Uczenie modeli kanonicznych w sieciach Bayesowskich - efektywne uczenie modelu Noisy OR/MAX z małych zbiorów danych

Krzysztof Nowak

Politechnika Białostocka

6 marca 2013

Wprowadzenie

Sieć Bayesowska - struktura służąca do przedstawiania zależności pomiędzy zdarzeniami, bazując na rachunku prawdopodobieństwa.

Wprowadzenie

Sieć Bayesowska - struktura służąca do przedstawiania zależności pomiędzy zdarzeniami, bazując na rachunku prawdopodobieństwa. W sieciach bayesowskich można wyróżnić:

Wprowadzenie

Sieć Bayesowska - struktura służąca do przedstawiania zależności pomiędzy zdarzeniami, bazując na rachunku prawdopodobieństwa.
W sieciach bayesowskich można wyróżnić:

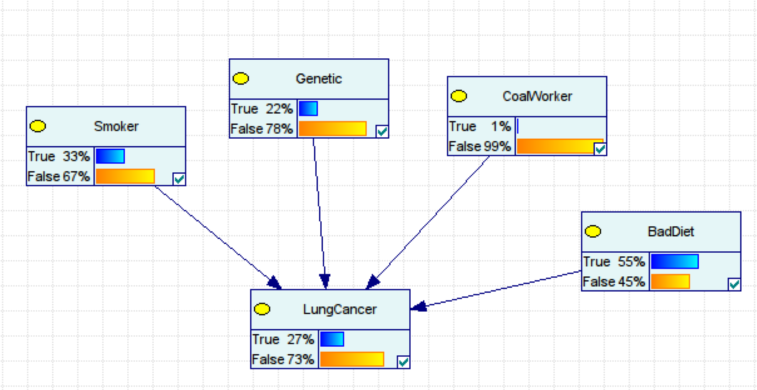
- Strukturę sieci - skierowany, acykliczny graf

Wprowadzenie

Sieć Bayesowska - struktura służąca do przedstawiania zależności pomiędzy zdarzeniami, bazując na rachunku prawdopodobieństwa.
W sieciach bayesowskich można wyróżnić:

- Strukturę sieci - skierowany, acykliczny graf
- Parametry sieci - wartości liczbowe określające prawdopodobieństwo poszczególnych zdarzeń

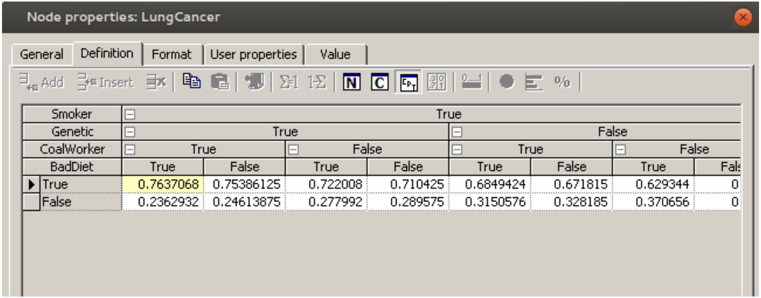
Struktura sieci



Rysunek: Przykładowa sieć bayesowska - Genie 2.0

Parametry sieci

CPT - Conditional Probability Table

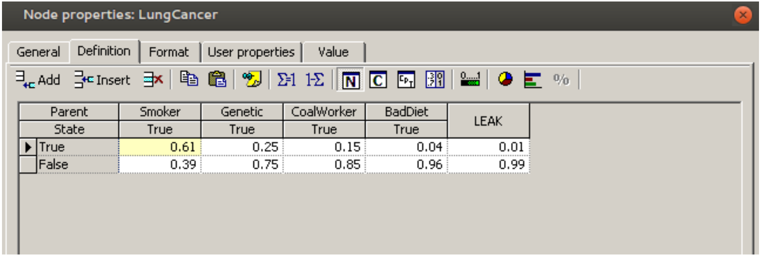


Smoker	True							
Genetic	True				False			
CoalWorker	True		False		True		False	
BadDiet	True	False	True	False	True	False	True	Fals
True	0.7637068	0.75386125	0.722008	0.710425	0.6849424	0.671815	0.629344	0
False	0.2362932	0.24613875	0.277992	0.289575	0.3150576	0.328185	0.370656	0

Rysunek: CPT - Genie 2.0

- Wykładniczy przyrost parametrów ze względu na ilość węzłów rodzicielskich.
- Przydatny model przy uczeniu sieci z dużych zbiorów danych.

Parametry sieci



Parent	Smoker	Genetic	CoalWorker	BadDiet	LEAK
State	True	True	True	True	
True	0.61	0.25	0.15	0.04	0.01
False	0.39	0.75	0.85	0.96	0.99

Rysunek: Noisy MAX/OR - Genie 2.0

- Liniowy przyrost parametrów ze względu na ilość węzłów rodzicielskich.
- Przydatny model przy uczeniu sieci z małych zbiorów danych, bądź wiedzy eskperta.
- Bramka Noisy OR jest szczególnym przypadkiem bramki Noisy MAX.
- Noisy OR nie zawsze dobrze odwzorowuje rzeczywistość

Modele kanoniczne - Noisy OR

- Najprostszy i najbardziej intuicyjny z modeli kanonicznych.

Bramka Noisy-OR wymaga podania prawdopodobieństwa wystąpienia danego zjawiska dla poszczególnych wartości węzłów rodzicielskich, niezależnie od pozostałych:

$$p_k = P(y|\bar{x}_1, \bar{x}_2, \ldots, x_k, \ldots, \bar{x}_{n-1}, \bar{x}_n). \tag{1}$$

Prawdopodobieństwo w bramce Noisy-OR przy wektorze wejściowym **X** wyliczamy następująco:

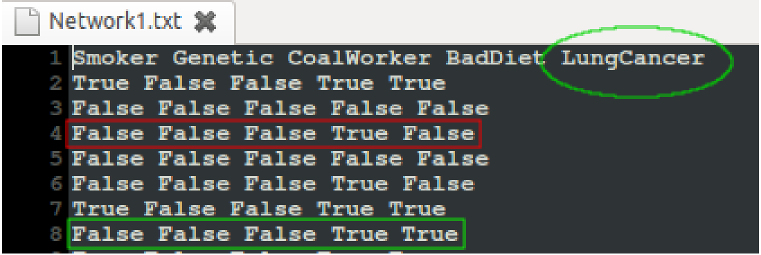
$$P(y|\mathbf{X}) = 1 - \prod_{i:x_i \epsilon \mathbf{X}} (1 - p_i) \tag{2}$$

Parametr "LEAK" oznacza prawdopodobieństwo wystąpienia danego zjawiska, pomimo braku wystąpienia jakiejkolwiek jawnej przyczyny.

$$p_{leak} = P(y|\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n). \tag{3}$$

Wyliczanie parametrów z danych - CPT

Standardowy węzeł w sieci bayesowskiej wymaga podania całej tabeli prawdopodobieństw warunkowych. Zliczamy poszczególne wystąpienia danych kombinacji parametrów w pliku z danymi i na ich podstawie wyliczamy prawdopodobieństwo.


Network1.txt
```
Smoker Genetic CoalWorker BadDiet LungCancer
True False False True True
False False False False False
False False False True False
False False False False False
False False False True False
True False False True True
False False False True True
```

Rysunek: Przykładowy plik z danymi

$$\frac{27}{27 + 311} = 0.079 \quad (4)$$

Wyliczanie parametrów z danych - Noisy-OR/MAX

- Węzeł typu Noisy-OR/MAX nie wymaga podania prawdopodobieństwa dla każdej możliwej kombinacji parametrów, a jedynie dla prawdopodobieństwa wystąpienia każdego z parametrów z osobna (niezależnie od innych)
- Sposób wyliczania jest identyczny, jednak ze względu na charakter bramki Noisy-OR/MAX potrzebujemy znacznie mniej parametrów

Usprawnienie wyliczania parametrów z danych - Noisy-OR/MAX

- W podanym wcześniej pliku z danymi, ilość rekordów składających się na wyliczenie wszystkich parametrów dla bramki Noisy-OR to około **47%** wszystkich rekordów
- Można to interpretować w taki sposób: Przy określaniu parametrów dla bramki Noisy-OR, pomijamy około **53%** informacji zawartych w danych
- Dla porównania, określenie parametrów (CPT) dla bramki standardowej wykorzystuje cały plik z danymi

Usprawnienie wyliczania parametrów z danych - Noisy-OR/MAX

- W podanym wcześniej pliku z danymi, ilość rekordów składających się na wyliczenie wszystkich parametrów dla bramki Noisy-OR to około **47%** wszystkich rekordów
- Można to interpretować w taki sposób: Przy określaniu parametrów dla bramki Noisy-OR, pomijamy około **53%** informacji zawartych w danych
- Dla porównania, określenie parametrów (CPT) dla bramki standardowej wykorzystuje cały plik z danymi
- W wypadku małych zbiorów danych niektóre parametry mogą być obciążone dużym błędem, bądź nie występować w ogóle

Usprawnienie wyliczania parametrów z danych - Noisy-OR/MAX

- W podanym wcześniej pliku z danymi, ilość rekordów składających się na wyliczenie wszystkich parametrów dla bramki Noisy-OR to około **47%** wszystkich rekordów
- Można to interpretować w taki sposób: Przy określaniu parametrów dla bramki Noisy-OR, pomijamy około **53%** informacji zawartych w danych
- Dla porównania, określenie parametrów (CPT) dla bramki standardowej wykorzystuje cały plik z danymi
- W wypadku małych zbiorów danych niektóre parametry mogą być obciążone dużym błędem, bądź nie występować w ogóle
- Czy da się lepiej uzyskać poszczególne parametry sieci Noisy-OR?

Usprawnienie wyliczania parametrów z danych.

Kombinacja, Ilość reprezentantów (mianownik), stosunek procentowy

```
1  False False False True   | 28330  | 0.2833%
2  False False False False  | 23400  | 0.2340%
3  True False False True    | 14192  | 0.1419%
4  True False False False   | 11297  | 0.1130%
5  False True False True    | 8162   | 0.0816%
6  False True False False   | 6589   | 0.0659%
7  True True False True     | 3791   | 0.0379%
8  True True False False    | 3230   | 0.0323%
9  False False True True    | 264    | 0.0026%
10 False False True False   | 253    | 0.0025%
11 True False True True     | 134    | 0.0013%
12 True False True False    | 122    | 0.0012%
13 False True True True     | 86     | 0.0009%
14 False True True False    | 79     | 0.0008%
15 True True True False     | 38     | 0.0004%
16 True True True True      | 33     | 0.0003%
```

Możemy użyć rekordów 1, 4, 6 oraz 10 do uzupełnienia tabeli Noisy-OR.

- Czy jest to optymalny sposób pozyskania parametrów?

Usprawnienie wyliczania parametrów z danych.

- Zamiast zliczać rekordy w danych określające każdy z parametrów bezpośrednio (Noisy-OR), wybieramy *n* kombinacji o największej liczbie "reprezentantów" w danych.
- Otrzymujemy w ten sposób pewien zbiór kombinacji parametrów, przykładowo:
  $x_1, \bar{x_2}, x_3, \bar{x_4} = p_{1,3}$
  $x_1, \bar{x_2}, \bar{x_3}, x_4 = p_{1,4}$
  $\bar{x_1}, \bar{x_2}, x_3, x_4 = p_{3,4}$
  $x_1, x_2, x_3, \bar{x_4} = p_{1,2,3}$

Usprawnienie wyliczania parametrów z danych.

- Zamiast zliczać rekordy w danych określające każdy z parametrów bezpośrednio (Noisy-OR), wybieramy n kombinacji o największej liczbie "reprezentantów" w danych.
- Otrzymujemy w ten sposób pewien zbiór kombinacji parametrów, przykładowo:

  $x_1, \bar{x_2}, x_3, \bar{x_4} = p_{1,3}$
  $x_1, \bar{x_2}, \bar{x_3}, x_4 = p_{1,4}$
  $\bar{x_1}, \bar{x_2}, x_3, x_4 = p_{3,4}$
  $x_1, x_2, x_3, \bar{x_4} = p_{1,2,3}$
- Zależy nam na uzyskaniu parametrów p_1, p_2, p_3, p_4.

Rozwiązywanie układu równań

- Korzystając ze wzoru (2) mamy:
  $1-(1-p_1)\cdot(1-p_3)=p_{1,3}$
  $1-(1-p_1)\cdot(1-p_4)=p_{1,4}$
  $1-(1-p_3)\cdot(1-p_4)=p_{3,4}$
  $1-(1-p_1)\cdot(1-p_2)\cdot(1-p_3)=p_{1,2,3}$

Rozwiązywanie układu równań

- Korzystając ze wzoru (2) mamy:
  $1 - (1 - p_1) \cdot (1 - p_3) = p_{1,3}$
  $1 - (1 - p_1) \cdot (1 - p_4) = p_{1,4}$
  $1 - (1 - p_3) \cdot (1 - p_4) = p_{3,4}$
  $1 - (1 - p_1) \cdot (1 - p_2) \cdot (1 - p_3) = p_{1,2,3}$
- Aby uprościć obliczenia, rozwiązujemy układ dla $1 - p_i$ zamiast p_i (używamy podstawienia $p_i' = 1 - p_i$).

Rozwiązywanie układu równań

- Korzystając ze wzoru (2) mamy:
  $1-(1-p_1)\cdot(1-p_3)=p_{1,3}$
  $1-(1-p_1)\cdot(1-p_4)=p_{1,4}$
  $1-(1-p_3)\cdot(1-p_4)=p_{3,4}$
  $1-(1-p_1)\cdot(1-p_2)\cdot(1-p_3)=p_{1,2,3}$
- Aby uprościć obliczenia, rozwiązujemy układ dla $1-p_i$ zamiast p_i (używamy podstawienia $p_i'=1-p_i$). $1-(p`_1)\cdot(p`_3)=p_{1,3}$
  $1-(p`_1)\cdot(p`_4)=p_{1,4}$
  $1-(p`_3)\cdot(p`_4)=p_{3,4}$
  $1-(p`_1)\cdot(p`_2)\cdot(p`_3)=p_{1,2,3}$

Rozwiązywanie układu równań

- Ostatecznie otrzymujemy:

  $(p`_1) \cdot (p`_3) = 1 - p_{1,3}$
  $(p`_1) \cdot (p`_4) = 1 - p_{1,4}$
  $(p`_3) \cdot (p`_4) = 1 - p_{3,4}$
  $(p`_1) \cdot (p`_2) \cdot (p`_3) = 1 - p_{1,2,3}$

- Nie jest to liniowy układ równań.
  1. Rozwiązanie za pomocą zmodyfikowanej eliminacji gaussa.
  2. Zlogarytmować obustronnie każde z równań.

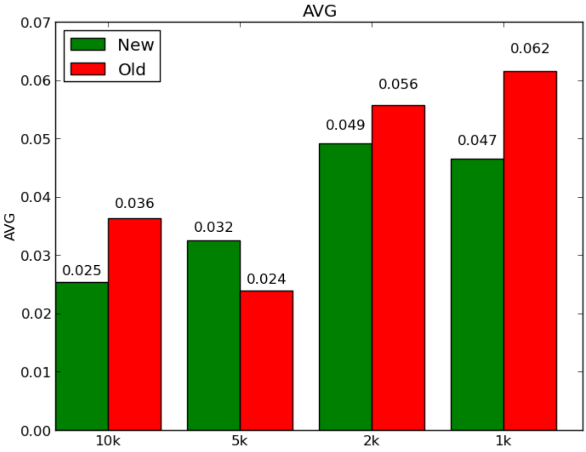
  $\log(p`_1) + \log(p`_3) = \log(1 - p_{1,3})$
  $\log(p`_1) + \log(p`_4) = \log(1 - p_{1,4})$
  $\log(p`_3) + \log(p`_4) = \log(1 - p_{3,4})$
  $\log(p`_1) + \log(p`_2) + \log(p`_3) = \log(1 - p_{1,2,3})$

Rozwiązywanie układu równań

- Uzyskane w ten sposób parametry $p_i'' = \log(p_i')$ dają nam szukane parametry p_i przez ponowne podstawienie:
  $p_i = 1 - e^{p_i''}$, gdzie e - podstawa logarytmu (dowolna)
- Rozwiązanie samego układu nie stanowi problemu
- Wybór odpowiednich równań jest kluczowy - Równania muszą być liniowo niezależne
- Sam wybór stanowi o skuteczności ostatecznego rozwiązania:
  1. Optymalizacja sumarycznej ilości "reprezentatów" w układzie równań.
  2. Rozwiązanie n układów, rozpatrując każdy z parametrów z osobna.
  3. Rozwiązanie m układów równań, i wyliczenie m zestawów parametrów, z których wyliczamy ostateczny zestaw parametrów za pomocą średniej ważonej.

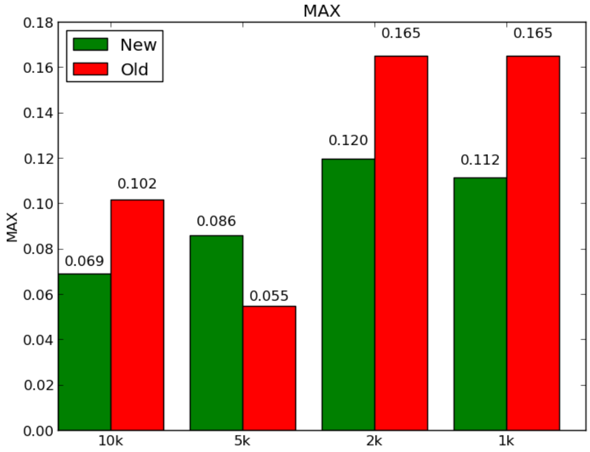
Przydatna będzie heurystyka wyboru równań uwzględniającą ich liniową niezależność. Dokładne rozwiązanie może być złożone $\binom{2^n}{n}$.

Wyniki



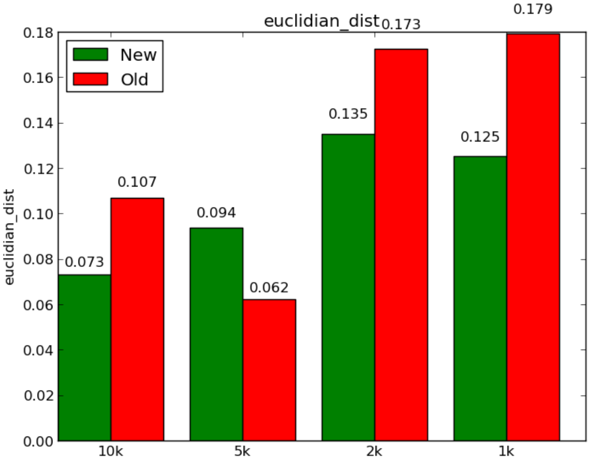
Rysunek: Średni błąd parametru

Wyniki



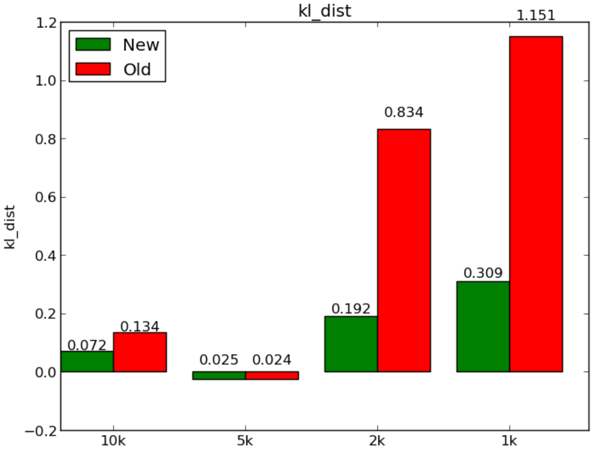
Rysunek: Maksymalny błąd parametru

Wyniki



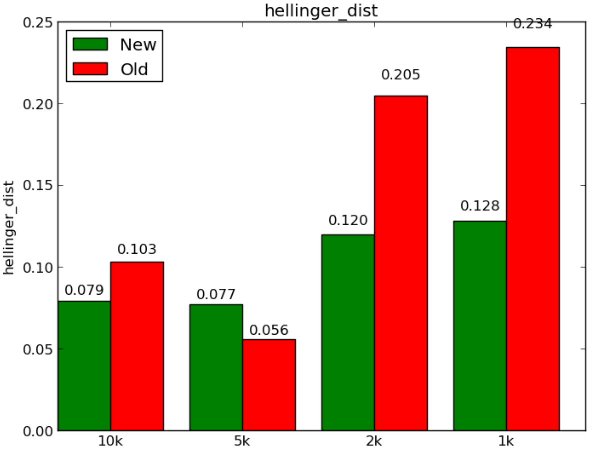
Rysunek: Odległość euklidesowa.

Wyniki



Rysunek: Dywergencja Kullbacka-Leiblera.

Wyniki



Rysunek: Odległość Hellingera.

Otwarte problemy

- Wybór równań do układu, tak aby minimalizować błąd, jednocześnie gwarantując jego rozwiązywalność
  - Metody heurystyczne
  - Sprowadzenie zagadnienia do problemu optymalizacyjnego (Simplex, metody gradientowe)
- Wyliczenie parametru LEAK przy przypadkach skrajnych

Źródła

- Francisco J. Diez, Marek J. Drużdżel - "Canonical Probabilistic Models for Knowledge Engineering" (28.4.2007)
- Nir Friedman, Moises Goldszmidt - "Learning Bayesian networks with local structure"
- Agnieszka Oniśko, Marek J. Drużdżel, Hanna Wasyluk - "Learning Bayesian network parameters from small data sets: application of Nosiy-OR gates" (1.3.2001)