Bayesian Switching Interaction Analysis Under Uncertainty

Zoran Dzunic
CSAIL, MIT

John W. Fisher III
CSAIL, MIT

Abstract

We introduce a Bayesian discrete-time framework for switching-interaction analysis under uncertainty, in which latent interactions, switching pattern and signal states and dynamics are inferred from noisy (and possibly missing) observations of these signals. We propose reasoning over full posterior distribution of these latent variables as a means of combating and characterizing uncertainty. This approach also allows for answering a variety of questions probabilistically, which is suitable for exploratory pattern discovery and post-analysis by human experts. This framework is based on a fully-Bayesian learning of the structure of a switching dynamic Bayesian network (DBN) and utilizes a state-space approach to allow for noisy observations and missing data. It generalizes the autoregressive switching interaction model of Siracusa et al. [1], which does not allow observation noise, and the switching linear dynamic system model of Fox et al. [2], which does not infer interactions among signals. Posterior samples are obtained via a Gibbs sampling procedure, which is particularly efficient in the case of linear Gaussian dynamics and observation models. We demonstrate the utility of our framework on a controlled human-generated data, and climate data.

1 Introduction

We consider the problem of inferring time-varying interactions over multi-dimensional time-series data. In contrast to previous work we address noisy and uncertain measurement models and the possibility of missing data. Furthermore, by construction of the model we allow for marginalization over model parameters enabling efficient Bayesian inference over discrete structure elements. Analyzing such interactions is important in many domains such as biology, finance, social networks, Earth sciences, transportation, games, etc. We formulate this as the problem of learning the structure of a switching dynamic Bayesian network as depicted in Figure 1.

Inferring interaction structures over multi-dimensional time-series presents a formidable challenge owing to the super-exponential number of possible directed graphs. With limited data available, there may exist a large number of structures that explain the data well. Structure point estimates (e.g., MAP) are likely to yield incorrect interactions. The problem is exacerbated when the structure varies over time and time-series state is not observed directly, but rather by some noisy observation process.

Here, we propose a Bayesian approach for reasoning over interaction structures. The resulting model allows for efficient calculation of marginal event probabilities corresponding to such questions as “Does object A depend on object B, given that it interacts with object C?” or “Which object is the most influential, i.e., has the most objects that depend on it?”.

Figure 1: Dynamic Bayesian Network (DBN) representation of switching interaction among four signals. They initially evolve according to interaction graph $E_1$. At time point 4, the interaction pattern changes, and they evolve according to interaction graph $E_2$. Self-edges are assumed.
To this end, we develop a Bayesian switching state-space interaction model (SSIM), presented in Section 5, that accounts for noisy and missing observations. We define interaction graphs in Section 3. To address computational challenges, we use a modular prior over structures with bounded in-degree, as described in Section 4. A Gibbs sampling inference procedure is presented in Section 6. In addition, we derive a novel numerically-stable message-passing algorithm for “batch” sampling of a state sequence in a linear Gaussian state-space model that allows deterministic dependence relationships (Section 6.1). In Section 7, we demonstrate utility of our approach on realistic human-generated data as well as real data.

2 Related work

The proposed model integrates inference over structures, dynamic switching, and latent state-space models. All have been the subject of extensive research. Change point detection was first a subject of interest in the area of quality control, but has since become an important problem in time-series analysis domains. A huge number of online and offline, Bayesian and non-Bayesian, parametric and nonparametric methods have been developed. Basseville and Nikiforov [3] and Polunchenko and Tartakovsky [4] provide an overview of these methods. Most of these methods assume segment independence. In contrast, switching dynamic systems (SDS) – also called state-space switching models (SSM) – allow coupling between segments through dynamics parameters, which is typically modeled via latent switching states. They combine state-space modeling with switching point detection. Inference in SDS models is done via approximate methods (Pavlovic et al. [5, 6]), EM algorithm (Oh et al. [7]), or sampling (Fox et al. [8, 2]). Most of related work deals with switching linear dynamic systems (SLDS) since they allow for simpler inference but are still widely applicable.

In recent years, a number of methods for learning changing structure among time-series have been suggested. For example, Xuan and Murphy [9] combine inference over undirected graphs with change-point detection. Optimization techniques have been used to estimate time-varying undirected networks (Kolar et al. [10]), as well as time-varying DBNs (Song et al. [11]). Jiang et al. [12] use EM algorithm to obtain the MAP estimate of a switching DBN. Lebre et al. [13] and Robinson and Hartemink [14] use MCMC sampling method to learn time-varying DBNs. However, the number of sampled structures may not be sufficient to adequately represent the posterior over structures. Siracusa and Fisher [1] develop a method based on prior modularity for efficient reasoning over the structure posterior. The model we propose is most closely related to the work of [1]. It differs (in fact, from most available methods) in that we do not assume direct observation and allow for missing data. The result is a more expressive and robust model at the cost of a more complex inference procedure.

3 Interaction graphs and DBN

Our goal is to reason over time-varying interactions (dependence structures) between N multivariate signals. We assume that signals evolve according to a Markov process over discrete time points \( t = 0, 1, \ldots, T \). The latent state associated with signal \( i \) at time point \( t \) depends on the state of a subset of signals \( pa(i, t) \) at time point \( t - 1 \). We refer to \( pa(i, t) \) as a parent set of signal \( i \) at time point \( t \). While the preceding implies a first-order Markov process, the approach extends to higher-ordered Markov processes. A collection of directed edges \( E_t = \{ (v, i) ; i = 1, \ldots, N, v \in pa(i, t) \} \) forms an interaction graph at time point \( t \), \( G_t = (V, E_t) \), where \( V = \{ 1, \ldots, N \} \) is the set of all signals. That is, there is an edge from \( j \) to \( i \) in \( G_t \) if and only if signal \( i \) at time point \( t \) depends on signal \( j \) at time point \( t - 1 \). We say that the parent signals \( pa(i, t) \) influence signal \( i \) at time \( t \).

Let \( X^t_i \) denote a (multivariate) random variable that describes the latent state associated to signal \( i \) at time point \( t \). Then, signal \( i \) depends on its parents at time \( t \) according to a probabilistic model \( p(X^t_i | X^{t-1}_{t-1}, \theta_t) \) parametrized by \( \theta_t \). Furthermore, we assume that conditioned on their parents at the previous time point, signals are independent of each other:

\[
p(X^t_i | X^{t-1}_{t-1}, E_t, \theta_t) = \prod_{i=1}^{N} p(X^t_i | X^{pa(i,t)}_{t-1}, \theta_t^i), \tag{1}
\]

where \( X_t = \{ X^t_i \}_{i=1}^{N} \) (i.e., \( X_t \) is a collection of variables of all signals at time point \( t \)) and \( \theta_t = \{ \theta_t^i \}_{i=1}^{N} \). Structure \( E_t \) and parameters \( \theta_t \) determine a dependence model at time \( t \), \( \mathcal{M}_t = (E_t, \theta_t) \). Finally, we express a joint probability of all variables at all time points, \( X \), as

\[
p(X) = p(X_0 | \theta_0) \prod_{t=1}^{T} p(X_t | X_{t-1}, E_t, \theta_t)
\]

\[
= \prod_{i=1}^{N} p(X^t_i | \theta_t^i) \prod_{t=1}^{T} \prod_{i=1}^{N} p(X^t_i | X^{pa(i,t)}_{t-1}, \theta_t^i). \tag{2}
\]

The stochastic process of Eq. 2 can be represented using a dynamic Bayesian network (DBN), such that
there is a one-to-one correspondence between the network and the collection of interaction graphs over time, as shown in Figure 1.

4 Prior on interaction structure

Let us assume for a moment that the dependence model is homogenous in time, i.e., \( E_t \equiv E, \) \( \rho a(t, t) \equiv \rho a(i) , \) and \( \theta_t \equiv \theta . \) Let \( p(E; \beta) \) be the prior probability of structure \( E, \) parameterized by \( \beta . \) In the most general form, \( \beta \) is a collection of parameters \( \{ \beta_E \} \) (one parameter for each structure), such that \( \beta_E \) is proportional to the prior probability of \( E: \)

\[
p(E; \beta) = \frac{1}{B} \beta_E \propto \beta_E ,
\]

where \( B = \sum \beta_E \) is a normalization constant.

Let \( p(\theta|E; \gamma) \) be the prior probability of \( \theta, \) parameterized by \( \gamma . \) For now, we do not assume any particular form of the dependence models, \( p(X_t|X_{t-1}^{pa(i)}, \theta^t) \). Note however that the prior on parameters, \( \theta, \) may depend on the structure. Since different structures may differ in the number of parents (for some signals), they may also require parameters of different dimensionality. Thus, \( \gamma \) is indeed a collection \( \{ \gamma_E \} \) of sets of hyperparameters, such that \( p(\theta|E; \gamma) = p(\theta|\gamma_E) . \)

Learning (static) Bayesian network structures (under reasonable assumptions) is NP hard [15]. The number of possible structures is superexponential in the number of nodes, which is also true for the number of interaction graphs \( (2^N)^N \). A modular prior on structure and parameters [16, 17, 18, 19] is typically introduced to reduce the complexity of inference over structures:

- \( p(E; \beta) = \prod_{i=1}^N p(\rho a(i); \beta) \) (structure modularity)
- \( p(\theta|E; \gamma) = \prod_{i=1}^N p(\theta^t|E; \gamma) \) (global parameter independence)
- \( p(\theta^t|E; \gamma) = p(\theta^t|\rho a(i); \gamma) \) (param. modularity).

This assumption is, for static BNs, typically combined with a known ordering of variables [16, 17] or a procedure for sampling such orderings [19]. Since loops are allowed in the interaction graph (Figure 1), ordering of variables/signals is irrelevant and parent sets can be chosen independently for each signal [1]. Therefore, the modular prior assumption directly reduces the number of structures to exponential \( (N2^N) \). As will be discussed in Section 6.2, modularity and independence of parent sets are reflected in the structure posterior. Note that as a result of modular prior assumption, \( \beta \) is no longer a collection of parameters per structure, but rather a collection of parameters \( \{ \beta_{i,pa(i)} \} \) (one parameter for each possible parent set of each signal), such that

\[
p(\rho a(i); \beta) = \frac{1}{B_i} \beta_{i,pa(i)} \propto \beta_{i,pa(i)} ,
\]

where \( B_i = \sum \beta_{i,s} \) are normalization constants.

If, in addition, the number of parents of each signal is bounded by some constant \( M \) (a structure with bounded in-degree [17, 18, 19]), the number of parent sets to evaluate is further reduced to \( O(N^{M+1}) \), which is polynomial in \( N \), while the total number of structures with non-zero probability is still superexponential \( (2^{O(NlogN)}) \). More generally, using any polynomially large subset of parent sets per signal leads to a polynomial complexity of structure inference.

5 State-space switching interaction model (SSIM)

In order to learn time-varying interaction from time-series data, we assume that the dependence model switches over time between \( K \) distinct models, \( \mathcal{M}_k = (\hat{E}_k, \hat{\theta}_k), k = 1, \ldots, K. \) More formally, for each time point \( t, \mathcal{M}_t = \mathcal{M}_{Z_t} \) for some \( k, 1 \leq k \leq K. \) One interaction may be active for some period of time, followed by a different interaction over another period of time, and so on, switching between a pool of possible interactions. This is illustrated in Figure 1. Let \( Z_t, 1 \leq t \leq T, \) be a discrete random variable that represents an index of a dependence model active at time point \( t; \) i.e., \( \mathcal{M}_t = \mathcal{M}_{Z_t}, Z_t \in \{1, \ldots, K\}. \) We can now rewrite the transition model (Equation 1) as

\[
p(X_t|X_{t-1}, Z_t, \hat{E}, \hat{\theta}) = p(X_t|X_{t-1}, \hat{E}_{Z_t}, \hat{\theta}_{Z_t})
\]

\[
= \prod_{i=1}^N p(X_t|X_{t-1}^{\rho a(i,Z_t)}, \hat{\theta}_{Z_t}) ,
\]

where \( \{(\hat{E}_k, \hat{\theta}_k)\}_{k=1}^K \) is a collection of all \( K \) models and \( \rho a(i,k) \) is a parent set of signal \( i \) in \( \hat{E}_k. \) We can also rewrite Equation 2 as

\[
p(X|Z, \hat{E}, \hat{\theta}) = p(X|\hat{\theta}_0) \prod_{t=1}^T p(X_t|X_{t-1}, Z_t, \hat{E}, \hat{\theta}) ,
\]

where \( Z = \{Z_t\}_{t=1}^T. \) To distinguish from signal state, we call \( Z_t \) a switching state (at time \( t \)) and \( Z \) a switching sequence. Furthermore, we assume that \( Z \) forms a first order Markov chain:

\[
p(Z) = p(Z_1) \prod_{t=2}^T p(Z_t|Z_{t-1}) = \pi Z_1 \prod_{t=2}^T \pi_{Z_{t-1},Z_t} ,
\]

where \( \pi_{i,j} \) is a transition probability from state \( i \) to state \( j \) and \( \pi_i \) is the initial probability of state \( i. \)

Finally, we model that the observed value \( Y^t_i \) of signal \( i \) at time \( t \) is generated from its state \( X^t_i \) via a probabilistic observation model \( p(Y^t_i|X^t_i, \xi^t_i) \) parametrized
Bayesian Switching Interaction Analysis Under Uncertainty

Z2 Z3 ZT- 1 ZT
X1 X2 X3 XT- 1 XTX0
Y1 Y2 Y3 YT- 1 YTY0

by \( \xi_i \). For simplicity, we assume that the observation model is independent of the state \( (\xi_i = \xi^t, \forall t, i) \),

\[
p(Y|X, \xi) = \prod_{t=0}^{T} \prod_{i=1}^{N} p(Y^t_i|X^t_i, \xi^t),
\]

where \( Y = \{ Y^T_{i=1} \} \) is the observation sequence and \( \xi \) is the collection of parameters \( \{ \xi^t \}_{t=1}^{N} \).

The full SSIM generative model, shown in Figure 2, incorporates probabilistic models described above along with priors on structures and parameters:

- Multinomials \( \pi \) are sampled from Dirichlet priors parametrized by \( \alpha \) as
  \[
  (\pi_1, \ldots, \pi_K) \sim \text{Dir}(\alpha_1, \ldots, \alpha_K),
  \]
  \[
  (\pi_1, \ldots, \pi_i, K) \sim \text{Dir}(\alpha_1, \ldots, \alpha_i, K) \forall i.
  \]
- \( K \) structures \( E_k \) and parameters \( \theta_k \) are sampled from the corresponding priors as
  \[
  E_k \sim p(E; \beta), \quad \theta_k \sim p(\theta; E_k; \gamma), \quad \forall k.
  \]
- Parameters of the observation model are sampled as \( \xi^t \sim p(\xi^t; \delta), \quad \forall i. \)
- Initial values \( X_0 \) and \( Y_0 \) are generated as \( X_0 \sim p(X_0|\theta_0) \) and \( Y_0 \sim p(Y_0|X_0, \xi). \)
- For each \( t = 1, 2, \ldots, T \) (in that order), values of \( Z_t, X_t \) and \( Y_t \) are sampled as
  \[
  Z_t \sim \text{Mult}(p_{Z_{t-1}, \ldots, Z_{t-1}, K}) \quad \text{or}
  \]
  \[
  Z_t \sim \text{Mult}(p_{Z_{t}, K}) \quad \text{if } t = 1,
  \]
  \[
  X_t \sim p(X_t|X_{t-1}, E_{Z_t}, \theta_{Z_t}) \quad \text{and} \quad Y_t \sim p(Y_t|X_t, \xi).
  \]

The choice of dependence and observations models is application specific and will impact the complexity of some of the inference steps. For example, commonly used linear Gaussian models (Section 5.1) allow efficient inference in state space models, which is a subprocedure in our sampling algorithm (step 1 in Algorithm 1). Also, the choice of conjugate priors on parameters of dependence and observation models results in closed form expressions for sampling steps 4 and 5 in Algorithm 1, respectively. In this paper, we focus on linear Gaussian models and their conjugate priors, as described in Section 5.1.

5.1 Linear Gaussian SSIM (LG-SSIM)

Linear Gaussian state-space switching interaction models (LG-SSIM) are an instance of SSIM in which the dependence and observation models of each signal \( i \) at each time point \( t \) are linear and Gaussian:

\[
X_t^i = A^i_{Z_t} Y^i_{t-1}(Z_t) + w_t^i, \quad w_t^i \sim \mathcal{N}(0, Q^i_{Z_t})
\]

\[
Y_t^i = C^i X_t^i + v_t^i, \quad v_t^i \sim \mathcal{N}(0, R_t^i).
\]

\( A^i_k \) and \( Q^i_k \) are the dependence matrix and the noise covariance matrix of signal \( i \) in the \( k \)th dependence model (i.e., \( \theta_k^i = (A^i_k, Q^i_k) \)), while \( C^i \) and \( R_t^i \) are the observation matrix and the noise covariance matrix of the observation model of signal \( i \) (i.e., \( \xi^i = (C^i, R^i) \)).

We adopt a commonly used matrix normal inverse Wishart distribution as a conjugate prior on the parameters \( (A, Q) \) of a linear Gaussian model:

\[
p(A, Q; M, \Omega, \kappa, \Psi) = \mathcal{M}N(A; M, \Omega, Q) \mathcal{IW}(Q; \kappa, \Psi).
\]

Here, \( \kappa \) and \( \Psi \) are the degree of freedom and the inverse scale matrix parameters of the inverse Wishart distribution, while \( M, \Omega \) and \( Q \) are the mean, the row covariance and the column covariance parameters of the matrix normal distribution. Note that the two distributions are coupled. The matrix normal distribution of the parameter \( A \) depends on the parameter \( Q \) that is sampled from the inverse Wishart distribution.

6 Inference in SSIM and LG-SSIM

Exact inference for the SSIM is generally intractable. Consequently, we develop a Gibbs sampling procedure as described in Algorithm 1.

**Algorithm 1 SSIM Gibbs sampler**

1. \( X \sim p(X|Z, Y, E, \theta, \xi) \)
2. \( Z \sim p(Z|X, E, \theta, \pi) \)
3. \( \pi \sim p(\pi|Z; \alpha) \)
4. \( E, \theta \sim p(E, \theta|Z, X; \beta, \gamma) \)
5. \( \xi \sim p(\xi|X, Y; \delta) \)

This algorithm is similar to that of [1], but with two additional steps: sampling a latent state sequence \( X \) (step 1), and sampling parameters \( \xi \) of the observation model (step 5). Sampling parameters \( \pi \) of multinomials given the switching sequence \( Z \) (step 3) is straightforward as the Dirichlet distribution is conjugate to
the multinomial. The complexity of sampling parameters $\xi$ (step 5) depends on the particular observation model used. When a conjugate prior is available, as in the LG-SSIM, this step is similarly straightforward. Given the state sequence $X$ and the dependence models $\{\tilde{E}_k, \tilde{\theta}_k\}_{k=1}^K$, a sample of a switching sequence (step 2) is generated via a backward message-passing forward sampling algorithm, as in [1].

6.1 Sampling state sequence (step 1)

Conceptually, sampling a state sequence $X$ when all other variables in the model are known can be performed via the same backward message-passing forward sampling algorithm as in step 2. Note that the meaning of a backward message is $m^t(x) \propto P(Y_{t+1}, ..., Y_T | X_t = x, Z, E, \theta, \xi)$. If analytic expressions for messages are not available, MCMC methods such as particle filtering (e.g., [20]) may be used. However, in the case of linear Gaussian dependence and observation models, as in LG-SSIM, all messages have a form of a Gaussian distribution and can be compactly represented with their mean and covariance, which can be computed using the following standard information filter recursive equations (e.g., as in Fox et al. [2]):

$$
(S^m_t)^{-1} = A_{Z_t+1}^T (Q_{Z_t+1}^{-1} - Q_{Z_t+1}^{-1} \Sigma_t^i Q_{Z_t+1}^{-1}) A_{Z_t+1},
$$

$$
(S^m_t)^{-1} \mu^m_t = A_{Z_t+1}^T Q_{Z_t+1}^{-1} \Sigma_t^i \mu^o_t,
$$

where $\Sigma_t^o = C^T R^{-1} C + \Sigma_t^{-1}, \Sigma_t^{-1} \mu^o_t = C^T R^{-1} Y_{t+1} + \Sigma_t^{-1} \mu^m_t$, and $\Sigma_t^o = Q_{Z_t+1}^{-1} + \Sigma_t^{-1}$. For long sequences of missing data, $S^m_t$ approaches $Q_{Z_t+1}^{-1}$, and intermediate values $Q_{Z_t+1}^{-1} - Q_{Z_t+1}^{-1} \Sigma_t^i Q_{Z_t+1}^{-1}$ are close to singular. Via the matrix equality $(A + B)^{-1} = A^{-1} - (I + A^{-1} B)^{-1} A^{-1} B A^{-1}$, we derive alternative recursive equations that yields a numerically stable algorithm and allows for singular covariance matrices, which we exploit to impose deterministic constraints between variables across time:

$$
(S^m_t)^{-1} = A_{Z_t+1}^T \Sigma_t^i \Sigma_t^o A_{Z_t+1},
$$

$$
(S^m_t)^{-1} \mu^m_t = A_{Z_t+1}^T \Sigma_t^i \Sigma_t^o \mu^o_t,
$$

where $\Sigma_t^o$ and $\Sigma_t^{-1} \mu^o_t$ are as above, and $\Sigma_t^i = (I + \Sigma_t^{-1} Q_{Z_t+1})^{-1}$.

In the forward sampling procedure, at each step $t$ a sample of $X_t$ is drawn from a distribution $P(X_t | X_{t-1}, Z, Y, E, \theta) \propto P(X_t | X_{t-1}, Z, X_{t-1}, \hat{\theta}_k) P(Y_t | X_t, \xi) m^t(X_t)$. Similarly, this distribution has a closed form expression (Gaussian) in the case of a LG-SSIM model. Again, we derive equations for the mean $\mu_t^i$ and the covariance matrix $\Sigma_t^i$ that do not require inversion of dependence covariance matrices:

$$
\mu^i_t = G_t^{-1} (G_t^{-1} \mu^i_t), \quad \Sigma_t^i = G_t Q_{Z_t}, \quad \text{where } G_t^i = I + Q_{Z_t} C^T R^{-1} C + Q_{Z_t} (\Sigma_t^{-1})^{-1}
$$

$$
G_t^{-1} \mu^i_t = A_{Z_t} X_{t-1} + Q_{Z_t} C^T R^{-1} Y_t + Q_{Z_t} (\Sigma_t^{-1})^{-1} \mu^m_t.
$$

Note that missing observations require only a slight modification of the algorithm. Namely, at each step $t$ rows of matrix $C$ corresponding to missing observations at time $t+1$ should be set to zero.

6.2 Sampling dependence models (step 4)

By conditioning on $Z$ and assuming their prior independence, dependence models are also decoupled in the posterior, and each can be sampled as

$$
\tilde{E}_k, \tilde{\theta}_k \sim p(\tilde{E}_k, \tilde{\theta}_k | \{X_t, X_{t-1}\}_{t=Z_t=k}; \beta, \gamma).
$$

We adopt the same approach as [19] and [1], and assume a modular prior on each dependence model. It follows that dependence models of signals are also independent in the posterior:

$$
p(\tilde{E}_k, \tilde{\theta}_k | \{X_t, X_{t-1}\}_{t=Z_t=k}; \beta, \gamma)
$$

$$
= \prod_{i=1}^N p(\tilde{\rho}_a(i, k), \tilde{\theta}_k | \{X^i_t, X_{t-1}^i\}_{t=Z_t=k}; \beta, \gamma),
$$

and that, for each signal $i$, posterior over its dependence model is decomposed into a product of the posterior over parent set and the posterior over parameters given structure:

$$
p(\tilde{\rho}_a(i, k), \tilde{\theta}_k | \{X^i_t, X_{t-1}^i\}_{t=Z_t=k}; \beta, \gamma)
$$

$$
= p(\tilde{\rho}_a(i, k) | \{X_t^i, X_{t-1}^i\}_{t=Z_t=k}; \beta)
$$

$$
\times p(\tilde{\theta}_k | \{X^i_t, X_{t-1}^i\}_{t=Z_t=k}; \beta, \gamma).
$$

For a given parent set, the conditional posterior over parameters given the parent set and the marginal posterior of the parent set (both given on the right-hand side of Equation 15), can be computed efficiently when a conjugate prior is used.

7 Experiments

We present experimental results on two sets of data. The first is obtained in a controlled environment using joystick responses, while the second analyzes publicly available climate data. We use a latent-AR parametrization of the SSIM model, in which the latent state at time $t$ consists of a (possibly vector-valued) variable of interest $O_t$ at
Bayesian Switching Interaction Analysis Under Uncertainty

Figure 3: (top) Three assignments of tasks. Individual tasks can be: F – “follow”, M – “stay in the middle between”, and “move arbitrarily” (otherwise). (bottom) Order and duration of assignments.

time $t$ and the previous $r-1$ time points, i.e., $X_t = [O_t^k O_{t-1}^k \ldots O_{t-r+1}^k]^T$. In all experiments, we use the following prior on parent sets: $\beta_{i,pa(i)} = 1/(|pa(i)| + 1)^b$, where $|pa(i)|$ is the number of parents. When $b > 0$, the prior favors smaller parent sets. We typically set a weak prior on state transition probabilities that favors self-transitions. Finally, we set the parameters of the matrix-normal inverse-Wishart prior on the dependence and observation models similar to [2]. The exact settings of the parameters will be included in the posted code, at http://groups.csail.mit.edu/vision/sli/projects.php?name=structure_inference.

7.1 Joystick (human generated) data

Most available temporal data is not annotated for interactions. Furthermore, obtaining ground truth interactions is difficult and, in most cases, subjective. While that amplifies the importance of developing algorithms that aid in uncovering such interactions, it also makes the testing of these algorithms difficult. Consequently, we created a simple experiment, from so-called “joystick” data, where the structure is known (although the parameterization is not). In the experiment, five players control a joystick to move an object on the screen in order to accomplish a task. There are three different assignments of tasks shown in the top of Figure 3. Assignments switch over time over the duration of 4.5 minutes, as shown in the bottom of the figure. To remove bias, a player only sees the objects on which it depends. Positional (2D) data is recorded every 1/10 sec, so there is a total of 2701 time points, including the initial one. This data is realistic since it is human-generated and not synthesised from the model. In addition, it contains interaction annotations by design and is useful for validating the model.

We find that the best results are obtained when the data is downsampled 3 times (total of 901 time points) and AR order is 5, which we use in all experiments. This order corresponds to a lag of 1.5 seconds. A 3 times higher AR order would be required with the original data in order to capture the dependencies of the same length. However, the original data does not provide much additional information due to high correlation of samples at neighboring time points.

In all of the experiments, self-dependencies are assumed and are included in the count of parents. Results with $K = 3$, $b = 10$, and maximum number of parents set to 3 and 5, respectively, are shown in Figure 4. The top row presents the switching-state pairwise probability matrix, whose entry $(i, j)$ is the posterior probability that time points $i$ and $j$ are assigned the same switching state. There is an obvious switching pattern that coincides with the setup of the experiment. The bottom row shows the posterior probabilities of edges at 0.5, 1.25 and 2 min, which correspond to the three different assignments. The value in $i^{th}$ row and $j^{th}$ column is the probability of edge $i \rightarrow j$. Self-edges are “blacked out”, while the assignment (“correct”) edges are marked with a white dot. The algorithm assigns high posterior probability to all correct edges. In addition, a few spurious edges are
assigned medium to high probability. We note that these are typically edges between players that follow a common other player, possibly via intermediate players. For example, 2 and 3 both follow 5 in the first assignment, while 4 and 5 (via 2) both follow 3 in the second assignment. We also note that the results are better when fewer parents are allowed, since the number of possible incorrect choices of parents is reduced.

We set maximum number of parents to 3 in the in the rest of the experiments. Interestingly, when only two switching states are allowed, the switching pattern still indicates the presence of three states, as shown in Figure 5. Namely, states 1 and 2 are combined into a single state in some samples, while states 2 and 3 are combined in other samples. On the other hand, when \( K = 5 \) states are allowed, only 3 of them are actually used, yielding similar results as with \( K = 3 \).

Finally, we test our algorithm in the scenarios of higher uncertainty. In the first experiment, we add Gaussian noise of a fixed variance to all observations. Selection of variance \( 10^{-5} \) does not change the results. The results with variance \( 10^{-4} \) show higher uncertainty in some of the edges (Figure 6, left). Also, from the switching pattern we see that states 2 and 3 are not distinguished from each other in some of the samples. When noise variance is further increased to \( 10^{-3} \), none of the three states is recognized. In the second experiment, we treat a subset of the data as missing. When every 2\(^{nd} \) value is observed, the results do not change. The results when every 3\(^{rd} \) value is observed (Figure 6, right), show higher uncertainty of some edges.

\[ \text{noise variance } = 10^{-4} \]

7.2 Climate data

Here, we apply the LG-SSIM model to real-world climate data. In doing so, we wish to emphasize that one should be careful in drawing scientific conclusions from these results. In particular, the interactions amongst these data sets are likely not linear (as assumed by the LG-SSIM) and consequently, inferred structures may not necessarily be indicative of explicit causality. Nevertheless, the analysis may yield interesting details.

Following Jiang et al. [12], we use data on a subset of 16 climate indices from the repository maintained by the Earth System Research Laboratory of the National Oceanic and Atmospheric Administration (NOAA) [21], which are described in Table 1. These indices are compiled monthly and span various characteristics of the climate system. For the purpose of comparison, we use the data from 1951 to 2007, as in Jiang et al., and apply linear and quadratic detrending. Note that a small fraction of the data in this span is missing, which our model addresses naturally.

![Figure 6: Results on Joystick data when observation noise variance is 10\(^{-4} \) (left) and when every 3\(^{rd} \) value is observed (right). Top row are switching similarity matrices. Bottom row are edge posteriors at 0.5, 1.25 and 2 min.](image)

<table>
<thead>
<tr>
<th>#</th>
<th>abbrev.</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AMM</td>
<td>Atlantic Meridional Mode SST</td>
</tr>
<tr>
<td>2</td>
<td>AO</td>
<td>Arctic Oscillation</td>
</tr>
<tr>
<td>3</td>
<td>EP/NP</td>
<td>East Pacific/North Pacific Oscillation</td>
</tr>
<tr>
<td>4</td>
<td>GMT</td>
<td>Global Mean Lan/Ocean Temperature</td>
</tr>
<tr>
<td>5</td>
<td>Nino3</td>
<td>Eastern Tropical Pacific SST</td>
</tr>
<tr>
<td>6</td>
<td>Nino4</td>
<td>Central Tropical Pacific SST</td>
</tr>
<tr>
<td>7</td>
<td>Nino12</td>
<td>Extreme Eastern Tropical Pacific SST</td>
</tr>
<tr>
<td>8</td>
<td>Nino34</td>
<td>East Central Tropical Pacific SST</td>
</tr>
<tr>
<td>9</td>
<td>NOI</td>
<td>Northern Oscillation Index</td>
</tr>
<tr>
<td>10</td>
<td>ONI</td>
<td>Oceanic Nino Index</td>
</tr>
<tr>
<td>11</td>
<td>PDO</td>
<td>Pacific Decadal Oscillation</td>
</tr>
<tr>
<td>12</td>
<td>PNA</td>
<td>Pacific North American Index</td>
</tr>
<tr>
<td>13</td>
<td>SOI</td>
<td>Southern Oscillation Index</td>
</tr>
<tr>
<td>14</td>
<td>Solar</td>
<td>Solar Flux (10.7cm)</td>
</tr>
<tr>
<td>15</td>
<td>SWM</td>
<td>South Western USA Monsoon</td>
</tr>
<tr>
<td>16</td>
<td>WP</td>
<td>Western Pacific Index</td>
</tr>
</tbody>
</table>

Table 1: Description of climate indices.

We run inference using the SSIM latent-AR model with two switching states. We bound the number of parents per node to 3 and require a minimum of 1 parent with enforcing self-edges. The top row of Figure 7 shows the switching-state pairwise probability matrix. Unlike Jiang et al., whose results suggest a single switch point in 1978, this result suggests that there is a cyclic behavior. Figure 9 shows two matrices of posterior probabilities of edges that correspond to June 1963 (left) and August 1992 (right), which belong to the opposite phases of the cycle. We observe that Nino indices and ONI index are the most influential overall, confirming that they are important predictors of climate [22]. Interestingly, the only significant dependence of ONI index is on Southern Oscillation Index.
Bayesian Switching Interaction Analysis Under Uncertainty

8 Conclusion

We presented a state-space switching interaction model (SSIM), which represents interactions as directed edges of a dynamic Bayesian network, allows switching between interactions, and allows arbitrary observation processes and missing data. Furthermore, we employed Bayesian reasoning over structures to deal with uncertainty in the data and due to the large number of possible structures. Efficient inference is enabled by limiting the number of parents per signal, and is done via a Gibbs sampling procedure. This model is expressive and can uncover different aspects of interactions among time-series and their patterns, as we have demonstrated by experiments.

Acknowledgements

ZD was partially supported by the Office of Naval Research Multidisciplinary Research Initiative (MURI) program, award N000141110688. JF was partially supported by the Defense Advanced Research Projects Agency, award FA8650-11-1-7154.
References


