НЕЙРОСЕТЕВАЯ МОДЕЛЬ МНОГОМЕРНЫХ ДВОИЧНЫХ ВРЕМЕННЫХ РЯДОВ

С. А. Шибалко

Белорусский государственный университет, пр. Независимости, 4,
220030, г. Минск, Беларусь, shibalko2003@bk.ru
Научный руководитель — Ю. С. Харин, доктор физико-математических наук,
профессор, академик НАН Беларуси

В статье рассматривается анализ многомерных двоичных временных рядов на основе малопараметрической нейросетевой модели эргодической цепи Маркова порядка *s*. Построены состоятельные статистические оценки параметров модели и алгоритмы компьютерного анализа данных с использованием нейросетевой модели: алгоритм оценивания параметров и алгоритм прогнозирования.

Ключевые слова: многомерный двоичный временной ряд; цепь Маркова порядка *s*; нейросетевая модель; статистическое оценивание параметров; статистическое прогнозирование.

Определим на вероятностном пространстве $(\Omega, \mathcal{F}, \mathrm{P})$ N-мерный ($N \in \mathbb{N}$) двоичный временной ряд $X_t = (x_{t1}, \ldots, x_{tN})' \in V^N$, порожденный семейством условных распределений вероятностей:

$$\mathrm{P}\{X_t = J_t \mid \mathcal{F}_{t-1}\} = \mathrm{P}\{X_t = J_t \mid X_{t-1} = J_{t-1}, \ldots, X_{t-s} = J_{t-s}\}, t \in \mathbb{Z}, \tag{1}$$

где $x_{tl} \in V = \{0, 1\}$ – двоичная случайная величина, задающая компоненту номер l временного ряда в момент времени $t \in \mathbb{Z}$, $\mathcal{F}_{t-1} = \sigma\{X_\tau : \tau \le t-1\}$ – σ-алгебра случайных событий, $J_t = (j_{tl}) \in V^N$ – значение двоичного случайного вектора X_t в момент времени $t \in \mathbb{Z}$, s – глубина предыстории, $s \in \mathbb{N}$.

Рассмотрим ситуацию, когда при фиксированной s-предыстории $X_{t-1} = J_{t-1}, \ldots, X_{t-s} = J_{t-s}$ случайные величины $x_{t1}, \ldots, x_{tN}$ условно независимы:

$$\begin{gathered}\mathrm{P}\{X_t = J_t \mid X_{t-1} = J_{t-1}, \ldots, X_{t-s} = J_{t-s}\} = \\ = \prod_{l=1}^{N} \mathrm{P}\{x_{tl} = j_{tl} \mid X_{t-1} = J_{t-1}, \ldots, X_{t-s} = J_{t-s}\}, J_t = (j_{tl}) \in V^N,\end{gathered} \tag{2}$$

где условное распределение l-го бита x_{tl} при условии, что фиксирована s-предыстория, представимо в виде:

$$\mathrm{P}\left\{x_{tl} = j_{tl} \mid X_{t-1} = J_{t-1}, \mathrm{K}\ , X_{t-s} = J_{t-s}\right\} = \begin{cases} p_l(J_{t-1:t-s}), & j_{tl} = 1, \\ 1 - p_l(J_{t-1:t-s}), & j_{tl} = 0. \end{cases} \qquad (3)$$

Для l-ой (l=1,…,N) компоненты аналогично [1] определим нейросетевую модель (индекс l опущен для упрощения обозначений):

$$p = p(J_{s:1}) = F(B'\boldsymbol{F}(A'J_{s:1})), J_{s:1} \in V^{Ns}, \qquad (4)$$

где $\boldsymbol{F} = \left(F_1, \ldots, F_m\right)' : \text{Ў}^m \to \text{Ў}^m$ – векторная функция, осуществляющая функциональное преобразование по-компонентно; $B \in \text{Ў}^m$ – вектор-столбец параметров модели; A=(A_1,…, A_m)∈ $\mathbb{R}^{Ns\times m}$ – матрица параметров модели.

Два набора параметров $\theta^{(1)} = (B^{(1)}, A_1^{(1)}, ..., A_m^{(1)}) \in \text{Ў}^{m(Ns+1)}$ и $\theta^{(2)} = (B^{(2)}, A_1^{(2)}, ..., A_m^{(2)}) \in \text{Ў}^{m(Ns+1)}$ являются эквивалентными для модели (1)-(4), если $F(B^{(1)\prime}\boldsymbol{F}(A^{(1)\prime}J_{s:1})) \equiv F(B^{(2)\prime}\boldsymbol{F}(A^{(2)\prime}J_{s:1})), J_{s:1} \in V^{Ns}$.

Л е м м а 1. Для любого набора параметров $\theta^{(1)} = (B, A_1, ..., A_m) \in \text{Ў}^{m(Ns+1)}$ модели (1)-(4) существует m! эквивалентных ему наборов параметров, отличающихся перестановкой пар (b1, A1),…,(b_m, A_m).

Статистическое оценивание параметров модели. Для оценивания параметров модели применим FBE-метод [2]. Пусть наблюдается реализация длины T двоичного N-мерного временного ряда $X_{1:T} = (X_1, ..., X_T) \in V^{TN}$. Примем следующие обозначения: $\mathbf{1}\{C\}$ – индикаторная функция события C, $(1; J_{s:1}) = (1, J_s' J_{s-1}' ... J_1')' \in V^{Ns+1}$ - «расширенный» двоичный вектор-столбец,

$$v_s{}^T(J_{s:1}) = \sum_{t=s}^{T} \mathbf{1}\{X_t = J_s, ..., X_{t-s+1} = J_1\},$$

$$v_{s+1}{}^T(1; J_{s:1}) = \sum_{t=s}^{T-1} \mathbf{1}\{x_{t+1,l} = 1, X_t = J_s, ..., X_{t-s+1} = J_1\},$$

где $J^{(s)} = \left\{J_{s:1} \in V^{Ns}, t = s, s+1, ..., T : v_s^T\left(J_{s:1}\right) > 0\right\} \subseteq V^{Ns}$.

Построим частотную оценку условной вероятности:

$$\hat{p}(J_{s:1}) = \begin{cases} \dfrac{T-s}{T-s+1} \cdot \dfrac{v_{s+1}^T(1; J_{s:1})}{v_s^T(J_{s:1})}, & J_{s:1} \in J^{(s)}, \\ \dfrac{1}{2}, & J_{s:1} \notin J^{(s)}; \end{cases}$$

и статистики:

$$\hat{u}(J_{s:1}) = F^{-1}(\hat{p}(J_{s:1})) \in \check{Y}, E = (e_k) = \sum_{J_{s:1} \in J^{(s)}} \hat{u}(J_{s:1}) \boldsymbol{F}'(A'J_{s:1}) \in \check{Y}^{m\times 1},$$
$$D = (d_{ij}) = \sum_{J_{s:1} \in J^{(s)}} \boldsymbol{F}(A'J_{s:1}) \boldsymbol{F}'(A'J_{s:1}) \in \check{Y}^{m\times m}, D^{-1} = (\bar{d}_{ij}) \in \check{Y}^{m\times m}. \quad (5)$$

Определим функцию потерь:

$$W(B, A) = \sum_{J_{s:1} \in J^{(s)}} (\hat{u}(J_{s:1}) - B'\boldsymbol{F}(A'J_{s:1}))^2. \quad (6)$$

Чтобы построить статистические оценки $\hat{A}, \hat{B}$ параметров A, B модели (1)-(4) необходимо решить следующую экстремальную задачу:

$$W(B, A) \to \min_{B,A}.$$

Т е о р е м а 1. Если в (5) определитель $|D| \neq 0$ и известны параметры $A_i = \left(a_{i,1}, \ldots, a_{i,ns}\right)' \in R^{Ns}, i = 1, \ldots m,$ то FBE-оценка для B имеет вид :

$$\hat{B} = \left(b_k\right) = D^{-1}E; \quad (7)$$

при $T \to \infty$ она состоятельна:

$$\hat{B} \xrightarrow{\mathrm{P}} B^0.$$

Примем дополнительные обозначения: δ_{ij} – символ Кронекера, $\otimes$ – произведение Кронекера, $I_m \in \mathbb{R}^{m\times m}$ – единичная матрица, G_f – матрица Якоби функции $\boldsymbol{F}$, $J_{s:1}$ для краткости обозначим J, Kn,m – коммутационная матрица [3]. Блочная матрица $K_{n,m} \in \mathbb{R}^{nm\times nm}$, состоящая из $n \times m$ блоков, называется коммутационной, если в (i, j) блоке элемент (j, i) равен 1, а все остальные элементы в этом блоке равны 0. ($i \in \{1,\ldots, n\}, j \in \{1,\ldots, m\}$).

Л е м м а 2. В принятых обозначениях справедливы следующие соотношения:

$$\frac{d\boldsymbol{F}(A'J)}{dA} = (I_m \otimes J)K_{1,m}G_F, \frac{d\boldsymbol{F}'(A'J)}{dA} = \frac{d\boldsymbol{F}(A'J)}{dA}K_{1,m}; \frac{dE}{dA} = \sum_{J\in J^{(s)}} \hat{u}(J)\frac{d\boldsymbol{F}(A'J)}{dA};$$

$$\frac{dD}{dA}=\sum_{J\in J^{(s)}}\frac{d\boldsymbol{F}(A'J)}{dA}(\boldsymbol{F}'(A'J)\otimes I_m)+\frac{d\boldsymbol{F}'(A'J)}{dA}(I_m\otimes \boldsymbol{F}'(A'J));$$

$$\frac{dD^{-1}}{dA}=-\frac{dD}{dA}D^{-1}\otimes D^{-1}.$$

Л е м м а 3. Частные производные первого порядка функции $W(\hat{B},A)=\sum_{J_{s:1}\in J^{(s)}}\hat{u}^2(J_{s:1})-E'D^{-1}E$ вычисляются по формуле:

$$\frac{dw}{dA}=-(\frac{dE'}{dA}+\frac{dE}{dA})\hat{B}-\frac{dD^{-1}}{dA}(E\otimes I_m)E.$$

Используя Леммы 2-3, построим итерационный алгоритм градиентного спуска для статистического оценивания параметров модели (1)-(4):

1) инициализируем начальное значение матричного параметра A: $\hat{A}^{(0)}$.
2) вычисляем оценку $\hat{B}$ (7): $\hat{B}^{(\tau-1)}=D^{-1}\left(\hat{A}^{(\tau-1)}\right)E\left(\hat{A}^{(\tau-1)}\right),\tau=1,2,\ldots;$
3) обновляем значение оценки $\hat{A}$:

$$\hat{A}^{(\tau)}=\hat{A}^{(\tau-1)}-\alpha\frac{dw\left(\hat{B}^{(\tau-1)},\hat{A}^{(\tau-1)}\right)}{dA}, \qquad (8)$$

где $\hat{B}^{(\tau)}$ - оценка параметра B на τ-ой итерации, $\hat{A}^{(\tau)}$ - оценка параметра A на τ-ой итерации, α – задаваемая величина шага итерации.

Для нахождения глобального минимума (6) необходимо осуществить итерационный процесс (8) для $L\in\mathbb{N}$ различных вариантов задания $\hat{A}^{(0)}$.

Пусть (B^0, A^0) – истинные значения параметров, (B_π^0,A_π^0) – переставочный набор параметров, где $\pi=(\pi_1,\ldots,\pi_m)\in\Pi_m$ – произвольная подстановка из множества $m!$ всевозможных подстановок на $\{1,\ldots,m\}$.

Т е о р е м а 2. Если имеет место модель (1)-(4) с истинными значениями параметров B^0, A^0, $|D|\neq 0$ и $(\hat{B},\hat{A})$ – статистические оценки параметров (B^0, A^0), то при $L\to+\infty$ и $T\to+\infty$ имеет место сходимость:

$$(\hat{B},\hat{A})\xrightarrow{\mathrm{P}}(B_\pi^0,A_\pi^0),$$

для некоторой подстановки $\pi\in\Pi_m$.

Подстановочный алгоритм оптимального прогнозирования на один шаг определяется явным соотношением [4]:

$$\hat{x}_{tl} = \mathbf{1}\left\{ \hat{p}_l(X_{t-1:t-s}) - \frac{1}{2} > 0 \right\}, l = 1, \dots, N,$$

$$\hat{p}_l\left(X_{t-1:t-s}\right) = F\left(\hat{B}'_{(l)} \boldsymbol{F}(\hat{A}'_{(l)} X_{t-1:t-s})\right),$$

где $(\hat{B}_{(l)}, \hat{A}_{(l)})$ - оценки параметров для l-го бита.

Заключение. В работе получены следующие основные результаты: для многомерных двоичных временных рядов предложена малопараметрическая нейросетевая модель; построена состоятельная статистическая оценка параметров нейросетевой модели; разработаны алгоритм оценивания параметров и подстановочный алгоритм прогнозирования многомерного двоичного временного ряда.

Библиографические ссылки

1. Харин Ю. С. Нейросетевые модели биномиальных временных рядов в задачах анализа дананых / Ю. С. Харин // Докл. Нац. акад. наук Беларуси. – 2021. – Т. 65, № 6. – С. 654– 660. https://doi.org/10.29235/1561- 8323-2021-65-6-654-660
2. Kharin Yu. Robust estimation for binomial conditionally nonlinear autoregressive time series based on multivariate conditional frequencies / Yu. Kharin, V. Voloshko // J. Multivar. Anal. 2021. Vol. 185. P. 104777. https://doi.org/10.1016/j. jmva.2021.104777
3. Kollo T., Rosen D. Advanced Multivariate Statistics and Matrices / T. Kollo, D. Rosen. – Springer, Dordrecht. 2005. 506 s.
4. Харин Ю.С. Оптимальность и робастность в статистическом прогнозировании / Ю. С. Харин - Минск: БГУ, 2008. 263 с.