Hybrid Time Bayesian Networks

Manxia Liu
Arjen Hommersom, Maarten van der Heijden,
Peter J.F. Lucas

Software Science, ICIS, Radboud University
email: mliu@cs.ru.nl
September 7, 2016

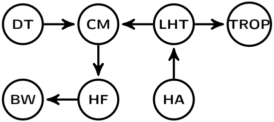
A motivating example

Heart attack (HA)

- heart attack causes heart muscle loss (LHT)
- dying muscle increases muscle protein levels in blood (TROP)
- loss of muscle affects the contractability of myocardium (CM)

Heart failure (HF)

- measurement: body weight (BW)
- treatment: digitalis (DT)



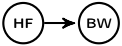
+ regular variables: DT, BW, HF, CM
+ irregular variables: HA, LHT, TROP

Dynamic Bayesian Networks (DBNs)

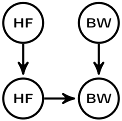
Representation

- initial model

$\alpha = 0$



- transition model



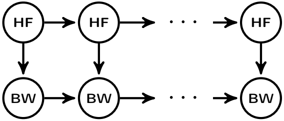
α

$s(\alpha)$

Unrolling

- standard BN

$\alpha = 0$ $\alpha = 1$ $\alpha = |A| - 1$



- the distribution of variables at time $s(\alpha)$ is conditional on the state of system at time α

Continuous Time Bayesian Networks (CTBNs)

Homogeneous Markov processes

- $n \times n$ matrix, n is the number of possible values
- entries represent transition intensities
- main diagonal makes each row sum up to zero
- transition intensities are independent from time

Example

- state space: low, normal, high

$$Q_{\text{BW}} = \begin{array}{ccc} -0.13 & 0.09 & 0.04 \\ 0.13 & -0.23 & 0.1 \\ 0.07 & 0.16 & -0.23 \end{array}$$

- entry $(3, 2)$ means that a transition from high at time β to normal at time $\beta + \epsilon$ with a probability of 0.16/0.23=0.696

Continuous Time Bayesian Networks (CTBNs)

Conditional Markov processes

- inhomogeneous
- depend on the values of other variables
- defined by conditional intensity matrices (CIMs)
- simply a set of intensity matrices

Amalgamation

- produce a larger CIMs by combining two CIMs
- assume no two transitions at the same time
- each entry determined by the changing variable corresponding to its intensity matrix
- entire system is a homogeneous Markov process

Continuous Time Bayesian Networks (CTBNs)

Queries

- for variable X, given an initial distribution P_0 and an intensity matrix Q_X

We can compute

- a distribution at a time point: $P(X_t) = P_0 \exp(Q(t))$
- distribution at two time points: $P(X_t, X_s) = P(X_s) \exp(Q(t-s))$
- conditional probability: $P(X_t \mid X_s) = \exp(Q(t-s))$ $s < t$

Comparison

DBNs

- Upsides
  + parameterized by CTPs
  + easy to explain
- Downsides
  - finest time granularity
  - information missing on discrete time points

CTBNs

- Upsides
  + time is continuous
  + capture different transition rates
- Downsides
  - indirect representation of probability knowledge
  - hard to interpret
  - assume time before a transition is exponentially distributed

Hybrid Time Bayesian Networks (HTBNs)

Formally

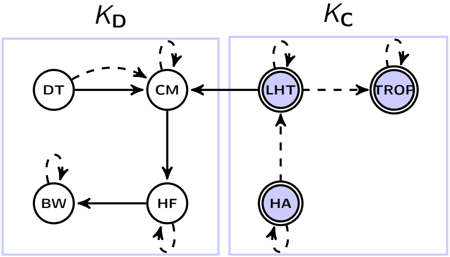
- a triple $\mathcal{H} = (G, \Phi, \Lambda)$
- $G = (V(G), E^t(G), E^a(G))$
- Φ: conditional probabilities
- Λ: intensity matrices and initial distribution

An HTBN can be naively looked as a combination of a number of dynamic Bayesian networks (DBNs) and continuous-time Bayesian networks (CTBNs).

HTBNs (Heart Failure Example)

Heart Failure Example

- continuous-time variables: *LHT*, *TROE*, *HA*
- discrete-time variables: *DT*,*CM*,*BW*,*HF*



Hybrid Time Bayesian Networks (Factorization)

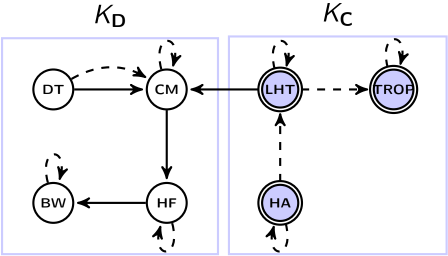
Components

- connected components
+ $K_{\mathbf{C}}$: continuous connected components
+ $K_{\mathbf{D}}$: discrete connected components

- parents of components
+ $\pi(V(K_{\mathbf{C}}))$: parents of $K_{\mathbf{C}}$
+ $\pi(V(K_{\mathbf{D}}))$: parents of $K_{\mathbf{D}}$

Example

- $\pi(V(K_{\mathbf{C}})) = \emptyset$
- $\pi(V(K_{\mathbf{D}})) = \{LHT\}$



Hybrid Time Bayesian Networks (Factorization)

Distribution for components

- distribution for $K_{\mathbf{C}}$

$$P(V(K_{\mathbf{C}})_B \mid \pi(V(K_{\mathbf{C}}))_A) = P(V(K_{\mathbf{C}})_0) \prod_{\beta \in B \setminus \{\max B\}} \exp(Q_{V(K_{\mathbf{C}}) \mid \pi(V(K_{\mathbf{C}}))_a}(s(\beta) - \beta))$$

$$a = \max\{\alpha \mid \alpha < \beta, \alpha \in A\}$$

- distribution for $K_{\mathbf{D}}$

$$P(V(K_{\mathbf{D}})_A \mid \pi(V(K_{\mathbf{D}}))_A) = \prod_{D \in V(K_{\mathbf{D}})} (P(D_0 \mid \pi^a(D)_0) \prod_{\alpha \in A \setminus \{0\}} P(D_\alpha \mid \pi^a(D)_\alpha, \pi^t(D)_{\alpha-1}))$$

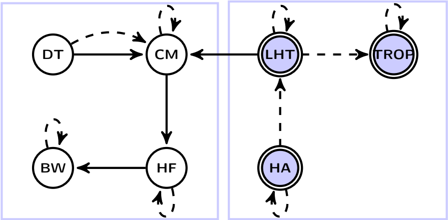
Hybrid Time Bayesian Networks (Factorization)

Joint distribution

- given an HTBN with an associated time points A, B

$$P(V(G)_B) = \prod_{K_\mathbf{C} \in \mathbf{K_C}} P(V(K_\mathbf{C})_B \mid \pi(V(K_\mathbf{C}))_A)$$
$$\prod_{K_\mathbf{D} \in \mathbf{K_D}} P(V(K_\mathbf{D})_A \mid \pi(V(K_\mathbf{D}))_A)$$

$K_\mathbf{D}$ $K_\mathbf{C}$



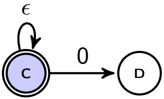
Likelihood

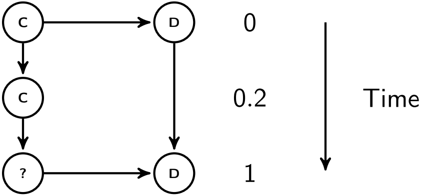
Given time points A, B for an HTBN $\mathcal{H}$ with graph G, parameter $\boldsymbol{\theta}$, we can compute the log-likelihood of $\mathcal{H}$ given a trajectory $\boldsymbol{\mathcal{D}}$ by summing out unobserved variables.

$$\ell_{\mathcal{H}}(\boldsymbol{\theta} : \boldsymbol{\mathcal{D}}) = \ln \sum_{V(G)_{\mathbf{B}} \setminus \boldsymbol{\mathcal{D}}} P(V(G)_{\mathbf{B}})$$

- Missing values: continuous-time variables are observed at arbitrary in time

Incomplete data

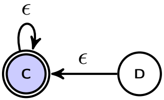


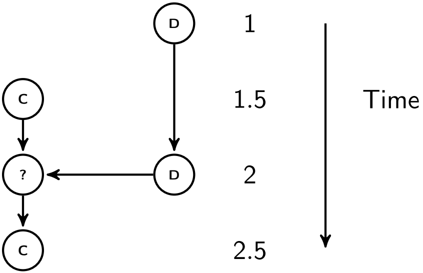


Likelihood of D at time 1:

$$\ell_{\mathcal{H}}(\boldsymbol{\theta} : d_1, d_0, c_{0.2}, c_0) = \ln P(c_0 : \boldsymbol{\theta}) + \ln P(c_{0.2} | c_0 : \boldsymbol{\theta}) + \ln P(d_0 : \boldsymbol{\theta}) + \ln \sum_{C_1} P(d_1 | d_0, C_1 : \boldsymbol{\theta}) P(C_1 | c_{0.2} : \boldsymbol{\theta})$$

Incomplete data





The value of C at time 2.5 is dependent on D at time 2:

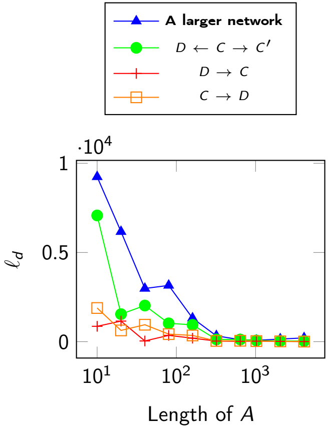
$$\ell_{\mathcal{H}}(\boldsymbol{\theta} : c_{2.5}, d_2, c_{1.5}, d_1) = \ln P(d_1 : \boldsymbol{\theta}) + \ln P(d_2 | d_1 : \boldsymbol{\theta}) + \ln P(c_{1.5} : \boldsymbol{\theta}) + \ln \sum_{C_2} P(c_{2.5} | d_2, C_2 : \boldsymbol{\theta}) P(C_2 | c_{1.5}, d_1 : \boldsymbol{\theta})$$

Experiments

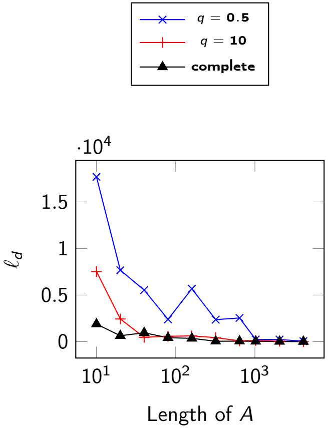
Set up

- complete data: exact MAP estimates
- incomplete data
  + MCMC sampling
  + Rstan: R interface for stan
  + iteration: 1000
  + Dirichlet distributions: one's
  + Gamma distribution: two's
- evaluation
  + log-likelihood: $\ell_d = |\ell_t - \ell_e|$

Results

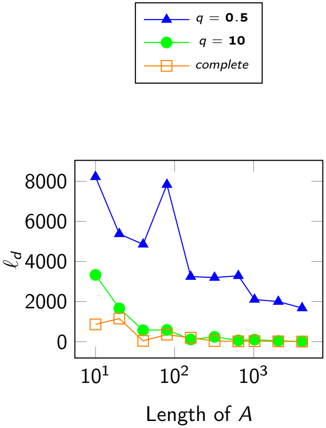


complete data

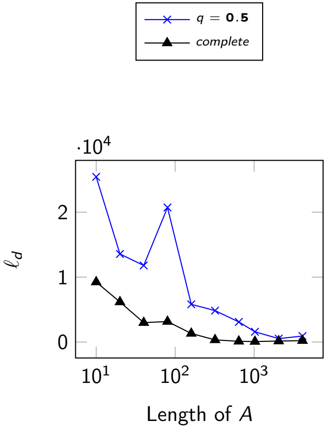


$C \rightarrow D$

Results



$D \to C$



A larger network

Conclusion

Future work

- whether the estimated parameters of discrete-time and continuous-time variables converge to the true parameters at same pace
- Investigate the quality of learned models for HTBNs with a varying number of components.
- combine different discrete-time granularities within the hybrid-time framework
- optimize the BN construction to only take into account dependencies that have significant influence on the temporal process