

可拓熵在信息科学中的应用

常运航

PB22000086

少年班学院 3 班

2023 年 6 月 17 日

可拓熵在信息科学中的应用

常运航

1 背景

熵是热力学中常见的一个物理量，用来表征一个热力学系统的混乱程度。克劳修斯在 1854 年引入熵，并通过宏观量 (温度, 热量) 给出了熵所满足的关系式。而玻尔兹曼注意到熵和系统的无序程度的关系，并给出了熵与热力学概率的关系 $S = k \ln W$, k 为玻尔兹曼常数，上式为著名的玻尔兹曼关系式。热力学第二定律用熵的语言可以表述为系统的熵在绝热过程中永不减小。它对应着系统总是从热力学概率小的状态向热力学概率大的状态过渡。热力学概率的增大意味着无序度的增大。这便是熵的概念在热力学中的含义。

随着时代的发展，熵这一概念被赋予了越来越多的含义。1948 年，香农在《通信的数学原理》[1] 中，香农借助比特和信息熵的概念为通信系统建立了完整的数学体系。熵这一概念便不再局限于热力学中对无序度的表征，添加了描述一个系统所需的信息量的含义。

本文将从热力学中熵的概念谈起，过渡到信息熵的建立过程，了解熵这一概念在信息学中的应用。

2 热力学熵

2.1 克劳修斯熵

考虑克劳修斯不等式的积分形式

$$\oint \frac{dQ}{T} \leq 0 \quad (1)$$

等号的取值条件为此过程为可逆过程。对两点 x 和 x_0 ，很容易看出积分

$$\int_{x_0}^x \frac{dQ}{T} \quad (2)$$

沿着任何一条可逆过程的路径都相等。因此我们可以引入态函数 S ，满足

$$dS = \frac{dQ}{T} \quad (3)$$

从而我们得到

$$S - S_0 = \int_{x_0}^x \frac{dQ}{T} \quad (4)$$

这便是熵的宏观定义，它在可逆过程中不变，在不可逆过程中增大。

2.2 玻尔兹曼熵

玻尔兹曼从微观的角度来定义熵。对一个由大量微观粒子组成的系统，假设每个微观状态出现的概率是相等的，那么最有可能出现的宏观态就是包含微观态数目最多的宏观态，也就是系统最无序的状态。定义宏观状态对应的微观状态数为该宏观状态出现的热力学概率 W ，玻尔兹曼证明了熵与热力学概率之间存在如下关系

$$S = k \ln W \quad (5)$$

这就是著名的玻尔兹曼关系式，它体现了熵在热力学中最重要的含义：熵是宏观系统无序程度的表征。

2.3 麦克斯韦妖

熵是否仅仅只能表示宏观系统的无序程度呢？1867 年麦克斯韦给出了一个十分有意思的思想实验：想像两个相邻的装有气体的盒子 A 和 B，两者中间有一个很小的可以开合的门，由一个小妖怪掌控。这个小妖怪可以快速分辨气体分子运动的快慢，这样它可以让 A 中运动速度快的气体分子在接近门时打开门让其通过，运动速度慢的气体分子在接近门时关上门，让其保留在 A 中。经过一段时间之后，A 中气体的温度降低，B 中气体的温度升高，在这个过程中，气体的熵是减小的，这明显违反了热力学第二定律。物理学家尝试了许多解释的方案，最终他们意识到，分辨气体分子运动快慢这一操作，即获取信息的过程，本身就伴随着另一种熵的增加：信息熵。

3 信息熵

3.1 信息熵的定义

在讨论信息熵之前，我们先要探讨信息的含义。对一个随机事件，在它发生之前会具有不确定度，当随机事件发生时，我们获得了信息 (即随机事件的结果)，同时系统的不确定度降低。香农认为，获得信息的过程，即系统不确定度降低的过程。我们用信息熵来度量不确定度的降低，可以得到如下结论：信息的输入量等于事件不确定度的减少的大小，也就是信息熵的大小。

接下来我们借助抛硬币来具体体验信息熵的含义。如果我们现在要抛一枚硬币，在抛之前会有正面向上和反面向上两种可能，系统具有一定的不确定度。在抛完硬币之后，我们获得了信息 (如结果是正面向上)，从而信息熵降低，我们不妨以这个减少量为一个单位，记为 1bit。现在如果一个随机事件有 4 个等概率的可能情况，那么就相当于前后抛了两枚硬币，信息熵的减少量为 2bit，以此类推，如果一个随机事件有 n 中等概率的可能情况，那么就相当于抛了 $\log_2 n$ 次硬币，信息熵的减少量为 $\log_2 n$ bit。

以上讨论的情况均为等概率的情况，那么如果随机事件的各个情况发生的概率是不相等的，我们又如何计算信息熵呢？接下来我们将从数学角度考虑信息熵的定量描述。

3.2 信息熵的定量描述

我们设随机事件的各个概率为 p_1, p_2, \dots, p_n ，设信息熵为 p_1, p_2, \dots, p_n 的函数 $H(p_1, p_2, \dots, p_n)$ ，香农指出此函数要具有一下一些性质：

1. H 应该对 p_i 连续，这是为了是函数有好的性质；
2. 当 p_i 相等时， n 越大，信息熵越大，因为这代表可能的情况越多，系统的不确定度就越大；
3. 对两个相继的随机事件 A、B，两个事件复合后的信息熵应满足 $H(AB) = H(A) + H(B)$ 。

接下来我们将推出 H 的表达式。记 $H(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}) = A(n)$ ，取两种不同的概率分布 $A(s), A(t)$ 。对任意 n 我们都可以找出 m 满足

$$s^m \leq t^n \leq s^{m+1} \quad (6)$$

上式可以化为

$$\frac{m}{n} \leq \frac{\log t}{\log s} \leq \frac{m}{n} + \frac{1}{n} \quad (7)$$

这样任意 ε ，我们都可以找到 n, m 满足

$$\left| \frac{m}{n} - \frac{\log t}{\log s} \right| < \varepsilon \quad (8)$$

而由性质 2 和性质 3 我们可以得到

$$A(s^m) \leq A(t^m) \leq A(s^{m+1}) \quad (9)$$

$$mA(s) \leq nA(t) \leq (m+1)A(s) \quad (10)$$

和上式一样我们可以得到

$$\left| \frac{m}{n} - \frac{A(t)}{A(s)} \right| < \varepsilon \quad (11)$$

从而

$$\left| \frac{A(t)}{A(s)} - \frac{\log t}{\log s} \right| \leq 2\varepsilon \quad (12)$$

由上式易知 $A(t) = -K \log t, K > 0$ 。

当 p_i 不相同时，先假设 p_i 为有理数，将其表示为 $p_i = \frac{n_i}{\sum n_i}$ 。考虑先后的两个随机事件，首先一个随机事件的概率分布为 p_1, p_2, \dots, p_n ，假设结果为第 i 个，那么第二个随机事件的概率分布为等概率分布 $1/n_i$ 。由第 3 个条件，我们得到

$$K \log \sum n_i = H(p_1, p_2, \dots, p_n) + K \sum p_i \log n_i \quad (13)$$

所以

$$H = -K \sum p_i \log \frac{n_i}{\sum n_i} = -K \sum p_i \log p_i \quad (14)$$

再由条件 1 的连续性，当 p_i 不为有理数时，上式仍然能成立。当底数取为 2 时，可以令 $K = 1$ ，这样信息熵的单位就是 bit。

3.3 信息熵的应用

香农的信息熵理论奠定了信息论的基础，在其他领域也有许多实际的应用。如信息熵理论为数据压缩提供了理论指导：数据压缩的本质就是去掉信息中的冗余，保留之前不确定的信息，去除确定的 (或可推知的) 信息，保

留最本质的信息量 (不确定因素), 从而尽可能的提高信息熵。同时, 我们也可以利用信息熵来比较不同语言所蕴含的信息量, 如香农估算了英语的信息熵约为 0.6bit/字至 1.3bit/字之间 [2], 而科学家计算汉字的信息熵将近 10bit/字, 由此可见汉字所蕴含的信息量更大, 即表示同样的信息, 汉语通常比英语使用的字数少。

4 总结

信息熵的概念是热力学中熵的拓展。从数学表达式我们可以看出, 当满足等概率条件时, 信息熵的表达式就过渡到热力学熵。而如果我们的观测精度足够大时, 微观系统的状态数就表示系统所包含的信息, 这时热力学熵就过渡到信息熵。信息熵的发现极大地拓宽了人们对熵这个抽象概念的理解, 相信随着科学知识的积累, 会有更多的领域出现熵的应用, 让人们更接近熵的本质。

参考文献

- [1] C.E.Shannon. A mathematical theory of communication. *The Bell System Technical*, 27(3), 1948.
- [2] C.E.Shannon. Prediction and entropy of printed english. *The Bell System Technical*, 30(1), 1951.