Systemy ekspertowe

Wykład 5
Reprezentacja wiedzy
Probabilistyczne wyrażanie niepewności — systemy regułowe
Współczynniki wiartygodności

Joanna Kołodziejczyk

7 maja 2020

Plan wykładu

Probabilistyczne systemy regułowe

Reguła probabilistyczna

Sieci bayesowskie można przedstawić jako zbiór reguł. Ogólny schemat reguły:

IF D(owód) jest TRUE
THEN H(ipoteza) jest TRUE (z prawdopodobienstwem P)

Wnioskowanie jest wynikiem zastosowania reguły Bayes'a (przykład dla zmiennych boolowskich)

$$P(H|D) = \frac{P(D|H)p(H)}{P(D|H)p(H) + P(D|\neg H)P(\neg H)}$$

Wskaźnik wiarygodności i iloraz szans

Wskaźnik wiarygodności (Likelihood ratio)

$$L(H|D) = \frac{P(D|H)}{P(D|\neg H)}$$

Iloraz szans (Odds ratio)

1. Dla hipotezy

$$O(H) = \frac{P(H)}{P(\neg H)} = \frac{P(H)}{1 - P(H)}$$

2. Warunkowy

$$O(H|D) = \frac{P(H|D)}{P(\neg H|D)}$$

Reguła Bayesa oparta na L i O

$$O(H|D) = L(H|D)O(H)$$

Obliczanie wniosku i stopnia przekonania — przykład

- W środku nocy budzi nas alarm informujący o włamaniu.
- Istnieje 95% szansy, że włamanie włączy alarm: $P(alarm|wlamanie) = 0.95$.
- W oparciu o wcześniejsze fałszywe alarmy obliczono, że jest niewielka szansa (1%), że alarm włączył się bez przyczyny: $P(alarm|\neg wlamanie) = 0,01$.
- W oparciu o dane o przestępczości jest szansa 1 na 1000, że danej nocy do danego domu nastąpi włamanie: $p(wlamanie) = 0.0001$
- Jaki jest stopień przekonania o włamaniu?

Obliczanie wniosku i stopnia przekonania — przykład

1. $O(wlamanie|alarm) = L(alarm|wlamanie)O(wlamanie)$
2. $O(wlamanie|alarm) = \frac{0.95}{0.01} \cdot \frac{0.0001}{1-0.0001} = 0.0095$
3. $P(wlamanie|alarm) = \frac{O(wlamanie|alarm)}{1+O(wlamanie|alarm)} = 0.00941$

Choć to samo można uzyskać z twierdzenia Bayesa:

$$P(wlamanie|alarm) = \frac{0.95 \cdot 0.0001}{0.95 \cdot 0.0001 + 0.01 \cdot (1 - 0.0001)} = 0.00941$$

Wniosek

Retrospektywne wsparcie dla hipotezy o włamaniu dawane przez dowód w postaci włączonego alarmu zostało wzmocnione ponad stukrotnie (0.0001 do 0.0094).

Prognozowanie pogody

Zadanie:

Określić czy będzie, lub czy nie będzie padał deszcz następnego dnia.

Dane:

- minimalna temperatura w ciągu dnia
- maksymalna temperatura w ciągu dnia,
- opady w mm
- aktualny stan: pada lub nie pada itd...

Reguły:

1. IF dziś pada $[LS = 2.5\, LN = 0.6]$
   THEN jutro pada $\{P(jp) = 0.5\}$
2. IF dziś nie pada $[LS = 1.6\, LN = 0.4]$
   THEN jutro nie pada $\{P(jnp) = 0.5\}$

Dane

Table 3.3 London weather summary for March 1982

Day of month	Min. temp. °C	Max. temp. °C	Rainfall mm	Sunshine hours	Actual weather	Weather forecast
1	9.4	11.0	17.5	3.2	Rain	–
2	4.2	12.5	4.1	6.2	Rain	Rain
3	7.6	11.2	7.7	1.1	Rain	Rain
4	5.7	10.5	0.0	4.3	Dry	Rain*
5	3.0	12.0	0.0	9.5	Dry	Dry
6	4.4	9.6	0.0	3.5	Dry	Dry
7	4.8	9.4	4.6	10.1	Rain	Rain
8	1.8	9.2	5.5	7.8	Rain	Rain
9	2.4	10.2	4.8	4.1	Rain	Rain
10	5.5	12.7	4.2	3.8	Rain	Rain
11	3.7	10.9	4.4	9.2	Rain	Rain
12	5.9	10.0	4.8	7.1	Rain	Rain
13	3.0	11.9	0.0	8.3	Dry	Rain*
14	5.4	12.1	4.8	1.8	Rain	Dry*
15	8.8	9.1	8.8	0.0	Rain	Rain
16	2.4	8.4	3.0	3.1	Rain	Rain
17	4.3	10.8	0.0	4.3	Dry	Dry
18	3.4	11.1	4.2	6.6	Rain	Rain
19	4.4	8.4	5.4	0.7	Rain	Rain
20	5.1	7.9	3.0	0.1	Rain	Rain
21	4.4	7.3	0.0	0.0	Dry	Dry
22	5.6	14.0	0.0	6.8	Dry	Dry
23	5.7	14.0	0.0	8.8	Dry	Dry
24	2.9	13.9	0.0	9.5	Dry	Dry
25	5.8	16.4	0.0	10.3	Dry	Dry
26	3.9	17.0	0.0	9.9	Dry	Dry
27	3.8	18.3	0.0	8.3	Dry	Dry
28	5.8	15.4	3.2	7.0	Rain	Dry*
29	6.7	8.8	0.0	4.2	Dry	Dry
30	4.5	9.6	4.8	8.8	Rain	Rain

Co to jest LS i LN?

LS Likelihood of Sufficiency

Jest miarą ufności w hipotezę, jeżeli przedstawiony został dowód.

$$LS = \frac{P(D|H)}{P(D|\neg H)}$$

LN Likelihood of Necessity

Jest miarą ufności w hipotezę, jeżeli dowód nie został przedstawiony.

$$LN = \frac{P(\neg D|H)}{P(\neg D|\neg H)}$$

LN nie może być wyprowadzony z LS.

Jak ekspert określa LS i LN?

- Zaleta: nie musi opierać się na prawdopodobieństwach warunkowych.
- Ekspert podaj LS i LN bezpośrednio pilnując się zasad:
  1. duże $LS >> 1$ silnie potwierdza hipotezę H, gdy dany jest dowód D
  2. niskie $LN (0 < LN < 1)$ oznacza, że reguła silnie osłabia hipotezę $(\neg H)$, gdy nie ma dowodu D.
- Z LN i LS łatwo wyprowadzić prawdopodobieństwa warunkowe i zastosować regułę Bayes'a.
- W przykładzie:
  1. $LS = 2.5$ oznacza, że jeżeli dziś pada, to z dużym prawdopodobieństwem będzie padać jutro,
  2. $LN = 0.6$ oznacza, że istnieje szansa, że pomimo, że dziś nie pada to jutro będzie padać.

Jak oblicza się prawdopodobieństwo hipotezy?

- Prawdopodobieństwo a priori konsekwencji p(H) jest konwertowane na szansę a priori. $O(H) = P(H)/1 - P(H)$.
- Szansa warunkowa jest obliczana przy użyciu LS jeżeli poprzednik(przesłanka) reguły jest prawdziwa lub przy użyciu LN jeżeli przesłanka jest fałszywa: $O(H|D) = LS \cdot O(H)$ i $O(H|\neg D) = LN \cdot O(H)$
- Szansa warunkowa jest wykorzystana do obliczenia prawdopodobieństwa warunkowego: $P(H|D) = \frac{O(H|D)}{1+O(H|D)}$ i $P(H|\neg D) = \frac{O(H|\neg D)}{1+O(H|\neg D)}$

Obliczenia dla przykładu — odpalenie reguły 1

- Dowód D: dziś pada.
- $O(jp) = O(jp)/1 - O(jp) = 1$
- Dowód D zwiększa szansę na opad przez LS:
  $O(jp|dp) = LS \cdot O(jp) = 2.5 \cdot 1 = 2.5$
  $P(jp|dp) = O(jp|dp)/(1 + O(jp|dp)) = 0.71$
- Wniosek: Prawdopodobieństwo jutrzejszego opadu wzrasta z 0.5 do 0.71 poprzez przedstawienie dowodu D.

Obliczenia dla przykładu — odpalenie reguły 2

- Dowód D: dziś pada.
- $O(jnp) = O(jnp)/1 - O(jnp) = 1$
- Dowód D zmniejsza szansę na brak opadu przez LN:
  $O(jnp|dp) = LN \cdot O(jnp) = 0.4 \cdot 1 = 0.4$
  $P(jnp|dp) = O(jnp|dp)/(1 + O(jnp|dp)) = 0.29$
- Wniosek: prawdopodobieństwo, że jutro nie będzie padać, gdy dziś pada jest 0.29, czyli zmniejszyło się z powody zaistnienia D.

Obliczenia dla przykładu — „Dziś nie pada"

Obliczenia wykonuje się w sposób analogiczny. Oblicza się prawdopodobieństwa, że jutro nie będzie padać, lub będzie deszczowo, gdy dowód D mówi, że dziś nie pada.

1. Z prawdopodobieństwom 0.62 nie będzie padać, gdy dziś nie pada.
2. Z prawdopodobieństwom 0.38 będzie padało, gdy dziś nie padało

Rozbudowa systemu prognozowania pogody

- Powiększamy bazę wiedzy na temat prognozowania pogody o inne parametry.
- Wskaźniki są uzyskiwane poprzez analizę danych zbieranych z lat poprzednich. Pewne zależności są widoczne i można je ująć w reguły, które uszczegółowią analizę i sprawią, że wnioskowanie będzie bardziej wiarygodna.
- Reguły 1 i 2 takie jak we wcześniejszym przykładzie.

Rozbudowa systemu prognozowania pogody

Dalsze reguły:

3. IF dziś pada AND wilgotność jest niska $[LS = 10 \, LN = 1]$
   THEN jutro nie pada $\{P(jnp) = 0,5\}$
4. IF dziś pada AND wilgotność jest niska AND temperatura jest niska $[LS = 1.5 \, LN = 1]$
   THEN jutro nie pada $\{P(jnp) = 0,5\}$
5. IF dziś nie pada AND temperatura jest wysoka $[LS = 2 \, LN = 0.9]$
   THEN jutro pada $\{P(jp) = 0,5\}$
6. IF dziś nie pada AND temperatura jest wysoka AND niebo jest pochmurne $[LS = 5LN = 1]$
   THEN jutro pada $\{P(jp) = 0,5\}$

Działanie SE d.1

- Pytanie: Jaka jest dziś pogoda?
- Odp: PADA
- Obliczenia (wcześniej na slajdach): na podstawie reguły 1 $P(jp|dp) = 0.71$
- Obliczenia (wcześniej na slajdach): na podstawie reguły 2 $P(jnp|dp) = 0.29$
- Wniosek: jutro: pada [0.71], nie pada [0.29]

Działanie SE d.2

- Pytanie: Jaka jest dziś wilgotność?
- Odp: NISKA
- Obliczenia: prawdopodobieństwo a priori zmienione po poprzednim kroku stąd: $O(jnp) = 0.29/(1 - 0.29) = 0.41$
- Obliczenia: na podstawie reguły 3: $O(jnp|dp \wedge nw) = 10 \cdot 0.41 = 4.1$
- Obliczenia: $P(jnp|dp \wedge nw) = 4.1/(1 + 4.1) = 0.81$
- Wniosek: jutro: pada [0.71], nie pada [0.81]

Działanie SE d.3

- Pytanie: Jaka jest dziś temperatura?
- Odp: NISKA
- Obliczenia: prawdopodobieństwo a priori zmienione po poprzednim kroku stąd: $O(jnp) = 0.81/(1-0.81) = 4.26$ i $O(jp) = 0.71/(1-0.71) = 2.45$
- Obliczenia: na podstawie reguły 4: $O(jnp|dp \wedge nw \wedge nt) = 1.5 \cdot 4.26 = 6.4$
- Obliczenia: $P(jnp|dp \wedge nw \wedge nt) = 6.4/(1+6.4) = 0.86$
- Obliczenia: na podstawie reguły 5: $O(jp|dp \wedge nt) = 0.9 * 2.45 = 2.2$
- Obliczenia: $PO(jp|dp \wedge nt) = 2.2/(1+2.2) = 0.69$
- Wniosek: jutro: pada [0,69], nie pada [0,86]

Działanie SE d.4

- Pytanie: Jakie jest niebo?
- Odp: NIEZACHMURZONE
- Obliczenia: prawdopodobieństwo a priori zmienione po poprzednim kroku stąd: $O(jp) = 0.69/(1-0.69) = 2.23$
- Obliczenia: na podstawie reguły 6: $O(jp|dp \wedge nw \wedge nt \wedge np) = 1 * 2.23 = 2.23$
- $P(jp|dp \wedge nw \wedge nt \wedge np) = 2.23/(1+2.23) = 0.69$
- Wniosek: jutro: pada [0,69], nie pada [0,86]

Zastosowania sieci bayesowskich

- diagnoza medyczna ((Heckerman 1990)
- rozumienie języka (Charniak and Goldman 1989)
- widzenie (Levitt, Mullin, and Binford 1989)
- przeszukiwanie heurystyczne (Hansson and Mayer 1989)

Plan wykładu

Współczynnik wiarygodności

- Alternatywa dla niepewności probabilistycznej czy też opartej na logice.
- MYCIN – system, w którym wprowadzono współczynniki wiarygodności
- Zastosowanie, gdy brak spójnych danych statystycznych.
- Naśladują procesy myślowe przy wnioskowaniu prowadzonym przez eksperta - człowieka.

Współczynnik wiarygodności

- Współczynnik wiarygodności *cf* jest miarą przekonań eksperta.
- Ma wartość z przedziału -1(całkowity fałsz) do 1(całkowita prawda).
- Wartości >0 są miarą przekonania o hipotezie.
- Wartości <0 są miarą braku przekonania o hipotezie.

Współczynnik wiarygodności

Table 3.4 Uncertain terms and their interpretation

Term	Certainty factor
Definitely not	−1.0
Almost certainly not	−0.8
Probably not	−0.6
Maybe not	−0.4
Unknown	−0.2 to +0.2
Maybe	+0.4
Probably	+0.6
Almost certainly	+0.8
Definitely	+1.0

Konstrukcja reguł

Budowa reguły

IF D <dowód>
THEN H<hipoteza {cf}

Interpretacja

cf jest miarą ufności w hipotezę H, gdy dany jest dowód D.

Na czym oparty jest *cf*

Wykorzystuje się dwie funkcje oceny (dają wartości z przedziału $[0, 1]$):

1. MB — measure of belief — miara przekonania o hipotezie H w jej definicji wykorzystuje się prawdopodobieństwa
2. MD — measure of disbelief — miara braku przekonania w hipotezę H w jej definicji wykorzystuje się prawdopodobieństwa

Measure of belief

$$MB(H,D) = \begin{cases} 1 & \text{if } P(H) = 1 \\ \frac{P(H|D)-P(H)}{1-P(H)} & \text{wpp} \end{cases}$$

- Ekspert ocenia jak bardzo przedstawiony dowód D redukuje jego wątpliwości w wypadku braku dowodu $(1 - p(H))$.
- Jeżeli dowód jest słaby, to $p(H|D) - p(H)$ jest bliskie 0 i jego wątpliwości nie ulegną zmianie.

Measure of disbelief

$$MD(H,D) = \begin{cases} 1 & \text{if } P(H) = 0 \\ \frac{P(H) - P(H|D)}{P(H)} & \text{wpp} \end{cases}$$

- Ekspert ocenia jak bardzo przedstawiony dowód D redukuje jego przekonanie o hipotezie $p(H)$.
- $p(H)$ prawdopodobieństwo a priori, że H jest prawdziwe.

Jak wyrazić całkowite przekonanie o hipotezie?

- Używa się do tego współczynnika *cf* wyrażonego przez funkcje *MB* i *MD*.

$$cf = \frac{MB(H, D) - MD(H, D)}{1 - min[MB(H, D), MD(H, D)]}$$

- *cf* ma wartości z zakresu $[-1, 1]$, które oznacza całkowite przekonanie o hipotezie H
- Ekspert zazwyczaj potrafi oszacować *cf* skojarzony z każdą wartością B gdy dane jest $A = X$
  IF A is X
  THEN B is Y {cf 0.7}
  B is Z {cf 0.2}

Interpretacja *cf*

1. W przypadku przedstawionej reguły: gdy dane jest, że A ma wartość X, to B będzie równe Y w 70% przypadków i Z w 20% przypadków.
2. 10% przypadków może mieć dowolną wartość.
3. Przekonanie przypisane do reguły jest propagowane w trakcie wnioskowania.

cf dla reguły z jedną przełanką

$cf(H, D)$ dla reguły z jednym dowodem (jedną przesłanką) *cf* oblicza się jako mnożenie $cf(D)$ dowodu i *cf* reguły:

$$cf(H, D) = cf(D) \cdot cf$$

Przykład:

IF niebo czyste
THEN prognozujemy słońce {cf 0.8}

$cf(D)$ dla czyste niebo to 0.5
zatem $cf(H, D) = 0.5 * 0.8 = 0.4$
co czytamy, że „może być słonecznie".

Obliczanie *cf* przy koniunkcji wielu dowodów

Ogólna budowa reguły

IF <dowód D1>
AND <dowód D2> ...
AND <dowód Dn>
THEN <hipoteza> {cf}

Obliczanie *cf*

$$cf(H, D_1 \wedge D_2 \wedge \cdots \wedge D_n) = min[cf(D_1), cf(D_2) \ldots cf(D_n)] \cdot cf$$

Obliczanie *cf* przy koniunkcji wielu dowodów

Przykład:

IF niebo czyste
AND prognozujemy słońce
THEN noś okulary {cf 0.8}

$cf(D_1) = 0,9$ i $cf(D_2) = 0.7$

zatem $cf(H, D_1 \wedge D_2) = min[0.9, 0.7] * 0.8 = 0.56$
co czytamy „dziś, prawdopodobnie dobrym pomysłem będzie noszenie okularów przeciwsłonecznych".

Obliczanie *cf* przy dysjunkcji wielu dowodów

Ogólna budowa reguły

IF <dowód D1>
OR <dowód D2> ...
OR <dowód Dn>
THEN <hipoteza> {cf}

Obliczanie *cf*

$$cf(H, D_1 \vee D_2 \vee \cdots \vee D_n) = max[cf(D_1), cf(D_2) \ldots cf(D_n)] \cdot cf$$

Obliczanie *cf* przy dysjunkcji wielu dowodów

Przykład:

IF niebo zachmurzone
OR prognozujemy deszcz
THEN weź parasol {cf 0.9}

$cf(D_1) = 0,6$ i $cf(D_2) = 0.8$

zatem $cf(H, D_1 \vee D_2) = max[0.6, 0.8] * 0.9 = 0.72$
co czytamy „dziś, prawie na pewno powinna zostać zabrana parasolka”.

Obliczanie *cf*, gdy wiele zdarzeń wypływa na tą samą hipotezę

Ogólna budowa reguły

IF A jest X
THEN B is Y {cf1}

IF A jest X
THEN B is Y {cf2}

Obliczanie *cf*

$$cf(cf_1, cf_2) = \begin{cases} cf_1 + cf_2(1 - cf_1) & \text{if } cf_1 > 0 \wedge cf_2 > 0 \\ \frac{cf_1 + cf_2}{1 - min[|cf_1|, |cf_2|]} & \text{if } cf_1 > 0 \vee cf_2 > 0 \\ cf_1 + cf_2(1 + cf_1) & \text{if } cf_1 < 0 \wedge cf_2 < 0 \end{cases}$$

Analiza wzoru

- Jeżeli oba zdarzenia wzmacniają przekonanie o hipotezie przekonanie wzrasta przy złożeniu obu warunków.
- Jeżeli jedno zdarzenie wzmacnia, a drugie osłabia przekonanie o hipotezie to całkowity współczynnik wiarygodności zmniejszy się.
- Jeżeli oba zdarzenia osłabiają hipotezę to ich złożenie da bardziej osłabiający współczynnik wiarygodności.

Obliczanie *cf*, gdy wiele zdarzeń wypływa na tą samą hipotezę

Przykład:

IF A jest X
THEN B is Y {cf1 0.8}

IF A jest X
THEN B is Y {cf2 0.6}

$cf(D_1) = cf(D_2) = 1$

zatem $cf(H, D_1) = cf_1 \cdot cf(D_1) = 0.8$
$cf(H, D_2) = cf_2 * cf(D_2) = 0.6$

Wspólny *cf*
$cf = cf_1(H, D_1) + cf_2(H, D_2) \cdot (1 - cf_1(H, D_1)) = 0.8 + 0.6 \cdot (1 - 0.8) = 0.92$
Zatem przekonanie o hipotezie wzrosło.

Prognozowanie pogody

Zadanie:

Określić czy będzie, lub czy nie będzie padał deszcz następnego dnia. Zamiast współczynników LS i LN zastosowane zostaną *cf*.

Dane:

- minimalna temperatura w ciągu dnia
- maksymalna temperatura w ciągu dnia,
- opady w mm
- aktualny stan: pada lub nie pada itd...

Reguły

1. IF dziś pada
   THEN jutro pada {cf 0.5}
2. IF dziś nie pada
   THEN jutro nie pada {cf 0.5}
3. IF dziś pada AND wilgotność jest niska
   THEN jutro nie pada {cf 0.6}
4. IF dziś pada AND wilgotność jest niska AND temperatura jest niska
   THEN jutro nie pada {cf 0.7}
5. IF dziś nie pada AND temperatura jest wysoka
   THEN jutro pada {cf 0.65}
6. IF dziś nie pada AND temperatura jest wysoka AND niebo jest pochmurne
   THEN jutro pada {cf 0.65}

Działanie SE d.1

- Pytanie: Jaka jest dziś pogoda?
- Odp: PADA
- Pytanie: Podaj swoje przekonanie [-1,1]
- Odp: 1
- Obliczenia na podstawie reguły 1 $cf(jp, dp) = 1 \cdot 0.5 = 0.5$ przekonanie, że jutro pada.
- Wniosek: jutro: pada [0.5]

Działanie SE d.2

- Pytanie: Jaka jest dziś wilgotność?
- Odp: NISKA
- Pytanie: Jaki jest stopień wilgotności [-1,1]?
- Odp: 0.8
- Obliczenia: na podstawie reguły 3:

$$cf(jnp, dp \wedge nw) = min[cf(dp), cf(nw)] \cdot cf = min[1, 0.8] \cdot 0.6 = 0.48$$

- Wniosek: jutro: pada [0.5], nie pada [0.48]

Działanie SE d.3

- Pytanie: Jaka jest dziś temperatura?
- Odp: NISKA
- Pytanie: Podaj swoje przekonanie [-1,1]
- Odp: 0.9
- Obliczenia: na podstawie reguły 4:

$$cf(jnp, dp \wedge nw \wedge nt) = min[cf(dp), cf(nw), cf(nt)] \cdot cf =$$

$$= min[1, 0.8, 0.9] \cdot 0.7 = 0.56$$

- Złożenie reguł 3 i 4 w ocenie hipotezy nie pada:

$$cf(cf_{R3}, cf_{R4}) = cf_{R3} + cf_{R4} \cdot (1 - cf_{R3}) = 0.48 + 0.56 \cdot (1 - 0.48) = 0.77$$

- Wniosek: jutro: pada [0,5], nie pada [0,77]

Cechy systemów z *cf*

- Zastosowanie: planowanie, prognozowanie, tam gdzie brak danych statystycznych
- Brak matematycznej poprawności. Dane pochodzą z ocen ekspertów z dziedziny.
- Propagowanie przekonań wzrasta wykładniczo, dlatego nie nadaje się dla bardzo dużych baz wiedzy.
- Problemem jest znalezienie właściwej metody określania prawdopodobieństw i współczynników, ponieważ ludzie są tendencyjni w ocenach.

Plan wykładu

Wykorzystana literatura (do samodzielnego studiowania)

Eugene Charniak
Bayesian Networks without Tears.
AI MAGAZINE, WINTER 1991

S.J. Russel, P. Norvig
Artificial Intelligence. A modern approach.
Pearson Education wyd. 2, p.111-116