

Úloha 4 – Bayesovské sítě

Zadání:

Data:

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

Legenda:

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.

Zadání:

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:
 $P(Est)$
 $P(Est | Pat)$
 $P(Rain)$
 $P(Rain | Fri)$
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?

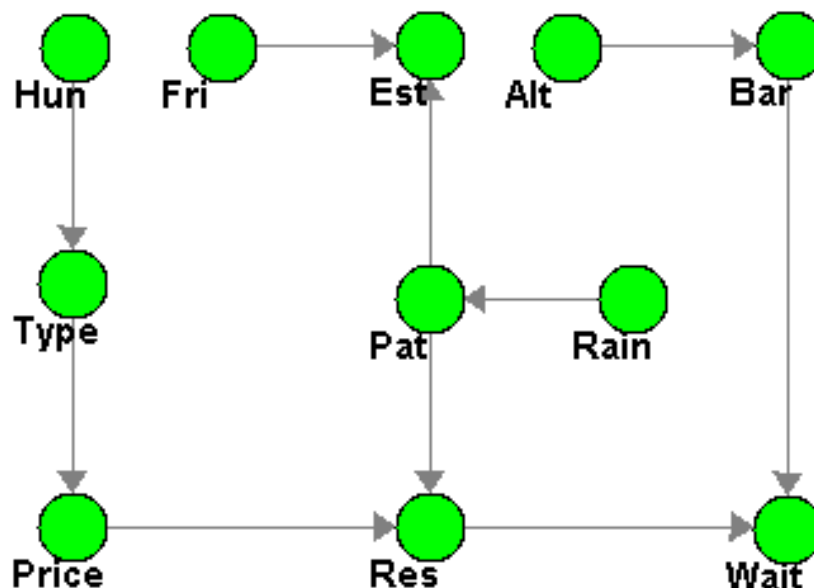
Použité nástroje:

Applet javaBayes – nástroj doporučený v dokumentaci předmětu.
 Dostupný z WWW <<http://www.cs.cmu.edu/~javabayes/>>.

Zpracování:

1. Návrh struktury Bayesovské sítě

Podle zadání jsem zpracovala následující Bayesovskou síť:



Legenda ke grafu:

- Alt** - může mít vliv na to, zda má restaurace bar (Alt → Bar)
- Bar** - může mít vliv na čekání (Bar → Wait)
- Fri** - může mít vliv na to, zda budeme čekat (Fri → Est)
- Hun** - může mít vliv na type restaurace (Hun → Type)
- Type** - může mít vliv na cenu (Type → Price)
- Price** - může mít vliv na množství rezervací (Price → Res)
- Res** - může mít vliv na to, zda budeme chtít čekat (Wait → Est)
- Rain** - může mít vliv na obsazenost restaurace (Rain → Pat)
- Pat** - může mít vliv na to, zda budeme čekat (Pat → Est)
- může mít vliv na množství rezervací (Pat → Res)

2. Výpočet podmíněných pravděpodobností

Nejprve jsem vypočetla nepodmíněné pravděpodobnosti pro tato data ALT, FRI, HUN, RAIN. Výpočet uvádím v následující tabulce (výsledky jsem zaokrouhlila na 2 desetinná místa):

	ALT	FRI	HUN	RAIN
<i>P(true)</i>	0,5	0,42	0,58	0,25
<i>P(false)</i>	0,5	0,58	0,42	0,75

Pro ostatní hodnoty jsem vypočetla podmíněné pravděpodobnosti (výsledky jsem opět zaokrouhlila na dvě desetinná místa).

Edit Function

$p(\text{Type} \mid \text{Hun})$

	Hun	true	false
French		0.14	0.2
Thai		0.43	0.2
Burger		0.14	0.4
Italian		0.29	0.2

Apply Dismiss

Edit Function

$p(\text{Price} \mid \text{Type})$

Type	French	Thai	Burger	Italian
1	0.0	0.75	1.0	0.0
2	0.0	0.25	0.0	0.33
3	1.0	0.0	0.0	0.67

Apply Dismiss

Edit Function

$p(\text{Res} \mid \text{Price}, \text{Pat})$

Values for parents:

Pat:

Price	1	2	3
true	0.0	0.0	0.0
false	1.0	0.0	0.0

Apply Dismiss

Edit Function

$p(\text{Res} \mid \text{Price}, \text{Pat})$

Values for parents:

Pat:

Price	1	2	3
true	0.0	1.0	1.0
false	1.0	0.0	0.0

Apply Dismiss

Edit Function

$p(\text{Res} \mid \text{Price}, \text{Pat})$

Values for parents:

Pat:

Price	1	2	3
true	0.0	0.0	1.0
false	1.0	0.0	0.0

Apply Dismiss

Edit Function

$p(\text{Wait} \mid \text{Bar}, \text{Res})$

Values for parents:

Res:

Bar	true	false
true	0.33	0.67
false	0.67	0.33

Apply Dismiss

Edit Function

$p(\text{Wait} \mid \text{Bar}, \text{Res})$

Values for parents:

Res:

Bar	true	false
true	0.67	0.33
false	0.33	0.67

Apply Dismiss

Edit Function

$p(\text{Bar} \mid \text{Alt})$

Alt	true	false
true	0.33	0.67
false	0.67	0.33

Apply Dismiss

Edit Function

$p(\text{Pat} \mid \text{Rain})$

Rain	true	false
None	0.33	0.11
Some	0.67	0.22
Full	0.0	0.67

Apply Dismiss

3. Výpočet třech zadaných pravděpodobností pomocí počítače

Výpočty jsem provedla pomocí nástroje JavaBayes. Výsledky jsou vloženy formou obrázku.

P(Est) Posterior distribution:
 probability ("Est") { //1 variable(s) and 3 values
 table
 0.3647683458694141 // p(0az10 | evidence)
 0.5818721951836168 // p(nad30 | evidence)
 0.05335945894696923; // p(10az30 | evidence);}

P(Est / Pat)

Pat = None Posterior distribution:
 probability ("Est") { //1 variable(s) and 3 values
 table
 1.0 // p(0az10 | evidence)
 0.0 // p(nad30 | evidence)
 0.0; // p(10az30 | evidence);}

Pat = Some Posterior distribution:
 probability ("Est") { //1 variable(s) and 3 values
 table
 1.0 // p(0az10 | evidence)
 0.0 // p(nad30 | evidence)
 0.0; // p(10az30 | evidence);}

Pat = Full Posterior distribution:
 probability ("Est") { //1 variable(s) and 3 values
 table
 0.0 // p(0az10 | evidence)
 0.916 // p(nad30 | evidence)
 0.084; // p(10az30 | evidence);}

P(Rain) Posterior distribution:
 probability ("Rain") { //1 variable(s) and 2 values
 table
 0.25 // p(true | evidence)
 0.75; // p(false | evidence);}

4. Výpočet poslední pravděpodobnosti $P(\text{Rain} \mid \text{Fri})$ ručně

V BS není žádná podmíněná vazba mezi Rain a Fri, tzn. není určena pravděpodobnost, že bude pršet právě v pátek. Tzn., že hodnoty pro $P(\text{Rain} \mid \text{Fri})$ budou stejné jako pro $P(\text{Rain})$:

```

P(Rain)    Posterior distribution:
            probability ( "Rain" ) { //1 variable(s) and 2 values
            table
                0.25          // p(true | evidence )
                0.75;         // p(false | evidence );}

```

Ruční výpočet pravděpodobnosti $P(\text{Rain} \mid \text{Fri})$ pak bude:

```

P(Rain = true | Fri)  = P(Rain = true)  = 0.25
P(Rain = false | Fri) = P(Rain = false) = 0.75

```

5. Porovnání výsledků 1 s 2 a dále 3 s 4

Ovlivňuje dotazovanou proměnou informace o proměnné v podmínce?

Z výpočtů 1 a 2, tj. $P(\text{Est})$ a $P(\text{Est} \mid \text{Pat})$ je patrné, že podmíněný vztah mezi Est a Pat ovlivnil výsledky, které jsou u výpočtů 1 a 2 jsou rozdílné.

U výpočtů 3 a 4, tj. $P(\text{Rain})$ a $P(\text{Rain} \mid \text{Fri})$ jsou výsledky stejné, protože mezi Rain a Fri není žádná podmíněná závislost.

6. Význam využití BS při výpočtu podmíněných pravděpodobností.

Přímý výpočet podmíněných pravděpodobností z dat může být v některých případech zbytečně rozsáhlý, protože některé proměnné nejsou v konečném důsledku důležité. Na počátku výpočtu je ale nelze v tabulce rozeznat. Použitím Bayesovské sítě lze počet výpočtů velmi snadno zoptimalizovat.

Zdroje:

[1] *Stránky předmětu A7B36VYD: Vytěžování dat* [online]. Poslední aktualizace neuvedena [cit. 2011-11-25]. Dostupné z WWW <<http://cw.felk.cvut.cz/doku.php/courses/a7b36vyd/cviceni/start>>.