

сивностью $\psi_l(n_l)$ прибор l -го узла переходит в режим $j_l - 1$, а с интенсивностью $\nu_l(n_l)$ - в режим $j_l + 1$. Время пребывания в последнем τ_l режиме имеет показательное распределение с параметром $\psi_l(i_l, \tau_l)$, после чего прибор переходит в $\tau_l - 1$ режим. Во время переключения прибора с одного режима на другой число заявок в очереди не меняется.

Состояние сети в момент времени t будет характеризоваться вектором $n(t) = (n_1(t), \dots, n_N(t))$, где $n_l(t) = (i_l(t), j_l(t))$ - состояние l -го узла в момент времени t . В соответствии с указанным выше $i_l(t)$ - число заявок в l -м узле в момент времени t , $j_l(t)$ - номер режима работы l -го узла в момент времени t .

Предположим, что все величины $\mu_l(n_l), \nu_l(n_l), \psi_l(n_l)$ строго положительны, а уравнение трафика

$$\varepsilon_j = \pi_{0j} + \sum_{i=1}^N \varepsilon_i \pi_{ij} \quad (j = \overline{1, N})$$

имеет единственное решение $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)$, для которого $\varepsilon_1 > 0, \dots, \varepsilon_N > 0$ (для этого достаточно, чтобы матрица $(\pi_{kl}, k, l = \overline{0, N}, \text{ где } \pi_{00} = 0)$ была неприводимой. Тогда $n(t)$ -неприводимый Марковский процесс на фазовом пространстве $X = n_1 \times n_2 \times \dots \times n_N$

Если во всех узлах сети выполняется условие

$$f_i(m_i, \tau_i) \nu_i(m_{i+1}, \tau_i) \psi_i(m_i, \tau_{i+1}) \mu_i(m_{i+1}, \tau_{i+1}) = f_i(m_i, \tau_{i+1}) \nu_i(m_i, \tau_i) \psi_i(m_{i+1}, \tau_{i+1}) \mu_i(m_{i+1}, \tau_i)$$

где $i = \overline{1, N}$, то $p(m, \tau) = p_1(m_1, \tau_1) p_2(m_2, \tau_2) \dots p_N(m_N, \tau_N)$ (30), где

$$p_i(m_i, \tau_i) = (\varepsilon_i \lambda)^{m_i} \prod_{k=1}^{\tau_i} \frac{\nu_i(0, k-1)}{\psi_i(0, k)} \prod_{r=1}^{m_i} \frac{f_i(r-1, j)}{\mu_i(r, j)} p(0, 0)$$

$$p_i(0, 0) = \left(\sum_{m_i=1}^{\infty} \sum_{\tau_i=0}^{l_i} ((\varepsilon_i \lambda)^{m_i} \prod_{k=1}^{\tau_i} \frac{\nu_i(0, k-1)}{\psi_i(0, k)} \prod_{r=1}^{m_i} \frac{f_i(r-1, j)}{\mu_i(r, j)} p(0, 0)) \right)^{-1}$$

Для эргодичности достаточно, чтобы

$$\sum_{m_i=1}^{\infty} \sum_{\tau_i=0}^{l_i} (\varepsilon_i \lambda)^{m_i} \prod_{k=1}^{\tau_i} \frac{\nu_i(0, k-1)}{\psi_i(0, k)} \prod_{r=1}^{m_i} \frac{f_i(r-1, j)}{\mu_i(r, j)} < \infty; \quad i = \overline{1, N}$$

ПРОВЕРКА ГИПОТЕЗЫ О ЧИСТОЙ СЛУЧАЙНОСТИ С ИСПОЛЬЗОВАНИЕМ ПРЕДИКТОРОВ: СЛУЧАЙ ЦЕПИ МАРКОВА

Костевич А. Л., Шилкин А. В.

НИИЦ ППМИ Белорусский государственный университет, г. Минск

Рассмотрим задачу проверки гипотезы H_0 о том, что бинарная последовательность описывается моделью независимых симметричных испытаний Бернулли – гипотезу о чистой случайности. Данная задача возникает в различных областях: криптографии, имитационном моделировании и др.

Построение статистического критерия для проверки гипотезы H_0 против широкого класса альтернатив в рамках традиционных подходов к тестированию бинарных последовательностей является затруднительным. Поэтому в литературе развивается альтернативный подход к проверке гипотезы H_0 на базе различных мер сложности последовательности: последовательность считается чисто случайной, если она не может быть “сжата” с помощью выбранного алгоритма ([4]).

Известно большое число универсальных алгоритмов сжатия данных, и предпринимаются попытки их использования для проверки гипотезы о чистой случайности: например, в критериях Лемпеля-Зива и Маурера [1]. Однако сложность исследования вероятностных свойств алгоритмов сжатия приводит к использованию в таких критериях статистических оценок неизвестных параметров вероятностных характеристик.

В данной статье описана процедура проверки гипотезы о чистой случайности на базе *любого* универсального предиктора. Выбор данного подхода обусловлен тесной связью между универсальными предикторами, универсальными алгоритмами сжатия данных, классическими статистическими методами, а также теоремой А. Яо об универсальности теста предсказания следующего бита. В статье исследуется эффективность рассматриваемого подхода в случае альтернативной гипотезы, описываемой цепью Маркова.

Пусть наблюдаются бинарные случайные величины X_1, X_2, \dots , описываемые набором условных вероятностей $\{P\{X_{t+1} | X_t, \dots, X_1\}\}$ из некоторого класса M , $X_t \in A = \{0,1\}$. Пусть по первым t наблюдениям X_1, \dots, X_t требуется сделать прогноз относительно значения X_{t+1} . Если вероятностная модель последовательности известна, то оптимальный прогноз X_{t+1}^* определяется максимумом условной вероятности и достигается минимальная вероятность ошибочного прогноза π_t^* :

$$X_{t+1}^* = \operatorname{argmax}_{a \in A} P\{a | X_t, \dots, X_1\} \quad (1)$$

$$\pi_t^*(X_t, \dots, X_1) = P\{X_{t+1}^* \neq X_{t+1} | X_t, \dots, X_1\} \leq 0.5$$

Если модель последовательности неизвестна, то должен быть построен предиктор $\{\widehat{P}\{X_{t+1} | X_t, \dots, X_1\}\}$, и прогноз строится аналогично (1) с большей вероятностью ошибки прогноза $\widehat{\pi}_t$:

$$\widehat{X}_{t+1} = \operatorname{argmax}_{a \in A} \widehat{P}\{a | X_t, \dots, X_1\} \quad (2)$$

$$\widehat{\pi}_t(X_t, \dots, X_1) = P\{\widehat{X}_{t+1} \neq X_{t+1} | X_t, \dots, X_1\} \geq \pi_t^*(X_t, \dots, X_1)$$

Определение ([5]). *Предиктор (2) называется универсальным для класса M , если $\widehat{\pi}_t(X_t, \dots, X_1) - \pi_t^*(X_t, \dots, X_1)$ сходится по вероятности к 0 при $t \rightarrow \infty$ для любого набора условных вероятностей из класса M .*

Отметим, что если для двух предикторов для всех X_t, \dots, X_1 верно:

$$\operatorname{argmax}_{a \in A} \widehat{P}_{(1)}\{a | X_t, \dots, X_1\} = \operatorname{argmax}_{a \in A} \widehat{P}_{(2)}\{a | X_t, \dots, X_1\},$$

то они делают одинаковые прогнозы и в этом смысле являются эквивалентными. Это ослабляет требования по асимптотической несмещенности оценок условных вероятностей.

В [3] предложена следующая схема построения критерия для проверки гипотезы о чистой случайности на базе произвольного предиктора.

Пусть дана выборка $X = \{x_t\}_{t=1}^n$ объема n . Для каждого $t = \overline{1, n-1}$ по первым t наблюдениям строится прогноз для x_{t+1} с использованием предиктора (2) и вычисляется индикатор успеха прогноза: $Y_t = I\{\widehat{X}_t = x_t\}$. Если для X верна гипотеза о чистой случайности H_0 , то последовательность индикаторов $\{Y_t\}$ также будет описываться H_0 . Если же для X верна гипотеза H_1 и $E\{\pi_t^*\} < 0.5$, то в случае использования универсального для H_1 предиктора, последовательность $\{Y_t\}$ будет иметь некоторое (за-

висящее от H_1 и предиктора) совместное распределение, но со следующими маргинальными вероятностями:

$$H_1^{(Y)} : P\{Y_t = 1\} = \frac{1}{2} + \varepsilon_t, \text{ причем } \varepsilon_t > 0 \text{ при } t \rightarrow \infty.$$

В качестве статистического критерия для проверки H_0 против уже альтернативы $H_1^{(Y)}$ (и H_1 соответственно) в [3] предложен следующий статистический критерий:

$$\text{принять } \begin{cases} H_0, & \text{если } 2\sqrt{n} (S - \frac{1}{2}) < \Phi^{-1}(1 - \alpha), \\ H_1, & \text{иначе} \end{cases} \quad S = \frac{1}{n} \sum_{t=1}^n Y_t, \quad (3)$$

где $\Phi()$ – функция распределения стандартного нормального распределения, α – уровень значимости.

В [3] показано, что при определенных условиях, в частности при выполнении для последовательности индикаторов успехов прогнозов ЦПТ, предложенный критерий будет обладать заданным уровнем значимости и являться состоятельным. Проверка необходимых условий, в общем случае, является затруднительной. Поэтому в данной статье исследуется эффективность описанного подхода в случае альтернативы H_1 , описываемой моделью стационарной цепи Маркова:

$$H_1 : P\{X_1 = j\} = \pi_j, \quad P\{X_{t+1} = j | X_t = i\} = p_{ij} \neq 0.5, \quad i, j \in A, \quad (4)$$

а в предикторе (2) используется некоторая оценка матрицы вероятностей одношаговых переходов $\hat{P} = (\hat{p}_{ij})$, т.е.

$$\hat{X}_{t+1} = f(X_t) = \operatorname{argmax}_{a \in A} \hat{p}_{X_t j} \quad (5)$$

Определение универсальности для данного предиктора будет выполняться, если для всех $i \in A$: $\operatorname{argmax}_{j \in A} \hat{p}_{ij} = \operatorname{argmax}_{j \in A} \hat{p}_{ij}$.

Для модели (4), (5) индикаторы $\{Y_t\}$ являются значениями функционала от состояний цепи Маркова и обладают следующими характеристиками:

$$E\{Y_t\} = \sum_{k,i=0}^1 k \pi_i p_{i,k \oplus f(i)}, \quad E\{Y_t^2\} = \sum_{k,i=0}^1 k^2 \pi_i p_{i,k \oplus f(i)},$$

$$\operatorname{cov}\{Y_1, Y_t\} = \sum_{u,v,i,j} uv \pi_i p_{i,u \oplus f(i)} [p_{u \oplus f(i),j}^{t-1} - \pi_j] p_{j,v \oplus f(j)}$$

Теорема. В асимптотике $t \rightarrow \infty$ при верной гипотезе H_1 и использовании универсального для H_1 предиктора (5) статистика S имеет предельное нормальное распределение с параметрами $\mu_1 = \sum_{k,i=0}^1 k \pi_i p_{i,k \oplus f(i)}$, $\sigma_1^2 = (1/n)(D\{Y_1\} + 2 \sum_{t=2}^{\infty} \operatorname{cov}\{Y_1, Y_t\}) < \infty$, а критерий (3) является состоятельным, мощность критерия имеет вид:

$$1 - \Phi\left(\Delta \frac{\sigma_0}{\sigma_1} + \frac{1}{\sigma_1} (\mu_0 - \mu_1)\right), \quad \mu_0 = \frac{1}{2}, \quad \sigma_0^2 = \frac{1}{4n}.$$

Были проведены численные эксперименты по оценке вероятности ошибки первого рода и мощности критерия (3) в случае H_1 : $p_{01} = p_{10} = 0.6$ по выборкам различного объема. На рис.1 приведены оценки вероятности ошибки I рода (\bullet) и мощности (\circ) с 95% доверительным интервалом (пунктир) предложенного критерия, а также теоретическое

значение мощности марковского предиктора (сплошная линия). Предиктор (5) строился по дополнительным выборкам. Из рис.1 можно видеть, что экспериментальные результаты согласуются с утверждением теоремы. На рис.2 для той же модели дополнительно приведены оценки мощности критериев, построенных на базе универсальных предиктора Лемпеля-Зива (◦) и Sampled Pattern Matching (SPM) предиктора (•) [2], а также теоретическое значение мощности марковского предиктора (сплошная линия). Можно заметить, что универсальные предикторы Лемпеля-Зива и SPM даже без априорной информации о виде альтернативы выявляют отклонение от H_0 , хотя и с меньшей мощностью, чем марковский предиктор, являющийся оптимальным (при наличии априорной информации о виде H_1).

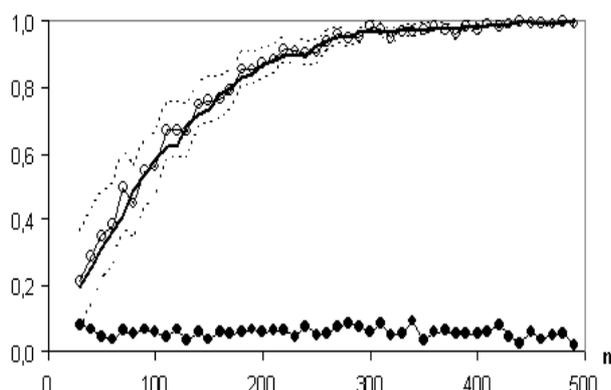


Рис.1: Оценки вероятности ошибки I рода и мощности критерия (3)

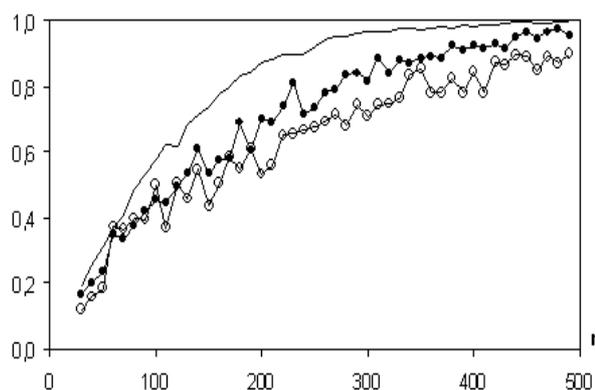


Рис.2: Оценки мощности при использовании универсальных предикторов

Литература

1. NIST Special Publication 800-22. A statistical test suite for random and pseudorandom number generators for cryptographic applications. 2001.
2. Katsaros D., Manolopoulos Ya. Suffix Tree Based Prediction Scheme for Pervasive Computing Environments // Lecture Notes on Computer Science. 2005. Vol. 3746. – P. 267-277.
3. Kostevich A.L., Shilkin A.V. Approach to Randomness Testing on the base of the Universal Predictors // Proc. of the 8 Int. Conf “Computer Data Analysis and Modeling” – Minsk 2007. – Vol. 1. – P. 256-259.
4. Ryabko B.Ya., Monarev V.A. Using information theory approach to randomness testing // Journal of Statistical Planning and Inference. – 2005. – Vol. 133 (1). – P. 95-110.
5. Suzuki J. Universal prediction and universal coding. Systems and Computers. – 2003. – Vol. 34 (6). – P. 1-11.

ЕЩЁ ОБ ОДНОМ ПОДХОДЕ К РЕШЕНИЮ ЗАДАЧИ ДУФФИНГА

Кузин В. В., Веремчук А. П.

Брестский государственный университет им. А. С. Пушкина, г. Брест

Среди методов решения задачи Дуффинга в настоящее время наиболее эффективными являются разностные методы, основным недостатком которых является необходимость восстановления полученного сеточного решения. Процедура восстановления позволяет получить приближенное решение в аналитическом виде с точностью порядка $1e-13$ – $1e-14$ для периодической задачи и с точностью $1e-8$ – $1e-9$ для непериодической задачи. Предлагаемый нами метод решения задачи Дуффинга позволяет достичь точности по невязке $1e-14$ – $1e-15$ как в периодическом, так и в непериодическом случае, что позволяет говорить об общности подхода к решению задачи Дуффинга.