

ПРОГНОЗИРОВАНИЕ ДВОИЧНЫХ ВРЕМЕННЫХ РЯДОВ С ДЛИННОЙ ПАМЯТЮ НА ОСНОВЕ NVIDIA CUDA

В. В. Пьянов

ВВЕДЕНИЕ

Случайные последовательности и их генераторы являются неотъемлемыми элементами современных криптосистем[1]. Случайные последовательности используются для построения гаммы в поточных криптосистемах, сеансовых и других ключей в блочных криптосистемах. Отметим, что для криптографических приложений требуются равномерно распределенные случайные последовательности значительной длины. Поэтому возникает важная задача статистического тестирования таких последовательностей. 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 $x = (x_1, \dots, x_T) \in V^T = \{0,1\}^T, x_t \in V = \{0,1\}, t = 1, \dots, T$, длины T , являющийся цепью Маркова порядка $s, s \gg 1$, обладающий следующим гипотетическим свойством:

$$\frac{1}{2} - \varepsilon \leq \left| P\{x_t = j_0 | x_{t-1} = j_1, \dots, x_{t-s} = j_s\} - \frac{1}{2} \right| \leq \frac{1}{2}, j_0, j_1, \dots, j_s \in V, \quad (1)$$

где $\varepsilon \in [0, \frac{1}{2})$ – некоторое достаточно малое число. Это свойство означает, что при некотором достаточно большом s в выходной последовательности $\{x_t\}$ криптографического генератора существует статистически значимая зависимость глубины s . Прогнозированию подлежит следующий бит $x_{T+1} \in V$.

Так как распределение вероятностей, входящее в (1), на практике неизвестно, а его статистическое оценивание имеет вычислительную сложность порядка $O(2^{s+1})$, то необходимы другие подходы к прогнозированию, использующие малопараметрические модели цепи Маркова

высокого порядка[1–4]. Выберем в качестве такой модели цепь Маркова порядка s с r частичными связями [2].

Пусть $r \in \mathbb{N}, 1 \leq r \leq s$ – число связей, $M = \{m_1, m_2, \dots, m_r\}$ – целочисленный r -вектор с упорядоченными компонентами $1 \leq m_1 \leq m_2 \leq \dots \leq m_r \leq s$, который будем называть шаблоном, $P = (p_{j_0, \dots, j_{s-1}, j_s}), j_0, \dots, j_s \in V$ – $(s+1) \times (s+1)$ матрица вероятностей одношаговых переходов цепи Маркова x_t :

$$p_{j_0, \dots, j_{s-1}, j_s} = P\{x_t = j_0 | x_{t-1} = j_1, \dots, x_{t-s} = j_s\},$$

$Q = (q_{j_0, \dots, j_{r-1}, j_r}), j_0, \dots, j_r \in V$ – некоторая $(r+1) \times (r+1)$ стохастическая матрица.

Цепь Маркова $x_t \in V$ принято называть цепью Маркова s -го порядка с r частичными связями и обозначать ЦМ(s, r), если ее вероятности одношаговых переходов допускают следующее малопараметрическое представление:

$$p_{j_0, \dots, j_{s-1}, j_s} = q_{j_0, j_{m_1}, \dots, j_{m_r}}, \quad j_0, \dots, j_s \in V.$$

В дальнейшем будем считать, что двоичный временной ряд $x = (x_1, \dots, x_T) \in V^T = \{0, 1\}^T, x_t \in V = \{0, 1\}, t = 1, \dots, T$, является цепью Маркова s -го порядка с r частичными связями.

АЛГОРИТМ ПРОГНОЗИРОВАНИЯ НА ОСНОВЕ ЦМ(s, r) И ЕГО РЕАЛИЗАЦИЯ НА NVIDIA CUDA

Выберем $r \in \mathbb{N}, 1 \leq r \leq s$ и зафиксируем произвольный упорядоченный набор r номеров координат шаблона $M = \{m_1, m_2, \dots, m_r\}$, где $1 \leq m_1 \leq m_2 \leq \dots \leq m_r \leq s$. Введем в рассмотрение условную вероятность события:

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

где $j_1, \dots, j_r \in V$ – фиксированные наблюдаемые значения $x_{T-m_1}, \dots, x_{T-m_r}$. В силу (1) будем строить прогноз для x_t на основе $x_{T-m_1}, \dots, x_{T-m_r}$. Согласно [3], оптимальная прогнозирующая статистика для модели ЦМ(s, r) примет вид:

$$\hat{x}_{T+1} = \arg \max_{j_0} p(r, j_0, M). \quad (2)$$

Сложность вычисления прогнозирующей статистики (2) имеет порядок $O(T + 2^{r+1})$. Точность прогноза оценивается величиной [3]

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

Точность прогнозирования можно увеличить, максимизируя по $r \in [r_{min}, r_{max}]$ и шаблону [4]:

$$p_+(r; M) \rightarrow \max_{r, M}, \quad (3)$$

где $r_{min}(r_{max})$ – минимальное (максимальное) возможное число связей.

Решением экстремальной задачи (3) будет являться набор r^*, M^* наиболее информативных компонент шаблона. Если решать данную максимизацию (3) перебором, то вычислительная сложность будет иметь порядок $O(C_s^r T + 2^{r+1})$. Для уменьшения вычислительной сложности будем использовать алгоритм последовательной максимизации, позволяющий приближенно решать задачу максимизации по M . Пусть r изменяется в пределах $1 \leq r_{min} \leq r \leq r_{max} \leq s$. Решим задачу максимизации по M для $r = r_{min}$ перебором всех возможных шаблонов длины r_{min} и получим $M^{r_{min}}$. Затем будем последовательно увеличивать r , добавляя к $M^{r_{min}}$ ещё одну не задействованную компоненту:

$$M^{r_{min}} = \arg \max_M p_+(r, M),$$

$$M^r = \arg \max_{m \in \{1, \dots, s\} \setminus M^{r-1}} p_+(r, (M^{r-1}, m)).$$

И так далее увеличиваем r пока не достигнем $r = r_{max}$.

В качестве оценки M^* принимаем $M^{r_{max}}$. Подставляя M^* в (2), получаем оценку для будущего значения x_{T+1} .

Для эффективной реализации алгоритма использовалась технология NVIDIA CUDA[5], которая благодаря архитектуре SIMT позволяет эффективно производить перебор при решении экстремальных задач (3), (4).

ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Параметр s , означающий порядок цепи Маркова (глубину стохастической зависимости), в представленном алгоритме предполагался известным. Отметим, что в [4] предложен метод статистического оценивания параметров s, r на основе байесовского информационного критерия и информационного критерия Акаике.

Для тестирования разработанного алгоритма по точности, характеризуемой вероятностью ошибки:

$$p_{ош.} = P\{\hat{x}_{T+1} \neq x_{T+1}\},$$

и вычислительной сложностью, характеризуемой затратами машинного времени $t_{\text{маш.}}$ для вычисления прогноза $\hat{x}_{T+1}$, проведены две серии компьютерных экспериментов.

В первой серии обрабатывалось 10 реализаций двоичного дискретного временного ряда длины $T = 2^{15}$, порождаемого линейной рекуррентной:

$$x_t = x_{t-16} \oplus x_{t-14} \oplus x_{t-13} \oplus x_{t-11},$$

соответствующего модели ЦМ(s, r) при $s = 16, r = 4, M^* = \{11, 13, 14, 16\}$ и удовлетворяющего свойству (1) при $\varepsilon = 0$. При $r_{\min} = 3, r_{\max} = 6$ прогнозирование всех 10 реализаций произошло безошибочно: $\hat{p}_{\text{ош.}} = 0$; время обработки одной реализации составило $t_{\text{маш.}} = 1$ сек. Результаты обработки всех 10 реализаций представлены в таблице 1 (истинные компоненты шаблона в оценке $\hat{M}^{r_{\max}}$ выделены жирным шрифтом).

Во второй серии обрабатывалось 10 реализаций двоичного дискретного временного ряда длины $T = 2^{15}$, порождаемого нелинейной рекуррентой:

$$x_t = x_{t-1} \oplus x_{t-2} \oplus x_{t-8} x_{t-11} \oplus x_{t-10} x_{t-16},$$

соответствующего модели ЦМ(s, r) при $s = 17, r = 6, M^* = \{1, 2, 8, 10, 11, 16\}$ и удовлетворяющего свойству (1) при $\varepsilon = 0$. При $r_{\min} = 4, r_{\max} = 6$ прогнозирование всех 10 реализаций произошло безошибочно: $\hat{p}_{\text{ош.}} = 0$; время обработки одной реализации составило $t_{\text{маш.}} = 1.8$ сек. Результаты обработки всех 10 реализаций представлены в таблице 2.

Таблица 1

Результаты прогнозирования для первой серии

номер эксперимента	1	2	3	4	5	6	7	8	9	10
x_{T+1}	0	1	1	0	1	0	1	1	0	1
$\hat{x}_{T+1}$	0	1	1	0	1	0	1	1	0	1
$\hat{M}^{r_{\max}}$	16,14, 13,11, 10, 4	16,15, 14,13, 11, 5	16,14, 13,11, 7, 4	16,14, 13,11, 10, 8	16,14, 13,12, 11, 9	16,14, 13,11, 9, 6	16,14, 13,12, 11, 4	16,15, 14,13, 11, 8	16,14, 13,11, 9, 6	16,14, 13,11, 10, 7

Таблица 2

Результаты прогнозирования для второй серии

номер эксперимента	1	2	3	4	5	6	7	8	9	10
x_{T+1}	1	0	1	0	0	1	0	0	1	1
$\hat{x}_{T+1}$	1	0	1	0	0	1	0	0	1	1
$\bar{M}^{r_{max}}$	1, 2, 10,11, 14, 16	1, 2, 4, 9, 11, 16	1, 2, 10,11, 13, 16	1, 2, 3, 5, 8, 10	1, 2, 6, 8, 10,16	1, 2, 8,10, 11,17	1, 2, 11,14, 16, 17	1, 2, 5, 8, 10,16	1, 2, 11,12, 15, 16	1, 2, 7, 8, 10, 11

Таким образом, компьютерные эксперименты показали эффективность разработанного алгоритма прогнозирования.

Литература

1. Криптология / Ю. С. Харин [и др.]. – Минск : БГУ, 2014. – 512 с.
2. Харин, Ю. С. Цепи Маркова с g -частичными связями и их статистическое оценивание / Ю. С. Харин // Доклады НАН Беларуси. – 2004. Т. 48, № 1. – С. 40–44.
3. Ю. С. Харин Оптимальность и робастность в статистическом прогнозировании: монография / Ю.С. Харин. – Минск : БГУ, 2008. – 263 с.
4. Ю. С. Харин, А. И. Петлицкий, “Идентификация двоичной цепи Маркова s -го порядка с g частичными связями при наличии аддитивных искажений”, Дискрет. матем., 22:4 (2010), 138–155
5. CUDA Toolkit documentation. <http://docs.nvidia.com/>

ОПТИМАЛЬНОЕ ПЛАНИРОВАНИЕ ВНЕДРЕНИЯ НОВЫХ ТЕХНОЛОГИЙ В МОДЕЛИ ФИРМЫ

Е. Н. Скурат

ПОСТАНОВКА ЗАДАЧИ

В работе рассматривается обобщение задачи, предложенной в [1], где фирма внедряет одну инновацию.

На конечном промежутке времени $[0, z]$, рассмотрим фирму, которая планирует последовательно использовать N технологий, где $N > 2$ фиксировано. Моменты перехода к новой, более производительной технологии будем обозначать через $t_1, t_2, \dots, t_{N-1}$, где $0 = t_0 < t_1 < t_2 < \dots < t_{N-1} < t_N = z$. Вектор $T = (t_1, t_2, \dots, t_{N-1})$, состоящий из моментов внедрения новых технологий, необходимо определить оптимальным образом наряду с инвестиционной политикой фирмы.

Обозначим через $K = K(t), t \in [0, z]$, основной капитал фирмы. Будем полагать, что до внедрения первой технологии единичные инвестиции