

Research on the Production of the Amount of Information and Information Entropy

Meng-yi Li, Hong-yan Zang, and Jiu-ru Dai

Schools of Mathematics and Physics, University of Science and Technology Beijing, Beijing 100083, China

Abstract—This paper presents the essentiality of the process of mining for knowledge under the demand of the cultivation of innovative talents. In traditional classroom teaching, more focus are put on the explanation of knowledge points itself. However, the first thing of the cultivation of innovative talents is to retrace the root our predecessors had taken to innovate and restore the thought course of the generation of knowledge. We analyze the background of definition about the amount of information and information entropy and the methods generated from them in the information theory course, what's more, we also study the generating process of them.

Keywords—The cultivation of innovative talents, the amount of information, information entropy, probability

关于信息量和信息熵产生过程的研究

李梦义 臧鸿雁 戴九如

北京科技大学数理学院, 北京, 中国

摘要 本文提出在创新型人才培养的需求下, 挖掘知识生产过程的重要性。传统课堂教学中, 比较注重知识点本身的讲解。而培养创新型人才, 首先要重走前人创新之路, 还原知识产生过程中的思想历程。本文分析了信息论课程中信息量和信息熵定义产生的背景和产生的方法, 研究了信息量和信息熵的产生过程。

关键词 创新型人才培养, 信息量, 信息熵, 概率

1. 引言

传统课堂教学中, 通常比较注重知识点本身的讲解, 这样有助于培养知识储备型人才。在知识创新型人才培养的需求下, 在课堂教学中, 适当地引导学生重走前人创新之路, 还原知识产生过程中的思想历程是重要的环节。只有适当地了解前人创新知识的过程, 才能具备在当前条件下进一步创新有价值的新知识的能力。

在《信息理论基础》这门课中, 事件的信息量和信息熵两个是非常基础的概念, 信息理论的迅猛发展首先是建立在这两个概念的基础上的。对事件的信息量和信息熵这两个概念的产生过程进行研究是有意义的。

信息具有不确定性, 信息的核心问题是它的度量问题, 即从量的关系上来精确地刻画信息, 要对它给出一个统一的度量是很困难的。最早在这方面取得进展的是 Hartly, 但他的结果是一类特殊的情况。真正取得重大意义结果的是 Shannon, 他对一类重要的信息即概率性语言信息进行了深入的研究, 给出了合理的度量^[1]。

本文首先对 Hartly 提出的信息度量方法的函数形式给出了证明。Shannon 在 Hartly 的基础上, 用概率论的方法描述了由于随机性引起的不确定性, 并利用给定的条件证明了 Shannon 信息熵的函数形式。最后利用单个事件的不确定性所具有的性质推出了自信息量的函数表达式形式。

2. Hartly 方法度量信息量

假设任何通信系统的发信端都有一个符号表, 发信者发出信息的过程即为按某种方式从这个符号表中选出一个特定符号序列的过程。假设符号表共有 S 个不同的符号, 发信者选定的符号序列一共包含 N 个符号 (称 N 为此符号序列的“长度”), 因此可以把发信过程看作是从 S^N 个不同的序列中选出一个特定序列的过程。

由实际情况可知, 通信设备的传递能力(即信息量)随 N 线性增长。因此, 不妨设信息量为 $H = KN$, 其中 K 是一个待定的比例系数, 与可供选择的符号数 S 有关^[2]。

现假设两个通信系统, 它们各自的符号表所包含的符

号数分别为 S_1 和 S_2 。令 K_1 和 K_2 分别为与这两个系统相对应的比例系数，则可根据下面假设来求出比例系数：不管这两个系统所进行的选择次数 N_1 和 N_2 是多少，只要它们所具有的可供选择的序列数相同，所产生的信息量就相同。也就是说，只要等式

$$S_1^{N_1} = S_2^{N_2} \quad (1)$$

成立，就应当有

$$H = K_1 N_1 = K_2 N_2 \quad (2)$$

于是，结合式 (1) 和 (2) 得：

$$\frac{K_1}{\log S_1} = \frac{K_2}{\log S_2} \quad (3)$$

则对于任何 S ，式 (3) 成立的条件必是

$$K = K_0 \log S \quad (4)$$

式中， K_0 对任意系统（任意 S ）均为相同的任意常数。

既然 K_0 为任意常数，不妨令它为 1，将 K_0 的任意性转换为对数底的任意性。这样式 (4) 就成为

$$K = \log S$$

则信息量为：

$$H = N \log S = \log S^N$$

可见，这时信息量 H 已确定与选择次数 N 成正比例关系。

3. Shannon 概率熵

Shannon 把信息定义为消除不定性的东西。既然如此，那么信息量就可以用被消除的不定性的大小来表示。而这种不定性是由随机性引起的，因此可用概率论方法来描述。这就是 Shannon 信息度量方法的基本思想^{[3][4]}。

假设随机事件 x_1, x_2, \dots, x_N ，它们概率分别为 p_1, p_2, \dots, p_N ($0 \leq p_i \leq 1, i=1, \dots, N; \sum_{i=1}^N p_i = 1$)。设用 $H_s(p_1, p_2, \dots, p_N)$ 来表示不定性的程度，为了确定 $H_s(p_1, p_2, \dots, p_N)$ 的具体形式，Shannon 提出了如下三个基本条件：

(1) H_s 应当是对 p_i ($i=1, \dots, N$) 连续的函数；

(2) 如果所有的 p_i 都相等，即 $p_i = \frac{1}{N}$ ($i=1, \dots, N$)，那么

H_s 应是 N 的单调增函数；

(3) 如果选择分为相继的两步，那么原来的 H 应等于分步选择的各个 H 值的加权和。

由以上三个条件，Shannon 推出了函数 $H_s(p_1, p_2, \dots, p_N)$

的具体形式，即：

$$H_s(p_1, p_2, \dots, p_N) = -K \sum_{i=1}^N p_i \log p_i$$

式中 K 为正常数。

证 先考虑等概率选择的情况，令

$$H_s\left(\frac{1}{N}, \dots, \frac{1}{N}\right) = A(N)$$

由条件(3)可得

$$\begin{aligned} A(MN) &= H_s\left(\frac{1}{MN}, \dots, \frac{1}{MN}\right) \\ &= H_s\left(\frac{1}{M}, \dots, \frac{1}{M}\right) + \sum_{i=1}^M \frac{1}{M} H_s\left(\frac{1}{N}, \dots, \frac{1}{N}\right) = A(M) + A(N) \end{aligned}$$

从而

$$A(N^2) = 2A(N)$$

一般的有

$$A(S^\alpha) = \alpha A(S) \quad (5)$$

或

$$A(t^\beta) = \beta A(t)$$

对于给定的 β ，总能找到恰当的 α ，使其满足如下关系：

$$S^\alpha \leq t^\beta < S^{\alpha+1} \quad (6)$$

对式(6)取对数，并除以 $\beta \log S$ ，则：

$$\frac{\alpha}{\beta} \leq \frac{\log t}{\log S} < \frac{\alpha}{\beta} + \frac{1}{\beta} \quad (7)$$

令一方面，结合条件(2)和式(6)可得

$$A(S^\alpha) \leq A(t^\beta) < A(S^{\alpha+1})$$

再由式(5)可知

$$\alpha A(S) \leq \beta A(t) < (\alpha+1)A(S) \quad (8)$$

不等式(8)除以 $\beta A(S)$ ，则得

$$\frac{\alpha}{\beta} \leq \frac{A(t)}{A(S)} < \frac{\alpha}{\beta} + \frac{1}{\beta} \quad (9)$$

由式(7)和(9)可以得出

$$\left| \frac{A(t)}{A(S)} - \frac{\log t}{\log S} \right| < \frac{1}{\beta} \quad (10)$$

若取 β 足够大, 则式(10)可以写为

$$\left| \frac{A(t)}{A(S)} - \frac{\log t}{\log S} \right| < \varepsilon$$

式中 ε 是任意小的正数。于是有

$$A(t) = K \log t$$

根据条件(2), K 为正数。

再考虑非等概率选择的情况, 事件的概率分别为 p_1, \dots, p_N , 它们都是有理数, 又设 ε' 是 p_1, \dots, p_N 的一个公因数, 则有

$$p_i = \frac{n_i}{\sum_{i=1}^N n_i} = n_i \varepsilon' \quad i = 1, \dots, N$$

其中 n_i 是正整数, $i = 1, \dots, N$ 。这样, N 个非等概率事件就可以看作是 $\sum_{i=1}^N n_i$ 个等概率事件, 其中每个事件的出现概率都等于 ε' 。利用条件(3)和前面的结果可以写出

$$\begin{aligned} H_s \left[\frac{1}{\sum_{i=1}^N n_i}, \dots, \frac{1}{\sum_{i=1}^N n_i} \right] &= H_s(p_1, p_2, \dots, p_N) + \sum_{i=1}^N p_i H_s \left[\frac{1}{n_i}, \dots, \frac{1}{n_i} \right] \\ &= H_s(p_1, p_2, \dots, p_N) + \sum_{i=1}^N p_i A \left(\frac{1}{n_i} \right) \\ &= H_s(p_1, p_2, \dots, p_N) + K \sum_{i=1}^N p_i \log n_i \end{aligned} \quad (11)$$

于是有

$$\begin{aligned} H_s(p_1, p_2, \dots, p_N) &= K \left[\log \sum_{i=1}^N n_i - \sum_{i=1}^N p_i \log n_i \right] \\ &= -K \sum_{i=1}^N p_i \log \frac{n_i}{\sum_{i=1}^N n_i} = -K \sum_{i=1}^N p_i \log p_i \end{aligned}$$

由条件(1)知 H_s 是 p_i 的连续函数, 所以当 p_i 为无理数时可近似地用有理数来代替, 式(11)仍然成立。

同样可将常数 K 是任意性转换为对数底的任意性。当对数底数为 2 时, 信息单位为二进单位, 也叫比特, 此时 $K = 1$; 当对数底数为 e 时, 则称自然单位, 也叫奈特; 当底数为 10 时, 称为迪特。

在上述公式中, 当 $p_1 = \dots = p_i = \dots = p_N = \frac{1}{S^N}$ 时, 有

$$H_s \left(\frac{1}{S^N}, \dots, \frac{1}{S^N} \right) = \log S^N, \text{ 这正是 Hartly 的结果。即: Hartly}$$

的公式是 Shannon 公式在等概率选择条件下的特例。事实上, Hartly 没有考虑概率的因素, 假定各个序列出现的概

率都相等。

4. 自信息量函数的确定

在 Shannon 熵公式中,

$$H_s(p_1, \dots, p_N) = \sum_{i=1}^N p_i (-\log p_i) = \sum_{i=1}^N p_i I(p_i)$$

式中

$$I(p_i) = -\log p_i$$

可以理解为具有出现概率 p_i 的单个事件的不定性, 或为消除这个不定性所需要的消息量。下面利用单个事件的性质来确定自信息量的函数形式^[5]。

当信源 X 发出概率为 p_i 的信息时, 用 $I(p_i)$ 来表示此单个事件的不定性, 称为自信息量。由信息的概率含义, 可知其具有以下性质:

- ① 不定性具有非负性;
- ② 以概率 1 发生的事件具有 0 的不确定性;
- ③ 若第一个事件发生的概率比第二个事件发生的概率小, 则第一个事件有比第二个事件大的不确定性;
- ④ 两个相互独立的事件同时发生的不确定性, 是单个事件发生的不确定性之和。

利用以上性质, 不妨设 $I(p_i)$ 是定义在 $(0, 1]$ 上的实值函数, 并满足下列条件:

- (1) $I(p_i) > 0, \forall p_i \in (0, 1)$;
- (2) $I(1) = 0$;
- (3) $I(p_i) > I(p_j), p_i < p_j$;
- (4) $I(p_i \cdot p_j) = I(p_i) + I(p_j), \forall p_i, p_j \in (0, 1]$;

则 $I(p_i) = c \log(p_i), \forall p_i \in (0, 1]$ 。

证 设 $0 < p_i < p_j \leq 1$, 由条件(4)知

$$I(p_i) = I(p_j \cdot \frac{p_i}{p_j}) = I(p_j) + I(\frac{p_i}{p_j})$$

因而

$$I(p_i^2) = 2I(p_i)$$

一般的有

$$I(p_i^n) = nI(p_i) \quad (n \text{ 为正整数}) \quad (12)$$

或

$$I(p_i^r) = rI(p_i) \quad (r \text{ 为正有理数})$$

$\forall p_i, p_j \in (0, 1)$, 这时区间 $(0, 1]$ 可用

$(p_j, p_j^0], (p_j^2, p_j], \dots, (p_j^{l+1}, p_j^l], \dots$ 的并来表示, 即有 $(0, 1] = \bigcup_{l=0}^{+\infty} (p_j^{l+1}, p_j^l]$, 且对自然数 k , 这时 p_i^k 必属于上述区间之一。因此, 总存在非负整数 n , 有

$$p_j^{n+1} < p_i^k \leq p_j^n \quad (13)$$

对式(13)等号两边取对数, 再除以 $-k \log p_j$, 得

$$-\frac{n+1}{k} < -\frac{\log p_i}{\log p_j} \leq -\frac{n}{k} \quad (14)$$

由条件(3)和式(13), 有

$$I(p_j^{n+1}) > I(p_i^k) \geq I(p_j^n)$$

再由式(12)可知

$$(n+1)I(p_j) > kI(p_i) \geq nI(p_j) \quad (15)$$

将式(15)等号两边同时除以 $I(p_j)$, 得

$$\frac{n}{k} \leq \frac{I(p_i)}{I(p_j)} < \frac{n+1}{k} \quad (16)$$

结合式(14)和式(16), 知

$$-\frac{1}{k} < \frac{I(p_i)}{I(p_j)} - \frac{\log p_i}{\log p_j} < \frac{1}{k}$$

当 $k \rightarrow \infty$ 时, 有 $\frac{I(p_i)}{\log p_i} = \frac{I(p_j)}{\log p_j} = c$ (c 为常数), 即有

$$I(p_i) = c \log(p_i)$$

对 $p_i \in (0, 1]$, 有 $\log p_i \leq 0$, 而 $I(p_i) \geq 0$, 所以 $c < 0$ 。

不妨设 $c = -1$, 则可将 c 的任意性转换为对数底的任意性。

$$I(p_i) = -\log(p_i)$$

5. 结论

要想培养知识创新型人才, 在课堂教学中, 适当地引导学生重走前人创新之路, 还原知识产生过程中的思想历程是重要的环节。只有适当地了解前人创新知识的过程, 才能具备在当前条件下进一步创新有价值的新知识的能力。

事件的信息量和信息熵这两个概念是信息理论中非常基础和重要的概念。信息理论得以快速发展与这两个概念的成功建立是分不开的。本文通过对事件的信息量和信息熵这两个概念的产生过程的研究还原了这两个数学概念给出时的思想历程和数学推导过程, 对知识产生过程的挖掘有助于重要的创新思想的传承和发展。

6. 致谢

本文获北京科技大学 2012 研究型教学示范课项目(KC2012YJX23) 和北京科技大学 2013 年青年教学骨干人才培养计划(06200016)资助。

参考文献(References)

- [1] Y.X. Zhong, *Informatics Rambling*, Beijing: Popular Science Press, 1984.
- [2] Y.X. Zhong, *The Principles of Information Science*, Beijing: Beijing University of Posts and Telecommunications Publishing House, 2002.
- [3] W. Li and J.M. Xi, "Research of information measure of random variable", *Systems Engineering-Theory Methodology Applications*, vol. 10, no. 3, pp. 177-201, 2001.
- [4] Y. Chen, "The historical evolution of the concept of entropy", *Physical Bulletin*, no. 4, pp. 49-52, 2009.
- [5] X.X. Yang and J. Yang, *Information Basis*, Anhui: University of Science and Technology of China Press, 2011.