

可计算性理论

杨睿之

复旦大学哲学学院

2021 年春季

随机性概念的对象

定理 (无穷猴子定理 (Borel, 1913))

一只猴子随机击打键盘, 在无穷时间内一定会打出任何一部著作, 例如莎士比亚全集。

随机性概念的对象: 无穷 01 序列

随机性概念的刻画

三类刻画

- 基于不可压缩性的刻画
- 基于统计学测试的刻画
- 基于不可预测性的刻画

柯尔莫哥洛夫复杂度

接下来, 我们试图谈论一个有穷 01 字符串是否是“随机的”

柯尔莫哥洛夫复杂度

例 (压缩与解压缩)

一个非常大, 却可以用很少的空间描述的数

$$1 \underbrace{0 \cdots 0}_{10^{1000} \text{个}} 1$$

我们称以 01 串为输入和 (可能的) 输出的图灵机 (或部分递归函数) $M : 2^{<\omega} \rightarrow 2^{<\omega}$ 是一个解压缩程序。对任意 01 串 σ 和 τ , 如果 $M(\sigma) = \tau$, 称 σ 是 τ 的一个 M 描述, 即 M 将 σ 还原为 τ 。

柯尔莫哥洛夫复杂度

例 (压缩与解压缩)

一个非常大, 却可以用很少的空间描述的数

$$1 \underbrace{0 \cdots 0}_{10^{1000} \text{个}} 1$$

我们称以 01 串为输入和 (可能的) 输出的图灵机 (或部分递归函数) $M : 2^{<\omega} \rightarrow 2^{<\omega}$ 是一个 **解压缩程序**。对任意 01 串 σ 和 τ , 如果 $M(\sigma) = \tau$, 称 σ 是 τ 的一个 **M 描述**, 即 **M 将 σ 还原为 τ** 。

柯尔莫哥洛夫复杂度

定义

给定解压缩程序 $M : 2^{<\omega} \rightarrow 2^{<\omega}$, 我们定义 01 串 τ 在 M 下的柯尔莫哥洛夫复杂度 (Kolmogorov complexity) 为

$$C_M(\tau) = \min\{|\sigma| : M(\sigma) = \tau\},$$

即被解压缩后能够还原为 τ 的最短的 01 串 σ 的长度。

柯尔莫哥洛夫复杂度

注意:

- 如果 $\tau \notin \text{ran } M$, 那么 $C_M(\tau) = \infty$
- 柯尔莫哥洛夫复杂度 C_M 受 M 的选取影响过大

柯尔莫哥洛夫复杂度

定义

通用解压缩程序 我们称 $U : 2^{<\omega} \rightarrow 2^{<\omega}$ 是 **通用程序**，当且仅当对任意程序 $M : 2^{<\omega} \rightarrow 2^{<\omega}$ 都存在固定的 01 串 ρ_M ，使得对任意 01 串 σ ，都有

$$U(\rho_M \sigma) = M(\sigma)$$

此时，我们称 ρ_M 是 M 的 **编码串**， $|\rho_M|$ 是 M 在 U 中的 **编码常量**

柯尔莫哥洛夫复杂度

事实

每个通用程序 U 都是 **最优的**。即对任意可能的程序 $M : 2^{<\omega} \rightarrow 2^{<\omega}$ 都存在其编码常量 c_M , 使得

$$\forall \tau \forall \sigma [M(\sigma) = \tau \rightarrow \exists \theta (U(\theta) = \tau \wedge |\theta| \leq |\sigma| + c_M)].$$

即存在一个常量 c_M , 使得对任意字符串 τ , 都有

$$C_U(\tau) \leq C_M(\tau) + c_M.$$

柯尔莫哥洛夫复杂度

事实

存在通用程序

证明.

令 $\{\Phi_e\}_{e \in \mathbb{N}}$ 是对所有程序的可计算枚举。可以编写程序 U 使得, 对任意 $e \in \mathbb{N}$, 任意 $\sigma \in 2^{<\omega}$,

$$U(\underbrace{0 \cdots 0}_e 1 \sigma) = \Phi_e(\sigma).$$

柯尔莫哥洛夫复杂度

接下来，我们固定一个通用程序

定义

对任意 01 串 τ ，定义 τ 的 **柯尔莫哥洛夫复杂度** 为

$$C(\tau) = C_U(\tau)。$$

注意：考虑到存在等同程序 $M_{id}(\sigma) = \sigma$ ， C 在所有 01 串上都有有穷值

柯尔莫哥洛夫复杂度

事实

存在常量 c_1, c_2, c_3 , 使得对任意字符串 τ , 都有

- $C(\tau) \leq |\tau| + c_1$;
- $C(\tau\tau) \leq C(\tau) + c_2$;
- $C(h(\tau)) \leq C(\tau) + c_3$, 其中 $h : 2^{<\omega} \rightarrow 2^{<\omega}$ 是一个部分可计算函数。

柯尔莫哥洛夫复杂度

$C(\tau) \leq |\tau| + c_{id}$ (我们固定等同程序在 U 中的编码常数为 c_{id} , 根据通用程序的选取, 亦可固定为 1) 告诉我们 $C(|\tau|)$ 的一个粗略的上界, 由此我们可以给出 C 函数一个“自上而下”的可计算逼近: 对任意 $\tau \in 2^{<\omega}$ 、自然数 $s \in \mathbb{N}$, 定义

$$C_s(\tau) = \min \left(\{|\sigma| : U_s(\sigma) \downarrow = \tau\} \cup \{|\tau| + c_{id}\} \right)$$

显然, $C(\tau) \leq C_{s+1}(\tau) \leq C_s(\tau)$, 且 $C(\tau) = \lim_{s \rightarrow \infty} C_s(\tau)$ 。

柯尔莫哥洛夫复杂度

我们称 $\langle n, \tau \rangle \in \mathbb{N} \times 2^{<\omega}$ 是一个 **压缩请求**

定理 (压缩程序存在)

令 W 是一个 c.e. 的压缩请求集。并且对任意 n , 至多有 2^n 个 $\tau \in 2^{<\omega}$ 使得 $\langle n, \tau \rangle \in W$ 。那么存在一个程序 M 使得对任意 n, τ

$$\langle n, \tau \rangle \in W \leftrightarrow \exists \sigma (|\sigma| = n \wedge M(\sigma) = \tau)$$

柯尔莫哥洛夫复杂度

回忆：自然数与 01 串的对应：我们把 $\sigma \in 2^{<\omega}$ 等同于自然数 n 使得 $n + 1$ 的二进制表示是 $\sigma 1$

例

- $\text{number}(\emptyset) = 0$
- $\text{number}(\langle 0 \rangle) = 1$
- $\text{number}(1100) = (2^0 + 2^1 + 2^4) - 1 = 17$
- $\text{string}(12) = \text{string}(2^0 + 2^2 + 2^3 - 1) = 101$

柯尔莫哥洛夫复杂度

定义

$$\log n = \max \{k \in \mathbb{N} \mid 2^k \leq n\}$$

用 $\log_2 n$ 表示实数值的 \log 函数, 则

- $\log n = \lfloor \log_2 n \rfloor$
- $n/2 \leq 2^{\log n} \leq n$
- 若 $\sigma = \text{string}(n)$, 则 $|\sigma| = \log(n + 1)$

柯尔莫哥洛夫复杂度

接下来, 我们直接将自然数 n 等同于 $\text{string}(n)$ 。由此, 存在常量 c 使得

$$C(n) \leq \log n + c$$

柯尔莫哥洛夫复杂度

对任意 01 串 τ , 假设 $C(\tau) = n$, 那么, 总存在最左边的 (即 $<_L$ 下最小的) $\sigma \in 2^n$, 使得 $U(\sigma) = \tau$, 我们将其记作

$$\tau_C^*$$

直观上它是 τ “最小” 的 U 描述。

柯尔莫哥洛夫复杂度

给定一个能行可计算的 01 串有序对的编码 $\langle \sigma, \tau \rangle \mapsto \mu$, 我们用 $C(\sigma, \tau)$ 表示 $C(\langle \sigma, \tau \rangle)$

事实

存在常量 c , 对任意 01 串 τ , 有

$$C(\tau, C(\tau)) \leq C(\tau_C^*) + c$$

柯尔莫哥洛夫复杂度

尝试定义有穷字符串的某种随机程度

定义

令 $d \in \mathbb{N}$ 是一个常量。我们称 01 串 τ 是 d -C-随机的，当且仅当

$$C(\tau) \geq |\tau| - d$$

显然，有穷字符串的“是否随机”与通用程序的选取相关

柯尔莫哥洛夫复杂度

可以证明, “随机的” 字符串很多

事实

- 对任意自然数 n , 存在字符串 τ , 满足 $|\tau| = n$ 且 $C(\tau) \geq n$ 。
- 给定 $d \in \mathbb{N}$ 。对任意 $n \in \mathbb{N}$, 存在至少 $2^n - 2^{n-d} + 1$ 个长度为 n 的 d -C-随机字符串。

柯尔莫哥洛夫复杂度

例

我们称字符串 σ 是 **半随机的**，当且仅当 $C(\sigma) \geq \frac{|\sigma|}{2}$ 。根据前述事实，对任意自然数 n ，长度为 n 的半随机字符串的个数超过 $2^n(1 - 2^{-\frac{n}{2}})$ 。也就是说，当 n 越来越大的时候，几乎所有或更准确地说，占比 $(1 - 2^{-\frac{n}{2}})$ 的长度为 n 的字符串都是半随机的。

柯尔莫哥洛夫复杂度的问题

直观上我们认为字符串 $\sigma\tau$ 所含的信息量不应超过 σ 和 τ 的信息量之和，从后者似乎很容易得到前者。因此，我们希望存在常量 c ，使得对任意字符串 σ 和 τ ，都有

$$C(\sigma\tau) \leq C(\sigma) + C(\tau) + c$$

柯尔莫哥洛夫复杂度的问题

定理

对任意 $d \in \mathbb{N}$, 存在足够长的字符串 μ , 使得 $C(\mu) \geq |\mu|$, 并且对所有这样的 μ , 存在 $\sigma < \mu$, 使得 $\mu = \sigma\tau$, 且

$$C(\mu) > C(\sigma) + C(\tau) + d$$

我们先证明后面的引理

柯尔莫哥洛夫复杂度的问题

引理

存在常量 $c_M \in \mathbb{N}$, 对任意 $k \in \mathbb{N}$ 、任意足够长的字符串 μ (更准确地说, 我们要求其长度不小于 $2^{k+c_M+1} + k + c_M + 1$), 都存在 $\sigma < \mu$, 使得 $C(\sigma) < |\sigma| - k$ 。

前情回顾

- 相对柯尔莫哥洛夫复杂度
- 无前束柯尔莫哥洛夫复杂度

习题

- 2.1.3 (不能再压缩) ,2.1.4 - 2.1.7
- 2.1.17, 2.1.19

下期预告

