

Vytěžování dat, cvičení 4:

Bayesovské sítě

Radomír Černoch



Evropský sociální fond
Praha & EU: Investujeme do vaší budoucnosti

Fakulta elektrotechnická, ČVUT

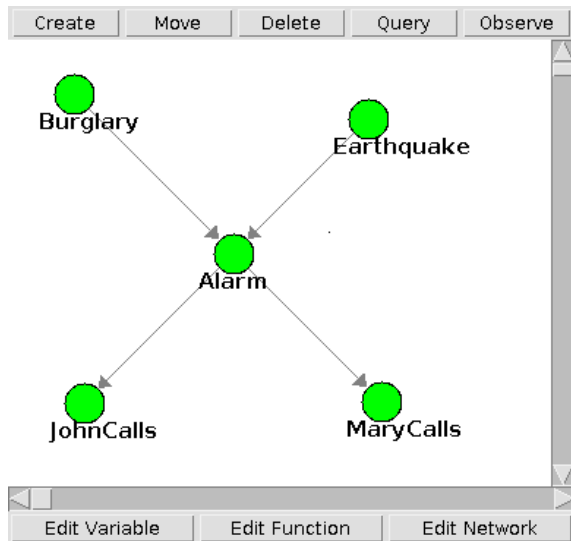
Consider the following situation. You have a new burglar alarm installed at home. It is fairly reliable at detecting a burglary, but also responds on occasion to minor earthquakes.¹ You also have two neighbors, John and Mary, who have promised to call you at work when they hear the alarm. John always calls when he hears the alarm, but sometimes confuses the telephone ringing with the alarm and calls then, too. Mary, on the other hand, likes rather loud music and sometimes misses the alarm altogether. Given the evidence of who has or has not called, we would like to estimate the probability of a burglary. Zdroj: AIMA

¹This example is due to Judea Pearl, a resident of Los Angeles; hence the acute interest in earthquakes.

Krok 1: Vytvoření struktury Bayesovské sítě

- ▶ Mějme 4 náhodné proměnné:
Burglary, Earthquake, Alarm, JohnCalls, MaryCalls.
- ▶ Základní princip: Šipka vede od X do Y právě tehdy když X má přímý vliv na Y .
- ▶ Navrhněte vazby mezi proměnnými!
- ▶ Applet: <http://www.cs.cmu.edu/~javabayes/>

Krok 1: Správná odpověď



Krok 3: Zadání parametrů

Name:

Values:

Types:

☒ Chance node ☐ Single distribution

☐ Explanatory node ☐ Credal set with extreme points

Variable properties:

Next

Function properties:

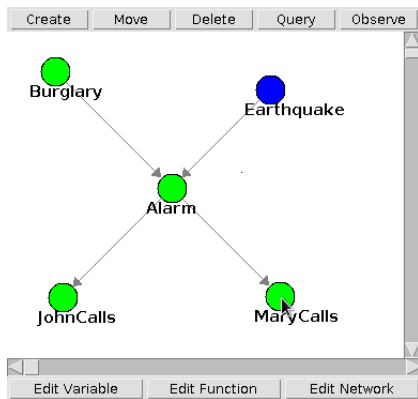
Next

p(Burglary)

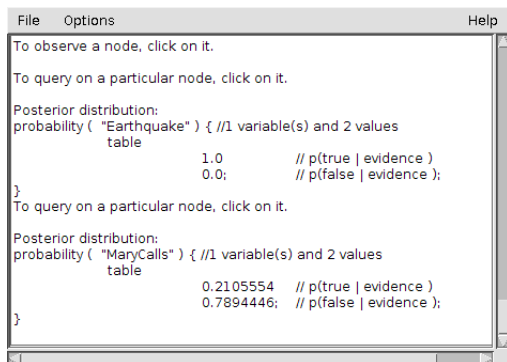
| | |
|-------|------------------------------------|
| true | <input type="text" value="0.001"/> |
| false | <input type="text" value="0.999"/> |

- ▶ $P(\textit{Burglary} = \textit{true}) = 1\%$
- ▶ $P(\textit{Earthquake} = \textit{true}) = 0.2\%$
- ▶ $P(\textit{Alarm} = \textit{true} \mid \textit{Burglary} = \textbf{true}, \textit{Earthquake} = \textbf{true}) = 95\%$
- ▶ $P(\textit{Alarm} = \textit{true} \mid \textit{Burglary} = \textbf{true}, \textit{Earthquake} = \textbf{false}) = 94\%$
- ▶ $P(\textit{Alarm} = \textit{true} \mid \textit{Burglary} = \textbf{false}, \textit{Earthquake} = \textbf{true}) = 29\%$
- ▶ $P(\textit{Alarm} = \textit{true} \mid \textit{Burglary} = \textbf{false}, \textit{Earthquake} = \textbf{false}) = 0.1\%$
- ▶ $P(\textit{JohnCalls} = \textit{true} \mid \textit{Alarm} = \textbf{true}) = 90\%$
- ▶ $P(\textit{JohnCalls} = \textit{true} \mid \textit{Alarm} = \textbf{false}) = 5\%$
- ▶ $P(\textit{MaryCalls} = \textit{true} \mid \textit{Alarm} = \textbf{true}) = 70\%$
- ▶ $P(\textit{MaryCalls} = \textit{true} \mid \textit{Alarm} = \textbf{false}) = 1\%$

Krok 4: Evidence



Krok 4: Výpočet podm. p.



The screenshot shows a window with a menu bar (File, Options, Help) and a text area containing the following text:

```
To observe a node, click on it.  
  
To query on a particular node, click on it.  
  
Posterior distribution:  
probability ( "Earthquake" ) { //1 variable(s) and 2 values  
    table  
        1.0          // p(true | evidence )  
        0.0;         // p(false | evidence );  
}  
To query on a particular node, click on it.  
  
Posterior distribution:  
probability ( "MaryCalls" ) { //1 variable(s) and 2 values  
    table  
        0.2105554    // p(true | evidence )  
        0.7894446;   // p(false | evidence );  
}
```


Krok 5: Ruční výpočet (1/3)

$$P(\text{MaryCalls} \mid \text{Earthquake} = \text{true}) = P(M \mid E) = \frac{P(M, E)}{P(E)} = \dots \quad (1)$$

$$P(M, E) = \sum_A \sum_B \sum_J P(B, E, A, M, J) \quad (2)$$

$$= \sum_A \sum_B \sum_J P(B) \cdot P(E) \cdot P(A \mid B, E) \cdot P(M \mid A) \cdot P(J \mid A) \quad (3)$$

$$= \sum_A P(M \mid A) \sum_B \sum_J P(B) \cdot P(E) \cdot P(A \mid B, E) \cdot P(J \mid A) \quad (4)$$

$$= \dots = P(E) \sum_A P(M \mid A) \sum_B P(B) \cdot P(A \mid B, E) \cdot \sum_J P(J \mid A) \quad (5)$$

$$= P(E) \sum_A P(M \mid A) \sum_B P(B) \cdot P(A \mid B, E) \quad (6)$$

Tahák: $\sum_X P(X) = 1$; $\sum_X P(X, Y) = P(Y)$;

$P(M, E) = P(E) \cdot P(M \mid E)$; $\sum_Y f(X) \cdot g(Y) = f(X) \sum_Y g(Y)$.

Krok 5: Ruční výpočet (2/3)

$$\begin{aligned}
 P(\text{MaryCalls} \mid \text{Earthquake} = \text{true}) &= \frac{P(M, E = \text{true})}{P(E = \text{true})} = \\
 &= \frac{P(E = \text{true}) \sum_A P(M \mid A) \sum_B P(B) \cdot P(A \mid B, E = \text{true})}{P(E = \text{true})} \\
 &= \sum_A P(M \mid A) \sum_B P(B) \cdot P(A \mid B, E = \text{true}) \\
 &= \sum_A P(M \mid A) \sum_B \left(B = \begin{array}{c} \text{true} \\ \text{false} \end{array} \mid \begin{array}{c} 0.01 \\ 0.99 \end{array} \right) \times \left(\begin{array}{c|cc} & \begin{array}{cc} A \\ \text{true} & \text{false} \end{array} \\ \begin{array}{c} B = \\ \text{true} \\ \text{false} \end{array} & \begin{array}{cc} 0.95 & 0.05 \\ 0.29 & 0.71 \end{array} \end{array} \right) \\
 &= \sum_A P(M \mid A) \sum_B \left(\begin{array}{c|cc} & \begin{array}{cc} A \\ \text{true} & \text{false} \end{array} \\ \begin{array}{c} B = \\ \text{true} \\ \text{false} \end{array} & \begin{array}{cc} 0.0095 & 0.0005 \\ 0.2871 & 0.7029 \end{array} \end{array} \right) \\
 &= \sum_A P(M \mid A) \left(\begin{array}{c|cc} & \begin{array}{cc} A \\ \text{true} & \text{false} \end{array} \\ & \begin{array}{cc} 0.2966 & 0.7034 \end{array} \end{array} \right)
 \end{aligned}$$

Krok 5: Ruční výpočet (2/3)

$$\begin{aligned}
 &= \sum_A P(M|A) \times \left(\begin{array}{c|c} & A \\ \hline & \text{true} & \text{false} \\ \hline & 0.2966 & 0.7034 \end{array} \right) \\
 &= \sum_A \left(\begin{array}{c|c} & A \\ \hline M = \text{true} & 0.75 & 0.01 \\ M = \text{false} & 0.30 & 0.99 \end{array} \right) \times \left(\begin{array}{c|c} & A \\ \hline & \text{true} & \text{false} \\ \hline & 0.2966 & 0.7034 \end{array} \right) \\
 &= \sum_A \left(\begin{array}{c|c} & A \\ \hline M = \text{true} & 0.22245 & 0.007034 \\ M = \text{false} & 0.08898 & 0.696366 \end{array} \right) \\
 &= \underline{\underline{\left(\begin{array}{c|c} M = \text{true} & 0.229484 \\ M = \text{false} & 0.785346 \end{array} \right)}}
 \end{aligned}$$

| Alt | Bar | Fri | Hun | Pat | Price | Rain | Res | Type | Est | Wait |
|-----|-----|-----|-----|------|-------|------|-----|---------|-------|------|
| Yes | No | No | Yes | Some | 3 | No | Yes | French | 0-10 | Yes |
| Yes | No | No | Yes | Full | 1 | No | No | Thai | >30 | No |
| No | Yes | No | No | Some | 1 | No | No | Burger | 0-10 | Yes |
| Yes | No | Yes | Yes | Full | 1 | No | No | Thai | 10-30 | Yes |
| Yes | No | Yes | No | Full | 3 | No | Yes | French | >30 | No |
| No | Yes | No | Yes | Some | 2 | Yes | Yes | Italian | 0-10 | Yes |
| No | Yes | No | No | None | 1 | Yes | No | Burger | 0-10 | No |
| No | No | No | Yes | Some | 2 | Yes | Yes | Thai | 0-10 | Yes |
| No | Yes | Yes | No | Full | 3 | No | Yes | Italian | >30 | No |
| Yes | Yes | Yes | Yes | Full | 3 | No | Yes | Italian | >30 | No |
| No | No | No | No | None | 1 | No | No | Thai | 0-10 | No |
| Yes | Yes | Yes | Yes | Full | 1 | No | No | Burger | >30 | Yes |

1. Alt: whether there is a suitable alternative restaurant nearby.
2. Bar: whether the restaurant has a comfortable bar area to wait in.
3. Fri: true on Fridays and Saturdays.
4. Hun: whether we are hungry.
5. Pat: how many people are in the restaurant (values are None, Some, and Full).
6. Price: the restaurant's price range (\$, \$\$, \$\$\$).
7. Rain: whether it is raining outside.
8. Res: whether we made a reservation.
9. Type: the kind of restaurant (French, Italian, Thai, or Burger).
10. Est: the wait estimated by the host (0-10 minutes, 10-30, 30-60, >60).
11. Wait: whether we decided to wait

1. Navrhnete strukturu Bayesovské sítě. Snažte se respektovat kauzální vazby mezi náhodnými proměnnými.
2. Z dodaných dat vypočtete podmíněné pravděpodobnosti, které odpovídají struktuře BS.
3. Pomocí počítače vypočtete následující pravděpodobnosti z Bayesovské sítě:
 - 3.1 $P(Est)$
 - 3.2 $P(Est | Pat)$
 - 3.3 $P(Rain)$
 - 3.4 $P(Rain | Fri)$
4. Podmíněnou pravděpodobnost 4 vypočtete navíc ručně.
5. Porovnejte výsledky 1 s 2 a dále 3 s 4. Ovlivňuje dotazovanou proměnou informace o proměnné v podmínce?
6. Proč nelze počítat podmíněné pravděpodobnosti přímo z dat a je dobré využít „mezikrok“ podmíněných pravděpodobností BS?