Chương 3. Các đại lượng thông tin

3.1. Lượng tin riêng

3.2. Entropy

3.3. Lượng tin tương hỗ

3.4. Nguồn tin

3.5. Kênh

3.6. Phối hợp nguồn – kênh

3.1. Lượng tin riêng

Lưu ý:

- Một nguồn có mô hình là một biến ngẫu nhiên
- Thông tin là một khái niệm trừu tượng mô tả sự hiểu biết về đối tượng xung quanh ta. Thông tin thu được thông qua sự làm mất đi sự chưa biết hay sự bất ngờ (bất định) về đối tượng
  - Lượng tin riêng của tin sẽ bằng độ bất định về tin
  - Tính toán lượng tin riêng thông qua tính toán độ bất định
- Lượng tin riêng là số đơn vị thông tin chứa trong tin, hay còn gọi là độ lớn thông tin của tin

- Độ đo độ bất định của một sự kiện đã được shannon đề xuấttừ 1948.
- Theo lý thuyết độ đo:
  - Độ bất định sẽ tỷ lệ nghịch với xác suất xuất hiện của sự kiện. Tức nó là hàm f(1/p(x)). P(x) là xác suất xuất hiện của sự kiện x
  - Để đảm bảo tính tuyến tính, độ bất định phải được đo bởi hàm log(1/(p(x))
  - Hai sự kiện x và y độc lập với nhau có xác suất xuất hiện đồng thời p(x,y) = p(x)p(y). Vậy log(1/(p(x)p(y)) = log(1/(p(x)) + log(1/(p(y))
  - 0<= p(x) <= 1 cho sự kiện rời rạc nên độ bất định đảm bảo không âm
  - Lượng tin riêng của tin x được tính bằng độ bất định và được ký hiệu là I(x)

- Cho một nguồn rời rạc có bảng chữ X ={x1,..,xk,..,xm}, với xác suất xuất hiện của mỗi ký tự của nguồn là p(X=xk) =pk, lượng tin riêng của tin xk sẽ là: I(xk) = log(1/p(xk)) = - log(p(xk))
  - Nếu log có cơ số là 2, đơn vị sẽ là bit (đơn vị nhị phân)
  - Nếu log có cơ số là e, đơn vị sẽ là nat (đơn vị tự nhiên)
  - Nếu log có cơ số là 10, đơn vị này sẽ là Hartley.
- Các đơn vị trên được gọi là các đơn vị thông tin

3.1. Amount of information (Cont.)

- Ví dụ, bảng dưới cho độ bất ngờ về kết quả của một phép thử và lượng tin chưa trong kết quả đó (chú ý, sự kiện (event) là tập các kết quả của phép thử).

Event	Probability	Surprise
1 = 1	1	0 bits
kết quả sai trong câu hỏi có 4 đáp án	3/4	0.415 bits
kết quả đúng trong câu hỏi 2 đáp án	1/2	1 bit
kết quả đúng trong câu hỏi 4 đáp án	1/4	2 bits
kết quả 7 khi gieo 2 con súc sắc	6/36	2.58 bits
Thắng trong trò chơi Jackpot	≈−1/76 million	≈26 bits

- Lượng tin của bản tin (chuỗi liên tiếp các tin) sẽ là tổng lượng tin riêng của các tin nếu **các tin độc lập** hay không chứa lượng tin của nhau. Nếu các tin không độc lập lượng tin của bản tin nhỏ hơn tổng lượng tin riêng của các tin.

- Thường, trong lý thuyết thông tin, khi tính lượng tin của bản tin, các tin của bản tin sẽ được coi là độc lập với nhau.

- Trong nhiều trường hợp, chúng ta cần xác định lượng tin của bản tin, nhưng chỉ biết số tin của bản tin mà không biết bản tin. Trong trường hợp này, người ta coi lượng tin của bản tin bằng số tin của bản tin nhân với lượng tin trung bình chứa trong các tin có trong nguồn.

- Lượng tin trung bình của các tin có trong một nguồn được gọi là lượng tin của nguồn'

- Với nguồn rời rác X = {xk}, k =1..m, p(X=xk) =pk, lương tin trung bình được ký hiêu: I(X) = E{I(xk)} **=** ∑pk.I(xk) . Đơn vị tính là đơn vị thông tin/ tin.

3.2. Entropy

3.2.1. Định nghĩa

3.2.2. Entropy của nguồn nhị phân

3.2.3. Entropy đồng thời

3.2.4. Entropy có điều kiện

3.2.5. Quan hệ giưa các entropy

3.2.6. Ví dụ

3.2.7. Entropy tương hỗ: khoảng cách Kullback-Leibler

3.2.1.Định nghĩa Entropy

Entropy là độ bất định có thể định nghĩa cho mỗi tin và cho nguồn.
Đại lượng Entropy được ký hiệu là H
Entropy của mỗi tin là độ bất định của mỗi tin và nó có giá trị bằng lượng tin của tin
Entropy của tin x là H(x) = -log p(x)
Entropy của nguồn X= {x} là H(X) = - ∑p(x) log p(x). Thường thì lý thuyết thông tin chỉ quan tâm đến Entropy của nguồn và gọi nó là Entropy.
Đơn vị của Entrpy là đơn vị lương tin
Tính chất của Entropy :
0<= H(X) <= H(X)max
H(X)max = log||X|| với điều kiện nguồn có phân bố xác suất đều (bằng nhau)

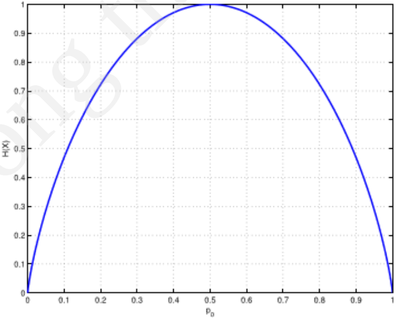
Các ví dụ

Nguôn X = (a,b); P(X) = 0.5, 0.5)
Entropy H(X) = 1 bit/ tin
Nguồn X = a,b); P(X) = (0.25, 0.75)
Entropy H(X) = 0.75 bit/ tin
Nguồn X = (a,b,c,d); P(X) = (0.25, 0.25, 0.25, 0.25)
Entropy H(X) = 2 bit/ tin
Nguồn có Entropy lớn hơn thì mỗi khi tạo ra một tin sẽ tạo ra được một lượng tin lớn hơn và tốc độ truyền tin từ nguồn này sẽ cao hơn

3.2.2. Entropy của nguồn nhị phân

- Cho một nguồn nhị phân (nguồn có 2 tin) và xác suất xuất hiện của hai tin tương ứng là p0 và p1
- Hàm Entropy của nguồn nhị phân

$$\begin{aligned} H(X) &= -p_0 \log p_0 - p_1 \log p_1 \\ &= -p_0 \log p_0 - (1-p_0)\log(1-p_0) \end{aligned}$$



3.2.3. Entropy đồng thời

(Entropy đồng thời của một cặp nguồn X = {x) và Y = {y) cho bởi:

- The joint entropy of a pair of random variables X and Y is given by:

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x,y)$$

3.2.4. Entropy có điều kiện

- Average amount of information of a random variable given the occurence of other

$$H(X|Y) = -\sum_{x \in X}\sum_{y \in Y} p(x,y)\log p(x|y)$$

$$H(Y|X) = -\sum_{x \in X}\sum_{y \in Y} p(x,y)\log p(y|x)$$

3.2.5. Quan hệ giữa các entropies

- Quan hệ giữa các Entropy xác định trên cặp nguồn X và Y là:

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

- Mở rộng:

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

3.2.5. Quan hệ giữa các entropies (Cont.)

$$H(X_1, X_2, \ldots, X_M) = \sum_{j=1}^{M} H(X_j | X_1, \ldots, X_{j-1})$$

- M: Số biến ngẫu nhiên
- Xj/x1,..,xj-1 là biến ngẫu nhiên xj xuất hiện với điều kiện các biến x1,..,xj-1 đã xuất hiện

3.2.6. Examples

- Source X,Y = with probability P(X,Y)=
- Joint entropy H(X,Y)

H(X,Y) = - P(x0,y0)logP(x0,y0) - P(x0,y1)logP(x0,y1) - P(x1,y0)logP(x1,y0) P(x1,y1)logP(x1,y1) = 4 - log0.25 - log0.25 = 2 bits/information

- Entropy H(X)

H(X) = -P(x0)logP(x0) - P(x1)logP(x1)

  - P(x0) = P(x0,y0) + P (x0,y1) = 0.25 + 0.25 = 0.5 (marginal probability)
  - P (x1) = 1 - P(x0) = 0.5

→H(X) = 2 - log0.5 - log0.5 = 1 bit/information

- Entropy H(Y)

H(Y) = - P(y0)logP(y0) - P(y1)logP(y1)

  - P(y0) = P(x0,y0) + P (x1,y0) = 0.25 + 0.25 = 0.5 (marginal probability)
  - P (y1) = 1 - P(y0) = 0.5

→H(Y) 2 - log0.5 - log0.5 = 1 bit/information

3.2.6. Examples (Cont.)

- Conditional entropy H(X|Y)

H(X|Y) = - P(x0,y0)logP(x0|y0) - P(x0,y1)logP(y1|x0)

- P(x1,y0)logP(x1|y0) - P(x1,y1)logP(x1|y1)

P(x0|y0) = = = 0.5

P(x0|y1) = = = 0.5

P(x1|y0) = = = 0.5

P(x1|y1) = = = 0.5

→ H(X|Y) = 4 x log0.25 - log0.5 = 1 bit/information

3.2.6. Examples (Cont.)

- Conditional entropy

$H(Y|X) = - P(x0,y0)\log P(y0|x0) - P(x1,y0)\log P(y0|x1)$

$- P(x0,y1)\log P(y1|x0) - P(x1,y1)\log P(y1|x1)$

$P(y0|x0) = = = 0.5$

$P(y0|x1) = = = 0.5$

$P(y1|x0) = = = 0.5$

$P(y1|x1) = = = 0.5$

→ $H(X|Y) = 4 \times \log 0.25 - \log 0.5 = 1$ bit/information

- $H(X,Y) = H(X) + H(Y|X) = 1 + 1 = 2$ bit/information

3.2.7. Entropy quan hệ: Quãng cách Kullback-Leibler

- Là độ đo quãng cách giữa hai phân bố xác suất
- Entropy tương hỗ giữa hai hàm mật độ xác suất pX(*x*) and qX(*x*) dduwwocj định nghĩa:

$$D(p_X(x)||q_X(x)) = \sum_{x \in \mathcal{X}} p_X(x) \log \frac{p_X(x)}{q_X(x)}$$

- *D*(pX(*x*)// qX(*x*)) = 0 nếu và chỉ nếu pX(*x*) = qX(*x*)
- *D*(pX(*x*)// qX(*x*)) *D*(qX(*x*)// pX(*x*))
  - D càng lớn, pX(*x*) qX(*x*) khá nhâu càng nhiều
  - pX(*x*) : phân bố xác suất thứ nhất trong miền *X*
  - qX(*x*) : phân bố xác suất có quan hệ với pX(*x*)

3.3. Lượng tin tương hỗ

- The mutual information of two random variables X and Y is defined as the relative entropy between the joint probability density $p_{XY}(x, y)$ and the product of the marginals $p_X(x)$ and $p_Y(y)$

$$
\begin{aligned}
I(X;Y) &= D(p_{XY}(x,y) \| p_X(x) p_Y(y)) \\
&= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x,y) \log \frac{p_{X,Y}(x,y)}{p_X(x) p_Y(y)}
\end{aligned}
$$

- $0 \leq I(X;Y) \leq H(X)$
- Mutual information with X as input, Y as output of a channel is the information that X transfers to Y or the information can be transmitted through the channel

3.3. Mutual information (cont.)

- Reducing uncertainty of X due to the knowledge of Y:

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

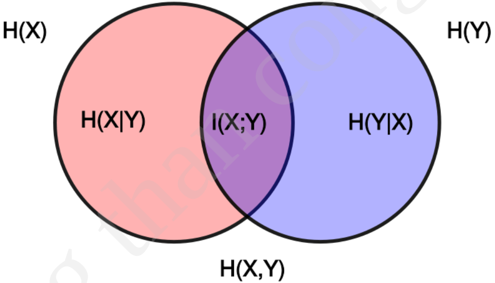
- Symmetry of the relation above: $I(X;Y) = H(Y) - H(Y|X)$
- Sum of entropies:

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

- "Self" Mutual Information:

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

3.3. Mutual information (cont.)



$$
\begin{aligned}
I(X;Y) &= H(X) - H(X \mid Y) \\
&= H(Y) - H(Y \mid X) \\
&= H(X) + H(Y) - H(X,Y)
\end{aligned}
$$

3.3. Mutual information (cont.)

Mở rộng:

- Conditional Mutual Information:

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$$

- Chain Rule for Mutual Information

$$I(X_1, X_2, \ldots, X_M; Y) = \sum_{j=1}^{M} I(X_j; Y|X_1, \ldots, X_{j-1})$$

Exercises:

1) X,Y = {xi,yj} i = 1..3; j = 1..3

P(X,Y) = {P(xi,yj)}

Calculate entropies and mutual information?

2) A source can generate only one message that have content: “information theory” is represented in form of a string without space between words, case insensitive. Each character in the message is an information. Probability of each information is calculated using ratio of the number of occurrences of information in the message divided by the total number of information in the message. Calculate entropy of the source and amount of information of the message ?

Solution

1) Joint entropy

H(X,Y) = - P(x0,y0)logP(x0,y0) - P(x0,y1)logP(x0,y1) - P(x0,y2)logP(x0,y2)

- P(x1,y0)logP(x1,y0) - P(x1,y1)logP(x1,y1) - P(x1,y2)logP(x1,y2)
- P(x2,y0)logP(x2,y0) - P(x2,y1)logP(x2,y1) - P(x2,y2)logP(x2,y2)

H(X) = -P(x0)logP(x0) - P(x1)logP(x1) - P(x2)logP(x2)

P(x0) = P(x0,y0)+ P(x0,y1) + P(x0,y2)

P(x1) = P(x1,y0)+ P(x1,y1) + P(x1,y2)

P(x2) = 1 – P(x1) – P (x2)

H(Y) = -P(y0)logP(y0) - P(y1)logP(y1) - P(y2)logP(y2)

P(y0) = P(x0,y0)+ P(x1,y0) + P(x2,y0)

P(y1) = P(x0,y1)+ P(x1,y1) + P(x2,y1)

P(y2) = 1 – P(y1) – P (y2)

Solution (cont.)

$H(X|Y) = - P(x0,y0)\log P(x0|y0) - P(x0,y1)\log P(x0|y1) - P(x0,y2)\log P(x0|y2)$

$- P(x1,y0)\log P(x1|y0) - P(x1,y1)\log P(x1|y1) - P(x1,y2)\log P(x1|y2)$

$- P(x2,y0)\log P(x2|y0) - P(x2,y1)\log P(x2|y1) - P(x2,y2)\log P(x2|y2)$

$P(x0|y0) =$ $P(x0|y2) =$

$P(x1|y0) =$ $P(x1|y2) =$

$P(x2|y0) =$ $P(x2|y2) =$

Solution (cont.)

- $H(X|Y) = - P(x0,y0)\log P(x0|y0) - P(x0,y1)\log P(x0|y1) - P(x0,y2)\log P(x0|y2)$
  $- P(x1,y0)\log P(x1|y0) - P(x1,y1)\log P(x1|y1) - P(x1,y2)\log P(x1|y2)$
  $- P(x2,y0)\log P(x2|y0) - P(x2,y1)\log P(x2|y1) - P(x2,y2)\log P(x2|y2)$

$P(x0|y0) = \frac{P(x0,y0)}{P(y0)}$ $P(x0|y1) = \frac{P(x0,y1)}{P(y1)}$ $P(x0|y2) = \frac{P(x0,y2)}{P(y2)}$

$P(x1|y0) = \frac{P(x1,y0)}{P(y0)}$ $P(x1|y1) = \frac{P(x1,y1)}{P(y1)}$ $P(x1|y2) = \frac{P(x1,y2)}{P(y2)}$

$P(x2|y0) = \frac{P(x2,y0)}{P(y0)}$ $P(x2|y1) = \frac{P(x2,y1)}{P(y1)}$ $P(x2|y2) = \frac{P(x2,y2)}{P(y2)}$

Solution (cont.)

$H(Y|X) = - P(x0,y0)\log P(y0|x0) - P(x0,y1)\log P(y1|x0) - P(x0,y2)\log P(y2|x0)$

$- P(x1,y0)\log P(y0|x1) - P(x1,y1)\log P(y1|x1) - P(x1,y2)\log P(y2|x1)$

$- P(x2,y0)\log P(y0|x2) - P(x2,y1)\log P(y1|x2) - P(x2,y2)\log P(y2|x2)$

$P(y0|x0) =$ $P(y2|x0) =$

$P(y0|x1) =$ $P(y2|x1) =$

$P(y0|x2) =$ $P(y2|x2) =$

Solution (cont.)

- H(Y|X) = - P(x0,y0)logP(y0|x0) - P(x0,y1)logP(y1|x0) - P(x0,y2)logP(y2|x0)
  - P(x1,y0)logP(y0|x1) - P(x1,y1)logP(y1|x1) - P(x1,y2)logP(y2|x1)
  - P(x2,y0)logP(y0|x2) - P(x2,y1)logP(y1|x2) - P(x2,y2)logP(y2|x2)

$P(y0|x0) = \frac{P(x0,y0)}{P(x0)}$ $\quad P(y1|x0) = \frac{P(x0,y1)}{P(x0)}$ $\quad P(y2|x0) = \frac{P(x0,y2)}{P(x0)}$

$P(y0|x1) = \frac{P(x1,y0)}{P(x1)}$ $\quad P(y1|x1) = \frac{P(x1,y1)}{P(x1)}$ $\quad P(y2|x1) = \frac{P(x1,y2)}{P(x1)}$

$P(y0|x2) = \frac{P(x2,y0)}{P(x2)}$ $\quad P(y1|x2) = \frac{P(x2,y1)}{P(x2)}$ $\quad P(y2|x2) = \frac{P(x2,y2)}{P(x2)}$

Solutions (Cont)

$$I(X;Y) = H(X) - H(X|Y)$$
$$= H(Y) - H(Y|X)$$
$$= H(X) + H(Y) - H(X,Y)$$

$$I(X;Y) = \sum_i \sum_j P(x_i, y_j) log \frac{P(x_i|y_j)}{P(x_i)}$$
$$= \sum_i \sum_j P(x_i, y_j) \log P(x_i|y_j)$$
$$- \sum_i \sum_j P(x_i, y_j) log P(x_i)$$
$$= -H(X|Y) + H(X)$$

Solution (cont.)

2) Message "informationtheory"

X = {i,n,f,o,r,m,a,t,h,e,r,y }

P(X) = {2/17, 2/17, 1/17,3/17,2/17,1/17,1/17,2/17,1/17,1/17,1/17}

H(X)= -3 x (2/17) x log_2 (2/17) – 6 x (1/17) log_2 (1/17) - (3/17) log_2 (3/17) (bit/information)

Amount of information of message is estimated by

I(message) = 17 x H(X) (bits)

- 17 is number of the information in message
- H(X) is average amount of information contained in an information