

Interaction Analysis using Switching Structured Autoregressive Models

Michael R. Siracusa and John W. Fisher III

Computer Science and Artificial Intelligence Laboratory
 Massachusetts Institute of Technology, Cambridge MA 02139
 {siracusa,fisher}@csail.mit.edu

Abstract—This paper explores modeling the dependency structure among multiple vector time-series. We focus on a large classes of structures which yield efficient and tractable exact inference. Specifically, we use directed trees and forests to model causal interactions among time-series. These models are incorporated in a dynamic setting in which a latent variable indexes evolving structures. We demonstrate the utility of the method by analyzing the interaction of multiple moving objects.

I. INTRODUCTION

Consider a scene in which multiple objects are moving around in an environment. Given measurements of their position over time we wish to identify any interactions among them and describe how this interaction changes over time. We treat this task of labeling interaction as a problem of inference of dependency structures in a model for multiple time-series. Here the statistical dependence is of primary interest and the underlying parameters of the model are treated as nuisances. We show how to perform exact Bayesian inference on directed tree and forest structured vector autoregressive models. While the number of directed trees and forests is super-exponential in the number of time-series, the Matrix-Tree theorem allows one to calculate sums of multiplicative and additive functions on directed edges in polynomial time. This, combined with a conjugate prior allows for efficient and exact calculation of posterior event probabilities.

We examine the utility of our model within a dynamic setting. An evolving latent variable is introduced to index changing structure over time. We present empirical results on two data sets evaluating the interaction of multiple moving objects.

Spectral methods for point estimation of graphical models over stationary time-series is considered in [2]. The related work of [3], [4], [5], [6] consider alternative models in which structure is changing over time. The methodology here differs in that while we restrict ourselves to directed models, we marginalize over parameters generating a tractable posterior over structures. These models as well as the model proposed here fall into the general class of multinets [7] or Contingent Bayesian Networks [8].

II. STRUCTURED VECTOR AUTOREGRESSIVE MODEL

Here, we present a structured vector autoregressive model for multivariate time-series. This model is an r th order Markov model with additional structural constraints. We begin by introducing some notation for the purpose of explicitly denoting individual time-series, past values of individual time-series as well as sets thereof.

Consider N time-series / data streams and let \mathbf{x}_t^v be a random vector representing the value of the v th time-series at time t . The r past values of time-series v is defined as $\tilde{\mathbf{x}}_t^v$. That is, $\tilde{\mathbf{x}}_t^v$ is a stacked vector of \mathbf{x}_{t-1}^v through \mathbf{x}_{t-r}^v . As we will be explicit on the model order, r is suppressed in the notation, $\tilde{\mathbf{x}}_t^v$, for brevity. Furthermore, the vector $\tilde{\mathbf{x}}_t^{\mathbf{S}}$ indexed by set \mathbf{S} is $\tilde{\mathbf{x}}_t^{S(1)}, \dots, \tilde{\mathbf{x}}_t^{S(m)}$ stacked in a vector where $|\mathbf{S}| = m$ (e.g. $\tilde{\mathbf{x}}_t^{v,u}$ is $\tilde{\mathbf{x}}_t^v$ and $\tilde{\mathbf{x}}_t^u$ stacked). The random variables denoting the present and past of all time-series at time t are defined as \mathbf{X}_t and $\tilde{\mathbf{X}}_t$ respectively.

Multiple time points can be indexed by a vector $\mathbf{t} = [t_1, t_2, \dots, t_T]$ such that $\mathbf{X}_{\mathbf{t}} = [\mathbf{X}_{t_1}, \dots, \mathbf{X}_{t_T}]$. Note that the collection of past values, $\tilde{\mathbf{X}}_{\mathbf{t}}$, can be formed from $\mathbf{X}_{\mathbf{t}}$ and a set \mathcal{C} containing values not available in $\mathbf{X}_{\mathbf{t}}$ (initial conditions). Given a directed structure \bar{E} , a set of parameters Θ and \mathcal{C} , the model is r th order Markov:

$$p(\mathbf{X}_{\mathbf{t}} | \bar{E}, \Theta, \mathcal{C}) = \prod_{i=1}^T p(\mathbf{X}_{\mathbf{t}(i)} | \tilde{\mathbf{X}}_{\mathbf{t}(i)}, \bar{E}, \Theta). \quad (1)$$

In order to simplify notation we will drop the \mathcal{C} when it is clear from the context. \bar{E} is a directed structure on N nodes/vertices defining the factorization of the model:

$$p(\mathbf{X}_{\mathbf{t}} | \bar{E}, \Theta) = \prod_{i=1}^T \prod_{v=1}^N p(\mathbf{x}_{\mathbf{t}(i)}^v | \tilde{\mathbf{x}}_{\mathbf{t}(i)}^{v, \text{pa}(v, \bar{E})}, \Theta_{v | \text{pa}(v, \bar{E})}) \quad (2)$$

where $\text{pa}(v, \bar{E})$ returns the parents of vertex v given the structure \bar{E} . We will drop the \bar{E} and use $\text{pa}(v)$ when it is clear from the context. Each time-series v at time t is dependent on its own past $\tilde{\mathbf{x}}_t^v$ as well as the past of its parent set $\mathbf{S} = \text{pa}(v)$, $\tilde{\mathbf{x}}_t^{\mathbf{S}}$. Note that we use $\Theta_{v | \mathbf{S}}$ rather than the more explicit notation $\Theta_{v | v, \mathbf{S}}$ to represent the parameters of this relationship for brevity.

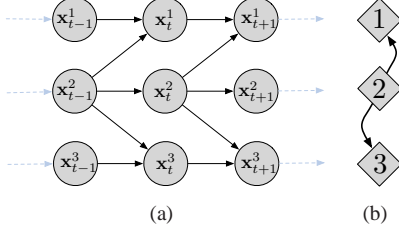


Fig. 1. Example directed tree SVARM(1) model (a) and corresponding interaction graph (b)

In general we refer to this class of models as r th order temporal interaction models, TIM(r). In this paper we focus on the special case in which \bar{E} describes a directed tree or forest and Θ parameterizes a linear Gaussian relationships. We refer to such models as directed tree or forest structured vector autoregressive models, SVARM(r).

Figure 1(a) illustrates a SVARM(1) for three time-series with \bar{E} describing a directed tree with the second time-series as the root. Here $\text{pa}(1) = \text{pa}(3) = 2$ and $\text{pa}(2) = \emptyset$. Figure 1(b) shows an alternative and more compact view for this model in which a single node represents a time-series/data stream over all time. We refer to this as the *interaction graph* and use diamond shaped nodes to emphasize it is not meant to be interpreted as a directed Bayesian network, though there is a one to one mapping between these graphs. A directed edge from u to v in the interaction graph implies a directed edge from \tilde{x}_t^u and x_t^v in the SVARM.

III. CONJUGATE PRIOR

In this section we introduce a conjugate prior on the parameters, θ , and structure, \bar{E} of the SVARM(r). The conjugacy of the prior yields a simple form for posterior updates. Tractable calculation of the exact posterior depends on the ability to efficiently evaluate a partition function which sums over a super-exponential number of structures. We show that directed trees and forests are classes of structures for which this can be done in polynomial time. This allows one to efficiently calculate the posterior on structure given data, while integrating out parameters.

The prior we use is related to those presented in [9] and [10]. It factors as

$$p_0(\bar{E}, \Theta) = p_0(\bar{E}) p_0(\Theta | \bar{E}). \quad (3)$$

The parameters are assumed to be independent and *modular* given a structure \bar{E} and hyper-parameters Γ . That is, they factorize according to the edges in \bar{E} :

$$p_0(\Theta | \bar{E}) = \prod_{v=1}^N p_0(\Theta_{v|\text{pa}(v)} | \Gamma) \quad (4)$$

and $p_0(\Theta_{v|\mathbf{S}} | \Gamma)$ is the same for all structures \bar{E} for which \mathbf{S} is the parent set of v . Thus, for each time-series v one needs to specify a parameter prior for all potential parent sets. Given this finite set of priors for each v one is able to supply a full prior on parameters for any given structure.

We place a prior on directed structures which has the form:

$$p_0(\bar{E}) = \frac{1}{Z(\beta)} \prod_{v=1}^N \beta_{\text{pa}(v),v} \quad (5)$$

where the partition function $Z(\beta)$ ensures proper normalization. Each scalar hyper-parameter $\beta_{\mathbf{S},v}$ can be interpreted as a weight on the parent set \mathbf{S} for v and the prior is simply a proportional to the product of these weights.

Let $\mathcal{D} = \{\mathbf{X}_1, \dots, \mathbf{X}_T\}$ be a set of T complete observations. We use the notation $\mathcal{D}^{\mathbf{S}} = \{\mathbf{x}_1^{\mathbf{S}}, \dots, \mathbf{x}_T^{\mathbf{S}}\}$ to denote observations of a set of time-series. $\mathcal{D}_t = \mathbf{X}_t$ and $\mathcal{D}_t^u = \mathbf{x}_t^u$. $\tilde{\mathcal{D}} = \{\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_T\}$ can be formed using \mathcal{D} and past information in \mathcal{C} . Given data \mathcal{D} , the posterior on parameters given structure has the form:

$$p(\Theta | \bar{E}, \mathcal{D}) = \prod_{v=1}^N p(\Theta_{v|\text{pa}(v)} | \mathcal{D}^v, \mathcal{D}^{\text{pa}(v)}, \Gamma) \quad (6)$$

That is, the prior on parameters given a structure is fully conjugate. If one chooses a conjugate prior for each $\Theta_{v|\mathbf{S}}$ in Equation 4 the posterior has the same form and can be updated by sufficient statistic calculations from the data. In addition, the posterior on structure is:

$$p(\bar{E} | \mathcal{D}) = \frac{1}{Z(\beta \circ W)} \prod_{v=1}^N \beta_{\text{pa}(v),v} W_{\text{pa}(v),v} \quad (7)$$

where \circ is an element wise / Hadamard product and

$$W_{\mathbf{S},v} = \int p(\mathcal{D}^v | \tilde{\mathcal{D}}^{v,\mathbf{S}}, \Theta_{v|\mathbf{S}}) p_0(\Theta_{v|\mathbf{S}} | \Gamma) d\Theta_{v|\mathbf{S}} \quad (8)$$

is the evidence, $p(\mathcal{D}^v | \tilde{\mathcal{D}}^{v,\mathbf{S}})$, for time-series v given the time-series of its parent set defined by \mathbf{S} . That is, the prior is updated by modifying β with a set of evidence weights W . The proof follows from the fact that the prior on structure factorizes in the same manner as Equation 2.

In this paper, we assume continuous observations and a linear gaussian model with parameters $\Theta_{v|\mathbf{S}}$. We model $p_0(\Theta_{v|\mathbf{S}} | \Gamma)$ to be a matrix-normal-inverse-Wishart distribution with hyper-parameters Γ . This yields efficient updates for Equation 6 and $W_{v|\mathbf{S}}$ will be the evaluation of a Matrix-T distribution [11]. However, tractable calculation of Equation 7 requires one to evaluate $Z(\beta)$ efficiently. We show this can be done for directed trees and forests in the following section.

A. Directed Trees and Forests

In general there are 2^{N-1} possible parent sets for each v . This yields a super-exponential number of possible structures, 2^{N^2-N} , implying $Z(\beta)$ requires summing over a super-exponential number of terms. In this paper we focus on two particular subclasses of these structures, while still super-exponential in number, that can be reasoned over in polynomial time.

The first class, directed trees (spanning arborescences), yield structures in which all the time-series are dependent on each other with no temporal feedback. That is, directed trees restrict the interaction graph to be spanning and acyclic with no more than a single parent for each time-series. Thus, in the directed tree case, β is simply an $N \times N$ matrix and each hyper-parameter $\beta_{u,v \neq u}$ can be interpreted as a weight on the edge $u \rightarrow v$, and $\beta_{\emptyset,v} = \beta_{v,v}$ is a weight on a node being a root (having no parents). The edge set corresponding to the nonzero entries of β form a support graph. We assume this support graph is connected and contains at least one directed tree.

While there are N^{N-1} possible directed trees on N nodes, the Matrix Tree Theorem allows one to calculate $Z(\beta)$ in polynomial time. This theorem was used by Meila and Jaakkola [9] for reasoning over undirected trees. The undirected version of theorem is a special case of the often rediscovered real valued directed version developed by Kirchhoff [12]. The theorem allows one to calculate the weighted sum over all directed trees rooted at r , $Z_r(\beta)$ via:

$$Z_r(\beta) = \sum_{\bar{E} \text{ rooted at } r} \prod_{u \rightarrow v} \beta_{u,v} = \text{Cof}_{r,r}(\bar{Q}(\beta)) \quad (9)$$

where $\bar{Q}(\beta)$ is the Kirchhoff matrix with its u, v entry defined as:

$$\bar{Q}_{uv}(\beta) = \begin{cases} -\beta_{u,v} & 1 \leq u \neq v \leq N \\ \sum_{u'=1}^N \beta_{u',v} & 1 \leq u = v \leq N \end{cases} \quad (10)$$

and $\text{Cof}_{i,j}(M)$ is the i, j cofactor of matrix M . $\text{Cof}_{i,j}(\bar{Q}(\beta))$ is invariant to i and gives the sum over all weighted trees rooted at j . A proof can be found in [13].

By summing over all N possible roots one obtains

$$Z(\beta) = \sum_{v=1}^N \beta_{v,v} Z_v(\beta). \quad (11)$$

Thus, a straight forward implementation yields $O(N^4)$ time for calculating the partition function. However, as pointed out in [14], using the invariance of $\text{Cof}_{i,j}(\bar{Q}(\beta))$ allows for $O(N^3)$ time computation of $Z(\beta)$. That is, $Z(\beta)$ can be calculated by replacing any row of the matrix $\bar{Q}(\beta)$ with $[\beta_{1,1}, \dots, \beta_{N,N}]$ and taking its determinant.

Directed forests, remove the fully connected assumption and can have multiple roots. There are $(N+1)^{(N-1)}$ directed forest for N nodes, but $Z(\beta)$ can still be calculated

in $O(N^3)$ time [14]. Some intuition as to why this is true is that any directed forest can be turned into a directed tree by the addition of one virtual super root node which has no parents and connects to all the roots in the forest.

B. Event Probabilities and Expectations

The ability to compute the partition function and conjugacy of the prior allows one to calculate a wide variety of useful prior and/or posterior event probabilities. For example, the probability of a particular edge being present is:

$$p(\mathbb{I}_{u \rightarrow v} = 1) = \mathbb{E}[\mathbb{I}_{u \rightarrow v}] = 1 - \frac{Z(\beta^{-(u \rightarrow v)})}{Z(\beta)} \quad (12)$$

where $\mathbb{I}_{u \rightarrow v}$ is an indicator variable that has value 1 when the edge $u \rightarrow v$ is present. $\beta^{-(u \rightarrow v)}$ is β with all elements involving edge from u to v set to zero. Using the same approach one can calculate the joint edge appearance probability of one set of edges conditioned on another set. Similarly one can calculate the probability a time-series/node has no parents (is a root):

$$p(\mathbb{I}_v \text{ is a root}) = \frac{Z(\beta^{-\text{edgesin}(v)})}{Z(\beta)}, \quad (13)$$

where $\text{edgesin}(v)$ and return the set of all edges into time-series/node v . β^{-e} indicates all elements of β which involve any edge in the set e are zero. Similarly one can use $\beta^{-\text{edgesout}(v)}$ to calculate the probability of a time-series having no children (is a leaf).

The indicator variables used in the examples above can be expressed as general multiplicative functions of the form $g(\bar{E}) = \prod_{v=1}^N g_{\text{pa}(v),v}$. The expected value of a general multiplicative function can be calculated by:

$$\mathbb{E}[g(\bar{E})] = \frac{Z(\beta \circ g)}{Z(\beta)}. \quad (14)$$

Note that variance or other higher order moments of multiplicative functions can also be calculated in this manner. Additionally, one can calculate the expectation of additive functions of the form $f(\bar{E}) = \sum_{v=1}^N f_{\text{pa}(v),v}$. For directed trees $\mathbb{E}[f(\bar{E})] =$

$$\sum_{r=1}^N \frac{Z_r(\beta)}{Z(\beta)} \text{tr} \left(M_{r,r}(\bar{Q}(\beta \circ f)) M_{r,r}(\bar{Q}(\beta))^{-1} \right) \quad (15)$$

where $M_{i,j}(M)$ is the matrix M with its i th row and j th column removed. A proof follows that of [9] substituting in the directed tree partition function in place of the undirected version. A similar form is obtained for directed forests. Additive functions allow calculation of quantities such as the expected number of children of a particular time-series/node.

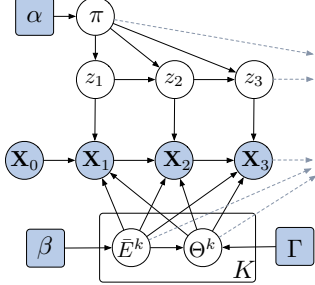


Fig. 2. The structure of a SSVARM(1,K)

C. Sampling directed trees and forests

Given data, we have shown how to calculate the posterior $p(\bar{E}|\mathcal{D})$ and various event probabilities. Another important task is sampling from this posterior. Random sampling of directed spanning trees (and forests with some modification) is a well studied problem. This paper uses Wilson’s random walk based algorithm [15]. If a point estimate of structure is desired, the maximum a posteriori structure can also be found. The Chu-Liu/Edmonds/Bock [16] algorithm can be used for both directed trees and forests

IV. DYNAMIC SWITCHING STRUCTURES

The SVARM(r) model is stationary and assumes a single \bar{E} and parameters Θ over all time. Here, we introduce a switching SVARM, SSVARM(r, K), which allows structure to change over time. Let z_t be a hidden state at time t which indexes a specific structure, E^{z_t} , and parameters, Θ^{z_t} . Figure 2 shows a first order model. Given a set of K structures $\mathbf{E} = \{\bar{E}^1, \dots, \bar{E}^K\}$, parameters $\Theta = \{\pi^0, \pi^1, \dots, \pi^K, \Theta^1, \dots, \Theta^K\}$ and an observations over the time period $\mathbf{t} = [1, \dots, T]$:

$$\begin{aligned} p(\mathbf{X}_{\mathbf{t}}, z_{\mathbf{t}}|\mathbf{E}, \Theta) &= p(z_{\mathbf{t}})p(\mathbf{X}_{\mathbf{t}}|z_{\mathbf{t}}, \mathbf{E}, \Theta) \\ &= \prod_{t=1}^T p(z_t|\pi^{z_{t-1}})p(\mathbf{X}_t|\tilde{\mathbf{X}}_t, E^{z_t}, \Theta^{z_t}) \end{aligned} \quad (16)$$

where $z_0 = 0$, the transition distribution is multinomial $p(z_t|\pi^{z_{t-1}}) = \text{Mult}(z_t; \pi^{z_{t-1}})$ and is given a Dirichlet prior $p_0(\pi) = \prod_{k=1}^K \text{Dir}(\pi^k; \alpha_1^k, \dots, \alpha_K^k)$.

Exact inference on this model is complex due to the fact there are K^T possible state sequences. Thus, we turn to a Monte Carlo Markov chain (MCMC) approach in which samples are drawn from this model using a Gibbs sampler. The sampler has three main steps. Step 1 samples the state sequence given previous estimate of structures and parameters. This is done efficiently with backward message passing followed by forward sampling. This step can be modified when initializing to sample the state sequence from its transition prior. During step 2 the sampled counts of state transitions are noted and transition probabilities

are then sampled given these counts. In Step 3 a vector \mathbf{t}_k is formed with all the time points with $z_t = k$. The structure and parameters are then sampled given $\mathbf{X}_{\mathbf{t}_k}$ for each k . It is important to note that given a state sequence, one can efficiently calculate exact event probabilities and posterior over structures.

V. EXPERIMENTS

In this section we present two illustrative experiments focusing on the calculation of posterior event probabilities. Each experiment analyzes the interaction of tracked moving objects. Specifically, we are interested in quantifying uncertainty in the dependence structure among time-series rather than obtaining point estimates. When analyzing data we do not assume there is a “true/correct structure” one would like to discover. Our goal is to obtain a full characterization of the posterior uncertainty.

A. Interacting People

We begin by analyzing the motion of two individuals in an enclosed environment. Individuals are instructed to switch between moving independently and following each other. A simple tracker outputs the (x, y) position of each individual over all time and ground-truth indicating when interaction occurred was created.

In [6] a dependence test was performed on this data using a hidden factorization Markov model (HFactMM). An HFactMM allows switching between latent states each of which having a known/specified fixed structure. For this data two states were used and the structures were set to be independent in state 1 and fully connected/dependent in state 2. The dependence test was performed by finding the parameters that maximized the likelihood of the data assuming these fixed structures and then Viterbi decoding was performed to segment the sequence. A probability of error of .035 was obtained when comparing to the ground-truth.

Here, we will not assume known structures. We place a uniform prior on all directed forests and weak matrix-normal-inverse-Wishart prior on parameters (small degrees of freedom, and larger variance). Using a two state, first order model, SSVARM(1,2), we generate 100 Gibbs samples. Burn in was quick and required approximately 15 iterations of the sampler.

Figure 3(a) summarizes the results. The ground truth state sequence is shown on top. The individuals start out being independent and then randomly switch between that and the following behavior. The 100 sampled state sequences are shown below the ground truth ranked by the log probability of the data given the sampled parameters and structures, with the top being the most likely. The state sequence labels were permuted to give a consistent coloring with the ground truth segmentation. The side of the plot shows the normalized Hamming distance of the

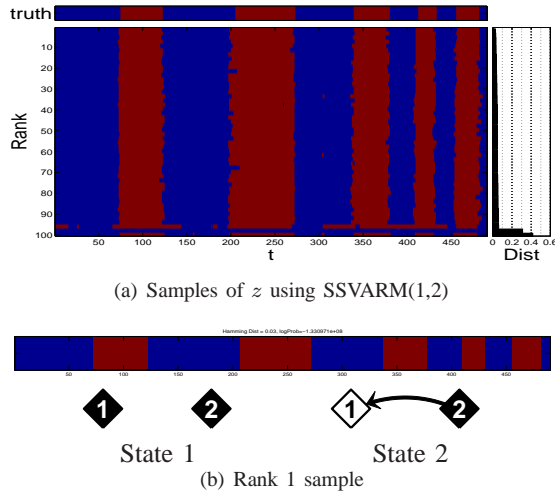


Fig. 3. Results for 2 person interaction experiment.

best mapping to the ground truth. Note that these samples indicated a very sharp posterior. Almost all of the samples give a very consistent segmentation of the data and have an average Hamming distance of .05. This distance can be interpreted as the probability of error if we were only concerned with segmenting the data.

However, Figure 3(a) only shows the sampled state sequence. Figure 3(b) shows a more detailed breakdown of the sampled model with the highest log probability. The top row shows the sampled state sequence. The second row represents the posterior probability of the structure for each state depicted as weighted interaction graphs. In these graphs edge color represents the posterior probability of that edge. Node color represents the probability a time-series has no parents/is a root (white = 0.0, black = 1.0). Recall that while the state sequence is MCMC sampled, we obtain an exact posterior conditioned on this sequence. We see that for this data the posterior is very sharp. The posterior for the first state is sharp around independent structure. The second state has a strong posterior on the second time-series/individual influencing the first. This is consistent with the fact that first person follows the second when they are interacting. The Hamming distance for this particular state sequence is .03. That is, we obtain comparable results to those obtained in [6] without fixing the two possible structures a priori.

B. Follow The Leader

Next we explore a dataset comprised of recordings of three individuals playing a simple interactive computer game. Each player’s computer mouse controls a specific dot/marker on their screen. All players can see three dots on their screen representing the other players in game and themselves. The individuals are instructed to play a game

of follow the leader. One player is designated the leader. The leader moves his or her dot randomly around the screen while the other players are instructed to follow. The designated leader changes throughout the game. That is, a fourth person observing the game tells the players when to switch leaders. Thus, the latent variable indicating the change of leader is observable, and consequently, nominal ground-truth is available by which to evaluate performance.

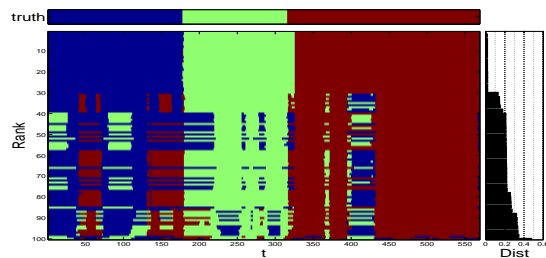
We begin by using SSVARM(1,3) with directed trees to analyze this sequence. That is, we will use the fact there are only three players and knowledge that directed trees may sufficiently describe the interaction among players. A uniform prior on structures and equivalently weak prior on parameters is used. A weak self biased prior on the state transition distribution is imposed with a bias towards self transition.

Given the data and the prior model, 100 samples of the structure, parameters and the hidden state sequence are obtained with a Gibbs sampler. Burn in required approximately 60 iterations. A detailed view of the results can be found in Figure 4(a) in the same format as shown in Figure 3(a). The ground-truth state sequence is shown on top with players taking turns being the leader in order. The sampled state sequences show some posterior uncertainty. Each sample falls within two general categories. The top third of the samples match the ground truth closely, the bottom two thirds suggests a consistent alternative explanation.

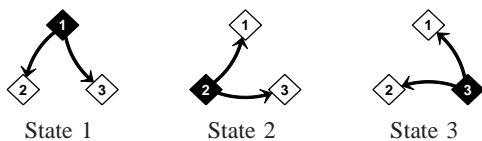
Again, the state labels alone simply provide a segmentation. Given this segmentation we look at the posterior on structure to analyze the interaction among the players. Figures 4(b) and 4(c) show a more detailed breakdown of two sampled models. Figure 4(b) is a sample with low Hamming distance and high log probability. Notice that the posterior on structure for each state is basically a delta function on three distinct structures. These structures agree with our intuition in that each root is consistent with who was designated as the leader and the followers are conditionally independent given the root. Figure 4(c) is a sample with a mid-range Hamming distance (ranked 34 out the 100 samples). It has errors consistent with the majority of sequences shown in 4(a). The confusion between the first and third state is most noticeably reflected in posterior uncertainty in the structure for state 3.

While the above analysis assumed three states, consistent with our knowledge of the ground truth, Figure 4(c) gives evidence for additional modes/states. That is, for each phase of the game a better model may be a mixture of processes each with similar structure but different parameters. We repeat the experiment using $K = 6$ states. Figure 4(d) is a sample from this model.

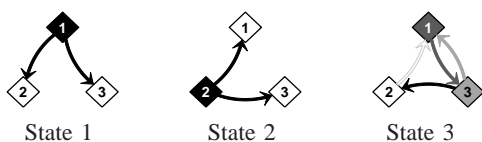
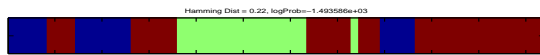
The second row shows the occurrence of the learned states. Interestingly, state 4 indicates uncertainty in the structure. However, this state is never used and thus its



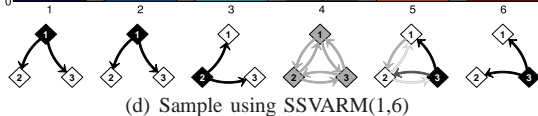
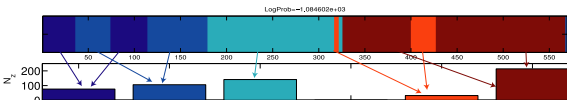
(a) Samples of z using SSVARM(1,3)



(b) Rank 1 sample



(c) Rank 34 sample



(d) Sample using SSVARM(1,6)

Fig. 4. Results for follow the leader experiment.

posterior remains uniform. The remaining structures are consistent with the ground truth indicating who is the leader with little uncertainty.

VI. CONCLUSION

We have presented a framework for Bayesian inference of the dependency structure among multiple vector time-series. Assuming linear Gaussian dynamics we presented a structure autoregressive model in which directed trees and forests were used to describe causal interactions among time-series. We showed how a super-exponential number of structures can be reasoned over in polynomial time and how exact posterior event probabilities can be obtained. By introducing an evolving latent variable to index these models we allowed for dynamically switching structures. We demonstrated the utility of our method by analyzing the interaction of moving objects.

Here we have focused on acyclic structures in which each time-series has at most one parent. However, the framework presented is general and can be used for any class of directed structures. Tractable calculation of exact posterior probabilities is dependent on the ability to efficiently evaluate the partition function $Z(\beta)$. Other structures for which $Z(\beta)$ can be tractably calculated are discussed in [17].

The SSVARM(r, K) model used in this paper assumes a known number of states, K , and that structures can be revisited over time. Such assumptions were reasonable in our applications of interest. However, the SVARM(r) can easily be embedded into alternative dynamic models in which the number of states are unknown such as those presented in [5] and [18].

REFERENCES

- [1] F. Heider and M. Simmel, "An experimental study of apparent behavior," in *American Journal of Psychology*, 1944, vol. 57.
- [2] F.R. Bach and M. I. Jordan, "Learning graphical models for stationary time series," in *Trans. on signal processing*, 2004, vol. 52, pp. 2189–2199.
- [3] J. A. Bilmes, "Dynamic bayesian multinets," in *UAI*, 2000.
- [4] S. Kirshner, P. Smyth, and A. Robertson, "Conditional chow-liu tree structures for modeling discrete-valued vector time series," in *Technical Report UCI-ICS*, 2004.
- [5] X. Xuan and K. Murphy, "Modeling changing dependency structure in multivariate time series," in *ICML*, 2007.
- [6] M.R. Siracusa and J.W. Fisher, "Inferring dynamic dependency with applications to link analysis," in *Asilomar Conference on Signals, Systems, and Computers*, 2006.
- [7] D. Geiger and D. Heckerman, "Knowledge representation and inference in similarity networks and bayesian multinets," in *Artificial Intelligence*, 1996, vol. 82, pp. 45–74.
- [8] B. Milch, B. Marthi, D. Sontag, S. Russell, D.L. Ong, and A. Kolobov, "Approximate inference for infinite contingent bayesian networks," in *AIStats*, 2005.
- [9] M. Meila and T. Jaakkola, "Tractable bayesian learning of tree belief networks," in *Statistical Computing*, 2006, pp. 77–92.
- [10] N. Friedman and D. Koller, "Being bayesian about network structure: a bayesian approach to structure discovery in bayesian networks," *Machine Learning*, vol. 50, pp. 95–125, 2003.
- [11] M. West and J. Harrison, *Bayesian Forecasting and Dynamic Models*, Springer, 1997.
- [12] G. Kirchhoff, "ueber die auflsung der gleichungen, auf welche man bei der untersuchung der linearen vertheilung galvanischer strme gefhrt wird," in *Annalen der Physik und Chemie*, 1847, vol. 72, pp. 497–508.
- [13] W.T. Tutte, *Graph Theory*, Addison-Wesley Publishing Co., 1984.
- [14] T. Koo, A. Globerson, X. Carreras, and Michael Collins, "Structured prediction models via the matrix-tree theorem," in *EMNLP*, 2007.
- [15] D. B. Wilson, "Generating random spanning trees more quickly than the cover time," in *Theory of Computing*, 1996, pp. 296–303.
- [16] Y. J. Chu and T. H. Liu, "On the shortest arborescence of a directed graph," *Science Sinica*, vol. 14, pp. 1396–1400, 1965.
- [17] M.R. Siracusa and J.W. Fisher, "Tractable bayesian inference of time-series dependence structure," in *LIDS Tech Report*, MIT, 2008.
- [18] Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei, "Hierarchical dirichlet processes," *Journal of the American Statistical Association*, vol. 101(476), 2006.

This research was partially supported by the Air Force Office of Scientific Research via Grant FA9550-06-1-0324 and the Army Research Office via grant W911NF-06-1-0076.