Event-Driven Continuous Time Bayesian Networks

Debarun Bhattacharjya[†], Karthikeyan Shanmugam[†], Tian Gao[†],
Nicholas Mattei[§], Kush R. Varshney[†], Dharmashankar Subramanian[†]

[†] Research AI, IBM T. J. Watson Research Center, Yorktown Heights, NY, USA
[§] Department of Computer Science, Tulane University, New Orleans, LA, USA
{debarunb, tgao, krvarshn, dharmash}@us.ibm.com
Karthikeyan.Shanmugam2@ibm.com, nsmattei@tulane.edu

Abstract

We introduce a novel event-driven continuous time Bayesian network (ECTBN) representation to model situations where a system's state variables could be influenced by occurrences of events of various types. In this way, the model parameters and graphical structure capture not only potential "causal" dynamics of system evolution but also the influence of event occurrences that may be interventions. We propose a greedy search procedure for structure learning based on the BIC score for a special class of ECTBNs, showing that it is asymptotically consistent and also effective for limited data. We demonstrate the power of the representation by applying it to model paths out of poverty for clients of CityLink Center, an integrated social service provider in Cincinnati, USA. Here the ECTBN formulation captures the effect of classes/counseling sessions on an individual's life outcome areas such as education, transportation, employment and financial education.

Introduction

Real-world decision situations often involve uncertainties that interact with each other in a way that requires modeling complex dynamic inter-dependencies. The uncertain variables of interest of a system can be modeled as state variables whose evolution is captured by some dynamic process. In many cases, state variables are only observed at irregular epochs in time, unlike time series models.

A continuous-time Bayesian network (CTBN) (Nodelman, Shelton, and Koller 2002) representation handles modeling joint trajectories of a system's state variables where irregular state variable transitions are modeled as homogeneous Markov processes. While CTBNs offer a simple way to model the dynamics of such transitions, they are only suitable when a dynamical system is observed in an isolated context. There are many scenarios that have the following added complexity: external occurrences of events could influence the manner in which the system evolves. In this paper, we consider modeling situations where occurrences of various types of *events* influence evolution of a set of *state variables*, or synonymously *system variables*. We provide a small set of potential applications where external events occurring irregularly over time could affect the evolution of a system:

- *Health* – a diabetic patient's blood glucose level and mental well-being are influenced by events such as insulin intake, meals and physical activity.
- *Finance* – stock prices for a set of companies in an industry are affected by natural events such as disasters or political events such as trade deals.
- *Social Impact* – social services such as counseling sessions and classes have an impact on a person's level of education, employment, and well-being.

Event datasets are sequences of labels on a time line where each label indicates the type of event. For example, labeled time stamps of medication, exercise, and meals would indicate events that could be relevant for a patient's health outcomes. To capture the influence of events on state variables, we introduce a new model – event-driven continuous time Bayesian networks (ECTBNs) – where, in addition to state variables driving transitions of other state variables, a time stamped event history could influence the time to transition as well as the probability of transition of state variables.

Including events in the scope of the model requires a fundamental extension to CTBNs and cannot be reduced to an expanded CTBN with proxy state variables for events. This is because the intensity function that determines time to next transition in a CTBN only depends on the current configuration of parent state variables; it does not depend on *when* the configuration of these state variables attained their current configuration. However, when event sequences influence the intensity functions of state transitions, their previous times of occurrence could matter, making the influence non-Markov because it does not only depend on the current state.

As an example, consider the case where the frequency of meals in the recent history affects transitions of a patient's blood sugar levels. This is illustrated in the sketch in Figure 1(a) where a blood sugar state variable with two states is influenced by exercise and meal events. Even if the events were modeled as state variables and the sequences of events were tracked with memory, the intensity function would be unable to capture the notion that only the number of meals within a certain time window matters.

Contributions. In this paper we make three major inter-related contributions: (1) we introduce a novel, interpretable yet analytically sophisticated graphical model that captures
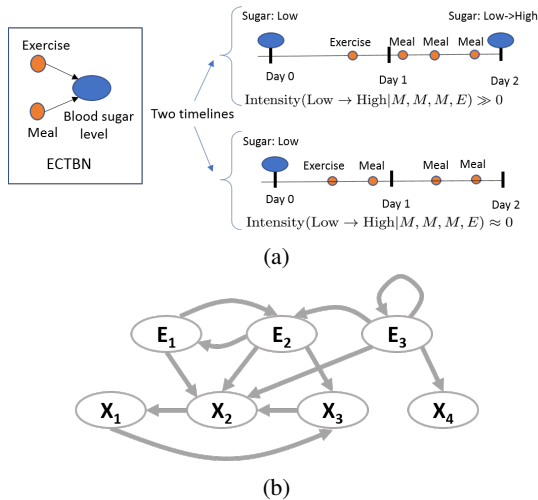
Figure 1: (a) A sketch involving a blood sugar state variable with two possible states (High and Low), influenced by meal and exercise events. Three meals within a day without exercise causes the blood sugar level to increase while two meals maintains the level; (b) Illustrative ECTBN graph representing a dynamic process involving 4 state variables (X_1 to X_4) and 3 event labels (E_1 to E_3).

joint dynamics involving both event occurrences, modeled as a multivariate point process, and state variables, modeled as Markov processes; (2) we propose an algorithm for learning the structure and parameters of this model from data involving traces of events and transitions of state variables. We prove its consistency in the asymptotic data case whilst also demonstrating its effectiveness for limited data through experiments with synthetic data; (3) We conduct a detailed study applying our model to a real-world dataset pertaining to social service. This is work in partnership with CityLink Center – a non-profit organization in Cincinnati, USA that works with approximately 15 on-site agencies to provide a suite of services to help adults in poverty meet their goals in areas such as education, employment and transportation.

Related Work

Continuous time processes model the dynamics of events occurring irregularly over time. Conditional intensity functions capture the instantaneous rate of occurrence of an event given the history of other events. There has been a lot of work on various parametric models for learning conditional intensity functions for event streams. Notable amongst them are Poisson networks (Rajaram, Graepel, and Herbrich 2005), Poisson cascades (Simma and Jordan 2010), piecewise-constant conditional intensity models (Gunawardana, Meek, and Xu 2011), forest-based point processes (Weiss and Page 2013), etc. Recent work has also considered neural network architectures (Du et al. 2016; Xiao et al. 2017).

When an event's conditional intensity depends only on the past history of a set of parent events, this can be represented using graphical event models (GEMs) (Didelez 2008;

Meek 2014; Gunawardana and Meek 2016). These are different from time series graphical models (Eichler 1999; Dahlhaus 2000) and dynamic Bayesian networks (Dean and Kanazawa 1989; Murphy 2002) as they represent continuous time processes.

CTBNs are closely related models that represent joint trajectories of discrete variables, as opposed to models of event streams in continuous time. In this work we introduce a model that can be viewed as a novel combination of joint trajectories in CTBNs and the effect of event arrivals. CTBNs have been deployed in diverse applications, including reliability analysis (Boudali and Dugan 2006), cardiogenic heart failure (Gatti, Luciani, and Stella 2012), cybersecurity (Xu and Shelton 2008) and gene network reconstruction (Acerbi et al. 2014). Extensions to CTBNs have been proposed, such as more general transition times like Erlang-Coxian distributions and explicit negative evidence (Gopalratnam, Kautz, and Weld 2005), as well as adding standard Bayesian network static chance nodes (Portinale and Codetta-Raiteri 2009).

ECTBNs: Model Description and Learning

We introduce the ECTBN model, capturing processes involving state variables and event arrivals that combine elements of CTBNs and GEMs in a non-trivial way. We start with a more general formulation and then specify the case that we use for learning and experimental investigation.

Notation

Consider a set of discrete state variables $\mathcal{X} = \{X_i\}_{i=1}^I$. Let $\mathrm{Val}(X_i)$ be the domain of variable X_i. The states of these variables are assumed to be known at all times between initial time $t_0 = 0$ to the end time T. Data about each variable is of the form of state transitions, $D_{X_i} = (t_k, x_k)_{k=0}^{N_i}$ where the state at time t_0 is the initial state and $x_{k+1} \neq x_k \ \forall k, x_k \in \mathrm{Val}(X_i)$. Data for all state variables taken together is denoted $D_{\mathcal{X}} = \bigcup_{X \in \mathcal{X}} D_X$.

We assume there is also data about events occurring over time, $D_{\mathcal{E}} = (t_k, e_k)_{k=1}^{N_E}$, where t_k are time stamps and e_k belong to an event label set $\mathcal{E} = \{E_j\}_{j=1}^J$. All the data taken together is $D = D_{\mathcal{X}} \cup D_{\mathcal{E}}$. We use $h(\cdot)$ to denote historical occurrences of events. $h_{\mathcal{B}}(t) = \{(t_k, e_k) \in D_{\mathcal{B}} : t_k < t\}$ represents the history of events in the set $\mathcal{B} \subset \mathcal{E}$ until time t.

Formulation

Definition 1. *An ECTBN $\mathcal{N}$ includes:*
- *A directed graph $\mathcal{G}$ where $\mathbf{U}_E \subset \mathcal{E}$ are the parents of an event label E and $\mathbf{U}_X \subset \{\mathcal{X} \setminus X\} \bigcup \mathcal{E}$ are the parents of a state variable $X \in \mathcal{X}$. We decompose the latter into two sets: parents that are state variables $\mathbf{U}_{X(\mathcal{X})} \subseteq \mathcal{X} \setminus X$ and parents that are event labels $\mathbf{U}_{X(\mathcal{E})} \subseteq \mathcal{E}$.*
- *An initial distribution $\mathbf{P}_{\mathcal{X}}^0$ over state variables.*
- *Conditional intensity matrices for every $X \in \mathcal{X}$, $Q_{X|\mathbf{u}_{X(\mathcal{X})}, h_{\mathbf{U}_{X(\mathcal{E})}}}(t)$, which model state transitions. This depends on the current state $\mathbf{u}_{X(\mathcal{X})}$ of the parents $\mathbf{U}_{X(\mathcal{X})}$ at time t and history of labels in $\mathbf{U}_{X(\mathcal{E})}$ till time t, denoted $h_{\mathbf{U}_{X(\mathcal{E})}}(t)$. A matrix $Q(\cdot)$ is equivalent to considering waiting times $q_{x|\mathbf{u}_{X(\mathcal{X})}, h_{\mathbf{U}_{X(\mathcal{E})}}}(t)$ in state $X = x$ before*

transitioning to some other state $x' \neq x$, as well as the probabilities of transitioning from state x to state x' at time t, $\theta_{xx'|\mathbf{u}_{X(\mathcal{X})}, h_{\mathbf{U}_{X(\mathcal{E})}}}(t)$.

- Conditional intensity rates for every event label $E \in \mathcal{E}$, $\lambda_{E|h_{\mathbf{U}_E}}(t)$, which model event arrivals. The history of event labels in parent set $\mathbf{U}_E$ at time t is denoted $h_{\mathbf{U}_E}(t)$.

Figure 1(b) shows an illustrative ECTBN graph. Note that there may be cycles and even self-loops for an event label because its occurrence rate could depend on its own history. State variables could have event labels as parents but not vice versa. Our motivation here is to study situations where events could probabilistically influence the uncertainties in a system but not the other way around.

It should be evident from the complex inhomogeneous history-dependence in Definition 1 that it is impractical to consider all possible histories for modeling the influence of events on state variables; one cannot learn arbitrary dependencies with finite data as it would be difficult to generalize for learning. We simplify the historical dependence by making an assumption that results in an important special case, motivated by Bhattacharjya, Subramanian, and Gao (2018):

Assumption 2. *Consider a set $\mathcal{W}$ of time windows for every edge from event label E directed into state variable X in graph $\mathcal{G}$, each denoted $w_{E,X}$. Assume that the rates and probabilities associated with state variable transitions depend only on whether a parent event label $E \in \mathbf{U}_{X(\mathcal{E})}$ occurred at least once in some recent time window $w_{E,X}$.*

The above assumption is the *recency* or *proximal* assumption, where recent events matter more than older ones; this simplifies parent conditions to be binary for each parent. Specifically, if $\mathbf{u}_{X(\mathcal{E})}$ denotes a vector of indicators, one each for whether an event label in $\mathbf{U}_{X(\mathcal{E})}$ occurs or not, then Assumption 2 simplifies the dependence of $q_{(\cdot)}$ and $\theta_{(\cdot)}$ as:

$$\theta_{xx'|\mathbf{u}_{X(\mathcal{X})}, h_{\mathbf{U}_{X(\mathcal{E})}}}(t) = \theta_{xx'|\mathbf{u}_{X(\mathcal{X})}, \mathbf{u}_{X(\mathcal{E})}}$$

$$q_{x|\mathbf{u}_{X(\mathcal{X})}, h_{\mathbf{U}_{X(\mathcal{E})}}}(t) = q_{x|\mathbf{u}_{X(\mathcal{X})}, \mathbf{u}_{X(\mathcal{E})}} \quad (1)$$

The number of parameters can now be ascertained for any state variable. As an example, for the ECTBN in Figure 1(b), if state variable X_3 has 3 values in its domain $\mathrm{Val}(X_3)$, then X_2 has $2^3 * 3 = 24$ parental conditions $(\mathbf{u}_{X(\mathcal{X})}, \mathbf{u}_{X(\mathcal{E})})$ since it has 3 event labels as parents, $\mathbf{U}_{X_2(\mathcal{E})} = \{E_1, E_2, E_3\}$, along with 1 state variable parent, $\mathbf{U}_{X_2(\mathcal{X})} = \{X_3\}$.

We note that our work extends easily to the case where state variable parameters are a piece-wise constant function of the history of events. Indeed, we show in a subsequent section that our theoretical results apply in the case of this generalization, where we can choose a general class of functions to model dependence on event histories instead of a function involving only the most recent time window (Gunawardana, Meek, and Xu 2011). The piece-wise constant model is general enough to approximate arbitrary histories (Gunawardana and Meek 2016). We only consider recent windows for much of this paper to avoid the notation from getting unwieldy. We also highlight that the recency assumption is often suitable in practice due to the nature of real-world causal influences, and that the simplification avoids overfitting (Bhattacharjya,

Subramanian, and Gao 2018), which is why we make the assumption for our experiments.

Learning ECTBNs

We pose the learning problem as: given data D about state transitions and event occurrences and a complete set of hyperparameters of windows for every edge from E to X, $\mathcal{W}^c$, find the ECTBN graph $\mathcal{G}$ and model parameters. We focus on learning state variable parameters here in this work as their dependence on events is the main novelty. In the remainder of this section, we begin with likelihood decomposition. We show how the parameter estimation problem is similar to a regular CTBN, given a graph, and then present a search procedure to learn the optimal graph.

Data Likelihood and Parameter Learning. Let $\boldsymbol{Q} = \{\mathbf{q}, \boldsymbol{\Theta}\}$ represent the collection of $q_{(\cdot)}$ and $\theta_{(\cdot)}$ parameters that model the state variable transitions. Similarly, let $\boldsymbol{\Lambda}$ represent the the collection of $\lambda_{(\cdot)}$ parameters for the arrival process of events. Then, the likelihood of observed data factorizes according to the graph $\mathcal{G}$, given by:

$$L(D|\boldsymbol{Q}, \boldsymbol{\Lambda}) = \left[\prod_{X \in \mathcal{X}} L(D_X|\boldsymbol{Q}, D_{\mathbf{U}_{X(\mathcal{X})}}, D_{\mathbf{U}_{X(\mathcal{E})}}) \right]$$

$$\left[\prod_{E \in \mathcal{E}} L(D_E|\boldsymbol{\Lambda}, D_{\mathbf{U}_E}) \right] \quad (2)$$

The data likelihood for a state variable X is a function of the parameters for waiting times and probabilities of transitions. In the general case, these depend on the history of events. For brevity in the following equation, we use $\mathbf{h}(t)$ as shorthand for the joint historical condition $\mathbf{u}_{X(\mathcal{X})}, h_{\mathbf{U}_{X(\mathcal{E})}}(t)$. The likelihood can be factored as:

$$L(D_X|\mathbf{Q}, D_{\mathbf{U}_{X(\mathcal{X})}}, D_{\mathbf{U}_{X(\mathcal{E})}}) = \prod_{(t_k, x_k) \in D_X} \theta_{x_k x_{k+1}|\mathbf{h}(t_{k+1})}$$

$$\prod_{(t_k, x_k) \in D} q_{x_k|\mathbf{h}(t_{k+1})} * e^{\left(-\int_{t_k}^{t_{k+1}} q_{x_k|\mathbf{h}(t)} dt \right)}$$

$$(3)$$

The data likelihood for arrivals of an event label E depends on the event arrival rates:

$$L(D_E|\boldsymbol{\Lambda}, D_{\mathbf{U}_E}) = \prod_{(t_k, E) \in D_{\mathcal{E}}} \left(\lambda_{E|h_{\mathbf{U}_E}}(t_k) * \right.$$

$$\left. \exp \left(-\int_{t_k}^{t_{k+1}} \lambda_{E|h_{\mathbf{U}_E}}(t) dt \right) \right) \quad (4)$$

The above expression is quite general and covers most reasonable multivariate point processes (Didelez 2008). Here we focus solely on learning state variable parameters $\boldsymbol{Q}$ given a graph $\mathcal{G}$, omitting details about learning event arrival process parameters $\boldsymbol{\Lambda}$, but we stress that any of a number of models could be deployed for this purpose; we refer the reader to the related work section, which describes several such temporal event models.

Let $\mathbf{u}$ be a vector that takes values in $\mathrm{Val}(\mathbf{u}_{X(\mathcal{X})}) \times \mathrm{Val}(\mathbf{u}_{X(\mathcal{E})})$ for any $X \in \mathcal{X}$. According to the X variable, $\mathbf{u}$

vector taking appropriate values is implicit to avoid clutter in notation. Then, we have:

$$L(D_{\mathcal{X}}|\boldsymbol{Q}, D_{\mathcal{E}}) = \left[\prod_{X \in \mathcal{X}} L(D_X | \mathbf{q}, \boldsymbol{\Theta}, D_{\mathbf{U}_{X(\mathcal{X})}}, D_{\mathbf{U}_{X(\mathcal{E})}}) \right]$$

$$\stackrel{a}{=} \left[\prod_{x \in \text{Val}(X), X \in \mathcal{X}} \prod_{\mathbf{u}} q_{x|\mathbf{u}}^{M[x|\mathbf{u}]} e^{-T[x|\mathbf{u}]q_{x|\mathbf{u}}} \right]$$

$$\times \left[\prod_{x \in \text{Val}(X), X \in \mathcal{X}} \prod_{x' \neq x} \prod_{\mathbf{u}} \theta_{xx'|\mathbf{u}}^{M[x,x'|\mathbf{u}]} \right] \quad (5)$$

The summary statistics for X are defined as follows:

- $M[x, x'|\mathbf{u}]$: the number of times the variable transitions from state x to state x' and the condition $\mathbf{u}$ is true at those times, i.e. when $\mathbf{u}_{X(\mathcal{X})}$ and $\mathbf{u}_{X(\mathcal{E})}$ take values in $\mathbf{u}$.
- $M[x|\mathbf{u}]$: the number of times the variable transitions from state x and the condition $\mathbf{u}$ is true at those times, i.e. when $\mathbf{u}_{X(\mathcal{X})}$ and $\mathbf{u}_{X(\mathcal{E})}$ take values in $\mathbf{u}$.
- $T[x|\mathbf{u}]$: the total amount of time where the variable is in state x and the condition $\mathbf{u}$ is true at those times, i.e. when $\mathbf{u}_{X(\mathcal{X})}$ and $\mathbf{u}_{X(\mathcal{E})}$ take values in $\mathbf{u}$.

The maximum likelihood estimates for parameters $\mathbf{q}$ and $\boldsymbol{\Theta}$ given the structure $\mathcal{G}$ are:

$$\hat{q}_{x|\mathbf{u}} = \frac{M[x|\mathbf{u}]}{T[x|\mathbf{u}]}; \hat{\theta}_{xx'|\mathbf{u}} = \frac{M[x, x'|\mathbf{u}]}{M[x|\mathbf{u}]} \quad (6)$$

This is due to the simplification in (1) from Assumption 2 and the likelihood expression in (3).

Structure Learning. One of the main goals of modeling with an ECTBN is to discover the (in)dependencies and causal effects of events on state variables. Discovering the underlying true graph would reveal information about events that change the current state to some other state.

The structure learning problem is as follows: to find the optimal graph $\mathcal{G}^* = \max_{\mathcal{G}} s(\mathcal{G}, \mathcal{D})$, where $s(\mathcal{G}, \mathcal{D})$ is a scoring function that measures the goodness of fit between any graph $\mathcal{G}$ and data D. We use the BIC score, adapted to ECTBNs, defined for state variable X as: $BIC(X) = \log L(D_{\mathcal{X}}) - \left[\frac{\log |D|}{2} Dim(\boldsymbol{Q}(X)) \right]$, where $|D|$ is the size of the data. $Dim(\boldsymbol{Q}(X))$ is the dimensionality of the parameters for X, which in our case is the number of independent parameters in $\mathbf{q}$ and $\boldsymbol{\Theta}$ that are associated with X: $Dim(\boldsymbol{Q}(X)) = |\text{Val}(X)|^2 * 2^{|\mathbf{U}_{X(\mathcal{E})}|} * \prod_{Z \in \mathbf{U}_{X(\mathcal{X})}} |\text{Val}(Z)|$. In the next section, we show that the BIC score is asymptotically consistent with fixed known windows $\mathcal{W}$.

Note that the learning procedure for ECTBNs enjoys similar benefits to CTBNs. Since there are no constraints around acyclicity, structure learning can be decomposed into learning individual optimal sub-graphs and then combining them to form the global optimal graph. We use a sub-graph learning approach that finds the optimal parent set of each state variable X with a hill climbing search. At each iteration, we choose the highest scoring graph among the set of graphs consisting of the current graph and all graphs that are one operation away from the current graphs. The operations under

consideration include adding an edge and deleting an edge. The search for the parents for each node continues until there is no improvement in scores.

Structure Identifiability

In this section, we study the stochastic processes that can be modeled by an ECTBN and demonstrate that under some assumptions and with enough data, our proposed structure learning approach with the BIC score can recover the underlying graph from which the data is generated.

The key enhancement of an ECTBN over a CTBN is that the former is able to incorporate historical dependencies of event arrivals. We introduced a special case with the recency assumption (Assumption 2), i.e. rates and state transitions depend on $\mathbf{u}_{X(\mathcal{E})}$ that denotes whether the individual events $E \in \mathbf{U}_{E(X)}$ occurred in time window $w_{E,X}$ or not. This vector can not be viewed as a typical Markovian state variable since the intensity function for transitions would include the memory of *when* the last event happened. To resolve this issue, we prove certain lemmas that guarantee BIC score consistency following the broad outline in Nodelman, Shelton, and Koller (2003). For the theoretical structure identifiability results, we assume a more general dependence on event histories as outlined below.

Piecewise Constant Model for Event Histories: We define a mapping function $\sigma_{E,X}$ one for every $(E, X) \in \mathcal{G}$ which takes the event history $\mathbf{h}_E(t)$ and produces a categorical value from the discrete alphabet $\Sigma_{E,X}$. We assume that the rates and probabilities of state transition depend on the categorical output of these functions $\sigma_{E,X}$, i.e. $\mathbf{u}_{X(\mathcal{E})}$ is a vector of function values $\sigma_{E,X}(h_E(t))$ for all $E \in \mathbf{U}_{X(\mathcal{E})}$. A special case of this is our recency assumption where the function $\sigma_{E,X}$ is an indicator for whether event E occurs within time period $[t - w_{E,X}, t)$. The assumption of a general sigma function and discrete output states is one of the most popular models for modeling event histories in event datasets (without state variables) – this is the piecewise constant intensity model (PCIM) (Gunawardana, Meek, and Xu 2011).

To model this during search procedure formally, we introduce new state variables that are based on event occurrences, given a complete set of functions $\{\sigma_{E,X}(\cdot)\}$ for every possible E, X pair. Let $S_{E,X} \in \Sigma_{E,X}$ be categorical output produced by $\sigma_{E,X}$ acting on the history of the event E, i.e. $h_E(t)$. Let $s_{E,X}$ be the realization of $S_{E,X}$ at time t. We use $\mathbf{S} = \{S_{E,X} : \forall E \in \mathcal{E}, \mathcal{X} \in \mathcal{X}\}$ to denote the set of all these categorical valued variables and $\mathbf{S}_X = \{S_{E,X} : \forall E \in \mathbf{U}_{X(\mathcal{E})}\}$ for a particular X in a graph $\mathcal{G}$. Let the vector $\mathbf{s}_X$ denote the realization for $\mathbf{S}_X$ and let $\mathbf{s}$ denote the realization of all categorical valued variables $\mathbf{S}$.

It is important to note that the new state variables introduced have transitions that are not memoryless, because transitions are driven by occurrences (and non-occurrences) of historical events; they cannot therefore be nodes in a regular CTBN with an exponential clock.

Now, we provide definitions to incorporate the new state variables and also help specify parameters of an internally consistent stochastic process involving state variable transitions that are also potentially influenced by event arrivals:

Definition 3. *The modified graph $\tilde{\mathcal{G}}$ from $\mathcal{G}$ consists of nodes $\mathbf{S} \cup \mathcal{X}$ where the parents of $X \in \mathcal{X}$ are $\mathbf{U}_{X(\mathcal{X})} \cup \mathbf{S}_X$ and nodes $S_{E,X} \in \mathbf{S}$ have no parents.*

Definition 4. $\mathbf{s}$ *is an admissible categorical vector if for given E and for all X, $s_{E,X} = \sigma_{E,X}(h_E(t))$, for some common history of event E, i.e. $h_E(t)$. In other words, the set of values $s_{E,X}$ $\forall X$ is produced by different functions $\sigma_{E,X}$ acting on a common history $h_E(t)$. $\mathcal{A}$ is the subset of admissible categorical vectors $\mathbf{s}$ from the set $\prod_{E,X} \Sigma_{E,X}$ (Cartesian product of the discrete alphabet $\Sigma_{E,X}$).*

Let $\tilde{q}_{xx'|\mathbf{u}_{X(\mathcal{X})}, s_X}$ be the effective intensity function of transitioning from state $X = x$ to $X = x'$ combining the parameters $q_{x|\mathbf{u}_{X(\mathcal{X})}, \mathbf{u}_{X(\mathcal{E})}}$ and $\theta_{xx'|\mathbf{u}_{X(\mathcal{X})}, \mathbf{u}_{X(\mathcal{E})}}$.

Definition 5. *A graph $\tilde{\mathcal{G}}$ with conditional intensities $\tilde{q}_{xx'|\mathbf{u}_{X(\mathcal{X})}, s_x}$ for subsets $\mathbf{U}_{X(\mathcal{X})}$ and $\mathbf{S}_X$ induces extended conditional intensities wherein the vector $\mathbf{s}$ is consistent with the vector $\mathbf{s}_X$, i.e. $\tilde{q}_{xx'|\mathbf{u}_{X(\mathcal{X})}, \mathbf{s}} = \tilde{q}_{xx'|\mathbf{u}_{X(\mathcal{X})}, \mathbf{s}_X}$.*

Let the collection of these extended conditional intensities be in the form of matrices $\tilde{\mathbf{Q}}_{X|\mathbf{U}_{X(\mathcal{X})}, \mathbf{s}}$. The various extended conditional intensities of different variables X and their parent state variables $\mathbf{U}_{X(\mathcal{X})}$ can be multiplied though the "amalgamation" operation (Nodelman, Shelton, and Koller 2002) into a conditional intensity for the entire graph $\tilde{\mathcal{G}}$ given by: $\tilde{\mathbf{Q}}_{\tilde{\mathcal{G}}|\mathbf{s}} = \prod_X \tilde{\mathbf{Q}}_{X|\mathbf{U}_{X(\mathcal{X})}, \mathbf{s}}$. This is similar to the amalgamation done in a CTBN except that it is done for a specific $\mathbf{s}$.

The set of possible admissible vectors $\mathcal{A}$ allows us to characterize potential stochastic processes that a proximal ECTBN could model. Consider a process over $\prod_X \text{Val}(X)$ defined by the set of conditional intensity matrices $\tilde{\mathbf{Q}}_{\mathbf{X}|\mathbf{s}}$ $\forall \mathbf{s} \in \mathcal{A}$. Each matrix has all the corresponding conditional intensities $\tilde{q}_{\mathbf{xx'}|\mathbf{s}}$ over all state vectors $\mathbf{x}, \mathbf{x'} \in \prod_X \text{Val}(X)$ for a specific $\mathbf{s}$.

Definition 6. $\tilde{\mathcal{G}}$ *is an S-map for the stochastic process for conditional intensity matrices $\tilde{\mathbf{Q}}_{\mathbf{X}|\mathbf{s}}$ if and only if $\tilde{\mathbf{Q}}_{\mathbf{X}|\mathbf{s}} = \tilde{\mathbf{Q}}_{\tilde{\mathcal{G}}|\mathbf{s}}$, $\forall \mathbf{s} \in \mathcal{A}$ for some set of conditional intensity matrices $\tilde{\mathbf{Q}}_{\tilde{\mathcal{G}}|\mathbf{s}}$ that respect $\tilde{\mathcal{G}}$.*

Definition 7. *A stochastic process represented by conditional intensity matrices $\tilde{\mathbf{Q}}_{\mathbf{X}|\mathbf{s}}$ is called variable based if $\tilde{q}_{\mathbf{xx'}|\mathbf{s}} = 0$ whenever $\mathbf{x}$ differs from $\mathbf{x'}$ in more than one co-ordinate for all $\mathbf{s} \in \mathcal{A}$.*

Consider the following graph $\mathcal{G}^c$ where there is a directed edge between any two state variables X and X' and there is a directed edge from $S_{E,X'}$ to X for all E, X, X'. It is almost fully connected except amongst $S_{E,X}$ state variables.

The following lemma is a variation of one in Nodelman, Shelton, and Koller (2003) as adapted to our case.

Lemma 8. $\mathcal{G}^c$ *is an S-map for any stochastic process with intensity matrices $\tilde{Q}_{\mathbf{X}|\mathbf{s}}$, $\mathbf{s} \in \mathcal{A}$.*

This is simply because all possible connections are found in the graph $\mathcal{G}^c$. This can easily be seen by generalizing Theorem 5.3 in Nodelman, Shelton, and Koller (2003) for every $\mathbf{s} \in \mathcal{A}$.

Now we state our main theorem which is a variation of Theorem 5.5 in Nodelman, Shelton, and Koller (2003). Proofs for the major theorems are in Appendix A[1].

Theorem 9. *The modified graph $\tilde{\mathcal{G}}$ is an S-map for the stochastic process with conditional intensities $\tilde{\mathbf{Q}}_{\mathbf{X}|\mathbf{s}}$, $\mathbf{s} \in \mathcal{A}$, if and only if $\tilde{\mathbf{Q}}_{\mathbf{X}|\mathbf{s}}$ satisfies the following two conditions:*
1. $\forall X \in \mathcal{X}$, for any two full state vectors $\mathbf{x}$ and $\mathbf{x'}$ that agree on X and $\mathbf{U}_{X(\mathcal{X})}$, then $\tilde{q}_{\mathbf{xx}|\mathbf{s}} = \tilde{q}_{\mathbf{x'x'}|\mathbf{s}}$, $\forall \mathbf{s} \in \mathcal{A}$
2. $\forall X \in \mathcal{X}$, for any two admissible $\mathbf{s}$ and $\mathbf{s'}$ that agree on $\mathbf{S}_X$ and for any two $\mathbf{x}$ and $\mathbf{x'}$ that differ in only the state variable X, we have : $\tilde{q}_{\mathbf{xx'}|\mathbf{s}} = \tilde{q}_{\mathbf{xx'}|\mathbf{s'}}$

Remark. The above theorem is different from Theorem 5.5 in Nodelman, Shelton, and Koller (2003) because of the second condition which is needed due to conditioning by the states $\mathbf{s}$. This is the crucial piece of difference in the proofs.

By the same arguments as in Nodelman, Shelton, and Koller (2003), due to Theorem 9, one can only add vacuous edges and still maintain S-map relationships.

Theorem 10. *If $\tilde{\mathcal{G}}$ is an S-map for a variable based stochastic process with intensities $\tilde{\mathbf{Q}}_{\mathbf{X}|\mathbf{s}}$, $\forall \mathbf{s} \in \mathcal{A}$, then there is a minimal unique minimal S-map $\hat{\mathcal{G}} \subset \tilde{\mathcal{G}}$.*

Once the existence of a minimal S-map is established and we have a parameter learning procedure that is asymptotically accurate given the graph, the following result holds:

Theorem 11. *Consider state transition data generated by an ECTBN stochastic process with the set of functions $\sigma_{E,X}$ for every edge (E, X) in $\mathcal{G}$. For every X, all admissible $\mathbf{u}$ and state value $X = x$ associated with the graph where $q_{x|\mathbf{u}} > 0$, suppose that $P(\lim_{N \to \infty} M[x|\mathbf{u}]/N > 0) \to 1$, where N is the number of state transitions observed. Then, the hill climbing structure learning approach with the BIC score outlined earlier, given the set of functions that include the generating set of functions $\{\sigma_{E,X}\}$, is consistent with the minimal S-map $\hat{\mathcal{G}}$ of the data generating ECTBN.*

Remark: The theoretical results of this section rely primarily on showing how an ECTBN resembles an ensemble of CTBNs. The results are valid for a fairly general assumption on modeling event histories' influence on state transitions using $\sigma.(\cdot)$ that map event histories into discrete states. Previously we presented a learning algorithm that uses windows $w_{E,X}$ as hyper parameters and finds the graph with the best BIC score. We note that for the more general case, we could follow the same learning paradigm and choose $\sigma_{E,X}$ from a pre-defined set of basis functions as hyper-parameters, like in Gunawardana, Meek, and Xu (2011).

Synthetic Data Experiments

We test the proposed learning approach using synthetic data where the ground truth ECTBN graph and parameters are known. We generate 3 models, each with 5 state variables and 5 event label variables. The models differ in the structural relations among the state variables: they include a *chain*, a *star* (naive Bayes like structure), and a *cycle*.

---

[1]All appendices can be found in the arXiv version of the paper.

Graph Structure. The *chain* model has a chain graph structure among state variables: $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow X_5$. Each state variable has 3 random event label parents. The *star* model has a naive Bayes graphical structure among variables: $X_1 \rightarrow X_2, X_1 \rightarrow X_3, X_1 \rightarrow X_4$, and $X_1 \rightarrow X_5$. Again, each state variable has 3 random event label parents. Lastly, the *cycle* model forms a circle with its state variables: $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow X_5 \rightarrow X_1$. In this model, each state variable has 2 random event label parents. In all three models, each of 5 event labels can have 2 to 4 other event labels as parents, but with no state variables as parents as per the ECTBN assumptions.

Model Parameters and Data Generation. In all three models, each state variable has three states. State variable parameters $q_{(\cdot)}$ and $\theta_{(\cdot)}$ are generated randomly from a uniform distribution between 0.1 to $\frac{1}{3}$ and a Dirichlet distribution with hyperparameter $\alpha = (1, 1)$ respectively. Event traces are generated from a proximal graphical event model (PGEM) (Bhattacharjya, Subramanian, and Gao 2018) with windows ranging from 10 to 30 and rate of 0.5. Other parameters follow default values. Windows from event parents to state variables are set to 15. For each model, we generate 10 datasets over time period $T = 10K$ that include PGEM generated event traces as well as state variable transitions which are unique to an ECTBN. Details about the synthetic data generator are provided in Appendix B.

Results. Table 1 shows graph structure recovery results of the ECTBN learner for all variables' parents (both state variables and event labels) in these 3 synthetic models. We use the average precision and recall of each variable's parents as the performance measure for the learned graph structure against the ground truth. We observe that the precision is excellent for all models but the recall varies and is model dependent. There is perfect recall for the *cycle* model. Structure recovery is in general a challenging task and while this is also the case for ECTBNs, we see that the learner has very low false positive rates along with reasonable false negative rates with sufficient data.

Application to Tracking Life Outcomes

We apply the ECTBN model to study the effect of a set of services (events) on an individual's life outcome areas (state variables) in an integrated social services initiative. The data used for this section comes from our partnership with the CityLink Center in Cincinnati, Ohio, USA – a city-wide initiative launched in 2013 by a group of social service agencies and churches who recognized the need for a systemic approach to poverty. In contrast to the typical siloed social service delivery model, CityLink recognizes that different realms of clients' lives are interrelated, so their case management team works with clients in a holistic manner which leads to integrated longitudinal data. While there has been some related work on helping homeless individuals using planning, e.g. with MDPs (Gehlbach et al. 2006; Yi, Finkel, and Goldsmith 2008; Dekhtyar et al. 2009) and POMDPs (Yadav et al. 2016a;

Model	Precision	Recall
Chain	97%	47.4%
Star	84.6%	57.9%
Cycle	100%	100%

Table 1: ECTBN structure recovery for synthetic data generated from the three simple models.

Outcome Area	ECTBN	CTBN	CTBN-EV
Anxiety	**-5530**	-7282	-5639
Depression	**-5079**	-6715	-5122
Education	**-2100**	-2776	-2105
Employment	**-7486**	-7877	-7877
Financial Education	**-1400**	-1795	-1564
Transportation	**-1481**	**-1481**	**-1481**

Table 2: Log likelihood for the models on the CityLink data.

2016b), as well as other studies of intervention allocation (Kube, Das, and Fowler 2018), none of these efforts use the dynamic events setting considered here.

About the Data. For our analysis we use the approximately 1400 clients who have had more than 15 total interactions with CityLink out of a total of over 2900 total clients. We consider 6 outcome areas that are tracked through CityLink's data: education, employment, financial education, transportation, anxiety, and depression. These are dimensions of an individual's progress in attaining a self-sustainable way out of poverty; details about the levels and their descriptions are in Appendix C. Each of these six outcome areas has between three and six levels. We consider 11 types of services provided by CityLink and its partners, which are treated as events: 6 of them are group classes/sessions and 5 are one-on-one. The services include group industrial training, group classes on education, employment, financial education, transportation and wellness, as well as one-on-one sessions on employment, wellness, and financial education.

Structure Learning. We adopt the following learning procedure on the data, conducted separately for each state variable (outcome area) X. First, we configure a hyper-parameter setting for windows in $\mathcal{W}^c$ associated with incoming edges into X by uniformly randomly choosing a window from the list $\{15, 30, 60, 90, 180\}$ days for each event label. We repeat this procedure 100 times to build various window hyper-parameter configurations. Using 5-fold cross validation, we determine the optimal hyper-parameter setting by maximizing the average BIC score across folds. Finally, this optimal hyper-parameter setting is used to learn the optimal graph and parameters for X using all the training data. Although this may lead to some bias, it is necessary due to the paucity of data.

Figure 2 presents the learned graphical structure and windows for the CityLink data. This graph was learned using at most three levels for outcomes areas and with a slightly reduced weight for the penalty term in the BIC score, again due to limited data. There are several interesting results that
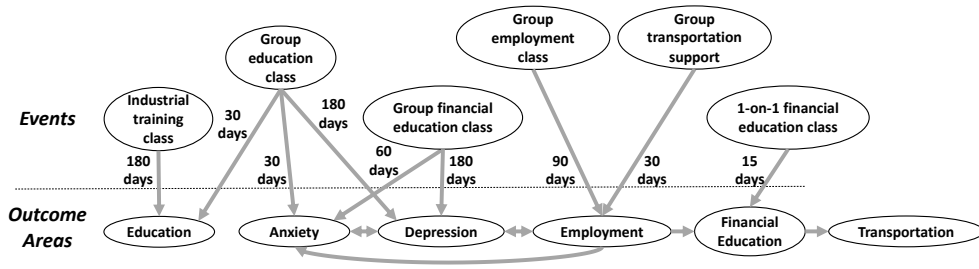
Figure 2: Learned ECTBN graph from CityLink data.

Outcome Area	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Education	group edu class	group edu class	group edu class; indus. training	indus. training	indus. training	N/A
Employment	group emp class; group transp. class	group emp class; group transp. class	group emp class; group transp. class	-	-	N/A
Financial Education	1-on-1 fin-ed; group fin-ed	1-on-1 fin-ed; group fin-ed	1-on-1 fin-ed	1-on-1 fin-ed	N/A	N/A

Table 3: Event label parents in ECTBN state variable transition analysis for three outcome areas. New state variables are created with states: current level, next higher level and other level, showcasing important events for specific level increments.

can be gleaned from the graph, potentially affecting the way CityLink operates. Firstly, group education classes have a direct and lasting effect on the Anxiety and Depression outcome areas, as do group financial education classes. Industrial training classes have a longer duration of effect (180 days) on the Education outcome area than the other group education classes (30 days). One-on-one financial education classes have more impact on the Financial Education outcome area than group financial education classes. Employment has a direct effect on Anxiety, Depression, and Financial Education. It is interesting to see that Anxiety, Depression, and Employment are critical, reinforcing the importance of a holistic approach to case management.

State Variable Transition Analysis. We also conducted a study to better identify influential events that affect transitions from a particular outcome area level to the next level, which was of interest to CityLink. We did this by creating additional state variables to track when the level of an outcome area increased; this new state variable has three states – the current level (not the maximum level), the next higher level and some other level of the outcome area under consideration. An ECTBN is learned for each new state variable while considering other outcome areas and events.

Table 3 summarizes the ECTBN event parents for three outcome areas determined from this transition analysis, enabling us to foreground local effects that were not evident previously. Selecting a few of these additional insights: (1) core education classes are important for transitions at lower levels of education whereas industrial training is important for transitions at higher levels; (2) the impact of group employment classes is particularly felt on low to mid levels of employment transitions; and (3) group financial education classes affect lower level transitions whereas the one-on-one classes are influential throughout the progression. For this

analysis, all windows were set to 180 days during learning.

ECTBN vs. CTBN(s). We also checked how ECTBN fit the data as compared to the following two baselines: (1) CTBN: This is a regular CTBN that only considers state variables and does not see the event occurrences. (2) CTBN-EV (event variables): This is a CTBN where new state variables are introduced, one each for the $|\mathcal{E}|$ event labels. These are binary state variables which are active when the corresponding event is the most recent one observed and inactive otherwise. Table 2 compares the log likelihood on the data across models, demonstrating that ECTBN generally performs better than the baselines in this application except for the Transportation outcome area.

Conclusions

We have proposed a novel, interpretable graphical model that captures joint dynamics of both event occurrences and state variable transitions. Our focus was primarily on learning causal relations from events to state variables. In that regard, we presented an algorithm for learning the model structure and parameters, and demonstrated the model's effectiveness through experiments on synthetic data as well as a real-world application with data from CityLink Center. Potential directions for future work include models that incorporate even more complex historical dependencies of events and a decision making framework using parameters obtained from structure learning to optimally guide system evolution to achieve a desired goal.

Acknowledgments

References

Acerbi, E.; Zelante, T.; Narang, V.; and Stella, F. 2014. Gene network inference using continuous time Bayesian networks: a comparative study and application to Th17 cell differentiation. *BMC Bioinform.* 15(1):387.

Bhattacharjya, D.; Subramanian, D.; and Gao, T. 2018. Proximal graphical event models. In *Adv. Neur. Inf. Process. Syst. (NIPS)*.

Boudali, H., and Dugan, J. B. 2006. A continuous-time Bayesian network reliability modeling, and analysis framework. *IEEE T. Reliab.* 55(1):86–97.

Dahlhaus, R. 2000. Graphical interaction models for multivariate time series. *Metrika* 51:157–172.

Dean, T., and Kanazawa, K. 1989. A model for reasoning about persistence and causation. *Computational Intelligence* 5:142–150.

Dekhtyar, A.; Goldsmith, J.; Goldstein, B.; Mathias, K. K.; and Isenhour, C. 2009. Planning for success: The interdisciplinary approach to building Bayesian models. *Int. Journal Approx. Reason.* 50(3):416–428.

Didelez, V. 2008. Graphical models for marked point processes based on local independence. *J. R. Stat. Soc. B* 70(1):245–264.

Du, N.; Dai, H.; Trivedi, R.; Upadhyay, U.; Gomez-Rodriguez, M.; and Song, L. 2016. Recurrent marked temporal point processes: Embedding event history to vector. In *Proc. SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 1555–1564.

Eichler, M. 1999. *Graphical Models in Time Series Analysis*. Ph.D. Dissertation, University of Heidelberg, Germany.

Gatti, E.; Luciani, D.; and Stella, F. 2012. A continuous time Bayesian network model for cardiogenic heart failure. *Flex. Serv. Manuf. J.* 24(4):496–515.

Gehlbach, K. R.; Laracuente, B.; Isenhour, C.; Goldsmith, J.; Goldstein, B.; and Truszczyński, M. 2006. A benchmark model for decision-theoretic planning with constraints. In *Proc. Bayesian Model. Appl. Workshop*, 68–73.

Gopalratnam, K.; Kautz, H.; and Weld, D. S. 2005. Extending continuous time Bayesian networks. In *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 981–986.

Gunawardana, A., and Meek, C. 2016. Universal models of multivariate temporal point processes. In *Proc. Int. Conf. Artif. Intell. Stat. (AISTATS)*, 556–563.

Gunawardana, A.; Meek, C.; and Xu, P. 2011. A model for temporal dependencies in event streams. In *Adv. Neur. Inf. Process. Syst. (NIPS)*, 1962–1970.

Kube, A. R.; Das, S.; and Fowler, P. J. 2018. Allocating interventions based on counterfactual predictions: A case study on homelessness services. In *Proc. Workshop Mechan. Des. Social Good*.

Meek, C. 2014. Toward learning graphical and causal process models. In *Proc. UAI Workshop Causal Inference: Learning and Prediction*, 43–48.

Murphy, K. 2002. *Dynamic Bayesian Networks: Representation, Inference and Learning.* Ph.D. Dissertation, University of California Berkeley, USA.

Nodelman, U.; Shelton, C. R.; and Koller, D. 2002. Continuous time Bayesian networks. In *Proc. Conf. Uncertainty Artif. Intell. (UAI)*, 378–378.

Nodelman, U.; Shelton, C. R.; and Koller, D. 2003. Learning continuous time Bayesian networks. In *Proc. Conf. Uncertainty Artif. Intell. (UAI)*, 451–458.

Portinale, L., and Codetta-Raiteri, D. 2009. Generalizing continuous time Bayesian networks with immediate nodes. In *Proc. Workshop on Graph Structure for Knowledge Representation and Reasoning*, 12–17.

Rajaram, S.; Graepel, T.; and Herbrich, R. 2005. Poisson-networks: A model for structured point processes. In *Proc. Int. Workshop Artif. Intell. Stat. (AISTATS)*, 277–284.

Simma, A., and Jordan, M. I. 2010. Modeling events with cascades of Poisson processes. In *Proc. Conf. Uncertainty Artif. Intell. (UAI)*, 546–555.

Weiss, J. C., and Page, D. 2013. Forest-based point process for event prediction from electronic health records. In *Machine Learning and Knowledge Discovery in Databases*, 547–562.

Xiao, S.; Yan, J.; Yang, X.; Zha, H.; and Chu, S. M. 2017. Modeling the intensity function of point process via recurrent neural networks. In *Proc. AAAI Conf. Artif. Intell. (AAAI)*, volume 17, 1597–1603.

Xu, J., and Shelton, C. R. 2008. Continuous time Bayesian networks for host level network intrusion detection. In *Joint Eur. Conf. Mach. Learn. Knowl. Disc. Databases*, 613–627.

Yadav, A.; Chan, H.; Jiang, A.; Rice, E.; Kamar, E.; Grosz, B.; and Tambe, M. 2016a. POMDPs for assisting homeless shelters - computational and deployment challenges. In *Proc. IDEAS Workshop Int. Conf. Auton. Agents Multiagent Syst.*, 67–87.

Yadav, A.; Kamar, E.; Grosz, B.; and Tambe, M. 2016b. HEALER: POMDP planning for scheduling interventions among homeless youth. In *Proc. Int. Conf. Auton. Agents Multiagent Syst. (AAMAS)*, 1504–1506.

Yi, L.; Finkel, R.; and Goldsmith, J. 2008. Planning for welfare to work. In *Proc. Int. FLAIRS Conf.*, 696–701.