# Time-Series Classification Using Mixed-State Dynamic Bayesian Networks*


Vladimir Pavlović,† Brendan J. Frey,‡ and Thomas S. Huang
ECE Department and Beckman Institute
University of Illinois, Urbana, IL 61801
vpavlovic@ieee.org



## Abstract

*We present a novel mixed-state dynamic Bayesian network (DBN) framework for modeling and classifying time-series data such as object trajectories. A hidden Markov model (HMM) of discrete actions is coupled with a linear dynamical system (LDS) model of continuous trajectory motion. This combination allows us to model both the discrete and continuous causes of trajectories such as human gestures. The model is derived using a rich theoretical corpus from the Bayesian network literature. This allows us to use an approximate structured variational inference technique to solve the otherwise intractable inference of action and system states. Using the same DBN framework we show how to learn the mixed-state model parameters from data. Experiments show that with high statistical confidence the mixed-state DBNs perform favorably when compared to decoupled HMM/LDS models on the task of recognizing human gestures made with a computer mouse.*


## 1. Introduction

Analysis and classification of temporal sequences has been a focus of research for many decades. Many techniques have been developed in fields such as signal processing, computer vision and finance, that deal with analysis and classification of different time-series. For instance, Kalman state estimation [11] is the basis for estimation in continuous state linear dynamic systems (LDS) while hidden Markov models (HMMs) [14] excel at classification of discrete-state sequences. Recently, a new statistical approach from the perspective of *Bayesian networks* was proposed for temporal series modeling[16, 5]. It was shown that estimation in LDSs and inference in HMMs can be viewed as special case of Bayesian inference in a general time-series model known as a *dynamic Bayesian network* (DBN). Besides allowing one to view seemingly different models (e.g. LDSs and HMMs) as special cases of DBNs, the framework also enables one to apply a corpus of exact and approximate statistical inference and learning techniques from the BN literature to time-series modeling. This has resulted in new approaches to inference and in novel complex temporal models such as factorial HMMs [8], coupled HMMs [2, 13], switching-state space models [7], mixtures of DBNs [13], etc.

We consider an instance of a complex DBN that arises as a combination of discrete-state HMMs and continuous-state LDSs. We call such DBNs *mixed-state DBNs*. Namely, a mixed-state DBN is a HMM coupled with a LDS (see Figure 1). The output of a HMM is the *driving* input to a linear system. What motivates one to consider such a combination of systems? Several examples may easily come to mind. Suppose we observe an autonomous moving target. The target motion is governed by Newtonian physics as well as the input force (thrust) imposed upon it by its (human) operator. Assume we have some knowledge of what sequences of actions the operator may take in time. In other words, we know that there are dependencies between levels of control thrust at successive time instances. So, it is plausible to model the thrust controlled by an operator as

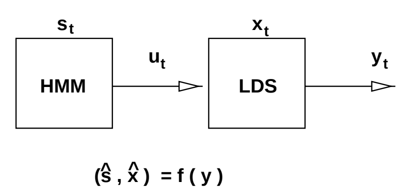


**Figure 1.** Block diagram of a physical system (LDS) driven by an input generated from a hidden Markov model (HMM). The system has an equivalent representation as a mixed state dynamic Bayesian network.

---

†Present affiliation: Compaq Computer Corporation, Cambridge Research Lab, Cambridge, MA 02139

‡Present affiliation: Department of Computer Science, University of Waterloo, Waterloo, Ontario N2L 3G1


*This work was supported in part by National Science Foundation grant IRI-96-34618 and in part by the Army Research Laboratory Cooperative Agreement No. DAAL01-96-2-0003.

a HMM where the hidden states correspond to a number possible actions the operator may take and the observables model the thrust value at those action instances. Knowledge of object motion under Newtonian laws of physics is embedded in the LDS model. Another example of a physical system that can be modeled as a mixed DBN is the human hand/arm motion during gestural communication. The physical arm motion can be described using different kinematic and dynamic models of simple or articulated structures while the language concepts that influence the arm motion can be modeled using HMMs.

In both of the above examples, the aim of modeling the systems in the mixed DBN framework is to be able to *infer* what the underlying action that drives the physical system is. This can help us distinguish among different motion patterns of hand gestures observed by a computerized interactive kiosk or a target observed by a radar. Moreover, the actions need not only be inferred – they can also be predicted. Consequently, the motion of the human hand or the target can be predicted based on the *predicted action* and used for tracking of the physical system. Finally, the DBN framework provides us with a well-defined basis for *learning* the model parameters from data.

### 1.1. Previous Work

Models similar to mixed-state DBN have been considered in the past, although from a different perspective. The mixed-state DBN can be most directly related to different models of *maneuvering* targets [1]. However, the majority of maneuvering target models have their origins in the classical LDS theory and are focused on the estimation of the physical system states (and not the actions). They also sometimes employ approximations to exact inference that are not well justified or without strict error bounds.

## 2. Mixed-State Dynamic Bayesian Network

Consider a coupled system described by the block diagram in Figure 1. The system can be described using the following set of state-space equations:

$$x_{t+1} = Ax_t + Bu_{t+1} + v_{t+1}, \quad (1)$$
$$y_t = Cx_t + w_t, \quad (2)$$
$$x_0 = Bu_0 + v_0, \quad (3)$$

for the physical system, and

$$Pr(s_{t+1}|s_t) = s_{t+1}' \, \Pi \, s_t, \quad (4)$$
$$u_t = Ds_t + r_t, \quad (5)$$
$$Pr(s_0) = \pi_0, \quad (6)$$

for the driving actions. The meaning of the variables is as follows: $x_t \in \Re^N$ denotes the hidden state of the LDS, $u_t$ is an input to this system, $v_t$ is the state noise process. Similarly, $y_t \in \Re^M$ is the observed measurement and $w_t$ is the measurement noise. Parameters $A$, $B$ and $C$ are the typical LDS parameters: the state transition matrix, the input matrix and the observation matrix, respectively. The action generator is modeled by a HMM. State variables of this model are written as $s_t$. They belong to the set of $S$ discrete symbols $\{e_0, \dots, e_{S-1}\}$, where $e_i$ is the unit vector of dimension $S$ with a non-zero element in the $i$-th position. The HMM is defined with the state transition matrix $\Pi$ whose elements are $\Pi(i,j) = Pr(s_{t+1} = e_i | s_t = e_j)$, observation matrix $D$, and an initial state distribution $\pi_0$. The HMM observation noise process is denoted by $r_t$. Note that *the input to the LDS ,$u$, is the output of the action HMM.*

The mixed state space representation is equivalently depicted by the dependency graph in Figure 2 and can be written as the *joint distribution* $P$:

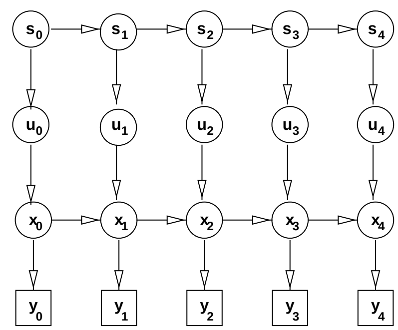


**Figure 2.** Bayesian network representation (dependency graph) of the mixed-state DBN. $s$ denote instances of the discrete valued action states driving the continuous valued physical system states $x$ and observations $y$.

$$P(\mathcal{Y}, \mathcal{X}, \mathcal{U}, \mathcal{S}) = Pr(s_0) \prod_{t=1}^{T-1} Pr(s_t|s_{t-1}) \prod_{t=0}^{T-1} Pr(u_t|s_t)$$
$$Pr(x_0|u_0) \prod_{t=1}^{T-1} Pr(x_t|x_{t-1}, u_t) \prod_{t=0}^{T-1} Pr(y_t|x_t), \quad (7)$$

where $\mathcal{Y}, \mathcal{X}, \mathcal{U},$ and $\mathcal{S}$ denote the sequences (with length $T$) of observations and hidden state variables. For instance, $\mathcal{Y} = \{y_0, \dots, y_{T-1}\}$. Terms $v_t$ and $w_t$ in the physical system formulation are used to denote random noise. We can write an equivalent representation of the physical system in the probability space assuming that the following conditional pdfs are defined:

$$Pr(x_{t+1}|x_t, u_{t+1}) = P_x(x_{t+1} - Ax_t - Bu_{t+1}) \quad (8)$$
$$Pr(y_t|x_t) = P_y(y_t - Cx_t), \quad (9)$$

where $P_x$ and $P_y$ are known, parametric or non-parametric, pdfs. Similarly, the observation pdf of the HMM can be written:

$$Pr(u_t|s_t) = P_u(u_t - Ds_t). \tag{10}$$

Throughout the rest of this paper we assume without loss of generality that the state noise $v$ of the physical system is zero w.p.1 because the HMM observation noise $r_t$ can account for it. The observation noise processes of both the physical system and the HMM are modeled as i.i.d. zero-mean Gaussian:

$$\begin{aligned} w_t &\sim \mathcal{N}(0, R), \\ r_t &\sim \mathcal{N}(0, Q). \end{aligned}$$

Also, assume $B$ to be identity, $B = I$. Input variable $u_t$ can be eliminated from Equations 1 and 5 as an auxiliary variable. Given the above assumptions, the joint pdf of the mixed-state DBN of duration T (or, equivalently, its Hamiltonian[1]) can be written as in Equation 11.

### 2.1. Hidden State Inference

The goal of inference in mixed-state DBNs is to estimate the posterior probability of the hidden states of the system ($s_t$ and $x_t$) given some known sequence of observations $\mathcal{Y}$ and the known model parameters. Namely, we need to find the posterior

$$P(\mathcal{X}, \mathcal{S}|\mathcal{Y}) = Pr(\mathcal{X}, \mathcal{S}|\mathcal{Y}).$$

In fact, it suffices to find the *sufficient statistics* [3] of the posterior. Given the form of $P$ it is easy to show that these statistics are $\langle [x_t s_t] \rangle$, $\langle [x_t s_t][x_t s_t]' \rangle$, and $\langle [x_t s_t][x_{t-1} s_{t-1}]' \rangle$. The operator $\langle \cdot \rangle$ denotes conditional expectation with respect to the posterior distribution, e.g. $\langle x_t \rangle = \sum_{\mathcal{S}} \int_{\mathcal{X}} x_t P(\mathcal{X}, \mathcal{S}|\mathcal{Y})$.

If there were no action dynamics, the inference would be straightforward – we could infer $\mathcal{X}$ from $\mathcal{Y}$ using LDS inference (RTS smoothing [15]). However, the presence of action dynamics embedded in matrix $\Pi$ makes exact inference more complicated. To see that, assume that the initial distribution of $x_0$ at $t = 0$ is Gaussian, at $t = 1$ the pdf of the physical system state $x_1$ becomes a mixture of $S$ Gaussian pdfs since we need to marginalize over $S$ possible but unknown input levels. At time $t$ we will have a mixture of $S^t$ Gaussians, which is clearly intractable for even moderate sequence lengths. So, it is more plausible to look for an approximate, yet tractable, solution to the inference problem.

[1]Hamiltonian $H(x)$ of a distribution $P(x)$ is defined as any positive function such that $P(x) = \frac{\exp(-H(x))}{\sum_{\psi} \exp(-H(\psi))}$.

### 2.2. Approximate Inference Using Structured Variational Inference

Structured variational inference techniques [10] consider a parameterized distribution which is in some sense close to the desired conditional distribution, but is easier to compute. Namely, for a given set of observations $\mathcal{Y}$, a distribution $Q(\mathcal{X}, \mathcal{S}|\eta, \mathcal{Y})$ with an additional set of *variational parameters* $\eta$ is defined such that Kullback–Leibler divergence between $Q(\mathcal{X}, \mathcal{S}|\eta, \mathcal{Y})$ and $P(\mathcal{X}, \mathcal{S}|\mathcal{Y})$ is minimized with respect to $\eta$:

$$\eta^* = \arg\min_{\eta} \sum_{\mathcal{S}} \int_{\mathcal{X}} Q(\mathcal{X}, \mathcal{S}|\eta, \mathcal{Y}) \log \frac{P(\mathcal{X}, \mathcal{S}|\mathcal{Y})}{Q(\mathcal{X}, \mathcal{S}|\eta, \mathcal{Y})}. \tag{12}$$

The dependency structure of $Q$ is chosen such that it closely resembles the dependency structure of the original distribution $P$. However, unlike $P$ the dependency structure of $Q$ *must* allow a computationally efficient inference. In our case we decouple the HMM and LDS as indicated in Figure 3. The two subgraphs of the original network are a HMM $Q_S$ with variational parameters $\{q_0, \ldots, q_{T-1}\} \in \Re^S$ and a LDS $Q_X$ with variational parameters $\{u_0, \ldots, u_{T-1}\} \in \Re^N$. More precisely, the Hamiltonian of the approximating network is defined in Equation 13. The subgraphs are *decoupled*, thus allowing for independent inference, $Q(\mathcal{X}, \mathcal{S}|\eta, \mathcal{Y}) = Q_X(\mathcal{X}|\eta, \mathcal{Y}) Q_S(\mathcal{S}|\eta)$. This is also reflected in the sufficient statistics of the posterior defined by the approximating network. They are $\langle s_t \rangle$, $\langle s_t s_{t-1}' \rangle$ for the HMM subgraph, and $\langle x_t \rangle$, $\langle x_t x_t' \rangle$, and $\langle x_t x_{t-1}' \rangle$ for the LDS.

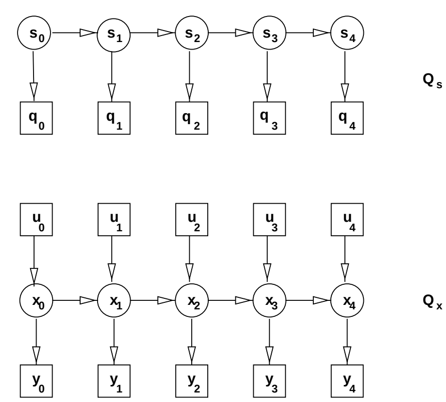


**Figure 3.** Factorization of the original mixed-state DBN. Factorization reduces the coupled network into a decoupled pair of a HMM ($Q_s$) and a LDS ($Q_x$).

The optimal values of the variational parameters $\eta = \{q_0, \ldots, q_{T-1}, u_0, \ldots, u_{T-1}\}$ can be obtained by setting the derivative of the KL-divergence w.r.t. $\eta$ to zero. Alternatively, one can employ the theorem of Ghahramani [6] to

$$H(\mathcal{X},\mathcal{S},\mathcal{Y}) = \frac{1}{2}\sum_{t=1}^{T-1}\left(x_t - Ax_{t-1} - Ds_t\right)' Q^{-1}\left(x_t - Ax_{t-1} - Ds_t\right) + \frac{1}{2}\left(x_0 - Ds_0\right)' Q^{-1}\left(x_0 - Ds_0\right)$$
$$+\frac{T}{2}\log|Q| + \frac{NT}{2}\log 2\pi + \frac{1}{2}\sum_{t=0}^{T-1}\left(y_t - Cx_t\right)' R^{-1}\left(y_t - Cx_t\right) + \frac{T}{2}\log|R| + \frac{MT}{2}\log 2\pi$$
$$+\sum_{t=1}^{T-1} s_t'(-\log\Pi)s_{t-1} + s_0'(-\log\pi_0) \qquad (11)$$

$$H_Q(\mathcal{X},\mathcal{S},\mathcal{Y}) = \frac{1}{2}\sum_{t=1}^{T-1}\left(x_t - Ax_{t-1} - u_t\right)' Q^{-1}\left(x_t - Ax_{t-1} - u_t\right) + \frac{1}{2}\left(x_0 - u_0\right)' Q^{-1}\left(x_0 - u_0\right)$$
$$+\frac{T}{2}\log|Q| + \frac{NT}{2}\log 2\pi + \frac{1}{2}\sum_{t=0}^{T-1}\left(y_t - Cx_t\right)' R^{-1}\left(y_t - Cx_t\right) + \frac{T}{2}\log|R| + \frac{MT}{2}\log 2\pi$$
$$+\sum_{t=1}^{T-1} s_t'(-\log\Pi)s_{t-1} + s_0'(-\log\pi_0) + \sum_{t=0}^{T-1} s_t'(-\log q_t) \qquad (13)$$

arrive at the following optimal variational parameters:

$$u_t^* = D\langle s_t\rangle \qquad (14)$$
$$q_t^*(i) = e^{d_i' Q^{-1}\left(\langle x_t\rangle - A\langle x_{t-1}\rangle - \frac{1}{2}d_i\right)} \qquad (15)$$

where $d_i$ denotes the $i$-th column of $D$ and $\langle x_{-1}\rangle \triangleq 0$. To obtain the expectation terms $\langle s_t\rangle = Pr(s_t|q_0,\ldots,q_{T-1})$ we use the inference in the HMM [14] with output "probabilities" $q_t$. Similarly, to obtain $\langle x_t\rangle = E[x_t|u_0,\ldots,u_{T-1},y_0,\ldots,y_{T-1}]$ we perform LDS inference (Rauch-Tung-Streiber smoothing [15]) with inputs $u_t$. Since $u_t$ in subgraph $Q_X$ depends on $\langle s_t\rangle$ from subgraph $Q_S$ and $q_t$ depends on $\langle x_t\rangle$, Equations 14 and 15 together with the inference solutions form a set of *fixed-point equations*. Solution of this fixed-point set yields a tractable approximation to the intractable original posterior. Error bounds of the approximation are easy to derive and can be found in [13].

The variational inference algorithm for mixed-state DBNs can now be summarized as:

```
error = ∞;
Initialize ⟨x⟩;
while (error > maxError) {
    Find q_t from ⟨x_t⟩ using Equation 15;
    Estimate ⟨s_t⟩ from q_t using HMM inference;
    Find u_t from ⟨s_t⟩ using Equation 14;
    Estimate ⟨x_t⟩ from y_t and u_t using LDS
        inference;
    Update approximation error;
}
```

From Equation 14 and the factorization of the network defined in Equation 13 it is evident that $u_t$ can be viewed as the estimated input of the LDS, based on the estimates of the hidden states of the HMM subgraph. The input at time $t$ is estimated to be a linear combination of all possible inputs $d_i$ weighted by their corresponding probabilities $\langle s_t(i)\rangle$, $D\langle s_t\rangle = \sum_{i=0}^{N-1} d_i\langle s_t(i)\rangle$.

The meaning of $q_t$ is not immediately obvious. Based on Equation 13, $q_t$ can be viewed as the probabilities of some fictional discrete-valued inputs presented to the HMM subgraph. These probabilities are related to the estimates of the states $x_t$ of the LDS through Equation 15. To better understand the meaning of this dependency consider the plot in Figure 4 of $q_t$ versus $d = d_i$ for a fixed value of the difference $\langle x_t\rangle - A\langle x_{t-1}\rangle$ and unit variance $Q$. Clearly, the function assumes a maximum value for $d = \langle x_t\rangle - A\langle x_{t-1}\rangle$. If we have a set of discrete values of $d$ corresponding to $N$ possible LDS input levels $d_i$, $q_t(i)$ would be maximized for $d_i$ closest to the estimated difference $\langle x_t\rangle - A\langle x_{t-1}\rangle = \langle u_t\rangle$. Thus, those states of the HMM are favored which are

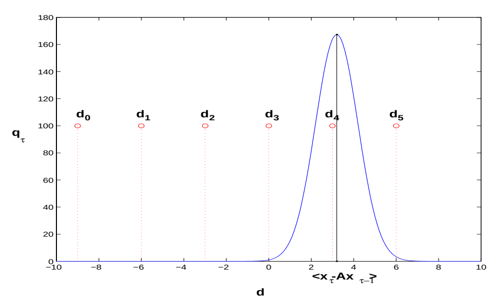


**Figure 4.** Variational parameter $q_t$ as a function of input level $d$. Shown is a set of six different input levels $d_0$ through $d_5$. The function attains maximum for input level $d_4$ which is closest to the global maximum at $d^* = \langle x_t\rangle - A\langle x_{t-1}\rangle$.

produce inputs “closer” to the ones estimated from the LDS dynamics.

## 3. Maximum Likelihood Learning of Mixed-State DBNs

Learning in mixed-state DBNs can be formulated as the problem of ML learning in general Bayesian networks. It was shown in [10] that structured variational inference can be viewed as the *expectation* step of a generalized EM algorithm [9, 12]. The maximization step then yields

$$\theta^* = \arg\max_{\theta} \sum_{\mathcal{S}} \int_{\mathcal{X}} Q(\mathcal{X}, \mathcal{S}|\mathcal{Y}, \eta^*) \log P(\mathcal{X}, \mathcal{S}, \mathcal{Y}),$$

where $\theta$ is the set of parameters of pdf $P$. In our case, the parameters are $\{A, C, D, Q, R, \Pi, \pi_0\}$.

Given the sufficient statistics obtained in the inference phase, it is easy to show that the following *parameter update equations* result from the Maximization step:

$$\begin{aligned}
A^* &= \left( \sum_{t=1}^{T-1} \left\langle x_t x'_{t-1} \right\rangle - D^* \left\langle s_t \right\rangle \left\langle x'_{t-1} \right\rangle \right) \\
&\quad \left( \sum_{t=1}^{T-1} \left\langle x_{t-1} x'_{t-1} \right\rangle \right)^{-1} \\
Q^* &= \frac{1}{T} \sum_{t=0}^{T-1} \left\langle x_t x'_t \right\rangle - A^* \left\langle x_{t-1} x'_t \right\rangle - \\
&\quad D^* \left\langle s_t \right\rangle \left\langle x'_t \right\rangle \\
D^* &= \left( \sum_{t=0}^{T-1} \left\langle x_t \right\rangle \left\langle s'_t \right\rangle - A^* \left\langle x_{t-1} \right\rangle \left\langle s'_t \right\rangle \right) \\
&\quad \left( \sum_{t=0}^{T-1} \left\langle s_t s'_t \right\rangle \right)^{-1} \\
C^* &= \left( \sum_{t=0}^{T-1} y_t \left\langle x'_t \right\rangle \right) \left( \sum_{t=0}^{T-1} \left\langle x_t x'_t \right\rangle \right)^{-1} \\
R^* &= \frac{1}{T} \sum_{t=0}^{T-1} \left( y_t y^t_t - C^* \left\langle x_t \right\rangle y'_t \right) \\
\Pi^* &= \left( \sum_{t=1}^{T-1} \left\langle s_t s'_{t-1} \right\rangle \right) \left( \sum_{t=1}^{T-1} \left\langle s_t s'_t \right\rangle \right)^{-1} \\
\pi_0^* &= \left\langle s_0 \right\rangle .
\end{aligned}$$

All the variable statistics are evaluated before updating any parameters. Notice that the above equations represent a generalization of the parameter update equations of zero-input LDS models [5].

## 4. Analysis and Recognition of Hand Gestures Acquired by a Computer Mouse

To demonstrate feasibility of the mixed-state DBN framework we consider the task of classifying a set of symbols drawn using a computer mouse. We defined four classes of symbols: arrow, erase, circle, and wiggle (see Figure 5.) The task in question was to model each of the four symbols with a combination of LDS and HMM. The LDS part modeled the Newtonian dynamics of the mouse motion. Namely, we assumed that the mouse motion can be modeled as a planar motion of a point-mass particle with piece-wise constant acceleration:

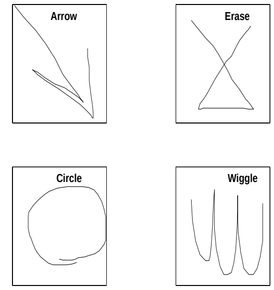


**Figure 5.** Examples of four symbols produced by computer mouse motion.

$$\begin{aligned}
\frac{d\, position(t)}{dt} &= velocity(t) \\
\frac{d\, velocity(t)}{dt} &= u(t) + noise(t).
\end{aligned}$$

This leads to a discrete-time LDS with known $A$, $B$ and $C$ and unknown $Q$ and $R$ (cf [1]). On the other hand the HMM models the driving force (action) that causes the motion. The mixed-state DBN model is contrasted with two decoupled model

- Decoupled adapted LDS and HMM. Namely, the LDS is adapted to “best” model the dynamics of the mouse motion of each symbol when the driving force $u_t$ is assumed to be quasi-constant with additive white noise, $u_t = u_{t-1} + n_{u,t}$. The HMM is consequently employed to model the quasi-constant driving force $\langle u_t \rangle$ inferred by the LDS.

- Decoupled fixed LDS and HMM. In this case, the LDS is assumed to be fixed for all four symbols. In particular, we estimated the driving force using numerical gradient approximation: $u_t = grad(grad(x_t))$, where $grad(x_t) = \frac{x_{t+1} - x_{t-1}}{2 \cdot \Delta T}$. Again, an HMM is used to model the estimated driving force.

All three model classes are depicted in Figure 6.

For each of the three models the same action state spaces are assumed. The number of action states is proportional to the number of strokes necessary to produce each symbol. Thus, the action model of the arrow symbol had eight states (two times four strokes), erase has six states, circle four states, and wiggle six. Furthermore, each symbol’s state transitions are limited to left-to-right: from current state the

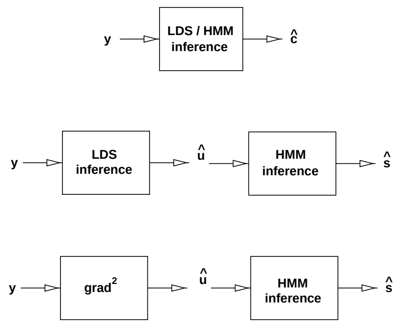


**Figure 6.** Three ways of modeling mouse acquired symbols. From top to bottom: completely coupled LDS and HMM (mixed-state HMM), decoupled adapted LDS and HMM, and decoupled fixed LDS and HMM.

action can only transition back to itself or to only one other not-yet-visited state. In the two decoupled symbol models, we model the observations $u_t$ of the action models as variable mean Gaussian processes with identical variances at every action state[2]. Model parameters are learned from data using the ML learning framework.

The data set consists of 136 examples of each symbol (a total of $4 \times 136$ examples). Symbols were acquired from normalized[3] mouse movements sampled at $\Delta T = 100ms$ intervals. To test the models' performance we used rotation error counting (cross-validation) method with four rotational sets [4]. For each test sample and each symbol model the likelihood of the sample was appropriately obtained. For instance, in the case of symbols modeled by mixed-state DBNs, variational inference with a relative error threshold of $10^{-3}$ was used to estimate the lower bound on likelihood. One example of mixed-state DBN-based decoding of the "arrow" symbol is shown in Figure 7. For the fixed LDS and gradient-based LDS/HMM models, likelihood was obtained using the standard HMM and LDS inference.

Classification test were performed on two sets of data: noise-free and noisy. Classification of noisy symbols is of particular interest since it introduces variability that may pose a challenge to decoupled classification models. The noisy data set was constructed by adding i.i.d. zero mean Gaussian noise with standard deviation of 0.01 to noise-free examples (see Figure 8). Models of the four symbols trained on noise-free samples were now tested on the noisy data. Classification results are summarized in Table 1 and Figure 9. Table 1 and Figure 9 in this case indicate that with 95% confidence completely coupled mixed-state HMM models had significantly better performance that both fixed and adapted decoupled LDS/HMM classifiers (with the exception of mixed-state and fixed LDS "circle" models). Of course, the tradeoff is as always in increased computational complexity of the mixed-state models. We note, however, that on the average the iterative scheme of the mixed-state models required only about 5 to 10 iterations to converge.

[2]Even though it is a usual practice to allow the variance to vary from action state to action state, for sake of compatibility with the fixed variance mixed-state HMM we decided to keep the other HMMs' observation variances fixed.

[3]Symbol were scaled to $[0, 1] \times [0, 1]$ unit area and directionally aligned.

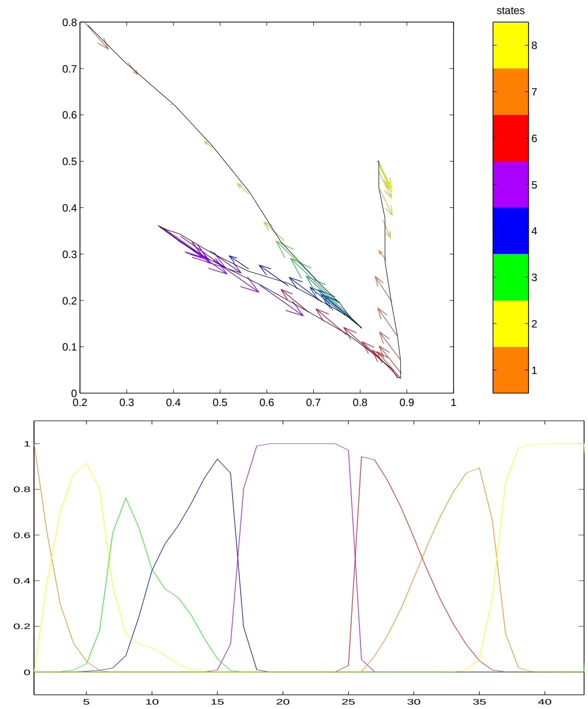


**Figure 7.** Estimates of action states for "arrow" symbol. Top graph depicts the symbol and the estimated driving force. Bottom graph shows estimates of action states obtained using variational inference. Colors in both graphs indicate action states.

## 5. Summary and Conclusions

We formulated a novel *mixed-state dynamic Bayesian network* (DBN) framework for modeling of time-series that *fuses* the typical models of driving actions (HMMs) with continuous state models of physical systems (LDSs). The model was developed under the auspices of the DBN theory allowing us to employ a well-founded set of statistical estimation and learning techniques. In particular, we employed an approximate iterative solution to otherwise intractable inference of action and system states. Using the

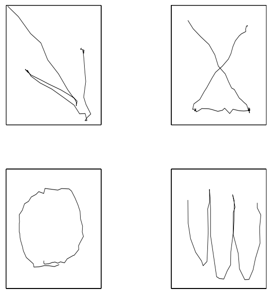

**Figure 8.** Samples of four mouse acquired symbols corrupted by additive zero mean white noise with standard deviation of 0.01.

| Model | Arrow | Erase | Circle | Wiggle |
|---|---|---|---|---|
| mixed-state HMM | 4.36 (0.89) | 4.36 (0.89) | 0.18 (0.26) | 0.18 (0.26) |
| gradient fixed LDS/HMM | 9.45 (1.25) | 14.55 (1.51) | 14.73 (1.51) | 8.18 (1.18) |
| decoupled adapted LDS/HMM | 24.91 (1.84) | 25.09 (1.85) | 0.55 (0.36) | 35.64 (2.04) |

**Table 1.** Error estimates [%] and error estimate variances ([%]) for noisy mouse symbol classification.

same DBN framework, we derived the ML learning equations for model parameter updates. The proposed model was tested on a mouse drawn symbol classification task. With high statistical confidence the model performed favorably when compared to several classical decoupled action/physical system models.

## References


[1] Y. Bar-Shalom and T. E. Fortmann. *Tracking and data association*. Academic Press, Inc., Orlando, FL, 1988.

[2] M. Brand and N. Oliver. Coupled hidden Markov models for complex action recognition, 1997.

[3] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, New York, NY, 1991.

[4] P. A. Devijver and J. Kittler. *Pattern recognition: A statistical approach*. Prentice Hall, London, 1982.

[5] Z. Ghahramani. Learning dynamic Bayesian networks. In C. L. Giles and M. Gori, editors, *Adaptive processing of temporal information*, Lecture notes in artificial intelligence. Springer-Verlag, 1997.

[6] Z. Ghahramani. On structured variational inference. Technical Report CRG-TR-97-1, Department of Computer Science, University of Toronto, 1997.

[7] Z. Ghahramani and G. E. Hinton. Switching state-space models. submitted for publication, 1998.

[8] Z. Ghahramani and M. I. Jordan. Factorial hidden markov models. *Machine learning*, 29:245–273, 1997.


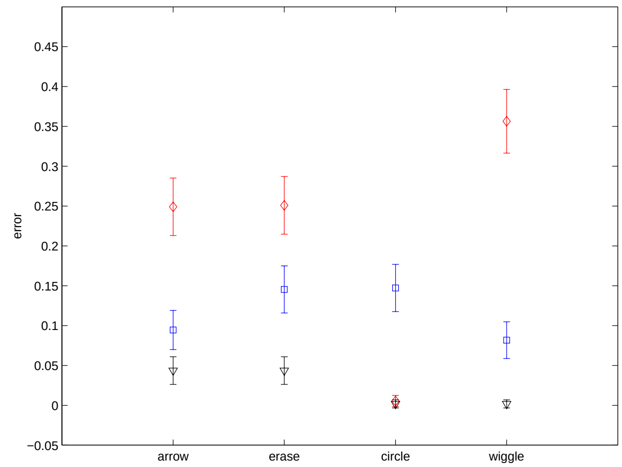


**Figure 9.** Classification error estimates of four noisy mouse symbols. Show are 95% confidence intervals for error counts. The coupled mixed-state DBN performs significantly better than the decoupled adapted and fixed LDS/HMM models. (▽ - mixed-state DBN, □ - fixed (gradient) LDS/HMM, ◇ - adapted LDS/HMM.)


[9] R. J. Hathaway. Another interpretation of the EM algorithm for mixture of distribution. *Statistics and Probability Letters*, 4:53–56, 1986.

[10] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. In M. I. Jordan, editor, *Learning in graphical models*. Kluwer Academic Publishers, 1998.

[11] R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction. *Journal of Basic Engineering (ASME)*, D(83):95–108, 1961.

[12] R. M. Neal and G. E. Hinton. A new view of the EM algorithm that justifies incremental and other variants. Technical report, Department of Computer Science, Univeristy of Toronto, 1993.

[13] V. Pavlović. *Dynamic Bayesian Networks for Information Fusion with Applications to Human–Computer Interfaces*. PhD thesis, University of Illinois at Urbana-Champaign, 1999.

[14] L. R. Rabiner and B. Juang. *Fundamentals of Speech Recognition*. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1993.

[15] H. E. Rauch. Solutions to the linear smoothing problem. *IEEE Trans. Automatic Control*, AC-8(4):371–372, October 1963.

[16] S. T. Roweis and Z. Ghahramani. A unifying review of linear gaussian models. submitted for publication, 1997.