Modelowanie niepewności

Wprowadzenie do sztucznej inteligencji

Plan wykładu

Wprowadzenie do sztucznej inteligencji

- Zastosowanie rachunku prawdopodobieństwa do modelowania niepewności
- MYCIN – system ekspercki wnioskujący na podstawie wiedzy niepewnej
- Model współczynnika pewności
- Metoda Dempstera-Shafera
- Zbiory rozmyte
- Zbiory przybliżone

Prawdopodobieństwo zdarzenia

Wprowadzenie do sztucznej inteligencji

- Podejście z dziedziny gier losowych:
  prawdopodobieństwo to stosunek liczby wyników spełniających warunki określonego zdarzenia do liczby wszystkich wyników
- Podejście częstotliwościowe:
  prawdopodobieństwo jest granicą częstości występowania określonego zdarzenia, po wykonaniu dużej liczby eksperymentów
- Podejście subiektywne:
  prawdopodobieństwo jest pewną liczbą charakterystyczną dla danego zdarzenia losowego – miarą przekonania o wystąpieniu zdarzenia, czy też miarą odczuwania niepewności

Aksjomaty rachunku prawdopodobieństwa

Wprowadzenie do sztucznej inteligencji

Aksjomat I Każdemu zdarzeniu *A* odpowiada określona liczba P(A), zwana prawdopodobieństwem zdarzenia *A*, spełniająca nierówność: $0 \leq P(A) \leq 1$

Aksjomat II Prawdopodobieństwo zdarzenia pewnego równa się jedności.

Aksjomat III Prawdopodobieństwo sumy zdarzeń parami wyłączających się równa się sumie prawdopodobieństw tych zdarzeń.

Prawdopodobieństwo warunkowe

Wprowadzenie do sztucznej inteligencji

oznaczane symbolem P(A|B) wyraża przekonanie o prawdziwości zdarzenia A przy założeniu, że zdarzenie B zachodzi z całkowitą pewnością

Definicja

Niech prawdopodobieństwo zdarzenia B będzie dodatnie. Prawdopodobieństwo warunkowe zdarzenia A pod warunkiem zdarzenia B wyznaczamy następująco:

$$P(A\,|\,B) = \frac{P(A,B)}{P(B)}, \qquad \text{gdzie } P(B) > 0$$

Twierdzenie o prawdopodobieństwie całkowitym

Wprowadzenie do sztucznej inteligencji

Jeżeli zdarzenia $A_1, A_2, \ldots$ wyłączają się parami i wyczerpują zbiór zdarzeń elementarnych Θ, przy czym $P(A_i) > 0$ dla $i = 1, 2, \ldots$ to dla dowolnego zdarzenia B zachodzi równość:

$$P(B) = \sum_i P(B\,|\,A_i) P(A_i)$$

Twierdzenie Bayesa

Thomas Bayes (1701-1761)

Jeżeli zdarzenia A_1, A_2,... spełniają założenia twierdzenia o prawdopodobieństwie całkowitym oraz P(B) > 0, to dla i =1, 2,... zachodzi równość:

$$P(A_i\,|\,B) = \frac{P(B\,|\,A_i)P(A_i)}{\sum_i P(B\,|\,A_i)P(A_i)}$$

Twierdzenie Bayesa

Jeśli uwzględnimy twierdzenie o prawdopodobieństwie całkowitym oraz użyjemy symboli H i E dla zdarzeń oznaczających odpowiednio hipotezę oraz obserwację (przesłankę), to reguła Bayesa przyjmuje postać:

$$P(H\,|\,E) = \frac{P(E\,|\,H)P(H)}{P(E)}$$

Wady wnioskowania z zastosowaniem całkowitego rozkładu prawdopodobieństwa

- konieczność zdefiniowania i zapamiętania łącznego rozkładu prawdopodobieństwa
- wysoka złożoność obliczeniowa wyznaczania prawdopodobieństw brzegowych i warunkowych

Wymagania odnośnie do modelu niepewności w systemach regułowych

- W systemach wnioskowania logicznego reguła postaci $A \Rightarrow B$ pozwala wywnioskować B, gdy tylko zachodzi A, niezależnie od innych faktów. W systemach probabilis-tycznych trzeba wziąć pod uwagę wszystkie dostępne przesłanki.
- Jeżeli przeprowadzimy dowód jakiejś tezy, to tezy tej można użyć w kolejnych dowodach bez potrzeby ponownego jej dowodzenia. W systemach probabilis-tycznych przesłanki użyte do dowodu mogą ulec zmianie.
- W logice prawdziwość zdań złożonych można wywnioskować na podstawie wartości logicznej termów. Wnioskowanie probabilistyczne zachowuje tej własności, chyba, że nałożymy silne ograniczenia o niezależności.

Współczynnik pewności

- Buchanan, Shortliffe 1975
- Model opracowany dla regułowego systemu MYCIN

Jeżeli E to H

obserwacja

hipoteza

Belief

- MB[H, E] – miara zwiększenia przekonania (ang. *belief*) o wystąpieniu hipotezy H, na podstawie obserwacji E.

$$MB[H,E] = \begin{cases} 1 & \text{gdy } P(H) = 1 \\ \dfrac{\max\{P(H\,|\,E), P(H)\} - P(H)}{\max\{1,0\} - P(H)} & \text{w przeciwnym wypadku} \end{cases}$$

Disbelief

- MD[H, E] – miara zwiększenia braku przekonania (ang. *disbelief*) o wystąpieniu hipotezy H, na podstawie obserwacji E.

$$MD[H,E] = \begin{cases} 1 & \text{gdy } P(H)=0 \\ \dfrac{\min\{P(H|E),P(H)\} - P(H)}{\min\{1,0\} - P(H)} & \text{w przeciwnym wypadku} \end{cases}$$

Certainty factor

$$CF(H,E) = MB[H,E] - MD[H,E]$$

$$CF \in [-1\ ,\ 1]$$

Interpretacja współczynnika pewności

Współczynnik pewności jest skojarzony z regułą postaci:

Jeżeli *przesłanka* to *hipoteza*

i oznacza *zmianę* przekonania o prawdziwości hipotezy w wyniku wystąpienia przesłanki (obserwacji).

$$E \xrightarrow{CF(H,\ E)} H$$

Propagacja niepewności

$$E_1 \xrightarrow{CF(H,\ E_1)} H \qquad E_2 \xrightarrow{CF(H,\ E_2)} H$$

$$E_1,\ E_2 \xrightarrow{CF(H,\ E_1\&E_2)} H$$

Równoległe połączenie reguł

$$MB(H,E_1 \& E_2) = \begin{cases} 0 & \text{gdy } MD(H,E_1 \& E_2) = 1 \\ MB(H,E_1) + MB(H,E_2) * [1 - MB(H,E_1)] & \text{w przeciwnym razie} \end{cases}$$

$$MD(H,E_1 \& E_2) = \begin{cases} 0 & \text{gdy } MB(H,E_1 \& E_2) = 1 \\ MD(H,E_1) + MD(H,E_2) * [1 - MD(H,E_1)] & \text{w przeciwnym razie} \end{cases}$$

Propagacja niepewności

$$E_1 \xrightarrow{CF(E_2,\ E_1)} E_2 \xrightarrow{CF(H,\ E_2)} H \qquad E_1 \xrightarrow{CF(H,\ E_1)} H$$

Szeregowe połączenie reguł

$$CF(H,E_1) = \begin{cases} CF(E_2,E_1)CF(H,E_2) & \text{gdy } CF(E_2,E_1) \geq 0 \\ CF(E_2,E_1)CF(H,\neg E_2) & \text{w przeciwnym razie} \end{cases}$$

Gdy $CF(H,\neg E_2)$ nie jest zdefiniowane, przyjmuje się że jest ono równe 0.

Współczynnik pewności – definicja probabilistyczna

Heckerman 1986

$$CF(H,E) = \begin{cases} \dfrac{P(H|E) - P(H)}{P(H|E)(1 - P(H))} & \text{gdy} \quad P(H|E) > P(H) \\ \dfrac{P(H|E) - P(H)}{P(H)(1 - P(H|E))} & \text{gdy} \quad P(H) > P(H|E) \end{cases}$$

Miara pewności

Grzymała-Busse 1991

C(E) C(H)

$$E \xrightarrow{CF(H,\ E)} H$$

$$C'(H) = \begin{cases} C(H) + (1 - C(H))CF'(E \to H) & \text{gdy } C(H), CF'(E \to H) \geq 0 \\ C(H) + (1 + C(H))CF'(E \to H) & \text{gdy } C(H), CF'(E \to H) \leq 0 \\ \dfrac{C(H) + CF'(E \to H)}{1 - \min\{|C(H)|, |CF'(E \to H)|\}} & \text{gdy } C(H)CF'(E \to H) \leq 0 \end{cases}$$

$$CF'(E \to H) = CF(E \to H) * \max\{0, C(E)\}$$

Miara pewności

Jeżeli reguła ma w części warunkowej nie jedną przesłankę ale wyrażenie logiczne, stosowane jest podejście znane z teorii zbiorów rozmytych:

$$C(E_1 \wedge E_2) = \min\{C(E_1), C(E_2)\}$$
$$C(E_1 \vee E_2) = \max\{C(E_1), C(E_2)\}$$
$$C(\neg E_1) = 1 - C(E_1)$$

Przykład 1

C(s1) = 0,2

s1, s2 $\xrightarrow{CF(h,\ s1 \wedge s2) = 0,4}$ h

C(s2) = – 0,1

C(h) = 0,3

C(s1 ∧ s2) = min(0,2; – 0,1) = – 0,1

CF'(h, s1 ∧ s2) = 0,4 * 0 = 0

C'(h) = 0,3 + (1– 0,3) * 0 = 0,3 + 0 = 0,3

$$CF'(E \to H) = CF(E \to H) * \max\{0, C(E)\}$$

$$C'(H) = \{C(H) + (1 - C(H))CF'(E \to H) \quad \text{gdy } C(H), CF'(E \to H) \geq 0$$

Przykład 2

C(s1) = 0,2

s1, s2 $\xrightarrow{CF(h,\ s1 \wedge s2) = 0,4}$ h

C(s2) = 0,8

C(h) = 0,3

C(s1 ∧ s2) = min(0,2; 0,8) = 0,2

CF'(h, s1 ∧ s2) = 0,4 * 0,2 = 0,08

C'(h) = 0,3 + (1– 0,3) * 0,08 = 0,3 + 0,7 * 0,08 = 0,356

Teoria Dempstera-Shafera

Wiedza reprezentowana jest w postaci *funkcji przekonania*, stąd też model ten nazywany jest także teorią funkcji przekonania.

Każdemu zdaniu logicznemu przypisuje nie jedną lecz dwie wartości.

Użycie dwóch wartości zamiast jednej umożliwia, oprócz modelowania niepewności, reprezentację *ilości* pozyskanej informacji.

Proste modelowanie wspierania zbioru hipotez przez obserwacje.

Funkcja gęstości prawdopodobieństwa

(*density probability function*)

$$m : 2^{\Theta} \to [0,1]$$
$$m[\varnothing] = 0$$
$$\sum_{A \subseteq \Theta} m(A) = 1$$

Przekonanie (belief)

Przekonanie oznaczane w skrócie Bel $\in$ [0,1] mierzy siłę pozyskanych obserwacji wspierających przekonanie o prawdziwości rozważanego zbioru hipotez.

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B)$$

Wyobrażalność (plausibility)

Wyobrażalność oznaczana w skrócie Pl $\in$ [0,1] określa na ile przekonanie o prawdziwości **A** jest ograniczone przez dowody wspierające $\neg$**A**.

$$\mathrm{Pl}(A) = 1 - \mathrm{Bel}(\neg A)$$

Łączenie obserwacji pochodzących z różnych źródeł

Załóżmy, że dane są dwa źródła obserwacji **X** i **Y** reprezentowane odpowiednio przez podzbiory zbioru Θ: $\mathbf{X}_1,\ldots,\mathbf{X}_m$ oraz $\mathbf{Y}_1,\ldots,\mathbf{Y}_n$. Ponadto na zbiorach **X** i **Y** zdefiniowane są funkcje gęstości prawdopodobieństwa, odpowiednio m_1 i m_2. Połączenie obserwacji z obu źródeł pozwala obliczyć nową wartość $m_3(\mathbf{Z})$ dla każdego podzbioru zbioru Θ następująco:

$$m_3(Z) = \frac{\sum_{X_i \cap Y_j = Z} m_1(X_i) m_2(Y_j)}{1 - \sum_{X_i \cap Y_j = \varnothing} m_1(X_i) m_2(Y_j)}$$

Przykład

A – allergy
F – flu
C – cold
P - pneumonia

Θ ={A, F, C, P} $m_1(\Theta) = 1$

Obserwacja 1

$m_2(\{A, F, C\}) = 0,6$

$m_2(\Theta) = 0,4$

	$m_2(\{A, F, C\}) = 0,6$	$m_2(\Theta) = 0,4$
$m_1(\Theta) = 1$	$m_3(\{A, F, C\}) = 0,6$	$m_3(\Theta) = 0,4$

Przykład

$m_3(\{A, F, C\}) = 0,6$ $m_3(\Theta) = 0,4$

Obserwacja 2

$m_4(\{F,C,P\}) = 0,8$

$m_4(\Theta) = 0,2$

	$m_4(\{F,C,P\}) = 0,8$	$m_4(\Theta) = 0,2$
$m_3(\{A,F,C\}) = 0,6$	$m_5(\{F,C\}) = 0,48$	$m_5(\{A,F,C\}) = 0,12$
$m_3(\Theta) = 0,4$	$m_5(\{F,C,P\}) = 0,32$	$m_5(\Theta) = 0,08$

Przykład

$m_5(\{F,C\}) = 0,48$	$m_5(\{A,F,C\}) = 0,12$
$m_5(\{F,C,P\}) = 0,32$	$m_5(\Theta) = 0,08$

Obserwacja 3

$m_6(\{A\}) = 0,75$

$m_6(\Theta) = 0,25$

$m_7(\{A\}) = 0,15$

	$m_6(\{A\}) = 0,75$	$m_6(\Theta) = 0,25$
$m_5(\{F,C\}) = 0,48$	$m_7(\varnothing) = 0,36$	$m_7(\{F,C\}) = 0,12$
$m_5(\{A,F,C\}) = 0,12$	$m_7(\{A\}) = 0,09$	$m_7(\{A,F,C\}) = 0,03$
$m_5(\{F,C,P\}) = 0,32$	$m_7(\varnothing) = 0,24$	$m_7(\{F,C,P\}) = 0,08$
$m_5(\Theta) = 0,08$	$m_7(\{A\}) = 0,06$	$m_7(\Theta) = 0,02$

Przykład

$m_5(\{F,C\}) = 0{,}48$	$m_5(\{A,F,C\}) = 0{,}12$
$m_5(\{F,C,P\}) = 0{,}32$	$m_5(\Theta) = 0{,}08$

Obserwacja 3

$m_6(\{A\}) = 0{,}75$

$m_6(\Theta) = 0{,}25$

$m_7(\emptyset) = 0{,}6$

	$m_6(\{A\}) = 0{,}75$	$m_6(\Theta) = 0{,}25$
$m_5(\{F,C\}) = 0{,}48$	$m_7(\emptyset) = 0{,}36$	$m_7(\{F,C\}) = 0{,}12$
$m_5(\{A,F,C\}) = 0{,}12$	$m_7(\{A\}) = 0{,}09$	$m_7(\{A,F,C\}) = 0{,}03$
$m_5(\{F,C,P\}) = 0{,}32$	$m_7(\emptyset) = 0{,}24$	$m_7(\{F,C,P\}) = 0{,}08$
$m_5(\Theta) = 0{,}08$	$m_7(\{A\}) = 0{,}06$	$m_7(\Theta) = 0{,}02$

Przykład

$m_7(\{A\}) = 0{,}15$
$m_7(\{F,C\}) = 0{,}12$
$m_7(\{A,F,C\}) = 0{,}03$
$m_7(\{F,C,P\}) = 0{,}08$
$m_7(\Theta) = 0{,}02$

$m_7(\{A\}) = 0{,}375$
$m_7(\{F,C\}) = 0{,}3$
$m_7(\{A,F,C\}) = 0{,}075$
$m_7(\{F,C,P\}) = 0{,}2$
$m_7(\Theta) = 0{,}05$

{*A*}: [0,375, 0,500] — 1 – 0,3 – 0,2

{*F*}: [0, 0,625] — 1 – 0,375

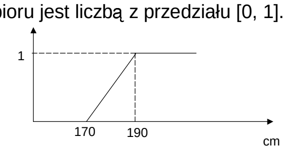
{*C*}: [0, 0,625]

{*P*}: [0, 0,250] — 1 – 0,375 – 0,3 – 0,075

Teoria zbiorów rozmytych (Zadeh)

Miara przynależności do zbioru jest liczbą z przedziału [0, 1].

Zbyszek jest wysoki.



$$T(E_1 \wedge E_2) = \min\{T(E_1), T(E_2)\}$$

$$T(E_1 \vee E_2) = \max\{T(E_1), T(E_2)\}$$

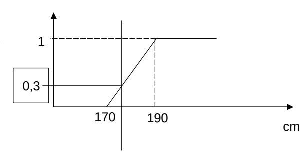
$$T(\neg E_1) = 1 - T(E_1)$$

$$T(A \vee \neg A) \neq 1$$

Teoria zbiorów rozmytych

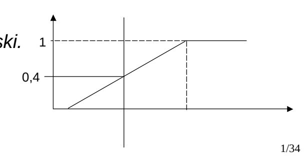
Miara przynależności do zbioru jest liczbą z przedziału [0, 1].

Zbyszek jest wysoki.



oraz

Zbyszek jest towarzyski.



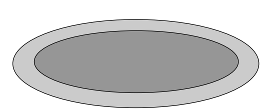
Teoria zbiorów przybliżonych (Pawlak)

Głównym celem zbiorów przybliżonych jest wprowadzenie koncepcji aproksymacji pojęć. Stosowane są głównie w odkrywaniu wiedzy.

System informacyjny (U, A) składa się z niepustego zbioru obiektów U oraz niepustego zbioru atrybutów A, $a: U \rightarrow V_a$, dla każdego $a \in A$ gdzie V_a jest nazywany zbiorem wartości A.

Teoria zbiorów przybliżonych (Pawlak)

Niech T = (U, A), $B \subseteq A$ oraz $X \subseteq U$. Aproksymujemy X używając tylko informacji zawartej w B poprzez skonstruowanie B-dolnej i B-górnej aproksymacji X:



Podsumowanie

- Rachunek prawdopodobieństwa jest daleki od intuicyjnego rozumienia niepewności
- Wnioskowanie z zastosowaniem twierdzenia Bayesa jest bardzo złożone obliczeniowo
- Współczynnik pewności, teoria Dempstera-Shefera teoria zbiorów rozmytych i teoria zbiorów przybliżonych są alternatywnymi narzędziami modelowania wiedzy niepewnej dla celów automatycznego wnioskowania
- Sieci bayesowskie stanowią metodę wnioskowania z zastosowaniem twierdzenia Bayesa redukującą złożoność obliczeń