О ПРОГНОЗИРОВАНИИ ДВОИЧНЫХ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ МОДЕЛИ ЦМ (s, r)

В.В. Пьянов, Ю.С. Харин

Белгосуниверситет, факультет прикладной математики и информатики, Минск, Беларусь
fpm.pyanovvv@bsu.by, kharin@bsu.by

Введение. Одним из направлений статистического тестирования является проверка свойства невозможности прогнозирования выходных последовательностей криптографических генераторов статистическими методами. В данной статье представляется эффективный вычислительный алгоритм статистического прогнозирования, основанный на нахождении оптимального шаблона прогнозирования в классе малопараметрических цепей Маркова высокого порядка.

Математическая модель временного ряда. Пусть на вероятностном пространстве (Ω, F, P) наблюдается двоичный временной ряд

$$x = (x_1, \dots, x_T) \in V^\top = \{0,1\}^\top, \quad x_t \in V = \{0,1\}, \quad t = 1, \dots, T,$$

длины T, являющийся цепью Маркова s-го порядка, $s \gg 1$. Так как распределение вероятностей для x_t на практике неизвестно, а его статистическое оценивание имеет большую вычислительную сложность, то необходимы другие подходы к прогнозированию, использующие малопараметрические модели цепи Маркова высокого порядка. В качестве такой модели выберем цепь Маркова s-го порядка с r-частичными связями [1], обозначаемой ЦМ (s, r).

Алгоритм прогнозирования на основе ЦМ (s, r). Выберем $r \in \mathbb{N}$, $1 \leqslant r \leqslant s$, и зафиксируем произвольный упорядоченный набор r номеров координат шаблона $M = \{m_1, \dots, m_r\}$, где $1 \leqslant m_1 \leqslant \ldots \leqslant m_r \leqslant s$. Введем в рассмотрение условную вероятность события:

$$p(r, j_0, M) ::= P\{x_{T+1} = j_0 \mid x_{T-m_1} = j_1, \dots, x_{T-m_r} = j_r\},$$

где $j_1, \dots, j_r \in V$ — фиксированные наблюдаемые значения $x_{t-m_1}, \dots, x_{t-m_r}$. Будем строить прогноз для x_t на основе $x_{T-m_1}, \dots, x_{T-m_r}$. Оптимальная прогнозирующая статистика примет вид:

$$\hat{x}_{T+1} = \arg\max_{j_0} p(r; j_0, M). \tag{1}$$

Точность прогноза оценивается величиной [2]:

$$p_+(r; j_0, M) ::= \max_{j_0 \in V} \hat{p}(r; j_0, M).$$

Точность прогнозирования можно увеличить, максимизируя по $r \in [r_{\min}, r_{\max}]$ и шаблону:

$$p_+(r; j_0, M) \to \max_{r, M}.$$

Если решать данную задачу максимизации перебором, то вычислительная сложность будет иметь порядок $O(C_r^s T + 2^{r+1})$. Для уменьшения вычислительной сложности будем использовать алгоритм последовательной максимизации, позволяющий приближенно решать задачу максимизации по M. Пусть r изменяется в пределах

$$1 \leqslant r_{\min} \leqslant r \leqslant r_{\max} \leqslant s.$$

Решим задачу максимизации по M для $r = r_{\min}$ перебором всех возможных шаблонов длины $r_{\min}$ и получим оценку шаблона $M^{r_{\min}}$. Затем будем последовательно увеличивать r, добавляя к $M^{r_{\min}}$ ещё одну не задействованную компоненту. В качестве оценки для M^* примем $M^{r_{\max}}$. Подставляя оценку для M^* в (1), получаем оценку для будущего значения x_{T+1}. Для эффективной реализации алгоритма использовалась технология NVIDIA CUDA.

Литература

1. Харин Ю. С. *Цепи Маркова с r-частичными связями и их статистическое оценивание* // Докл. НАН Беларуси. 2004. Т. 48, № 1. С. 40–44.

2. Харин Ю.С. *Оптимальность и робастность в статистическом прогнозировании.* Мн.: БГУ, 2008. 263 с.