

УДК 519.246.8+ 512.643.4 +004.8 (075)+ 004.422.632.4.

ТЕНЗОРНЫЕ ДЕКОМПОЗИЦИИ КАК ТЕХНОЛОГИИ DATA MINING: АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

Ю. Н. Минаев, д-р техн. наук, проф.; **Н. Н. Гузий**, канд. техн. наук, доц.
Национальный авиационный университет

О. Ю. Филимонова, канд. техн. наук, доц.; **Ю. И. Минаева**, канд. техн. наук, доц.
Киевский национальный университет строительства и архитектуры.

e-mail: min_14@ukr.net

Запропоновано метод інтелектуального аналізу числових рядів (ЧР), який реалізовано у вигляді розв'язку матричних рівнянь, що зв'язують послідовні тензор-вікна ЧР на рівні одновимірних моделей. Створено узагальнену характеристику мінімального фрагменту ЧР. Розв'язки матричних рівнянь дозволяють обчислити інваріанти (сліди, норми та ін.), матриці сингулярних величин, Фробеніусові відстані між матрицями, визначити тренди слідів, норм, відстаней, ознаки аномалії ЧР.

Ключові слова: часовий ряд, тензор, матричне рівняння, аномалії, тренд, слід, інваріанти.

In the article method of intellectual analysis TS, realised in the manner of deciding the matrix equations, link consequent tensor-windows TS, at a level rate of univariate models is considered. Incorporated generalise feature of minimum fragment. Deciding the matrix equations allow calculate invariants (traces, norms and others), matrixes of singular values, Frobenius distances between matrixes, define trands of traces, norms, distances; change of trand a sign of anomaly TS or falling out of conditions a rate.

Keywords: time row, tensor, matrix equation, anomalies, trand, trace, invariants.

Постановка проблемы

Исследование многомерных временных рядов (МВР) в последние годы стало объектом пристального внимания учёных. В системах реального времени рост объёма обрабатываемой информации не сопровождается пропорциональным ростом получаемых знаний, (мониторинг космической информации, анализ трафика в компьютерных сетях (КС), обработка полётной информации в авиации и др.) [1; 2; 3]. Создание новой технологии обработки сверхбольших объёмов данных, базирующейся на идеологии data mining (DM, извлечение знаний) на основе *априорного знания* о системе, является актуальной проблемой [4]. Возникает необходимость разработки формализованных эвристик, извлечения знаний из баз фактов (knowledge discovery), порождения гипотез с помощью интеллектуальных технологий. Интеллектуальные процедуры имитируют способности идеального типа интеллекта и реализованы в приложениях [5].

В представленной работе авторы используют *когнитивные правдоподобные рассуждения* (КПР), которые подразделяются на подклассы: вероятностные; приближённые и правдоподобные рассуждения. Класс формализованных эвристик, реализуемых в компьютерных системах посредством КПР-рассуждений, основан на парадигме *качественного анализа данных*.

Современный подход к исследованию и анализу ВР состоит в представлении его *векторами* состояний, в частности собственными значениями матриц 2D-тензоров, сформированных на окнах

ВР. Для обработки многомерной информации сверхбольших объёмов применяют тензорный подход, методы и модели тензорных декомпозиций [6; 7].

Современное состояние исследований

В классических работах [8; 9] ВР рассматривают как объект статистического моделирования. В работе [10] предложен метод анализа, основанный на преобразовании одномерного ВР в многомерный ряд с последующим применением к полученному МВР метода главных компонент. Характерная особенность метода — развёртка одномерного ряда в многомерный.

Интеллектуальный анализ данных (ИАД) ВР (или ИА темпоральных данных, Temporal Data Mining, Time series Data mining) фокусируется на автоматизированной обработке больших объёмов последовательностей данных для получения новой информации и новых знаний о взаимодействии между элементами этих последовательностей (контекстуальных данных, упорядоченных по временному критерию).

Основные понятия интеллектуального анализа темпоральных данных, общая терминология методов ИА ВР раскрыты в работах [11; 12]. Временной ряд \mathbf{X} представляет собой упорядоченную последовательность из n действительных чисел

$$\mathbf{X}(t) = \{x(t_1), x(t_2), \dots, x(t_n)\}, \quad x(t) \in \mathbb{R}^n$$

или $\mathbf{X} = \{x_1, x_2, \dots, x_n\}, \quad x_i \in \mathbb{R}^n$.

В общем случае это результаты наблюдения анализируемого процесса, следующие друг за

другом в определённые моменты времени в соответствии с заданным разделением на примеры. Последовательность ВР $Y \subseteq X$ (или *паттерн* ВР X) представляет собой ряд упорядоченных моментов времени из $X:Y = \{x_k, x_{k+1}, \dots, x_{k+m-1}\}$, $m \leq n, 1 \leq k \leq n - m + 1$.

Выявление аномалий в темпоральных данных имеет ряд особенностей: понятие аномалии является *уникальным* для различных предметных областей; при детектировании возникают ситуации, когда рационально ограничиваться фрагментами ВР различной длины; мера Евклидова расстояния в ряде случаев неприменима для анализа ВР, но реально выполнима для Фробениусовых расстояний.

Временной ряд можно представить в виде последовательности многомерных массивов — *тензоров* [13]. Полученные многомерные массивы содержат *скрытые* знания, задача состоит в извлечении этих знаний, используя и сохраняя структуру тензора. Предложенный авторами тензорный метод анализа ВР основан на формировании обобщенного показателя ВР, элементы

которого (окно, фрагмент, сегмент ВР) представлены в общем случае 3D-тензорами.

Для выявления скрытых связей и закономерностей ВР предложено использовать матричные уравнения. Указанное обстоятельство требует представления 3D-тензора в виде совокупности rank-1-тензоров.

Постановка задачи

В общем случае рассматривается многомерный (многокомпонентный) ВР, заданный совокупностью своих компонент:

$$X^{(t)} = \{^{(i)}X^{(t)}\}, i = 1, n;$$

где i — номер компоненты ВР, $^{(i)}X^{(t)} = \{^{(i)}x(t_j)\}$, $j=1, J$; j — номер сканирования; компоненты рассматриваются в n -мерном пространстве $\mathbb{R}^{x_1 \times x_2 \times \dots \times x_n}$.

Отметим, что это может быть многомерный ВР (компоненты которого ортогональны), многокомпонентный ВР, расположенный в трёхмерном пространстве (рис. 1).

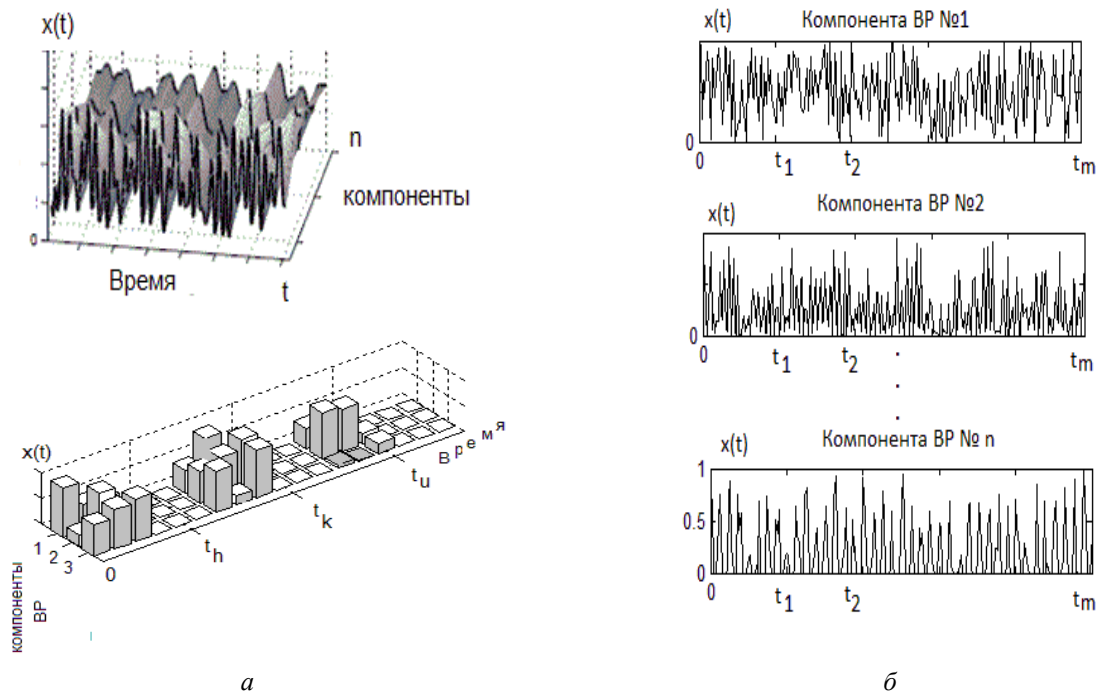


Рис. 1. Стилизованное представление многомерного ВР: a — многомерный (многокомпонентный) ВР, b — покомпонентное представление

Тензорные модели многомерных МВР рассматриваются как трёхмерные; многомерный (трёхмерный массив (тензор) или 3D-матрица) \rightarrow слайс (2D-матрица). Тензорная модель ВР принята в виде:

$$\mathfrak{X}(T) = \{\mathfrak{X}_1(T_1), \mathfrak{X}_2(T_2), \dots, \mathfrak{X}_w(T_w), \dots\},$$

где $\mathfrak{X}_i(T_i)$ — тензорная модель окна ВР длиной n элементов, при этом $n = m \cdot m$:

$$\mathfrak{X}_1 = \begin{pmatrix} x_1 & x_2 & \dots & x_m \\ x_{m+1} & x_{m+2} & \dots & x_{2m} \\ \vdots & & & \\ x_{m(m-1)+1} & x_{m(m-1)+2} & \dots & x_{m^2} \end{pmatrix}, T_1 = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix},$$

$$\mathfrak{K}_2 = \begin{pmatrix} x_1 & x_2 & \dots & x_m \\ x_{m+1} & x_{m+2} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m(m-1)+1} & x_{m(m-1)+2} & \dots & x_{m^2} \end{pmatrix}, T_2 = \begin{pmatrix} t_{n+1} \\ t_{n+2} \\ \vdots \\ t_{2n} \end{pmatrix}$$

$$\mathfrak{K}_w = \begin{pmatrix} x_1 & x_2 & \dots & x_m \\ x_{m+1} & x_{m+2} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m(m-1)+1} & x_{m(m-1)+2} & \dots & x_{m^2} \end{pmatrix}, T_w = \begin{pmatrix} t_{(w-1)n+1} \\ t_{(w-1)n+2} \\ \vdots \\ t_{wn} \end{pmatrix}$$

Индексация элементов 2D-тензора принята в границах окна.

Сложность возникающих задач заставляет искать новые методы анализа ВР [15], в т. ч. расширение PCA на многомерные ВР. Идея Data Mining with Matrix Decompositions — извлечение (новых) знаний через тензорные декомпозиции (ТД) [16].

В работе МВР имеет вид фрагмента, представленного покомпонентно, для отдельной компоненты аномалии рассматриваются как *впадина/всплеск* (рис. 2).

Метод решения основной задачи исследования

На рис. 3 представлены последовательные фрагменты исследуемого ВР-графика КС.

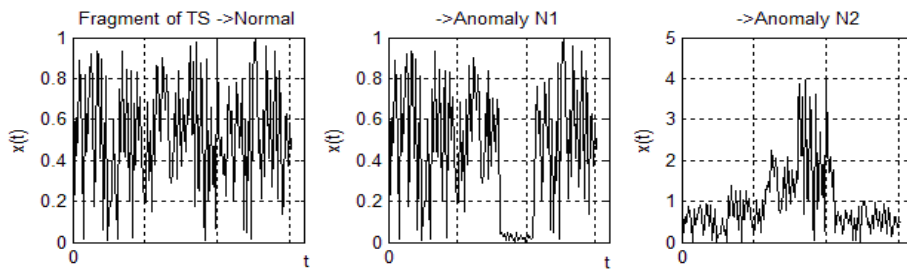


Рис. 2. Фрагменты одной компоненты ВР, содержащей аномалии

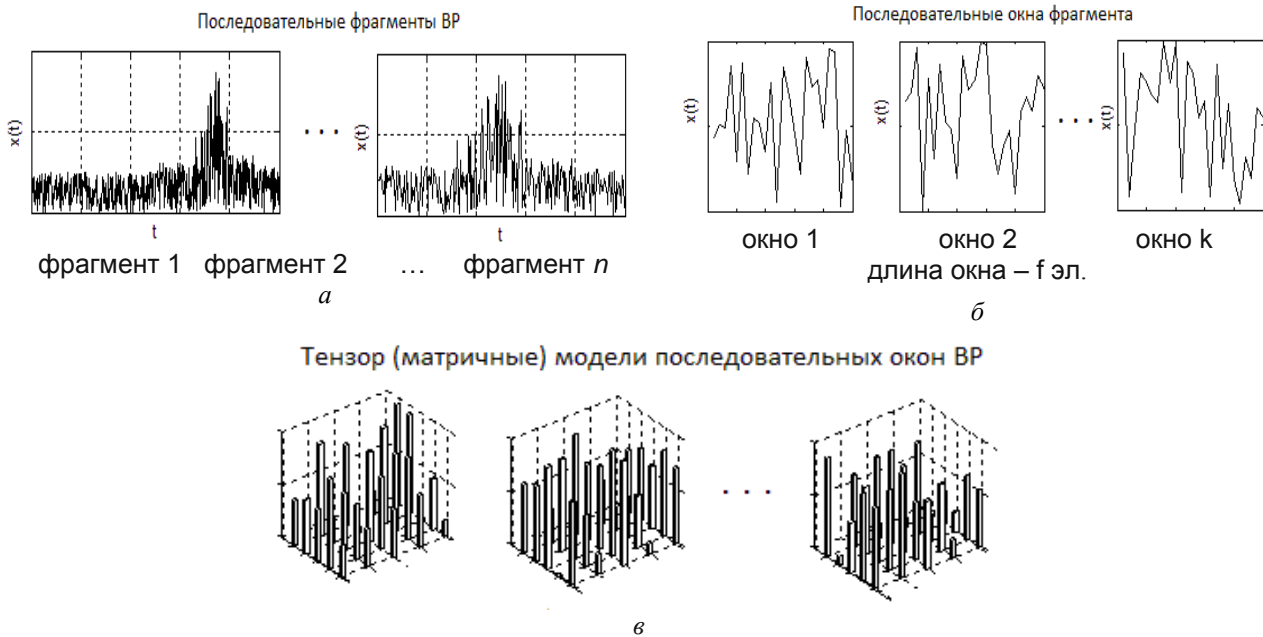


Рис. 3. Тензорные модели одномерного ВР: а — общий вид рассматриваемого ВР; б, в — представление окна (фрагмента) ВР в виде 2D- тензора с матрицей $m \times m$, $m \cdot m = f$

Задача — выявить наличие аномалий на ранних стадиях (минимальная априорная информация), например, анализируя окна t_n, t_{n+1}, t_{n+2} или фрагменты ВР; подзадача — получить максимально возможную информацию относительно аномалий, располагая информацией только о 2-х предыдущих объектах (окна, фрагмента).

Параметры нормального состояния ВР (по крайней мере, 3-х последовательных окон/фраг-

ментов) известны. Формирование тензор-окна $\mathfrak{K}(X), X \in X$

$$X = \{x(t_i)\}_{i=1,f} \rightarrow \mathfrak{K} = \text{reshape}(X, m, m), m \cdot m = f.$$

Для \mathfrak{K} выполняется сингулярная декомпозиция $[u \ s \ v] = \text{svd}(\mathfrak{K})$, где $u(:,1) \cdot s(1,1), v(:,1)^T$ — левый и правый сингулярный векторы, для которых $u(:,1) \cdot s(1,1) \otimes v(:,1)^T \cong \mathfrak{K}$, где \otimes — символ

Кронекерова (тензорного) произведения. SVD $m_1 \times m_2$ матрицы $\mathbf{B} = (b_{i_1 i_2})$ ранга r — это представление матрицы \mathbf{B} в форме $\mathbf{B} = \mathbf{U}\Sigma\mathbf{V}^T$, где \mathbf{U}, \mathbf{V} — унитарные матрицы ($\mathbf{U}\mathbf{U}^* = \mathbf{I}_{m_1}, \mathbf{V}\mathbf{V}^* = \mathbf{I}_{m_2}$), Σ — диагональная матрица размерности $m_1 \times m_2$ с неотрицательными числами на диагонали, которые известны как сингулярные величины (СВ).

Обозначим сумму квадратов СВ матрицы \mathbf{B} как $SS(\mathbf{B})$, т.е. $SS(\mathbf{B}) = \sum_{s=1}^r \sigma_s^2$, СВ матриц имеют целый ряд важных свойств, которые до нашего времени практически не использовались при анализе ВР, приведем некоторые из них [18]:

- i) $SS(\mathbf{B}) = \text{trace}(\mathbf{B}^T \mathbf{B})$;
- ii) $SS(\mathbf{B}) = \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} b_{i_1 i_2}^2$;

iii) если \mathbf{B} есть распределённым (вдоль каждой из своих размерностей) на pq подматриц с размерностями $p_i \times q_j$ $i=1, p; j=1, q$ (с $\sum_{i=1}^p p_i = m_1, \sum_{j=1}^q q_j = m_2$) тогда $\sum_{i=1}^p \sum_{j=1}^q SS(\mathbf{B}_{ij}) = SS(\mathbf{B})$;

iiii) Ф-норма определяется как $\|\mathbf{B}\| = \left(\sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} b_{i_1 i_2}^2 \right)^{1/2}$.

Пусть \mathbf{A} — трёхмерная (пространственная) матрица размерности $m_1 \times m_2 \times m_3$ с рангом $\rho = (\rho_1, \rho_2, \dots, \rho_{m_3})$ и СВ $\sigma^0 = (\sigma_1^0, \sigma_2^0, \dots, \sigma_{m_3}^0)$, где

$$\sigma_{i_3}^0 = (\sigma_{i_3 1}^0, \sigma_{i_3 2}^0, \dots, \sigma_{i_3 r_{i_3}}^0), \quad i_3 = 1, m_3.$$

Пусть $SS(\mathbf{A}) = \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} \sum_{i_3=1}^{m_3} a_{i_1 i_2 i_3}^2$. Характеристики

СВ \mathbf{A} определены следующим образом:

- i) $SS(\mathbf{A}) = \text{trace}(\mathbf{A}^T \mathbf{A})$;
- ii) $SS(\mathbf{A}) = \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} \sum_{i_3=1}^{m_3} a_{i_1 i_2 i_3}^2$;

iii) пусть матрица будет распределенной (вдоль каждого из 3-х измерений в 2-частях) так, что имеем 8 подматриц \mathbf{A}_{ijk} размерностями $p_i \times q_j \times r_k$ для $i, j, k = 1, 2$ таких, что $p_1 + p_2 = m_1, q_1 + q_2 = m_2, r_1 + r_2 = m_3$, тогда $\sum_{i,j,k=1}^2 SS(\mathbf{A}_{ij})_w = SS(\mathbf{A})$;

iiii) матричная норма даётся как $\|\mathbf{A}\| = \left(\sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} \sum_{i_3=1}^{m_3} a_{i_1 i_2 i_3}^2 \right)^{1/2}$.

Тензорная модель фрагмента ВР основана на формировании тензор-фрагментов размерностью $p \times q \times k$, k — количество окон в фрагменте, тензор-окно — фронтальный слайс 3D-тензор-фрагмента (рис. 4); 3D-матрица порядка $m_1 \times m_2 \times m_3$ — это массив из $m_1 \cdot m_2 \cdot m_3$ скаляров, собранных в m_1 строк, m_2 колонок и m_3 слоев. 3D-матрица размерностью $m_1 \times m_2 \times m_3$ таким образом состоит

из 2D-матриц размерностью $m_1 \times m_2$, собранных в массив размерностью m_3 [18].

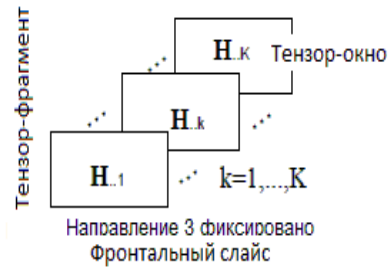


Рис.4. Формирование трёхмерной тензорной модели одномерного ВР

Матрица \mathbf{A} представлена в форме массива матриц $\mathbf{A}(:, :, i_3)$ размерностью $m_1 \times m_2$ и элементов из $\mathbf{A}(:, :, i_3)$, как показано ниже:

$$\mathbf{A}(:, :, i_3) = \begin{pmatrix} a_{11i_3} & a_{12i_3} & \dots & a_{1m_2i_3} \\ a_{21i_3} & a_{22i_3} & \dots & a_{2m_2i_3} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m_1i_3} & a_{m_12i_3} & \dots & a_{m_1m_2i_3} \end{pmatrix}, \quad i_3 = 1, m_3.$$

По результатам сканирования одномерного ВР (в частности, трафика КС) $\mathbf{X} = \{x_j(t)\}$, $j=1, J_x$, формируется минимальный фрагмент, состоящий из 3-х последовательных окон: $^{(i-1)}\mathbf{X}, ^{(i)}\mathbf{X}, ^{(i+1)}\mathbf{X}$, обозначенных индексами $i-1, i, i+1$; каждое окно состоит из отдельных элементов $\{x_j(t)\}$, $j=1, 2, \dots$; $\mathbf{X}_f = \{^{(i-1)}\mathbf{X}, ^{(i)}\mathbf{X}, ^{(i+1)}\mathbf{X}\}$, $f=1, F$ — фрагмент ВР, $\mathbf{X}_f \in \mathbf{X}$. Для каждого окна ВР сформированы матрицы тензоров $\mathbf{T}_{d_w}, \mathbf{T}_{d_f}, \mathbf{T}_{d_s}$, группы тензорных признаков: их сингулярные разложения и инварианты.

В процессе исследования необходимо выявить состояние норма или выпадение из нормы (аномалия) путем сравнения характеристик минимального фрагмента (три последовательных окна — предыдущее, настоящее и последующее).

Минимальная длина окна составляет 4 или 9 элементов, что позволяет использовать 2D-тензор с матрицей 2×2 или 3×3 соответственно, минимальное количество окон фрагмента — три.

Тензоры с матрицами разной размерности могут быть приведены к тензорам с матрицами одинаковой размерности при помощи Кронекерова произведения (КП) [16]. Отметим, что совокупность 2D-матриц-окон, представляющих фрагмент ВР, может быть представлена как 3D-тензор, 2D-матрица-фронтальный слайс (рис. 4). МВР анализируется на уровне отдельных компонент, окна/фрагменты которых преобразуются в 2D- или 3D-тензоры. МВР в общем случае представляется 3D-тензором, который при помощи PARAFAC-декомпозиции [18] преобразуется в совокупность rank-1 тензоров.

На уровне тензорных гранул (состоящих из КП векторов $A \otimes B \otimes C$), МВР анализируется на уровне методов анализа одномерных ВР.

Предлагается обобщенная характеристика минимального фрагмента (состоящего из 3-х

окон) в виде *решения* матричных уравнений (МУ): $AXB=C$ или/и $AX+XB=C$, где A, B, C — 2D-тензоры (с матрицами $m \times m$) соседних окон или фрагментов (рис. 5).

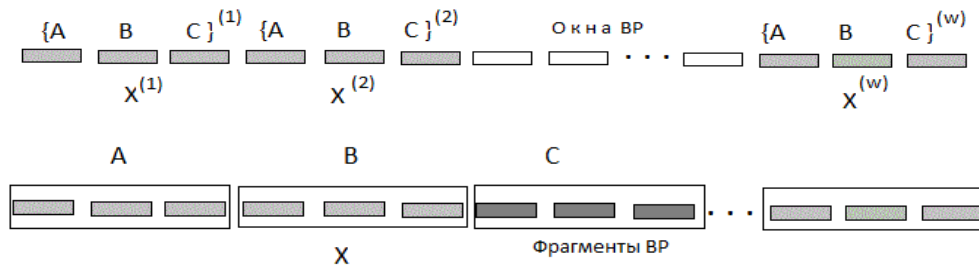


Рис. 5. Обобщенная характеристика минимального фрагмента в виде матричных уравнений: $AXB = C$ или/и $AX + XB = C$

Для схемы на рис. 5 имеем совокупность МУ: 1-й кортеж $(AXB=C)^{(1)}$, 2-й кортеж $(AXB=C)^{(2)}$, ..., w-й кортеж $(AXB=C)^{(w)}$. Наличие решений МУ в виде матриц $X^{(1)}, X^{(2)}, \dots, X^{(w)}$ для кортежей тензор-окон $\{A B C\}^{(1)}, \{A B C\}^{(2)}, \dots, \{A B C\}^{(w)}$ позволяет вычислять инварианты (следы, нормы и др.) — решений соответствующих МУ, (диагональные) матрицы сингулярных величин $S^{(i)}, i = 1, w$, Фробениусовы расстояния между матрицами $X^{(i)}$ и $S^{(i)}$ — $ds(X^{(i-1)}, X^{(i)})$ и $ds(S^{(i-1)}, S^{(i)})$ соответственно. Это, в свою очередь, даёт возможность определить тренды следов, норм, расстояний. Изменение тренда может свидетельствовать либо о возможном наличии аномалии или выпадении из состояния *норма*.

Отметим, что для МВР совокупность матриц X-решений уравнений $AXB=C$ или/и $AX+XB=C$ -кортежей из тензор-окон рассматривается при условии, что каждая матрица-решение $X^{(i)}, i=1, w$ — это фронтальный слайс — $X(:, :, 1), \dots, X(:, :, w)$, что позволяет сформировать 3D-тензор \mathfrak{X} , PARAC-декомпозиция тензора \mathfrak{X} в виде 2D матриц $\{A, \bar{B}, \bar{C}\}$ с последующим формированием упорядоченных троек $\{\bar{A}(:, 1), \bar{B}(:, 1), \bar{C}(:, 1)\}$, из которых в свою очередь может быть сформировано НМ 2-типа, представленное в виде утверждения (нечёткого правила). Тензорная модель МВТ позволяет более объективно идентифицировать аномалию.

Матрично-матричные уравнения в анализе ВР. Матричные уравнения (МУ) — уравнения, в которых неизвестные являются матрицами, широко применяются в решении разнообразных задач, в частности, в теории управления [19]. Практическое большинство МУ решаются путём их эквивалентного преобразования в векторно-матричную форму на основании Кронекерова

произведения. Например, МУ $BXA^T=C$ с известными $n \times n$ матрицами A, B, C , и неизвестной $n \times n$ матрицей X при условии, что все матрицы принимаются имеющими комплексно-значные входные, а вектор-колонка $\text{vec}(X) = X \in \mathbb{C}^{n^2}$ включает все столбцы X .

Вектор $\text{vec}(X) = X = (X_{11}, \dots, X_{n1}, X_{12}, \dots, X_{n2}, \dots, X_{1n}, \dots, X_{nn})^T$, прямой подстановкой может быть записан в матрично-векторной форме для неизвестного вектора X в виде $(A \otimes B)X = C, C = \text{vec}(C)$, где $A \otimes B$ — КП матриц A, B , представляющее блочную матрицу с размерностью $n^2 \times n^2$ и блоками $A_{ij} \cdot B$. Это означает справедливость выражения $\text{vec}(BXA^T) = (A \otimes B) \cdot \text{vec}(X)$.

Данное МУ имеет единственное решение при $BB^+ \otimes A^+A = I$, где верхний индекс $+$ означает псевдоинверсию. Решение МУ $AX + XB = C$, где $A \in \mathbb{C}^{n \times n}, B \in \mathbb{C}^{m \times m}, C, X \in \mathbb{C}^{n \times m}$ при использовании процедуры векторизации [20] имеет вид $(I_m \otimes A) \text{vec}(X) + (B^T \otimes I_n) \text{vec}(X) = \text{vec}(C)$.

Решение в представленном виде позволяет сформулировать так называемую Кронекерову сумму (\oplus), которая для квадратных матриц $A \in \mathbb{C}^{n \times n}, B \in \mathbb{C}^{m \times m}$, определяется как $A \oplus B = (I_m \otimes A) + (B \otimes I_n)$.

Отметим, что в работе рассмотрен случай, когда все размерности одинаковы. В общем случае размерности и порядки A и B могут быть другими, выбор размерности определяется стратегией решения.

Общее условие существования решения уравнения таково: уравнение $AX+XB=C$ допускает решение тогда и только тогда, когда матрицы $\begin{pmatrix} A & -C \\ 0 & B \end{pmatrix}$ и $\begin{pmatrix} A & 0 \\ 0 & -B \end{pmatrix}$ подобны; матрица преобразования подобия (сходства) имеет вид $\begin{pmatrix} I & X \\ 0 & I \end{pmatrix}$, где матрица X — решение МУ.

Тензорные характеристики ВР-трафика в нормальном состоянии и при наличии аномалий. Рассмотрим ВР, модели нормального состояния которого приведены на рис. 6.

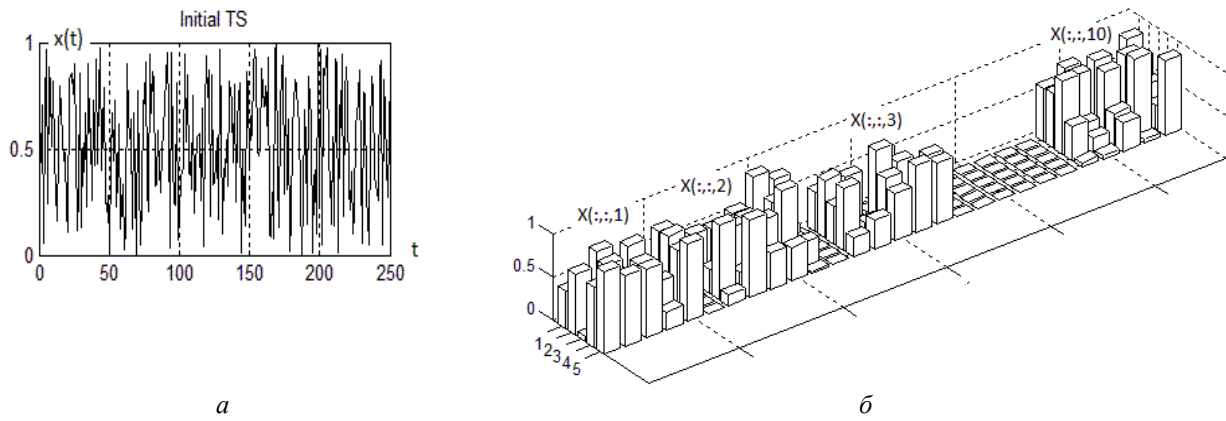


Рис. 6. Исходный ВР (а) и его тензорная модель ВР (б)

Тензорная модель ВР: 2D-тензор-окна с матрицами 5×5 приведена ниже.

$x(:,:,1) =$ 0.44 0.24 0.47 0.68 0.11 0.71 0.88 0.30 0.54 0.83 0.06 0.80 0.16 0.44 0.85 0.70 0.51 0.51 0.48 0.75 0.96 0.82 0.79 0.21 0.90	$x(:,:,2) =$ 0.52 0.50 0.58 0.92 0.79 0.12 0.39 0.59 0.43 0.28 0.28 0.67 0.25 0.77 0.21 0.86 0.23 0.73 0.98 0.28 0.13 0.90 0.42 0.40 0.02	$x(:,:,3) =$ 0.46 0.36 0.03 0.64 0.38 0.67 0.61 0.11 0.12 0.05 0.44 0.12 0.90 0.67 0.26 0.73 0.23 0.65 0.01 0.73 0.23 0.33 0.55 0.77 0.71
$x(:,:,4) =$ 0.91 0.86 0.41 0.95 0.24 0.30 0.82 0.29 0.66 0.24 0.20 0.32 0.49 0.03 0.81 0.95 0.33 0.58 0.95 0.07 0.67 0.59 0.71 0.45 0.17	$x(:,:,5) =$ 0.37 0.84 0.19 0.69 0.87 0.38 0.39 0.40 0.04 0.18 0.42 0.44 0.55 0.67 0.85 0.94 0.31 0.57 0.94 0.33 0.53 0.75 0.26 0.74 0.34	$x(:,:,6) =$ 0.44 0.10 0.82 0.61 0.26 0.55 0.41 0.72 0.74 0.14 0.17 0.37 0.44 0.81 0.86 0.96 0.54 0.80 0.01 0.78 0.63 0.32 0.25 0.49 0.41
$x(:,:,7) =$ 0.65 0.49 0.80 0.30 0.13 0.56 0.46 0.61 0.08 0.70 0.88 0.50 0.93 0.97 0.85 0.97 0.88 0.10 0.99 0.94 0.90 0.70 0.08 0.38 0.28	$x(:,:,8) =$ 0.83 0.58 0.27 0.15 0.36 0.31 0.83 0.01 0.84 0.40 0.64 0.80 0.26 0.72 0.10 0.02 0.67 0.68 0.64 1.00 0.16 0.64 0.53 0.01 0.24	$x(:,:,9) =$ 0.97 0.41 0.92 0.95 0.31 0.05 0.39 0.90 0.18 0.21 0.87 0.15 0.01 0.76 0.39 0.85 0.55 0.71 0.89 0.10 0.98 0.14 0.60 0.76 0.19

Решение уравнений $AX+XB=C$: 1 — окна 1, 2, 3; 2 — 4, 5, 6; 3 — 7,8,9

X1	X2	X3
-1.49 -0.80 1.72 0.80 -0.90	-1.89 3.38 -2.81 0.90 -0.20	-0.13 -0.35 0.77 -1.01 1.86
-0.71 0.08 -0.33 -0.07 1.69	7.40 -4.53 0.49 -5.92 4.31	-0.81 0.12 1.32 -1.67 1.46
-0.49 0.76 0.39 0.20 -0.53	0.16 3.99 -2.24 -1.27 -0.60	1.54 -2.05 -3.03 3.40 0.84
2.52 0.64 0.31 -2.93 2.56	1.51 0.45 2.46 -1.15 -2.05	-0.83 2.56 1.69 -0.83 -3.01
0.03 -0.48 -0.68 2.14 -2.39	-5.32 -3.46 2.68 6.68 -0.26	0.60 0.35 1.56 -1.27 -0.18

Сингулярные величины	Нормы матрицы связей:	Квадрат следа su =
5.86 13.51 6.78	n = 6.71 16.80 8.06	45.03 282.10 64.96
2.51 9.06 4.14		
1.81 4.15 1.25		
1.01 0.50 0.44		
0.37 0.08 0.09		

Решения матричных уравнений $AX+XB=C$, где A, B, C — последовательные тензор-окна ВР: $x(:,:,1), x(:,:,2), x(:,:,3)$ и т.д. соответственно с матрицами 5×5. ВР при наличии аномалий приведены на рис. 7.

Матрицы X вычислены для кортежей: $X1 - \{x(:,:,1), x(:,:,2), x(:,:,3)\}$, $X2 - \{x(:,:,4), x(:,:,5), x(:,:,6)\}$, $X3 - \{x(:,:,7), x(:,:,8), x(:,:,9)\}$, их Φ -нормы, сингулярные векторы и следы приведены на рис. 8.

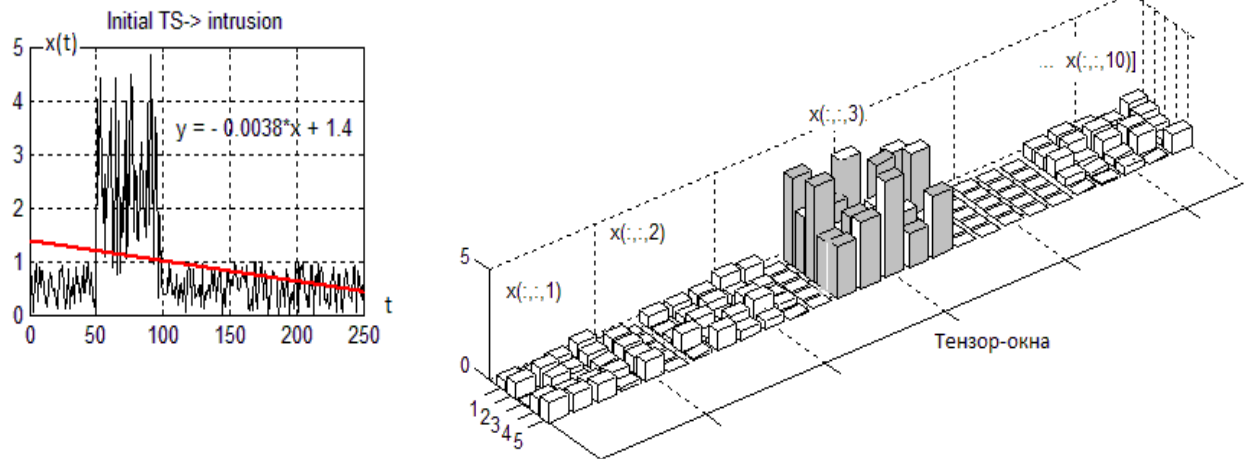


Рис. 7. ВР с внедрением аномалий (а); тензорная модель ВР (наличие аномалий) (б)

Тензорная модель ВР: 2D-тензор-окна с матрицами 5x5

$x(:,:,1) =$ 0.44 0.24 0.47 0.68 0.11 0.71 0.88 0.30 0.54 0.83 0.06 0.80 0.16 0.44 0.85 0.70 0.51 0.51 0.48 0.75 0.96 0.82 0.79 0.21 0.90	$x(:,:,2) =$ 0.52 0.50 0.58 0.92 0.79 0.12 0.39 0.59 0.43 0.28 0.28 0.67 0.25 0.77 0.21 0.86 0.23 0.73 0.98 0.28 0.13 0.90 0.42 0.40 0.02	$x(:,:,3) =$ 4.03 1.08 3.86 0.74 3.24 2.62 2.62 0.88 3.60 1.05 4.41 1.81 1.34 0.78 4.00 2.22 2.87 2.91 2.08 1.27 2.35 3.02 4.43 1.54 2.80
$x(:,:,4) =$ 4.49 1.46 1.84 4.86 0.78 4.07 1.37 1.43 3.22 0.33 1.54 2.45 3.97 0.95 1.86 2.69 3.23 3.08 3.67 0.54 2.55 2.18 1.67 3.11 0.45	$x(:,:,5) =$ 0.37 0.84 0.19 0.69 0.87 0.38 0.39 0.40 0.04 0.18 0.42 0.44 0.55 0.67 0.85 0.94 0.31 0.57 0.94 0.33 0.53 0.75 0.26 0.74 0.34	$x(:,:,6) =$ 0.44 0.10 0.82 0.61 0.26 0.55 0.41 0.72 0.74 0.14 0.17 0.37 0.44 0.81 0.86 0.96 0.54 0.80 0.01 0.78 0.63 0.32 0.25 0.49 0.41
$x(:,:,7) =$ 0.65 0.49 0.80 0.30 0.13 0.56 0.46 0.61 0.08 0.70 0.88 0.50 0.93 0.97 0.85 0.97 0.88 0.10 0.99 0.94 0.90 0.70 0.08 0.38 0.28	$x(:,:,8) =$ 0.83 0.58 0.27 0.15 0.36 0.31 0.83 0.01 0.84 0.40 0.64 0.80 0.26 0.72 0.10 0.02 0.67 0.68 0.64 1.00 0.16 0.64 0.53 0.01 0.24	$x(:,:,9) =$ 0.97 0.41 0.92 0.95 0.31 0.05 0.39 0.90 0.18 0.21 0.87 0.15 0.01 0.76 0.39 0.85 0.55 0.71 0.89 0.10 0.98 0.14 0.60 0.76 0.19

AX+XB=C : матрицы X:

X1					X2					X3				
6.34	-7.55	-9.76	6.04	-5.46	0.08	0.01	-0.13	-0.41	0.45	-0.13	-0.35	0.77	-1.01	1.86
0.23	-1.31	6.52	-1.98	-1.79	-0.94	0.02	-0.07	1.04	-0.10	-0.81	0.12	1.32	-1.67	1.46
-0.82	-7.24	-4.02	14.39	-11.14	1.27	-0.79	0.24	-0.81	0.49	1.54	-2.05	-3.03	3.40	0.84
-24.85	5.11	10.92	5.95	-1.55	-3.47	1.79	-0.88	2.28	-0.13	-0.83	2.56	1.69	-0.83	-3.01
14.42	8.26	-4.66	-15.70	15.71	4.37	-1.44	1.40	-2.51	-0.63	0.60	0.35	1.56	-1.27	-0.18

Сингулярные величины			Нормы матриц X			Квадрат следа		
38.72	7.41	6.78	n =	50.29	7.52 8.06	2529.08	56.55	64.96
31.10	1.07	4.14						
7.63	0.72	1.25						
2.13	0.19	0.44						
0.16	0.05	0.09						

Решения матричных уравнений $AX+XB=C$, где A, B, C — последовательные тензор-окна ВР: $x(:,:,1), x(:,:,2), x(:,:,3)$ и т. д. соответственно с матрицами 5×5 ; матрицы X вычислены для кортежей: $X1$ — $\{x(:,:,1), x(:,:,2), x(:,:,3)\}$, $X2$ — $\{x(:,:,4), x(:,:,5), x(:,:,6)\}$, $X3$ — $\{x(:,:,7), x(:,:,8),$

$x(:,:,9)\}$, их Ф-нормы, сингулярные векторы и следы приведены на рис. 8, б.

Сравнительная визуальная оценка свойств тензорных моделей нормального ВР и ВР с аномалиями приведена на рис. 8.

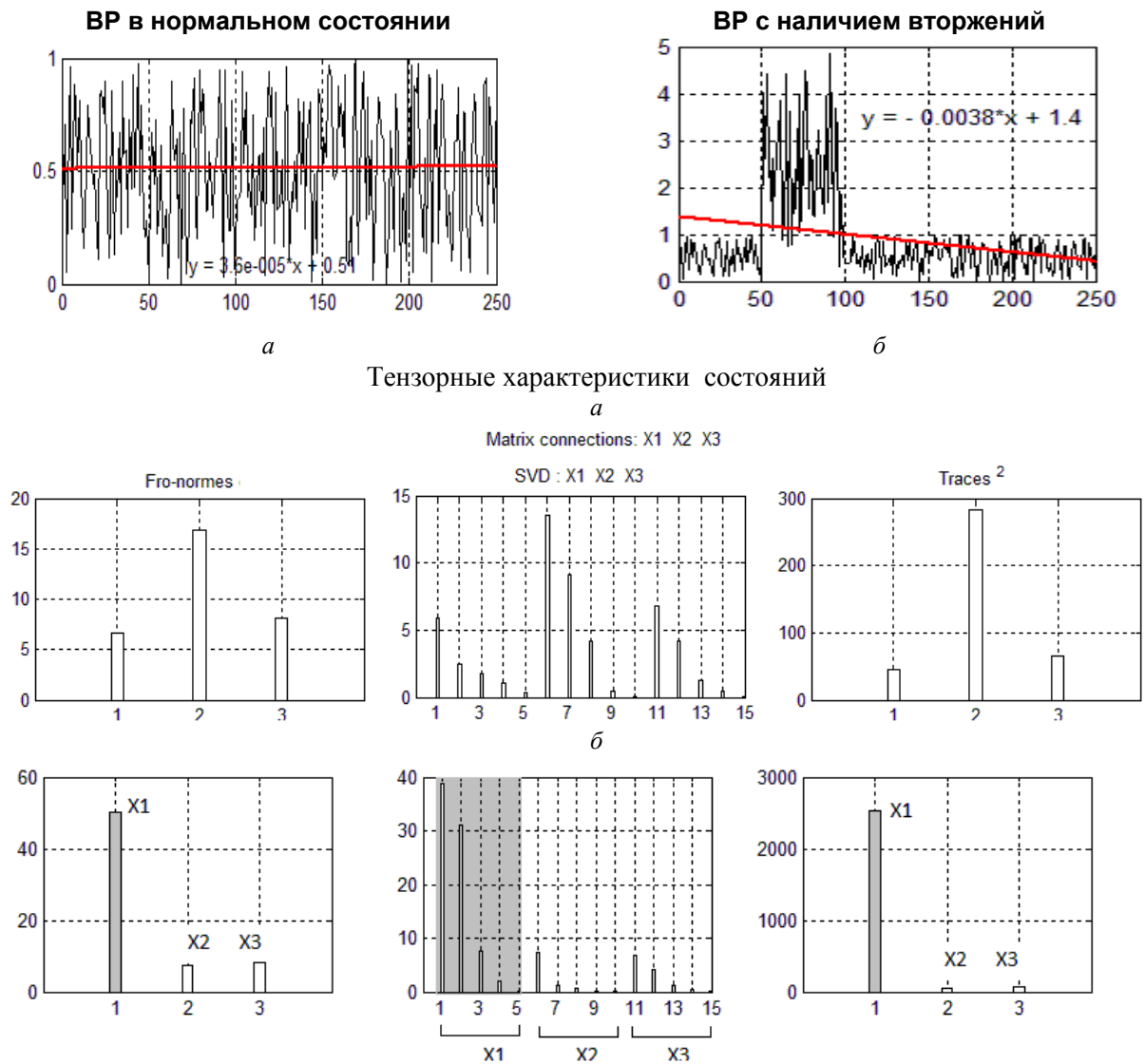


Рис. 8. Сравнительная визуальная оценка тензорных моделей ВР:
 а — нормальное состояние ВР; б — наличие аномалий ВР.

Сравнивая соответствующие характеристики (Ф-нормы: $a \rightarrow \bar{b}$, компоненты сингулярных векторов: $a \rightarrow \bar{b}$, квадраты следов: $a \rightarrow \bar{b}$), можно однозначно идентифицировать наличие (факт) и место (окно) аномалии ВР.

Выводы

1. Предложена тензорная модель ВР, основанная на формировании тензор-фрагментов размерностью $m_1 \times m_2 \times m_3$, где m_3 — количество окон в фрагменте ВР, тензор-окно — фронтальный слайс 3D-тензор-фрагмента, где 3D-матрица размерностью $m_1 \times m_2 \times m_3$ состоит из 2D-матриц размерностью $m_1 \times m_2$, собранных в массив размерностью m_3 . Для каждого окна, фрагмента (сегмента) ВР сформированы матрицы 2D- и 3D-тензоров $T_{d_w}, T_{d_f}, T_{d_s}$ соответственно, вычислены сингулярные разложения и инварианты тензоров.

2. Интеллектуальный анализ ВР реализован в виде решения матричных уравнений, связываю-

щих последовательные тензор-окна или тензор-фрагменты ВР. Обобщенная характеристика минимального фрагмента ВР (состоящего из 3-х окон 2D-тензоров) представлена в виде решения матричных уравнений: $AXB=C$ или/и $AX+XB=C$, где A, B, C — 2D-тензоры (с матрицами $m \times m$) соседних окон (или фрагментов).

3. На основании полученных матриц $X^{(1)}, X^{(2)}, \dots, X^{(w)}$ — решений соответствующих МУ, вычисляются инварианты, матрицы сингулярных величин $S^{(i)}$ $i=1,w$; Фробениусовы расстояния между матрицами $X^{(i)}$ и $S^{(i)}$ — $ds(X^{(i-1)}, X^{(i)})$ и $ds(S^{(i-1)}, S^{(i)})$ соответственно, определяются тренды следов, норм, расстояний.

4. Изменение тренда свидетельствует о возможном наличии аномалии или выпадении из состояния норма. Результаты моделирования подтверждают эффективность предложенного метода выявления аномалий трафика для предупреждения вторжений в КС.

ЛИТЕРАТУРА

1. *Cheboli D.* Anomaly Detection of Time Series. Facility Of The Graduate School Of The Uni-versity Of Minnesota. — 2010. — 75 с. — [Электронный ресурс] — Режим доступа: http://con-servancy.umn.edu/bitstream/11299/92985/1/Cheboli_Deepthi_May2010.pdf (дата обращения: 20. 04.2014).
2. *Foslien W., Guralnik V., Haigh K.* — Z. Data Mining For Space Applications. SpaceOps 2004 — Conference. Montreal, Canada — May 17–21 2004. Honeywell Laboratories, 3660 Techno-logy Drive, Minneapolis, MN 55418.
3. *Binglin X., Zhanhuai L.* An Anomaly Detection Method for Spacecraft Using ICA Technology. International Conference on Advanced Computer Science and Electronics Information (ICACSEI 2013). — P. 50–54
4. *Mahoney Mat. V. and Chan Ph. K.* Trajectory Boundary Modeling of Time Series for Anomaly Detection. *KDD-05*, Aug. 21, 2005, Chicago IL, USA. Copyright 2005 ACM 1-58113-000-0/00/ 0004.
5. *Арский Ю. М.* Принципы конструирования интеллектуальных систем / Ю. М. Арский, В. К. Финн // Информационные технологии и вычислительные системы. — № 4/2008. — С. 1–37.
6. *Cichocki A.* Tensor Decompositions: A New Concept in Brain Data Analysis arXiv: 1305. 0395v1 [cs.NA] 2 May 2013. — 19 pp.
7. *Iverson D. L.* System Health Monitoring for Space Mission Operations. Интернет-ресурс: [http://ti.arc.nasa.gov/m/pub-archive/1389h/1389%20\(Iverson\).pdf](http://ti.arc.nasa.gov/m/pub-archive/1389h/1389%20(Iverson).pdf)
8. *Кендалл М. Дж.* Курс статистики (в 3-х томах) / М.Дж. Кендалл, А. Стьюарт. — М. : Наука, Физматлит: 1966, 1973, 1976. — С.: 588+466+375. Т. 3. Многомерный статистический анализ и временные ряды.
9. Хеннан Э. Многомерные временные ряды / Э. Хеннан. — М. : Мир, 1974. — 576 с.
10. *Главные компоненты временных рядов: метод «Гусеница»;* под ред. Д. Л. Данилова, А. А. Жиглявского. СПб. : Пресском, 1997. —308 с.
11. *Эсбенсен К.* Анализ многомерных данных. Избранные главы / К. Эсбенсен; пер. с англ. С. В. Кучерявского; под ред. О. Е. Родионовой. — Черногловка : ИПХФ РАН, 2005. — 160 с.
12. *Han J., Kamber M.* Data Mining: Concepts and Techniques, 3rd ed. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers, March 2011.
13. *Rogers M., Li L., Russell St.* Multilinear Dynamical Systems for Tensor Time Series. In book “Advances in Neural Information Processing Systems” (NIPS 2013). — 2013. — P. 2634–2642.
14. *Dobos L., Abonyiu J.* On-line detection of homogeneous operation ranges by dynamic principal component analysis based time-series segmentation // Chemical Engineering Science 75 (2012). — P. 96–105.
15. *Ringberg H., Soule A., Rexford J., Diot Ch.* Sensitivity of PCA for Traffic Anomaly Detection // *SIGMETRICS’07*, June 12–16, 2007, San Diego, California, USA. Copyright 2007 ACM 978-1-59593-639-4/07/0006.
16. *Skillicorn D.* Data Mining and Knowledge Discovery Series. Understanding Complex Datasets. Data Mining with Matrix Decompositions. Chapman & Hall/CRC.- 2007.- 257 pp.
17. *Laub A. J.* Matrix Analysis for Scientists and Engineers. — 2005. — 158 p. Интернет-ресурс: www.c-securehost.com/SIAM/ot91.html
18. *Kamalja K. K., Khangar N. V.* Singular Value Decomposition for Multidimensional Matrices Int. Journal of Engineer. Research and Applications: 2248-9622, Vol. 3, Issue 6, Nov-Dec 2013. — P. 123–129.
19. *Lototsky S. V.* Simple spectral bounds for sums of certain Kronecker products. Linear Algebra and its Applications // 469 (2015). — P. 114–129.
20. *COM521500 Math. Methods for SP I Lecture 11: Matrix Equations and the Kronecker Product.* Режим доступа : [http://www.oz.nthu.edu.tw/~d915691/files_com521_5/lecture11\(1222\).pdf](http://www.oz.nthu.edu.tw/~d915691/files_com521_5/lecture11(1222).pdf) (дата обращения 14.04.2016).

Стаття надійшла до редакції 28.05.2016