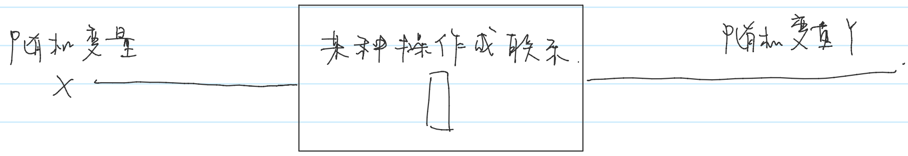
互信息

2020年3月30日 16:39

1. 定义

事物是普遍存在联系的。作为通信工程，我们也很关系当一组信号通过系统后，信息量被消除了多少。



- 单独观察 X，得到的信息量是 $H(X)$
- 已知 Y 之后，观察 X，得到的信息量是 $H(X|Y)$
- $H(X)-H(X|Y)$ 就是信息的减少量，定义为互信息

定义离散随机变量 X 与 Y 之间的互信息为 $I(X;Y)=H(X)-H(X|Y)$，那么 X 和 Y 之间的互信息理应是一种对称的量，因为这 X 和 Y 之间的联系是一定的。

也即 $I(X;Y)=H(X)-H(X|Y)=H(Y)-H(Y|X)$

我们可以简单的推导一下。

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

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

从数学的表达形式上来说，互信息也是对称的。这也间接证明信息论的数学模型的正确性。

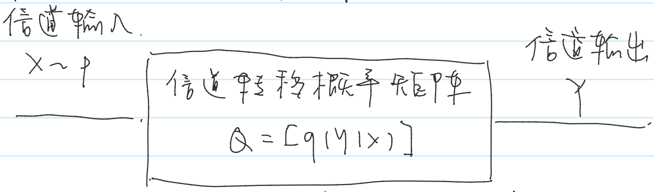
或者说 $I(X;Y)=H(X)-H(X|Y)=H(X)-[H(X,Y)-H(Y)]=H(X)+H(Y)-H(X;Y)$

即互信息是两者信息熵之和减去两者的联合信息熵；

这是结论从直觉上是比较难以理解的。$H(X)+H(Y)$ 代表了我们不知道 X 和 Y 之间关联的信息量，而 $H(X,Y)$ 是指我们知道了两者关联的信息量，两者的差别就是信息的减少量，所以进一步分析的话，在直觉上也是成立的。

我们还可以得知。
- 若 X, Y 独立，那么互信息为 0。
- 若 X, Y 一一映射，则 $I(X;Y)=H(X)$，因为 $H(X)=H(Y)$，知道了 X 的话，Y 对于我们则毫无信息量。

我们也可以从通信的角度观察互信息



$$I(X;Y)=\sum_{x\in X}\sum_{y\in Y}p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

$$=\sum_{x\in X}\sum_{y\in Y}p(x)q(y|x)\log\frac{q(y|x)}{\sum_{x\in X}p(x)\cdot q(y|x)}$$

$I(X;Y)$ 则也可以代表信道传输的能力。若误码率较低，互信息则较大，那么最大则为 $H(X)$。也就是信道是有容量的。$H(X)$ 是一个上凸函数，有最大值。

换句话说，我们在信道给定的情况下，可以通过设计 $p(x)$ 达到最大的互信息量。那么我们要考虑的就是互信息的凸优化问题

对于多变量的情形，例如。

$I(X;Y,Z)=H(X)-H(X|Y,Z)=H(Y,Z)-H(Y,Z|X)$，这是从定义出发的。

也可以展开描述

$$I(X;Y,Z)=\sum_{x\in X}\sum_{y\in Y}\sum_{z\in Z}p(x,y,z)\log\frac{p(x,y,z)}{p(x)p(y,z)}$$

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

我们还可以定义条件互信息. 例如:

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

或

$$I(X;Y|Z) = \sum_{x\in X}\sum_{y\in Y}\sum_{z\in Z} p(x,y,z)\log\frac{p(x,y|z)}{p(x|z)p(y|z)}.$$

即

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

即 X,Y 在已知 Z 与 X,Y 关联时的互信息, 减去 Z 的信息熵, 从而得到条件下的互信息.

这里稍加思考就可以直接给出. 我们对于他们的理解并不突出.
我们还可以发现条件互信息非负, 但 $I(X;Y)$ 和 $I(X;Y|Z)$ 没有确定的不等关系. 条件减少熵. 但不能减少熵的差值.

2. 性质.

首先总结一下简单的性质, 介绍定义时也已经推导或论述过.

1). 对称性.

$$I(X;Y) = I(Y;X)$$

2). 非负性

$$I(X;Y) \geq 0$$

3). 极值性.

$$I(X;Y) \leq \min(H(X), H(Y)).$$

4). 可加性

$$I(X_1, X_2, \cdots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y | X_1, \cdots X_{i-1})$$

简单证明:

理论基础一定是信息熵的可加性.

$$H(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} H(X_i | X_1 \cdots X_{i-1})$$

从定义出发.

$$\begin{aligned} I(X_1,\cdots,X_n;Y) &= H(X_1,\cdots,X_n) - H(X_1,\cdots,X_n|Y) \\ &= \sum_{i=1}^{n} H(X_i|X_1,\cdots,X_{i-1}) - \sum_{i=1}^{n} H(X_i|X_1,\cdots X_{i-1},Y) \\ &= \sum_{i=1}^{n} H(X_i|X_1,\cdots,X_{i-1}) - H(X_i|X_1,\cdots X_{i-1},Y) \\ &= \sum_{i=1}^{n} I(X_i;Y|X_1,\cdots X_{i-1}). \end{aligned}$$

通俗的语言来说, 互信息可以分步获得.

3. 互信息的凸性.

之前我们也写出了适用于通信模型的互信息数学结果, 指出了其在信源和信道的优化方向. 这里我们尝试用凸优化来解决下这个问题.

$$I(X;Y) = I(p;Q) = \sum_{x\in X}\sum_{y\in Y} p(x)q(y|x)\log\frac{q(y|x)}{\sum_{x\in X} p(x)q(y|x)}.$$

我们分两种情况讨论.

① 讨论信源: 即固定 $q(y|x)$

我们分两种情况讨论。

① 讨论信源：即固定 $q(y|x)$

$$I(X;Y)=H(Y)-H(Y|X)=H(Y)-\sum_{x\in X}p(x)H(Y|X=x)$$

第一项 $H(Y)$ 我们在讨论信息熵时已经讨论过是 $p(y)$ 的上凸函数，由于 $q(y|x)$ 给定，也是 $p(x)$ 的上凸函数。

第二项 $H(Y|X)=\sum_{x\in X}p(x)\sum_{y\in Y}q(y|x)\log\frac{1}{q(y|x)}$ 而 $q(y|x)$ 是给定的，那么 $H(Y|X)$ 是关于 $p(x)$ 的线性组合，线性并不影响凸性。

所以，我们证明了 $I(X;Y)$ 是关于 $p(x)$ 上凸。那么我们还需要找到 $I(X;Y)$ 的最大值，而这也从反面证明了是一个凸优化问题。我们以后在介绍香农第二定理时再讨论。

② 讨论信道，即固定 $p(x)$，

$$I(X;Y)=\sum_{x\in X}\sum_{y\in Y}p(x)q(y|x)\log\frac{p(x,y)}{p(x)p(y)}$$

利用对数求和不等式

$$I(X;Y)=\sum_{x\in X}p(x)\sum_{y\in Y}q(y|x)\log\frac{q(y|x)}{p(y)}\geqslant\sum_{x\in X}p(x)\log\frac{\sum_{y\in Y}q(y|x)}{1}=0.$$

所以这是一个下凸函数，我们在证明对数求和不等式时也证明了。

下凸函数则具有最小的互信息。即我们可以找到一个转移矩阵，在卡尔给定的情况下，设计一个最优的转移矩阵。那么，我们可以当转移矩阵看成是编码器，而互信息最小，而X和Y的关联度最小，即失真很大。我们可以在确定最大失真的情况下，利用下凸特性找到最合适的编码。这也就是香农第三定理，后面会详细的展开。

4. 连续随机变量的互信息

我们采用类似微分熵的推广方法求随机变量的互信息

$$I(X;Y)=\lim_{\substack{\Delta x\to0\\ \Delta y\to0}}\sum_{i=-\infty}^{+\infty}\sum_{j=-\infty}^{+\infty}p(x_i,y_j)\Delta x_i\Delta y_j\log\frac{p(x_i,y_j)\Delta x_i\Delta y_j}{p(x_i)\Delta x_i p(y_j)\Delta y_j}$$

$$=\iint p(x,y)\log\frac{p(x,y)}{p(x)p(y)}dxdy.$$

于是我们定义连续随机变量 X, Y 之间的互信息为

$$I(X;Y)=\iint p(x,y)\log\frac{p(x,y)}{p(x)p(y)}dxdy.$$

历史背景说明：

最早的信息熵由 Shannon 在 1948 年的论文中给出。

严格的对于微分熵和连续变量互信息的定义由 Kolmogorov 和 Pinsker 给出。

这也给连续的信道的信道容量提供了理论基础。