

Wnioskowanie Probabilistyczne

Na podstawie: AIMA, ch14

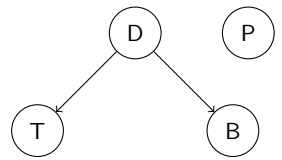
Wojciech Jaśkowski

Instytut Informatyki,
Politechnika Poznańska

8 kwietnia 2016

Na poprzednim wykładzie

	<i>ból</i>		\neg <i>ból</i>	
	<i>test</i>	\neg <i>test</i>	<i>test</i>	\neg <i>test</i>
<i>dziura</i>	0.108	0.012	0.072	0.008
\neg <i>dziura</i>	0.016	0.064	0.144	0.576



- 1 Łączny rozkład prawd. pozwala odpowiedzieć na każde pytanie. Ale:
 - liczba „danych” (wart. prawd) rośnie wykładniczo z liczbą zmiennych losowych
 - podawanie tych prawd. jest nienaturalne

2016-04-08

Wnioskowanie Probabilistyczne

↳ Wstęp

↳ Na poprzednim wykładzie

Na poprzednim wykładzie

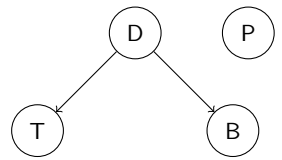
	<i>ból</i>	\neg <i>ból</i>
	<i>test</i>	\neg <i>test</i>
<i>dziura</i>	0.108	0.012
\neg <i>dziura</i>	0.016	0.064

• Łączny rozkład prawd. pozwala odpowiedzieć na każde pytanie. Ale:

- liczba „danych” (wart. prawd) rośnie wykładniczo z liczbą zmiennych losowych
- podawanie tych prawd. jest nienaturalne

Na poprzednim wykładzie

	<i>ból</i>		\neg <i>ból</i>	
	<i>test</i>	\neg <i>test</i>	<i>test</i>	\neg <i>test</i>
<i>dziura</i>	0.108	0.012	0.072	0.008
\neg <i>dziura</i>	0.016	0.064	0.144	0.576



- 1 Łączny rozkład prawd. pozwala odpowiedzieć na każde pytanie. Ale:
 - liczba „danych” (wart. prawd) rośnie wykładniczo z liczbą zmiennych losowych
 - podawanie tych prawd. jest nienaturalne
- 2 **Niezależność i niezależność warunkowa** może znacznie zmniejszyć liczbę potrzebnych wartości:

$$P(P, D) = P(P)P(D)$$

$$P(B, T|D) = P(B|D)P(T|D)$$

2016-04-08

Wnioskowanie Probabilistyczne

↳ Wstęp

↳ Na poprzednim wykładzie

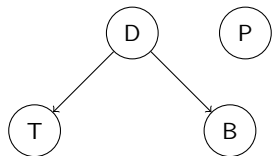
Na poprzednim wykładzie

	<i>ból</i>		\neg <i>ból</i>	
	<i>test</i>	\neg <i>test</i>	<i>test</i>	\neg <i>test</i>
<i>dziura</i>	0.108	0.012	0.072	0.008
\neg <i>dziura</i>	0.016	0.064	0.144	0.576

- Łączny rozkład prawd. pozwala odpowiedzieć na każde pytanie. Ale:
 - liczba „danych” (wart. prawd) rośnie wykładniczo z liczbą zmiennych losowych
 - podawanie tych prawd. jest nienaturalne
- **Niezależność i niezależność warunkowa** może znacznie zmniejszyć liczbę potrzebnych wartości:
 - $P(P, D) = P(P)P(D)$
 - $P(B, T|D) = P(B|D)P(T|D)$

Sieć baysowska

Definicja



Definicja:

- 1 Graf skierowany acykliczny (DAG)
 - Węzeł — zmienna losowa
 - Łuk $D \rightarrow T$ — D jest rodzicem T
- 2 Z każdym węzłem X_i związane jest prawd. warunkowe:

$$P(X_i | Rodzice(X_i))$$

2016-04-08

Wnioskowanie Probabilistyczne

└ Sieć baysowska

└ Sieć baysowska



Definicja:

- Graf skierowany acykliczny (DAG)
 - Węzeł — zmienna losowa
 - Łuk $D \rightarrow T$ — D jest rodzicem T
- Z każdym węzłem X_i związane jest prawd. warunkowe:

$$P(X_i | Rodzice(X_i))$$

Sieć baysowska

Interpretacja

- $X_i \rightarrow X_j$ mówi, że X_i ma **bezpośredni wpływ na X_j** .
- Rozkład prawd. łącznego:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Rodzice(X_i))$$

2016-04-08

Wnioskowanie Probabilistyczne

└ Sieć baysowska

└ Sieć baysowska

1. Ekspertowi w danej dziedzinie łatwo ten bezpośredni wpływ określić. Trudniej za to określić same prawd.

Sieć baysowska

- $X_i \rightarrow X_j$ mówi, że X_i ma **bezpośredni wpływ na X_j** .
- Rozkład prawd. łącznego:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | Rodzice(X_i))$$

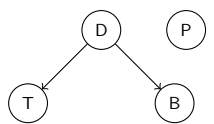
Sieć baysowska

Interpretacja

- $X_i \rightarrow X_j$ mówi, że X_i ma **bezpośredni wpływ** na X_j .
- Rozkład prawd. łącznego:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Rodzice}(X_i))$$

- Przykład:



$$\begin{aligned}
 P(D, T, B, P) &= P(P)P(D, T, B) \text{ (niezależność)} \\
 &= P(P)P(T, B|D)P(D) \text{ (reguła produkcji)} \\
 &= P(P)P(B|D)P(T|D)P(D) \text{ (niezależność warunkowa)}
 \end{aligned}$$

2016-04-08

Wnioskowanie Probabilistyczne

- └ Sieć baysowska
- └ Sieć baysowska

Sieć baysowska

- $X_i \rightarrow X_j$ mówi, że X_i ma **bezpośredni wpływ** na X_j .
- Rozkład prawd. łącznego:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{Rodzice}(X_i))$$
- Przykład:
- $P(D, T, B, P) = P(P)P(D, T, B)$ (niezależność)
- $= P(P)P(T, B|D)P(D)$ (reguła produkcji)
- $= P(P)P(B|D)P(T|D)P(D)$ (niezależność warunkowa)

1. Ekspertowi w danej dziedzinie łatwo ten bezpośredni wpływ określić. Trudniej za to określić same prawd.

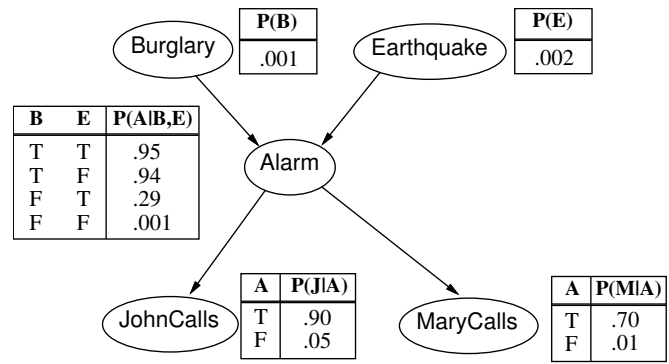
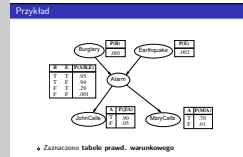
Przykład

2016-04-08

Wnioskowanie Probabilistyczne

└ Sieć baysowska

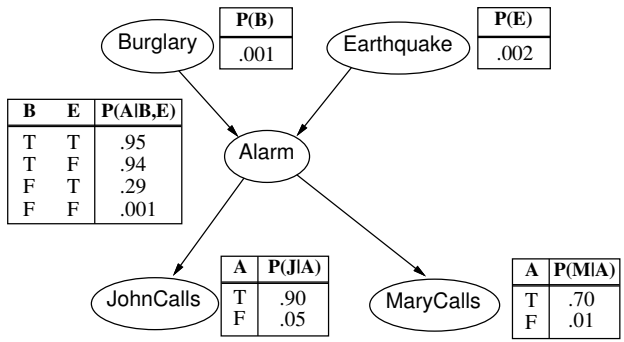
└ Przykład



- Zaznaczono **tabele prawd. warunkowego**

1. Widać naszą ignorancję i lenistwo: Nie ma prawd., że Marysia słucha głośnej muzyki. To jest już ukryte w węźle „Mary”. Czyli świat robi się stochastyczny ze względu na naszą ignorancję (lub brak wiedzy/danych)
2. Nie uwzględnia również wielu innych czynników

Przykład



$$P(B, J, M, E, A) = P(B)P(E)P(A|B, E)P(J|A)P(M|A)$$

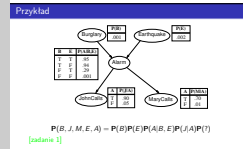
[zadanie 1]

2016-04-08

Wnioskowanie Probabilistyczne

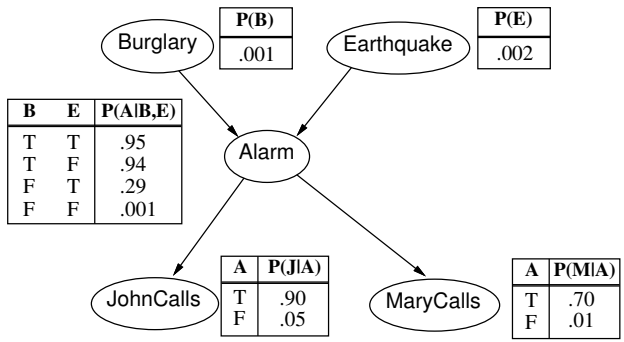
└ Sieć bayesowska

└ Przykład



1. Odp. $P(M|A)$
2. Odp. 0.000628
3. (Zauważmy, że J i M nie mają strzałki to i tak są zależne (marginalnie), tj. $J \not\perp M$, natomiast są niezależne warunkowo $J \perp M|A$).

Przykład



$$P(B, J, M, E, A) = P(B)P(E)P(A|B, E)P(J|A)P(M|A)$$

[zadanie 1]

$$P(\neg b, j, m, \neg e, a) = ?$$

[zadanie 2]

2016-04-08

Wnioskowanie Probabilistyczne

└ Sieć baysowska

└ Przykład

Przykład

$P(B, J, M, E, A) = P(B)P(E)P(A|B, E)P(J|A)P(M|A)$

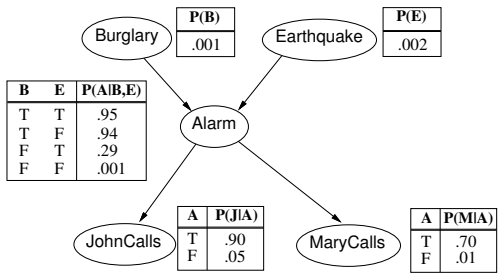
[zadanie 1]

$P(\neg b, j, m, \neg e, a) = ?$

[zadanie 2]

1. Odp. $P(M|A)$
2. Odp. 0.000628
3. (Zauważmy, że J i M nie mają strzałki to i tak są zależne (marginalnie), tj. $J \not\perp M$, natomiast są niezależne warunkowo $J \perp M|A$).

Niezależność warunkowa w sieciach



Które z poniższych zdań są prawdziwe? [zadanie 3]

- $P(M|J, A, E, B) = P(M|A, E, B)?$
- $= P(M|J, A)?$
- $= P(M|J, E)?$
- $= P(M|J, E, B)?$
- $= P(M|A)?$

2016-04-08

Wnioskowanie Probabilistyczne

└ Sieć baysowska

└ Niezależność warunkowa w sieciach

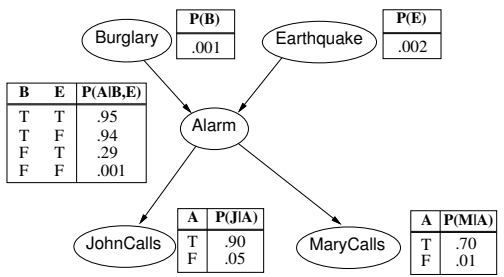
Niezależność warunkowa w sieciach

Które z poniższych zdań są prawdziwe? [zadanie 3]

- $P(M, J, A, E, B) = P(M|A, E, B)?$
- $= P(M|J, A)?$
- $= P(M|J, E)?$
- $= P(M|J, E, B)?$
- $= P(M|A)?$

1. $P(M|J, A, E, B) = P(M|A) = P(M|A, E, B) = P(M|J, A)$
2. I właśnie $P(M|J, A, E, B) = P(M|A)$ jest cechą sieci — wystarczy znać rodzica i reszta się wyjaśnia. Nawet więcej: koniecznie trzeba znać rodzica!

Ile wartości potrzeba?



- W tym przypadku? [zadanie 4]

2016-04-08

Wnioskowanie Probabilistyczne

└ Sieć baysowska

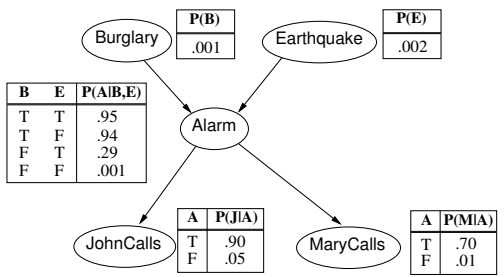
└ Ile wartości potrzeba?

1. 10
2. $\leq n2^k$



• W tym przypadku? [zadanie 4]

Ile wartości potrzeba?



- W tym przypadku?[\[zadanie 4\]](#)
- W ogólnym przypadku, jeśli mamy n węzłów i liczba rodziców ograniczona jest przez k ?[\[zadanie 5\]](#)

2016-04-08

Wnioskowanie Probabilistyczne

└ Sieć baysowska

└ Ile wartości potrzeba?

1. 10
2. $\leq n2^k$

- W tym przypadku?[\[zadanie 4\]](#)
- W ogólnym przypadku, jeśli mamy n węzłów i liczba rodziców ograniczona jest przez k ?[\[zadanie 5\]](#)

Niezależność zwykła vs. niezależność warunkowa

- $T_1 \perp T_2 | C \implies T_1 \perp T_2?$ [zadanie 6]

2016-04-08

Wnioskowanie Probabilistyczne

└ Niezależność warunkowa w sieciach

└ Niezależność zwykła vs. niezależność warunkowa

Niezależność zwykła vs. niezależność warunkowa

- $T_1 \perp T_2 | C \implies T_1 \perp T_2?$ [zadanie 6]

1. Nie. Przykład: rak i dwa testy. Jest niezależność warunkowa, ale nie ma niezależności absolutnej, bo $P(t_1|t_2) > P(t_1)$ (czyli, pozytywny wynik testu T_2 wpływa na prawd., że otrzymamy pozytywny wynik testu T_1).
2. Też nie. I zaraz to wyjaśnimy

Niezależność zwykła vs. niezależność warunkowa

- $T_1 \perp T_2 | C \implies T_1 \perp T_2?$ [zadanie 6]
- $T_1 \perp T_2 \implies T_1 \perp T_2 | C?$ [zadanie 7]

2016-04-08

Wnioskowanie Probabilistyczne

└ Niezależność warunkowa w sieciach

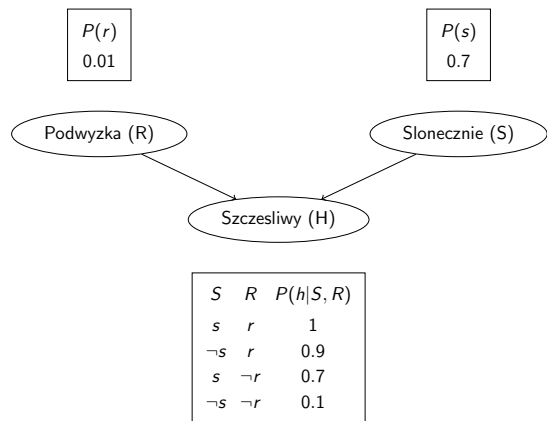
└ Niezależność zwykła vs. niezależność warunkowa

Niezależność zwykła vs. niezależność warunkowa

- $T_1 \perp T_2 | C \implies T_1 \perp T_2?$ [zadanie 6]
- $T_1 \perp T_2 \implies T_1 \perp T_2 | C?$ [zadanie 7]

1. Nie. Przykład: rak i dwa testy. Jest niezależność warunkowa, ale nie ma niezależności absolutnej, bo $P(t_1|t_2) > P(t_1)$ (czyli, pozytywny wynik testu T_2 wpływa na prawd., że otrzymamy pozytywny wynik testu T_1).
2. Też nie. I zaraz to wyjaśnimy

Dwie przyczyny



• $P(r|s) = ?$ [zadanie 8]

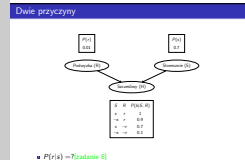
2016-04-08

Wnioskowanie Probabilistyczne

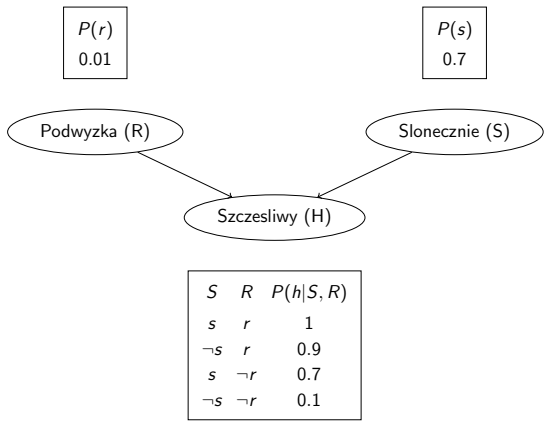
└ Niezależność warunkowa w sieciach

└ Dwie przyczyny

1. 0.01, bo R i S są niezależne (bezwarunkowo!)



Explaining away



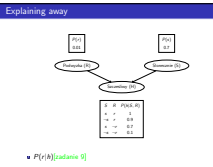
• $P(r|h)$ [zadanie 9]

2016-04-08

Wnioskowanie Probabilistyczne

Niezależność warunkowa w sieciach

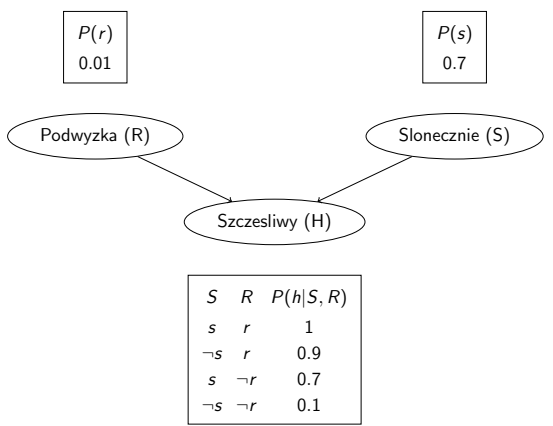
Explaining away



1.

$$\begin{aligned}
 \mathbf{P}(R|h) &= \alpha \mathbf{P}(R, h) = \alpha \sum_s \mathbf{P}(R, h, s) \text{ (marginalizacja)} \\
 &= \alpha \sum_s \mathbf{P}(h|s, R) \mathbf{P}(R, s) \text{ (reguła produkcji)} \\
 &= \alpha \sum_s \mathbf{P}(h|s, R) \mathbf{P}(R) \mathbf{P}(s) \text{ (niezależność)} \\
 &= \alpha \mathbf{P}(R) \sum_s \mathbf{P}(h|s, R) \mathbf{P}(s) \text{ (wyciągnięcie przed nawias)} \\
 &= \langle 0.01 \times (1 \times 0.7 + 0.9 \times 0.3), 0.99 \times (0.7 \times 0.7 + 0.1 \times 0.3) \rangle \\
 &= \langle 0.0185, 0.9815 \rangle
 \end{aligned}$$

Explaining away



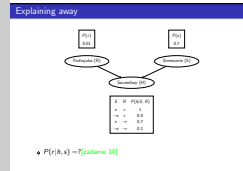
• $P(r|h, s) = ?$ [zadanie 10]

2016-04-08

Wnioskowanie Probabilistyczne

└ Niezależność warunkowa w sieciach

└ Explaining away

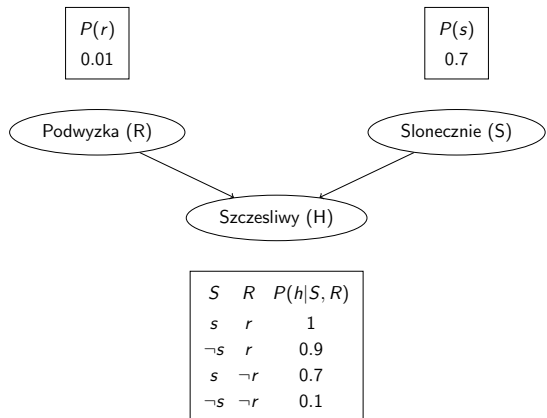


1.

$$\begin{aligned}
 \mathbf{P}(R|h, s) &= \alpha \mathbf{P}(h|R, s) \mathbf{P}(R|s) \text{ (ugólniony Bayes)} \\
 &= \alpha \mathbf{P}(h|R, s) \mathbf{P}(R) \text{ (niezależność } R \text{ i } S) \\
 &= \langle 0.0142, 0.9858 \rangle
 \end{aligned}$$

2. Analogicznie jak $P(r|h, s)$. Wychodzi: 0.0833.

Explaining away



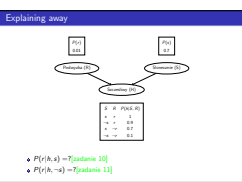
- $P(r|h, s) = ?$ [zadanie 10]
- $P(r|h, \neg s) = ?$ [zadanie 11]

2016-04-08

Wnioskowanie Probabilistyczne

↳ Niezależność warunkowa w sieciach

↳ Explaining away

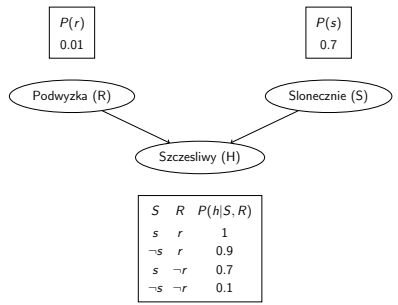


1.

$$\begin{aligned}
 \mathbf{P}(R|h, s) &= \alpha \mathbf{P}(h|R, s) \mathbf{P}(R|s) \text{ (ugólniony Bayes)} \\
 &= \alpha \mathbf{P}(h|R, s) \mathbf{P}(R) \text{ (niezależność } R \text{ i } S) \\
 &= \langle 0.0142, 0.9858 \rangle
 \end{aligned}$$

2. Analogicznie jak $P(r|h, s)$. Wychodzi: 0.0833.

Explaining away — podsumowanie



$$P(r|s) = P(r) = 0.01$$

$$P(r|h) = 0.0185$$

$$P(r|h, s) = 0.0142$$

$$P(r|h, \neg s) = 0.0833$$

- Jeśli jestem radosny, pogoda może wyjaśnić przyczynę radości. Jeśli jednak okazuje się, że jest słonecznie, to prawd., że dostałem podwyżkę maleje.
- Wniosek: R i S są **niezależne**, ale H powoduje, że **stają się zależne** (pod warunkiem, że H)!


2016-04-08

Wnioskowanie Probabilistyczne

Niezależność warunkowa w sieciach

Explaining away — podsumowanie

Explaining away — podsumowanie

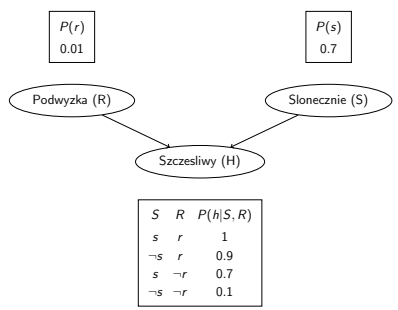


$P(r) = P(r) = 0.01$
 $P(r|h) = 0.0185$
 $P(r|h, s) = 0.0142$
 $P(r|h, \neg s) = 0.0833$

- Jeśli jestem radosny, pogoda może wyjaśnić przyczynę radości. Jeśli jednak okazuje się, że jest słonecznie, to prawd., że dostałem podwyżkę maleje.
- Wniosek: R i S są niezależne, ale H powoduje, że stają się zależne (pod warunkiem, że H)!

1. Z kolei jeśli jestem radosny i mówię, że nie dostałem podwyżki, to nie trzeba wyglądać za okno, żeby wiedzieć, że prawd. jest dziś słonecznie.

Explaining away — podsumowanie



$$P(r|s) = P(r) = 0.01$$

$$P(r|h) = 0.0185$$

$$P(r|h, s) = 0.0142$$

$$P(r|h, \neg s) = 0.0833$$

- Jeśli nie znamy H , to $P(r|s) = P(r) \rightarrow R$ i S są niezależne,
- Jeśli znamy H , to R i S stają się zależne ($P(r|h, s) \neq P(r|h)$)
- Wniosek: niezależność nie implikuje niezależności warunkowej.

2016-04-08

Wnioskowanie Probabilistyczne

Niezależność warunkowa w sieciach

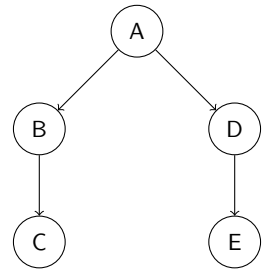
Explaining away — podsumowanie

Explaining away — podsumowanie

$P(r|s) = P(r) = 0.01$
 $P(r|h) = 0.0185$
 $P(r|h, s) = 0.0142$
 $P(r|h, \neg s) = 0.0833$

- Jeśli nie znamy H , to $P(r|s) = P(r) \rightarrow R$ i S są niezależne.
- Jeśli znamy H , to R i S stają się zależne ($P(r|h, s) \neq P(r|h)$)
- Wniosek: niezależność nie implikuje niezależności warunkowej.

Niezależność warunkowa — przykłady



- Czy:
 - $C \perp A$? [zadanie 12]
 - $C \perp A|B$? [zadanie 13]
 - $C \perp D$? [zadanie 14]
 - $C \perp D|A$? [zadanie 15]
 - $E \perp C|D$? [zadanie 16]

2016-04-08

Wnioskowanie Probabilistyczne

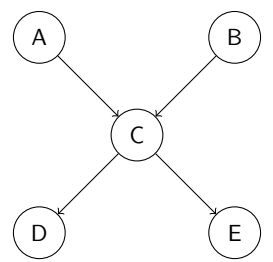
- └ Niezależność warunkowa w sieciach
- └ Niezależność warunkowa — przykłady

Niezależność warunkowa — przykłady

- Czy:
 - $C \perp A$? [zadanie 12]
 - $C \perp A|B$? [zadanie 13]
 - $C \perp D$? [zadanie 14]
 - $C \perp D|A$? [zadanie 15]
 - $E \perp C|D$? [zadanie 16]

1. Nie
2. Tak
3. Nie
4. Tak
5. Tak

Niezależność warunkowa — przykłady



- Czy:
 - $A \perp E?$ [zadanie 17]
 - $A \perp E|B?$ [zadanie 18]
 - $A \perp E|C?$ [zadanie 19]
 - $A \perp B?$ [zadanie 20]
 - $A \perp B|C?$ [zadanie 21]

2016-04-08

Wnioskowanie Probabilistyczne

└ Niezależność warunkowa w sieciach

└ Niezależność warunkowa — przykłady

Niezależność warunkowa — przykłady

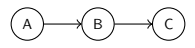
• Czy:

- $A \perp E?$ [zadanie 17]
- $A \perp E|B?$ [zadanie 18]
- $A \perp E|C?$ [zadanie 19]
- $A \perp B?$ [zadanie 20]
- $A \perp B|C?$ [zadanie 21]

1. Nie
2. Nie
3. Tak
4. Tak
5. Nie (explaining away!)

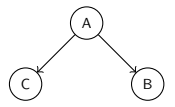
Niezależność warunkowa (zasady ogólne uproszczone)

Seryjne



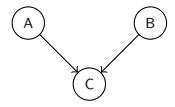
A i C są zależne, ale $A \perp C|B$

Rozchodzące się



B i C są zależne, ale $B \perp C|A$

Schodzące się



$A \perp B$, ale $A \not\perp B|C$

2016-04-08

Wnioskowanie Probabilistyczne

└ Niezależność warunkowa w sieciach

└ Niezależność warunkowa (zasady ogólne uproszczone)

Niezależność warunkowa (zasady ogólne uproszczone)

Seryjne

 A i C są zależne, ale $A \perp C|B$

Rozchodzące się

 B i C są zależne, ale $B \perp C|A$

Schodzące się

 $A \perp B$, ale $A \not\perp B|C$

1. podobnie w przypadku ogólnym: $A, B \rightarrow X_1 \dots \rightarrow \dots X_n \rightarrow C$

D-separacja (zasady ogólne)

X , Y i Z są zbiorami zmiennych losowych.

- 1 $X \perp Y | Z \iff X$ i Y są d-odseparowane.
- 2 X , Y są **d-odseparowane** jeśli każda ścieżka prowadząca od węzła $X \in X$ do węzła $Y \in Y$ jest „zablokowana”, czyli jeśli istnieje na niej jakiś węzeł $V \notin X \cup Y$, taki, że:
 - 1 Połączenie przez V jest typu **rozchodzącego się** lub **seryjnego**, oraz $V \in Z$
 - 2 Połączenie przez V jest typu **schodzącego się** oraz ani $V \notin Z$ ani żaden potomek węzła V w grafie $\notin Z$

Aplikacja do testowania d-seperowalności:
<http://www.aispace.org/downloads.shtml> (Belief and Decision Networks)

2016-04-08

Wnioskowanie Probabilistyczne

- └ Niezależność warunkowa w sieciach
- └ D-separacja (zasady ogólne)

- 1. Ważna uwaga: typ węzła V rozpatrujemy tylko w kontekście danej ścieżki (w ten sposób V nie może być równocześnie wielu typów).

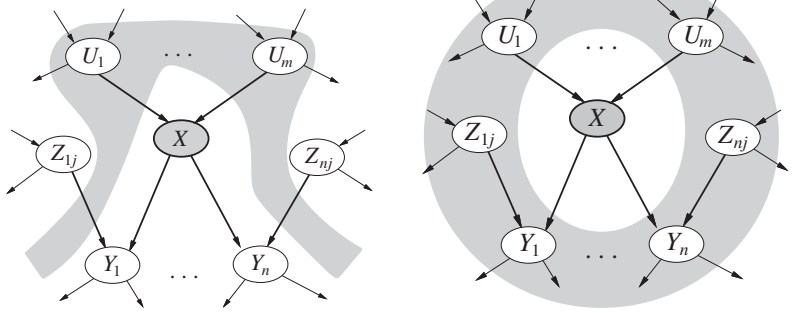
D-separacja (zasady ogólne)

X, Y i Z są zbiorami zmiennych losowych.

- 1 $X \perp Y | Z \iff X$ i Y są d-odseparowane.
- 2 X, Y są **d-odseparowane** jeśli każda ścieżka prowadząca od węzła $X \in X$ do węzła $Y \in Y$ jest „zablokowana”, czyli jeśli istnieje na niej jakiś węzeł $V \notin X \cup Y$, taki, że:
 - 1 Połączenie przez V jest typu **rozchodzącego się** lub **seryjnego**, oraz $V \in Z$
 - 2 Połączenie przez V jest typu **schodzącego się** oraz ani $V \notin Z$ ani żaden potomek węzła V w grafie $\notin Z$

Aplikacja do testowania d-seperowalności:
<http://www.aispace.org/downloads.shtml> (Belief and Decision Networks)

Właściwości D-separacji



- **Lewa:** X jest niezależny od swoich „niepotomków” (np. Z) pod warunkiem swoich rodziców (U)
- **Prawa (Koc Markowa):** X jest niezależny od pozostałych węzłów pod warunkiem swojej rodziny, tj. swoich rodziców (U), swoich dzieci (Y) oraz małżonków (czyli rodziców swoich dzieci (Z)).

2016-04-08

Wnioskowanie Probabilistyczne

- └ Niezależność warunkowa w sieciach
- └ Właściwości D-separacji

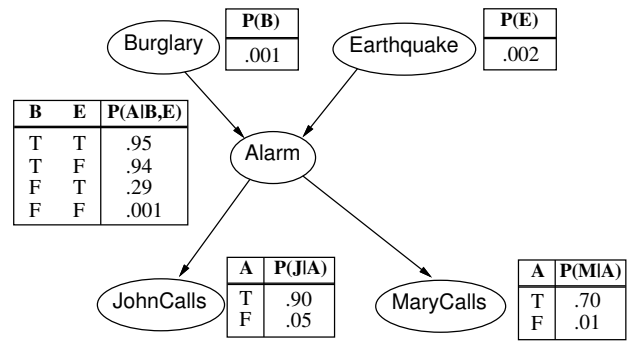
Właściwości D-separacji

• **Lewa:** X jest niezależny od swoich „niepotomków” (np. Z) pod warunkiem swoich rodziców (U)

• **Prawa (Koc Markowa):** X jest niezależny od pozostałych węzłów pod warunkiem swojej rodziny, tj. swoich rodziców (U), swoich dzieci (Y) oraz małżonków (czyli rodziców swoich dzieci (Z)).

1. Lewa: Z łączy się z X albo poprzez U albo poprzez Y . Jeśli poprzez U , to jest zablokowany w sposób naturalny (warunek a). Jeśli poprzez Y , to jest zablokowany ze względu na rozchodzenie się (warunek b). Należy zwrócić uwagę fakt iż, w tym przypadku także X jest niezależne (bezwarunkowo) od Z (o ile oczywiście nie ma krawędzi $U \rightarrow Z$, ale to zupełnie inna historia).
2. Prawa: Dzięki temu, że Z należy do koca, wyeliminowano wszystkie potencjalne V typu schodzącego się, które nie należą do Z .

Wnioskowanie



- Pytanie:
 - $P(J, M | \neg b, e)$
 - B, E : obserwowane, A : ukryte, J, M : zapytanie

2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania
- Wnioskowanie

Wnioskowanie

Pytanie:

- $P(J, M | \neg b, e)$
- B, E : obserwowane, A : ukryte, J, M : zapytanie

Wnioskowanie

W ogólności pytania, które możemy stawiać mają postać:

1 Rozkład łączny

$$P(Q_1, Q_2, \dots | E_1 = e_1, E_2 = e_2)$$

- wynik: łączny rozkład a posteriori (znając wartości \mathbf{e})

2 Estymacja największej wiarygodności (ang. *maximum likelihood estimation*):

$$\operatorname{argmax}_{\mathbf{q}} P(Q_1 = q_1, Q_2 = q_2, \dots | E_1 = e_1, E_2 = e_2)$$

- wynik: wektor wartości \mathbf{q} (wartości najbardziej prawdopodobne).

2016-04-08

Wnioskowanie Probabilistyczne

Metody wnioskowania

Wnioskowanie

Wnioskowanie

W ogólności pytania, które możemy stawiać mają postać:

- Rozkład łączny**

$$P(Q_1, Q_2, \dots | E_1 = e_1, E_2 = e_2)$$
 - wynik: łączny rozkład a posteriori (znając wartości \mathbf{e})
- Estymacja największej wiarygodności** (ang. *maximum likelihood estimation*):

$$\operatorname{argmax}_{\mathbf{q}} P(Q_1 = q_1, Q_2 = q_2, \dots | E_1 = e_1, E_2 = e_2)$$
 - wynik: wektor wartości \mathbf{q} (wartości najbardziej prawdopodobne).

Enumeracja

- Obliczamy wprost, np.

$$P(b|j, m) = P(b, j, m) / P(j, m)$$

2016-04-08

Wnioskowanie Probabilistyczne

└ Metody wnioskowania

└ Enumeracja

Enumeracja

- Obliczamy wprost, np.

$$P(b|j, m) = P(b, j, m) / P(j, m)$$

1. Uwaga: tutaj b, j i m są stałymi, ale e i a zmiennymi (wynika to z kontekstu)

Enumeracja

- Obliczamy wprost, np.

$$P(b|j, m) = P(b, j, m) / P(j, m)$$

- Obliczmy np. $P(b, j, m)$:

$$\begin{aligned}
 P(b, j, m) &= \sum_{e \in E} \sum_{a \in A} P(b, j, m, e, a) \text{ (marginalizacja)} \\
 &= \sum_{e \in E} \sum_{a \in A} P(b)P(e)P(a|b, e)P(j|a)P(m|a)
 \end{aligned}$$

2016-04-08

Wnioskowanie Probabilistyczne

- └ Metody wnioskowania
 - └ Enumeracja

Enumeracja

- Obliczamy wprost, np.

$$P(b|j, m) = P(b, j, m) / P(j, m)$$
- Obliczmy np. $P(b, j, m)$:

$$\begin{aligned}
 P(b, j, m) &= \sum_{e \in E} \sum_{a \in A} P(b, j, m, e, a) \text{ (marginalizacja)} \\
 &= \sum_{e \in E} \sum_{a \in A} P(b)P(e)P(a|b, e)P(j|a)P(m|a)
 \end{aligned}$$

1. Uwaga: tutaj b, j i m są stałymi, ale e i a zmiennymi (wynika to z kontekstu)

Enumeracja

Koszt obliczeniowy

$$P(b, j, m) = \sum_{e \in E} \sum_{a \in A} P(b)P(e)P(a|b, e)P(j|a)P(m|a)$$

- Zmienne są binarne, więc mamy $2 \times 2 = 4$ „wierszy” do policzenia
 ⇒ w ogólności: $O(2^n)$ w najgorszym przypadku przy n zmiennych binarnych.
- Kiedy występuje najgorszy przypadek? [zadanie 22]

2016-04-08

Wnioskowanie Probabilistyczne

└ Metody wnioskowania

└ Enumeracja

$$P(b, j, m) = \sum_{e \in E} \sum_{a \in A} P(b)P(e)P(a|b, e)P(j|a)P(m|a)$$

- Zmienne są binarne, więc mamy $2 \times 2 = 4$ „wierszy” do policzenia
 ⇒ w ogólności: $O(2^n)$ w najgorszym przypadku przy n zmiennych binarnych.
- Kiedy występuje najgorszy przypadek? [zadanie 22]

1. To zależy od liczby zmiennych ukrytych, więc najgorszy przypadek jest wtedy, gdy szukamy prawd. marginalnego, np. $P(m)$

Przyspieszanie enumeracji wyciąganie przed nawias

- Niektóre wartości można wyciągnąć przed nawias:

$$\begin{aligned}
 P(b, j, m) &= \sum_e \sum_a P(b)P(e)P(a|b, e)P(j|a)P(m|a) \\
 &= P(b) \sum_e P(e) \sum_a P(a|b, e)P(j|a)P(m|a)
 \end{aligned}$$

2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania

- Przyspieszanie enumeracji

- Niektóre wartości można wyciągnąć przed nawias:

$$\begin{aligned}
 P(b, j, m) &= \sum_e \sum_a P(b)P(e)P(a|b, e)P(j|a)P(m|a) \\
 &= P(b) \sum_e P(e) \sum_a P(a|b, e)P(j|a)P(m|a)
 \end{aligned}$$

Przyspieszanie enumeracji wyciąganie przed nawias

- Niektóre wartości można wyciągnąć przed nawias:

$$\begin{aligned}
 P(b, j, m) &= \sum_e \sum_a P(b)P(e)P(a|b, e)P(j|a)P(m|a) \\
 &= P(b) \sum_e P(e) \sum_a P(a|b, e)P(j|a)P(m|a)
 \end{aligned}$$

- Niewielkie przyspieszenie → nadal zbyt wiele „wierszy”

2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania

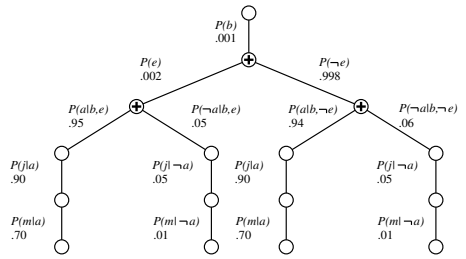
- Przyspieszanie enumeracji

- Niektóre wartości można wyciągnąć przed nawias:

$$P(b, j, m) = \sum_e \sum_a P(b)P(e)P(a|b, e)P(j|a)P(m|a) = P(b) \sum_e P(e) \sum_a P(a|b, e)P(j|a)P(m|a)$$
- Niewielkie przyspieszenie → nadal zbyt wiele „wierszy”

Przyspieszanie enumeracji eliminacja zmiennych

$$P(b, j, m) = P(b) \sum_e P(e) \sum_a P(a|b, e) P(j|a) P(m|a)$$



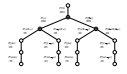
2016-04-08

Wnioskowanie Probabilistyczne

└ Metody wnioskowania

└ Przyspieszanie enumeracji

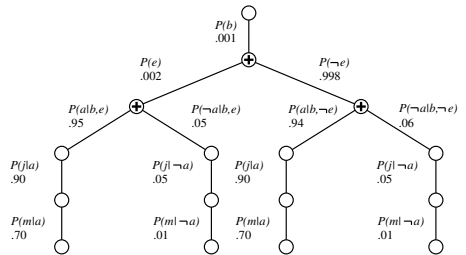
$$P(b, j, m) = P(b) \sum_e P(e) \sum_a P(a|b, e) P(j|a) P(m|a)$$



1. Niestety, to nie zmienia liczby wierszy

Przyspieszanie enumeracji eliminacja zmiennych

$$P(b, j, m) = P(b) \sum_e P(e) \sum_a P(a|b, e) P(j|a) P(m|a)$$



Obliczenia się powtarzają, np. $P(j|a)P(m|a)$
 ⇒ eliminacja zmiennych poprzez **programowanie dynamiczne**
 (spamiętywanie wyników pośrednich)

2016-04-08

Wnioskowanie Probabilistyczne

└ Metody wnioskowania

└ Przyspieszanie enumeracji

$P(b, j, m) = P(b) \sum_e P(e) \sum_a P(a|b, e) P(j|a) P(m|a)$

Obliczenia się powtarzają, np. $P(j|a)P(m|a)$
 ⇒ eliminacja zmiennych poprzez **programowanie dynamiczne**
 (spamiętywanie wyników pośrednich)

1. Niestety, to nie zmienia liczby wierszy

Próbkowanie bezpośrednie (ang. *direct sampling*)

- Estymuje łączny rozkład prawdopodobieństwa
- Algorytm:
 - Wylosuj N próbek zgodnie ze znanymi prawdopodobieństwami.
 - Liczbę pasujących próbek podziel przez N

2016-04-08

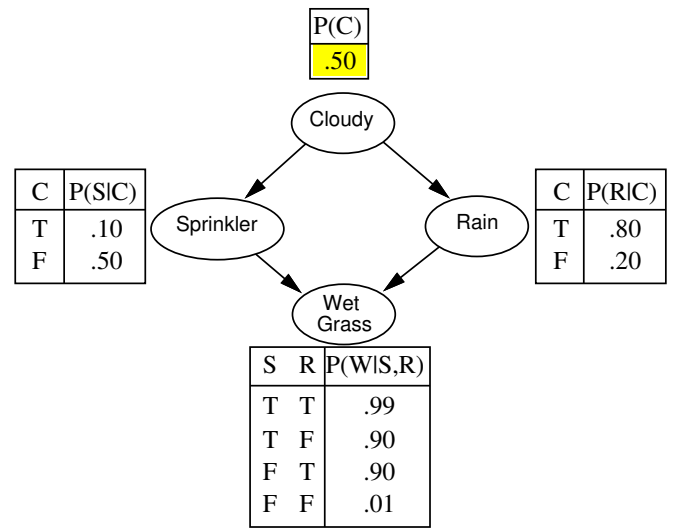
Wnioskowanie Probabilistyczne

└─ Metody wnioskowania

└─ Próbkowanie bezpośrednie (ang. *direct sampling*)

- Estymuje łączny rozkład prawdopodobieństwa
- Algorytm:
 - Wylosuj N próbek zgodnie ze znanymi prawdopodobieństwami.
 - Liczbę pasujących próbek podziel przez N

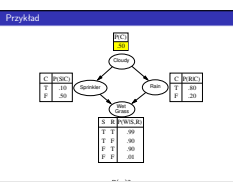
Przykład



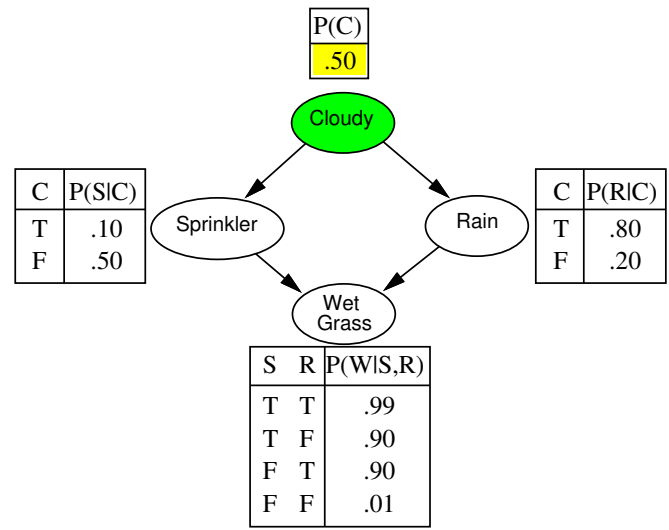
2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania
- Przykład



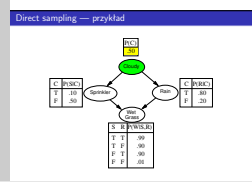
Direct sampling — przykład



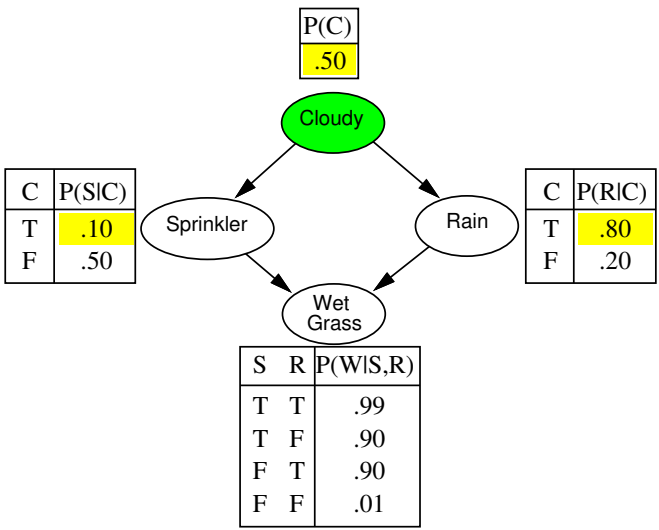
2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania
 - Direct sampling — przykład



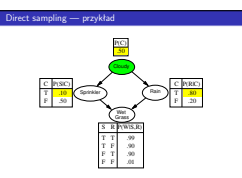
Direct sampling — przykład



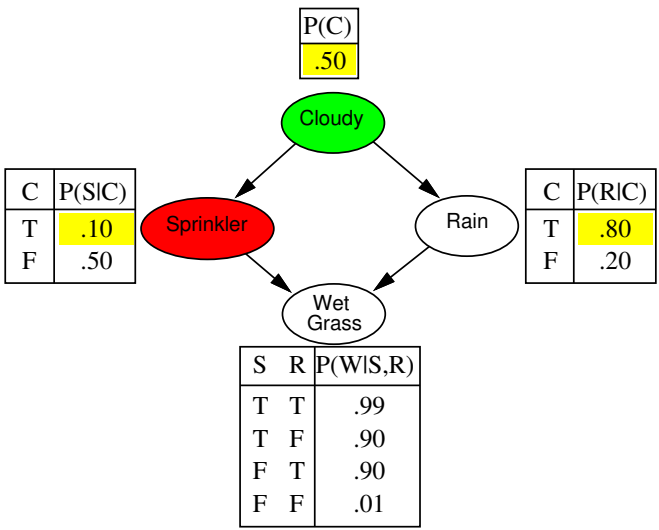
2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania
 - Direct sampling — przykład



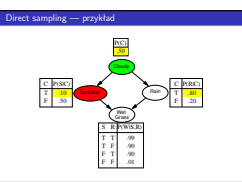
Direct sampling — przykład



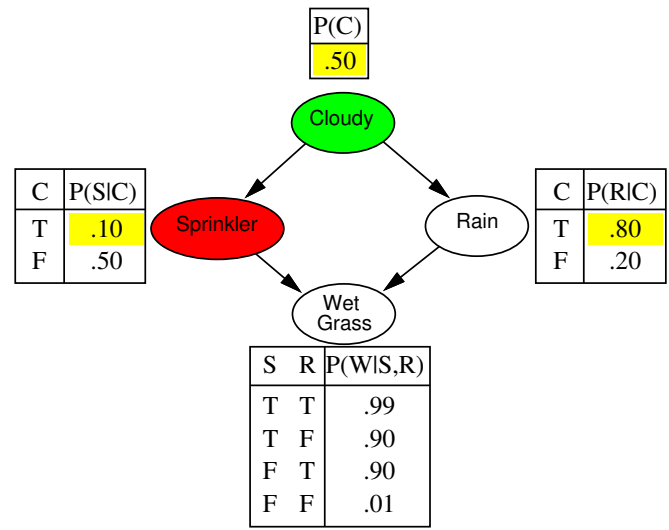
2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania
 - Direct sampling — przykład



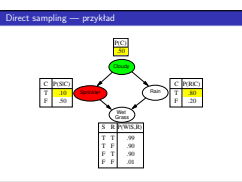
Direct sampling — przykład



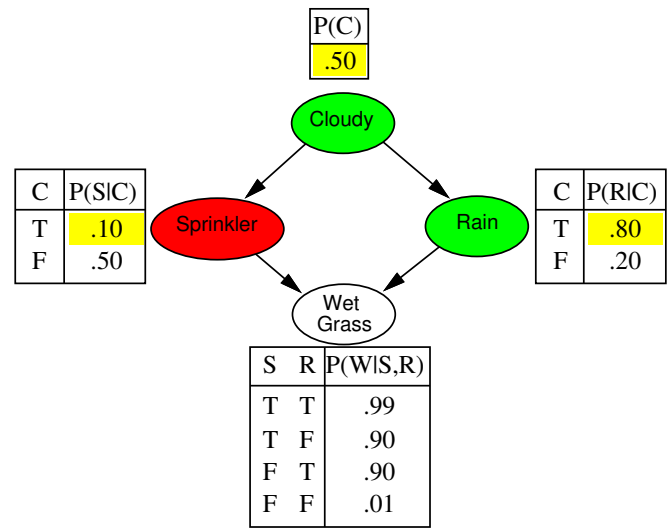
2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania
 - Direct sampling — przykład



Direct sampling — przykład

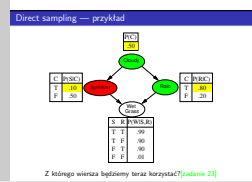


Z którego wiersza będziemy teraz korzystać? [zadanie 23]

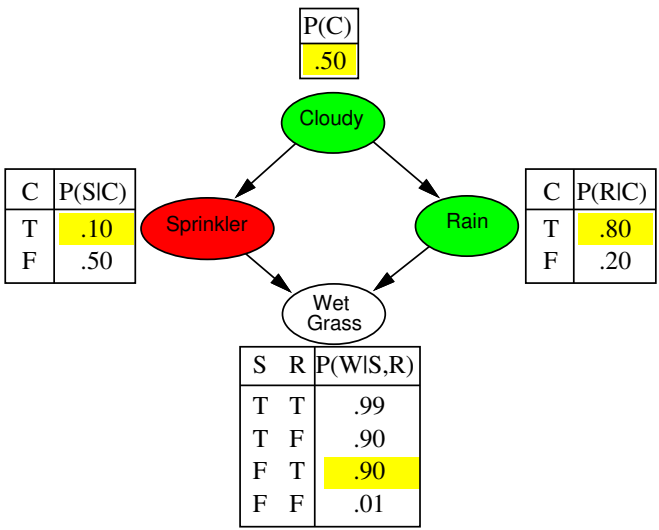
2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania
 - Direct sampling — przykład



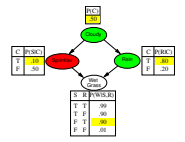
Direct sampling — przykład



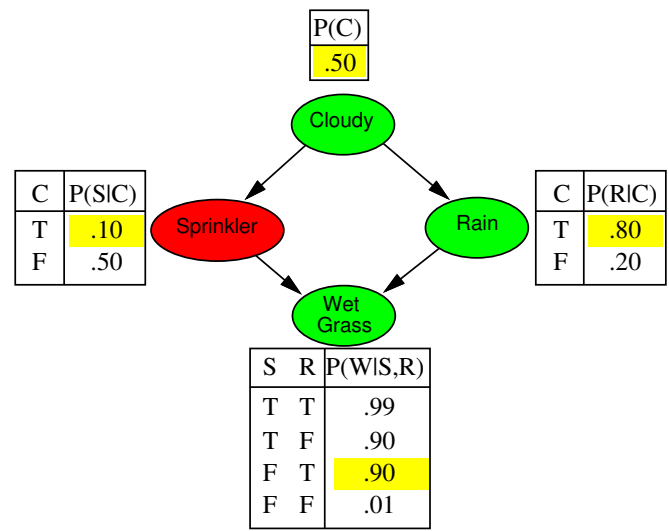
2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania
 - Direct sampling — przykład



Direct sampling — przykład

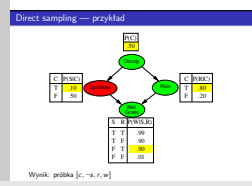


Wynik: próbka [c, ¬s, r, w]

2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania
 - Direct sampling — przykład



Direct sampling — przykład

Wyniki kolejnych próbkowań:

- 1 [c, ¬s, r, w]
- 2 [¬c, s, ¬r, w]
- 3 [c, ¬s, r, ¬w]
- 4 [c, ¬s, ¬r, ¬w]
- 5 [¬c, ¬s, ¬r, ¬w]
- 6 [¬c, ¬s, ¬r, ¬w]
- 7 [c, s, r, w]
- 8 [¬c, s, ¬r, w]
- 9 [c, ¬s, r, w]
- 10 [¬c, ¬s, r, w]

$P(w) = ?$

[zadanie 24]

2016-04-08

Wnioskowanie Probabilistyczne

Metody wnioskowania

Direct sampling — przykład

1. Wynik: 6 / 10

Direct sampling — przykład

Wyniki kolejnych próbkowań:

- [c, ¬s, r, w]
- [¬c, s, ¬r, w]
- [c, ¬s, r, ¬w]
- [c, ¬s, ¬r, ¬w]
- [¬c, ¬s, ¬r, ¬w]
- [¬c, ¬s, ¬r, ¬w]
- [c, s, r, w]
- [¬c, s, ¬r, w]
- [c, ¬s, r, w]
- [¬c, ¬s, r, w]

[zadanie 24] P(w) = ?

Direct sampling — zalety

- Zalety:
 - w krótkim czasie możemy uzyskać aproksymację łącznego rozkładu prawdopodobieństwa
 - możemy je zastosować nawet jeśli nie znamy prawdopodobieństw opisujących model (ale potrafimy symulować proces)

2016-04-08

Wnioskowanie Probabilistyczne

- └ Metody wnioskowania
 - └ Direct sampling — zalety

- Zalety:
 - w krótkim czasie możemy uzyskać aproksymację łącznego rozkładu prawdopodobieństwa
 - możemy je zastosować nawet jeśli nie znamy prawdopodobieństw opisujących model (ale potrafimy symulować proces)

1. Zaleta próbkowania jest taka, że dokładne wnioskowanie dla całego łącznego rozkładu prawdopodobieństwa ma złożoność wykładniczą, a tutaj możemy przynajmniej w skończonym czasie uzyskać jakieś aproksymacje.

Próbkowanie z odrzucaniem (ang. rejection sampling)

- Pozwala obliczyć rozkład prawdopodobieństwa warunkowego np. $P(C, R|s, w)$
- Idea:
 - 1 Wylosuj próbkę
 - 2 Odrzuć jeśli nie pasuje do wartości zmiennych warunkowych
- Wynik: liczba pasujących / liczba niepasujących

2016-04-08

Wnioskowanie Probabilistyczne

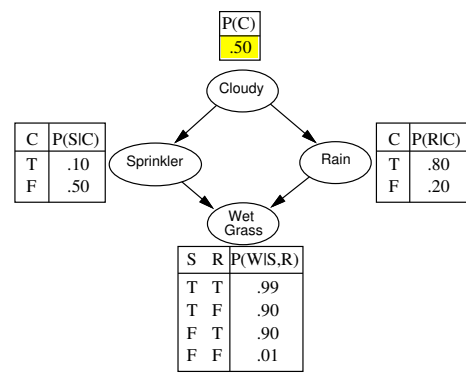
└─ Metody wnioskowania

└─ Próbkowanie z odrzucaniem (ang. rejection sampling)

Próbkowanie z odrzucaniem (ang. rejection sampling)

- Pozwala obliczyć rozkład prawdopodobieństwa warunkowego np. $P(C, R|s, w)$
- Idea:
 - 1 Wylosuj próbkę
 - 2 Odrzuć jeśli nie pasuje do wartości zmiennych warunkowych
- Wynik: liczba pasujących / liczba niepasujących

Przykład



$P(R|s) = ?$ [zadanie 25]

- 1 [c, ¬s, r, w]
- 2 [¬c, s, ¬r, w]
- 3 [c, ¬s, r, ¬w]
- 4 [c, ¬s, ¬r, ¬w]
- 5 [¬c, ¬s, ¬r, ¬w]
- 6 [¬c, ¬s, ¬r, w]
- 7 [c, s, r, w]
- 8 [¬c, s, ¬r, w]
- 9 [c, ¬s, r, w]
- 10 [¬c, ¬s, r, w]

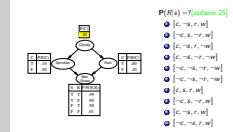
Wnioskowanie Probabilistyczne

Metody wnioskowania

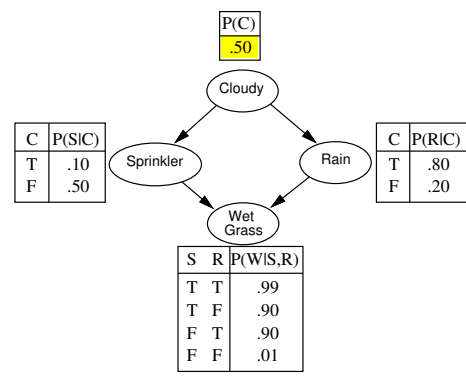
Przykład

2016-04-08

1. $\langle \frac{1}{3}, \frac{2}{3} \rangle$
2. Jeśli jakieś prawd. jest bardzo małe (np. prawd. włamania) albo bardzo duże musimy zebrać bardzo dużo próbek, żeby estymacja była dobra.
3. I niestety większość próbek musimy odrzucić



Przykład



$P(R|s) = ?$ [zadanie 25]

1. $[c, \neg s, r, w]$
2. $[\neg c, s, \neg r, w]$
3. $[c, \neg s, r, \neg w]$
4. $[c, \neg s, \neg r, \neg w]$
5. $[\neg c, \neg s, \neg r, \neg w]$
6. $[\neg c, \neg s, \neg r, w]$
7. $[c, s, r, w]$
8. $[\neg c, s, \neg r, w]$
9. $[c, \neg s, r, w]$
10. $[\neg c, \neg s, r, w]$

• Jakie są wady próbkowania z odrzucaniem? [zadanie 26]

2016-04-08

Wnioskowanie Probabilistyczne

- └ Metody wnioskowania
 - └ Przykład

Przykład

$P(R|s) = ?$ [zadanie 25]

- $[c, \neg s, r, w]$
- $[\neg c, s, \neg r, w]$
- $[c, \neg s, r, \neg w]$
- $[\neg c, s, \neg r, \neg w]$
- $[c, \neg s, \neg r, \neg w]$
- $[\neg c, \neg s, \neg r, \neg w]$
- $[c, s, r, w]$
- $[\neg c, s, \neg r, w]$
- $[c, \neg s, r, w]$
- $[\neg c, \neg s, r, w]$

• Jakie są wady próbkowania z odrzucaniem? [zadanie 26]

1. $\langle \frac{1}{3}, \frac{2}{3} \rangle$
2. Jeśli jakieś prawd. jest bardzo małe (np. prawd. włamania) albo bardzo duże musimy zebrać bardzo dużo próbek, żeby estymacja była dobra.
3. I niestety większość próbek musimy odrzucić

Ważone próbkowanie (ang. likelihood weighting)

- Ustalamy wartości zmiennych warunkowych \Rightarrow nie trzeba będzie ich odrzucać
- Wazymy próbki odpowiednimi prawdopodobieństwami

2016-04-08

Wnioskowanie Probabilistyczne

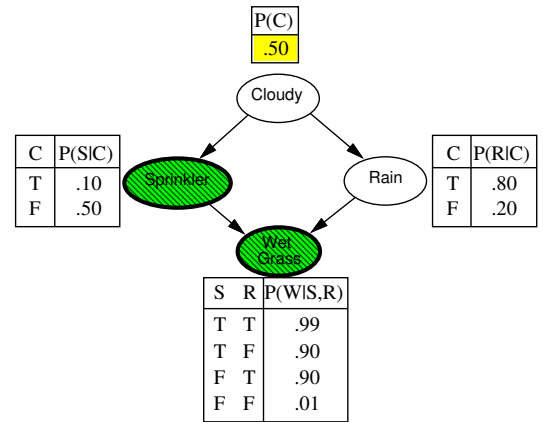
└ Metody wnioskowania

└ Ważone próbkowanie (ang. likelihood weighting)

- Ustalamy wartości zmiennych warunkowych \rightarrow nie trzeba będzie ich odrzucać
- Wazymy próbki odpowiednimi prawdopodobieństwami

Przykład

$$P(c, \neg r | s, w) = ?$$



$$w = 1.0$$

2016-04-08

Wnioskowanie Probabilistyczne

└ Metody wnioskowania

└ Przykład

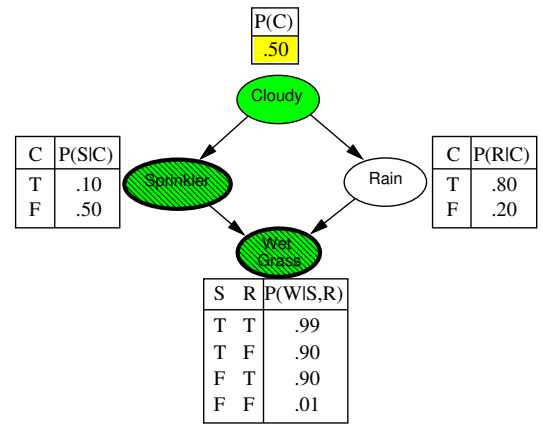
Przykład

$P(c, \neg r | s, w) = ?$

$w = 1.0$

Przykład

$$P(c, \neg r | s, w) = ?$$



$$w = 1.0$$

2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania

- Przykład

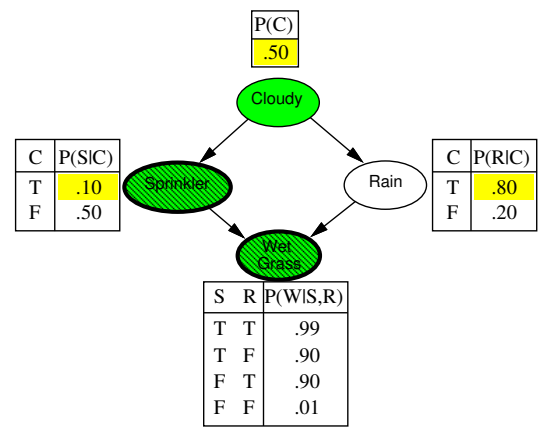
Przykład

$P(c, \neg r | s, w) = ?$

$w = 1.0$

Ważone próbkowanie — przykład

$$P(c, \neg r | s, w) = ?$$



w = 1.0

2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania
 - Ważone próbkowanie — przykład

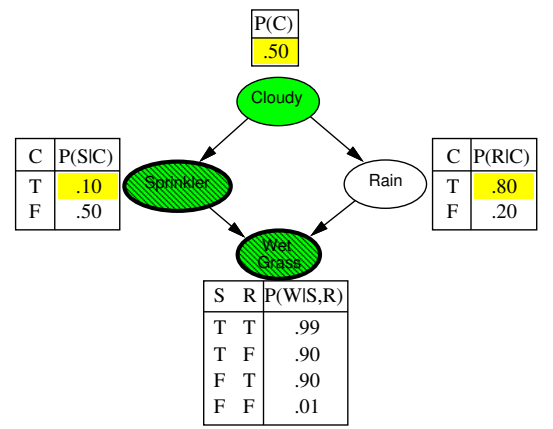
Ważone próbkowanie — przykład

$P(c, \neg r | s, w) = ?$

w = 1.0

Ważone próbkowanie — przykład

$$P(c, \neg r | s, w) = ?$$



$$w = 1.0 \times 0.1$$

2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania
 - Ważone próbkowanie — przykład

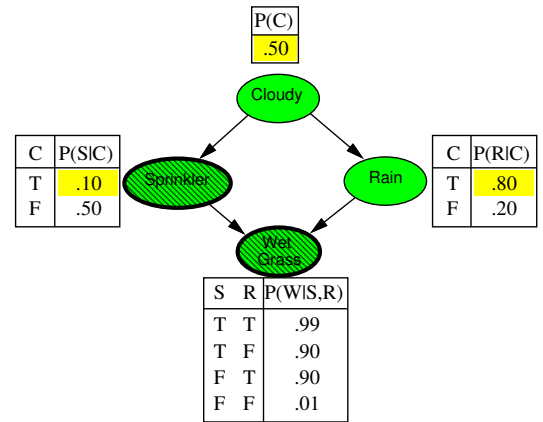
Ważone próbkowanie — przykład

$P(c, \neg r | s, w) = ?$

$w = 1.0 \times 0.1$

Ważone próbkowanie — przykład

$$P(c, \neg r | s, w) = ?$$



$w = 1.0 \times 0.1$ (Jaki będzie kolejny czynnik?) [zadanie 27]

2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania
- Ważone próbkowanie — przykład

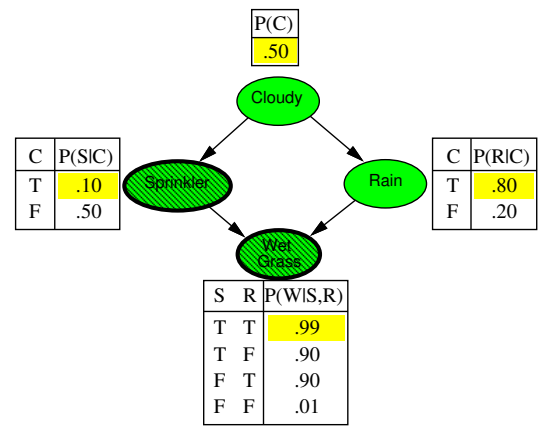
Ważone próbkowanie — przykład

$P(c, \neg r | s, w) = ?$

$w = 1.0 \times 0.1$ (Jaki będzie kolejny czynnik?) [zadanie 27]

Ważone próbkowanie — przykład

$$P(c, \neg r | s, w) = ?$$



2016-04-08

Wnioskowanie Probabilistyczne

- Metody wnioskowania
- Ważone próbkowanie — przykład

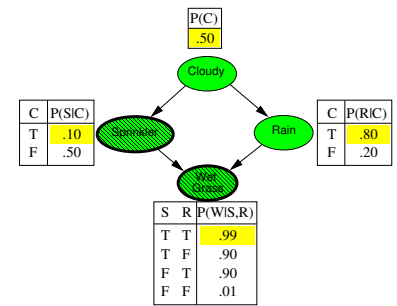
Ważone próbkowanie — przykład

$P(c, \neg r | s, w) = ?$

Ważone próbkowanie — przykład

- 1 $[c, s, r, w], w = 0.099$
- 2 $[\neg c, s, r, w], w = ?$
- 3 $[c, s, \neg r, w], w = ?$
- 4 $[c, s, \neg r, w], w = ?$
- 5 $[c, s, r, w], w = 0.099$
- 6 $[\neg c, s, \neg r, w], w = ?$
- 7 $[c, s, r, w], w = 0.099$
- 8 $[c, s, r, w], w = 0.099$
- 9 $[c, s, r, w], w = 0.099$
- 10 $[c, s, r, w], w = 0.099$

$P(c, \neg r | s, w) = ?$ [zadanie 28]



2016-04-08

Wnioskowanie Probabilistyczne

Metody wnioskowania

Ważone próbkowanie — przykład

1. $0.5 \times 0.99 = 0.495$
2. $0.1 \times 0.90 = 0.09$
3. $0.5 \times 0.90 = 0.45$
4. $(2 \times 0.09) / (6 \times 0.099 + 2 \times 0.09 + 0.495 + 0.45) = 0.1047$

Ważone próbkowanie — przykład

- 1 $[c, s, r, w], w = 0.099$
- 2 $[\neg c, s, r, w], w = ?$
- 3 $[c, s, \neg r, w], w = ?$
- 4 $[c, s, \neg r, w], w = ?$
- 5 $[c, s, r, w], w = 0.099$
- 6 $[\neg c, s, \neg r, w], w = ?$
- 7 $[c, s, r, w], w = 0.099$
- 8 $[c, s, r, w], w = 0.099$
- 9 $[c, s, r, w], w = 0.099$
- 10 $[c, s, r, w], w = 0.099$

$P(c, \neg r | s, w) = ?$ [zadanie 28]

Próbkowanie Gibbs'a – Markov Chain Monte Carlo (MCMC)

- „Stan” sieci = aktualne wartości węzłów
- Generuj następny stan próbując wartość jednej zmiennej pod warunkiem koca Markova
- Próbuj jedną zmienną naraz (kolejno albo losowo), nie ruszaj zmiennych warunkowych
- Po pewnej liczbie kroków oblicz proporcję interesujących stanów.

2016-04-08

Wnioskowanie Probabilistyczne

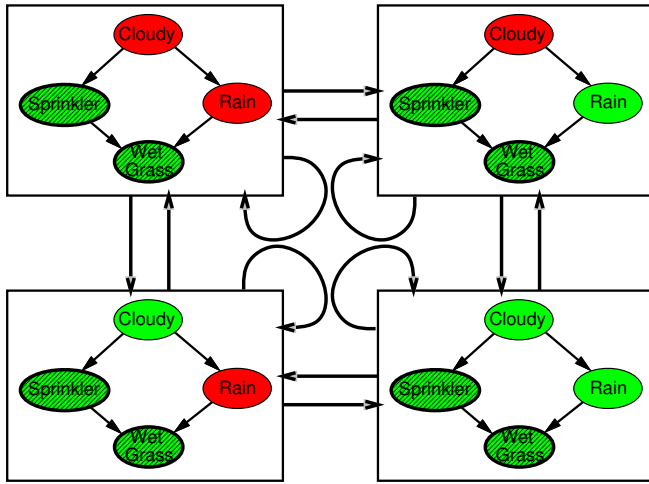
Metody wnioskowania

Próbkowanie Gibbs'a – Markov Chain Monte Carlo (MCMC)

- „Stan” sieci = aktualne wartości węzłów
- Generuj następny stan próbując wartość jednej zmiennej pod warunkiem koca Markova
- Próbuj jedną zmienną naraz (kolejno albo losowo), nie ruszaj zmiennych warunkowych
- Po pewnej liczbie kroków oblicz proporcję interesujących stanów.

Próbkowanie Gibbs'a – przykład

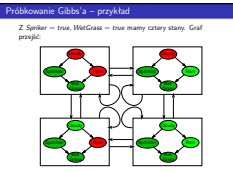
Z *Spriker = true*, *WetGrass = true* mamy cztery stany. Graf przejść:



2016-04-08

Wnioskowanie Probabilistyczne

- └ Metody wnioskowania
- └ Próbkowanie Gibbs'a – przykład



Próbkowanie Gibbs'a – przykład

Wyznaczmy $\mathbf{P}(Rain|Sprinkler = true, WetGrass = true)$:

- Powtarzaj: próbuj *Cloudy* lub *Rain* pod warunkiem koca Markov'a
- Policz $Rain = true|false$

Np. na 100 stanów mamy 31 $Rain = true$ i 69 $Rain = false$.

Wtedy:

$$\mathbf{P}(Rain|Sprinkler = true, WetGrass = true) = \langle 0.31, 0.69 \rangle$$

Twierdzenie: łańcuch dąży do rozkładu stacjonarnego: długoterminowa część czasu spędzona w każdym stanie jest proporcjonalna do jej rozkładu a posteriori.

2016-04-08

Wnioskowanie Probabilistyczne

Metody wnioskowania

Próbkowanie Gibbs'a – przykład

Próbkowanie Gibbs'a – przykład

Wyznaczmy $\mathbf{P}(Rain|Sprinkler = true, WetGrass = true)$:

- Powtarzaj: próbuj *Cloudy* lub *Rain* pod warunkiem koca Markov'a
- Policz $Rain = true|false$

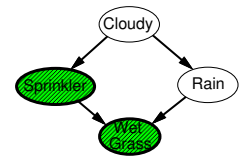
Np. na 100 stanów mamy 31 $Rain = true$ i 69 $Rain = false$.
Wtedy:

$$\mathbf{P}(Rain|Sprinkler = true, WetGrass = true) = \langle 0.31, 0.69 \rangle$$

Twierdzenie: łańcuch dąży do rozkładu stacjonarnego: długoterminowa część czasu spędzona w każdym stanie jest proporcjonalna do jej rozkładu a posteriori.

Próbkowanie Gibbs'a – jak próbkować?

- Kocem Markova dla *Cloudy* jest $\{Sprinkler, Rain\}$
- Kocem Markova dla *Rain* jest [zadanie 29]



Można łatwo pokazać, że:

$$P(x|mb(X)) = P(x|rodzice(X)) \prod_{Z \in \text{dzieci}(X)} P(z|rodzice(Z))$$

Np.

$$P(Cloudy|sprinkler, wetgrass) = P(Cloudy)P(sprinkler|Cloudy)P(rain|Cloudy)$$

2016-04-08

Wnioskowanie Probabilistyczne

Metody wnioskowania

Próbkowanie Gibbs'a – jak próbkować?

Próbkowanie Gibbs'a – jak próbkować?

- Kocem Markova dla *Cloudy* jest $\{Sprinkler, Rain\}$
- Kocem Markova dla *Rain* jest [zadanie 29]

Można łatwo pokazać, że:

$$P(x|mb(X)) = P(x|rodzice(X)) \prod_{Z \in \text{dzieci}(X)} P(z|rodzice(Z))$$

Np.

$$P(Cloudy|sprinkler, wetgrass) = P(Cloudy)P(sprinkler|Cloudy)P(rain|Cloudy)$$

Podsumowanie wnioskowania

- Wnioskowanie **dokładne** (eliminacja zmiennych):
 - w ogólności NP-trudne
 - bardzo wrażliwe na topologię sieci i pytania, które zadajemy
- Wnioskowanie **aproksymacyjne** (próbkiwanie):
 - generalnie, nie jest wrażliwe na topologię sieci
 - **próbkiwanie ważne**: zbieżność może być powolna, gdy prawd. mają skrajne wartości (0.0 lub 1.0)
 - **próbkiwanie Gibbs'a** ma wiele efektywnych rozszerzeń. Najlepsze, ale wady:
 - warunek stopu (zbieżność do rozkładu stacjonarnego)?
 - może być kosztowne gdy koc Markowa jest duży

2016-04-08

Wnioskowanie Probabilistyczne

Metody wnioskowania

Podsumowanie wnioskowania

Podsumowanie wnioskowania

- Wnioskowanie **dokładne** (eliminacja zmiennych):
 - w ogólności NP-trudne
 - bardzo wrażliwe na topologię sieci i pytania, które zadajemy
- Wnioskowanie **aproksymacyjne** (próbkiwanie):
 - generalnie, nie jest wrażliwe na topologię sieci
 - **próbkiwanie ważne**: zbieżność może być powolna, gdy prawd. mają skrajne wartości (0.0 lub 1.0)
 - **próbkiwanie Gibbs'a** ma wiele efektywnych rozszerzeń. Najlepsze, ale wady:
 - warunek stopu (zbieżność do rozkładu stacjonarnego)?
 - może być kosztowne gdy koc Markowa jest duży

Jak konstruować sieci?

- Najgorszy przypadek dla węzła:
 - to $O(2^k)$ liczb, gdzie k = liczba rodziców węzła.
- Zwykle jednak możemy zastosować jeden ze schematów. Rozważymy dwa z nich:
 - **węzeł deterministyczny**
 - **noisy-OR**

2016-04-08

Wnioskowanie Probabilistyczne └ Inżynieria sieci

└ Jak konstruować sieci?

- Najgorszy przypadek dla węzła:
 - to $O(2^k)$ liczb, gdzie k = liczba rodziców węzła.
- Zwykle jednak możemy zastosować jeden ze schematów. Rozważymy dwa z nich:
 - **węzeł deterministyczny**
 - **noisy-OR**

Węzeł deterministyczny

- Bez niepewności.
- Węzeł zdeterminowany w pełni przez rodziców (przez wyrażenie logiczne), np.
 - *Polak, Niemiec, Hiszpan* → *Europejczyk*

P	N	H	P(Europejczyk)
F	F	F	0.0
F	F	T	1.0
F	T	F	1.0
F	T	T	1.0
T	F	F	1.0
T	F	T	1.0
T	T	F	1.0
T	T	T	1.0

2016-04-08

Wnioskowanie Probabilistyczne

Inżynieria sieci

↳ Węzeł deterministyczny

Węzeł deterministyczny

- Bez niepewności.
- Węzeł zdeterminowany w pełni przez rodziców (przez wyrażenie logiczne), np.
 - *Polak, Niemiec, Hiszpan* → *Europejczyk*

P	N	H	P(Europejczyk)
F	F	F	0.0
F	F	T	1.0
F	T	F	1.0
F	T	T	1.0
T	F	F	1.0
T	F	T	1.0
T	T	F	1.0
T	T	T	1.0

Noisy-OR

- Uogólnienie logicznej alternatywy
- Przykład *Przeziębienie, Grypa, Malaria* → *Gorączka*
 - Zakładamy, że znamy wszystkie możliwe przyczyny (lub nowy węzeł *NieznanaPrzyczyna*)
 - Zakładamy, że związek $X \rightarrow Go$ może być w pewnym zakresie „hamowany”
 - **Hamowania są niezależne**, tzn. „hamowanie” $P \rightarrow Go$ jest niezależne od „hamowania” $G \rightarrow Go$.

2016-04-08

Wnioskowanie Probabilistyczne

- └ Inżynieria sieci
- └ Noisy-OR

Noisy-OR

- Uogólnienie logicznej alternatywy
- Przykład *Przeziębienie, Grypa, Malaria* → *Gorączka*
 - Zakładamy, że znamy wszystkie możliwe przyczyny (lub nowy węzeł *NieznanaPrzyczyna*)
 - Zakładamy, że związek $X \rightarrow Go$ może być w pewnym zakresie „hamowany”
 - **Hamowania są niezależne**, tzn. „hamowanie” $P \rightarrow Go$ jest niezależne od „hamowania” $G \rightarrow Go$.

Noisy-OR

- Uogólnienie logicznej alternatywy
- Przykład *Przeziębienie, Grypa, Malaria* \rightarrow *Gorączka*
 - Zakładamy, że znamy wszystkie możliwe przyczyny (lub nowy węzeł *NieznanaPrzyczyna*)
 - Zakładamy, że związek $X \rightarrow Go$ może być w pewnym zakresie „hamowany”
 - **Hamowania są niezależne**, tzn. „hamowanie” $P \rightarrow Go$ jest niezależne od „hamowania” $G \rightarrow Go$.
- Stąd: gorączka nie występuje $(\neg go) \iff \neg p \wedge \neg g \wedge \neg m$

2016-04-08

Wnioskowanie Probabilistyczne

Inżynieria sieci

Noisy-OR

Noisy-OR

- Uogólnienie logicznej alternatywy
- Przykład *Przeziębienie, Grypa, Malaria* \rightarrow *Gorączka*
 - Zakładamy, że znamy wszystkie możliwe przyczyny (lub nowy węzeł *NieznanaPrzyczyna*)
 - Zakładamy, że związek $X \rightarrow Go$ może być w pewnym zakresie „hamowany”
 - **Hamowania są niezależne**, tzn. „hamowanie” $P \rightarrow Go$ jest niezależne od „hamowania” $G \rightarrow Go$.
- Stąd: gorączka nie występuje $(\neg go) \iff \neg p \wedge \neg g \wedge \neg m$

Noisy-OR: Przykład

- Załóżmy, że prawd. „hamowania” są następujące:
 - $P(\neg go|p, \neg g, \neg m) = 0.6$
 - $P(\neg go|\neg p, g, \neg m) = 0.2$
 - $P(\neg go|\neg p, \neg g, m) = 0.1$
- Na tej podstawie budujemy całą tabelkę:

P	G	M	P(<i>Gorączka</i>)	P(\neg <i>Gorączka</i>)
F	F	F	0.0	1.0
F	F	T	0.9	0.1
F	T	F	0.8	0.2
F	T	T	0.98	0.02=0.2*0.1
T	F	F	0.4	0.6
T	F	T	0.94	0.06=0.6*0.1
T	T	F		?[zadanie 30]
T	T	T		?[zadanie 31]

2016-04-08

Wnioskowanie Probabilistyczne

↳ Inżynieria sieci

↳ Noisy-OR: Przykład

1. 0.88 i 0.12=0.6*0.2 oraz 0.988 i 0.012=0.6*0.2*0.1

Noisy-OR: Przykład

- Załóżmy, że prawd. „hamowania” są następujące:
 - $P(\neg go|p, \neg g, \neg m) = 0.6$
 - $P(\neg go|\neg p, g, \neg m) = 0.2$
 - $P(\neg go|\neg p, \neg g, m) = 0.1$
- Na tej podstawie budujemy całą tabelkę:

P	G	M	P(<i>Gorączka</i>)	P(\neg <i>Gorączka</i>)
F	F	F	0.0	1.0
F	F	T	0.9	0.1
F	T	F	0.8	0.2
F	T	T	0.98	0.02=0.2*0.1
T	F	F	0.4	0.6
T	F	T	0.94	0.06=0.6*0.1
T	T	F		?[zadanie 30]
T	T	T		?[zadanie 31]

Zmienne ciągłe

2016-04-08

Wnioskowanie Probabilistyczne

└ Inżynieria sieci

└ Zmienne ciągłe

Z definicji: nie można skonstruować tabeli prawdopodobieństw warunkowych. Rozwiązania:

1 Dyskretyzacja:

- Czasem OK
- Duże tabele

2 (parametryczne) funkcje gęstości prawdopodobieństwa, np. $N(\mu, \sigma^2)(x)$

3 (nieparametryczna) reprezentacja za pomocą przykładów

Hybrydowe sieci Baysowskie: rozkłady $\mathbf{P}(\text{ciągła}|\text{dykretne, ciągłe})$

2016-04-08

i $P(\text{dyskretna}|\text{dyskretne, ciągle})$

Prawd. całkowite

D_i - pogoda i -tego dnia. $D_i = \{\text{słoneczny, deszczowy}\}$

- $P(D_i = s) = 0.9$
- $P(D_{i+1} = s|D_i = s) = 0.8$
- $P(D_2 = d|D_1 = s) = ?$ [zadanie 32]
- $P(D_2 = s|D_1 = d) = 0.6$
- $P(D_2 = d|D_1 = d) = ?$ [zadanie 33]
- $P(D_2 = s) = ?$ [zadanie 34]
- $P(D_3 = s) = ?$ [zadanie 35]
- Narysuj sieć bayesowską opisującą ten świat [zadanie 36]

2016-04-08

Wnioskowanie Probabilistyczne

└ Zadania

└ Prawd. całkowite

1. 0.2
2. 0.4
3. (prawd. całkowite) = 0.78
4. (prawd. całkowite) = 0.756

- D_i - pogoda i -tego dnia. $D_i = \{\text{słoneczny, deszczowy}\}$
- $P(D_i = s) = 0.9$
 - $P(D_{i+1} = s|D_i = s) = 0.8$
 - $P(D_2 = d|D_1 = s) = ?$ [zadanie 32]
 - $P(D_2 = s|D_1 = d) = 0.6$
 - $P(D_2 = d|D_1 = d) = ?$ [zadanie 33]
 - $P(D_2 = s) = ?$ [zadanie 34]
 - $P(D_3 = s) = ?$ [zadanie 35]
 - Narysuj sieć bayesowską opisującą ten świat [zadanie 36]

Zadania

- Obecność nowotworu: $C = \{c, \neg c\}$.
- Test wykrył raka: $T = \{+, -\}$
- $P(c) = 0.01$
- $P(\neg c) =$
- $P(+|c) = 0.9$
- $P(-|c) =$
- $P(+|\neg c) = 0.2$
- $P(-|\neg c) =$
- $P(+, c) =$
- $P(-, c) =$
- $P(+, \neg c) =$
- $P(-, \neg c) =$

2016-04-08

Wnioskowanie Probabilistyczne

Zadania

Zadania

Zadania

- Obecność nowotworu: $C = \{c, \neg c\}$.
- Test wykrył raka: $T = \{+, -\}$
- $P(c) = 0.01$
- $P(\neg c) =$
- $P(+|c) = 0.9$
- $P(-|c) =$
- $P(+|\neg c) = 0.2$
- $P(-|\neg c) =$
- $P(+, c) =$
- $P(-, c) =$
- $P(+, \neg c) =$
- $P(-, \neg c) =$

1. 0.99
2. 0.1
3. 0.8
4. (z definicji) 0.009
5. (z definicji) 0.001
6. (z definicji) 0.198
7. (z definicji) 0.792
8. $P(c|+) = \frac{P(c,+)}{P(+)} = \frac{P(c,+)}{P(+|c)P(c)+P(+|\neg c)P(\neg c)} = \frac{0.9*0.01}{0.9*0.01+0.2*0.99} = 0.043$

Zwróćmy uwagę jak mała jest to wartość (w porównaniu do 0.9). Powodem jest prawy składnik licznika (0.2*0.99). Co ciekawe zwiększenie 0.9 nawet do 1.0 niewiele zmienia. Dopiero zwiększenie 0.2 (największe znaczenie!) w dół albo zwiększenie występowania choroby w populacji (0.01) może polepszyć sprawę.

Zadania

- Obecność nowotworu: $C = \{c, \neg c\}$.
- Test wykrył raka: $T = \{+, -\}$
- $P(c) = 0.01$
- $P(\neg c) =$
- $P(+|c) = 0.9$
- $P(-|c) =$
- $P(+|\neg c) = 0.2$
- $P(-|\neg c) =$
- $P(+, c) =$
- $P(-, c) =$
- $P(+, \neg c) =$
- $P(-, \neg c) =$
- $P(c|+) = ?$ [zadanie 37]
 - Zwróć uwagę, jak mała jest to wartość w porównaniu z wartością $P(+|c) = 0.9$

2016-04-08

Wnioskowanie Probabilistyczne

Zadania

Zadania

1. 0.99
2. 0.1
3. 0.8
4. (z definicji) 0.009
5. (z definicji) 0.001
6. (z definicji) 0.198
7. (z definicji) 0.792
8. $P(c|+) = \frac{P(c,+)}{P(+)} = \frac{P(c,+)}{P(+|c)P(c)+P(+|\neg c)P(\neg c)} = \frac{0.9*0.01}{0.9*0.01+0.2*0.99} = 0.043$

Zwróćmy uwagę jak mała jest to wartość (w porównaniu do 0.9). Powodem jest prawy składnik licznika (0.2*0.99). Co ciekawe zwiększenie 0.9 nawet do 1.0 niewiele zmienia. Dopiero zwiększenie 0.2 (największe znaczenie!) w dół albo zwiększenie występowania choroby w populacji (0.01) może polepszyć sprawę.

Zadania

- Obecność nowotworu: $C = \{c, \neg c\}$.
- Test wykrył raka: $T = \{+, -\}$
- $P(c) = 0.01$
- $P(\neg c) =$
- $P(+|c) = 0.9$
- $P(-|c) =$
- $P(+|\neg c) = 0.2$
- $P(-|\neg c) =$
- $P(+, c) =$
- $P(-, c) =$
- $P(+, \neg c) =$
- $P(-, \neg c) =$
- $P(c|+) = ?$ [zadanie 37]
- Zwróć uwagę, jak mała jest to wartość w porównaniu z wartością $P(+|c) = 0.9$

Warunkowa niezależność zmiennych

- Obecność nowotworu: C
- Test T wykrył raka $\{+, -\}$. Wynik testu jest losowy. Tutaj przeprowadzamy go dwukrotnie (zmiennie losowe T_1, T_2), żeby zwiększyć szansę na prawidłowy wynik.
- Narysuj sieć zawierającą zmienne: C, T_1, T_2 [zadanie 38]
- $P(c) = 0.01, P(\neg c) =$
- $P(+|c) = 0.9$
- $P(+|\neg c) = 0.2$
- $P(C = c | T_1 = +, T_2 = +) = ?$ [zadanie 39]

2016-04-08

Wnioskowanie Probabilistyczne

Zadania

Warunkowa niezależność zmiennych

1. $\alpha P(++|c)P(c) = \alpha P(+|c)P(+|c)P(c) = 0.1698$
Zmienne T_1 i T_2 są niezależne warunkowo

Warunkowa niezależność zmiennych

- Obecność nowotworu: C
- Test T wykrył raka $\{+, -\}$. Wynik testu jest losowy. Tutaj przeprowadzamy go dwukrotnie (zmiennie losowe T_1, T_2), żeby zwiększyć szansę na prawidłowy wynik.
- Narysuj sieć zawierającą zmienne: C, T_1, T_2 [zadanie 38]
- $P(c) = 0.01, P(\neg c) =$
- $P(+|c) = 0.9$
- $P(+|\neg c) = 0.2$
- $P(C = c | T_1 = +, T_2 = +) = ?$ [zadanie 39]

Warunkowa niezależność zmiennych

• j.w. oblicz:

• $P(T_2 = + | T_1 = +) = ?$ [zadanie 40]

2016-04-08

Wnioskowanie Probabilistyczne

└ Zadania

└ Warunkowa niezależność zmiennych

• j.w. oblicz:
• $P(T_2 = + | T_1 = +) = ?$ [zadanie 40]

1. Można różnymi sposobami. Tutaj chodzi o pokazanie niezależności warunkowej, więc:

$$P(+|+) = P(+|+, c)P(c|+) + P(+|+, \neg c)P(\neg c|+) = P(+|c)P(c|+) + P(+|\neg c)P(\neg c|+) = 0.2301$$

Wnioskowanie w sieci BEAMJ

Oblicz prawdopodobieństwa w sieci BEAJM:

- 1 $P(b|j, m)$
- 2 $P(\neg j|m)$
- 3 $P(\neg j|a, m)$
- 4 $P(m|\neg a, j, b)$
- 5 $P(\neg j, b|m)$

2016-04-08

Wnioskowanie Probabilistyczne

Zadania

Wnioskowanie w sieci BEAMJ

- Oblicz prawdopodobieństwa w sieci BEAJM.
- $P(b|j, m)$
 - $P(\neg j|m)$
 - $P(\neg j|a, m)$
 - $P(m|\neg a, j, b)$
 - $P(\neg j, b|m)$

Ile parametrów potrzeba?

Ile jest parametrów (niestrukturalnych) w sieci?[zadanie 41]

- $A \rightarrow B$

2016-04-08

Wnioskowanie Probabilistyczne

└ Zadania

└ Ile parametrów potrzeba?

1. $P(a)$ oraz $P(b|a)$ i $P(\neg b|a)$, czyli 3.

Ile parametrów potrzeba?

- 1 $A \rightarrow B \rightarrow E$
 $A \rightarrow C \rightarrow F$
 $A \rightarrow D \rightarrow F$ [zadanie 42]

- 2
 - $A, B, C \rightarrow D$, [zadanie 43]
 - $D \rightarrow E, F, G$
 - $C \rightarrow G$

2016-04-08

Wnioskowanie Probabilistyczne

Zadania

Ile parametrów potrzeba?

1. $1+2+2+2+2+4=13$
2. $1+1+1+8+4+2+2=19$

Ile parametrów potrzeba?

- $A \rightarrow B \rightarrow E$
- $A \rightarrow C \rightarrow F$
- $A \rightarrow D \rightarrow F$ [zadanie 42]
- $A, B, C \rightarrow D$, [zadanie 43]
- $D \rightarrow E, F, G$
- $C \rightarrow G$

Niezależność warunkowa

- $A \rightarrow B \rightarrow D \rightarrow D \rightarrow G$
- $H \rightarrow G$
- $C \rightarrow B$
- $F \rightarrow E \rightarrow D \rightarrow G$
- Czy:
 - $F \perp A?$ [zadanie 44]
 - $F \perp A|D?$ [zadanie 45]
 - $F \perp A|G?$ [zadanie 46]
 - $F \perp A|H?$ [zadanie 47]

2016-04-08

Wnioskowanie Probabilistyczne

Zadania

Niezależność warunkowa

1. Narysować tę sieć.
2. Tak
3. Nie
4. Nie
5. Tak

Niezależność warunkowa

- $A \rightarrow B \rightarrow D \rightarrow D \rightarrow G$
- $H \rightarrow G$
- $C \rightarrow B$
- $F \rightarrow E \rightarrow D \rightarrow G$
- Czy:
 - $F \perp A?$ [zadanie 44]
 - $F \perp A|D?$ [zadanie 45]
 - $F \perp A|G?$ [zadanie 46]
 - $F \perp A|H?$ [zadanie 47]