Bài 8

TRI THỨC KHÔNG CHẮC CHẮN

8.1. Tri thức không chắc chắn và xác suất

- Tri thức có thể không chắc chắn, ví dụ: mệnh đề “hút thuốc thì bị ung thư” là không chắc chắn.

- Xác suấtcủa mệnh đề là khả năng mệnh đề có thể đúng và ký hiệu là p ($0 \leq p \leq 1$). Ví dụ xác suất “hút thuốc thì bị ung thư” là p=0,63.

- **Xác suất có thể dựa vào:**
  - thống kê,ví dụtheo thống kê cứ 100 người hút thuốc thì có 63 người bị ung thư, vậy xác suất “hút thuốc thì bị ung thư” là 0,63.
  - hoặc dựa vào sự hiểu biết, kinh nghiệm, ví dụ xác suất “Real Madrid thắng Valencia trong trận tới” là 0,58

- **Các ký hiệu xác suất:**
  Pr (probability): xác suất
  Pr(X): xác suất mệnh đề X đúng
  Pr(¬X): xác suất mệnh đề X sai.
  Pr(X∧Y) hoặc Pr(X, Y): xác suất cả hai mệnh đề X và Y đều đúng
  Pr(X∨Y): xác suất mệnh đề X hoặc mệnh đề Y là đúng.

- **Các tính chất cơ bản**
  - Mệnh đề chắc chắncó xác suất là 1,mệnh đề không thoả được có xác suất là 0.
  - $Pr(X \vee Y) = Pr(X) + Pr(Y) - Pr(X \wedge Y)$
  - $Pr(\neg X) = 1 - Pr(X)$

  Cm: $Pr(X \vee \neg X) = Pr(X) + Pr(\neg X) - Pr(X \wedge \neg X) \Rightarrow Pr(\neg X) = 1 - Pr(X)$

- **Xác suất có điều kiện**
  Xác suất có điều kiện của X khi cho trước Y được ký hiệu là Pr(X|Y).
  Ta có:$Pr(X,Y) = Pr(X|Y)\, Pr(Y)$

- **Biến ngẫu nhiên và phân phối xác suất**
  Biến ngẫu nhiênlà một biến nhận giá trị một cách ngẫu nhiên từ một tập các giá trị,tập giá trịnày được gọi là miền giá trị (hay không gian mẫu) của biếnngẫu nhiên. Ta chỉ xét các biến ngẫu nhiên rời rạc,tức là miền giá trị của nó là một tập các giá trị rời rạc.

  Giả sử X là biến ngẫu nhiên với miền giá trị Ωvà x∈Ω. Mệnh đề“X = x” là viết tắt của mệnh đề “biến X nhận giá trị x”và gọi là mệnh đề phân tử.

  <u>Phân phối xác suất của X là hàm Pr(X)</u> xác định trên miền giá trị Ω, ứngvới mỗi x ∈Ωvới xác suất Pr(X = x). Hàm Pr(X) cần thoả mãn điều kiện:

  $$\sum_{x \in \Omega} Pr(X = x) = 1$$

  Ví dụ: Phân phối xác suất của biến ngẫu nhiên Weather ∈ Ω=(sunny, Rain,Cloudy) cho trong bảngsau:

Weather	Pr(Weather)
Sunny	0,6
Rain	0,3
Cloudy	0,1

- **Phân phối xác suất có điều kiện**

Giả sử X và Y là hai biến ngẫu nhiên với miền giá trị tương ứng là Ω_X và Ω_Y.*Phân phối xác suất có điều kiện của X khi đã cho giá trị của Y là hàm Pr(X|Y)*ứng mỗi cặp giá trị x $\in\Omega_X$ và y $\in\Omega_Y$ với xác suất có điều kiện Pr(X = x|Y =y). Hàm Pr(X|Y) cần thoả mãn điều kiện:

$$\sum_{x\in\Omega} \Pr(X=x \mid Y=y)=1, \text{ với mọi } y \in \mathbf{\Omega}_y.$$

Ví dụ. Phân phối xác suất Pr (X|Y) với $\Omega_X = \Omega_Y$={true,false} đượccho trong bảng sau:

		X = true	X = false
Y=true	Pr (X|Y=true)	0,8	0,2
Y=false	Pr (X|Y=false)	0,3	0,7

Cần lưu ý rằng, tổng của tất cả các số trong một dòng cần bằng 1.

- **Phân phối xác suất kết hợp**

Xét một tập biến ngẫu nhiên{X_1, X_2, …, X_n} với $X_i$nhận giá trị trongmiền giá trị Ω_i(i = 1, …, n) tương ứng. Giả sử x_i(i = 1, …, n) là một giá trị thuộc Ω_i,xác suất của mệnh đề (X_1= x_1) ∧(X_2= x_2) ∧…∧(X_n= x_n) ký hiệu là Pr(X_1= x_1, X_2= x_2, …, X_n= x_n).

*Phân phối xác suất kết hợp của tập {X_1, ..., X_n} là hàm Pr(X_1, X_2, ..., X_n)*ứng mỗi mệnh đề (X_1= x_1) ∧(X_2= x_2) ∧…∧(X_n= x_n) với xácsuất Pr(X_1= x_1, X_2= x_2,…, X_n= x_n). HàmPr(X_1, X_2, …, X_n) cần thoả mãn điều kiện sau:

$$\sum_{x_1\in\Omega_1}\cdots\sum_{x_n\in\Omega_n} \Pr(X_1 = x_1,\ldots, X_n = x_n) = 1$$

Ví dụ.Giả sử X và Y là hai biến ngẫu nhiên boolean. Phân phối xácsuất kết hợp Pr(X,Y) được cho trong bảng sau:

	X = true	X = false
Y = true	0,1	0,2
Y = false	0,3	0,4

Cần lưu ý rằng, tổng của tất cả các số trong bảng phân phối xác suấtkết hợp cần bằng 1.

Nếu biến ngẫu nhiên X nhận giá trịlà vectơ (x_1, …, x_n), với $x_i \in \Omega_i$, nghĩa là X nhận giá trị trong miền giá trị $\Omega = \Omega_1 \times\ldots\times\Omega_n$, thì phân phối xác suất kết hợp Pr(X_1, …, X_n) chính là phân phối xác suấtcủa X.

8.2 Các công thức tính xác suất

- **Luật tập con:**

Nếu biết phân phối xác suất kết hợp Pr(X_1,…,X_n)của tập biến ngẫu nhiên {X_1, …, X_n}, ta có thể tính được phânphối xác suất kết hợp của một tổ hợp bất kỳ của các biến {X_1, …,X_n}.
Ví dụ:
Giả sử chúng ta đã biết phân phối xácsuất Pr (X,Y,Z), trong đó X, Y, Z là các biến thuộc các miền giá trịΩ_X, Ω_Y, Ω_Ztương ứng. Khi đó ta có thể tính được phân phối xác suất kết hợp củahai biến Pr(X,Y) bằng công thức sau:

$$\Pr(X,Y) = \sum_{z\in\Omega_Z} \Pr(X,Y,Z = z)$$

và tính được phân phối xác suất của một biến Pr(X) theo công thức sau:

$$\Pr(X) = \sum_{y \in \Omega_Y} \sum_{z \in \Omega_Z} \Pr(X, Y = y, Z = z)$$

Ví dụ: Giả sử X và Y là hai biến ngẫu nhiên boolean vàphân phối xác suất kết hợp Pr(X,Y) được cho trong bảng sau:

	Y = true	Y = false
X = true	0,1	0,2
X = false	0,3	0,4

Suy ra phân phối xác suất của X là Pr(X) và phân phối xác suất của Y là Pr(Y) cho trong bảng sau:

X	Pr(X)
true	0,3
false	0,7

Y	Pr(Y)
True	0,4
False	0,6

- **Luật tổng**

  Từ các công thức

$$\Pr(X) = \sum_{y \in \Omega_Y} \Pr(X, Y = y)$$

$$\Pr(X, Y = y) = \Pr(X|Y = y)\Pr(Y = y)$$

Suy ra luật tổng:

$$\Pr(X) = \sum_{y \in \Omega_Y} \Pr(X|Y = y)\Pr(Y = y)$$

- **Luật tích**

  Ta có:

  Pr(Y,Z) = Pr(Y|Z) Pr(Z)

  Pr(X,Y,Z) = Pr(X|Y,Z) Pr(Y,Z)

  Suy ra:

  Pr(X,Y,Z) = Pr(X|Y,Z) Pr(Y|Z) Pr(Z)

Tổng quát ta có luật tích:

$\Pr(X_1, X_2, \ldots, X_n) = \Pr(X_1|X_2, \ldots, X_n)\Pr(X_2|X_3, \ldots, X_n) \ldots \Pr(X_{n-1}|X_n)\Pr(X_n)$

Do thứ tự của các biến trong phân phối xác suất kết hợp là không quan trọng nênsẽ có n! cách phân tích một xác suất kết hợp thành tích của các xác suất có điều kiện.

- **Công thức bayes**

  Theo công thức tính xác suất có điều kiện ta có:

$$\Pr(X|Y) = \frac{\Pr(X,Y)}{\Pr(Y)}$$

Mặt khác, Pr(X,Y) = Pr(Y|X)Pr(X) => công thức Bayes:

$$\Pr(X|Y) = \frac{\Pr(Y \mid X)\Pr(X)}{\Pr(Y)}$$

Để tính Pr(X|Y) theocông thức Bayes, ta chỉ cần biết Pr(Y|X) vàPr(X), vì nếu biết các xác suất này thì theo luật tổng sẽ tínhđược Pr(Y).

Ví dụ:
Giả sử xác suất một bệnh nhân sâu răng (cavity) bịđau răng (toothache) là 0,65=> Pr(Toothache|Cavity) = 0,65 và xác suất của một người sâu răng 0,05 => Pr(Cavity) = 0,05 và xácsuất của một người đau răng là 0,04 => Pr(Toothache) = 0,04. Cần tính xác suất để một người đau răng bị sâu răng. Theocông thức Bayes:

$$\Pr(\text{Cavity}|\text{Toothache}) = \frac{\Pr(\text{Toothache} \mid \text{Cavity})\Pr(\text{Cavity})}{\Pr(\text{Toothache})}$$

$$= \frac{0,65 * 0,05}{0,04} = 0,81$$

Như vậy, xác suất để một người đau răng bị sâu răng là 0,81.

8.3 Sự độc lập có điều kiện của các biến ngẫu nhiên

Biến X gọi là độc lập vớibiến Y nếu:

Pr(X|Y) = Pr(X)

Biến X gọi là độc lập có điều kiện với biến Y khi cho trước Z, nếu:

Pr(X|Y,Z) = Pr(X|Z)

Nếu X độc lập có điều kiện với Y khi cho trước Z. Ta có công thức:

Pr(X,Y,Z) = Pr(X|Z) Pr(Y|Z) Pr(Z)

8.4. Mạng xác suất

Nếu biết được phân phối xác suất Pr(X1,…,Xn)thì áp dụng luật tập con ta có thể tính được xác suất của các biến Xi (i=1,…,n) và áp dụng luật tích ta tính được xác suất có điều kiện bất kỳ của các biến Xi.

Thường rất khó xác định được Pr(X1, …, Xn) và để lưu Pr(X1, …, Xn) ta cần bảng n chiều.Chỉ với n biến boolean thì bảng cũng đã chứa 2n số!

Mạng xác suất là mô hình thích hợp để biểu diễn tri thức không chắc chắn. Mô hình này giúp tagiảm bớt số các xác suất ban đầu cần biết trước và đơn giản sự tính toán để tìm ra câu trả lời cho các câu hỏi.

8.4.1. Định nghĩa

Mạng xác suất là một đồ thị có hướng, không có chu trình và thoả mãn các điều kiện sau:

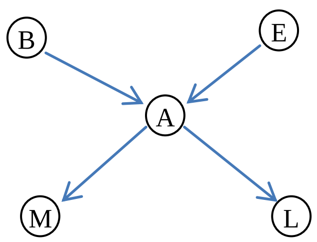
- Các đỉnh của đồ thị là các biến ngẫu nhiên.
- Mỗi cung từ đỉnh X đến đỉnh Y biểu diễn Y phụ thuộc trực tiếp vào X, X gọi là đỉnh cha của Y.
- Nếu $X_1, \ldots, X_n$ là các đỉnh cha của đỉnh Y thì đỉnh Y cần biết $\Pr(Y|X_1, \ldots, X_n)$.
- Nếu Y không có cha thì đỉnh Y cần biết Pr(Y).

Ví dụ:Nhà bạn có lắp đặt hệ thống báo động trộm. Nó sẽ kêu khi phát hiện ra trộm hoặc khi có động đất nhẹ. Khi bạn đi làm, bạndặn hai người hàng xóm là Lan và Mai hãy gọi cho bạn nếu nghe thấy chuông báo trộm kêu. Lan đôi khi nhầm lẫn chuông điện thoại với chuông báo trộm và cũng gọi cho bạn. Mai thì hay nghe nhạc to nên đôi khi không nghe thấy chuông báo trộm.

Đặt:

B: "có trộm"
E: "có động đất nhẹ"
A: "chuông báo trộm kêu"
L: "Lan gọi cho bạn"
M: "Mai gọi cho bạn"

Ta có mạng xác suất báo động trộm như hình 1.



Hình 1: mạng xác suất báo trộm

Trong mạng trên, ta mới chỉ đưa vào các quan hệ trực tiếp giữa các sự kiện,sự kiện còn Lan đôi khi nhầm lẫn chuông điện thoại với chuông báo trộm và Mai vì nghe nhạc to mà không nghe thấy chuông báo trộm, thìchưa được xét. Thực ra còn vô số nguyên nhân làm cho chuông báo trộm kêu hoặc không kêu (chẳng hạn, các nguyên nhân về thời tiết, về điện …). Cũng như vậy, còn có rất nhiều nguyên nhân khác làm cho Lan và Mai gọi hoặc không gọi cho bạn (chẳng hạn, Lan tình cờ nghe thấy tiếng kêu gì đó giống như tiếng chuông báo trộm và gọi cho bạn, hoặc có lúc chuông báo trộm thì Mai vừa ra khỏi nhà và không gọi cho bạn được, …).

Ta cần 5 bảng phân phối xác suấtgiả sử cho như sau:

B	Pr(B)
True	0,01
False	0,99

E	Pr(E)
True	0,02
False	0,98

B	E	Pr(A|B,E)	
		A = True	A = False
True	True	0,95	0,05
True	False	0,93	0,07
False	True	0,29	0,71
False	False	0,01	0,99

A	Pr(L|A)	
	L = True	L = False
True	0,92	0,08
False	0,05	0,95

A	Pr(M|A)	
	M = True	M = False
True	0,8	0,2
False	0,01	0,99

8.4.2. Lập luận trong mạng xác suất

Cho một tập các biến đã biết giá trị E (tập các biến bằng chứng) ta cần tínhphân phối xác suất có điều kiện của một biến X (biến hỏi) nào đó, nghĩa là tính Pr(X|E).
Xét lại ví dụ trên, giả sửLan và Mai đều gọi khi đótập các biến bằng chứng là E ={L = true, M = true} và giả sử muốn biết khả năng có trộm, khi đó biến hỏi là B vàcần tính Pr(B = true| L = true, M = true).

Có hai dạng lập luận:

• Lập luận chẩn đoán: các biến bằng chứng là hậu thế của biến hỏi.
Ví dụ Lan gọi cho bạn, tính khả năng có trộm. Ta có biến bằng chứng là L, biến hỏi là B và B->A->L.Cầntính Pr(B|L) = 0,016.

• Lập luận tiên đoán: các biến bằng chứnglà tiền thân của biến hỏi.
Ví dụbiết có trộm, tính khả năng Lan hoặc Mai gọi điệncho bạn.Ta cóbiến bằng chứng là B, các biến hỏi là L và M. Cần tính Pr(L|B) = 0,86 và Pr(M|B) = 0,67.

Trong trường hợp tổng quát, các biến bằng chứng có thể không phải làtiền thân hoặc không phải là hậu thế của biến hỏi. Chẳng hạn, chúng ta biếttrước rằng không có động đất và Lan gọi. Chúng ta cần đánh giá khả năngcó trộm khi được biết các thông tin đó. Có thể tính được:
Pr(B = true|E = false, L = true) = 0,03.

Vấn đề tính chính xác phân phối xác suất Pr(X|E) của biến hỏi X khiđược cho tập bằng chứng E là vấn đề NP khó. Tuy nhiên, trong một sốtrường hợp đặc biệt người ta có thể tính được Pr(X|E) trong thời gian đa thức. Trong các mục sau chúng ta sẽ trình bày các thuật toán cho phép ta tính chính xác Pr(X|E) với thời gian tỷ lệ với số đỉnh trong mạng, trongtrường hợp mạng xác suất có kết nối đơn. Sau đó chúng ta sẽ trình bày cácthuật toán cho phép ta tính xấp xỉ Pr(X|E) trong trường hợp tổng quát.