Chương 2

Lượng tin

Nguyễn Thanh Bình

Khoa CNTT&TT - Đại học Cần Thơ

01 - 2011

Nội dung

Ví dụ về đo lường thông tin

Ví dụ 1

- Xét biến ngẫu nhiên $X = \{1, 2, 3, 4, 5\}$ có phân phối xác suất đều:

$$p(X = i) = \frac{1}{5}$$

- Xét biến ngẫu nhiên $Y = \{1, 2\}$ cũng có phân phối xác suất đều:

$$p(Y = j) = \frac{1}{2}$$

- Khi đó ta nói Y cho lượng tin về nó nhiều hơn X, bởi vì nói đến Y ta biết nó chỉ có giá trị là 1 hoặc 2, nhưng khi nói đến X thì X có đến 5 giá trị có xác suất như nhau.

Ví dụ về đo lường thông tin

Ví dụ 2

- Xét tiếp biến ngẫu nhiên $W = \{1, 2\}$ với phân phối xác suất **không đều**:

$$p(W = 1) = \frac{1}{10} \qquad p(W = 2) = \frac{9}{10}$$

- Ta nói W cho lượng tin nhiều hơn Y, bởi vì khi nói đến W ta luôn biết khả năng $W = 2$ là lớn hơn nhiều so với $W = 1$.

Nhận xét

Các ví dụ trên cho ta biết về lượng tin nhiều hay ít, nhưng chưa biết là bao nhiêu $\Rightarrow$ cần xây dựng một đại lượng toán học để đo lượng tin chưa biết của một biến ngẫu nhiên bất kỳ.

Độ đo lượng tin

Các tính chất

1. Một sự kiện có xác suất càng nhỏ (ít xảy ra) $\Rightarrow$ tính không chắc chắn càng lớn (hay lượng tin không biết càng lớn).
2. Một biến ngẫu nhiên càng có nhiều sự kiện có phân phối xác suất bằng nhau thì tính không chắc chắc càng lớn.
3. Một biến ngẫu nhiên có phân phối xác suất càng lệch nhiều (có xác suất rất nhỏ và rất lớn) thì tính không chắc chắn càng ít.

Khái niệm về Entropy

Khái niệm

Entropy là một đại lượng toán học dùng để đo lượng tin không chắc (hay lượng ngẫu nhiên) của một sự kiện hay của một biến ngẫu nhiên với một phân phối xác suất biết trước.

Tính Entropy

Bài toán

Xét một thí nghiệm xác suất đòi hỏi quan sát một biến ngẫu nhiên rời rạc X trên không gian mẫu $A = \{a_1, a_2, \ldots, a_{M-1}, a_M\}$ có phân phối xác suất $p_i(i = 1, 2, \ldots, M)$ như sau:

$$X(a_i) = x_i(i = 1, 2, \ldots, M)$$

X	x_1	x_2	...	x_{M-1}	x_M
	p_1	p_2	...	p_{M-1}	p_M

$$\text{với}: \begin{cases} p_i \geq 0(i = 1, 2, \ldots, M) \\ p_1 + p_2 + \ldots + p_M = 1 \end{cases}$$

Tính Entropy

Hàm số Entropy $h(p)$

- Để đo lượng thông tin không chắc chắn của X, người ta xây dựng hàm số $h(p)$ ($p \subseteq [0,1]$), gọi là Entropy của một sự kiện cơ bản có xác suất p.
- Shannon đã đưa ra công thức cụ thể $h(p) = -C \log p$.

Hệ tiên đề

- $h(p)$ là hàm liên tục không âm và đơn điệu giảm
- giả sử A và B là 2 sự kiện độc lập nhau, có xác suất xuất hiện lần lượt là p_A và p_B. Khi đó $p(A,B) = p_A.p_B$, nhưng:

$$h(p_A, p_B) = h(p_A) + h(p_B).$$

Tính Entropy

Định lý dạng giải tích của Entropy

Người ta cũng xây dựng hàm $H_M(p_1, p_2, \ldots, p_{M-1}, p_M)$ gọi là Entropy trung bình của các sự kiện $\{X = x_i\}$(viết gọn là $H(X)$).

$$H(X) = H_M(p_1, p_2, \ldots, p_{M-1}, p_M) = \sum_{i=1}^{M} p_i.h(p_i)$$

Entropy cơ số 2

Khi $C = 1$ và cơ số logarithm là 2 *(đơn vị tính: bit)*, ta có:

$$H(X) = H(p_1, p_2, \ldots, p_{M-1}, p_M) = -\sum_{i=1}^{M} p_i \log_2(p_i)$$

Ví dụ minh họa

Ví dụ 1

- Xét biến ngẫu nhiên X có phân phối:

X	x_1	x_2	x_3
P	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

- Ta có:

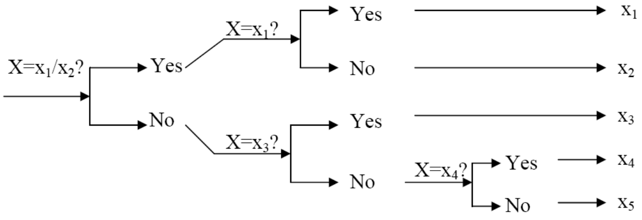
$$H(X) = H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) = \frac{3}{2}(bit)$$

Ví dụ minh họa

Ví dụ 2: Bài toán cây tìm kiếm nhị phân

- Số câu hỏi trung bình: 2.3
- Entropy H(X) = 2.27

$\Rightarrow$ sơ đồ câu hỏi là tối ưu.



Nhận xét về Entropy

- Qua ví dụ 2, ta có thể hiểu Entropy là **số bit trung bình tối thiểu** để có thể biểu diễn tất cả các giá trị của X.
- Khi xem xét công thức: $H(X) = -p_1 \log p_1 \ldots - p_M \log p_M$ có thể nói rằng $H(X)$ chính là trọng số trung bình theo các xác suất p_i của các giá trị $-\log p_i = \log \frac{1}{p_i}$.

X	x_1	$\ldots$	x_M
	p_1	$\ldots$	p_M

W	$\log \frac{1}{p_1}$	$\log \frac{1}{p_2}$	$\ldots$	$\log \frac{1}{p_M}$
	p_1	p_2	$\ldots$	p_M

- Đây chính là nghịch đảo của giá trị được thừa nhận bởi $X \Rightarrow$ tính không chắc chắn của X.

Tính chất 1

Phân tích

- Xét biến ngẫu nhiên rời rạc X có phân phối đều.
- Đặt $H(X) = H(\frac{1}{M}, \ldots, \frac{1}{M}) = f(M)$.
- Ta mong muốn: nếu $M < M' \Leftrightarrow \frac{1}{M} > \frac{1}{M'}$ thì $f(M) < f(M')$ *xem lại tính chất 1 của độ đo lượng tin: xác suất nhỏ, entropy lớn*.

Tính chất

$$f(M) = H(\frac{1}{M}, \ldots, \frac{1}{M}) \text{ đơn điệu tăng}$$

Tính chất 2

Phân tích

- Xét 2 biến ngẫu nhiên độc lập X, Y có phân phối đều với X có M sự kiện và Y có L sự kiện $\Rightarrow (X, Y)$ có ML sự kiện.
- Do X, Y độc lập việc biết X không giúp gì cho việc biết Y.
- Ta mong muốn: Entropy của (X, Y) mà trừ đi Entropy bị bỏ đi vì biết X (chính là $H(X) = f(M)$) sẽ là Entropy của Y.

Tính chất

$$f(ML) = f(M) + f(L)$$

Tính chất 3

Phân tích

- Xét biến ngẫu nhiên rời rạc X (một cách tổng quát)

X	x_1	...	x_r	x_{r+1}	...	x_M
P	p_1	...	p_r	p_{r+1}	...	p_M

- Phân hoạch X thành 2 nhóm: $A = \{x_1, \ldots, x_r\}$ và $B = \{x_{r+1}, \ldots, x_M\}$.
- Xét thí nghiệm chọn với không gian mẫu là A hoặc B.
- Xác suất của x_i trong không gian mẫu A: $p_i/(p_1 + \ldots + p_r)$.
- Tương tự, xác suất của x_i trong không gian mẫu B:

$$\frac{p_i}{\sum_{j=r+1}^{M} p_j}$$

Tính chất 3

Phân tích

- Gọi Y là biến ngẫu nhiên của thí nghiệm. Xét $p(Y = x_i)$:
  - Hoặc là:
    $p(Y = x_i) = p(A \text{ được chọn và } x_i \in A) = \sum_{j=1}^{r} p_j \frac{p_i}{\sum_{j=1}^{r} p_j} = p_i$
  - Hoặc là:
    $p(Y = x_i) = p(B \text{ được chọn và } x_i \in B) = \sum_{j=r+1}^{M} p_j \frac{p_i}{\sum_{j=r+1}^{M} p_j} = p_i$
- Vậy X và Y có cùng phân phối xác suất $\Rightarrow H(X) = H(Y)$.
- Phân tích thí nghiệm ta thấy việc chọn một phần tử x_i gồm:
  - Chọn không gian mẫu A hoặc B.
  - Chọn phần tử x_i khi không gian A (hoặc B) được chọn.

Tính chất 3

Phân tích

- Ta có: Entropy của X = Entropy của Y =

  Entropy của việc chọn không gian mẫu
  \+ Entropy của A nhân với xác suất chọn A
  \+ Entropy của B nhân với xác suất chọn B

Tính chất

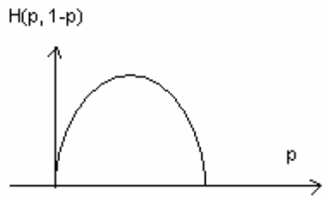
Đặt $P_A = \sum_{i=1}^{r} p_i$ và $P_B = \sum_{i=r+1}^{M} p_i$.

$$
\begin{aligned}
H(p_1, \ldots, p_M) &= H(P_A, P_B) \\
&+ P_A H\left(\frac{p_1}{P_A}, \ldots, \frac{p_r}{P_A}\right) + P_B H\left(\frac{p_{r+1}}{P_B}, \ldots, \frac{p_M}{P_B}\right)
\end{aligned}
$$

Tính chất 4

Phân tích

Về mặt toán học, ta yêu cầu $H(p, 1-p)$ phải là hàm liên tục theo p.



Tính chất

$$H(p, 1-p) \text{ là hàm liên tục theo } p$$

Tính chất cực đại của Entropy

Bổ đề

- Cho 2 bộ số dương $\{p_1, \ldots, p_M\}$ và $\{q_1, \ldots, q_M\}$ thỏa

$$\sum_{i=1}^{M} p_i = \sum_{i=1}^{M} q_i = 1$$

- Khi đó: $-\sum_{i=1}^{n} p_i \log_2(p_i) \leq -\sum_{i=1}^{n} p_i \log_2(q_i)$
- Đẳng thức xảy ra khi và chỉ khi: $p_i = q_i$ với $\forall i = 1 \ldots M$

Định lý

$$H(p_1, \ldots, p_M) \leq \log_2 M$$

Đẳng thức xảy ra khi và chỉ khi: $p_i = \frac{1}{M}$ với $\forall i$

Định nghĩa

- Giả sử X và Y là 2 biến ngẫu nhiên cho trước với:

$$p_{ij} = p(X = x_i, Y = y_j) \qquad \forall i = 1 \ldots M, \forall j = 1 \ldots L$$

- Khi đó, Entropy $H(X, Y)$ có dạng:

$$H(X,Y) = -\sum_{i=1}^{M}\sum_{j=1}^{L} p(x_i, y_j)\log_2 p(x_i, y_j) = -\sum_{i=1}^{M}\sum_{j=1}^{L} p_{ij}\log_2 p_{ij}$$

- Tổng quát:

$$H(X_1, \ldots, X_n) = -\sum_{\forall(x_1 \ldots x_n)} p(x_1, \ldots, x_n)\log p(x_1, \ldots, x_n)$$

Ví dụ

- Cho 2 biến ngẫu nhiên X và Y độc lập nhau.

X	0	1
P	0.5	0.5

Y	0	1	2
P	0.25	0.5	0.25

- Phân phối của $P(X, Y)$

X,Y	(0,0)	(0,1)	(0,2)	(1,0)	(1,1)	(1,2)
P	0.125	0.25	0.125	0.125	0.25	0.125

- $H(X, Y) = H(0.125, 0.25, 0.125, 0.125, 0.25, 0.125) = 2.5(bit)$

Định nghĩa

- Entropy của Y với điều kiện $X = x_i (i = 1 \dots M)$:

$$H(Y|X = x_i) = -\sum_{j=1}^{L} p(y_j|x_i) \log p(y_j|x_i)$$

- Entropy của Y với điều kiện X xảy ra:

$$H(Y|X) = \sum_{i=1}^{M} p(x_i) H(Y|X = x_i)$$

Ví dụ

X	1	2
P	0.5	0.5

Y | X=1	0	1	2
P	0.25	0.5	0.25

Y | X=2	0	1	2
P	0	0	1

Tính các Entropy

- $H(Y|X=1) = H(0.25, 0.5, 0.25) = 1.5(bit)$
- $H(Y|X=2) = H(0, 0, 1) = 0(bit)$
- $H(Y|X) = p(X=1)H(Y|X=1) + p(X=2)H(Y|X=2) = 0.5 \times 1.5 + 0.5 \times 0 = 0.75(bit)$

Quan hệ giữa $H(X, Y)$ với $H(X)$ và $H(Y)$

Định lý 1

$$H(X, Y) \leq H(X) + H(Y)$$

Đẳng thức xảy ra khi và chỉ khi X, Y độc lập.

Định lý 2

$$H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$$

Định lý 3

$$H(Y|X) \leq H(Y)$$

Đẳng thức xảy ra khi và chỉ khi X, Y độc lập.

Bài toán

Đồng tiền thật, đồng tiền giả

- Trò chơi tung đồng xu: có đầu hình *(thua)* - không có đầu hình *(thắng)*.
- Đồng tiền loại 1 *(thật)*: đồng chất, 1 mặt có đầu hình.
- Đồng tiền loại 2 *(giả)*: đồng chất, 2 mặt có đầu hình.
- Quá trình chọn đồng tiền là ngẫu nhiên.
- Tung đồng tiền 2 lần và đếm số đầu hình xuất hiện.

Xác định các phân phối

Phân phối của X

X	1	2
P	0.5	0.5

Phân phối có điều kiện của Y

Y | X=1	0	1	2
P	0.25	0.5	0.25

Y | X=2	0	1	2
P	0	0	1

Phân phối toàn phần của Y

Y	0	1	2
P	0.125	0.25	0.625

Tính các Entropy

- $H(X) = H(0.5, 0.5) = 1(bit)$
- $H(Y) = H(0.125, 0.25, 0.625)$
- $H(X, Y)$: lập bảng phân phối của (X, Y).
- $H(Y|X) = p(X = 1)H(Y|X = 1) + p(X = 2)H(Y|X = 2)$
  $$= 0.5\,H(0.25, 0.5, 0.25) + 0.5\,H(0, 0, 1)$$
- $H(X|Y)$: lập bảng phân phối xác suất cho các trường hợp
  $$(X|Y = 0), (X|Y = 1) \text{ và } (X|Y = 2)$$
- Kiểm tra quan hệ giữa các Entropy theo định lý 1, 2 và 3.

Lượng tin

Định nghĩa

Lượng tin (hay thông lượng) của X khi Y xảy ra, ký hiệu $I(X|Y)$, là sự chênh lệch giữa lượng tin không chắc chắn của X và lượng không chắc chắc của X khi Y đã xảy ra.

$$I(X|Y) = H(X) - H(X|Y)$$

Tính chất

$$I(X|Y) = I(Y|X)$$

Minh họa lượng tin

Ví dụ

Bài toán: đồng tiền thật, đồng tiền giả.

- $H(Y) = 1.3(bit)$
- $H(Y|X) = 0.75(bit)$
- $I(X|Y) = I(Y|X) = H(Y) - H(Y|X) = 0.55(bit)$