋮	⋮

(b) Task-specific data

Figure 1: Example of metadata (a) that can be used to describe tasks. Network models are learned from data for each task (b) comes from a different source than the metadata.

Topic: Unsupervised learning
Preference: Talk or poster

Problem Formulation We have a set of N subjects, where each subject has associated with it a metadata vector F_i and data $\mathcal{X}_i$. The metadata, F_i, are made up of clinical variates that describe the subject such as age, medication, alcohol use, family history of disease, symptoms, etc. The fMRI neuroimage data, $\mathcal{X}_i$, represents a set of multivariate samples of brain activity collected from a single subject. Each $\mathcal{X}_i = \{X_{i,1}, X_{i,2}, \ldots, X_{i,M}\}$ for M samples and each sample $X_{i,j} = [v_1, v_2, \ldots, v_d]$ for d variables representing activity level in regions of interest of the brain. Our goal is to group the subjects into K tasks according to metadata features and then to learn a network for each task. Formally, for each task $k \in 1, \ldots, K$ a network model G_k will be learned from the pooled data $D_k = \{\mathcal{X}_i\}$ for i that satisfy $f(F_i) = k$. We refer to the set of K learned networks as $\mathcal{G} = \{G_1, \ldots, G_K\}$ and the full dataset as $\mathcal{D} = \{D_1, \ldots, D_K\}$.

$$P(\mathcal{G}|\mathcal{D}) = \frac{P(\mathcal{G})}{P(\mathcal{D})} \prod_{k=1}^{K} P(D_k|G_k). = \frac{P(\mathcal{G})}{P(\mathcal{D})} \prod_{i=1}^{N} P(\mathcal{X}_i|G_{f(F_i)})$$

where $f : F \rightarrow k$ is a function that maps metadata to tasks. In this formulation, the objective is to learn $\mathcal{G}$ and f that maximizes $P(\mathcal{G}|\mathcal{D})$. In other words, we learn the best set of networks to fit the data under the constraint that subjects are divided into tasks according to their metadata features.

Algorithm Both $\mathcal{G}$ and f must be learned. The optimization of each one is straightforward if the other is fixed, therefore, the optimization of the objective is broken down into the following steps performed iteratively:

1. Learn networks based on tasks by maximizing $P(G_1, G_2, \ldots, G_K|D_1, D_2, \ldots, D_K)$
2. Re-assign task-labels according to each object's likelihood given the learned networks $Y_i = \arg\max_k P(\mathcal{X}_i|G_k)$.
3. Use labels Y to train a classifier $f(F_i) = \hat{Y}_i$ that maps metadata into tasks.
4. Use predicted labels $\hat{Y}_i$ to build task-data $D_k = \{\mathcal{X}_i\}$ for i s.t. $\hat{Y}_i = k$, and re-iterate.

We apply this algorithm to learning Bayesian networks from fMRI data. A multi-class logistic regression classifier is used to map the clinical metadata features to tasks. In these preliminary experiments, we explore settings of K, starting points for assigning subjects to tasks, and convergence of the algorithm.

References

[1] B. Bakker and T. Heskes. Task clustering and gating for Bayesian multitask learning. *Journal of Machine Learning Research*, 4:83–99, 2003.

[2] E. Eaton, M. desJardins, and T. Lane. Modeling transfer relationships between learning tasks for improved inductive transfer. In *Proceedings of the 2008 European Conference on Machine Learning and Knowledge Discovery in Databases - Part I (ECML PKDD -08)*, pages 317–332, 2008.

[3] J. Honorio and D. Samaras. Multi-task learning of Gaussian graphical models. In *Proceedings of the 27th International Conference on Machine Learning (ICML -10)*, 2010.

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

[5] D. Oyen and T. Lane. Exploiting task relatedness to learn multiple Bayesian network structures. Technical Report TR-CS-2010-08, Department of Computer Science, University of New Mexico, 2010.