

* 熵与信息简介

一. 信息、信息价值评估

1. 信息

早年：信息 — 消息

现代：信息 — • 人类所有的文化知识
• 五官所感受的一切

2. 信息价值评估

- 量的不同
- 质的差别

例如，有名的短诗与无味的小说的比较。

对信息价值在质的方面的评估有困难：

- 超出自然科学的范围
- 尚无为大家接受的客观准则

目前采用电报局的办法：

只计字数 不问内容

当代“信息论”的出发点：

只在信息量的问题上下功夫

二. 信息量与信息熵

1. 信息量

信息需要载体（语言文字、音符、图表...），
比较不同载体传递的信息量很困难。

1948年信息论的创始人 香农（Shannon）从
概率的角度给出信息的定义：

（1）信息的获得意味着在各种可能性中
概率分布的集中。

例如，在不同信息下，要去某楼找某人：

- 若只知某人住某楼（50间房），则在每间房找到他的概率为多少？
- 若又知某人住三楼（10间房），则在三楼的每间房找到他的概率为多少？
在其它楼层找到他的概率为多少？
- 若又知某人住 301室，则在301找到他的概率为多少？在其它房找到他的概率为多少？

缺乏信息（情况不明） — 概率分布分散
获得了信息（情况进一步明了）

— 概率分布集中

（2）信息量的度量

- 规定从两种可能性中作出判断所需信息量为1比特（bit）。例如要在“是和否”、“黑和白”、“有和无”、“二进制的0和1”中作出判断。在没有信息的情况下，以上每种可能性出现的概率均为1/2，要作出判断需1bit信息量。

- 从4种可能性中作出判断所需信息量为2bit。例如甲持一张扑克牌让乙猜是什么花色的？对乙的提问甲只能回答“是”和“否”那么乙提问次数最少而能猜中的问法应该如何？
错误问法：“是黑桃吗？”（为何不能这样问？）
正确问法：“是黑的吗？”“是桃吗？”
所以，从4种可能性中作出判断所需要的信息量为2 bit。

- 从 8 种可能性中作出判断所需信息量为 3 bit。
- 从 16 种可能性中作出判断所需信息量为 4 bit。
- 从 N 种可能性中作出判断所需信息量为：

$$n = \log_2 N$$

或

$$n = K \ln N$$

$$K = 1/\ln 2 = 1.4427$$

作出判断所需 bit 数越大，所缺信息量越多。

7

2. 信息熵

(1) 在对 N 种可能性完全无知的情况下，只能假定每种可能性出现的概率 P 都为 $1/N$ (概率都相等)，即 $P = 1/N$ ， $\ln P = -\ln N$ 这时作出完全的判断所需要的比特数为：

$$n = K \ln N = -K \ln P$$

记作

$$S = -K \ln P$$

香农称此 S 为信息熵，它意味着信息量的缺损。

8

(2) 各可能性概率不等情况信息熵定义为：

$$S = -K \sum_{i=1}^N P_i \ln P_i > 0$$

(若 $P_i = 1/N$ ，则过渡到各可能性等概率的情况)

例如天气预报：

- “明天有雨”，这给了 1 bit 的信息
- “明天有 80% 概率下雨”，有两种可能， $i = 1, 2$ ：

$$P_1 = 0.8 \text{ (有雨)}; P_2 = 0.2 \text{ (无雨)}$$

$$\text{信息熵 } S = -K(P_1 \ln P_1 + P_2 \ln P_2) = 0.722$$

这比全部确定所需信息 (1bit) 少 0.272 bit。

该天气预报所含信息量：

$$I = 1 - S = 0.278 \text{ bit}$$

- “明天有 90% 概率下雨”，可算出：

$$S = 0.469, \quad I = 1 - S = 0.531 \text{ bit}$$

所以信息熵 S 的减少意味着信息量 I 的增加。

在一个过程中，信息量的增量 $\Delta I = -\Delta S$

∴ 信息可转化为负熵 —— 信息的负熵原理

10

3. 信息熵公式和玻尔兹曼熵公式

信息熵 $S = -K \ln P > 0$ ($K = 1/\ln 2 = 1.4427$)

或 $S = K \ln N$ 信息熵单位：bit

玻尔兹熵 $S = k \ln \Omega$ ($k = 1.38 \times 10^{-23} \text{ J/K}$)

玻尔兹熵单位：J/K

两者相比： $K = k \rightarrow 1 \text{ bit} = k \ln 2 \text{ (J/K)}$

11

“1 bit = $k \ln 2$ (J/K)” 的物理意义：

要使计算机里信息量存储增加 1bit，则它的熵必定减少 $k \ln 2 = 0.957 \times 10^{-23} \text{ (J/K)}$ ，这至少要消耗 $kT \ln 2 = 0.693 kT \text{ (J)}$ 的能量。

例如 $T = 300 \text{ K}$ ，则信息量存储增加 1bit，要消耗能量为 $2.87 \times 10^{-21} \text{ J}$ (远小于热损耗)。

参考书：



—完—

《新概念物理教程》热学 赵凯华，罗蔚茵

12

