Infinite dynamic bayesian networks

Infinite Dynamic Bayesian Networks

Finale Doshi-Velez FINALE@MIT.EDU
David Wingate WINGATED@MIT.EDU
Joshua Tenenbaum JBT@MIT.EDU
Nicholas Roy NICKROY@MIT.EDU
Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139 USA

Abstract

We present the infinite dynamic Bayesian network model (iDBN), a nonparametric, factored state-space model that generalizes dynamic Bayesian networks (DBNs). The iDBN can infer every aspect of a DBN: the number of hidden factors, the number of values each factor can take, and (arbitrarily complex) connections and conditionals between factors and observations. In this way, the iDBN generalizes other nonparametric state space models, which until now generally focused on binary hidden nodes and more restricted connection structures. We show how this new prior allows us to find interesting structure in benchmark tests and on two real-world datasets involving weather data and neural information flow networks.

1. Introduction

Inferring structure in timeseries data has applications ranging from modeling neurological signals, processing language, and predicting weather patterns. For example, given data from neurological probes, we may wish to infer how different areas of the brain communicate or what the subject was doing; given meteorological data we may wish to infer regional patterns of weather. Learning causal structures of the world is also important for creating adaptive agents (as in reinforcement learning or robotics), where agents can use their world model for planning and control. In all of these examples,aspects of the world may be hidden from the agent: for example, the results of an agent's movement may depend on what (unknown) room it is in, rather than the agent's immediately observed surroundings.

Appearing in *Proceedings of the 28th International Conference on Machine Learning*, Bellevue, WA, USA, 2011.

Hidden Markov models (HMMs) provide one approach for modeling time-series data with hidden states. The HMM posits that, at every time step, there is a single hidden state (which can take on many values) that explains the observed data and determines how the state will evolve in the next time step. Dynamic Bayesian networks (DBNs) extend HMMs by encoding structure: they posit that there are a number of hidden variables at every time step, each of which can affect the observed data and causally affect the hidden nodes at the next time step. While a DBN can always be flattened into an HMM in which a single hidden state encodes the values of all the DBN's factors, the factored representation often allows for more efficient inference. The DBN's more structured explanation of the observed variables may also have inherent interest.

In some applications, the number of these hidden factors and their values may be known: for example, whether the robot moves as commanded may depend on hidden factors such as the state of its motors and brakes; we may even know that these factors have two states—on or off. Much work exists on learning DBN structure if all nodes are observed; work that allows for missing data (Ghahramani, 1998; Xing-Chen et al., 2007; Peña et al., 2005) still assumes knowledge about the number of hidden nodes and their values. However, in general it may be unclear how many hidden nodes are needed to explain the observed data, how they are connected, or even what values they may take.

Nonparametric extensions of the DBN have attempted to capture various structure in the data. The Infinite Factored HMM (Van Gael et al., 2009) posits that there are a potentially unbounded number of binary factors that explain the observed data, while the Infinite Hierarchical HMM (Heller et al., 2009) posits that there are a potentially unbounded number of discrete-valued factors that explain the observed data. Both of these models assume a fixed dependency structure: the iFHMM assumes that each factor evolves independently, while the iHHMM assumes that each factor is

affected by itself and a factor one level above it at the previous time step. The Infinite Latent Events Model (Wingate et al., 2009) posits that there are binary factors that can have time-varying sequences of causes. The Adaptive Dynamic Bayesian network (Ng, 2007) allows each factor to take on an unbounded number of values but assumes a fixed number of factors.

Generalizing these models, we present the Infinite DBN (iDBN), a nonparametric time-series model with a flexible number of factors, factor values, and factor connections. The model allows each factor to take on an arbitrary number of values (learned) as well as be connected in an arbitrary fashion to previous nodes (also learned). Setting concentration parameters lets designers manage trade-offs between models with more states and models with more factors without the hard model constraints assumed in previous work.

2. Dynamic Bayesian Networks

A regular dynamic Bayesian network (DBN) is a directed graphical model in which the state X_t at time t is represented through a set of factors $\{x_t^1, x_t^2, \ldots, x_t^K\}$. The value of a node—or state—x_{t+1}^k at time $t+1$ is sampled from $T(x_{t+1}^k | Pa_k(X^t))$, where $Pa_k(X^t)$ represents values of the parents of node k at time t. The parents of a node always come only from the previous time slice (there are no intra-slice connections).

The state of a DBN is generally hidden; values of the states must be inferred from a set of observed nodes $Y_t = \{y_t^1, y_t^2, \ldots, y_N^t\}$. The value of an observation y_t^n at time t is sampled from $\Omega(y_t^n | Pa_n(X^t))$, where $Pa_n(X^t)$ represents values of the parents of observed node n at time t. The parents of observed nodes at time t are hidden nodes at time t; given the values of the hidden nodes, the observations at different time steps are independent (see Murphy, 2002).

3. Infinite Dynamic Bayesian Networks

If the hidden factors in a DBN are truly hidden, knowing how many hidden factors exist may also be unknown. Our nonparametric DBN model places a prior over DBNs with unbounded numbers of hidden factors. Inference on this infinite structure is tractable only if the prior ensures that only a finite number of hidden nodes will be needed to explain a finite sample of time-series data. More generally, the following properties are desirable in a general nonparametric DBN model:

- A finite dataset should be generated by a finite number of hidden factors with probability one.
- The structure connecting the hidden nodes should be as general as possible (we do not wish to enforce a particular form of connections as the hierarchical or factorial HMM do).
- Each node should be able to take on multiple values (we do not wish to limit ourselves to binary nodes).

The first desideratum requires particular attention depending on how the hidden nodes at one time slice affect nodes at the next: care must be taken to ensure that inference for any particular hidden node k at time $t+1$ does not require knowing the values of an infinite number of hidden nodes at time t. There exist, of course, many priors that satisfy these desiderata; we present one here with an eye toward tractable inference. Our infinite DBN (iDBN) model posits that the world actually has an infinite number of hidden factors x_k^t at any time t. Only a finite number of factors are needed to explain a finite set of observed nodes; however, as we attempt to model more observed nodes, we expect that more hidden nodes will be required to explain the data.

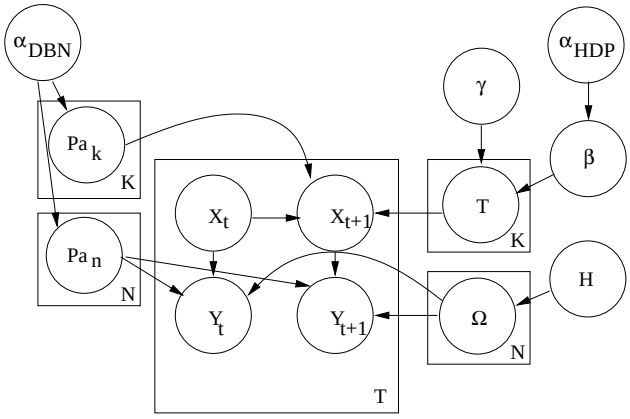

Figure 1. Graphical model for the iDBN prior: one concentration parameter, α_{DBN} controls the structure of the connections, while a second, α_{HDP} controls the number of values each hidden node is expected to take in a finite time-series.

The generative process, summarized in Fig. 1, for our iDBN model proceeds as follows: first, for each observed node, the generative model draws zero or more parent hidden factors via a non-parametric process with hyper-parameter α_{DBN}

$$Pa_n \sim \mathsf{NP}(\alpha_{DBN}) \tag{1}$$

where $Pa_n(k) = 1$ if hidden node k is a parent of observed node n. Once the observed nodes have chosen parents, all parent hidden nodes choose their own parent nodes via the same process:

$$Pa_k \sim \mathsf{NP}(\alpha_{DBN}) \tag{2}$$

where $Pa_k(j) = 1$ if hidden node j is a parent of hidden node k. This process is repeated for any newly instantiated parents until all hidden nodes that may affect the observed nodes have been instantiated. For example, suppose that there is only one observed node n, and it chooses hidden nodes i and j as its parents. Next, nodes i and j would choose their parents: suppose node i chooses only node j, but node j chooses itself and a new node k. Then we would have to again sample parents for node k: suppose it chooses nodes i and j. At this point, all nodes' parents are already-instantiated nodes, and we have a finite set of nodes (i,j,k) that are needed to predict observed node n.

The process NP should have a rich-get-richer property such that (1) nodes choose a finite number of parents with probability one and (2) when a new node is choosing its parents, there is always a finite probability that it not choose any new (uninstantiated parents). In this work, we use the Indian Buffet Process (IBP) (Griffiths & Ghahramani, 2005) as our nonparametric process NP. In the IBP, the n^{th} factor (the "customer") chooses $\mathsf{Poisson}(\alpha/n)$ new parents ("dishes"). The probability that the factor chooses no new parents is $\exp(-\alpha/n)$. Fig. 2 shows the how, when using the IBP as NP, the expected number of hidden factors grows logarithmically with the dimensions of the observation.

However, any nonparametric process satisfying (1) and (2) above will ensure number of observed nodes will be explained by a finite number of hidden nodes:

Proposition 1. *If* NP *is a nonparametric process such that the k^{th} node selects a new node chooses a new (uninstantiated) parent with probability less than some constant c for all k greater than some constant K, then the DBN is guaranteed to have a finite number of nodes with probability one.*

Proof. Once a new node selects no new parents, the process for growing the part of the DBN relevant to the observations is complete. Suppose that the probability that a new (uninstantiated) parent is chosen is always less than c after K nodes have already been instantiated. Then the distribution of number of new parent nodes that will be added to the DBN is dominated by a geometric distribution with parameter c. Since a geometric distribution outputs a finite value with probability one, only a finite number of nodes will be instantiated with probability one. □

Finally, we note that this process for sampling inter-factor connections is closely related to the cascading Indian Buffet Process (cIBP) (Adams et al., 2010). The key difference between the two structure models is that the cIBP uses an IBP used to winnow the number of factors in each layer of a deep belief network, while the iDBN uses its nonparametric prior to winnow the number of parents in a flatter two-layer network (representing the current and future time slices).

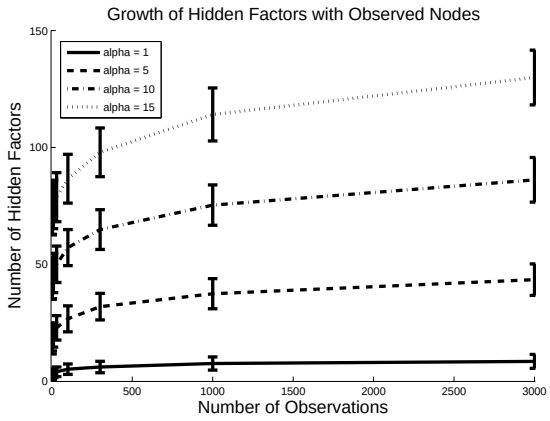

Figure 2. Expected number of hidden factors given different numbers of observed nodes, for varying alpha.

Once the connections of the iDBN are specified, the next step in specifying the iDBN model is describing the prior over transition distributions $T(x^k_{t+1}|Pa_k(X^t))$ and the emission distributions $\Omega(y^n_t|Pa_n(X^t))$. For the emission distribution, we simply specify some base distribution H_n for each observed node. For the transition distributions, we use the hierarchical construction of the hierarchical Dirichlet process HMM (HDP-HMM) (Teh et al., 2006): we first sample a base, or expected, transition distribution β from a Dirichlet process prior, and then use that distribution β as the base distribution for each transition distribution $T(x^k_{t+1}|Pa_k(X^t))$.[1]

The complete generative process for the iDBN prior is as follows:

- Sample parents X for all observed nodes y_n according to some nonparametric process $\mathsf{NP}(\alpha_{DBN})$: $Pa_n \sim \mathsf{NP}(\alpha_{DBN})$, where α_{DBN} is the concentration parameter of the nonparametric process.
- While there exist hidden nodes x_k without assigned parents, sample parents for them via the same process $\mathsf{NP}(\alpha_{DBN})$: $Pa_k \sim \mathsf{NP}(\alpha_{DBN})$.
- For each observed node y_n, sample emission dis-

[1]For simplicity, we used the same base distribution β for all hidden nodes k. While it may appear restrictive, evidence from the data still allowed the transitions T to vary; if needed, the hierarchy could easily be extended to sample a private base distribution β_k for each hidden node.

tributions $\Omega(y_t^n|Pa_n(X^t)) \sim H_n$ for each setting of the parent variables $Pa_n(X)$.

- Sample a base transition distribution $\beta \sim \mathsf{Stick}(\alpha_{HDP})$, where α_{HDP} is the concentration parameter of the base transition distribution.
- For each hidden node x_k, sample a transition distribution $T(x_{t+1}^k|Pa_k(X^t)) \sim \mathsf{DP}(\beta, \gamma)$, where γ is the concentration parameter for sampling the individual transition distributions.

Besides the properties induced by the nonparametric process, the choices of the concentration parameters adjust the biases of the iDBN prior regarding the number of hidden nodes (α_{DBN}), the number of values (or states) that a hidden node is likely to take (α_{HDP}), and the determinism of the transitions (γ). We also note that while the iDBN prior ensures that a finite number of hidden nodes will explain a finite number of observed nodes, as time goes on, those hidden nodes may take on new values (as sampled from the HDP prior on transitions) to explain new trends in the observations.

4. Inference

We sample potential DBNs from the iDBN posterior by cyclically resampling each of the hidden variables—the hidden factors X, the parent structure for the hidden nodes Pa_k and the observed nodes Pa_n, the base transition distribution β, and the transition and emission distributions T and Ω—one at a time conditioned on all of the other variables. Throughout this section and paper, we use the IBP as our nonparametric prior NP because of its straight-forward inference properties.

Resampling structure. We separate the process of resampling the structure Pa_n and Pa_k into two parts: resampling connections for already instantiated nodes and changing the number of hidden factors. Given the hidden state sequence X, it is straightforward to integrate out the transition or emission distribution and compute the probability of the hidden state sequence with or without an already-instantiated node as a parent (Heckerman, 1995), so that $p(Pa_n|Pa_k, X, \beta, T, \Omega, Y) = p(Pa_n|Pa_k, X, \beta)$ and $p(Pa_k|Pa_n, X, \beta, T, \Omega, Y) = p(Pa_k|Pa_n, X, \beta)$.

To add or delete factors, we use a Metropolis Hastings (MH) birth-death move of the following form:

- Choose whether to attempt adding or deleting a node with probability $p = .5$.
- If attempting to delete a node: only delete node whose hidden state sequences are constant.
- If attempting to add a node: add a node whose hidden state sequence is constant and connect it to existing nodes with probability p.

Computing the prior probability $p(Pa_k, Pa_n, \beta, T, \Omega)$ of the structure following this MH move is straightforward because adding or removing a node with a constant state sequence affects the structure of DBN but not the likelihood of the model with respect to the observations.

In addition to the MH-step above, we also sampled hidden state sequences from the iDBN prior for nodes unconnected to the currently-instantiated nodes; in following iterations these nodes may be connected to instantiated nodes that influence the observed nodes. Finally, we deleted a hidden node if it does not connect to an observed node or hidden nodes affecting observed nodes. While these hidden nodes are still part of the infinite DBN structure—and could have been connected to other nodes later—keeping them instantiated induced significant computational overhead.

Resampling transitions and observations. We now turn to resampling the parameters of the transition and emission distributions, $p(T|Pa_k, X, \beta)$ and $p(\Omega|Pa_n, X, \beta, Y)$, as well as the base transition distribution $p(\beta|Pa_k, X)$. The base transition vector β is infinite-dimensional; following Teh et al. (2006), we store it as $\{\beta_1, \beta_2, \ldots, \beta_N, \beta_u\}$, where each β_n is the base probability for some visited state n. The base probability of visiting any of the (infinite) unvisited states is β_u. We resample β using the restaurant-based sampler of Fox et al. (2010). Given the finite representation of β and the hidden node sequence X, resampling the transition distributions T is straightforward using Dirichlet-multinomial conjugacy; we can similarly resample the emission distributions Ω given the prior H_n and counts from the observed and hidden nodes. In the iDBN setting, where each hidden node can take on an infinite number of values, we obviously cannot sample distributions for all parent settings of a particular node. Instead, we only sample distributions for which we have data; additional distributions are instantiated on-demand as the sampler resamples the hidden node sequence.

Resampling states. Finally, we must resample the hidden node sequence $p(X|Pa_n, Pa_k, \beta, T, \Omega, Y)$. While exact inference in DBNs is generally computationally intractable, many approximation algorithms exist for inference over the hidden nodes. We applied the factored frontier algorithm (Murphy & Weiss, 2001), a form of loopy belief propagation with a forward-backward message-passing schedule. By representing the belief over states at every time step as a product of node marginals, the factored frontier adds one more approximation to our trun-

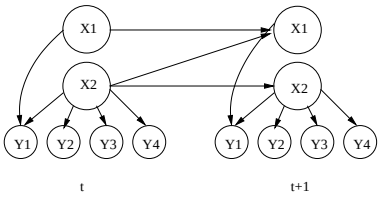

Figure 3. Simple DBN with 2 hidden nodes.

cated representation of β that groups all unvisited states into one extra node. However, we found no empirical difference between this computationally-efficient approximation and inference using a particle smoother (Doucet & Johansen, 2009) that did not require such an approximation to perform inference over the infinite-dimensional state space. We also found that the inference over the state sequence is required for structure-learning algorithms for finite DBNs (e.g. (Ghahramani, 1998)) as well; in our experiments almost 90% of the computational time was spent in this step. Thus, the iDBN prior does not add significant overhead to the inference.

5. Properties of the iDBN

We demonstrate various properties of using the iDBN prior using the simple DBN with 2 hidden nodes and 4 observed nodes (see Fig. 3). Fig. 4 plots the negative predictive log-likelihood of the finite DBN models and the iDBN on held-out test data, where the predictive likelihoods were computed by holding out 10% of the data from a time-series with 250 time-steps. Error bars show the standard error of the mean averaged from five 50-iteration runs of the sampler. As expected, increasing the number of hidden nodes helps initially because the flat model cannot fully explain the data. However, the larger finite models overfit the training data and thus make larger errors on the held-out test data. The iDBN prior infers a distribution over the number of hidden nodes (right pane of Fig. 4) and node values that generalizes to predict the held-out data well.

Many explanations can exist for a given sequence of observations: for example, suppose the "true" underlying model had 2 hidden nodes which took on 2 and 3 state values, respectively. While it would lose the structure, the model could also be represented by a flat model with a single hidden node with 6 state values. In Fig. 5, we show how adjusting the α_{DBN} and α_{HDP} in the iDBN prior biases the posterior toward more factors and more states, respectively. As expected, the number of hidden factors in the posterior increases with α_{DBN}, while the number of states the hidden factors taken on increases with α_{HDP} (though

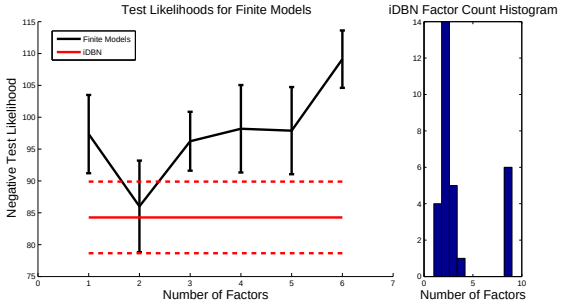

Figure 4. Negative log likelihoods for finite models compared to the iDBN. Error bars show the standard error of the mean.

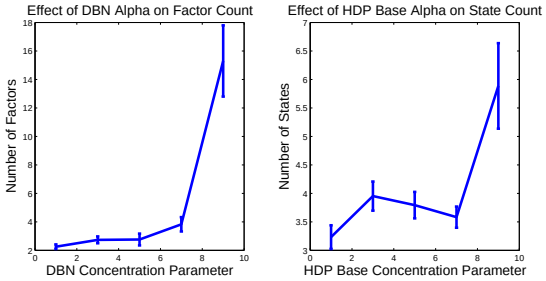

Figure 5. Number of hidden factors and number of states discovered by the iDBN; errorbars indicate one standard error of the mean.

with less sensitivity). However, the number of unique factor-state settings in the sequences' posterior stayed within a small range; changing the concentration parameters made biases for different structures but the posterior still captured the core variations in the data.

Overall, we found that good test likelihoods could be obtained over a variety of different concentration parameters. Over the parameter settings, the interquartile range for the test likelihoods was 21.8, suggesting that the iDBN could find a variety of likely models based on the biases given by the designer. Moreover, when empirically tested, using the same base distribution β for all of the transition distributions did not seem to be overly restrictive: the evidence from the data was able to shape the individual transition distributions to reasonable values.

6. Experiments

We first show that the iDBN prior generalizes well on several synthetic datasets from the literature. Next, we demonstrate how the iDBN produces interesting structure that can provide insights into the causality

patterns of time-series data. For all of the experiments, we set $\alpha_{DBN} = 1$, $\alpha_{HDP} = 1$, and $\gamma = 3$. The base emission distribution H_N was set to a uniform distribution with concentration 2. All three approaches compared—the iDBN, the iFHMM, and the DBN—were run using the same base software with various flags to constrain the numbers of factors and states. A full suite of repeated runs took between 1-4 hours depending on the size of the dataset.

Comparisons on Synthetic Datasets We applied the iDBN prior to several datasets from Wingate et al. (2009). The three network datasets consist of binary observations indicating whether various computers are up or down. Computers crash randomly, and crashes can propagate through the (unknown) network. Each dataset contains an unobserved node which affects the topology of the network. The jungle data set is a synthetic dataset containing a timeseries of noises in a jungle soundscape (where certain animal sounds cause other animals to also make sounds). Finally, the spike train dataset was derived from recordings of hippocampal place cells in a rat while running through a linear track. The data consisted of spike counts; we applied the IBP to reduce its dimensionality. The statistics of the datasets are summarized in table 2; however note that there are always ways of explaining the data with different numbers of factors or states.

Table 2. Description of Datasets

Domain	Factors	States	Length
Jungle	6	2	52
Spike Train	1	45	179
NW-Ring	4	2	1000
NW-Star	5	2	1000
NW-Tree	7	2	1000

We compared the iDBN prior ($\alpha_{DBN} = 1$, $\alpha_{HDP} = 1$) to a finite DBN initialized with the actual number of hidden factors and states from table 2 as well as an infinite factorial HMM (iFHMM) that assumed binary hidden factors. We chose these models as comparisons because, like the iDBN, they modeled stationary (non-changing with time) distributions over the hidden states and had somewhat complementary constraints: the DBN fixed the number of nodes but allowed non-binary-valued states, while the iFHMM fixed the number of states per node. The connections for the DBN and the iFHMM were initialized each hidden node with only itself as its parent and connecting to all observed nodes. In the case of the iFHMM, the number of hidden nodes was initially set to the number of observations. To speed up burnin, the iDBN was initialized with the final iFHMM model; completely random initializations tended to get caught in local optima.

As in section 5, we randomly held out different subsets of 10% of the observed data for 5 runs of the sampler. Each run consisted of 100 iterations, with more complex models initialized from less complex ones. The predictive test-likelihood of each approach was computed over the last 10 iterations of each runs. Table 1 summarizes the results: we see that the nonparametric models always outperform the finite model; in all cases the models proposed by the iDBN score either better or comparably to the iFHMM. The DBN—even though it has the "correct" number of states—does less well with limited data due to overfitting. We also emphasize that the structures found by the iDBN are designed to predict the provided data well, not find the "correct"—or even an interpretable—structure: indeed, especially with limited data, there will be many structures that describe the data well. The results show that even though the iDBN is a more flexible prior, it generalizes to unobserved data by finding structure in the model. By allowing for connections between hidden nodes, it can also model structures such as the network topologies better than the more constrained iFHMM.

6.1. Application: Weather Modeling

For this test, we downloaded historical weather data from the US Historical Climate Network[2]. In the first test, we used daily precipitation values for 5 different weather stations (one each in Rhode Island, Connecticut, New Jersey, Delaware and California) for 10 years between 1980-1989, resulting in 3,287 timepoints. Observations were evenly discretized into 7 values.

Fig. 6 shows the results on this small time-series: on the left is the learned DBN, which identified two independent weather systems for New England and California. This interpretation was stable across many samples from the posterior, as shown in the right hand side. An entry (i, j) in the matrix represents the percentage of samples in which there was a causal connection from parent j to child i (the model occasionally inferred one extra connection [square 2,3] which did not connect to any observation). Here we see the iDBN naturally picking out the independently evolving latent factors that the iFHMM is designed to model.

Fig. 7 shows the results of the iDBN applied on a time-series of 500 weather stations across the United States. As before, the algorithm does not have access to the weather station locations; it only sees a time-series of discretized precipitation data. The data can therefore be represented as a matrix with 500 rows (representing stations) and 3,287 columns (representing days). Fig. 7 shows the results. Not only does the iDBN

[2]From ftp://ftp.ncdc.noaa.gov/pub/data/ushcn/daily/

Table 1. Comparison of iDBN approach to other algorithms. Intervals represent the standard error of the mean.

	Negative Test Likelihood			Factors Discovered		
	DBN	iFHMM	iDBN	DBN	iFHMM	iDBN
NW Star	174.0 ± 8.2	165.2 ± 3.0	156.2 ± 3.0	5	12.8 ± 0.2	2.4 ± 0.2
NW Tree	255.6 ± 7.1	286.5 ± 2.9	216.2 ± 10.0	7	12.0 ± 0.0	4.0 ± 0.4
NW Ring	181.7 ± 16.0	154.3 ± 1.6	151.4 ± 2.8	4	9.0 ± 1.2	4.2 ± 1.0
Spike Train	142.4 ± 2.7	133.1 ± 2.1	136.0 ± 2.8	1	15.9 ± 0.1	18.1 ± 6.2
Jungle	14.8 ± 1.4	13.9 ± 1.5	14.2 ± 1.6	6	3.1 ± 0.1	29.5 ± 3.6

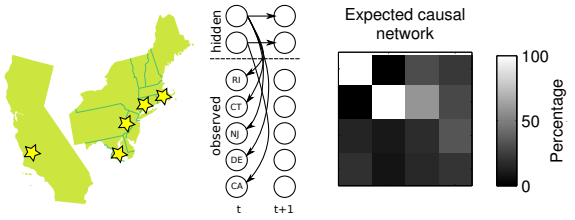

Figure 6. Results on the weather dataset. On the left: the weather stations. Middle: the inferred DBN. On the right: the expected causal connections between latent factors.

find geographically-localized clusterings of the observations, the west-to-east causal links are consistent with U.S. weather patterns (due to the jet stream). Fig. 8 shows that the iDBN finds models with lower training and test likelihoods than the iHMM, iFHMM or a flat HMM with up to 100 states.

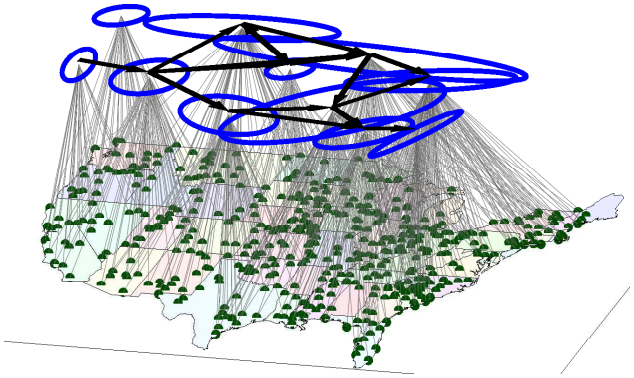

Figure 7. Sample network inferred by the iDBN based on 500 weather stations across the United States.

6.2. Application: Discovery of Neural Information Flow Networks

For our final application, we applied the iDBN to analyze neural activity recordings from the auditory pathway of zebra finches. First analyzed in Smith et al. (2006), the dataset corresponds to (possibly misplaced) electrodes put in the cerebral auditory regions of zebra finches. Raw data was discretized into three observations per electrode. The goal of the analysis

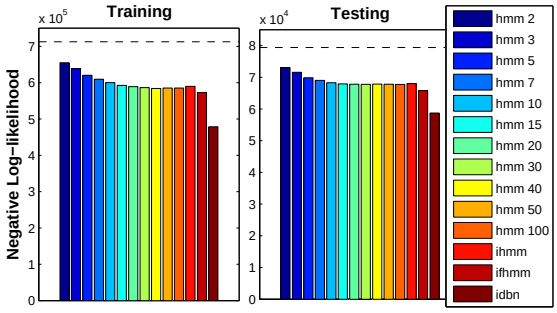

Figure 8. Training and test likelihoods for the iDBN, iHMM, iFHMM, and HMM models on the full weather data. Dashed line represents random guessing with the marginal empirical distribution.

was to infer functional connectivity between different brain regions given only a timeseries of electrode measurements. We expected factors to correspond to functional regions of the brain and causal connections to represent information or processing pathways.

We analyzed data for two birds (Black747 and LtGr841). We first tested the iDBN on temporally-scrambled versions of the datasets. It reliably inferred that no causal connections existed between the hidden factors, suggesting that the temporal connections found in the unscrambled dataset were not a product of chance. Fig. 9(B) shows clusterings found in the unscrambled time-series: each entry (i, j) in the square represents the frequency with which observation dimensions i and j were connected to the same parent. Over many runs of the iDBN, several observation factors were collapsed into a single state variable—implying that more often than not, the differences between some observations were not significant enough to justify their own factors.

The groupings are anatomically plausible: for example, in the LtGr841 block, we find that L2 and L3 were often grouped together into a single state variable; similarly, in Black747, we see that CMM and L2 were often grouped together. These observational clusterings correlate strongly with the inferred functional connectivity graphs from the original paper (Fig. 9(A)); the

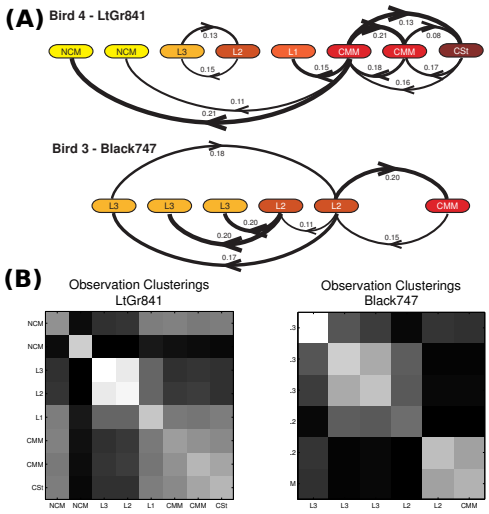

Figure 9. Results on the finch dataset. (A) Inferred functional connectivity from (Smith et al., 2006). (B) the iDBN's inferred observation clusterings (*figure (A) courtesy of V. Anne Smith*).

fully-observed-DBN approach of Smith et al. (2006) cannot infer the same collapsing of variables.

7. Discussion and Future Work

We presented the infinite DBN, a nonparametric prior over dynamic Bayesian networks that posits that the world contains an infinite number of hidden nodes as well as observed nodes; however, only a finite number of hidden nodes are needed to explain a finite number of observed nodes. By using the iDBN as a prior over hidden nodes, we automatically infer the number of hidden factors—and the number of state values they take on—to explain the observations. Adjusting concentration parameters lets us tune the models to the type of structures we prefer to find. On a variety of datasets, the iDBN finds reasonable structure, ranging from independent chains to highly connected subsets of latent factors. Importantly, this flexibility does not compromise the likelihood of the data, which is on par or better than more structurally constrained models.

The iDBN provides a very flexible way to model latent structure in observed time-series, and it also raises several interesting questions in non-parametric time-series modeling. For example, the nonparametric process that ensures that each child only has a finite number of parents also results in several popular parent factors that influence many parts of the network. While reasonable for many scenarios, one could also imagine a complementary model in which a few popular child nodes were affected by many other nodes in the network. One could also imagine models in which the expected number of hidden nodes needed to model a time-series grows with the length of the time-series, rather than the number of observed nodes. We chose the form of the iDBN prior as a balance between flexibility and tractable inference; developing other nonparametric time-series models—including those that can model non-stationary and relational data—and accompanying inference techniques for specific applications remains an interesting area for future work.

Acknowledgments

F.D. is funded by the Hugh Hampton Young Fellowship. D.W. is funded by AFOSR FA9550-07-1-0075.

References

Adams, R. P., Wallach, H. M., and Ghahramani, Z. Learning the structure of deep sparse graphical models. In *AISTATS*, 2010.

Doucet, A. and Johansen, A. M. A tutorial on particle filtering and smoothing: fifteen years later. In *In Handbook of Nonlinear Filtering (eds.* University Press, 2009.

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. Bayesian nonparametric inference of switching linear dynamical systems. Technical Report 2830, MIT Laboratory for Information and Decision Systems, March 2010.

Ghahramani, Z. Learning dynamic Bayesian networks. In *Adaptive Processing of Sequences and Data Structures, International Summer School on Neural Networks*, 1998.

Griffiths, T. and Ghahramani, Z. Infinite latent feature models and the Indian buffet process. In *TR 2005-001, Gatsby Computational Neuroscience Unit*, 2005.

Heckerman, D. A tutorial on learning with Bayesian networks. Technical report, Microsoft Reasearch MSR-TR-95-06, 1995.

Heller, K. A., Teh, Y. W., and Görür, D. Infinite hierarchical hidden Markov models. In *AISTATS*, 2009.

Murphy, K. *Dynamic Bayesian Networks: Representation, Inference and Learning.* PhD thesis, UC Berkeley, 2002.

Murphy, K. P. and Weiss, Y. The factored frontier algorithm for approximate inference in dbns. In *UAI*, 2001.

Ng, B. N. Adaptive dynamic Bayesian networks. In *Joint Statistical Meetings*, 2007.

Peña, J. M., Björkegren, J., and Tegnér, J. Learning dynamic Bayesian network models via cross-validation. *Pattern Recogn. Lett.*, 26, October 2005.

Smith, V. A., Yu, J., Smulders, T. V., Hartemink, A. J., and Jarvis, E. D. Computational inference of neural information flow networks. *PLoS Computational Biology*, 2(11), 2006.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101:1566–1581, 2006.

Van Gael, J., Teh, Y. W., and Ghahramani, Z. The infinite factorial hidden Markov model. In *NIPS*, 2009.

Wingate, D., Goodman, N. D., Roy, D. M., and Tenenbaum, J. B. The infinite latent events model. UAI, 2009.

Xing-Chen, H., Zheng, Q., Lei, T., and Li-Ping, S. Research on structure learning of dynamic Bayesian networks by particle swarm optimization. In *Artificial Life*, 2007.