

ПОДХОД К АНАЛИЗУ ДИНАМИЧЕСКИХ ДАННЫХ НА ОСНОВЕ СЕГМЕНТАЦИИ ГРАФИЧЕСКИХ ПРЕДСТАВЛЕНИЙ И СТРУКТУРНЫХ ПРИЗНАКОВ

Э. Г. Тунян^{1,2,a}, Р. С. Сазиков^{1,2,3,b}, Т. В. Гавриленко^{1,3,c}

¹ Сургутский государственный университет, г. Сургут, Российская Федерация

² ООО «ЕДРО», г. Сургут, Российская Федерация

³ Сургутский филиал федерального государственного автономного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Национального исследовательского центра «Курчатовский институт», г. Сургут, Российская Федерация

^a ORCID: <https://orcid.org/0009-0003-3260-1310>, ✉ tunyan@edro.su

^b ORCID: <https://orcid.org/0009-0005-0078-0013>, sazikov@edro.su

^c ORCID: <https://orcid.org/0000-0002-3243-2751>, taras.gavrilenko@gmail.com

Аннотация: представлен подход к анализу динамических данных, основанный на сегментации их графического представления и анализе структурных признаков. Исходные временные ряды преобразуются в форму графика (изображения), после чего график автоматически разбивается на сегменты по точкам, где существенно изменяются свойства его формы (наклон, кривизна, разрывы и др.). Для каждого сегмента подбирается модель аппроксимации (линейная, полиномиальная, экспоненциальная и др.), оптимально описывающая данные на этом участке. Выбор модели производится по критерию минимальной ошибки аппроксимации с учетом штрафа за сложность модели. Рассматриваются метрические характеристики, позволяющие количественно оценить разницу между соседними сегментами по их статистическим свойствам и параметрам моделей. Формально аномалия определяется как сегмент или точка, чье поведение значительно отклоняется от ожидаемого на основе моделей других сегментов. Предложенный метод позволяет выявлять как точечные аномалии, так и аномальные последовательности точек (коллективные аномалии) в динамических рядах. Приведен пример сегментации синтетического временного ряда с визуализацией обнаруженных сегментов и аномалий. Показано сравнение с классическими методами обнаружения изменений: предложенный подход дает интерпретируемое разбиение и обладает гибкостью в учете различных структурных признаков, что выгодно отличает его от сугубо статистических тестов.

Ключевые слова: сегментация временных рядов, структурные признаки, изменение режима, аппроксимация, аномалия.

Благодарности: работа выполнена в рамках государственного задания НИЦ «Курчатовский институт» — НИИСИ по теме № FNEF-2024-0001 «Создание и реализация доверенных систем искусственного интеллекта, основанных на новых математических и алгоритмических методах, моделях быстрых вычислений, реализуемых на отечественных вычислительных системах» (1023032100070-3-1.2.1).

Для цитирования: Тунян Э. Г., Сазиков Р. С., Гавриленко Т. В. Подход к анализу динамических данных на основе сегментации графических представлений и структурных признаков. *Успехи кибернетики*. 2026;7(1):123–129.

Поступила в редакцию: 02.01.2026.

В окончательном варианте: 15.01.2026.

DYNAMIC DATA ANALYSIS BASED ON SEGMENTATION OF GRAPHICAL REPRESENTATIONS AND STRUCTURAL FEATURES

E. G. Tunyan^{1,2,a}, R. S. Sazikov^{1,2,3,b}, T. V. Gavrilenko^{1,3,c}

¹ Surgut State University, Surgut, Russian Federation

² EDRO, OOO, Surgut, Russian Federation

³ Surgut Branch of Scientific Research Institute for System Analysis of the National Research Centre “Kurchatov Institute”, Surgut, Russian Federation

^a ORCID: <https://orcid.org/0009-0003-3260-1310>, ✉ tunyan@edro.su

^b ORCID: <https://orcid.org/0009-0005-0078-0013>, sazikov@edro.su

^c ORCID: <https://orcid.org/0000-0002-3243-2751>, taras.gavrilenko@gmail.com

Abstract: we studied a method for analyzing dynamic data based on segmentation of graphical representations and analysis of structural features. We first converted the original time series into a plotted

graph. The method then automatically divided the curve into segments at points where the shape changed noticeably, such as changes in slope, curvature, or small discontinuities. For each segment, we fitted an approximation model (linear, polynomial, exponential, and others) that best described the local behavior of the data. We selected the model by minimizing the approximation error while applying a penalty for excessive model complexity, which prevented overfitting.

We also applied a set of quantitative characteristics to compare neighboring segments. These characteristics describe both statistical properties of the data and the parameters of the fitted models. In this framework, an anomaly is a segment or individual point whose behavior differs substantially from the behavior predicted by models fitted to the other segments. The method, therefore, detects both isolated outliers and sequences of unusual observations, known as collective anomalies, in time-dependent data.

We demonstrated the approach using a synthetic time series and presented a visualization of the resulting segmentation and detected anomalies. A comparison with classical change-point detection methods shows that the proposed method provides interpretable segmentation and remains flexible in its treatment of different structural features, which distinguishes it from purely statistical tests.

Keywords: time series segmentation, structural features, mode change, approximation, anomaly.

Acknowledgements: this study is a part of the FNEF-2024-0001 government order contracted to the Scientific Research Institute for System Analysis of the National Research Centre “Kurchatov Institute”, project No. 1023032100070-3-1.2.1 Development and Implementation of Trusted Artificial Intelligence Systems Based on new Mathematical Methods and Algorithms, Fast Computing Models for Domestic Computing Systems.

Cite this article: Tunyan E. G., Sazikov R. S., Gavrilenko T. V. Dynamic Data Analysis Based on Segmentation of Graphical Representations and Structural Features. *Russian Journal of Cybernetics*. 2026;7(1):123–129.

Original article submitted: 02.01.2026.

Revision submitted: 15.01.2026.

Введение

При анализе временных рядов внимание, как правило, сосредотачивается не столько на отдельных значениях наблюдаемой величины, сколько на изменениях характера ее поведения во времени. Речь идет о ситуациях, когда процесс, остававшийся относительно устойчивым, начинает демонстрировать иные закономерности — не обязательно резкие, но статистически различимые. Сегментация временного ряда в этом случае выступает как средство выделения таких участков, внутри которых допустимо говорить о некоторой однородности поведения, пусть и условной. Границы между сегментами могут соответствовать сдвигам среднего уровня, изменению разброса значений или более тонким преобразованиям структуры колебаний, что делает их интерпретацию зависимой от конкретного прикладного контекста — от задач мониторинга до прогнозных и диагностических процедур.

Сегментация тесно связана с обнаружением точек изменения [1] и служит подготовительным этапом для анализа аномалий, поскольку такие точки в классических работах трактуются как моменты разладки случайного процесса [2]. Существующие алгоритмы [3] условно подразделяются на несколько классов. Подход «сверху-вниз» предполагает рекурсивное разбиение исходного ряда, начиная с его целостного представления, тогда как агломеративная стратегия «снизу-вверх» начинает с мелких сегментов и постепенно объединяет их, минимизируя рост ошибки аппроксимации. Отдельное место занимает метод скользящего окна, в котором границы сегментов формируются последовательно по мере превышения допустимого уровня ошибки. Каждый из этих подходов обладает собственными ограничениями, связанными либо с чувствительностью к ранним решениям, либо с вычислительными затратами, либо с необходимостью априорного задания порогов.

Помимо эвристических методов, применяются оптимизационные подходы, основанные на минимизации целевой функции с учетом штрафа за сложность модели. Такие методы обеспечивают глобально оптимальное разбиение, но на практике требуют специальных приемов ускорения. Обзорные исследования показывают, что эффективность алгоритмов существенно зависит от свойств анализируемых данных [4], что исключает существование универсального решения.

После сегментации временного ряда обычно переходят к выявлению аномалий [5]: наблюдений или фрагментов, не согласующихся с ожидаемым поведением процесса. Временные ряды допускают различие между точечными, групповыми и контекстуальными аномалиями, интерпретация которых определяется областью применения. Традиционные статистические критерии эффективны лишь при

простых и известных распределениях, тогда как современные методы все чаще опираются на машинное обучение для выявления нетипичных паттернов в сложных данных.

В работе используется подход к сегментации временных рядов [6], в котором ключевую роль играет анализ формы их графического представления. Такой выбор не является случайным и связан, в частности, с тем, что при рассмотрении реальных данных изменения поведения процесса далеко не всегда сводятся к сдвигам среднего уровня или изменению дисперсии. Во многих случаях динамика ряда меняется более сложным образом и эти изменения проявляются, прежде всего, в структуре траектории. Учет геометрических характеристик отдельных участков позволяет зафиксировать подобные различия, даже если они не выражаются в виде явных числовых скачков, и тем самым уточнить как характер локальных закономерностей, так и природу аномальных фрагментов.

Признаки графического представления и метод сегментации

В предлагаемом методе временной ряд на начальном этапе рассматривается через его графическое представление — зависимость значений показателя от времени. Форма такого графика интерпретируется как носитель структурной информации о поведении процесса. Переход к этому представлению позволяет описывать и использовать те особенности динамики, которые при визуальном анализе воспринимаются интуитивно, а в числовом виде проявляются не всегда явно. Сегментация строится на выявлении изменений формы графика, по которым выделяются участки со схожим характером поведения данных.

При анализе формы графика учитывается несколько признаков, отражающих различные аспекты поведения временного ряда. Одним из базовых является наклон, который соответствует первой производной ряда и позволяет судить о локальном направлении изменения значений: росте, спаде либо близком к горизонтальному ходу. Существенную роль играет и изменение наклона, то есть кривизна графика: по ней можно выявлять точки перегиба и отличать участки с почти линейным характером от сегментов, где динамика становится более сложной.

Отдельно рассматриваются резкие изменения уровня между соседними наблюдениями, включая скачки и разрывы, в том числе связанные с пропусками данных. Подобные ситуации часто сопровождают смену режимов и на практике нередко совпадают с границами сегментов. В отдельных ситуациях обращают внимание на то, как значения ряда ведут себя во времени: иногда наблюдаются колебания, при которых через сопоставимые промежутки данные вновь оказываются вблизи прежних уровней. Такой характер изменения отличает колебательные участки от фрагментов с устойчивым однонаправленным трендом. Кроме того, при анализе некоторых сегментов может учитываться форма их траектории в целом, например, насколько схожи по очертаниям участки роста и последующего снижения. Подобные соображения используются не как основной критерий, а, скорее, как вспомогательное описание структуры сегмента.

В ходе сегментации указанные признаки вычисляются вдоль временной оси, после чего выделяются моменты, в которых их поведение изменяется наиболее существенно. Эти моменты рассматриваются как возможные границы сегментов.

В основе метода лежит использование простых пороговых правил для отдельных признаков. Так, в точках, где изменение наклона временного ряда становится достаточно большим ($|\Delta \text{slope}|$ превышает T_{slope}), возможно проведение границы сегмента. Аналогичное решение принимается и при резком изменении значений: если разность между соседними отсчетами Δx оказывается больше порога T_{jump} , считается, что поведение ряда меняется, и с этого момента начинается новый сегмент.

Полученные таким образом точки не используются напрямую. На следующем этапе они уточняются с учетом взаимного расположения: слишком близкие точки могут отбрасываться или объединяться, например при нарушении требования к минимальной длине сегмента. После этой процедуры формируется итоговый набор границ, по которым временной ряд разбивается на сегменты S_i .

По своей логике данный подход близок к классическим методам обнаружения разладок, однако не ограничивается проверкой изменений среднего уровня или дисперсии, как это делается в традиционных тестах, включая CUSUM. Вместо этого используется расширенный набор признаков, чувствительных к изменениям формы сигнала. Аналогичная идея прослеживается и в более поздних работах: так, в методе RECURVE границы сегментов определяются на основе анализа кривизны траектории в пространстве признаков [7]. Показано, что учет кривизны позволяет уверенно выявлять как резкие,

так и постепенные смены режима, тогда как анализ простых разностей между соседними точками оказывается менее устойчивым.

Аппроксимация сегментов и выбор моделей

После сегментации временного ряда поведение данных на каждом участке описывается с помощью простой интерпретируемой модели. Такое описание используется не только для аппроксимации значений внутри сегмента, но и для последующего сравнения сегментов между собой, в том числе с целью выявления нетипичных участков.

В качестве базового варианта применяется кусочно-линейная аппроксимация [8]. В зависимости от характера сегмента могут использоваться и другие модели, включая полиномиальные, экспоненциальные, гармонические и стационарные модели шумовых процессов.

Оценка качества аппроксимации сегмента выполняется по сумме квадратов отклонений:

$$E_m = \sum_{t \in S_i} (x(t) - \hat{f}_m(t))^2, \quad (1)$$

где S_i — i -й сегмент временного ряда;

$x(t)$ — значение временного ряда в момент времени t ;

$\hat{f}_m(t)$ — аппроксимирующая функция модели типа m ;

E_m — ошибка аппроксимации сегмента моделью m .

Для учета сложности модели используется информационный критерий Акаике:

$$AIC_m = N \cdot \ln(E_m/N) + 2 \cdot k_m, \quad (2)$$

где N — число точек в сегменте;

k_m — число параметров модели типа m .

Критерий (2) используется для выбора модели с учетом как точности аппроксимации, так и ее параметрической сложности.

После выбора модели вычисляются количественные характеристики сегмента. Среднее значение сегмента определяется как:

$$\bar{x}_i = (1/|S_i|) \cdot \sum \cdot x(t), \quad (3)$$

где $|S_i|$ — число точек в сегменте.

Для сравнения соседних сегментов используется метрика различия:

$$D(S_i, S_{i+1}) = \omega_1 \cdot |\bar{x}_i - \bar{x}_{i+1}| + \omega_2 \cdot |a_i - a_{i+1}| + \omega_3 \cdot |\sigma_i \cdot \sigma_{i+1}| + \dots, \quad (4)$$

где a_i — параметр тренда сегмента;

σ_i — показатель разброса значений;

$\omega_1, \omega_2, \omega_3$ — весовые коэффициенты признаков.

Если значение метрики (4) превышает заданный порог, переход между сегментами трактуется как существенная смена режима.

Обнаружение и определение аномалий

После разбиения временного ряда на сегменты и построения аппроксимирующих моделей возникает задача выявления аномальных фрагментов. В рамках рассматриваемого подхода к аномалиям могут относиться как отдельные наблюдения, так и целые сегменты, поведение которых заметно отличается от основной структуры ряда.

Под аномалией в дальнейшем понимается такой фрагмент данных (точка или сегмент), для которого либо не удастся подобрать модель из принятого набора с приемлемым качеством аппроксимации, либо его параметры существенно отличаются от характеристик остальных участков ряда. Иными словами, речь идет о наблюдениях, не согласующихся с типичным режимом функционирования процесса.

На практике используются несколько простых критериев. Если для сегмента S_k минимальная ошибка аппроксимации оказывается значительно выше, чем для большинства других сегментов, либо превышает заранее заданный допустимый уровень, такой сегмент рассматривается как аномальный. Аналогичное решение может приниматься и в том случае, когда параметры модели или статистические

характеристики сегмента (\bar{x}_k , a_k , σ_k и др.) выходят за диапазон значений, характерный для остальных сегментов. Например, резкий тренд на фоне в целом стабильных участков может служить признаком аномального поведения.

Иногда аномалии бывают не только в целом сегменте, но и в отдельных точках. Если одна точка сильно выбивается из общей картины, скажем, отличается куда больше, чем обычная погрешность, — это может быть просто шум, ошибка измерения или разовый сбой.

Важно учитывать, что обнаружение аномалий тесно связано с процедурой сегментации. Длительные аномальные фрагменты, как правило, выделяются в отдельные сегменты за счет своей неоднородности. Если же аномалия носит локальный характер, она может оставаться внутри сегмента и выявляться только при анализе отдельных точек относительно модели. С этой точки зрения целесообразен двухуровневый подход, при котором сначала выявляются аномальные сегменты, а затем внутри них проводится более детальный анализ.

Отметим, что такое понимание аномалии опирается на наличие модели нормального поведения процесса. В данном случае ее роль играет совокупность моделей сегментов, описывающих основные режимы работы системы; все, что существенно отклоняется от этих моделей, рассматривается как потенциально аномальное.

Пример сегментации и выявления аномалий

Для иллюстрации работы метода рассмотрен упрощенный пример на искусственно сгенерированном временном ряде. В ряду последовательно реализованы несколько режимов: начальный линейный рост, затем колебания около постоянного уровня и далее убывающий тренд. Дополнительно в данные введены отдельные выбросы. На рисунке показан исходный временной ряд и результат его автоматической сегментации, первый сегмент соответствует участку с линейным ростом, второй — колебаниям около среднего уровня, третий — нисходящему тренду. Границы сегментов отмечены пунктирными линиями, а обнаруженные аномальные точки — отдельными маркерами.

В данном примере алгоритм корректно выделяет границы сегментов в моментах смены характера динамики. Первый сегмент (примерно до $t \approx 40$) описывается линейной моделью с положительным наклоном. Второй сегмент ($t \approx 40 - 80$) характеризуется колебательным поведением около постоянного уровня и аппроксимируется гармонической моделью. Третий сегмент (после $t \approx 80$) соответствует убывающему тренду и также хорошо описывается линейной зависимостью с отрицательным наклоном.

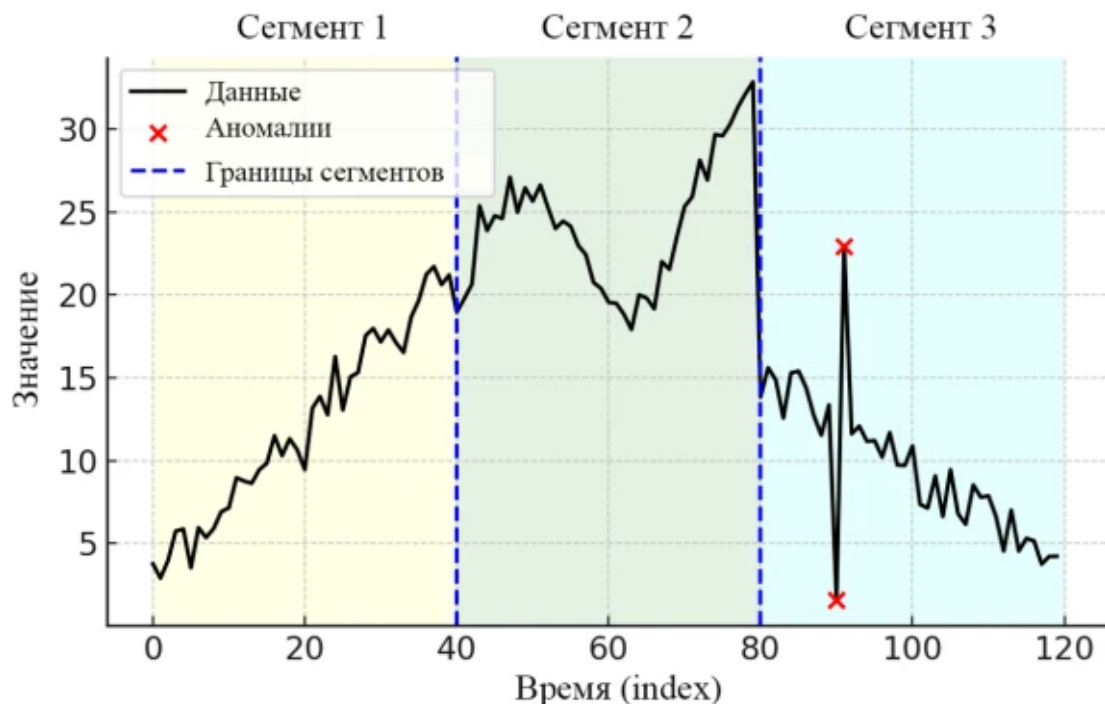


Рис. Разбиение ряда на три сегмента

В области третьего сегмента наблюдаются два выраженных выброса. Несмотря на то, что они расположены внутри сегмента, их отклонение от аппроксимирующей модели существенно превышает типичный уровень ошибки, вследствие чего эти точки помечаются как точечные аномалии. Если бы аналогичное отклонение сохранялось на протяжении более длительного интервала, такой участок был бы выделен в отдельный сегмент и рассматривался уже как сегментная аномалия.

Пример демонстрирует, что результаты сегментации можно интерпретировать содержательно: выделенные сегменты отражают смену режимов, а аномалии проявляются либо как отдельные участки, либо как точечные отклонения. Такой результат удобен для анализа и сопоставления с реальными процессами.

CUSUM и тесты на структурные сдвиги имеют четкую математическую базу, но на практике чувствительны, в основном, к резким «переломам» и требуют предположений о распределении данных. Если изменение носит плавный характер и проявляется, прежде всего, в форме сигнала, такие методы оказываются менее информативными.

Машинное обучение охватывает более сложные паттерны, однако после обучения выдает бинарный сигнал «есть/нет аномалия» без указания, какая именно характеристика процесса изменилась.

Предлагаемый подход состоит в том, чтобы разбить ряд на сегменты и сравнивать их параметры: наклон, дисперсию, спектральные веса и т. д. Это позволяет не только обнаружить момент изменения, но и сразу указать, в чем оно выразилось: появилась новая тенденция, увеличилась дисперсия или сместилась доминирующая частота. Метод остается интерпретируемым, не требует параметрических предположений и при необходимости легко комбинируется с более сложными моделями.

Заключение

В рамках работы анализ временного ряда строится вокруг его разбиения на участки, внутри которых поведение данных остается относительно однородным. Каждый из таких участков рассматривается отдельно и описывается своей моделью, что позволяет учитывать изменения формы сигнала, не сводящиеся к простым сдвигам отдельных статистических характеристик.

В рамках подхода данные сначала переводятся в графическое представление, после чего вычисляются признаки, описывающие форму графика, такие как наклон, кривизна, наличие разрывов, периодичность и другие характеристики. Далее временной ряд автоматически разбивается на сегменты в тех точках, где наблюдаются существенные изменения этих признаков. Для каждого сегмента из заранее заданного набора подбирается модель, наилучшим образом описывающая его поведение, с учетом не только ошибки аппроксимации, но и сложности модели. После этого вычисляются количественные характеристики сегментов, что позволяет проводить их сопоставление между собой. Аномалии при таком подходе трактуются как сегменты или отдельные наблюдения, которые не удается удовлетворительно описать ни одной из используемых моделей либо которые заметно отличаются по своим свойствам от основной части ряда.

Проведенный анализ показывает, что использование геометрических признаков формы графика позволяет более гибко выявлять смену режимов работы системы по сравнению с традиционными статистическими тестами, ориентированными, как правило, на изменение одного параметра. Применение нескольких типов моделей аппроксимации дает возможность выделять различные типы сегментов: трендовые, колебательные, стационарные — и сразу получать наглядное описание их поведения. При этом определение аномалий через отклонение от сегментной модели делает результаты интерпретируемыми: в каждом случае можно указать, в чем именно заключается необычность участка, будь то рост разброса значений, изменение характера колебаний или невозможность адекватной аппроксимации стандартными функциями.

Работа метода проиллюстрирована на синтетическом временном ряде, для которого корректно выявлены границы смены режимов и обнаружены точечные выбросы. По сравнению с традиционными тестами и «черными ящиками» машинного обучения получившаяся процедура оказывается чем-то средним: она все еще говорит на понятном языке «наклон уменьшился / дисперсия выросла», но при этом видит не только скачки среднего, но и более тонкие перестроения профиля кривой.

Следующий шаг — проверка на живых данных: технологические датчики, биржевые котировки, ЭКГ и ЭЭГ. Понадобится подобрать пороги и веса под каждую предметную область и посмотреть, насколько хорошо сегментные признаки дружат с обучаемыми моделями: удастся ли усилить чувствительность, не потеряв ясность объяснений.

Первые прогоны показывают, что разбивка ряда по форме сигнала превращается в удобный инструмент: быстро становится понятно, где именно поменялась динамика процесса, и можно сразу же принимать решение. В частности, она облегчает интерпретацию аномалий, поскольку позволяет рассматривать их в контексте поведения соседних участков ряда.

ЛИТЕРАТУРА

1. Killick R., Fearnhead P., Eckley I. A. Optimal Detection of Changepoints with a Linear Computational Cost. *Journal of the American Statistical Association*. 2012;107(500):1590–1598. DOI: 10.1080/01621459.2012.737745.
2. Воробейчиков С. Э., Конев В. В. Последовательный метод обнаружения разладок случайных процессов рекуррентного типа. *Автоматика и телемеханика*. 1984;5:27–38.
3. Truong C., Oudre L., Vayatis N. Selective Review of Offline Change Point Detection Methods. *Signal Processing*. 2020;167:107299. DOI: 10.1016/j.sigpro.2019.107299.
4. Fu T.-C. A Review on Time Series Data Mining. *Engineering Applications of Artificial Intelligence*. 2011;24(1):164–181. DOI: 10.1016/j.engappai.2010.09.007.
5. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey. *ACM Computing Surveys*. 2009;41(3):15. DOI: 10.1145/1541880.1541882.
6. Lovrić M., Milanović M., Stamenković M. Algorithmic Methods for Segmentation of Time Series: An Overview. *Journal of Contemporary Economic and Business Issues*. 2014;1(1):31–53.
7. Shin Y., Park J., Song H. et al. Exploiting Representation Curvature for Boundary Detection in Time Series. *Advances in Neural Information Processing Systems (NeurIPS 2024)*. 2024.
8. Keogh E., Chu S., Hart D., Pazzani M. Segmenting Time Series: A Survey and Novel Approach. *Data Mining in Time Series Databases*. Singapore: World Scientific; 2004:1–21.