

Применение принципа Риссанена MDL для марковских цепей

Ремизова Анна Петровна

08.05.2023

Постановка задачи: по последовательности 0 и 1 подобрать марковскую цепь, для которой наибольшая вероятность получить заданную траекторию. По Риссанену, для решения такой задачи необходимо сравнивать друг с другом гипотезы по их сложности, причём отдаётся преимущество простым гипотезам.

$$\mathcal{D}(\mathcal{M}, x) = C(\mathcal{M}) + \log_2 \frac{1}{P_{\mathcal{M}}(x)} \quad (1)$$

где \mathcal{M} – выбранная модель, $C(\mathcal{M})$ – сложность модели (complexity), $P_{\mathcal{M}}(x)$ – вероятность в модели \mathcal{M} получить реализацию x .

Марковские цепи с 2 состояниями

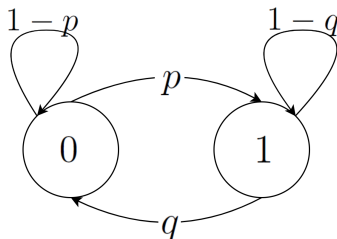


Рис.: Марковская цепь с 2 состояниями

$$P_{\mathcal{M}}(x) = p^{n(01)} \cdot (1-p)^{n(00)} \cdot q^{n(10)} \cdot (1-q)^{n(11)} \rightarrow \max \quad (2)$$

$$\mathcal{L}(\mathcal{M}, x) = \log_2 \frac{1}{P_{\mathcal{M}}(x)} = -(n(01) \cdot \log_2 p + n(00) \cdot \log_2 (1-p) + n(10) \cdot \log_2 q + n(11) \cdot \log_2 (1-q)) \quad (3)$$

Таблица: Таблица оптимальных эн-й p и q в двоичной записи для π

k / l	1	2	3	4	5	6
1	31.0	32.0	33.0	33.9891	34.9521	35.9521
	29.0	29.0	29.0	28.9891	28.9521	28.9521
	0.1	0.1	0.1	0.1	0.1	0.1
	0.1	0.10	0.100	0.1001	0.10001	0.100010
2	32.0	33.0	34.0	34.9891	35.9521	36.9521
	29.0	29.0	29.0	28.9891	28.9521	28.9521
	0.10	0.10	0.10	0.10	0.10	0.10
	0.1	0.10	0.100	0.1001	0.10001	0.100010
3	33.0	34.0	35.0	35.9891	36.9521	37.9521
	29.0	29.0	29.0	28.9891	28.9521	28.9521
	0.100	0.100	0.100	0.100	0.100	0.100
	0.1	0.10	0.100	0.1001	0.10001	0.100010
4	34.0	35.0	36.0	36.9891	37.9521	38.9521

