Bayesian Networks and Inference

Drew Kristensen

University of Puget Sound

Nov. 27, 2017

Outline

You live in California with a house that has a sensitive burglar alarm with two neighbors, Mary and John. You go on a trip out of state but while away, you get a call from neighbor John. You want to know what the probability that John is calling because of a burglary instead of an earthquake, but you don't know how to go about computing this.

What do you do?

You can create a Bayesian network and perform probabilistic inference on your model

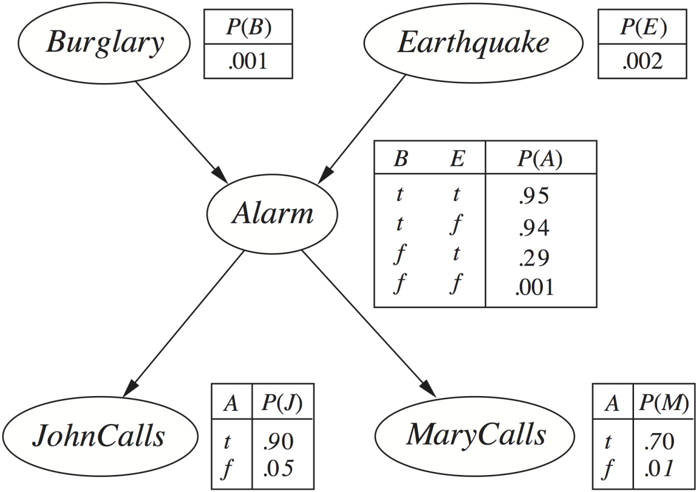


Figure 1: Simple Bayesian network for the example

What is a Bayesian Network?

Simply put, a Bayesian network represents your beliefs on dependencies between variables

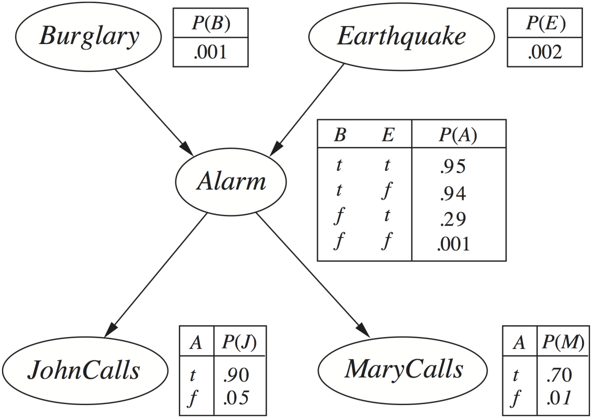
Create probabilistic models of events in order to predict outcomes

How does a Bayesian network work?

Each edge in our belief network represents a dependency that we think exists

For independent random variables, we can break down our joint probability equation into a product of simpler conditional probabilities.

Factoring Example



$$P(A, B, E, J, M) = P(A \mid B, E)P(B)P(E)P(J \mid A)P(M \mid A)$$

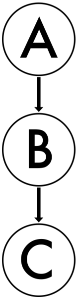
D-Separation

D-separation is an algorithm used to determine independence between variables

- Makes use of the three possible connections in a Bayesian network
- "Explains Away" outcomes based on observed variables

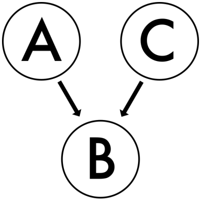
Connections

Linear



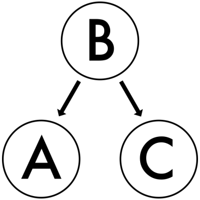
- A, C dependent
- Observing B renders A, C independent

Convergent



- A, C independent
- Observing B renders A, C dependent

Divergent



- A, C dependent
- Observing B renders A, C independent

Probabilistic Inference

Split your variables into one of three sets: Query, Evidence, Hidden.

Probabilistic inference is a four step process for a Bayesian network

1. Set the evidence variables to be consistent with the inference task
2. Marginalize over all the hidden variables
3. Compute the probability for each query variable taking on each possible value in it's domain
4. Normalize the computed probabilities to sum to 1 to be a proper probability distribution

Probabilistic Inference: Example

From our example at the start, what is the probability that John is calling because of a burglary instead of an earthquake?
First we need to break our random variables into our three sets

- Query: Since we want to know the probability that John is calling because of a burglary, we let J be our query variable
- Evidence: Since we want to know what the probability is that there was a burglary and not an earthquake, we let both B and E be our evidence variables
- Hidden: This leaves both Mary and Alarm as that random variables not in the sets above, so A and M become our hidden variables to marginalize over

Probabilistic Inference: Example

To compute this, we now have

$$
\begin{aligned}
P(j \mid \neg e, b) &\propto P(j, \neg e, b) \\
&= \sum_A \sum_M P(A, M, j, b, \neg e) \\
&= P(\neg e)P(b) \sum_A P(A \mid \neg e, b) P(j \mid A) \sum_M P(M \mid A) \\
&= (0.998)(0.001)\left[(0.94)(0.9) + (0.06)(0.05)\right] \\
&= 0.000847302
\end{aligned}
$$

When we compute $P(\neg j \mid \neg e, b)$ and normalize the two, we find that the probability of John calling because of a burglary and not an earthquake is 0.849

You work as a security guard in a top-secret underground lab. You notice that some days, researchers come in with umbrellas and some days they don't. Most of the days, if they bring an umbrella, it is dripping wet. You want to know what the probability that it was raining today given your observations for the past week.

What do you do?

Dyanmic Bayesian Networks

Dynamic Bayesian networks (DBNs) are useful ways to probabilistically model successive events that can cause different actions

For each time t, we let X_t be the set of hidden variables for t, and we let E_t be the set of evidence variables for time t.

Transmission model vs Emission model

Transmission Model

- Specifies $P(X_t \mid X_{0:t-1})$
- Markov Assumption: $P(X_t \mid X_{0:t-1}) = P(X_t \mid X_{t-1})$

$$T = \begin{bmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{bmatrix}$$

Emission Model

- Specifies $P(E_t \mid X_{0:t}, E_{0:t-1})$
- Sensor Markov Assumption: $P(E_t \mid X_{0:t}, E_{0:t-1}) = P(E_t \mid X_t)$

$$E = \begin{bmatrix} 0.7 & 0.15 \\ 0.3 & 0.85 \end{bmatrix}$$

Inference in the Dynamic case

To perform inference in for a DBN, we have to use different methods than for the static case.

We use a method known as filtering, where we want to know the state at time t given the evidence up to and including time t

Filtering: Overview

We use Bayes' Rule extensively when computing our probability, since

$$P(X_t \mid E_{1:t}) \propto P(E_t \mid X_t)P(X_t \mid E_{1:t-1})$$

Furthermore, we make use of marginalizing probabilities to break up more complicated problems, such as

$$P(X_t \mid E_{1:t-1}) = \sum_{X_{t-1}} P(X_t, X_{t-1} \mid E_{1:t-1})$$

We get that for time t,

$$P(X_t \mid E_{1:t}) = P(E_t \mid X_t) \sum_{X_{t-1}} \left[P(X_t \mid X_{t-1})P(X_{t-1} \mid E_{1:t-1})\right]$$

Conclusion

Using this, we can ask complicated questions about our system, like what is the probability of it raining today, given the past four days, we noticed an umbrella, no umbrella, no umbrella, and no umbrella?

This is the main use for dynamic Bayesian networks, since they have a large amount of probabilistic expression power given the relatively low amount of work it takes to create the model