Comprehensibility of Generative vs. Class Discriminative Dynamic Bayesian Multinets

John Burge, Terran Lane

University of New Mexico
lawnguy@cs.unm.edu, terran@cs.unm.edu

Abstract

We investigate the comprehensibility of dynamic Bayesian multinets (DBMs) and the dynamic Bayesian networks (DBNs) that compose them. Specifically, we compare the DBM structures resulting from searches employing generative and class discriminative scoring functions. The DBMs are used to model the temporal relationships among RVs and show how the relationships change between different classes of data. We apply our technique to the identification of dynamic relationships among neuro-anatomical regions of interest in both healthy and demented elderly patients based on functional magnetic resonance imaging (fMRI) data. The structures resulting from both generative and class discriminative scores were found to be useful by our collaborating neuroscientist, but for differing reasons. For example, generative scores result in structures that illuminate highly likely relationships and are more easily interpreted. Conversely, structures resulting from class discriminating scores are capable of representing more subtle changes and can illuminate important behavioral differences not apparent from structures learned from generative scores.

Introduction

We investigate the comprehensibility of Bayesian multinet (BM) structures used to model the interaction among neuroanatomical regions of interest (ROIs). Using functional magnetic imaging (fMRI) data collected by Buckner et al. (2000) on both healthy and demented elderly patients, our goal is to illuminate how ROI relationships change due to the onset of dementia.

BMs are a generalization of the Bayesian network (BNs) framework, which has long been recognized as a highly comprehensible modeling framework. BNs are commonly used to model systems in which the underlying relationships among random variables (RVs) is unknown. For these domains, the structure of the BN is also unknown and must be searched for.

Unfortunately, the comprehensibility of BN structures is limited by their static nature. This is because a BN's structure cannot change based on the values of its constituent RVs. E.g., if the independence between two RVs *A* and *B* is dependent on the value of a third RV *C*, a BN will include a link between *A* and *B*, and a link between *A* and *C*. However, the fact that *A* and *B* are conditionally dependent given *C* is not apparent from the structure. Instead, this conditional dependence must be teased out from the BN's parameters: a difficult task given the large number of parameters many BNs have.

Bayesian multinets (BMs) address this issue by modeling the system with multiple BNs. In our case, one BN with its own structure is used to represent each class of data. Since fMRI data is temporal in nature, we use BNs that explicitly model temporal dynamics: dynamic Bayesian networks (DBNs). Multinets that are composed of DBNs are referred to as dynamic Bayesian multinets (DBMs). For our purposes, the DBM is merely a collection of DBNs and we will mainly refer to a DBM's constituent DBNs as opposed to the DBMs themselves.

Once the structure has been found for a DBM, i.e., a structure has been found for each of the underlying DBNs, interpretation of the structure is a relatively straightforward process. The existence of a link clearly indicates a relationship between RVs. However, the nature of that relationship is entirely dependent on the method used to score candidate structures during the DBN structure search. We will refer to these methods as *scores*.

We divide scores into two categories: *generative* and *class discriminative*. Generative scores rank structures based on how well they identify highly likely relationships in the data. Conversely, discriminative scores rank structures based on how well they identify relationships whose dynamics change between classes. The type of score used actually changes the semantic meaning of the links in the resulting DBN structure. DBNs trained with generative scores, which we will simply refer to as generative models, select significantly different structures than DBNs trained with class discriminative scores, referred to as discriminative models.

Regardless of the score used, we find that structures mainly provide two types of useful information. 1) The set of parents for each ROI and how that parent set changes between each class's DBNs. 2) The number of times a ROI shows up as a parent in each class's DBN. Both sets of results give different information about the underlying system based on which type of score was used to train the structure.

We find that generative models are generally more comprehensible than their discriminative counterparts. Since they describe highly likely relationships in the data,

we have found that our collaborating neuroscientist can relatively easily compare structural results with putative neuroanatomical relationships. However, since many of these putative relationships are present in both healthy and demented patients, they tend not to be represented in discriminative models. This makes validation with discriminative models more difficult than with generative models. However, discriminative models more readily indicate changing behavior between demented and healthy populations.

Previously, we have also employed other machine learning techniques such as the Gaussian naïve Bayesian network (GNBN) and the support vector machine (SVM) to classify fMRI data. We briefly look at the comprehensibility of these models and find that they are not particularly helpful to our collaborating neuroscientist.

Background

Bayesian Networks

For an introductory overview of Bayesian networks (BNs), we refer the reader to Charniak's aptly titled "Bayesian Networks without Tears" (Charniak 1991). For a more detailed analysis, to (Jensen 2001; Heckerman, Geiger and Chickering 1995).

BNs are directed acyclic graphs that explicitly represent independence relationships among RVs. They contain nodes for each RV and a directed link between any two statistically correlated nodes. The node originating the directed link is a parent and the terminating node a child. A child and its set of parents are a family. The set of nodes and links composes a structural topology. Each node contains a conditional probability table[1] (CPT) that describes relationships between itself and its parents.

If the topology is unknown, i.e., the independence relations among RVs is unknown, an appropriate structure must be elicited from the data. This process is referred to as structure search and is well understood (Heckerman, Geiger & Chickering 1995) and known to be NP hard (Chickering, Geiger & Heckerman 1994). In essence, structure search boils down to proposing as many hypothesis structures as possible and measuring the wellness of fit between each structure and the data. The method used to measure this fit is called a structure scoring function, which we refer to simply as a score.

In many real-world domains, there are too many structures to score exhaustively. As such, greedy search heuristics are typically employed. One of the most commonly employed heuristic is as follows. Start with a topology with no links and iteratively search through all legal modifications to the current topology. A legal modification is a link addition, removal or inversion that does not result in a cycle. Calculate the score for each modification and chose the modification that resulted in the highest score. Repeat until no modifications yield improvements.

The semantic meaning of the resulting structures depends on the type of score used during the structure search. We divide scores into two general categories: *generative* and *class-discriminative*. Generative scores, such as BD (Cooper and Herskovits 1992), BDE (Heckerman, Geiger and Chickering 1995), MDL (Lam & Bacchus) or even just posterior likelihood, select structures that describe highly likely relationships among RVs. Highly likely relationships are defined to be relationships that have high posterior probabilities given the data.

However, if classification is the goal, highly likely structures are not ideal. Class-discriminating scores, such as class conditional likelihood (CCL) (Grossman & Domingos 2004), EAR (Bilmes et al. 2001) and approximate conditional likelihood (ACL) (Burge & Lane 2005), more effectively select structures that discriminate among classes.

Dynamic Bayesian Networks

We are particularly interested in modeling temporal processes via the *dynamic* BN (DBN) representation. For an overview of DBNs, we refer the reader to (Murphy 2002). We use the following notation.

$\boldsymbol{X}^{0:T}$ is an observation which is a set of T fully observable vectors of n RVs, $\{X_i^t : 1 \le i \le n, 0 \le t < T\}$ with arities r_i. Y is a RV used to represent the class associated with a given observation. A *data point* consists of observable RVs and a class RV: $d = \{\boldsymbol{X}^{0:T}, Y\}$. Notice that the class RV Y is associated with the entire time series. We will assume binary classification such that the domain of $Y = \{1, 2\}$. A *dataset*, D, is a collection of m fully observable data points, $\{d_1, \ldots, d_m\}$. D_j denotes a dataset containing all of (and only) the data points for a specific class, i.e., $D_j = \{d : Y_d = j\}$. $m_j = |D_j|$.

B denotes a DBN containing nodes for RVs $\{\boldsymbol{X}^{0:E}, Y\}$. E is used to denote time in a DBN instead of T since E may not equal T. The parent set for a RV X_i^e, $0 \le e < E$, is denoted $\mathrm{Pa}_B(X_i^e)$ or just $\mathrm{Pa}(X_i^e)$ if the DBN can be inferred. $q_{e,i}$ is the number of possible configurations for the RVs in $\mathrm{Pa}(X_i^e)$.

$\boldsymbol{\Theta}^B$ is the set of CPT parameters for DBN B. $\Theta^B_{e,i}$ is the CPT for node X_i^e in BN B, $\Theta^B_{e,i,j}$ is the multinomial $\mathrm{P}_B(X_i^e \mid \mathrm{Pa}(X_i^e)=j)$ and $\Theta^B_{e,i,j,k}$ is the single CPT parameter $\mathrm{P}_B(X_i^e = k \mid \mathrm{Pa}(X_i^e) = j)$.

The most general DBNs include T nodes for each RV in the system, i.e., one node for each of the RVs at each time point. For most real world problems, such DBNs are intractably large. To resolve this intractability, we make the *Markov order 1* and *stationary* assumptions and assume no isochronal links throughout the rest of this paper.

The Markov assumption states RVs are only dependent on other RVs in a small prior window of time w,

$$P\left(X^t \mid X^{t-1}, \ldots, X^0\right) = P\left(X^t \mid X^{t-1}, \ldots, X^{t-w}\right). \qquad (1)$$

[1] We assume discrete BNs.

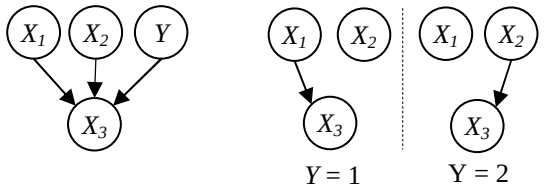

Figure 1. A system where X_3 is correlated with X_1 when $Y = 1$ and with X_2 when $Y = 2$. (left) The BN required to model the system. That X_3 is only correlated with either X_1 xor X_2 is not obvious from the structure and would need to be teased out of X_3's CPT. (right) The multinet required to represent the system. It is clearly obvious from the multinet's structures that X_3 is only correlated with either X_1 xor X_2.

The stationary assumption states the relationships among a RV and its p parents, $X^t_{Pa_1(i)}$, …, $X^t_{Pa_p(i)}$, do not change across time,

$$P\left(X_i^t \mid X^{t-1}_{Pa_1},...,X^{t-1}_{Pa_p}\right) = P\left(X_i^{\hat{t}} \mid X^{\hat{t}-1}_{Pa_1},...,X^{\hat{t}-1}_{Pa_p}\right), \forall t,\hat{t}. \quad (2)$$

The topological structure of these DBNs is composed of two columns of RVs. (For an example, fast forward to Figure 2g.) The RVs no longer represent absolute time points, but instead represent a RV's average statistical relationship with its parents across the entire time series. We denote RVs as $\boldsymbol{X}^{t:t+1}$ instead of $\boldsymbol{X}^{0:1}$ to make this more clear in the notation. Links are only allowed to originate in the $\boldsymbol{X}^t$ column and terminate in the $\boldsymbol{X}^{t+1}$ column.

We employ the following greedy structure search algorithm. For each X_i^{t+1} RV, the algorithm finds the parent node X_j^t that increases the DBN's score the most. It then finds the next parent, X_k^t, $k \neq j$, that increases the score the most in conjunction with the previously added parents. This process is repeated until a predetermined number of parents has been added to node X_i^{t+1}.

Multinets

Frequently, when using DBNs for class discrimination, a single DBN that includes a *class* node is learned. This is not optimal when the DBN's topology depends on the data's class. One fundamental weakness of the DBN framework (and the BN framework in general) is that the topology remains static, i.e., the existence of a link cannot change based on the values of other RVs. This decreases the comprehensibility of the DBN by requiring parameters to be analyzed in order to identify dependencies that only occur when RVs take on certain values.

Take for example, a non-temporal system containing four RVs, $\boldsymbol{X} = \{X_1, X_2, X_3, Y\}$. See Figure 1. Assume that X_3 depends on X_1 when $Y = 1$ but depends on X_2 when $Y = 2$. To model this system with a BN, the parent set of X_3 would include X_1, X_2 and Y. The conditional dependency is not apparent from the structure. Further, the number of parameters required to represent the BN is needlessly increased.

Instead of using a single BN, separate BNs can be used to model each class—B_1 for class 1 and B_2 for class 2. The parent set for X_3 would only include X_1 in B_1, and only X_2 in B_2. B, the collection of B_1 and B_2, *is* referred to as a multinet. The joint probability distribution (JPD) represented by B can be computed from B_1 and B_2's JPD,

$$P_B(\boldsymbol{X} = \boldsymbol{x}, Y = y) = P_{B_y}(\boldsymbol{X} = \boldsymbol{x}) \cdot P(Y = y). \quad (3)$$

Each class of data, Y ∈ {1,2}, is represented by a single dataset, D_1 and D_2. Each dataset (and in turn, each class) is represented by a DBN, B_1 for D_1 and B_2 for D_2. Dataset D_α is referred to as B_α's *intra-class* data, and D_β as B_α's *extra-class* data, $\beta \neq \alpha \in \{1,2\}$. An event $\xi^\beta_{i,j,k}$ is defined to be the occurrence within a data point $d \in D_\beta$ where $X_i^{t+1} = k$ and the set of X_i^{t+1}'s parent RVs (as defined by D_β's structure) are in their j^{th} configuration, $\forall\ t > 1$. $\eta^\beta_{i,j,k}$ is the number of times $\xi^\beta_{i,j,k}$ occurs.

Work in multinets has been formalized by Heckerman (1991) and Geiger and Heckerman (1996). They have also been specifically applied to DBNs by Bilmes (2000).

Structure Scores

To train generative models, we use Heckerman's BDE score (Heckerman, Geiger & Chickering 1995). The BDE score for DBN D_α is,

$$\prod_{i=1}^{n}\prod_{j=1}^{q_i}\left[\frac{\Gamma\left(\eta'^{\alpha}_{i,j}\right)}{\Gamma\left(\eta^{\alpha}_{i,j}+\eta'^{\alpha}_{i,j}\right)}\prod_{k=1}^{r_i}\frac{\Gamma\left(\eta^{\alpha}_{i,j,k}+\eta'^{\alpha}_{i,j,k}\right)}{\Gamma\left(\eta'^{\alpha}_{i,j,k}\right)}\right], \quad (4)$$

where a set of prior information is given by the virtual counts $\eta'^{\alpha}_{i,j}$ and $\eta'^{\alpha}_{i,j,k}$. We use uninformative priors by setting these values to r_i and 1, respectively.

The score that maximizes classification accuracy is the one that maximizes class conditional likelihood (CCL) (Duda & Hart, 1973). Unfortunately, CCL does not decompose into the aggregation of family scores. This yields application of CCL intractable for large BNs such as the ones we construct in the neuroscience domain. To overcome this limitation, Burge and Lane proposed a decomposable score, *approximate conditional likelihood* (ACL) (Burge & Lane 2005). The ACL score for DBM B, which consists of the two DBNs B_1 and B_2, is,

$$ACL(B \mid D) = ACL(B_1 \mid D) \cdot ACL(B_2 \mid D), \quad (5)$$

$$ACL(B_\alpha \mid D) = \prod_{d \in D_\alpha} P_{B_\alpha}\left(\boldsymbol{X}_d^{0:T}\right) \Big/ \prod_{d \in D_\beta} P_{B_\alpha}\left(\boldsymbol{X}_d^{0:T}\right), \alpha \neq \beta.$$

Unlike the BDE score, the ACL score is based on the choice of parameters for the DBNs. Thus, the method used to train the parameters is important. Burge and Lane experimented with three different parameter settings: the MLE parameters of the BN given the data, the parameters that maximize the ACL score and a smoothed version of the ACL-maximizing parameters, referred to as ACL-ML, ACL-Max and ACL-Mid, respectively. ACL-ML and ACL-Mid outperformed ACL-Max. For this work, we have decided to use the ACL-Mid parameters,

$$\Theta^{\alpha}_{i,j,k} = \frac{1}{\omega}\left[\left(\eta^{\alpha}_{i,j,k} - \eta^{\beta}_{i,j,k}\right) + \arg\min_k\left(\eta^{\alpha}_{i,j,k} - \eta^{\beta}_{i,j,k}\right) + \varepsilon\right]. \quad (6)$$

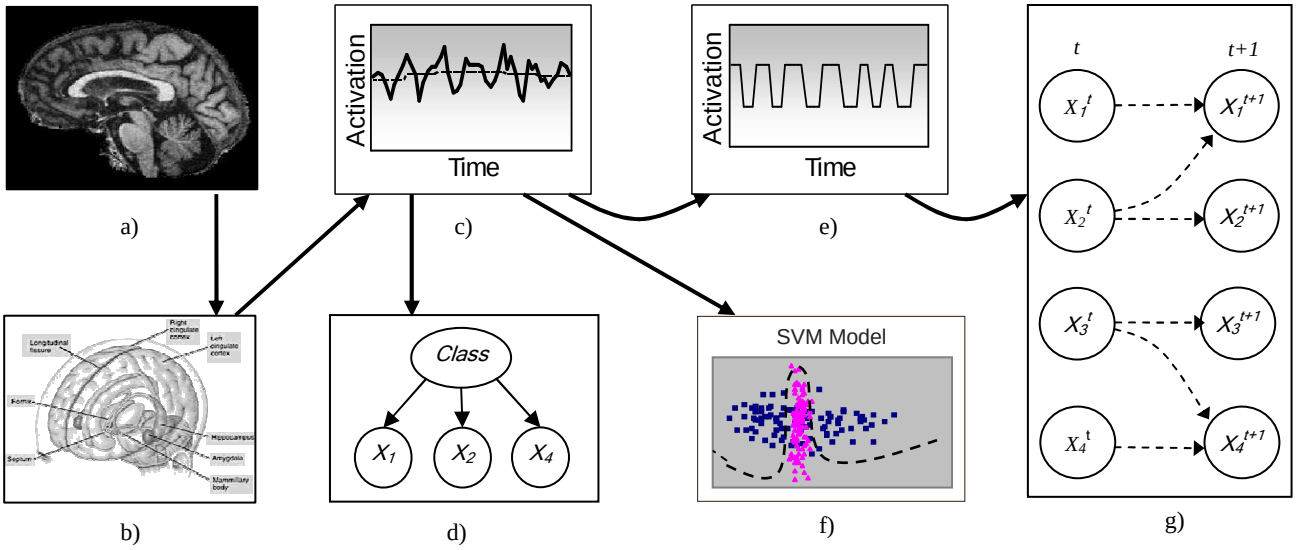

Figure 2. Experimental design flowchart. (a) Raw data is acquired for each patient. (b) The data is grouped into ROIs via the Talairach database. (c) The aggregate mean voxel activation (AMVA) of each ROI, and its sliding window trend, is computed. (d, f) The GNBN and SVM models are learnt from the AMVA time series. (e) The AMVA time series is quantized (two state quantization shown). (g) A discrete dynamic Bayesian network is learned for each class of data. The links are dashed to indicate that they were found during a structure search.

where ω is a normalization constant and ε is a Laplace smoothing constant. $\beta \neq \alpha \in \{1,2\}$.

Neuroanatomical Domain

Functional magnetic resonance imaging (fMRI) has recently become widely used in the study and diagnosis of metal illness. It is a non-invasive technique measuring the activity of small cubic regions of brain tissue (voxels). Psychologists frequently use fMRI data to test hypotheses about changing neural activity caused by mental illness.

An fMRI scanning session can result in hundreds of 3D images, each consisting of 65,000 or more voxels. As there is too much data collected to analyze directly, we abstract from the voxel level to a region of interest (ROI) level. To do this, we use the Talairach database (Lancaster et al. 2000) as it is widely accepted in the neuroscience community.

The database allows us to convert each 3D image into an activity snapshot detailing the momentary activation of 150 ROIs. After applying the database to each image, a detailed time series accounting for the activity of each ROI is built. Each ROI is treated as a temporal RV, and the system is modeled with a discrete stationary Markov order 1 DBN containing the nodes $\boldsymbol{X}^{t:t+1}=\{X_i^t, X_i^{t+1} : 1 \leq i \leq 150\}$.

Experiments

We analyze data collected by Buckner et al. (2000) in an experiment theorized to elicit different neural responses from healthy and demented elderly patients. Figure 2 shows our experimental layout. For each patient, fMRI images are taken while the patients undergo a visual recognition task. The resulting series of images are then converted into an activation time series for each ROI via the Talairach database. A support vector machine and a Gaussian naïve Bayesian network are then trained from the data. However, as the DBNs are discrete, the data must be quantized before the DBN structure search is employed.

To quantize the data, a sliding window mean for each ROI's activity is calculated and the ROI's activity is normalized around this mean. This removes low frequency drifts in the time series. Four discrete states are defined roughly corresponding to *high*, *very high*, *low* and *very low* activity. The continuous-valued time series is discretized based on how far the activity of the time series diverges from the sliding window mean at each point in time. Figure 2e shows a simple 2-state quantization for the time series in Figure 2c.

The data is then separated into two groups, one group for the demented patients and one group for the healthy patients. Generative and discriminative DBNs are learned for each group with the BDE and ACL scores, respectively.

Results

A DBN structure search results in a collection of family structures. With decomposable scores (such as BDE and ACL), each of the families in the DBN also has a score. This allows the families to be ranked in importance. We have found that two types of information are most helpful from the set of sorted families. 1) The parent set within each family, particularly high ranking families, and how that parent set changes between classes. Table I lists the parents for each of the DBN's five highest scoring families.

BDE Healthy	
ROI	Parents
Gray Matter	Cuneus
	Cingulate Gyrus
	Insula
	Brodmann area 28
Left Cerebrum	Cuneus
	Cingulate Gyrus
	Brodmann area 13
	Cerebellar Tonsil
Right Cerebrum	Right Cerebrum
	Cuneus
	Dentate
	Uncus
Brodmann area 6	Occipital Lobe
	Brodmann area 6
	Cerebellar Tonsil
	Subcallosal Gyrus
Brodmann area 4	Brodmann area 4
	Occipital Lobe
	Pyramis
	Rectal Gyrus

BDE Demented	
ROI	Parents
Parietal Lobe	Parietal Lobe
	Amygdala
	Uvula of Vermis
	Caudate Tail
BA 40	Parietal Lobe
	Amygdala
	Uvula of Vermis
	Lateral Dorsal Nucleus
BA 7	Inferior Parietal Lobule
	Declive
	Amygdala
	Inf. Semi-Lunar Lobule
Inf. Parietal Lobe	Inferior Parietal Lobule
	Uvula of Vermis
	Occipital Lobe
	Amygdala
BA 10	Brodmann area 10
	Brodmann area 33
	Cerebellar Tonsil
	Lateral Dorsal Nucleus

ACL Healthy	
ROI	Parents
Insula	BA 38
Med. Dorsal Nucleus	Lat. Globus Pallidus
	Lentiform Nucleus
Ventral Ant. Nucleus	Lat. Globus Pallidus
	Lentiform Nucleus
BA 13	BA 38
Angular Gyrus	Caudate Head
	Anterior Cingulate

ACL Demented	
ROI	Parents
BA 13	Thalamus
	BA 42
Insula	Thalamus
	BA 42
Amygdala	Caudate Body
Parahippocampal Gyrus	Mid Temporal Gyrus
Anterior Lobe	Mid Temporal Gyrus

Table I. The five top scoring families for the healthy and demented DBNs trained with the BDE or ACL score. The parents are listed in the order they were added to the family in the greedy structure search heuristic. The maximum number of parents allowed per family is four, but the ACL score generally resulted in fewer parents.

2) The number of times a ROI is found as a parent. Table II lists the top ten most common parents in each of the DBNs.

High Scoring Families

Families with higher scores will describe stronger relationships than families with weaker scores, though the nature of the relationship depends on the score used. The generative models are most useful in understanding strong relationships among ROIs, but they are not specifically trained to differentiate between classes.

For instance, in the healthy patient generative model, Brodmann areas (BAs) 4 and 6 were found as two of the top five highest scoring families. This indicates that their behavior was highly correlated to the behavior of their corresponding parents. BA 4 and BA 6 are the primary and premotor ROIs (respectively). Further, the occipital lobe, the location of the visual cortex, was a parent for both these ROIs. This allows neuroscientists to infer that a patient's visual processing is correlated with subsequent motor activity. This relationship is expected given the nature of the underlying experiment: recognition of a visual stimuli followed by pressing a button.

Finding meaningful relationships such as this one allows our collaborating neuroscientist to validate the generative model's results. Conversely, within healthy discriminative models, BA 4 and BA 6 are not both high scoring families. While BA 6 was the 12^{th} highest scoring family, BA 4 was the 109^{th}. This indicates that the behavior of BA 4 and BA 6 did not dramatically change between healthy and demented patients. Since many expected relationships among ROIs will be equally likely in both classes, they will not be included in discriminative structures. As a result, we have found that generative structural results are easier for our collaborating neuroscientist to compare with putative neuroanatomical relationships.

The order in which parents are found in the search is also important. Parents added earlier tend to have stronger relationships with the child. Further, due to the greedy search heuristic, all parents added after the first parent may or may not have a strong pair-wise correlation with the child. In fact, the first parent is the only one guaranteed to have a strong pair-wise correlation with the child. All other parents are only guaranteed to be correlated with the child given the value of previously selected parents. This has the unfortunate effect of reducing the comprehensibility of the models.

For example, assume ROI *X* and *Y* are parents for ROI *Z* such that *X* was the first parent added and *Y* was the second parent added. It could be that *Y* is completely uncorrelated with *Z* but that the combination of *X* and *Y* is highly correlated with *Z*. It is likely that domain experts fail to pick up on this subtle nuance and would tend to treat an ROI's parents as a list of ROIs who are all highly correlated with the child.

Even though generative scores select for structures that are highly likely given just a single class's data, differences between each class's resulting structures can be helpful in identifying differences between the classes. If a highly likely relationship exists in one class's data but not the other, the relationship will be found by only one of the class models. Neuroscientists can then infer a changing relationship between an ROI and its parents based on differing parent sets. For instance, four of the top five

BDE Demented						**ACL Demented**					
	N	1st	2nd	3rd	4th		N	1st	2nd	3rd	4th
Amygdala	77	6	10	30	31	Amygdala	37	32	5	0	0
Inferior Semi-Lunar Lobule	52	0	4	20	28	Brodmann area 42	26	13	12	1	0
Brodmann area 31	45	43	2	0	0	Parahippocampal Gyrus	23	15	6	2	0
Lateral Dorsal Nucleus	36	0	8	9	19	Medial Frontal Gyrus	20	6	13	1	0
Cerebellar Tonsil	30	0	5	15	10	Middle Frontal Gyrus	18	0	17	1	0
Uvula of Vermis	27	0	2	10	15	Thalamus	14	6	8	0	0
Uncus	17	3	2	4	8	Medial Dorsal Nucleus	11	7	4	0	0
Pyramis	17	1	8	7	1	Insula	7	6	1	0	0
Precuneus	11	10	1	0	0	Middle Temporal Gyrus	7	6	1	0	0
Insula	10	4	6	0	0	Brodmann area 47	7	0	2	5	0
BDE Healthy						**ACL Healthy**					
Cerebellar Tonsil	46	2	3	22	19	Amygdala	92	92	0	0	0
Dentate	33	2	3	8	20	Parahippocampal Gyrus	20	2	4	14	0
Rectal Gyrus	23	0	1	11	11	Inferior Occipital Gyrus	18	1	8	9	0
Brodmann area 38	22	0	15	5	2	Brodmann area 38	11	6	4	1	0
Subcallosal Gyrus	19	0	2	4	13	Brodmann area 10	9	0	7	2	0
Brodmann area 28	19	0	6	6	7	Vent. Post. Med. Nucleus	9	0	7	2	0
Cuneus	17	11	6	0	0	Caudate Tail	9	8	1	0	0
Insula	16	10	5	1	0	Thalamus	7	0	4	3	0
Uvula of Vermis	15	0	2	4	9	Middle Temporal Gyrus	7	1	3	3	0
Brodmann area 23	14	10	4	0	0	Brodmann area 36	7	2	5	0	0

Table II. The top ten most common parent ROIs. *N* is the number of times the ROI was a parent. The remaining columns indicate how many times the ROI was selected as the 1st, 2nd, 3rd or 4th parent. The BDE families had exactly 4 parents but the ACL families frequently selected only one or two (as additional parents did not increase the score).

families in the demented class's model specifically include portions of the parietal lobe (the parietal lobe itself, the inferior parietal lobe, BA 40 and BA 7), whereas none of the top families in the healthy model do. This suggests that parietal families may become more involved in visual-motor processing in early stages of dementia, as temporal brain regions become compromised. In addition, most of these regions do not receive input directly from posterior visual areas, but instead from a variety of parietal, cerebellar, and sub-cortical regions, as well as the amygdala.

However, the ability for generative models to differentiate between classes is limited to finding relationships that are highly likely in at least one of the class's data. Relationships that are relatively weak in both classes of data will not achieve high scores in the generative models regardless of whether the they change between classes. For instance, in the demented discriminative models, the parahippocampal gyrus has the 4th highest score. This suggests that the behavior of the parahippocampal gyrus's changes significantly in demented patients. This ROI is known to degenerate quickly in the presence of dementia. Yet, it was only the 96th highest scoring family in the healthy generative model since its relationship to its parents was significantly weaker in magnitude (as measured by the BDE score) than the relationships in many other families.

Parent Statistics

While specific ROIs' parent sets are informative, we have found the most comprehensible set of information to be the number of times an ROI has been a parent throughout a class's entire DBN. Table II lists the number of times each ROI is a parent in the generative and discriminative models. In generative models, if an ROI is found as a parent in many families, that ROI likely plays a significant role in the mental activity the class's corresponding population. In discriminative models, it indicates that an ROI's behavior significantly changes between classes.

The most profound neuroanatomical results involve the amygdala, an ROI responsible for emotional response. In the demented generative model, the amygdala shows up as a parent 77 times—that is more than half of the families! This is especially profound given that in the healthy generative model, it only shows up as a parent once. The structural results from the generative models clearly indicate a changing amygdalar relationship.

However, of the 77 families that the amygdala was a parent in, it was chosen as the third or fourth parent 61 times. In those families, the relationship among the child and two or three other parents was stronger than it was between the child and the amygdala. From these results, we cannot even be assured that the amygdala has a strong pair-wise correlation with many of its children. Further, if the structure search had been limited to finding just two parents (possibly due to computational and time constraints), the importance of the amygdala would have been obscured. We have also found that depending on the BDE parameter priors used, the structure search may actually stop improving the score before the third and fourth parents are added, further increasing the possibility the amygdalar relationships would be missed.

In the discriminative models, not only is the amygdala the most prevalent parent, but it is overwhelmingly chosen as the first parent. This resolves the ambiguity inherent in the generative models of whether the amygdala has a pair-wise correlation in its children. However, the interpretation of the amygdala as a parent in the discriminative models is more complex than in the generative models. The discriminative models indicate that the behavior of the amygdala dramatically changes between healthy and demented patients, but the nature of that change is not clear. Based solely on the discriminative results, it is not possible to tell whether the activity of the amygdala is more pronounced in healthy or demented patients. The generative structures clarify this ambiguity. As previously mentioned, in these structures the amygdala is frequently a parent in the demented networks, but not in the healthy networks. This indicates that the amygdala's behavior is significantly more correlated with other ROIs in the demented population.

Thus, combination of amygdalar results from both the discriminative and generative models is what was most useful.

SVM and GNBN Results

Previously, we also experimented with the ability for support vector machines (SVMs) and Gaussian naïve Bayesian networks (GNBNs) to classify dementia (Burge et al. 2005) based on the fMRI data. Overall, we found that the SVMs classified with the worst accuracy and that the GNBNs classified with an accuracy somewhere between the generative and descriptive models' accuracies; an impressive feat given the simplicity of the model.

But what do the results of these models indicate about the underlying neuroanatomical dynamics that change due to the onset of dementia? The GNBN has a static non-temporal structure with the class RV set as a parent of each of the ROI variables. The ROI RVs, X_i, can be ranked based on $P(X_i|Y,D)$ to determine which RVs were most strongly correlated with the class RV, however, we have found these results are not particularly helpful for our collaborating neuroscientist.

The results from the SVM are even less comprehensible. An SVM creates a hyperplane classification boundary in a high dimensional space based on the location of a set of support vectors. The SVM results indicated which of the patients were the most important in determining the classification boundary (i.e., which patients' data points corresponded to support vectors) but we found it difficult to associate underlying neuroanatomical behavior with the SVM's classification decisions.

Conclusions

We have examined the comprehensibility of dynamic Bayesian multinets (DBMs) and the dynamic Bayesian networks (DBNs) that compose them. Specifically, we have used DBMs to model the changing relationships among neuroanatomical regions of interest (ROIs) due to the onset of senile dementia.

We divided DBN scoring metrics into two categories: generative and class discriminative. Generative scoring functions were found to favor structures representing relationships that were highly likely given a single class's data. While this is not ideal for class discrimination, we have found the resulting structures to be quite helpful given their highly comprehensible nature. In particular, our collaborating neuroscientist found it easy to validate the generative results with putative neuroanatomical relationships.

Conversely, class discriminative scores favored structures representing relationships that discriminated between classes. These structures illuminated changing ROI relationships that were too subtle to be identified by the generative models. Indeed, classifiers based on discriminative models achieved higher accuracies than those based on generative models. However, interpretation of discriminative structures is not as straightforward as that of generative structures.

When analyzing the structure for both types of models, two types of information were most helpful. 1) The parent set for each ROI and how that set changed between classes. 2) The number of times a ROI was found as a parent in the entire DBN. Further, the order in which parents were found by the greedy structure search was important as only the first parent added was guaranteed to have a high pair-wise correlation with the child, a subtlety that we suspect is easily overlooked by domain experts.

Overall, we found that neither generative nor discriminative models were more helpful than the other but instead that a combination of results from both models was most meaningful.

Future Work

Regulatory systems are common in neural networks and the identification of feedback systems is an important issue in the neuroscience domain. Unfortunately, identification of feedback loops from DBNs making the stationary and Markov order one assumptions is difficult. Several problems arise, illustrated by Figure 3.

First, since the RVs in the DBNs represent the average temporal relationship between ROIs across the entire time series, it is difficult to determine if the relationship between two RVs temporally precedes a relationship between two other RVs. For example, the DBN in Figure 3a suggests there may be the feedback loop A→B→C→D→A. But it is also possible that the relationship between B and C does not occur after the relationship between A and B. This would indicate that the structure does not represent a feedback loop.

A second issue arises when many possible feedback loops can be inferred from a single structure. For example, the DBN in Figure 3b has many possible loops. But, which loops are indicative of actual feedback loops in the underlying system? In a large DBN with 300+ nodes, there

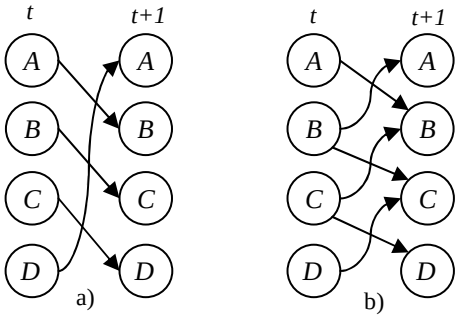

Figure 3. Identification of feedback loops in DBN structures is difficult. a) A structure with the possible feedback loop A→B→C→D→A. The structure alone cannot indicate whether this is truly a feedback loop. b) A structure with many possible feedback loops, e.g., A→B→A, B→C→B, A→B→C→D→C→B→A, etc.

may even be too many candidate feedback loops to realistically enumerate. The score used to select the structure also plays an important role. If a discriminative score is used, portions of the feedback loop that do not change between classes may not even appear in the structure.

Finally, we are interested in having our collaborating neuroscientist provide information that helps guide the DBN structure searches. In particular, we would like to take advantage of the hierarchical arrangement of ROIs in the brain. The approximately 150 ROIs defined by the Talairach database are arranged in 5 hierarchical levels. Large ROIs, such as the left and right cerebrum and cerebellum, are located in the topmost level. Small ROIs, such as the amygdala and the Brodmann areas, are located at the bottommost level. Currently, nodes in the DBNs are allowed to have parents from any level of the hierarchy. But many relationships that span hierarchical levels may be uninteresting to neuroscientists. Through interaction with our collaborating neuroscientist, we can eliminate some relationships that are not likely to be helpful and hopefully improve the structure search results. Unfortunately, given the sheer number of possible relationships, this type of interaction has not been significantly helpful in guiding structure search and further attempts to incorporate input from our collaborating neuroscientist continue.

Acknowledgements

We would like to thank Dr. Vincent P. Clark, our collaborating neuroscientist, for his invaluable assistance and analysis of our structural results. This project was partially funded through grant DA012852 from the National Institute of Drug Abuse, NIH. John Burge and Vincent P. Clark were supported in part by The MIND Institute.

References

Bilmes, J.A. Dynamic Bayesian Multinets. 2000. *Proceedings of the 16th conference on Uncertainty in Artificial Intelligence*. pg. 38-45.

Bilmes, J., Zweig, G., Richardson, T., Filali, K., Livescu, K., Xu, P., Jackson, K., Brandman, Y., Sandness, E., Holtz, E., Torres, J., & Byrne, B. 2001. Discriminatively structured graphical models for speech recognition (Tech. Report). Center for Language and Speech Processing, Johns Hopkins Univ., Baltimore, MD.

Buckner, R. L., Snyder, A., Sanders, A., Marcus, R., Morris, J. 2000. Functional Brain Imaging of Young, Nondemented, and Demented Older Adults. *Journal of Cognitive Neuroscience*, 12, 2. 24-34.

Burge, J., Clark, V.P., Lane, T., Link, H., Qiu, S. 2005 Bayesian Classification of fMRI Data: Evidence for Altered Neural Networks in Dementia. In submission to *Human Brain Mapping*. Alternate reference: Tech report TR-CS-2004-28, University of New Mexico, Department of Computer Science.

Burge, J., Lane, T. 2005. Learning Class Discriminative Dynamic Bayesian Networks. In submission to the *International Conference on Machine Learning*.

Cooper, G., & Herskovits, E. 1992. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9, 309–347.

Charniak, E. 1991. Bayesian Networks Without Tears. *AI Magazine*, 12 (4).

Chickering, D., Geiger, D., Heckerman, D. 1994. Learning Bayesian Networks is NP-Hard. Technical Report MSR-TR-94-17. Microsoft.

Duda, R. O., Hart, P. E. 1973. Pattern classification and scene analysis. New York, NY: Wiley.

Grossman, D., Domingos, P. 2004. Learning Bayesian Network Classifiers by Maximizing Conditional Likelihood. *International Conference on Machine Learning*, 21. 361-368

Heckerman, D. 1991. Probability Similarity Networks. MIT Press, 1991.

Heckerman, D., Geiger, D., Chickering, D.M. 1995. Learning Bayesian Networks: the Combination of Knowledge and Statistical Data. *Machine Learning*, 20, 197-243.

Jensen, F. V., 2001. Bayesian Networks and Decision Graphs. Springer-Verlag, New York.

Lam, W., Bacchus, F. 1992. Learning Bayesian Belief Networks An approach based on the MDL Principle. *Computational Intelligence*, Vol. 10, pg 269-293.

Lancaster J.L., Woldorff M.G., Parsons L.M., Liotti M., Freitas C.S., Rainey L., Kochunov PV, Nickerson D., Mikiten S.A., Fox P.T. 2000. Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping* 10,120-131.

Murphy, K. 2002. Dynamic Bayesian Networks: Representation, Inference and Learning. PhD dissertation, Berkeley, University of California, Computer Science Division.