

Применение MDL (Minimal Description length) принципа Рissanena для полумарковских процессов.

Ремизова Анна Петровна

17 марта 2023 г.

Введение

Для начала рассмотрим простые марковские цепи. Пусть марковская цепь состоит из 2 состояний. Есть данные, мы хотим подобрать марковскую цепь, для которой наибольшая вероятность получить '001'*300. По Рissanену, если мы хотим предсказать, что будет дальше, то должны сравнивать друг с другом гипотезы по их сложности, причём даём преимущество простым гипотезам.

$$C(\mu) + \log_2 \frac{1}{\mu(x)}$$

где $C(\mu)$ - complexity, μ - распределение вероятности.

Задача 1

Дана последовательность состояний Марковской цепи из 2 состояний: 0 и 1. Найти оптимальные переходные вероятности p из 0 в 1 и q из 1 в 0 по принципу Рissanена MDL.

Для решения этой задачи запишем вероятность получения заданной реализации: пусть n_{ij} - число переходов из состояния i в состояние j , тогда:

$$P_c(x) = p^{n_{01}} \cdot (1-p)^{n_{00}} \cdot q^{n_{10}} \cdot (1-q)^{n_{11}} \rightarrow \max$$

$$\log_2 \frac{1}{P_c(x)} = -(n_{01} \log_2 p + n_{00} \log_2 (1 - p) + n_{10} \log_2 q + n_{11} \log_2 (1 - q))$$

Сложность $C(\mu)$ будем определять как суммарную длину записи p и q в двоичной системе счисления. Пусть вероятность p имеет k знаков в двоичной системе, q - l знаков, тогда $C(\mu) = k + l$. Далее рассмотрим несколько реализаций Марковских цепей и исследуем, как меняются значения в зависимости от k и l .

В таблице в каждой ячейке представлены сначала оптимальные значения $C(\mu) + \log_2 \frac{1}{\mu(x)}$, затем p и q , при которых оно достигается, округлённые до тысячных. По горизонтали отмечены значения l - длина перебираемых q в двоичной системе, по вертикали - значения k - длина перебираемых p в двоичной системе.

Таблица 1: Таблица оптимальных значений p и q для π

k / l	5	6	7	8	9	10
5	46.258	46.256	46.255	46.255	46.255	46.255
	0.583	0.583	0.583	0.583	0.583	0.583
	0.625	0.609	0.617	0.617	0.615	0.615
6	46.258	46.256	46.255	46.255	46.255	46.255
	0.583	0.583	0.583	0.583	0.583	0.583
	0.625	0.609	0.617	0.617	0.615	0.615
7	46.258	46.256	46.255	46.255	46.255	46.255
	0.583	0.583	0.583	0.583	0.583	0.583
	0.625	0.609	0.617	0.617	0.615	0.615
8	46.258	46.256	46.255	46.255	46.255	46.255
	0.583	0.583	0.583	0.583	0.583	0.583
	0.625	0.609	0.617	0.617	0.615	0.615
9	46.258	46.256	46.255	46.255	46.255	46.255
	0.583	0.583	0.583	0.583	0.583	0.583
	0.625	0.609	0.617	0.617	0.615	0.615
10	46.258	46.256	46.255	46.255	46.255	46.255
	0.583	0.583	0.583	0.583	0.583	0.583
	0.625	0.609	0.617	0.617	0.615	0.615

Содержание