International Journal of Approximate Reasoning 70 (2016) 99–122

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Uncertain and negative evidence in continuous time Bayesian networks

Liessman Sturlaugson *, John W. Sheppard

Department of Computer Science, Montana State University, Bozeman, MT 59717, USA

ARTICLE INFO

Article history:
Received 16 January 2015
Received in revised form 16 December 2015
Accepted 17 December 2015
Available online 23 December 2015

Keywords:
Continuous time Bayesian network
Uncertain evidence
Negative evidence
Exact inference
Importance sampling

ABSTRACT

The continuous time Bayesian network (CTBN) enables reasoning about complex systems by representing the system as a factored, finite-state, continuous-time Markov process. Inference over the model incorporates evidence, given as state observations through time. The time dimension introduces several new *types* of evidence that are not found with static models. In this work, we present a comprehensive look at the types of evidence in CTBNs. Moreover, we define and extend inference to reason under uncertainty in the presence of uncertain evidence, as well as negative evidence, concepts extended to static models but not yet introduced into the CTBN model.

1. Introduction

Probabilistic models, such as Bayesian networks, provide a mathematically rigorous framework for reasoning under uncertainty. Given observations (i.e., evidence) about the system that the model represents, the model can update the posterior probabilities of other states of the system in light of that evidence. However, the representation of evidence in Bayesian networks has been further extended to allow for uncertainty in the evidence, in which there is uncertainty in the observations themselves. For example, a medical test might have false positive and false negative rates. Thus, the test result can be trusted only to a certain degree, which uncertain evidence is able to capture. Uncertain evidence provides one generalization of evidence, in which the evidence also has an associated likelihood.

Furthermore, one can think of negative evidence. Instead of an observation of the system being in a certain state, we can observe the system to *not* be in certain states, but it could be any of several other states. For example, we may be modeling a system in which sensors may only be able to detect a subset of the possible states. A negative sensor reading would imply that the system is in one of those other states. Negative evidence provides another generalization of evidence, in which multiple states can be ruled out instead of a single state being given.

By lifting restrictions on the types of evidence that the models can support, generalizations of evidence make the models more powerful and versatile. Rather than having to transform the observations into a form of supported evidence, such as treating all observations as certain or ignoring incomplete observations, generalizing evidence allows the observations to be used as evidence directly.

As the CTBN is a relatively new model, current CTBN inference algorithms only support certain and positive evidence, in which all of the temporal state evidence is trusted with complete confidence and in which the temporal state evidence

* Corresponding author.
E-mail addresses: listurlaugson@gmail.com (L. Sturlaugson), john.sheppard@cs.montana.edu (J.W. Sheppard).

http://dx.doi.org/10.1016/j.ijar.2015.12.013
0888-613X/

dictates what the state *is*, never what it is not. However, in many applications, the evidence may contain errors and can be trusted only to a certain degree. In some cases, only subsets of the states may be observable. Here, the model can benefit from ruling out certain states by reasoning with negative evidence. Uncertain and negative evidence for CTBNs has not yet been defined nor have CTBN inference algorithms been extended to allow for incorporating negativity and uncertainty into the temporal evidence. Therefore, this work proposes a definition for uncertain and negative evidence and shows how to support the definitions in the context of CTBNs. The introduction of timing information adds another dimension into the types of evidence we can apply. We show how combinations of certain/uncertain evidence and positive/negative evidence interact. By so doing, we show how the definitions of uncertain and negative evidence provide a generalization of certain positive evidence.

2. Background work

In this section, we give the formal definition of the Bayesian network (BN), as well as its temporal version, the dynamic Bayesian network (DBN). Next we give the formal definition of the CTBN model, compare and contrast the CTBN and the DBN, and then survey existing CTBN applications and extensions.

2.1. Bayesian networks

Bayesian networks are probabilistic graphical models that use nodes and arcs in a directed acyclic graph to represent a joint probability distribution over a set of variables [1]. Let $P(\mathbf{X})$ be a joint probability distribution over n variables $\mathbf{X} = \{X_1, \ldots, X_n\}$. A Bayesian network $\mathcal{B}$ is a directed, acyclic graph in which each variable X_i is represented by a node in the graph. Let $\mathbf{Pa}(X_i)$ denote the parents of node X_i in the graph. The graph representation of $\mathcal{B}$ factors the joint probability distribution as:

$$P(\mathbf{X}) = \prod_{i=1}^{n} P(X_i|\mathbf{Pa}(X_i)). \tag{1}$$

This factorization is induced by the conditional independences in the underlying distribution. Without any factorization, the number of parameters required to define the full joint probability distribution is exponential in the number of variables. By factoring the joint probability distribution to consider only the relevant variable interactions, represented by the parent–child relationships in the network, often the complexity of the distribution can be managed.

2.2. Dynamic Bayesian networks

The Bayesian network defined above is a static model. However, we can introduce the concept of time into the network by assigning discrete timesteps to the nodes to create a dynamic Bayesian network, a temporal version of a BN.

A dynamic Bayesian network (DBN) is a special type of Bayesian network that uses a series of connected timesteps, each of which contains a copy of a regular Bayesian network defined over $\mathbf{X}$ indexed by time t (denoted $\mathbf{X}_t$). The probability distribution of a variable at a given timestep can be conditioned on states of that variable (or even other variables) throughout any number of previous timesteps as well as on other variables within the same timestep. In first-order DBNs, the nodes in each timestep are conditionally independent of any nodes further back than the immediately previous timestep, given that previous timestep. Therefore, the joint probability distribution for a first-order DBN factors as:

$$P(\mathbf{X}_0, \ldots, \mathbf{X}_k) = P(\mathbf{X}_0) \prod_{t=0}^{k} P(\mathbf{X}_{t+1}|\mathbf{X}_t). \tag{2}$$

Spanning multiple timesteps, the DBN can include any evidence gathered throughout that time and use it to help reason about state probability distributions across different timesteps. Often, the conditional probability tables of the DBN can be defined compactly by defining a prior network over $\mathbf{X}_0$ and a single temporal network over $\mathbf{X}_t$ that represents $P(\mathbf{X}_{t+1}|\mathbf{X}_t)$ for every t. The temporal network $\mathbf{X}_t$ is then “unrolled” to consider $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k$ for k timesteps.

2.3. Continuous time Bayesian networks

The continuous time Bayesian network was first introduced in [2]. Although its name attempts to draw parallels between the conditional independence encoded by Bayesian networks, the CTBN represents a factored Markov process.

Let $\mathbf{X}$ be a set of conditional Markov processes $\{X_1, X_2, \ldots, X_n\}$, where each conditional process X_i has a finite number of discrete states. Formally, a continuous time Bayesian network $\mathcal{N} = \langle \mathcal{B}, \mathcal{G} \rangle$ over $\mathbf{X}$ consists of two components. The first is a Bayesian network $\mathcal{B}$ with nodes corresponding to $\mathbf{X}$. This Bayesian network is only used for determining $P(\mathbf{X}_0)$, the initial distribution of the process. The second is a continuous-time transition model $\mathcal{G}$, which describes the evolution of the process from its initial distribution, specified as:

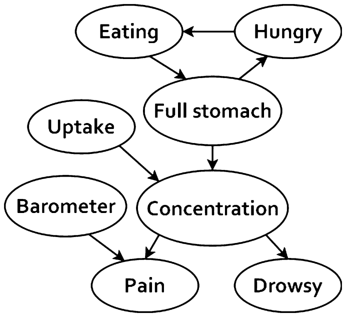


Fig. 1. Example CTBN.

- A directed graph with nodes $X_1, X_2, \dots, X_n$, where $\mathbf{Pa}(X_i)$ denotes the parents of X_i,
- A set of conditional intensity matrices (CIMs) $\mathbf{Q}_{X|\mathbf{Pa}(X)}$ associated with X for each possible state instantiation of $\mathbf{Pa}(X)$.

Each conditional intensity matrix $\mathbf{Q}_{X|\mathbf{Pa}(X)}$ gives the dynamics of node X when the states of $\mathbf{Pa}(X)$ are fixed. Each entry $q_{i,j} \geq 0$, $i \neq j$ as an element of $\mathbf{Q}_{X|\mathbf{Pa}(X)}$ gives the transition intensity of the node moving from state i to state j, and each entry $q_{i,i} \leq 0$ controls the amount of time the node remains in state i. With negative diagonal entries, the probability density function for the node remaining in state i is given by $|q_{i,i}| \exp(q_{i,i}t)$, with t being the amount of time spent in state i (called the sojourn time), making the probability of remaining in a state decrease exponentially with respect to time. The expected sojourn time for state i is $1/\left|q_{i,i}\right|$. Each row is constrained to sum to zero, $\sum_j q_{i,j} = 0 \ \forall \ i$, meaning that the transition probabilities from state i can be calculated as $q_{i,j}/\left|q_{i,i}\right| \ \forall \ j, i \neq j$.

Fig. 1 shows an example CTBN [2]. The initial distribution and intensity matrices for all the nodes can be found in the demo files for the Continuous Time Bayesian Network Reasoning and Learning Engine [3]. Each child node has multiple intensity matrices, one for each combination of states of its parent nodes. For example, the matrix denoted $\mathbf{Q}_{\text{Concentration}|u_1, f_0}$ defines the transition intensities of the node Concentration given that the state of Uptake is u_1 and that the state of Full stomach is f_0.

This model could be used to answer several interesting queries. For example, what is the expected proportion of time that the patient is in pain while drowsy? Or, given that the patient is initially in pain but that uptake occurred at time t_1 and that the patient finished eating at time t_2, what is the expected amount of time until the patient is not in pain? Or, given that the patient has been in pain from time t_2 to t_3, what is the expected number of transitions between the concentration levels that occurred during that time period? The user could query the probability of any state of any node at any real-valued time, conditioned on past and future observations about the system.

2.4. DBNs vs. CTBNs

Unlike a regular Bayesian network, cycles are allowed in the graph of $\mathcal{G}$. Only one state transition in one node is allowed at a time for the whole CTBN. A cycle in a CTBN model would be analogous to a dynamic Bayesian network with variables X and Y where arcs such as $X_t \to Y_{t+1}$ and $Y_t \to X_{t+1}$ would be valid.

Similar to the Bayesian network, exploiting the conditional independences allows for a more compact representation for the model. In a Bayesian network, the local conditionally independent distributions can be combined to form the full joint probability distribution. In the case of the CTBN, this is called the full joint intensity matrix, which describes the evolution of the entire process. However, just as in the Bayesian network, in which the number of entries in the full joint probability distribution grows exponentially in the number of variables, so too the number of states in the full joint intensity matrix grows exponentially in the number of variables for the CTBN.

Despite these few similarities, the CTBN model is fundamentally different from the DBN model. Although the network topologies for both models encode conditional independence, the models are differentiated by what the nodes represent. Whereas the nodes in a DBN are simple random variables, the nodes in a CTBN are conditional Markov processes. As a result, CTBNs can be queried about the state probabilities for any real-valued time. A DBN, unrolled for a discrete number of timesteps, can only be queried for state probabilities at these timesteps but not in-between adjacent timesteps. While the time interval between timesteps can be set with finer granularity, doing so multiplies the number of nodes needed to span the same amount of time as the original unrolled DBN. In fact, a DBN becomes asymptotically equivalent to a CTBN only as the interval of time between timesteps approaches zero [4].

2.5. CTBN inference and learning

The only exact inference algorithm that exists so far for CTBNs simply expands and works with the full joint intensity matrix, which is exponential in the number of nodes and the number of states [5]. However, this inference algorithm does not take advantage of the factored nature of the network. Thus, research on inference has focused on approximate methods.

There have been a number of sample-based inference algorithms developed for CTBNs, such as importance sampling [5–7] and Gibbs sampling [8,9]. In importance sampling, multiple samples are generated that are constrained to conform to the evidence. The particles are then weighted by their likelihood. Gibbs sampling, on the hand, takes a Markov Chain Monte Carlo (MCMC) approach. For each variable over each interval of evidence, the surrounding states (the node's parents, children, and children's parents) are held constant and a random walk is performed on the state of the node. The idea is that, on each interval of evidence, the distribution of the random walk will converge to the true distribution. Methods for expectation propagation [10,11] have also been developed. In these methods, for each interval of evidence, the nodes employ a message-passing scheme with their neighbors. The idea is to pass approximate "marginals" that are unconditional intensity matrices, valid for that interval, until all of the nodes have a consistent distribution over that interval. Recently, methods using mean-field approximation [12,13] and belief propagation [14] have also been developed, which propagate the product of inhomogeneous Markov processes to approximate the distribution through a system of ordinary differential equations. There have also been approaches specifically for continuous-time filtering in CTBNs [15,16].

As a data-driven model, algorithms have been developed for model learning, both with learning the network structure [17], as well as learning and maturing the model parameters [18]. These algorithms define the sufficient statistics for the nodes of the CTBN and define the log-likelihood that can be used to measure how well the network structure and the parameters match the data. They also support learning in the presence of missing data, leading to an expectation maximization algorithm for CTBNs. For a more comprehensive tutorial on CTBNs, see [19].

2.6. CTBN applications and extensions

CTBNs have found use in several applications. For example, CTBNs have been used for inferring computer users' presence, activity, and availability over time [20]; monitoring robotic systems [16]; modeling server farm failures [21]; modeling social network dynamics [22,23]; modeling sensor networks [24]; building intrusion detection systems [25–27]; predicting the trajectory of moving objects [28,29]; diagnosing cardiogenic heart failure and anticipating its likely evolution [30,31]; modeling and inferring about gene networks [32,33]; and diagnosing and prognosing equipment [34]. The CTBN has also been extended to support decision-making, resulting in structured continuous-time Markov decision processes [35].

The CTBN model has also undergone several specializations and generalizations. The Generalized CTBN (GCTBN) of [36] combines the conditional probability tables of BNs and the conditional intensity matrices of CTBNs, allowing nodes to be either what they call "delayed" variables or "immediate" variables. They show how inference can be performed when combining the conditional probabilities of the immediate nodes with the intensity matrices of the delayed nodes. The CTBN classifier (CTBNC) of [37–42] is an instance of the GCTBN, adding a parent-less immediate class node, with marginal probabilities over the class label, for classifying a static object given continuous-time evidence about that object. The work of [43,44] changes the representation of the CTBN to be partition-based, using what they call conditional intensity trees and conditional intensity forests. This is analogous to representing the conditional probabilities in BNs as decision trees [45]. The Erlang–Coxian CTBN (EC-CTBN) of [46] replaces the exponential distribution of the sojourn times with Erlang–Coxian distributions. Nodelman showed how combinations of nodes in the CTBN could represent Erlang–Coxian distributions [10, 47], but the EC-CTBN replaces this with a single node and introduces a generalized conditional intensity matrix. Lastly, the asynchronous dynamic Bayesian network (ADBN) of [48] maintains a CTBN, converting the conditional intensity matrices to conditional probability tables to perform inference. The idea is that the parameters of the DBN will be populated from the continuous-time evidence for the CTBN, avoiding the assumption of a uniform interval of time between all the timesteps of the DBN. After converting the conditional intensity matrices to conditional probability tables, inference can be performed over the DBN instead of the CTBN.

3. Types of evidence

While we use the terms "system" and "sensor" to describe the semantics of different types of evidence, these terms should be taken in their general sense through this paper. We use the term "system" to refer to whatever the model represents. These could be such varied systems as computer systems, vehicle systems, sensor networks, groups of people or individuals—just to name a few. We use the term "sensor" to refer to any means by which evidence about the system is gathered. These could be electrical or mechanical sensors (calibrated/drifting/noisy?), direct human observations (trustworthy/biased/adversarial?), educated guesses (with how much certainty?), or even hypothetical scenarios (if we observed the system to be in some given state, what would likely happen next?).

In this section we describe the types of evidence currently used in inference with BNs, which include certain, uncertain, and negative evidence. We then review the types of evidence currently defined for CTBNs. Finally, we define uncertain and negative types of continuous-time evidence for use in CTBNs.

3.1. Evidence in Bayesian networks

We start by reviewing the types of evidence used in Bayesian networks.

Fig. 2. Example of virtual evidence for node X.

3.1.1. Certain positive evidence

Traditional evidence in a BN corresponds to certain evidence. That is, the observations are trusted with complete confidence, and the observations are of specific states. Suppose that we had a set of observations $\mathbf{e}$. To perform inference with this evidence, we would compute the posterior probabilities $P(\mathbf{X}|\mathbf{e})$. Once we observe the state of a variable, the probability of that state for that variable becomes 1. From this, we can generalize to uncertain evidence, in which the probabilities of observed states can become less than 1.

3.1.2. Uncertain positive evidence

One approach for uncertain evidence in Bayesian networks is called virtual evidence [49]. One way to represent this type of uncertain evidence is to add a node as a child to an observed node, as shown in Fig. 2. This child node then becomes the node that is observed, and the conditional probabilities of this child are set to match the strength of the evidence. In effect, virtual evidence sets up a ratio of likelihoods to represent the confidence of an observation observing a particular state. Suppose the uncertain evidence for node X with states $x_1, \ldots, x_n$ is given as η. Then for x_i we set

$$P(\eta|X = x_i) = \lambda_i, \tag{3}$$

where λ_i is the likelihood of the evidence for x_i. The child node is conditionally independent of all other nodes given X. That is, for any event α in the probability space,

$$P(\eta|X = x_i, \alpha) = P(\eta|X = x_i). \tag{4}$$

Then the probability of event α given the uncertain evidence η is

$$P(\alpha|\eta) = \frac{\sum_{i=1}^{n} \lambda_i P(\alpha, X = x_i)}{\sum_{i=1}^{n} \lambda_i P(X = x_i)}. \tag{5}$$

By so doing, virtual evidence weights the marginal probabilities $P(X)$ by the likelihood of the evidence λ_i for x_i.

Another approach for representing uncertainty in an observation is through what is called soft evidence [50,51]. Soft evidence uses Jeffrey's rule as a generalization of conditioning on observed variables to condition on an observation of a probability distribution, in which the observed distribution holds the uncertainty of the observations. Suppose that the evidence is specified by a set of probabilities

$$P'(X = x_i) = q_i. \tag{6}$$

Then the new posterior distribution for event α is calculated as

$$P'(\alpha) = \sum_{i=1}^{n} q_i \frac{P(\alpha, X = x_i)}{P(X = x_i)} = \sum_{i=1}^{n} q_i P(\alpha|X = x_i). \tag{7}$$

As shown by Equation (6), soft evidence actually transforms the marginal distribution of X from $P(X)$ to $P'(X)$ such that it conforms exactly to the probability of the evidence.

Because each approach specifies uncertain evidence differently, virtual evidence and soft evidence yield different probabilities given the same values for q_i and λ_i, but conversions exist to derive the resulting probability of one given the other [49].

3.1.3. Certain negative evidence

In a Bayesian network, certain negative evidence is just a special case of uncertain positive evidence. If we observe the variable to *not* be in some states, this is the same as uncertain evidence in which there is zero probability of those states and the remaining probabilities for all other states are re-normalized.

Let A be a subset of states of X that can be ruled out. Negative evidence states that $X \notin A$. We can use virtual evidence to support negative evidence in Bayesian networks by choosing $c > 0$ and setting

$$\lambda_i = \begin{cases} 0 & \text{for } x_i \in A \\ c & \text{for } x_i \notin A \end{cases}. \tag{8}$$

It follows that for every event α given uncertain evidence η that the states of A can be ruled out with likelihoods λ_i given above,

$$P(\alpha|\eta) = \frac{\sum_{i=1}^{n} \lambda_i P(\alpha, X = x_i)}{\sum_{i=1}^{n} \lambda_i P(X = x_i)} = \frac{c \sum_{x \notin A} P(\alpha, X = x)}{c \sum_{x \notin A} P(X = x)} = \frac{P(\alpha, X \notin A)}{P(X \notin A)} = P(\alpha | X \notin A). \tag{9}$$

Thus, when dealing with a static Bayesian network, uncertain evidence is sufficient to represent negative evidence.

3.2. Evidence in CTBNs

While static BNs use evidence as observations of states, evidence in CTBNs must include temporal information. We use the notation $X(t)$ to denote the state of X at an instantaneous point in time t, while the notation $X([t_1, t_2))$ denotes the state of X on the half-open interval from t_1 to t_2. For example, $X([t_1, t_2)) = x$ means that $X(t) = x \ \forall t \in [t_1, t_2)$, while $X([t_1, t_2)) \neq x$ means that $X(t) \neq x \ \forall t \in [t_1, t_2)$. Three types of evidence in CTBNs have been defined in the literature so far: point, transition, and interval.

The first inference algorithms defined for CTBNs supported only point evidence [2]. Certain point evidence in a CTBN is defined formally as follows.

Definition 3.1 *(Certain point evidence).* Let t be an instantaneous point in real-valued time. Let X be a node in the CTBN, and let x be a state of X. Certain point evidence is an observation of an event of the form $X(t) = x$.

Point evidence would result when the system cannot be monitored continuously, and sensors can only "poll" the state at various instantaneous points in time. For example, blood test results might be most appropriately cast as point evidence, as they provide a current "snapshot" of the system but cannot continuously observe the state between tests.

Unlike a static model, the states of the CTBN could be changing throughout time. When monitoring a system, we might be able to detect changes in the state of the system. We would like to be able to incorporate this transition information into our inference procedures. Certain transition evidence in a CTBN is defined formally as follows.

Definition 3.2 *(Certain transition evidence).* Let t be an instantaneous point in real-valued time, and let ϵ be an arbitrarily small positive value. Let X be a node in the CTBN, and let x_1 and x_2 be two distinct states of X. Certain transition evidence is an observation of an event of the form, $X([t - \epsilon, t)) = x_1 \wedge X(t) = x_2$.

Transition evidence would result when sensors can detect certain changes in the system. For certain transition evidence, the sensors can detect exactly when and how the change occurs. Note that if the sensors can detect every state change, they can observe the complete path of the system.

For continuous-time systems, however, evidence at an instantaneous point in time is not powerful enough. We might be monitoring the system continuously and be able to make claims about the state of the system throughout an interval of time. When this idea was first introduced, it was called negative evidence, but the "negative" referred to transitions rather than states [46]. The idea was that a transition did *not* occur over an interval of time. Since then, it has changed to be referred to as continuous evidence or interval evidence. Certain interval evidence in a CTBN is defined formally as follows.

Definition 3.3 *(Certain interval evidence).* Let t_1 and t_2 be instantaneous points in real-valued time such that $t_1 < t_2$. Let X be a node in the CTBN, and let x be a state of X. Certain interval evidence is an observation of an event of the form $X([t_1, t_2)) = x$.

Interval evidence would result when the system is able to be monitored throughout a continuous interval of time. For example, a nurse could observe that a patient was breathing normally throughout the entire time the nurse was in the room. The evidence would be an interval of a single state (normal breathing). Outside of that interval (when the nurse was outside the room and no longer observing the patient), the state might have transitioned multiple times.

Certain interval evidence is able to approximate both certain point and certain transition evidence [6]. For point evidence, we set $t_2 = t_1 + \epsilon$ for some infinitesimal value ϵ. For certain transition evidence, this becomes two successive instances of infinitesimally short interval evidence such that on a real-valued interval of time $[t - \epsilon, t)$, node X was observed to be in

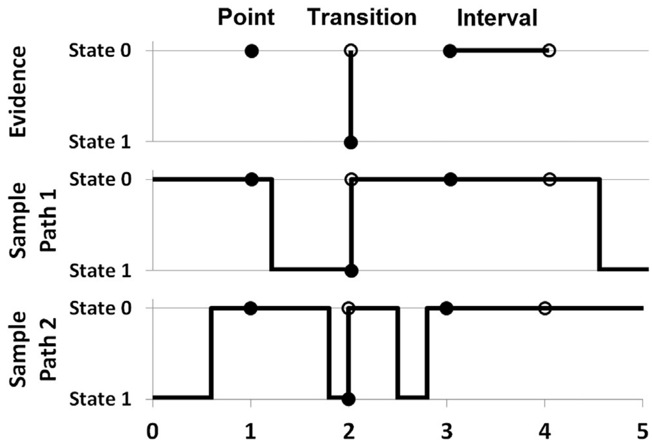


Fig. 3. Examples of evidence for a CTBN.

state x_1, while on a real-valued interval of time $[t, t+\epsilon)$, the state of node X was observed to be in state x_2. Because of this, in the remainder of this paper we focus on supporting interval evidence for CTBN inference algorithms.

Fig. 3 shows examples of these three evidence types for a two-state node. The figure also shows two possible complete sample paths that conform to the evidence. The evidence can be thought of as a partial sample path, a sequence of state and time pairs with gaps during which the state becomes unknown. In the example, point evidence at time $t = 1$ observes the node to be in state 0. Transition evidence at time $t = 2$ observes the state to transition from state 1 to state 0. Interval evidence from time $t = 3$ to $t = 4$ observes the state to remain in state 0.

3.3. *Extending evidence in CTBNs*

As seen with the progression from point evidence to transition and interval evidence, the temporal nature of the model gives rise to more varied types of continuous-time evidence. Uncertain evidence, as used in BNs, has not yet been extended to CTBNs. Furthermore, the introduction of time adds another dimension. We can differentiate between positive and negative evidence when the evidence becomes temporal, because now evidence must be defined in terms of both state *and* time. Temporal evidence introduces subtleties between uncertain evidence and negative evidence, and they become two distinct types of evidence when defined in continuous time. We now present the first definitions for these types of evidence in the CTBN.

3.3.1. *Uncertain positive evidence*

In this section, we define and describe uncertain positive evidence for CTBNs. First, we have uncertain point evidence, defined formally as follows.

Definition 3.4 *(Uncertain point evidence).* Let t be an instantaneous point in real-valued time. Let X be a node in the CTBN, and let λ_i be a likelihood for state x_i of X. Uncertain point evidence is an event η such that

$$P(\eta|X(t) = x_i) = \lambda_i \text{ for } i = 1, \ldots, n. \tag{10}$$

When we condition on this uncertain point evidence η for any event α, the probabilities are reweighted as

$$P(\alpha|\eta) = \frac{\sum_{i=1}^{n} \lambda_i P(\alpha, X(t) = x_i)}{\sum_{i=1}^{n} \lambda_i P(X(t) = x_i)}. \tag{11}$$

Uncertain point evidence would result when "polling" sensors can only be trusted to a certain degree. For example, many medical tests have quantifiable false positive and false negative rates. With uncertain point evidence, these uncertain snapshots of the system can be formalized and incorporated into the temporal inference process.

The likelihoods are analogous to λ_i of Equation (3). Because the observation is at a single instant in time, uncertain point evidence is also sufficient to represent negative point evidence.

We can also have uncertainty in an observed transition. Uncertain transition evidence is used when the source and/or destination of a transition is known only with some probability. Uncertain transition evidence in a CTBN is defined formally as follows.

Definition 3.5 *(Uncertain transition evidence).* Let t be an instantaneous point in real-valued time, and let ϵ be an arbitrarily small positive value. Let X be a node in the CTBN, and let $\lambda_{i,j}$ be the likelihood for a transition from x_i to x_j. Uncertain transition evidence is an event η such that

$$P(\eta|X([t-\epsilon,t))=x_i, X(t)=x_j)=\lambda_{i,j} \text{ for } i,j=1,\ldots,n. \tag{12}$$

When we condition on this uncertain transition evidence η for any event α, the probabilities are reweighted as

$$P(\alpha|\eta)=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_{i,j}P(\alpha, X([t-\epsilon,t))=_i, X(t)=x_j)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_{i,j}P(X([t-\epsilon,t))=x_i, X(t)=x_j)}. \tag{13}$$

Uncertain transition evidence would result when a sensor is able to determine only partially the source and/or destination state. The set of likelihoods can also be used to represent uncertainty that the state actually changed. That is, the uncertainty in the transition evidence could allow for non-zero probability that no transition occurred. This is done by setting $\lambda_{i,i}$ to be non-zero. Because a state does not transition to itself in a continuous-time model, this implies that the state simply remained in state x_i at time t. For example, the shell game could give rise to uncertain transition evidence. The player cannot distinguish between shells but relies on tracking changes (transitions) in their position (state). However, by the nature of the game, the player may become uncertain about which shell transitioned to which state. The player's confidence in knowing the final state will be influenced by the uncertainty of the observed transitions.

The definitions above represent uncertainty in the state. However, we could have uncertainty in the timing information as well. Therefore, for completeness, we can formally define temporally uncertain transition evidence as follows.

Definition 3.6 *(Temporally uncertain transition evidence).* Let t_1 and t_2 be instantaneous points in real-valued time such that $t_1 < t_2$. Let X be a node in the CTBN, and let x_i and x_j be two distinct states of X. Temporally uncertain transition evidence is an observation of an event of the form $X(t_1)=x_i \wedge X(t_2)=x_j$.

Temporally uncertain transition evidence would result when the sensor is able to detect state changes in the system, but not instantaneously. The sensor might have a non-constant time delay before the state change is detected. Note that temporally uncertain transition evidence does not rule out multiple transitions during the unknown period. The evidence merely gives the beginning and end states over that period.

Lastly, we have uncertain positive interval evidence. In this case, uncertain interval evidence cannot be used for negative interval evidence, because uncertain interval evidence holds the state constant over the interval while negative interval evidence only rules out states over the interval (but transitions may occur between other states). Therefore, uncertain interval evidence must be identified as either positive or negative. Negative interval evidence is discussed in the next section, while uncertain positive interval evidence is defined formally as follows.

Definition 3.7 *(Uncertain positive interval evidence).* Let t_1 and t_2 be instantaneous points in real-valued time such that $t_1 < t_2$. Let X be a node in the CTBN, and let λ_i be a likelihood of the evidence for state x_i. Uncertain positive interval evidence is an event η such that

$$P(\eta|X([t_1,t_2))=x_i)=\lambda_i \text{ for } i=1,\ldots,n, \tag{14}$$

$$P(\eta|\wedge_{i=1}^{n}\neg X([t_1,t_2))=x_i)=0. \tag{15}$$

When we condition on this uncertain positive interval evidence η for any event α, the probabilities are reweighted as

$$P(\alpha|\eta)=\frac{\sum_{i=1}^{n}\lambda_i P(\alpha, X([t_1,t_2))=x_i)}{\sum_{i=1}^{n}\lambda_i P(X([t_1,t_2))=x_i)}. \tag{16}$$

Uncertain positive interval evidence would result when the sensor is able to detect changes in the system but is unable to determine the state over the interval with certainty. For uncertain interval evidence, the state of node X is known to be in exactly one state over the entire interval, but the identity of that state is only known with some likelihood.

3.3.2. Negative evidence

In CTBNs, uncertain evidence is distinct from negative evidence. Uncertain interval evidence, for example, says that the system was in different states with different likelihoods, but whichever state it was, the system stayed in that state over the whole interval. Negative evidence is saying something different. Negative interval evidence, for example, says that the system was never in a certain state over an interval; however, the system could have experienced multiple transitions between the other states over that interval. Formally, we define negative point, transition, and interval evidence as follows.

Definition 3.8 *(Negative point evidence).* Let t be an instantaneous point in real-valued time. Let X be a node in the CTBN, and let X' be a proper subset of the states of X. Negative point evidence is an observation of an event of the form $X(t) \notin X'$.

Negative point evidence rules out one or more states at a single point in time, meaning it does not always positively identify the actual state at that point in time. Returning to the shell game example, suppose that the operator, while shuffling the shells, shows one of the shells as empty. The player can incorporate this as negative point evidence, because the player can rule out that shell at that instant with complete certainty.

Definition 3.9 *(Negative transition evidence).* Let t be an instantaneous point in real-valued time, and let ϵ be an arbitrarily small positive value. Let X be a node in the CTBN, and let X_S and X_E be non-empty proper subsets of the states of X such that $X_S \cup X_E$ includes every state of X. Negative transition evidence is an observation of the form $X([t-\epsilon, t)) \notin X_S \wedge X(t) \notin X_E$.

Negative transition evidence is an observation of a transition but can only rule out one or more source and/or destination states, meaning it does not always positively identify both source and destination states of the transition. For example, consider a security camera that is able to monitor a section of a hallway to multiple rooms. The camera observes only some of the doors along the hallway (states). If an individual walks into view of the camera from the hallway (transition), the camera observes the change, and we can rule out that the individual came from those rooms whose doors are in view of the camera.

Definition 3.10 *(Negative interval evidence).* Let t_1 and t_2 be instantaneous points in real-valued time such that $t_1 < t_2$. Let X be a node in the CTBN, and let X' be a proper subset of the states of X. Negative interval evidence is an observation of an event of the form $X([t_1, t_2)) \notin X'$.

Negative interval evidence rules out one or more states on an interval of time, meaning it does not always positively identify the actual state on that interval. Consider the security camera example again, assuming that adjoining rooms may have doors between them. While the security camera is on and observing an empty hallway, we can rule out anyone in the hallway during that period. However, an individual may be moving between rooms (states) that the camera cannot observe.

Note that uncertain and negative evidence are not mutually exclusive. We could have uncertainty in our negative evidence, in which we can rule out some states only with a certain probability. For instantaneous evidence, such as point evidence and transition evidence, uncertain evidence is sufficient to represent these, as shown earlier. Uncertain positive interval evidence, on the other hand, cannot represent uncertain negative interval evidence, which is defined formally as follows.

Definition 3.11 *(Uncertain negative interval evidence).* Let t_1 and t_2 be instantaneous points in real-valued time such that $t_1 < t_2$. Let X be a node in the CTBN, and let λ_i be a likelihood of the evidence for state x_i where $\exists i, \lambda_i < 1$. Uncertain negative interval evidence is an event η such that

$$P(\eta | X([t_1, t_2)) \neq x_i) = \frac{1}{2 - \lambda_i} \tag{17}$$

$$P(\eta | \neg X([t_1, t_2)) \neq x_i) = \frac{1 - \lambda_i}{2 - \lambda_i} \tag{18}$$

When we condition on this uncertain negative interval evidence η for any event α, the probabilities are reweighted as

$$P(\alpha | \eta) = \frac{\sum_{i=1}^{n} (1 - \lambda_i) P(\alpha, \neg X([t_1, t_2)) \neq x_i)}{\sum_{i=1}^{n} (1 - \lambda_i) P(\neg X([t_1, t_2)) \neq x_i)}. \tag{19}$$

Uncertain negative interval evidence rules out states on an interval of time but only with some level of certainty. Consider the security camera example again but assume that it also has facial recognition software that tries to track the movement of a specific individual. Suppose the facial recognition algorithm returns a confidence level for recognizing each face. While people walk through the hallway, the facial recognition software tries to track whether one of them is the specific individual. While the facial recognition software does not detect the individual's face, we can rule out the individual being in the hallway during that period with some level of certainty. We cannot be completely certain about the interval in this case, because there is some probability that the facial recognition software did not detect the individual when the individual walked through.

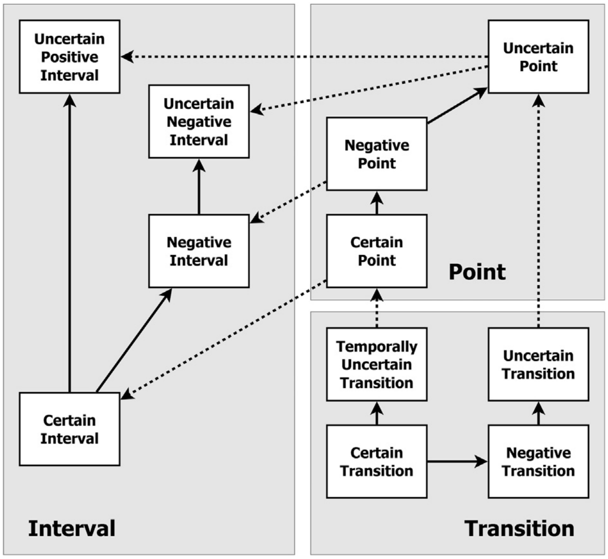


Fig. 4. Relationships between types of continuous-time evidence.

3.4. Relationships between types of evidence

Given these definitions, note that, unlike with Bayesian networks, uncertain evidence is insufficient to represent all types of negative evidence. Uncertain evidence introduces uncertainty in the state, not the duration of the evidence. In other words, for an interval of uncertain evidence $\mathbf{e}$ of node X on $[t_s, t_e)$,

$$P(X(t_1)|\mathbf{e}) = P(X(t_2)|\mathbf{e}), \ \forall\, t_1, t_2 \in [t_s, t_e). \tag{20}$$

On the other hand, for an interval of negative evidence $\mathbf{e}$ of node X on $[t_s, t_e)$, we can say that

$$P(X(t) \in A|\mathbf{e}) = 0, \ \forall\, t \in [t_s, t_e), \tag{21}$$

but the probabilities $P(X(t) = x|\mathbf{e})$ for $x \notin A$ could be changing throughout $[t_s, t_e)$. Therefore negative interval evidence is not necessarily a special case of uncertain positive interval evidence.

As already seen, the various types of evidence are related, whether through special cases or combinations. These relationships are summarized in Fig. 4. The solid arrows show when the parent type is a special case of the child type, while the dashed arrows show when the parent type can be approximated by the child type. Proof sketches for these relationships are given in Appendix A.

4. CTBN inference algorithms with extended evidence types

The types of continuous-time evidence defined up to this point are only useful if we can incorporate them when performing inference over the system. In this section, we show how exact inference and importance sampling can be extended to support the new types of evidence. Because interval evidence types can be used to approximate point and transition evidence types, we focus on extending inference algorithms to handle the uncertain and negative interval evidence. Once these are implemented, uncertain and negative variations for point and transition evidence types can also be simulated by special cases and combinations of the interval types.

We note that we cannot represent uncertain and negative evidence simply by modifying the network itself. This is in contrast to the dynamic Bayesian network (DBN) with virtual evidence, for example. When unrolled, the DBN has distinct nodes in each timestep to which the virtual evidence nodes can be attached. In a CTBN, there is a single node for each variable that persists throughout the entire process. Therefore, the uncertain evidence must be applied to the node at the right time during the inference process itself. We now show how exact inference and importance sampling can be extended to reason with uncertain and negative evidence.

We demonstrate the extended types of evidence on a real-world CTBN learned from the British Household Panel Survey dataset [52] as used previously by [5,15,17]. The network is shown in Fig. 5. The dataset recorded major life changes collected annually from a set of approximately 8000 British citizens. The nodes with an "H_" prefix are binary hidden variables that allow modeling of phase-type distributions, which can represent more complex sojourn time distributions than a single exponential distribution. The Smoking node has two states, {*non-smoker*, *smoker*}, the Married node has two states, {*single*, *married*}, the Children node has three states, {0, 1, 2+}, and the Employed node has three states, {*student*,

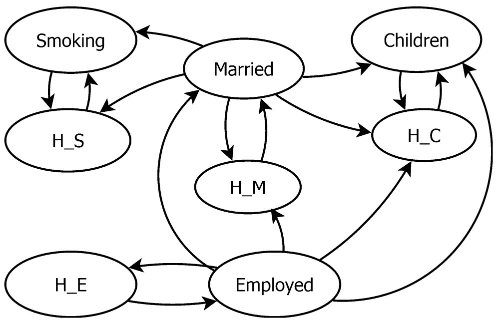


Fig. 5. British Household Panel Survey network.

employed, *unemployed*}. Three-state nodes allow us to show uncertain transitions and to show how negative evidence rules out one of those states and renormalizes the probability between the other two states. Nodes with parents and children also allow us to show how the new types of evidence affect the probabilities of nodes around the observed node. We assume that the initial states at time $t = 0$ are *non-smoker*, *single*, *0*, and *student*. For each demonstration of different evidence types, we query the probabilities at 25 points placed uniformly along 10-year span.

4.1. Exact inference for uncertain and negative evidence

Exact inference for CTBNs can be achieved by amalgamating all of the nodes into the full joint intensity matrix and performing the forward–backward algorithm for Markov processes [6]. Because the size of the full joint intensity matrix is exponential in the number of nodes, this algorithm quickly becomes intractable. We still show how to include uncertain and negative evidence with exact inference to provide a baseline for comparison to approximate methods, although even approximate inference in CTBNs, like exact inference, is known to be NP-hard [53].

First, consider when there is no evidence. Let $P(\boldsymbol{X}(0))$ be the initial distribution, and let $\boldsymbol{Q}$ be the full joint intensity matrix. The distribution at any point in time t can be calculated as

$$P(\boldsymbol{X}(t)) = P(\boldsymbol{X}(0)) \exp(\boldsymbol{Q}\, t), \tag{22}$$

where the matrix exponential is defined by the power series

$$\exp(\boldsymbol{Q}\, t) = \sum_{n=0}^{\infty} \frac{(\boldsymbol{Q}\, t)^n}{n!}. \tag{23}$$

Note that, although the matrix exponential is defined as above, there are more efficient and robust ways of computing it [54].

When continuous-time evidence is added, the matrix exponential is partitioned, resulting in a product of matrices that alternate between matrix exponentials over the segmented intervals and transition matrices between segments. The continuous-time evidence is given as a partial sample path. Let the partial sample path over the interval of time $[0, T)$ be denoted as $\sigma_{[0,T)}$. Suppose the evidence partitions σ into N segments, $[t_i, t_{i+1})$, for $i = (0, N-1)$. Let $\boldsymbol{Q}_i$ denote the full joint intensity matrix for segment i, meaning that the rows and columns of $\boldsymbol{Q}_i$ that do not conform to the evidence of segment i are zeroed out. Let $\boldsymbol{Q}_{i,j}$ denote the transition probabilities between segments i and j. If a transition is observed on the boundary between segments i and j, the rows and columns of $\boldsymbol{Q}_{i,j}$ are zeroed out except for the transition intensities from non-zero rows in $\boldsymbol{Q}_i$ to non-zero rows in $\boldsymbol{Q}_j$. Otherwise, segments i and j will differ only in what states are becoming observed or unobserved (instead of transitions being observed), in which case $\boldsymbol{Q}_{i,j}$ is the identity matrix. The segmentation of the matrix multiplications can then be defined recursively. Let $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ denote the forward and backward probability vectors, defined as

$$\boldsymbol{\alpha}_t = P(X(t), \sigma_{[0,t]}), \tag{24}$$

$$\boldsymbol{\beta}_t = P(\sigma_{[t,T)} | X(t)). \tag{25}$$

Let $\boldsymbol{\alpha}_0$ be the initial distribution $P(\boldsymbol{X}(0))$ over the states of $\boldsymbol{X}$, and let $\boldsymbol{\beta}_T$ be a vector of ones. Let $\Delta_{i,j}$ be an $n \times n$ matrix of zeros except for a one in position i, j. The recursive definitions for $\boldsymbol{\alpha}_{t_i}$ and $\boldsymbol{\beta}_{t_j}$ are

$$\boldsymbol{\alpha}_{t_{i+1}} = \boldsymbol{\alpha}_{t_i} \exp(\boldsymbol{Q}_i (t_{i+1} - t_i)) \boldsymbol{Q}_{i,i+1}, \tag{26}$$

$$\boldsymbol{\beta}_{t_i} = \boldsymbol{Q}_{i-1,i} \exp(\boldsymbol{Q}_i (t_{i+1} - t_i)) \boldsymbol{\beta}_{t_{i+1}}. \tag{27}$$

The distribution over state k of the CTBN at time $t \in [t_i, t_{i+1})$ given evidence $\sigma_{[0,T)}$ can be computed as

$$P(X(t)=k|\sigma_{[0,T)}) = \frac{1}{Z}\boldsymbol{\alpha}_{t_i}\exp(\boldsymbol{Q}_i(t-t_i))\Delta_{k,k}\exp(\boldsymbol{Q}_i(t_{i+1}-t))\boldsymbol{\beta}_{t_{i+1}}, \tag{28}$$

where Z is the normalizing constant.

The matrices $\boldsymbol{Q}_i$ and $\boldsymbol{Q}_{i-1,i}$ are modified to include the evidence for segment i. Let $\boldsymbol{Q}[l,m]$ denote entry (l,m) of matrix $\boldsymbol{Q}$. Let $\boldsymbol{Q}[l,*]$ and $\boldsymbol{Q}[*,m]$ denote row l and column m of matrix $\boldsymbol{Q}$, respectively. Let x denote a state of a node X in the CTBN, and let $I(x)$ denote the set of indices in Q_i and $Q_{i-1,i}$ that include state x in their state combination. Let $\mathbf{0}$ denote a row vector of zeros.

4.1.1. Exact inference with uncertain positive interval evidence

For uncertain evidence over node X on segment i, the rows and columns representing transitions between the states of X are zeroed out, because the uncertain evidence knows that the state is constant during the interval but is uncertain as to the identity of that state. For uncertain positive interval evidence on X, we iterate over all pairs of states x_j and x_k of X such that $x_j \neq x_k$. For each pair $l \in I(x_j)$ and $m \in I(x_k)$ such that $l \neq m$:

$$\boldsymbol{Q}_i[l,m] \leftarrow 0 \tag{29}$$

$$\boldsymbol{Q}_i[m,l] \leftarrow 0 \tag{30}$$

The matrix $\boldsymbol{Q}_{i-1,i}$ starts as the identity matrix. The diagonal elements corresponding to each state of X are multiplied by the uncertain evidence for that state. This weights the transition into each state according to the uncertainty in the evidence. For each $x_j \in X$ and for each $l \in I(x_j)$:

$$\boldsymbol{Q}_{i-1,i}[l,l] \leftarrow \lambda_j \cdot \boldsymbol{Q}_{i-1,i}[l,l] \tag{31}$$

We demonstrate uncertain positive interval evidence and certain negative interval evidence together in the next section as the latter is a special case of the former.

4.1.2. Exact inference with certain negative interval evidence

Negative evidence is a straightforward extension for exact inference. In this case, the rows and columns of $\boldsymbol{Q}_i$ are zeroed out in accordance with the negative evidence. For certain evidence, all but one state of X are zeroed out. For negative evidence, there can be multiple states of X that are not zeroed out in $\boldsymbol{Q}_i$. The matrix $\boldsymbol{Q}_{i-1,i}$ starts as the identity matrix as before. Formally, for each $x \in X'$, and then for each $l \in I(x)$:

$$\boldsymbol{Q}_i[l,*] \leftarrow \mathbf{0} \tag{32}$$

$$\boldsymbol{Q}_i[*,l] \leftarrow \mathbf{0}^T \tag{33}$$

$$\boldsymbol{Q}_{i-1,i}[l,l] \leftarrow 0 \tag{34}$$

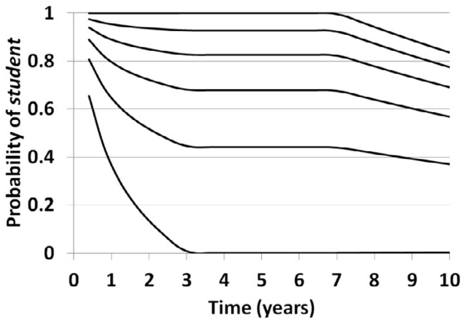
To demonstrate the effects of uncertain positive interval evidence and certain negative interval evidence, we vary the likelihood of the system being in state *student* over the interval $[3.0, 7.0)$ from 0.0 to 1.0 in increments of 0.2 and divide the remaining probability uniformly between *employed* and *unemployed* (the likelihood of 0.0 with uniformly dividing the likelihood among the remaining states demonstrates negative interval evidence). The bottom-most curve of Fig. 6a shows when *student* can be ruled out with complete certainty. The top-most curve of Fig. 6a shows when *student* is observed with complete certainty. As expected, the uncertain positive observation of Employed affects the probability of the states of other nodes as well. Fig. 6b shows the evolving probabilities of state *single* given the same varying uncertain positive observations to Employed. The bottom-most curve of Fig. 6b shows the effect on *single* when *student* can be ruled out with complete certainty, while the top-most curve shows the effect on *single* when *student* is observed with complete certainty.

4.1.3. Exact inference with uncertain negative interval evidence

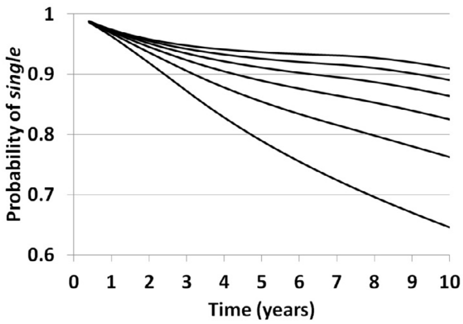
Uncertain negative interval must be handled differently, because we need to keep track of which states have been visited on the interval and weight the first-time transitions into the states according to the evidence. To do this, we create an augmented state matrix $\mathbf{Q}'_i$ representing the original states combined with flags for which states in the uncertain negative evidence have been visited so far on the interval.

Formally, let E be the set of indices for the uncertain negative evidence states over an interval of time for node X. Let V be a set of Boolean values defined over the indices in E, $\{v_j\}_{j\in E}$. A value for v_j of *true* means that state x_j has been visited, while value for v_j of *false* means that state x_j has not yet been visited. For each state $x_j \in X$, we create new states $x_{\langle j,V\rangle}$ for each possible combination of true/false assignments to the Boolean values in V (except for $\neg v_j$ when $j \in E$, because if the state is x_j then v_j must be *true*). Let $v_j \in V$ and $v'_k \in V'$. The entries of $\mathbf{Q}'_i$ are populated as follows:

$$q'_{\langle k,V\rangle,\langle k,V'\rangle} = \begin{cases} q_{j,j} & \text{if } j=k \wedge V=V' \\ q_{j,k} & \text{if } j\neq k \wedge V=V' \wedge (k\notin E \vee v'_k) \\ q_{j,k}\cdot(1-\lambda_k) & \text{if } j\neq k \wedge \left(\forall_{l\neq k} v_l = v'_l\right) \wedge \neg v_k \wedge v'_k \\ 0 & \text{otherwise} \end{cases}.$$



(a) Probability of *student* through time.



(b) Probability of *single* through time.

Fig. 6. Uncertain positive interval evidence.

The evidence does not change the sojourn rates, so the $q_{j,j}$ are copied over for each state as shown in the first case. The transition rates $q_{j,k}$ into state x_k are unchanged if the state is not part of the evidence or if the state has already been visited as shown by the second case. If the state is transitioning into state x_k for the first time, and if x_k is part of the evidence, then the transition is weighted according to the negative evidence for that state as shown by the third case. Note that the transition is forced into a state in which visited flag v_k is changed from *false* to *true* while all other flags remain the same (because the node cannot simultaneously transition into multiple states, changing multiple flags). All other transitions are set to zero as shown by the fourth case. These entries represent impossible transitions, such as a visited state transitioning to the non-visited state or states having multiple different flags that would imply simultaneous transitions.

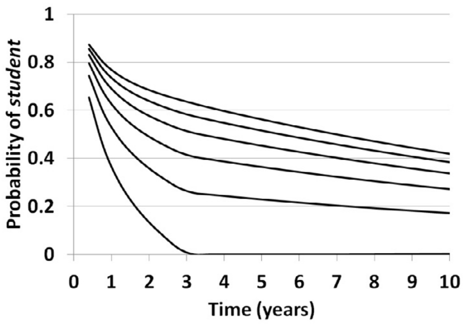
The matrix $\mathbf{Q}_{i-1,i}$ is expanded into $\mathbf{Q}'_{i-1,i}$ with the same states as $\mathbf{Q}'_i$. The diagonal entries of $\mathbf{Q}'_{i-1,i}$ are also weighted according to the uncertain evidence:

$$q'_{\langle j,V\rangle,\langle j,V'\rangle} = \begin{cases} 1 & \text{if } j \notin E \\ 1-\lambda_j & \text{if } j \in E \wedge \left(\forall_{k\neq j}\neg v_k\right) \wedge \left(\forall_{k\neq j}\neg v'_k\right) \wedge v_j \wedge v'_j \\ 0 & \text{otherwise} \end{cases}.$$

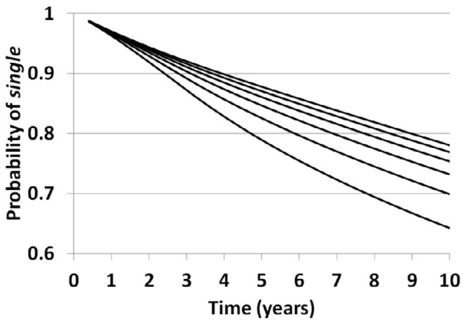
The vector $\boldsymbol{\alpha}_{t_i}$ is also expanded to match the dimensions with the new $\mathbf{Q}'_{i-1,i}$. For this, entries with state x_j map to states $\langle x_j, V\rangle$ such that $v_j \in V$ (if $j \in E$) and $\forall_{j\neq k,k\in E}\neg v_k \in V$. In other words, the flag that x_j has been visited is set to *true*, while all other states in the evidence are set to *false*. The probabilities for the interval of evidence are calculated by taking the matrix exponential of $\mathbf{Q}'_i$ instead of $\mathbf{Q}_i$. The state probabilities for each x_j are recovered by combining the probabilities across all of the flag settings $\langle x_j, V\rangle$. This approach is using to find probabilities of states within the interval of evidence as well as to recover $\boldsymbol{\beta}_{t_{i+1}}$ with compatible dimensions as the original matrix $\mathbf{Q}_{i+1,i}$.

Note that X may have parents and/or children in the network. When X has parents, the process is used to augment the states for each of the conditional intensity matrices of X. When X has children, each of the children's conditional intensity matrices remain conditionally dependent on each of the x_j's only and not on any of the values in V.

To demonstrate uncertain negative evidence, we vary the probability of state *student* being ruled out over the interval $[3.0, 7.0)$, from 0.0 to 1.0 in increments of 0.2. The top-most curve of Fig. 7a shows when *student* cannot be ruled out with any certainty, which is the same as no observation. The bottom-most curve of Fig. 7a shows when *student* can be ruled out with complete certainty. Hence, the probability of *student* is exactly 0.0 for the entire interval $[3.0, 7.0)$. As above, the uncertain negative observation of Employed affects the other nodes. Fig. 7b shows the evolving probabilities of state *single* given the same varying uncertain negative observations to Employed. The top-most curve of Fig. 7b shows the effect on *single* when *student* cannot be ruled out with any certainty, while the bottom-most curve shows the effect on *single* when *student* can be ruled out with complete certainty.



(a) Probability of *student* through time.



(b) Probability of *single* through time.

Fig. 7. Uncertain negative interval evidence.

As already discussed, uncertain and negative evidence become distinct types of evidence when the evidence is temporal. The difference between them becomes apparent when comparing Figs. 6 and 7. Uncertain positive evidence observes the state to be constant over the interval, shown by the constant probability of Fig. 6a on the interval $[3.0, 7.0)$. However, the uncertain negative evidence rules out states with some probability, but does not rule out state transitions. Therefore, as time progresses, the state probability might also vary, even if it has a non-zero probability of being ruled out.

4.1.4. *Exact inference with uncertain and negative point evidence*

For point evidence, we modify only the $\boldsymbol{Q}_{i-1,i}$ matrix where t_i is the time of the point evidence, but we modify it the same way as uncertain positive interval evidence or certain negative interval evidence. Specifically, Equation (31) is used for uncertain point evidence, and Equation (34) is used for negative point evidence.

To demonstrate the effects of uncertain and negative point evidence, we vary the likelihood of the system being in state *student* at $t = 5.0$ from 0.0 to 1.0 in increments of 0.2 and divide the remaining probability uniformly between *employed* and *unemployed* (the likelihood of 0.0 with uniformly dividing the likelihood among the remaining states demonstrates negative point evidence). The bottom-most curve of Fig. 8a shows when *student* can be ruled out with complete certainty. The top-most curve of Fig. 8a shows when *student* is observed with complete certainty. As before, the uncertain and negative observations of Employed affect the probability of the states of other nodes as well. Fig. 8b shows the evolving probabilities of state *single* given the same varying uncertain positive observations to Employed. The bottom-most curve of Fig. 8b shows the effect on *single* when *student* can be ruled out with complete certainty, while the top-most curve shows the effect on *single* when *student* is observed with complete certainty. The curves in Fig. 8b are understandably similar (but not identical) to Fig. 6b because the duration of the evidence is different (a point versus an interval).

4.1.5. *Exact inference with uncertain and negative transition evidence*

For transition evidence, we again modify only the $\boldsymbol{Q}_{i-1,i}$ matrix where t_i is the time of the transition. For uncertain transitions, we have likelihoods $\lambda_{j,k}$ for transitioning from x_j to x_k. For $\lambda_{j,j} > 0$, this implies a non-zero probability that no transition occurred if the state was in x_j. Negative transition evidence is accomplished by using $\lambda_{j,k} = 0$. For each $l \in (x_j)$ and for each $m \in I(x_k)$:

$$\boldsymbol{Q}_{i-1,i}[l,m] \leftarrow \lambda_{i,j} \cdot \boldsymbol{Q}_{i-1,i}[l,m] \tag{35}$$

Fig. 9 shows the state probabilities of Employed when an uncertain transition is observed at time $t = 5.0$. Fig. 9a shows the probabilities when the state is known to transition *from* state *student*, but the destination state is unknown (the destination state likelihoods are uniform). Fig. 9b shows the probabilities when the state is known to transition *to* state *employed*, but the source state is unknown (the source state likelihoods are uniform).

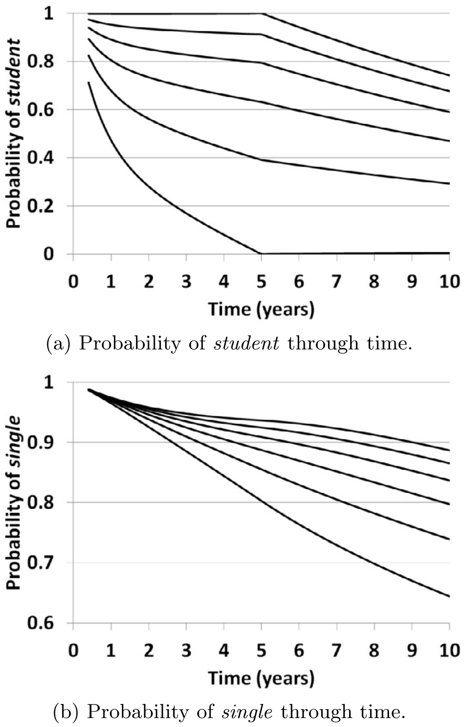


Fig. 8. Uncertain and negative point evidence.

The transition probabilities between the states of Employed are determined by the current state of its parent. Thus, Fig. 9a takes into account the probable state of the parent at time $t = 5.0$, in addition to Employed's own transition probabilities. Fig. 9b shows that, reasoning over the probable state of H_E at time $t = 5.0$, a transition to *employed* most likely originated from *student*.

4.2. Importance sampling for uncertain and negative interval evidence

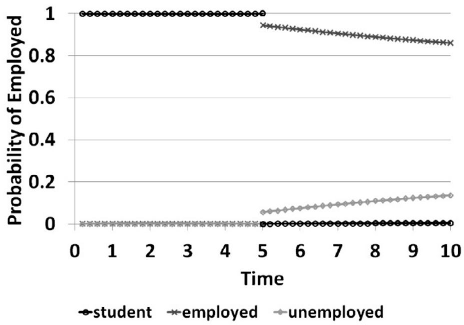
The size of the matrices in the forward–backward algorithm are exponential in the number of nodes in the CTBN. For uncertain and negative evidence to be useful, we need to extend existing approximation algorithms to be able to handle these new types of evidence.

In this paper, we extend the importance sampling algorithm [6]. The algorithm generates a set of weighted samples that conform to a partial sample path $\mathbf{e}$ taken as evidence. The algorithm samples a proposal distribution P' that conforms to the evidence to fill in the unobserved intervals, generating a complete sample path. Because the samples are drawn from P' to force each sample to be consistent with the evidence, each complete sample path σ is weighted by the likelihood of the evidence, calculated as $w(\sigma) = \frac{P(\sigma,\mathbf{e})}{P'(\sigma)}$, with the cumulative weight as $W = \sum_{\sigma \in \mathcal{S}} w(\sigma)$. After generating a set of i.i.d. samples $\mathcal{S}$, the algorithm approximates the conditional expectation of any function f given the evidence $\mathbf{e}$ as:

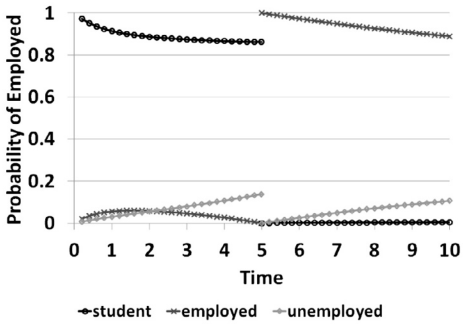
$$\hat{E}(f|\mathbf{e}) = \frac{1}{W} \sum_{\sigma \in \mathcal{S}} w(\sigma) f(\sigma) \tag{36}$$

While the extended importance sampling algorithm is similar in structure to the original algorithm of [6], the introduction of uncertain and negative evidence requires several modifications that must be made throughout the entire algorithm. Because of these differences, we present the extended importance sampling algorithm in full. The pseudocode for the main loop is given in Algorithm 1. This algorithm calls several helper methods, given in Algorithms 2 through 5. Asterisks mark substantial differences in the pseudocode from the original importance sampling algorithm that must be introduced to handle uncertain and negative evidence. The algorithm uses the following notation:

- t is the current time of the sample.
- σ is the sample path, consisting of a sequence of timestamp/state pairs.
- w is the likelihood (weight) of the sample.
- $\mathbf{e}'$ is a set of certain and/or uncertain observations, while $\mathbf{e}$ is a set containing only certain observations.
- $\langle X, E, [t_s, t_e), type\rangle$ is a certain observation and can be an element of either $\mathbf{e}'$ or $\mathbf{e}$. The variable *type* has a value of either *pos* or *neg* for positive and negative evidence, respectively. X identifies the node being observed, and E is the



(a) State probabilities of Employed through time.



(b) State probabilities of Employed through time.

Fig. 9. Uncertain transitions *to* and *from* a given state.

observed state (for positive evidence) or the set of states being ruled out (for negative evidence). The interval $[t_s, t_e)$ is the time during which this evidence holds for node X.

- $\langle X, \theta, [t_s, t_e), type\rangle$ is an uncertain observation and can be an element of $\mathbf{e}'$. The meaning of the variables match those above but with θ being a vector of size $|X|$. When *type* is *pos*, θ is a vector of likelihoods for observing each state on the interval. When *type* is *neg*, θ is a vector with values on $[0, 1]$, and each value gives the relative likelihood with which each state can be ruled out on the interval.
- $\mathbf{e}_X^{val}(t)$ is the value of X at time t according to the evidence or *null* if X has no evidence at time t. In the case of negative evidence, $\mathbf{e}_X^{val}(t)$ could be a set of values.
- $\mathbf{e}_X^{type}(t)$ is the type of evidence for X at time t, with values *pos* and *neg* for positive and negative evidence, respectively.
- $\theta^{\mathcal{B}}_{X|\mathbf{pa}_{\mathcal{B}}(X)}$ is the prior probability distribution of X given the parents of X in $\mathcal{B}$.
- $Time[X]$ is the proposed transition time of node X.
- $\mathbf{e}_X^{time}(t)$ is the first time after t when $\mathbf{e}_X^{val}(t)$ is defined.
- $\mathbf{e}_X^{end}(t)$ is the first time after or equal to t when $\mathbf{e}_X^{val}(t)$ changes value or becomes *null*.
- $q_{X(t)|\mathbf{u}_X(t)}$ is the exponential parameter of node X in state $X(t)$ given the parent states of X at time t.
- $\theta_{X(t)|\mathbf{u}_X(t)}$ is the transition probabilities out of state $X(t)$ given the parents' state of X at time t.
- $\mathbf{e}_X^{conf}(X(t), t)$ is the soonest time at which the current state conflicts with upcoming evidence, either positive or negative, or *null* if there is no future evidence or the current state matches the soonest positive evidence.

The method $Constrain(\theta, \mathbf{e})$ takes the probabilities for a multinomial distribution θ, zeroes out the states in θ listed in $\mathbf{e}$, and then re-normalizes θ. The method $Append(\sigma, \langle X, t\rangle)$ adds a transition, given as a state X at time t to the end of the sample path σ.

The main method *CTBN-Importance-Sample*$(\mathcal{N}, \mathbf{e}', t_{end})$ generates a single, weighted sample path σ that conforms to the evidence $\mathbf{e}'$ and is weighted according to the likelihood of that evidence. Line 1 initializes the current time t to 0, while line 2 initializes the set of proposed transition times. Line 3 samples any uncertain evidence in $\mathbf{e}'$. During the generation of this sample, the sampled evidence in $\mathbf{e}$ is treated as certain evidence. Line 4 generates the initial states for all nodes according to the prior distribution β while conforming to any evidence in $\mathbf{e}$ defined at $t = 0$. The weight of the sample is updated according to the likelihood of this evidence. Lines 5–36 continue to generate transitions until the duration of the sample path is at least t_{end}. Line 6 ensures all nodes have potential transition times. Line 7 gets the node with the soonest potential transition time. In lines 8–37, if that time is later than t_{end} the sample path is finished. The weight is updated through the last segment up until t_{end}, and the sample path and its likelihood are returned. Otherwise, lines 11–13 update the weight over the current segment. Line 14 updates the current time to the end of the segment. A transition may

Algorithm 1 *CTBN-Importance-Sample*($\mathcal{N}$, $\mathbf{e}'$, t_{end}).

1: $t \leftarrow 0$
2: $Time \leftarrow null$
3: $\mathbf{e} \leftarrow \textit{Sample-Evidence}(\mathbf{e}')$ *
4: $\langle \mathbf{X}, \sigma, w \rangle \leftarrow \textit{Sample-Initial-States}(\mathcal{N}, \mathbf{e})$
5: loop until termination
6: $Time \leftarrow \textit{Sample-Transition-Times}(Time, \mathcal{N}, \mathbf{e})$
7: $X \leftarrow \arg\min_{X \in \mathbf{X}}[Time[X]]$
8: if $Time[X] \geq t_{end}$
9: $w \leftarrow \textit{Update-Weight}(X, w, t, t_{end}, \mathcal{N}, \mathbf{e})$
10: break
11: else
12: $w \leftarrow \textit{Update-Weight}(X, w, t, Time[X], \mathcal{N}, \mathbf{e})$
13: end if
14: $t \leftarrow Time[X]$
15: if $\mathbf{e}_X^{end}(t) = t \wedge (\mathbf{e}_X^{val}(t) = null \vee X(t) = \mathbf{e}_X^{val}(t) \vee \mathbf{e}_X^{type}(t) = neg)$ *
16: $Time[X] \leftarrow null$
17: else
18: if $\mathbf{e}_X^{type}(t) = pos \wedge \mathbf{e}_X^{val}(t) \neq null \wedge X(t) \neq \mathbf{e}_X^{val}(t) \wedge \mathbf{e}_X^{end}(t) - t < \epsilon$ *
19: $w \leftarrow w \cdot \theta_{X(t)|\mathbf{u}_X(t)}[\mathbf{e}_X^{val}(t)]$
20: $X(t) \leftarrow e_X^{val}(t)$
21: else
22: $\theta \leftarrow \theta_{X(t)|\mathbf{u}_X(t)}$
23: if $(\mathbf{e}_X^{type}(t) = neg)$ *
24: $w \leftarrow w \cdot (1 - \sum_{e \in \mathbf{e}_X^{val}(t)} \theta[e])$ *
25: $\textit{Constrain}(\theta, \mathbf{e}_X^{val}(t))$ *
26: end if *
27: $X(t) \sim \text{Multinomial}(\theta)$
28: end if
29: $X \leftarrow X(t)$
30: $\textit{Append}(\sigma, \langle X, t \rangle)$
31: $Time[X] \leftarrow null$
32: for each Y for which $X \in \mathbf{Pa}(Y)$
33: $Time[Y] \leftarrow null$
34: end for
35: end if
36: end loop
37: return $\langle \sigma, w \rangle$

Algorithm 2 *Sample-Evidence*($\mathbf{e}'$).

$\mathbf{e} \leftarrow \emptyset$
for each $\langle X, \theta, [t_s, t_e), type \rangle \in \mathbf{e}'$
if $type = pos$
$x \sim \text{Multinomial}(\theta)$
$\mathbf{e} \leftarrow \mathbf{e} \cup \langle X, x, [t_s, t_e), pos \rangle$
else
do
$E \leftarrow \emptyset$
for $x \in X$
if $\text{Uniform}(0, 1) < \theta[x]$
$E \leftarrow E \cup \{x\}$
end if
end for
while $E = X$
$\mathbf{e} \leftarrow \mathbf{e} \cup \langle X, E, [t_s, t_e), neg \rangle$
end if
end for
return $\mathbf{e}$

not always occur at the end of the segment (hence the term "potential transition time"). A transition will not occur if the end of the segment falls on a change in evidence, such that the current time falls on the end or beginning of a segment of positive evidence for this node or the evidence is negative. In this case, lines 15–16 reset the potential transition time for this node and the process returns to line 5. Otherwise, a transition will occur. Lines 18–20 handle the case when the evidence specifies the transition. This occurs when there is positive evidence at the current time (or within ϵ of the current time) and the current state of the node does not match the evidence. Thus, to conform to the evidence, a transition is forced in line 20. Line 19 updates the weight with the likelihood of that transition. Otherwise, the node can transition to multiple states and the destination state must be sampled. If there is negative evidence, lines 23–26 zero out the transition probabilities for these states and update the weight with the likelihood that the node did *not* transition to these states. In line 29, the state of the node is updated. Because a transition has occurred, line 30 adds the transition to the sample path.

Algorithm 3 *Sample-Initial-States*($\mathcal{N}$, $\mathbf{e}$).

1: $\sigma \leftarrow null$, $w \leftarrow 1$
2: for each variable $X \in \mathbf{X}$
3:     if $\mathbf{e}_X^{val}(0) \neq null$
4:         if $\mathbf{e}_X^{type}(0) = pos$ *
5:             $X(0) \leftarrow \mathbf{e}_X^{val}(0)$
6:             $w \leftarrow w \cdot \theta^{\mathcal{B}}_{X(0)|\mathbf{pa}_{\mathcal{B}}(0)}$
7:         else *
8:             $\theta^{\mathcal{B}} \leftarrow \theta^{\mathcal{B}}_{X(0)|\mathbf{pa}_{\mathcal{B}}}$ *
9:             *Constrain*$(\theta^{\mathcal{B}}, \mathbf{e}_X^{val}(0))$ *
10:             $X(0) \sim$ Multinomial$(\theta^{\mathcal{B}})$ *
11:             $w \leftarrow w \cdot (1 - \sum_{e \in \mathbf{e}_X^{val}(0)} \theta^{\beta}[e])$ *
12:         end if *
13:     else
14:         $X(0) \sim$ Multinomial$(\theta^{\mathcal{B}}_{X|\mathbf{Pa}_{\mathcal{B}}(X)})$
15:     end if
16:     $X \leftarrow X(0)$
17:     *Append*$(\sigma, \langle X, 0 \rangle)$
18: end for
19: return $\langle \mathbf{X}, \sigma, w \rangle$

Furthermore, lines 31–34 reset the potential transition times for the node and all of its children, as the change of state in the parent changes the current intensity matrix of each child. The process returns to line 5 to generate new potential transition times.

The helper method *Sample-Evidence*($\mathbf{e}$) given by Algorithm 2 handles uncertainty in the evidence, whether positive or negative. The states of any uncertain evidence are re-sampled before the generation of each sample. This applies to both uncertain positive and uncertain negative evidence. The method also checks to make sure the evidence sampled is feasible. For example, uncertain negative evidence must not rule out every state. Thus, for each sample generated by the sampling algorithm, all of the evidence can be treated as certain. However, the certain evidence could change between samples, according to the uncertainty of the evidence, and the weights of the final set of samples will reflect this in the given query.

The helper method *Sample-Initial-States*($\mathcal{N}$, $\mathbf{e}$) is given in Algorithm 3. The method is responsible for sampling the initial states of the sample path while conforming to the evidence. Line 1 creates an initially empty sample path σ and initializes the weight w. Lines 2–18 loop over all nodes in $\mathcal{N}$. Lines 3–12 handle the case when the node has evidence set at $t = 0$. If the evidence is positive, lines 4–6 set the node to that state and update the weight with the likelihood of that evidence. If the evidence is negative, lines 7–12 zero out the transition probabilities for these states and update the weight with the likelihood that the node did *not* transition to these states. If no evidence is specified for this node at $t = 0$, line 14 samples from the prior distribution, and no weighting is necessary. Line 16 sets the initial state of the node, and line 17 adds the initial states to the sample path. The current states, the current sample path, and the current weight are returned in line 19.

The helper method *Sample-Transition-Times*(*Time*, $\mathcal{N}$, $\mathbf{e}$) is given in Algorithm 4. The method is responsible for generating proposed transition times that conform to the evidence. These are only proposed transition times, and transitions are not guaranteed to occur at these times. For example, whenever a parent node transitions, the children's proposed transition times will be re-sampled to account for their new conditional intensity matrix and the proposed transition times will change. If a node is currently within an interval of positive evidence or has upcoming positive evidence, the proposed transition times will be the start or end of the interval, respectively. However, a transition will not occur, because the state must be kept constant during the interval of positive evidence. Lines 1–19 loop over all nodes in $\mathcal{N}$, while line 2 checks whether the current node needs a new proposed transition time. Line 3 checks whether the node is currently within positive evidence. If so, line 4 sets the node's proposed transition time as the end of the evidence. This does not mean that the node will transition immediately after the interval of positive evidence, but will have its proposed transition time sampled again once it becomes unobserved. If the node is not currently observed, then line 6 gets the soonest time (if it exists) at which the current state of the node conflicts with upcoming evidence. Line 7 checks whether the current state conflicts with upcoming evidence (if the time of the soonest conflict is set). This could be positive evidence (the current state will need to transition *to* the observed state at some point) or negative evidence (the current state will need to transition *away* from the set of states that are ruled out). In either case, the proposed transition time must be sampled from a truncated exponential distribution, shown in line 8, to condition on the upcoming evidence. Otherwise, line 10 simply samples from an exponential distribution. While the current state could be conforming to upcoming evidence, the sampled transition time could be past the end of the upcoming evidence. Thus, line 11 gets the time of the next change in the evidence, and lines 12–14 make sure the sampled transition time does not exceed this time. Line 17 sets the proposed transition time for this node, and the set of proposed transition times for all nodes are returned in line 20.

The helper method *Update-Weight*(Y, w, t_1, t_2) is given in Algorithm 5. The method is responsible for weighting the likelihood of the transition times, whether the state was observed over an interval or the transition time was sampled from a truncated exponential to conform to upcoming evidence. Lines 1–14 loop over all nodes in $\mathcal{N}$. Line 2 gets the time of the next change in observation for this node. Line 3 checks whether the state of the node is currently known but will become unobserved before transitioning to another state. If this is the case, line 4 updates the weight by the likelihood that the

Algorithm 4 *Sample-Transition-Times*$(Time, \mathcal{N}, \mathbf{e})$.

for each $X \in \mathbf{X}$
if $Time[X] = null$
if $\mathbf{e}_X^{val}(t) \neq null \wedge \mathbf{e}_X^{type}(t) = pos$ *
$\Delta t \leftarrow \mathbf{e}_X^{end}(t) - t$
else
$t_{conf} \leftarrow \mathbf{e}_X^{conf}(X(t), t)$ *
if $t_{conf} \neq null$ *
$\Delta t \sim$ Exponential$(q_{X(t)|\mathbf{u}_X(t)})$ given $\Delta t < (t_{conf} - t)$ *
else
$\Delta t \sim$ Exponential$(q_{X(t)|\mathbf{u}_X(t)})$
$t_e = \mathbf{e}_X^{time}(t)$
if $X(t) = \mathbf{e}_X^{val}(t_e) \wedge t + \Delta t > t_e$
$\Delta t \leftarrow t_e - t$
end if
end if
end if
$Time[X] \leftarrow t + \Delta t$
end if
end for
return *Time*

Algorithm 5 *Update-Weight*$(Y, w, t_1, t_2, \mathcal{N}, \mathbf{e})$.

1: for each $X \in \mathbf{X}$
2: $t_e \leftarrow \mathbf{e}_X^{end}(t_1)$
3: if $\mathbf{e}_X^{val}(t_1) \neq null \wedge \mathbf{e}_X^{type}(t_1) = pos \wedge (\mathbf{e}_X^{val}(t_e) = null \vee \mathbf{e}_X^{type}(t_e) = neg)$ *
4: $w \leftarrow w \cdot \exp(-q_{X(t_1)|\mathbf{u}_X(t_1)}(t_2 - t_1))$
5: else
6: $t_{conf} \leftarrow \mathbf{e}_X^{conf}(X(t), t)$ *
7: if $t_{conf} \neq null$ *
8: if $X = Y$
9: $w \leftarrow w \cdot (1 - \exp(-q_{X(t_1)|\mathbf{u}_X(t_1)}(t_e - t_1)))$
10: else
11: $w \leftarrow w \cdot \frac{1-\exp(-q_{X(t_1)|\mathbf{u}_X(t_1)}(t_e - t_1))}{1-\exp(-q_{X(t_1)|\mathbf{u}_X(t_1)}(t_e - t_2))}$
12: end if
13: end if
14: end for
15: return w

node remained in the state for at least the interval observed. Otherwise, if the state was unknown, the transition time might have been sampled from a truncated exponential distribution. Line 6 gets the time at which the current state of this node conflicts with upcoming evidence, if it exists. Line 7 checks whether this time is set, i.e., whether the most recent proposed transition time for this node was sampled from a truncated exponential distribution. If the current node was the node with the soonest proposed transition time, the weight is updated with the likelihood of sampling the transition time from the truncated exponential, line 9. Otherwise, the proposed transition time must be later and the weight is updated according to line 11. The weight for this segment over all of the variables is finally returned in line 15.

The algorithm is extended to support uncertain and negative interval evidence. As mentioned, combinations of infinitesimal interval evidence can approximate uncertain and negative point and transition evidence. While the algorithm could be modified to handled points and transitions exactly, this introduces additional conditional branching for switching between points, transitions, and intervals. We also note that the algorithm could be extended further by incorporating uncertainty in the evidence into the generation of the sample through predictive look-ahead. This would address situations in which sampled combinations of uncertain evidence are highly unlikely given the network parameters and a high proportion of generated samples have relatively low weights. However, an efficient and general predictive look-ahead algorithm is unlikely to exist as, depending on the network parameters, future evidence for any ancestors and/or descendants could force an arbitrarily low weight, and the algorithm would have to check all combinations of uncertain evidence and compare each with the parameters of all ancestors/descendants to compute the likelihood of that evidence. Nevertheless, an extension for limited, local, and/or heuristic predictive look-ahead may be useful in certain cases but is left as future work.

5. Experiments

We demonstrate the extensions to the importance sampling algorithm by comparing the results with the exact inference evidence demonstrations from Section 4.1. We vary the number of samples from 100 to 1 000 000 and calculate the average KL divergence for all nodes and query times on the interval. Fig. 10 plots both the number of samples and the average KL divergence on a log scale to demonstrate convergence of importance sampling to the exact algorithm.

We also want to test how the evidence scales as more evidence is applied. We construct a ring network consisting of n three-state (s_0, s_1, s_2) nodes connected as follows:

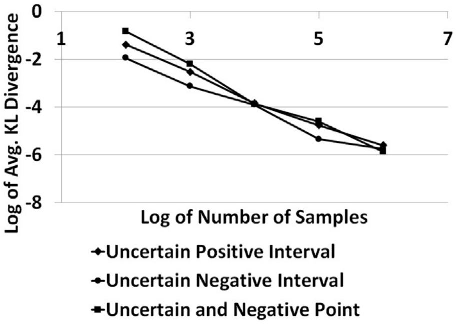


Fig. 10. Log of average KL divergence versus log of number of samples for demonstration evidence on BHPS network.

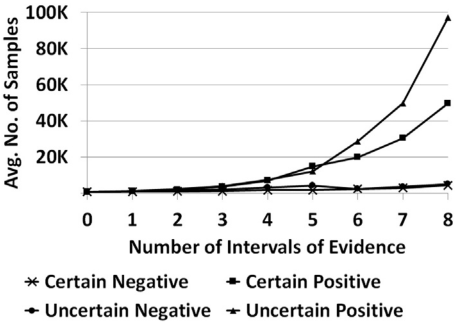


Fig. 11. Number of interval evidence versus average number of samples to reach 0.001 KL divergence across 100 trials on the ring network.

$$X_1 \to X_2 \to \cdots \to X_n \to X_1.$$

Let each $r^k_{i,j}$ be an independent sample from a uniform distribution over the interval $(0.5, 1.5)$. The conditional intensity matrices for the nodes are defined as follows (for ease of definition, X_0 and X_n denote the same node):

$$\mathbf{Q}_{X_k|X_{k-1}=s_0} = \begin{pmatrix} -r^k_{1,1} & \frac{1}{2}r^k_{1,1} & \frac{1}{2}r^k_{1,1} \\ \frac{2}{3}r^k_{1,2} & -r^k_{1,2} & \frac{1}{3}r^k_{1,2} \\ \frac{2}{3}r^k_{1,3} & \frac{1}{3}r^k_{1,3} & -r^k_{1,3} \end{pmatrix},$$

$$\mathbf{Q}_{X_k|X_{k-1}=s_1} = \begin{pmatrix} -r^k_{2,1} & \frac{2}{3}r^k_{2,1} & \frac{1}{3}r^k_{2,1} \\ \frac{1}{2}r^k_{2,2} & -r^k_{2,2} & \frac{1}{2}r^k_{2,2} \\ \frac{1}{3}r^k_{2,3} & \frac{2}{3}r^k_{2,3} & -r^k_{2,3} \end{pmatrix},$$

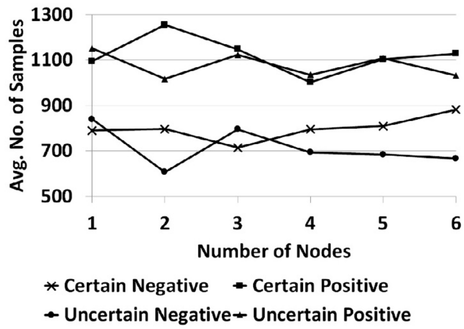
$$\mathbf{Q}_{X_k|X_{k-1}=s_2} = \begin{pmatrix} -r^k_{3,1} & \frac{1}{3}r^k_{3,1} & \frac{2}{3}r^k_{3,1} \\ \frac{1}{3}r^k_{3,2} & -r^k_{3,2} & \frac{2}{3}r^k_{3,2} \\ \frac{1}{2}r^k_{3,3} & \frac{1}{2}r^k_{3,3} & -r^k_{3,3} \end{pmatrix}.$$

Each node is given a uniform prior distribution.

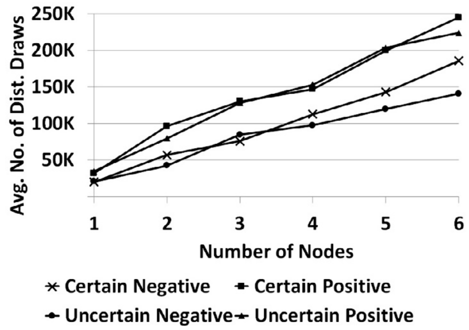
For the first scaling experiment, we test importance sampling with an increasing amount of evidence on a ring of size $n = 2$. We generate evidence uniformly along the interval $(0.0, 10.0)$. For each instance of evidence, we choose a random node and interval from a $(0.0, 0.5)$ uniform distribution. For certain evidence, we choose a random state for the chosen node. For uncertain evidence, we create an uncertain evidence vector in which each entry is drawn from a $(0.0, 1.0)$ uniform distribution. For the uncertain positive interval evidence, we normalize this vector to sum to 1.0.

For each type of evidence and count of evidence, we run 100 independent trials. Each trial is a newly sampled ring network and newly sampled set of evidence. For each trial, we sample a random time at which to query the probability of a random node. We compute the KL divergence between the importance sampling algorithm and the exact algorithm for this query and average the KL divergence across all trials. We continue to generate samples until the KL divergence is less than 0.001.

Fig. 11 shows the results for $n = 2$. The plot shows that negative evidence (both certain and uncertain) is easier to incorporate than tradition certain positive evidence. Negative evidence rules out one or more states, which can be less restrictive than certain positive evidence, which forces one particular state. Thus the samples with negative evidence are



(a) Average number of samples.



(b) Average number of multinomial and exponential draws.

Fig. 12. Size of network versus number of samples and sample complexity to reach 0.001 KL divergence across 100 trials with one interval of evidence.

likely to have a higher weight than samples with certain positive evidence, and fewer samples are needed to reach the target KL divergence. The opposite holds true between uncertain positive evidence and certain positive evidence, in which uncertain positive evidence is more difficult to sample and will likely have a lower weight. Thus more samples are usually needed for uncertain positive evidence to reach the same level of accuracy as those with certain positive evidence.

For the second scaling experiment, we test importance sampling with an increasing network size from $n = 1$ to $n = 6$. We generate one interval of evidence starting at $t = 5.0$. We choose a random node and interval from a $(0.0, 0.5)$ uniform distribution. For certain evidence, we choose a random state for the chosen node. For the uncertain positive interval evidence, we create an uncertain evidence vector in which each entry is drawn from a $(0.0, 1.0)$ uniform distribution, and then we normalize this vector to sum to 1.0. For uncertain negative interval evidence, we choose a random state and random evidence for this state from a $(0.0, 1.0)$ uniform distribution. We continue to generate samples until the KL divergence is less than 0.001.

Fig. 12a shows that the number of samples required to converge to the target KL divergence was roughly constant across network sizes. Nevertheless, Fig. 12b shows that the complexity of generating a single sample is directly proportional to the number of times that the algorithm has to sample from a multinomial or exponential distribution. In this experiment in which the amount of evidence is fixed, the complexity of generating the samples is shown to scale linearly with the size of the network.

6. Conclusion and future work

Continuous-time systems allow for much greater variety in the types of evidence (as opposed to DBNs, for example) that would be useful to know about the system. For CTBNs, evidence can be over an interval or at a single point, corresponding to the system being observed in a particular state. The types of continuous-time evidence can be generalized further to include uncertain and negative evidence. Uncertain and negative evidence in continuous-time systems represents a novel and useful generalization of evidence that makes the models more applicable and versatile.

We presented the first definitions for uncertain and negative variations with point, transition, and interval evidence in CTBNs and showed the relationships between these evidence types. We showed how to extend the forward–backward algorithm for CTBNs and the CTBN importance sampling algorithm to support all of these evidence types. As discussed in Section 2.5, there are several other CTBN inference algorithms that have been developed, each with their own strengths and weaknesses. Thus, to give the greatest flexibility to CTBN modelers who have evidence that is uncertain or negative (or both), it would be useful to extend these algorithms to support uncertain and negative evidence as well.

Acknowledgements

This work was funded in part by Phase II of NAVY STTR N10A-009-0292 and by Phase I of NASA STTR T13.01-9887. The authors would also like to thank members of the Numerical Intelligent Systems Laboratory at Montana State University for their helpful discussion throughout this work.

Appendix A. Proof sketches of relationships between types of evidence

Proposition 1. *Certain interval evidence is a special case of uncertain positive interval evidence.*

Proof. Let state x_i be the observed state. Set

$$\lambda_j = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}. \qquad \square$$

Proposition 2. *Negative interval evidence is a special case of uncertain negative interval evidence.*

Proof (*Sketch*). Set

$$\lambda_i = \begin{cases} 1 & \text{if } x_i \in X' \\ 0 & \text{otherwise} \end{cases}. \qquad \square$$

Proposition 3. *Certain interval evidence is a special case of negative interval evidence.*

Proof (*Sketch*). Let state x_i be the observed state. Set X' as all states of X except x_i. $\square$

Proposition 4. *Uncertain positive interval evidence approximates uncertain point evidence.*

Proof (*Sketch*). Set $t_2 = t_1 + \epsilon$ for infinitesimal value ϵ. $\square$

Proposition 5. *Uncertain negative interval evidence approximates uncertain point evidence.*

Proof (*Sketch*). Set $t_2 = t_1 + \epsilon$ for infinitesimal value ϵ. $\square$

Proposition 6. *Negative interval evidence approximates negative point evidence.*

Proof (*Sketch*). Set $t_2 = t_1 + \epsilon$ for infinitesimal value ϵ. $\square$

Proposition 7. *Certain interval evidence approximates certain point evidence.*

Proof (*Sketch*). Set $t_2 = t_1 + \epsilon$ for infinitesimal value ϵ. $\square$

Proposition 8. *Negative point evidence is a special case of uncertain point evidence.*

Proof (*Sketch*). Set

$$\lambda_i = \begin{cases} 0 & \text{if } x_i \in X' \\ \frac{1}{|X|-|X'|} & \text{otherwise} \end{cases}. \qquad \square$$

Proposition 9. *Certain point evidence is a special case of negative point evidence.*

Proof (*Sketch*). Let x_i be the observed state. Set X' as all states of X except x_i. $\square$

Proposition 10. *Two instances of certain point evidence represent temporally uncertain transition evidence.*

Proof (*Sketch*). Let distinct states x_i and x_j be observed by the two instances of certain point evidence. Set t_1 and t_2 as the times of the two instances of certain point evidence. $\square$

Proposition 11. *Two instances of uncertain point evidence approximate uncertain transition evidence.*

Proof (*Sketch*). Let t_1 and t_2 be the times of the two instances of uncertain point evidence. Set $t_1 = t_2 - \epsilon$ for infinitesimal value ϵ. □

Proposition 12. *Certain transition evidence is a special case of temporally uncertain transition evidence.*

Proof (*Sketch*). Set $t_1 = t_2$. □

Proposition 13. *Negative transition evidence is a special case of uncertain transition evidence.*

Proof (*Sketch*). Set

$$\lambda_{i,j} = \begin{cases} 0 & \text{if } x_i \in X_S \vee x_j \in X_E \\ \frac{1}{|X|^2 - |X_E| - |X_S|} & \text{otherwise} \end{cases}. \qquad \square$$

Proposition 14. *Certain transition evidence is a special case of negative transition evidence.*

Proof (*Sketch*). Let state x_S be the observed source state and x_E be the observed destination state. Set X_S as all states of X except x_S and X_E as all states of X except x_E. □

References

[1] D. Koller, N. Friedman, Probabilistic Graphical Models: Principles and Techniques, MIT Press, 2009.
[2] U. Nodelman, C. Shelton, D. Koller, Continuous time Bayesian networks, in: Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI), 2002, pp. 378–387.
[3] C.R. Shelton, Y. Fan, W. Lam, J. Lee, J. Xu, Continuous time Bayesian network reasoning and learning engine, J. Mach. Learn. Res. 11 (2010) 1137–1140.
[4] I. Cohn, T. El-Hay, N. Friedman, R. Kupferman, Mean field variational approximation for continuous-time Bayesian networks, J. Mach. Learn. Res. 11 (2010) 2745–2783.
[5] Y. Fan, J. Xu, C.R. Shelton, Importance sampling for continuous time Bayesian networks, J. Mach. Learn. Res. 99 (2010) 2115–2140.
[6] Y. Fan, C. Shelton, Sampling for approximate inference in continuous time Bayesian networks, in: Tenth International Symposium on Artificial Intelligence and Mathematics, 2008.
[7] J.C. Weiss, S. Natarajan, C.D. Page, Learning when to reject an importance sample, in: Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence, 2013.
[8] T. El-Hay, N. Friedman, R. Kupferman, Gibbs sampling in factorized continuous-time Markov processes, in: Proceedings of the Twenty-Fourth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-08), AUAI Press, Corvallis, OR, 2008, pp. 169–178.
[9] V. Rao, Y.W. Teh, Fast MCMC sampling for Markov jump processes and extensions, J. Mach. Learn. Res. 14 (1) (2013) 3295–3320.
[10] U. Nodelman, D. Koller, C. Shelton, Expectation propagation for continuous time Bayesian networks, in: Proceedings of the Twenty-First Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-05), AUAI Press, Arlington, VA, 2005, pp. 431–440.
[11] S. Saria, U. Nodelman, D. Koller, Reasoning at the right time granularity, in: Proceedings of the Twenty-Third Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-07), 2007.
[12] I. Cohn, Mean field variational approximations in continuous-time Markov processes, Ph.D. thesis, The Hebrew University, 2009.
[13] I. Cohn, T. El-Hay, N. Friedman, R. Kupferman, Mean field variational approximation for continuous-time Bayesian networks, in: Proceedings of the Twenty-Fifth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-09), AUAI Press, Corvallis, OR, 2009, pp. 91–100.
[14] T. El-Hay, I. Cohn, N. Friedman, R. Kupferman, Continuous-time belief propagation, in: Proceedings of the 27th International Conference on Machine Learning (ICML), 2010.
[15] E. Celikkaya, C. Shelton, W. Lam, Factored filtering of continuous-time systems, in: Proceedings of the Twenty-Seventh Annual Conference on Uncertainty in Artificial Intelligence (UAI-11), AUAI Press, Corvallis, OR, 2011, pp. 61–68.
[16] B. Ng, A. Pfeffer, R. Dearden, Continuous time particle filtering, in: International Joint Conference on Artificial Intelligence, vol. 19, 2005, p. 1360.
[17] U. Nodelman, C.R. Shelton, D. Koller, Expectation maximization and complex duration distributions for continuous time Bayesian networks, in: Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence (UAI-05), 2005.
[18] D. Shi, J. You, Update rules for parameter estimation in continuous time Bayesian network, in: PRICAI 2006: Trends in Artificial Intelligence, Springer, 2006, pp. 140–149.
[19] C.R. Shelton, G. Ciardo, Tutorial on structured continuous-time Markov processes, J. Artif. Intell. Res. 51 (2014) 725–778.
[20] U. Nodelman, E. Horvitz, Continuous time Bayesian networks for inferring users' presence and activities with extensions for modeling and evaluation, Tech. Report MSR-TR-2003-97, Microsoft Research, 2003.
[21] R. Herbrich, T. Graepel, B. Murphy, Structure from failure, in: Proceedings of the 2nd USENIX workshop on Tackling computer systems problems with machine learning techniques, USENIX Association, 2007, pp. 1–6.
[22] Y. Fan, C. Shelton, Learning continuous-time social network dynamics, in: Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI-09), AUAI Press, Arlington, VA, United States, 2009, pp. 161–168.
[23] Y. Fan, Continuous time Bayesian network approximate inference and social network applications, Ph.D. thesis, University of California, 2009.
[24] D. Shi, X. Tang, J. You, An intelligent system based on adaptive CTBN for uncertainty reasoning in sensor networks, Intell. Autom. Soft Comput. 16 (3) (2010) 337–351.
[25] J. Xu, C.R. Shelton, Intrusion detection using continuous time Bayesian networks, J. Artif. Intell. Res. 39 (1) (2010) 745–774.
[26] J. Xu, A continuous time Bayesian network approach for intrusion detection, Ph.D. thesis, University of California, 2010.
[27] J. Xu, C. Shelton, Continuous time Bayesian networks for host level network intrusion detection, in: W. Daelemans, B. Goethals, K. Morik (Eds.), Machine Learning and Knowledge Discovery in Databases, in: Lecture Notes in Computer Science, vol. 5212, Springer, Berlin/Heidelberg, 2008, pp. 613–627.
[28] S. Qiao, C. Tang, H. Jin, T. Long, S. Dai, Y. Ku, M. Chau, PutMode: prediction of uncertain trajectories in moving objects databases, Appl. Intell. 33 (3) (2010) 370–386.
[29] S. Qiao, X. Wang, L. Liu, X. Gong, Key techniques for predicting the uncertain trajectories of moving objects with dynamic environment awareness, J. Mod. Transp. 19 (3) (2011) 199–206.

[30] E. Gatti, D. Luciani, F. Stella, A continuous time Bayesian network model for cardiogenic heart failure, Flex. Serv. Manuf. J. (2011) 1–20.
[31] E. Gatti, Graphical models for continuous time inference and decision making, Ph.D. thesis, Università degli Studi di Milano-Bicocca, 2011.
[32] F. Stella, E. Acerbi, Continuous time Bayesian networks for gene network reconstruction: a comparative study on time course data, in: 10th International Symposium on Bioinformatics Research and Applications, 2014.
[33] E. Acerbi, T. Zelante, V. Narang, F. Stella, Gene network inference using continuous time Bayesian networks: a comparative study and application to Th17 cell differentiation, BMC Bioinform. 15 (1) (2014) 387.
[34] L. Perreault, J. Sheppard, H. King, L. Sturlaugson, Using continuous-time Bayesian networks for standards-based diagnostics and prognostics, in: AUTOTESTCON, IEEE, 2014, pp. 198–204.
[35] K.F. Kan, C.R. Shelton, Solving structured continuous-time Markov decision processes, in: Proceedings of Tenth International Symposium on Artificial Intelligence and Mathematics, 2008.
[36] D. Codetta-Raiteri, L. Portinale, Generalized continuous time Bayesian networks and their GSPN semantics, in: European Workshop on Probabilistic Graphical Models, 2010, pp. 105–112.
[37] F. Stella, Y. Amer, Continuous time Bayesian network classifiers, J. Biomed. Inform. 45 (6) (2012) 1108–1119.
[38] D. Codecasa, F. Stella, Conditional log-likelihood for continuous time Bayesian network classifiers, in: International Workshop NFMCP Held at ECML-PKDD, 2013.
[39] D. Codetta-Raiteri, L. Portinale, A Petri net-based tool for the analysis of generalized continuous time Bayesian networks, in: Theory and Application of Multi-Formalism Modeling, 2013, p. 118.
[40] D. Codecasa, F. Stella, Learning continuous time Bayesian network classifiers, Int. J. Approx. Reason. 55 (2014) 1728–1746.
[41] D. Codecasa, F. Stella, A classification based scoring function for continuous time Bayesian network classifiers, in: New Frontiers in Mining Complex Patterns, Springer, 2014, pp. 35–50.
[42] D. Codecasa, F. Stella, Classification and clustering with continuous time Bayesian network models, J. Intell. Inf. Syst. (2014) 1–34.
[43] J. Weiss, S. Natarajan, D. Page, Multiplicative forests for continuous-time processes, in: Advances in Neural Information Processing Systems, vol. 25, 2012, pp. 467–475.
[44] J.C. Weiss, D. Page, Forest-based point process for event prediction from electronic health records, in: Machine Learning and Knowledge Discovery in Databases, Springer, 2013, pp. 547–562.
[45] N. Friedman, M. Goldszmidt, Learning Bayesian networks with local structure, in: Learning in Graphical Models, Springer, 1998, pp. 421–459.
[46] K. Gopalratnam, H. Kautz, D.S. Weld, Extending continuous time Bayesian networks, Proceedings of the National Conference on Artificial Intelligence, vol. 20, AAAI Press/MIT Press, Menlo Park, CA, Cambridge, MA, London, 1999, 2005, p. 981.
[47] U. Nodelman, Continuous time Bayesian networks, Ph.D. thesis, Stanford University, Stanford, CA, 2007.
[48] A. Pfeffer, Asynchronous dynamic Bayesian networks, in: Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence (UAI-05), 2005.
[49] H. Chan, A. Darwiche, On the revision of probabilistic beliefs using uncertain evidence, Artif. Intell. 163 (1) (2005) 67–90.
[50] R. Pan, Y. Peng, Z. Ding, Belief update in Bayesian networks using uncertain evidence, in: 18th IEEE International Conference on Tools with Artificial Intelligence, IEEE, 2006, 2006, pp. 441–444.
[51] J. Bilmes, On virtual evidence and soft evidence in Bayesian networks, Report UWEETR-2004-0016, Department of Electrical Engineering, University of Washington, 2004.
[52] ESRC Research Centre on Micro-social Change, British household panel survey, Computer Data File and Associated Documentation, The Data Archive, Colchester, 2003, http://www.iser.essex.ac.uk/bhps.
[53] L. Sturlaugson, J.W. Sheppard, Inference complexity in continuous time Bayesian networks, in: Uncertainty in Artificial Intelligence, 2014.
[54] C. Moler, C. Van Loan, Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later, SIAM Rev. 45 (1) (2003) 3–49.