

УДК 004.93

DOI 10.22213/2413-1172-2018-4-194-199

ВЫБОР БАЗОВОЙ ФУНКЦИИ ПРИ АВТОМАТИЗИРОВАННОЙ ИДЕНТИФИКАЦИИ ВРЕМЕННЫХ РЯДОВ НА ОСНОВЕ АССОЦИАТИВНО-МАЖОРИТАРНОГО ПОДХОДА*

Т. З. Аралбаев, доктор технических наук, профессор, Оренбургский государственный университет, Оренбург, Россия

Т. В. Абрамова, аспирант, Оренбургский государственный университет, Оренбург, Россия

Р. Р. Галимов, кандидат технических наук, доцент, Оренбургский государственный университет, Оренбург, Россия

Д. А. Гайфулина, студент, Оренбургский государственный университет, Оренбург, Россия

Э. Р. Хакимова, студент, Оренбургский государственный университет, Оренбург, Россия

Предложена модель выбора базовых функций при автоматизированной идентификации временных рядов, алгоритм и программное средство идентификации временных рядов на основе ассоциативно-мажоритарного подхода, позволяющие идентифицировать вид базовых зависимостей динамических процессов.

Задача идентификации вида базовых функций решается с использованием классической теории распознавания образов. Идентификация производится путем сравнения исходного образа с эталонами, хранящимися в едином пространстве образов. Пространство образов представляет собой область памяти устройства (ассоциативной памяти), на котором производится идентификация. Особенностью предложенной модели является оперативность сравнения исходных образов с образами эталонов за счет использования единого признакового пространства и возможности сравнения всех образов за один такт. Для повышения оперативности идентификации в алгоритме выбора базовых функций также использован ассоциативно-мажоритарный подход к хранению и поиску идентификационных данных в электронной памяти.

Предложенный алгоритм и программное средство выбора базовых функций являются универсальными, так как позволяют идентифицировать вид базовой зависимости в любом динамическом процессе независимо от специфики исследуемой предметной области. Автоматизированная идентификация вида базовых функций сокращает время построения прогнозных моделей и позволяет оперативно прогнозировать дальнейшие варианты протекания процесса.

Ключевые слова: базовые зависимости, идентификация, временные ряды, прогнозирование, динамические процессы, ассоциативно-мажоритарный подход, распознавание образов.

Введение

В ходе современных научных исследований все чаще возникает необходимость наглядного представления статистического материала о значениях каких-либо параметров исследуемого процесса. Очень часто такая информация представляется в виде временных рядов, которые могут быть проанализированы и изучены аналитическими и программными способами. Но такие данные характеризуются сложными зависимостями, которые часто трудны для определения. Примером таких зависимостей могут являться данные интенсивности сетевого трафика (рис. 1). Так как подобные зависимости зачастую являются основой для построения прогнозных моделей, возникает

необходимость в определении вида функции временного ряда.

В основе моделирования и прогнозирования большинства динамических процессов лежит операция идентификации базовой функции (БФ), характеризующей основные зависимости в исследуемом процессе. В известной литературе, касающейся методов моделирования и прогнозирования процессов [1–7], распознавание БФ в большинстве случаев производится либо визуально, либо путем последовательного сравнения с эталонами. При этом достоверная идентификация БФ занимает значительное количество времени, что является проблемой в случае необходимости быстрой реакции на возникновение нерегламентированной ситуации (напри-

© Аралбаев Т. З., Абрамова Т. В., Галимов Р. Р., Гайфулина Д. А., Хакимова Э. Р., 2018

* Исследование выполнено при финансовой поддержке РФФИ и правительства Оренбургской области (проект № 18-47-560012 «Оптимизация методов контроля технического состояния распределенных автоматизированных систем в условиях воздействия пространственно-временных угроз на основе мониторинга сетевых информационных потоков»).

мер, сетевой атаки), распознаваемой и прогнозируемой с помощью тренда процесса. По этой причине автоматизированная идентификация базовых зависимостей является одной из актуальных задач при прогнозировании процессов.

Целью данной работы является автоматизация процесса выбора базовой функции при идентификации временных рядов. Для достижения данной цели были разработаны: модель распознавания БФ, алгоритм и программное средство, осуществляющее выбор БФ, и исследована программа при работе с реальными данными временных рядов. Для повышения оперативности идентификации в алгоритме выбора БФ ис-

пользован ассоциативно-мажоритарный подход к хранению и поиску идентификационных данных в электронной памяти. За счет поиска данных по ассоциациям ассоциативный подход увеличивает скорость доступа и сокращает время поиска информации [8, 9].

Модель распознавания базовых функций

В основе алгоритма идентификации тренда временного ряда лежит оценка степени соответствия исходного образа временного ряда эталонным образам элементарных базовых функций. Формулы и графики основных элементарных функций, используемых для прогнозирования, представлены в табл. 1.

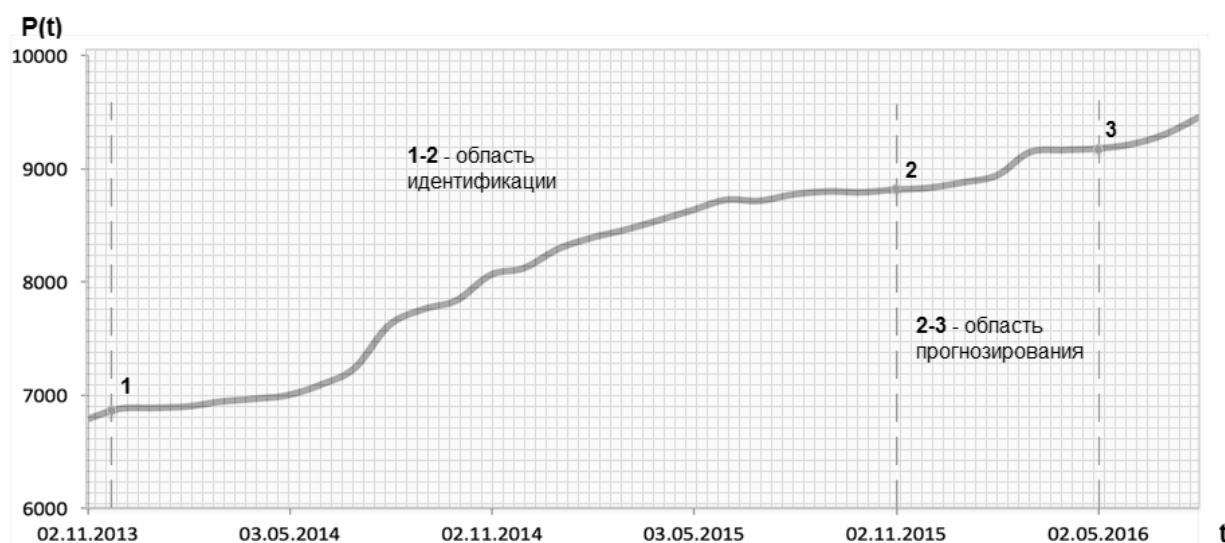


Рис. 1. Фрагмент сетевого трафика: t – время; $P(t)$ – функция зависимости интенсивности сетевого трафика от времени

Таблица 1. Базовые функции, используемые для прогнозирования

Название и формула функции	График функции	Название и формула функции	График функции
Линейная $y = kx + b$		Показательная $y = e^x$	
Квадратичная $y = ax^2 + bx + c$		Логарифмическая $y = \log_a x$	
Степенная $y = x^n$		$y = \sin x$ $y = \cos x$	

Задача идентификации БФ, как правило, решается с использованием классической теории распознавания образов. Известны:

$Q = \{Q_1, Q_2, \dots, Q_N\}$ – множество образов (эталонов) базовых функций;

N – количество образов БФ;

X – множество ассоциативных признаков распознавания $X = \{x_1, x_2, \dots, x_M\}$, характеризующих исходный временной ряд;

M – количество элементов временного ряда;

$D = \{D_{11}, D_{12}, \dots, D_{ij}\}$ – диапазон изменения i -го признака для j -го класса функций, в пределах которого лежат логические значения функций определенного эталона БФ: $D_{ij} = \{0,1\}$;

RP – разделяющее правило для идентификации;

RF – разделяющая функция идентификации БФ, определяющая меру близости распознаваемого образа с эталоном;

v – коэффициент ассоциативности.

Необходимо идентифицировать вид базовой функции исходного временного ряда для получения прогнозных оценок его изменения.

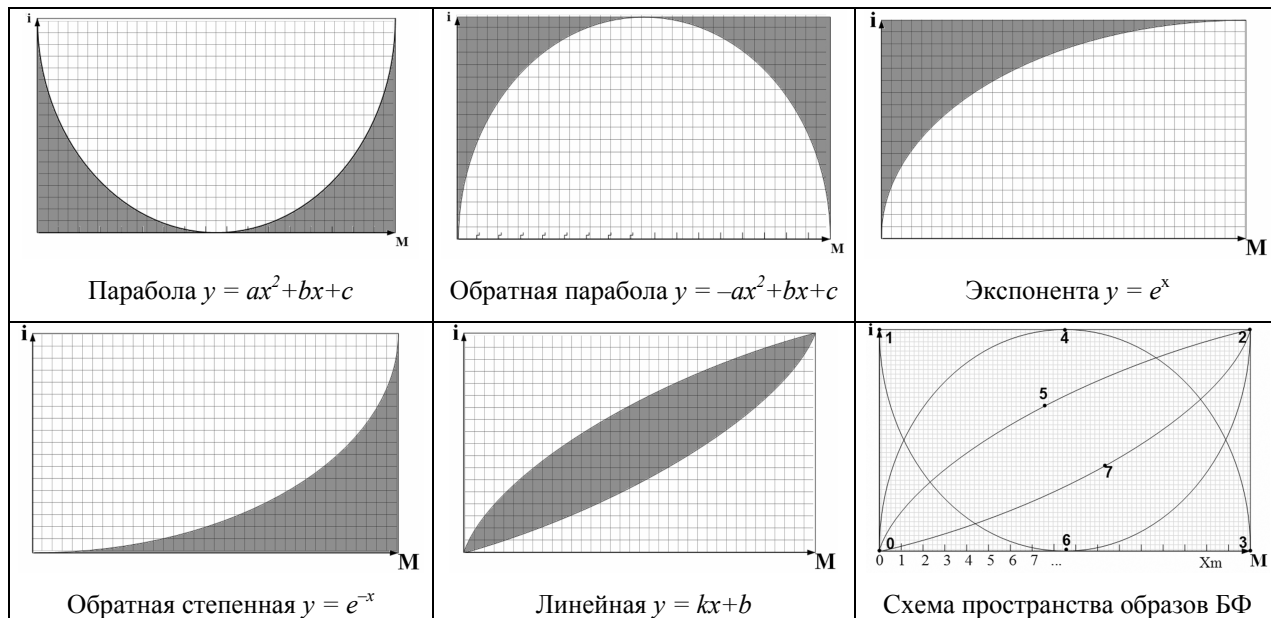
Для проведения идентификации на первом этапе необходимо составить библиотеку этало-

нов базовых функций (БФ), с каждым из которых будет проводиться сравнение исходного образа. В качестве данных эталонов могут использоваться эталоны одной из элементарных функций, представленных в табл. 1.

Графическая модель эталона представляет собой квадратную матрицу размерностью $n \times n$, заполняемую единицами в случае присутствия значения исходной функции в конкретной ячейке матрицы эталонов. Диапазоны изменения некоторых видов БФ представлены в табл. 2. Для примера были выбраны линейная, параболическая и экспоненциальная функции.

Все образы БФ хранятся в одном участке памяти, и идентификация того или иного вида функции производится в зависимости от степени нахождения исходного графика функции в диапазоне определенного образа. В правой нижней ячейки табл. 2 представлена схема пространства образов и показаны диапазоны нахождения базовых функций в едином пространстве образов: 0-5-2-7-0 – линейной функции, 1-6-2-3-6-0-1 и 0-4-3-2-4-1-0 – параболической и обратной параболической функций, 0-5-2-4-1-0 и 0-7-2-3-6-0 – экспоненциальной и обратной экспоненциальной функций.

Таблица 2. Графические образы БФ



Пространство образов представляет собой область памяти устройства, на котором производится идентификация. В качестве ЗУ для хранения эталонов используются микросхемы ассоциативной памяти, в которых в адресных полях записываются координаты каждой из клеток матрицы, а в полях данных значения 1 (в случае

заполненной клетки матрицы) или 0 (в случае пустой клетки матрицы).

Для распознавания подается неизвестный образ q^x , для удобства распознавания предварительно масштабированный в соответствии с размерами пространства образов, и сравнивается с хранящимися в нем образами.

В результате сравнения исходного тренда временного ряда с эталонами в пространстве образов определяется процентное соотношение вероятности совпадения с каждой базовой функцией по числу ассоциаций (единиц в матрицах). Вид функции исходного временного ряда определяется по полученным процентным соотношениям: если значение вероятности совпадения по конкретному виду БФ наибольшее – данному виду и принадлежит исходная функция.

В предложенной модели идентификации вида БФ используется разделяющая функция (1), позволяющая оценить степень соответствия исходного образа с эталонными образами базовых функций. Задача распознавания ставится следующим образом: необходимо идентифициро-

вать неизвестный образ функции q^x , отнеся его к одному из известных эталонов Q^x множества образов Q .

$$RP: q^x \in Q^x \in Q;$$

$$RF^* \equiv RF_{\max} \{RF\}; \quad (1)$$

$$RF_i = \sum_{j=1}^M v_{ij}, \quad v_{ij} = \begin{cases} 1, & \text{если } x_j \in d_{ij}, \\ 0, & \text{если } x_j \notin d_{ij}; \end{cases}$$

Вид базовой функции RF^* определяется по наибольшему числу соответствий значений RF_{\max} исходного образа со значениями диапазона определенного эталона D .

Функциональная схема модели распознавания на основе ассоциативно-мажоритарного подхода представлена на рис. 2.

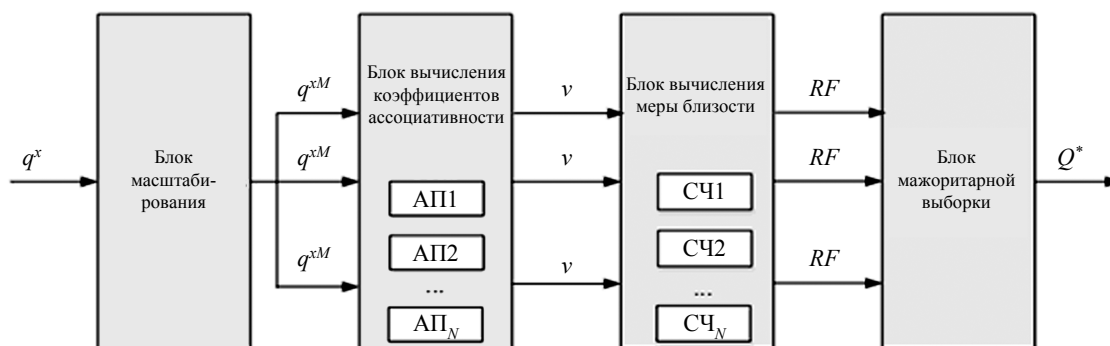


Рис. 2. Функциональная схема модели распознавания на основе ассоциативно-мажоритарного подхода

На вход устройства подается исходный идентифицируемый образ q^x . Для реализации предложенного алгоритма используется сглаженный ряд данных исходной функции, то есть линия тренда функции. Ввиду неограниченности размеров временных рядов исходный образ масштабируется в соответствии с размерностью пространства эталонов (q^{xM}), после чего сравнивается с каждым из эталонов в ассоциативных ЗУ (АП). В результате сравнения этих двух образов счетчиком (СЧ) подсчитывается суммарное значение совпадений (ассоциаций) между исходным образом и эталоном, после чего определяет вид идентифицируемой функции Q^* . Блок мажоритарной выборки производит идентификацию исходного образа по большинству ассоциаций исходного образа с эталонами. Критерий выбора базовой функции основан на мере близости Хемминга.

Особенностью предложенной модели является оперативность сравнения исходных образов с образами эталонов за счет использования единого признакового пространства и возможности сравнения всех образов за один такт.

Алгоритм и программное средство идентификации базовых функций

Для исследования работы предложенного алгоритма идентификации было разработано программное средство [10], осуществляющее выбор базовой функции при идентификации временных рядов на основе ассоциативно-мажоритарного подхода.

Исходными данными для работы прикладной программы являются: зарегистрированный временной ряд, образы базовых функций (БФ) эталонов трендов, представленные в матричной форме. Для проведения идентификации используется библиотека эталонов базовых функций. В качестве примеров выбраны линейная, параболическая и экспоненциальная функции.

При загрузке исходного образа (файла данных временного ряда) для корректной работы программы производится масштабирование исходных данных в требуемом диапазоне. Далее производится сглаживание графика функции с целью получения тренда. Полученный тренд переводится в матричную форму для последующего сравнения с эталонами базовых

функций. Матрица тренда имеет размерность 100×100 и заполняется единицами в зависимости от формы тренда.

Результатом выполнения программы является вывод о принадлежности функции временного ряда одному из эталонных образов, с указанием оценки вероятности принадлежности к каждому образцу. После сравнения матрицы

исходного тренда сетевого трафика и матриц базовых функций эталонов на форме отображается процентное соотношение вероятности совпадения с каждой базовой функцией. В качестве условия достоверности идентификации функции принят порог совпадения более чем на 50 %. Экранная форма работы программы представлена на рис. 3.

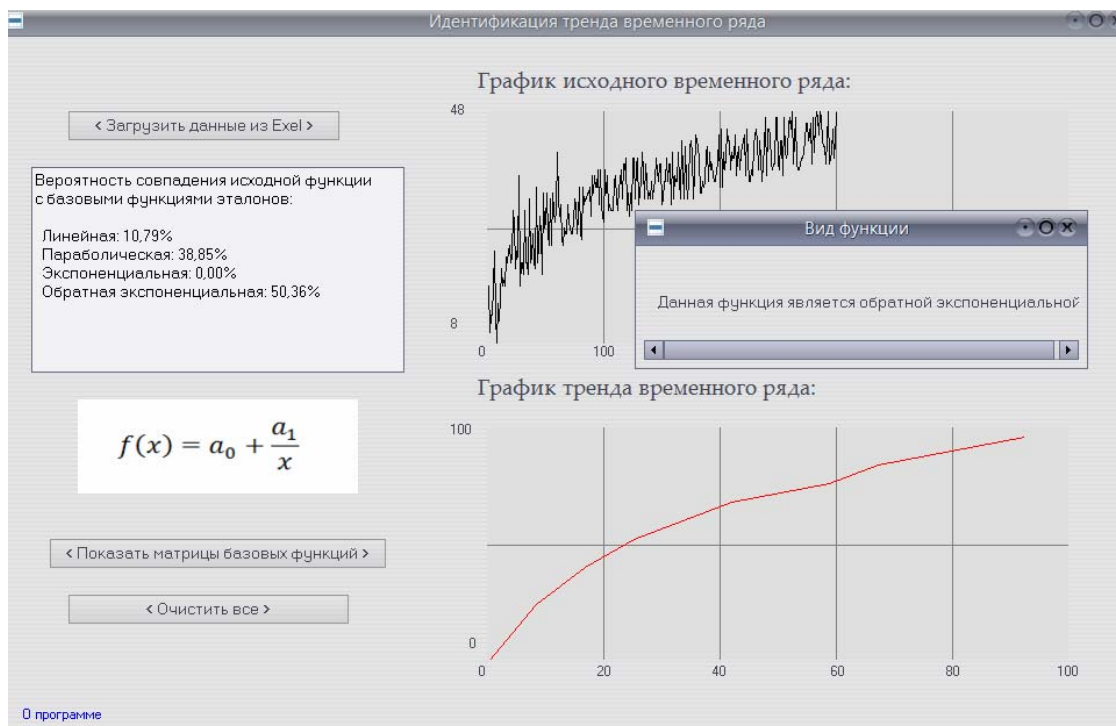


Рис. 3. Экранная форма программы идентификации БФ

Достоинствами разработанной программы являются: оперативность распознавания исходной функции временного ряда за счет автоматизации процесса идентификации и достоверность распознавания вида базовой функции за счет подсчета процентного соотношения исходной функции временного ряда с базовыми функциями эталонов. Прикладная программа также может быть использована в учебном процессе при проведении лабораторных работ, связанных с моделированием и прогнозированием динамических процессов.

Вывод

Предложенный алгоритм и программное средство являются универсальными, так как позволяют идентифицировать вид базовой зависимости в любом динамическом процессе независимо от специфики исследуемой предметной области. Автоматизированная идентификация вида БФ позволит оперативно прогнозировать дальнейшие варианты протекания процесса.

Библиографические ссылки

1. Афанасьев В. Н., Юзбашев М. М. Анализ временных рядов и прогнозирование : учебник. М. : Финансы и статистика; ИНФРА-М, 2010. 228 с. : ил.
2. Афанасьев В. Н., Лебедева Т. В. Моделирование и прогнозирование временных рядов: учеб.-метод. пособие для вузов. М.: Финансы и статистика, 2009. 292 с. : ил.
3. Подвальный Е. С. Модели индивидуального прогнозирования и классификации состояний в системах компьютерного мониторинга. Воронеж: Изд-во ВГТУ, 1998. 127 с.
4. Громова Н. М., Громова Н. И. Основы экономического прогнозирования. Старая Русса: Академия естествознания, 2006. 81 с.
5. Пат. 2430415 Российская Федерация, МПК G 06 K 9/00. Устройство для распознавания образов / Р. И. Хасанов, М. З. Масыгутов, Т. З. Аралбаев; заявитель и патентообладатель Оренбургский государственный университет. № 2010116601/08; заявл. 26.04.2010, опубл. 27.09.2011, Бюл. № 27. 21 с. : ил.
6. Chevaleyre Y., Endriss U., Maudet N. Simple negotiation schemes for agents with simple preferences: suffi-

ciency, necessity and maximality. *Autonomous Agents and Multi. Agent Systems*, 2010, no. 20, pp. 234-259.

7. *Abdo H., Kaouk M., Flaus J.-M., Masse F.* A safety/security risk analysis approach of Industrial Control Systems: A cyber bowtie - combining new version of attack tree with bowtie analysis. *Computers & Security*, 2018, vol. 72, pp. 175-195. DOI: 10.1016/j.cose.2017.09.004.

8. *Аралбаев Т. З., Абрамова Т. В.* Исследование эффективности методов мониторинга сетевого трафика на основе последовательного и ассоциативно-последовательного принципов поиска актуальной информации // СТИН. 2017. № 11. С. 2–5.

9. *Aralbaev T. Z., Abramova T. V.* Network Traffic Monitoring on the Basis of Sequential and Associative–Sequential Search Principles. *Russian Engineering Research*, 2018, vol. 38, no. 5, pp. 381-383.

10. *Аралбаев Т. З., Хакимова Э. Р., Гайфулина Д. А.* Идентификация тренда временного ряда: прикладная программа / Мин-во образования и науки РФ ; Оренбургский гос. ун-т. Электрон. текстовые данные (1 файл: 3 Мб). Оренбург : ОГУ, 2016.

References

1. *Afanasyev V. N., Yuzbashev M. M.* *Analiz vremennyh rjadov i prognozirovanie* [Time series analysis and forecasting: Tutorial]. Moscow, Finansy i statistika Publ., 2010, 228 p. (in Russ.).

2. *Afanasyev V. N., Lebedeva T. V.* *Modelirovanie i prognozirovanie vremennyh rjadov* [Modeling and forecasting time series]. Moscow, Finansy i statistika Publ., 2009, 292 p. (in Russ.).

3. *Podvalny E. S.* *Modeli individual'nogo prognozirovaniya i klassifikacii sostojanij v sistemah komp'yuternogo monitoringa* [Models of individual forecasting and

state classification in computer monitoring systems]. Voronezh, VGTU Publ., 1998, 127 p. (in Russ.).

4. *Gromova N. M., Gromova N. I.* *Osnovy jekonomicheskogo prognozirovanija*. [Basics of Economic Forecasting]. Staraja Russa, Akademija estestvoznaniya Publ., 2006, 81 p. (in Russ.).

5. *Khasanov R. I., Masyagutov M. Z., Aralbayev T. Z.* *Ustrojstvo dlja raspoznavanija obrazov* [A device for pattern recognition]. Pat. RF, no. 2010116601/08, 2011.

6. *Chevaleyre Y., Endriss U., Maudet N.* Simple negotiation schemes for agents with simple preferences: sufficiency, necessity and maximality. *Autonomous Agents and Multi. Agent Systems*, 2010, no. 20, pp. 234-259.

7. *Abdo H., Kaouk M., Flaus J.-M., Masse F.* A safety/security risk analysis approach of Industrial Control Systems: A cyber bowtie - combining new version of attack tree with bowtie analysis. *Computers & Security*, 2018, vol. 72, pp. 175-195. DOI: 10.1016/j.cose.2017.09.004.

8. *Aralbaev T. Z., Abramova T. V.* *Issledovanie jefektivnosti metodov monitoringa setevogo trafika na osnove posledovatel'nogo i associativno posledovatel'nogo principov poiska aktual'noj informacii* [Investigation of the effectiveness of methods for monitoring network traffic based on a consistent and associatively consistent principles of finding relevant information]. *STIN*, 2017, no. 11, pp. 2-5 (in Russ.).

9. *Aralbaev T. Z., Abramova T. V.* Network Traffic Monitoring on the Basis of Sequential and Associative - Sequential Search Principles. *Russian Engineering Research*, 2018, vol. 38, no. 5, pp. 381-383.

10. *Aralbaev T. Z., Khakimova E. R., Gaifulina D. A.* *Identifikacija trenda vremennogo rjada : prikladnaja programma* [Identification of the trend of a time series: application program]. Orenburg, Orenburg State University, 2016 (in Russ.).

Selection of Basic Function for Identification of Time Series Based on Associative Majority Approach

T. Z. Aralbaev, DSc in Engineering, Professor, Orenburg State University, Orenburg, Russia

T. V. Abramova, Post-graduate, Orenburg State University, Orenburg, Russia

R. R. Galimov, PhD in Engineering, Associate Professor, Orenburg State University, Orenburg, Russia

D. A. Gaifulina, Student, Orenburg State University, Orenburg, Russia

E. R. Khakimova, Student, Orenburg State University, Orenburg, Russia

The paper proposes a model for choosing basic functions for automated identification of time series, an algorithm and software for identifying time series based on an associative majority approach, which allow to identify the type of basic dependencies of dynamic processes.

The task of identifying the type of basic functions is solved using the classical theory of pattern recognition. Identification is performed by comparing the original image with the standards stored in a single image space. The image space is the area of device memory (associative memory) on which identification is performed. A feature of the proposed model is the efficiency of comparing the original images with the images of standards by using a single attribute space and the possibility of comparing all images in one measure. To increase the speed of identification in the algorithm for selecting basic functions, an associative majority approach to storing and searching identification data in electronic memory is also used.

The proposed algorithm and software for selecting basic functions are universal, since they allow identifying the type of basic dependence in any dynamic process, regardless of the specifics of the subject area under study. Automated identification of the type of basic functions reduces the time to build predictive models and allows you to quickly predict further variants of the process flow.

Keywords: basic dependencies, identification, time series, forecasting, dynamic processes, associative-majority approach, pattern recognition.

Получено 20.11.2018