Bayesian Networks - III : Dynamic Bayesian Networks

Philippe LERAY
philippe.leray@univ-nantes.fr

DUKe (Data User Knowledge) Research group
Laboratoire des Sciences du Numérique de Nantes – UMR 6004
Site de l'Ecole Polytechnique de l'université de Nantes

Introduction

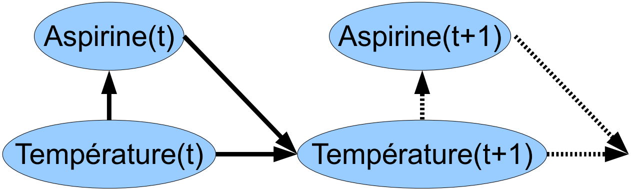
BN limitation

- BNs don't take time into account
- for this reason, BNs are not able to model some situations :
  - when a patient has a fever, he takes aspirin
  - taking aspirin makes the temperature falling
  - $\boxed{\text{Temperature} \rightarrow \text{Aspirin}}$ or $\boxed{\text{Temperature} \leftarrow \text{Aspirin}}$?

Introduction

Answer : dynamic BN

- Taking aspirin is an "immediate" decision :
  Temperature(t) $\rightarrow$ Aspirin(t)
- Temperature evolution depends on past temperature and past aspirin decision :
  { Temperature(t), Aspirin(t) } $\rightarrow$ Temperature(t+1)



Dynamic probabilistic graphical models

Some existing models

- Markov chain
- Hidden Markov Model (HMM)
- Dynamic Bayesian Networks (DBN)

Markov Chain

Principle : stochastic modeling of a random process

- X random process $= X_t$ random variable
- t discrete $(t = 1, 2, ...)$
- X discrete, described by n distinct states

- In order to estimate X_t, we need the whole log X_1 to X_{t-1}, and computation of $P(X_t|X_1 ... X_{t-1})$

Markov Chain

First order Markov Chain

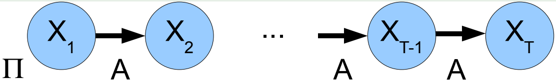
- Current state depends only on previous one

$$P(X_t|X_1 \dots X_{t-1}) = P(X_t|X_{t-1})$$

- stationary Markov chain : probability of transition is independent of t

$$P(X_t = j|X_{t-1} = i) = A_{ij}$$

- A : transition matrix
- Π : probability of chain initialization ($P(X_1)$)



Example

Albert the student

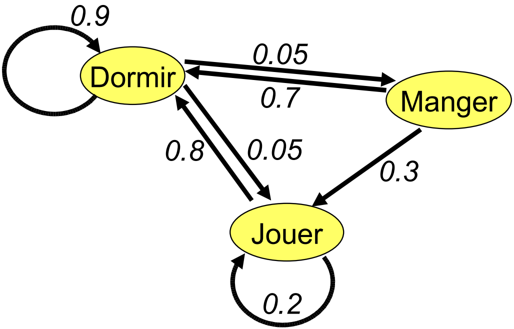
Albert the student knows only 3 places in his university : the table where he sleeps, the candy machine here he eats and the computer room where he plays. His days are all very similar, and his activity can be modeled by a Markov chain. Every minute, Albert can choose a new activity or continue the one he was doing :

- when he sleep, there is a 9/10 chance he keeps sleeping the next minute
- when he stops sleeping, there is a 1/2 chance he decides to eat and 1/2 chance he decides to play.
- eating a candy doesn't take so much time, so he stops eating after 1 minute.
- when he stops eating, there is a 3/10 chance he decides to play, and a 7/10 chance he decides to sleep.
- playing is exhausting, there is a 80 % chance he decides to sleep after one minute, otherwise he keeps playing and forgets he was exhausted.

Other representation

In state space

Another way to represent the transition matrix



Use

Several possible uses

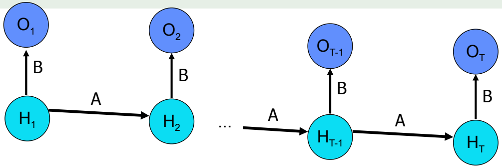
- Let's suppose that the initial state is "sleeping"
- What is the probability Albert sleeps for the 5 next minutes ?
- What is the probability he sleeps in 5 minutes ?

Hidden Markov Models (HMM)

Principle

- The observed variable O_t is not a Markovian process
- But this variable is "generated" by an **hidden** variable H_t
- and H_t is a Markovian process

- A : transition matrix for H, $P(H_t|H_{t-1})$
- B : emission matrix $P(O_t|H_t)$, independent from t
- Π : probability of chain initialization $(P(H_1))$



HMM use

Prediction $P(O_{t+1}|O_1 \ldots O_t)$?

- Forward-Backward algorithm == Message Passing !

Explanation $argmax_{H_1 \ldots H_t} P(H_1 \ldots H_t|O_1 \ldots O_t)$?

- Viterbi algorithm == abductive inference in BN

Learning ?

- $\mathcal{D}$: we observe one (or several) sequences $O_1 \ldots O_T$
- incomplete data : H never observed
- What are the parameters Π, A, B which maximise the likelihood ?
- Baum & Welch algorithm == adaptation of EM algorithm

Example

Albert the student is back

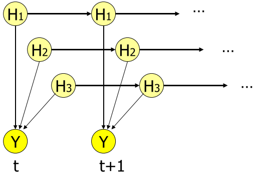
Albert and his colleague Bertrand are discussing in the same room. As they are not able to do two things at the same time, they have to stop talking in order to listen the other one. Their vocabulary is limited to 3 words : *Eat*, *Sleep*, *Wii*.

- Albert essentially speaks about *sleeping* (85%) or with the same probability about *eating* and *Wii*
- Bertrand essentially speaks about *Wii* (85%), then *eating* (10%) and *Wii* (5%)
- Albert is often listening Bertrand : when Albert says one word, there is a 2/3 chance he stops speaking in order to listen his colleague
- Bertrand keeps talking with a 80% probability
- Usually, Albert starts talking.

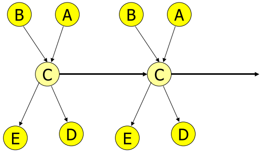
Application

- One microphone in the room records the discussion:
  Eat Eat Sleep Sleep Wii Wii Wii Wii Sleep Sleep Sleep Sleep Eat Eat
- What is the next word the most probable ?
- What is the most probable sequence of speakers ?

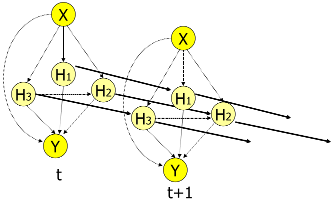
HMM extensions



- Factorial HMM



- Input Output HMM

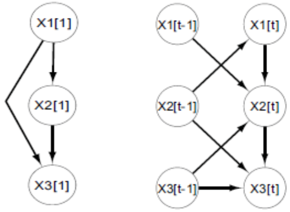


- Hidden Markov Decision Tree

Dynamic Bayesian networks (DBNs)

k slices temporal BN (k-TBN) [Murphy, 2002]

- $k-1$ Markov order
- prior graph G_0 + transition graph $G_{\rightarrow}$
- for example : 2-TBNs model [Dean & Kanazawa, 1989]



(a) Prior network (b) Transition network

Simplified k-TBN

- k-TBN with only temporal edges [Dojer, 2006][Vinh et al, 2012]

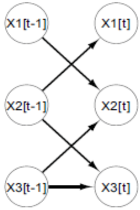
Dynamic Bayesian networks (DBNs)

k slices temporal BN (k-TBN) [Murphy, 2002]

- $k-1$ Markov order
- prior graph G_0 + transition graph $G_{\rightarrow}$
- for example : 2-TBNs model [Dean & Kanazawa, 1989]

Simplified k-TBN

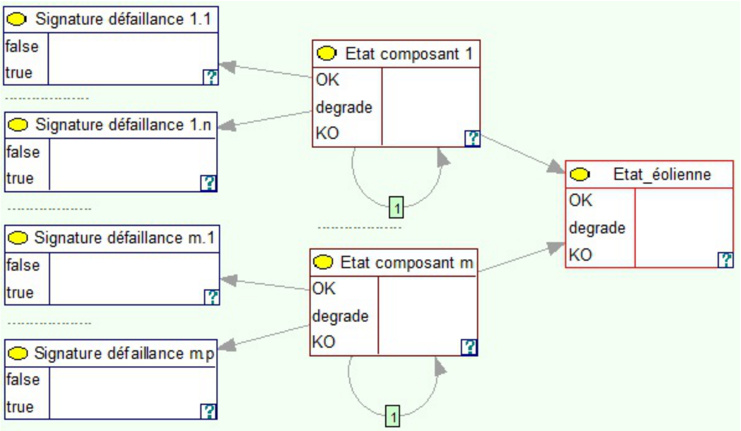
- k-TBN with only temporal edges [Dojer, 2006][Vinh et al, 2012]



(c) Transition network with only inter time-slice arcs

One example

FUI Hyperwind - windfarm modelling [Gherasim et al., 2016]



Inference in DBN = NP hard

Exact inference

- Adaptation of Forward-backward
  convert DBN in HMM, practical if N_H is low
- Unrolled junction tree : unroll the DBN over T and apply an usual inference algorithm, problem = size of cliques !
- Frontier algorithm [Zweig, 1996]
- Interface algorithm [Murphy, 2001]
- Kalman filtering [Minka, 1998]

Approximated Inference

- Deterministic algorithms
- Stochastic algorithms (sampling)

Inference in DBN = NP hard

Exact inference

- Adaptation of Forward-backward
  convert DBN in HMM, practical if N_H is low
- Unrolled junction tree : unroll the DBN over T and apply an usual inference algorithm, problem = size of cliques !
- Frontier algorithm [Zweig, 1996]
- Interface algorithm [Murphy, 2001]
- Kalman filtering [Minka, 1998]

Approximated Inference

- Deterministic algorithms
- Stochastic algorithms (sampling)

DBN structure learning

Score-based methods

- dynamic Greedy Search [Friedman et al., 1998], genetic algorithm [Gao et al., 2007], dynamic Simulated Annealing [Hartemink, 2005], ...
- for k-TBN (G_0 and $G_{\rightarrow}$ learning)
- but not scalable (high n)

Hybrid methods

- [Dojer, 2006] [Vinh et al., 2012] for simplified k-TBN, but often limited to $k = 2$ for scalability
- dynamic MMHC for "unsimplified" 2-TBNs with high n [Trabelsi et al., 2013]

DBN structure learning

Score-based methods

- dynamic Greedy Search [Friedman et al., 1998], genetic algorithm [Gao et al., 2007], dynamic Simulated Annealing [Hartemink, 2005], ...
- for k-TBN (G_0 and $G_\rightarrow$ learning)
- but not scalable (high n)

Hybrid methods

- [Dojer, 2006] [Vinh et al., 2012] for simplified k-TBN, but often limited to $k = 2$ for scalability
- dynamic MMHC for "unsimplified" 2-TBNs with high n [Trabelsi et al., 2013]