НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
сентябрь–октябрь 2026 Том 26 № 5 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
September–October 2026 Vol. 26 No 5 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2026-26-5-1026-1038
УДК 004.8:519.246

Автоматическое формирование объектов наблюдения в задаче оценки состояния системы по временным рядам

Илья Сергеевич Лебедев[1], Александр Алексеевич Макеенко[2]✉

[1,2] Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация

[1] isl_box@mail.ru, https://orcid.org/0000-0001-6753-2181
[2] aleks.mackeencko@yandex.ru✉, https://orcid.org/0009-0002-4453-9887

Аннотация

Введение. Повышение значений показателей качества классификации обуславливает необходимость совершенствования как моделей обработки данных, так и сопутствующих процессов формирования выборок данных и обучения. Реализация методов обработки информационных последовательностей зависит от представления данных и требует правильного выбора модели и настройки ее гиперпараметров. Существенное влияние на итоговый результат оказывает способ преобразования исходных временных рядов в объекты наблюдения, поскольку именно он определяет свойства обучающей выборки и условия последующего применения алгоритмов классификации. В этой связи задача формирования объектов наблюдения может рассматриваться как самостоятельный этап подготовки данных, влияющий на качество оценки состояния системы. **Метод.** Рассматривается задача автоматического формирования объектов наблюдения в задаче оценки состояния системы по временным рядам и информационным последовательностям. В отличие от классических подходов, где модель настраивается на данные, предлагаемый метод основан на выборе параметров данных, когда из временных рядов и информационных последовательностей формируются выборки объектов наблюдения, таким образом, чтобы модель достигала на них лучших значений заранее выбранного показателя качества. **Основные результаты.** Проведен эксперимент на модельных данных и на выборках наборов данных. Выборки объектов наблюдения, обладающих различными свойствами, формировались в результате изменения ширины окна и сдвига. Модель обучалась на созданных выборках и определялись характеристики объектов и выборки, где достигались лучшие результаты. Получены значения качественных показателей. Выполненный анализ показателей точности (Precision) для различных алгоритмов показал, что формирование объектов из последовательностей с определенными свойствами длины и сдвига позволяет повысить ее значения по сравнению со средними значениями по выборке на 3–8 %. **Обсуждение.** Предложенное решение может быть интегрировано в разработку новых методов обработки данных и использоваться как самостоятельный этап настройки моделей. Формирование выборок данных из объектов с разными свойствами позволяет использовать предложенное решение в процессах обучения моделей.

Ключевые слова

объекты наблюдения, временные ряды, состояние системы, классификация состояний, информационные последовательности, формирование выборки, адаптивная обработка, показатели качества

Благодарности

Исследование выполнено за счет гранта Российского научного фонда № 25-21-00269, https://rscf.ru/project/25-21-00269/.

Ссылка для цитирования: Лебедев И.С., Макеенко А.А. Автоматическое формирование объектов наблюдения в задаче оценки состояния системы по временным рядам // Научно-технический вестник информационных технологий, механики и оптики. 2026. Т. 26, № 5. С. 1026–1038. doi: 10.17586/2226-1494-2026-26-5-1026-1038

© Лебедев И.С., Макеенко А.А., 2026

Automatic formation of observation objects for system state assessment from time series

Ilya S. Lebedev[1], Alexander A. Makeenko[2]✉

[1,2] St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), Saint Petersburg, 199178, Russian Federation

[1] isl_box@mail.ru, https://orcid.org/0000-0001-6753-2181

[2] aleks.mackeencko@yandex.ru✉, https://orcid.org/0009-0002-4453-9887

Abstract

Improving classification quality metrics requires advances not only in data processing models but also in the related processes of dataset formation and training. The implementation of information sequence processing methods depends on data representation and requires an appropriate choice of model and tuning of its hyperparameters. The final result is significantly influenced by the way the original time series are transformed into observation objects, since this transformation determines the properties of the training dataset and the conditions for the subsequent application of classification algorithms. Therefore, the problem of observation object formation can be considered as an independent stage of data preparation affecting the quality of system state assessment. This paper considers the automatic formation of observation objects for system state assessment based on time series and information sequences. Unlike classical approaches, where the model is adapted to the data, the proposed method is based on solving the inverse problem, in which datasets of observation objects are formed from time series and information sequences so that a predefined model achieves the best value of a selected quality metric. Experiments were carried out using both model data and benchmark datasets. By varying the window width and shift, datasets of observation objects with different properties were generated from the time series. The models were trained on these datasets, and the characteristics of the observation objects and datasets yielding the best results were identified. Quality metrics were obtained and analysed. The results show that forming observation objects from sequences with specific window length and shift properties allows improving precision by 3–8 % compared with the average values obtained over the corresponding datasets. The proposed solution can be integrated into the development of new data processing methods and used to enhance existing models. The formation of datasets from observation objects with different properties extends the possibilities of model training and provides an additional mechanism for improving classification quality.

Keywords

observation objects, time series, system state, state classification, information sequences, dataset formation, adaptive processing, quality metrics

Acknowledgements

This research is supported by RSF (grant No. 25-21-00269, https://rscf.ru/project/25-21-00269/).

For citation: Lebedev I.S., Makeenko A.A. Automatic formation of observation objects for system state assessment from time series. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2026, vol. 26, no. 5, pp. 1026–1038 (in Russian). doi: 10.17586/2226-1494-2026-26-5-1026-1038

Введение

Совершенствование процессов хранения и обработки разнородной информации, фиксирующей поведение различных информационных и физических процессов, протекающих в технологических системах, обеспечивает возможность масштабного сбора измерительной информации в динамике. В результате накопления таких данных формируются выборки, представляющие из себя структурированные информационные последовательности и временные ряды. Анализ временных рядов приобретает все более значимую роль в исследовании разнообразных процессов. Однако на получение данных могут влиять различные факторы, связанные с внутренней сложностью изучаемых процессов, несовершенством измерительных систем, потенциальным воздействием вредоносного программного обеспечения. Совокупность указанных факторов нередко приводит к возникновению аномалий. Надежная оценка состояний критически важна для ряда областей, где цена ошибки особенно высока, например, в здравоохранении, финансах и при обнаружении промышленных неисправностей [1, 2]. В таких задачах основная цель заключается в выявлении образцов, которые отклоняются от «нормальной» структуры данных, сформированной на этапе обучения модели [3]. Однако имеется ряд проблемных вопросов, которые приходится учитывать при реализации. В большинстве случаев предлагаемые решения основываются на моделях, обладающих большой вычислительной сложностью и ресурсоемкостью, и могут использовать подключение к внешним источникам, что затрудняет их применение на относительно маломощных автономных вычислительных устройствах [4]. Идентификация события изменения состояния в информационной последовательности не всегда является однозначной и может по-разному трактоваться даже экспертами. Накопленные большие объемы последовательностей на сегодняшний день делают малоэффективной ручную разметку и требуют автоматизации этого процесса.

В связи с этим возникает необходимость разработки эффективных моделей и методов оценки состояния системы по временным рядам, осуществляющих определение текущего состояния на основе анализа последовательности измерений ее параметров, упорядоченных во времени.

Обзор методов оценки состояния по временным рядам

Решение обозначенной задачи позволяет предотвращать аварии, оптимизировать обслуживание оборудования, обеспечивать безопасность в критически важных системах, осуществлять поддержку принятия решений. Оно выполняется различными методами, основанными на статистических подходах, использующих машинное обучение, экспертные оценки. В современных исследованиях, посвященных оценке состояния систем по временным рядам, предложено значительное количество моделей, базирующихся на методах глубокого обучения. Наибольшее распространение получили нейросетевые архитектуры, ориентированные на анализ скрытых зависимостей и извлечение многомасштабных признаков [5]. Однако подобные методы обладают рядом ограничений, что обусловливает необходимость их дальнейшего развития.

Одной из основных групп моделей в задачах оценки состояния являются автокодировщики [6]. Их современное использование предполагает наличие глубоких архитектур для анализа, в которых состояния определяются на основе величины ошибки восстановления входных данных. Дальнейшее развитие данного подхода привело к появлению вариационных, глубоких сверточных сетей с механизмами внимания [7]. Вместе с тем они подвержены переобучению, имеют чувствительность к гиперпараметрам, что негативно сказывается на их способности к обобщению, обладают вычислительной сложностью.

С целью повышения устойчивости и качества оценки состояния были предложены интегрированные архитектуры, объединяющие автокодировщики с другими нейросетевыми подходами. В частности, в [8] разработаны состязательные сети с использованием модулей памяти, которые усиливают различия между нормальными и аномальными образцами за счет увеличения расхождений в ошибках реконструкции. Кроме того, были предложены архитектуры на основе трансформеров, извлекающие многомасштабные признаки путем изменения пространственного разрешения. Основным недостатком таких методов остается ориентация на ошибки прогнозирования, что может приводить к пропуску ряда событий, а также требования к вычислительным средствам [9]. Применение графовых нейронных сетей с механизмами внимания направлено на моделирование взаимосвязи между признаками, что позволяет улучшить качество оценки состояния. Однако подобные подходы, как правило, обладают сложностью отладки, требуют значительных затрат на вычисление коэффициентов внимания, сосредоточены на анализе временных характеристик и не учитывают информацию из частотной области [10]. Для преодоления этого недостатка разработаны методы, основанные на частотном анализе. Вместе с тем ограниченная способность этих подходов к моделированию взаимодействий между временными и частотными характеристиками может приводить к возникновению ложных отрицательных результатов [11].

Отдельного внимания заслуживают подходы, объединяющие генеративно-состязательные сети и трансформеры. Тем не менее такие модели, как правило, не учитывают корреляции между признаками [12]. Модели на базе архитектуры Transformer демонстрируют высокую эффективность в задачах оценки состояния. Они показали превосходство над многими существующими методами и часто используются в качестве базовых решений для сравнительного анализа [13]. Однако сравнение современных подходов выявляет ряд общих недостатков. К ним относятся ограниченные возможности по учету глобального контекста и долгосрочных зависимостей, высокая вычислительная сложность, требующая значительных ресурсов и времени обучения, а также чувствительность к шуму, приводящая к ложным срабатываниям при работе с зашумленными данными [14].

В отличие от рассмотренных подходов, в настоящей работе внимание сосредоточено не на усложнении архитектуры модели, а на автоматическом формировании объектов наблюдения и выборок данных, обладающих разными свойствами, из временных рядов. Особенностью подхода является выбор гиперпараметров оконного преобразования и анализ достижения лучших значений показателя качества на сформированных данных. В рассматриваемом методе основной акцент делается не на настройке модели на выборке данных, а на формировании и выборе с помощью перебора параметров оконного преобразования (длина окна, сдвиг) объектов наблюдения выборки. В результате становиться возможным определять свойства объектов наблюдения, на которых достигаются лучшие значения выбранного показателя качества модели обработки с предопределенными гиперпараметрами.

Предлагаемый метод предназначен для офлайн-предобработки временных рядов перед обучением модели, его применение в потоковом режиме в настоящей работе не рассматривается и требует отдельного исследования.

Постановка задачи

В ряде областей цена ошибки определения состояния системы может оказаться недопустимо высокой. В связи с этим возникает необходимость решения задачи классификации состояния, отнесения текущего состояния к одному из заранее определенных классов на основе ее наблюдаемых параметров и характеристик. Для этого необходима разработка методов адаптивного обучения, направленных на обнаружение, идентификацию состояния системы, позволяющих решать подобные задачи при ограниченных вычислительных ресурсах.

Реализация этих методов сильно зависит от представления данных. Для их успешного применения необходим правильный выбор модели и настройка ее гиперпараметров. В рассматриваемой задаче классификации состояния по временному ряду в дополнении к гиперпараметрам модели возможно изменение параметров настройки данных. Изменяя ширину окна из временного ряда, можно формировать различные выборки, содержащие объекты разного размера. Модель обучается на них и, в зависимости от достигаемого

показателя качества, происходит выбор модели и выборки, где достигнуты лучшие результаты.

В предлагаемом подходе автоматическая разметка строится на принципе разбиения последовательностей на отдельные патчи. Формальная постановка задачи представлена следующим образом.

X — обучающая выборка временных рядов, где последовательности получены в разных известных состояниях $\{c_1, \ldots, c_k\} \in C$. Временной ряд содержит значения $X^{cj} = (x_1^{cj}, \ldots, x_t^{cj}, \ldots, x_T^{cj})$, $x_t^{cj} \in R$, $t = 1, \ldots, T$, где T — длина ряда. $\varphi(l, s)$ — способ формирования патчей, использующий длину l и сдвиг s. Временной ряд X^{cj} разбивается на патчи $x_i^{\varphi(l,s)}$, где каждому патчу ставится в соответствие метка класса $x_i^{\varphi(l,s),cj} = (x_i^{cj}, x_{i+1}^{cj}, \ldots, x_{i+l-1}^{cj})$, $i = 1 + s, \ldots, T - l + 1 + s$, где $1 \le l \le T$ и $1 \le s \le l$. В результате для ряда X^{cj} получается множество патчей $X^{\varphi(l,s),cj} = \{x_i^{\varphi(l,s),cj}\ i | i = 1, \ldots, T - l + 1\}$, содержащих метку класса c_j. Разбивая на патчи последовательности $\{X^{c_1}, \ldots, X^{c_k}\} \in X$, получаем размеченную обучающую выборку $\{X^{\varphi(l,s),c_1}, \ldots, X^{\varphi(l,s),c_k}\} \in X^{\varphi(l,s)}$, где каждый патч представляется вектором признаков $x_i^{\varphi(l,s),cj} \in R^d$.

Обучающая выборка подается на вход заранее определенных моделей $\{a_1, \ldots, a_N\} \in A$, где $a_i(x_i^{\varphi(l,s)})$ — модель обработки поступающих патчей временного ряда. Рассматривая в дальнейшем решение задачи «обучения с учителем», становится возможным определить модели и параметры, при которых функционал качества достигает максимума.

$$Q(a_k(x_j^{\varphi(l,s)}), X^{\varphi(l,s)}) \to \max_{a_k \in A, l, s}. \qquad (1)$$

Определим $\theta = \varphi(l, s)$, $a_k \in A$, где l, s — параметры преобразования. Тогда выражая (1) через обобщенный параметр θ определим исследуемый функционал качества $Q(\theta) \to \max_{\theta \in \Phi}$, где множество $\Phi = \{\varphi(l, s) | a_k \in A, l, s\}$ задает область поиска оптимального решения. Параметры $\{\theta_1, \theta_2, \ldots, \theta_n\} \in \Theta$ принимают только дискретные значения и ограничиваются вычислительными ресурсами. В этом случае допустимое множество — конечное счетное множество изолированных точек. Если множество допустимых наборов параметров конечно, то гарантировано существование решения, где максимум функционала достигается в одной из этих точек и определяется перебором всех вариантов.

Единственность решения может не выполняться, при этом множество оптимальных решений $\Theta_{\text{opt}} = \{\theta \in \Theta | Q(\theta) = \max_{\theta'} Q(\theta')\}$, содержит более одного элемента. Перебор с последующим анализом допускает достижение схожих значений качества на нескольких точках длины и сдвига. Это зависит от включения в состав объекта наблюдения «значащих» паттернов и шумовых компонент, которые оказывают влияние на достигаемые показатели качества. Нахождение единственного решения может быть основано на правилах приоритета, например минимального размера окна.

Предлагаемый метод ориентирован на автоматическую генерацию обучающих выборок. Патч длины l рассматривается как объект наблюдения. Состояние, в котором получена информационная последовательность, определяет метку класса для всех объектов, сформированных из нее. Каждая модель обучается по выборке объектов наблюдения, полученных из подпоследовательностей в разных режимах работы. Определяются значения функционала качества. Затем изменяются параметры длины патча, сдвига и происходит очередное формирование объектов наблюдения и выборки данных, повторяются процессы обучения и вычисления значений функционала качества. Таким образом, формируя объекты наблюдения и определяя функционал качества возможно определить свойства патчей (длины и сдвига), где модель достигает лучших значений показателя качества.

Формирование объектов наблюдения является работой с входными данными, направленной на создание и определение обучающих примеров со свойствами, позволяющими модели лучше улавливать закономерности. Особенность предлагаемого метода — перебор параметров данных при формировании объектов, что может использоваться как один из этапов для настройки модели.

Предложенный метод может быть классифицирован как частный случай метода Grid Search (поиск по сетке) для данных, если изначально ограничиваться только параметрами, которые можно извлечь из временных рядов и информационных последовательностей. Формирования объектов наблюдения локализует пространство гиперпараметров данными, что позволяет уменьшить вычислительную сложность Grid Search при исследовании каждой возможной комбинации. Кроме метода Grid Search, возможно применение метода оптимизации Random Search (случайного поиска) для определения параметров настройки данных. Random Search более предпочтителен в сценариях с большим пространством гиперпараметров, где нецелесообразно проверять каждую возможную комбинацию. Отметим, что в рассматриваемом случае при использовании метода Random Search необходимо определять дополнительные стратегии выбора, чтобы не пропустить значимые значения.

При полном переборе для N моделей и максимальной длины окна $L_{\max}$, когда для каждой длины l рассматривается $n_s(l)$ вариантов сдвига s, число процедур формирования выборки, обучения и оценки качества составляет $N \sum_{l=1}^{l_{\max}} l n_s(l)$. В рассматриваемом случае $n_s(l) = l$, поэтому число процедур составляет $N \sum_{l=1}^{l_{\max}} l = O(NL_{\max}^2)$. Для сокращения перебора может использоваться метод Random Search: ранняя остановка при отсутствии улучшения качества на валидационной выборке и сужение диапазона длин окна на основе априорных данных. Один из возможных способов определения остановки описан в [15], где проводится анализ свойств выборки.

Эксперимент

В рамках эксперимента исследовалась возможность автоматического формирования выборок данных из информационных последовательностей для оценки со-

стояния. Рассматриваемые значения, иллюстрирующие обработку данных предлагаемым методом выбраны из набора данных [16]. Основная цель выборки заключалась в анализе влияния способов формирования и получаемых для них объектов наблюдения на качественные показатели моделей. Самоорганизующаяся модель получала на вход временные ряды, сформированные для различных состояний. Пример таких рядов представлен на рис. 1. Из них изменением длины и сдвига формировались выборки объектов наблюдения, которые подавались на вход моделям обучения.

В ходе эксперимента при оценке качества моделей учитывались ресурсные ограничения. В результате были выбраны алгоритмы с высокой скоростью обучения и небольшими требованиями к вычислительным ресурсам. Проводилось обучение и оценка качественных показателей алгоритмов наивного Байеса (Naive Bayes, NB), логистической регрессии (Logistic Regression, LR) и метода опорных векторов (Support Vector Machine, SVM), градиентного бустинга (Extreme Gradient Boosting, XGBoost), рекуррентной нейронной сети (Recurrent Neural Network, RNN) и нейронной сети с долговременной краткосрочной памятью (Long Short-Term Memory, LSTM). Гиперпараметры всех моделей использовались по умолчанию в среде MATLAB. Патчи информационных последовательностей формировались посредством разбиения исходных данных на неперекрывающиеся окна. При изменении длины l и сдвига s происходило деление последовательности на дискретные блоки, в результате чего создавались выборки объектов различной размерности. Для каждой заданной длины l выполнялся последовательный сдвиг на один отсчет, в результате чего формировалось $n_s(l)$ обучающих выборок, где $n_s(l)$ — количество рассматриваемых вариантов сдвига для патча длины.

Свойства объектов внутри этих выборок могут различаться, вследствие чего обучение одного и того же алгоритма на выборках одинаковой размерности, но полученных при разных сдвигах, дает неодинаковые значения показателя качества. На рис. 2 приведены кривые Receiver Operating Characteristic (ROC) и значения Area Under the Curve (AUC) для двух выборок, полученных сдвигом.

Полученные значения на рис. 2 показывают влияние сдвига на формирование выборки данных. Варьирование сдвига может как увеличивать количество шумовых примеров на границах классов, так и уменьшать его.

Оценка эффективности предлагаемого метода проводилась на задаче бинарной классификации. Сформированные выборки разделялись на обучающую и тестовую части в соотношения 90:10. Обучающая часть выборки подвергалась кросс-валидации (K Fold Cross Validation с K = 10), тестовая часть использовалась только для проверки. Длина объектов наблюдения, представленных подпоследовательностями, варьировалась в диапазоне от 2 до 17 отсчетов, а количество сдвигов для каждой длины соответствовало самой этой длине. На рис. 3 показаны значения Precision для классифицирующих алгоритмов.

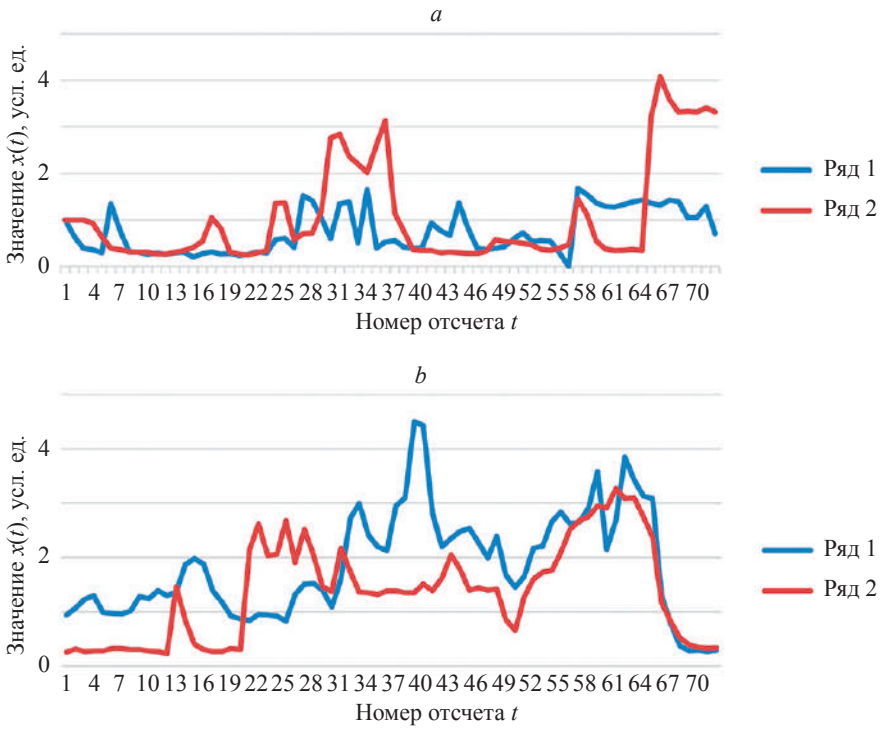


Рис. 1. Последовательности состояний 1 (*a*) и 2 (*b*).
$x(t)$ — значение временного ряда при дискретном отсчете t

Fig. 1. Sequences for states 1 (*a*) and 2 (*b*)
$x(t)$ is the time series value at discrete sample t

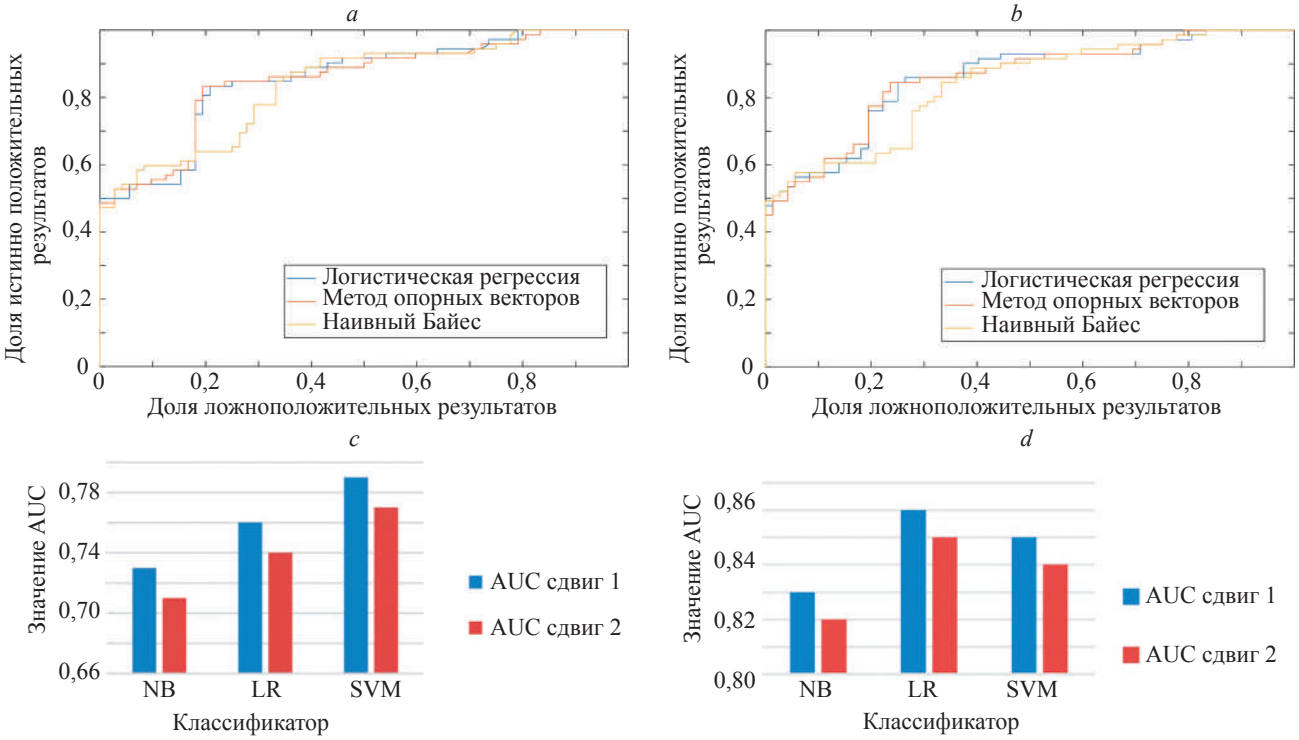


Рис. 2. ROC-кривые и значения AUC для первой (*a*) и (*c*), для второй (*b*) и (*d*) выборок

Fig. 2. ROC curves and AUC values for two samples: ROC curves for the first sample (*a*); ROC curves for the second sample (*b*); AUC values for the first sample (*c*); AUC values for the second sample (*d*)

На рис. 3 видно, что длина последовательности оказывает влияние на обобщающую способность моделей обработки данных. При малой длине патча модель не видит информативных признаков, при большой длине происходит включение нерелевантной информации (шума), рост размерности признакового пространства, увеличение вычислительных затрат.

Для оценки устойчивости к шуму было применено зашумление данных. Для каждой выборки поочередно зашумлялись определенные столбцы вектора данных. Результаты достигаемых значений Precision по длинам с зашумлением отдельных элементов и без представлены на рис. 4.

Анализ значений метрики Precision на рис. 4 при разных длинах окна и в условиях добавления шума демонстрирует, что наиболее выраженное влияние наблюдается при длине окна 4, что свидетельствует о том, что в соответствующих позициях содержатся информативные признаки. При этом воздействие шума при длине 4 оказывается более существенным по сравнению с длинами 8, 12 и 16. Такая динамика обусловлена вовлечением в анализ дополнительных элементов при увеличении длины. Степень ухудшения показателей зависит от того, какой компонент зашумляется. При поражении «значащего» элемента паттерна снижение значений более выражены, при воздействии на менее информативные компоненты эффект оказывается менее заметным.

Сходимость выборочного максимума определялось изменением размера обучающей части выборки данных. Графики рис. 5 экспериментов для различных длин и сдвигов показывает повторение общих трендов.

На рис. 5 видно, что изменение размера обучающей выборки данных в большей степени влияют на более слабые классификаторы. Объем рассматриваемых данных выборки и свойства позволяют сильным моделям быстро схватывать закономерности. Однако при изменении длины последовательности для нейросетевых моделей приходится изменять параметры размерности входного слоя, что влечет дополнительные расходы.

Использование сдвига дает возможность для анализа детального покрытия всей последовательности, что позволяет модели анализировать данные с высокой плотностью охвата. Модель может обнаруживать короткие по длительности события, которые могли бы остаться незамеченными при большем размере сдвига, лучше распознавать быстрые изменения в информационных последовательностях.

Для оценки статистической надежности полученных результатов использовался доверительный интервал Уилсона. Такой подход позволяет избежать систематического занижения ширины доверительного интервала и обеспечивает более достоверную оценку неопределенности экспериментальных результатов.

Доверительный интервал (CI) для показателя Precision рассчитывался по формуле:

$$CI = \frac{p + z^2/2n \pm z\sqrt{p(1-p)/n + z^2/4n^2}}{1 + z^2/n}, \qquad (2)$$

где p — значение показателя Precision; $z = 1{,}96$ — критическое значение стандартного нормального распределения для доверительной вероятности 95 %;

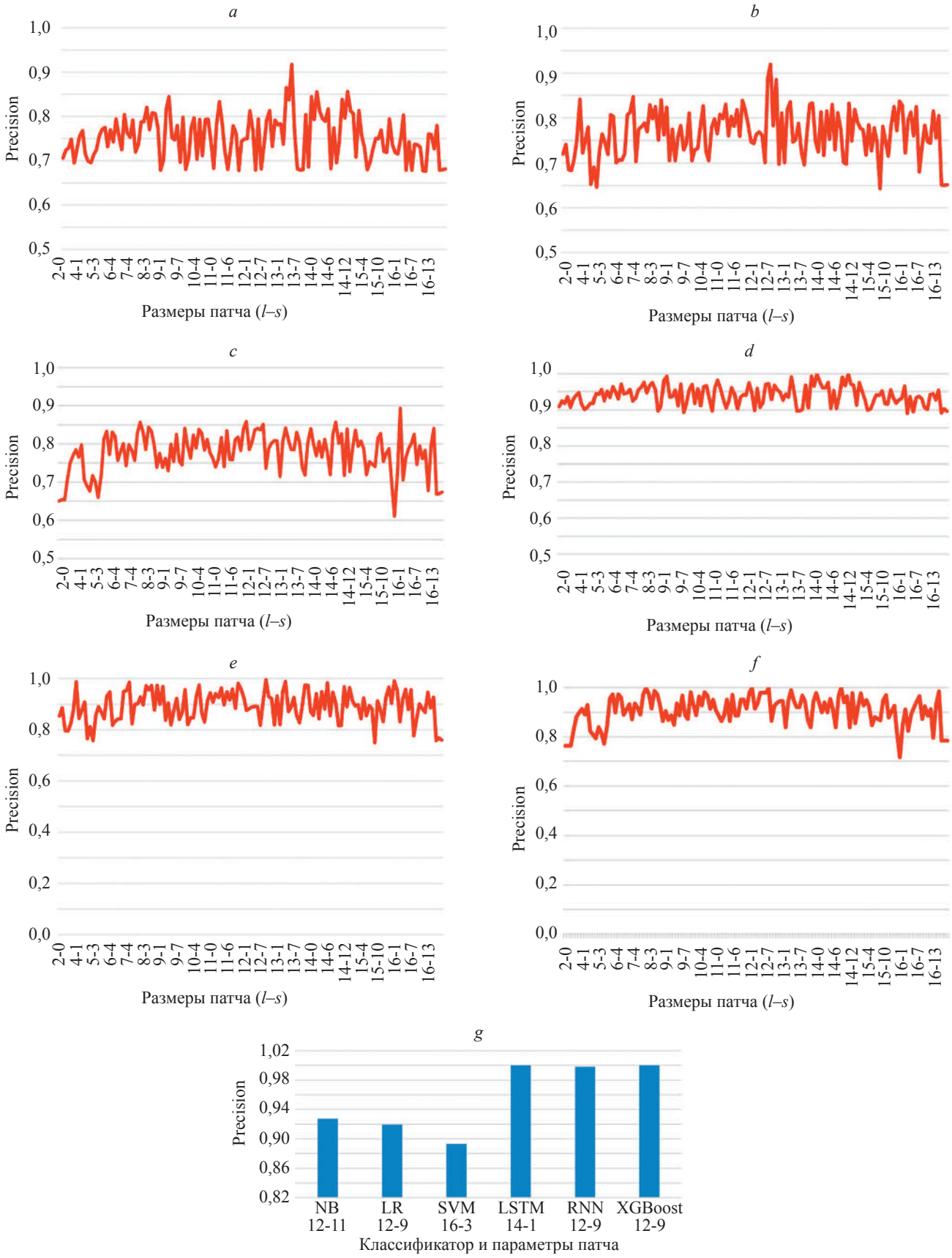


Рис. 3. Значения Precision классификаторов в зависимости от параметров оконного преобразования: NB (*a*); LR (*b*); SVM (*c*); LSTM (*d*); RNN (*e*); XGBoost (*f*); лучшие значения (*g*)

Fig. 3. Precision values of classifiers depending on the window transformation parameters: NB (*a*); LR (*b*); SVM (*c*); LSTM (*d*); RNN (*e*); XGBoost (*f*); the best values (*g*)

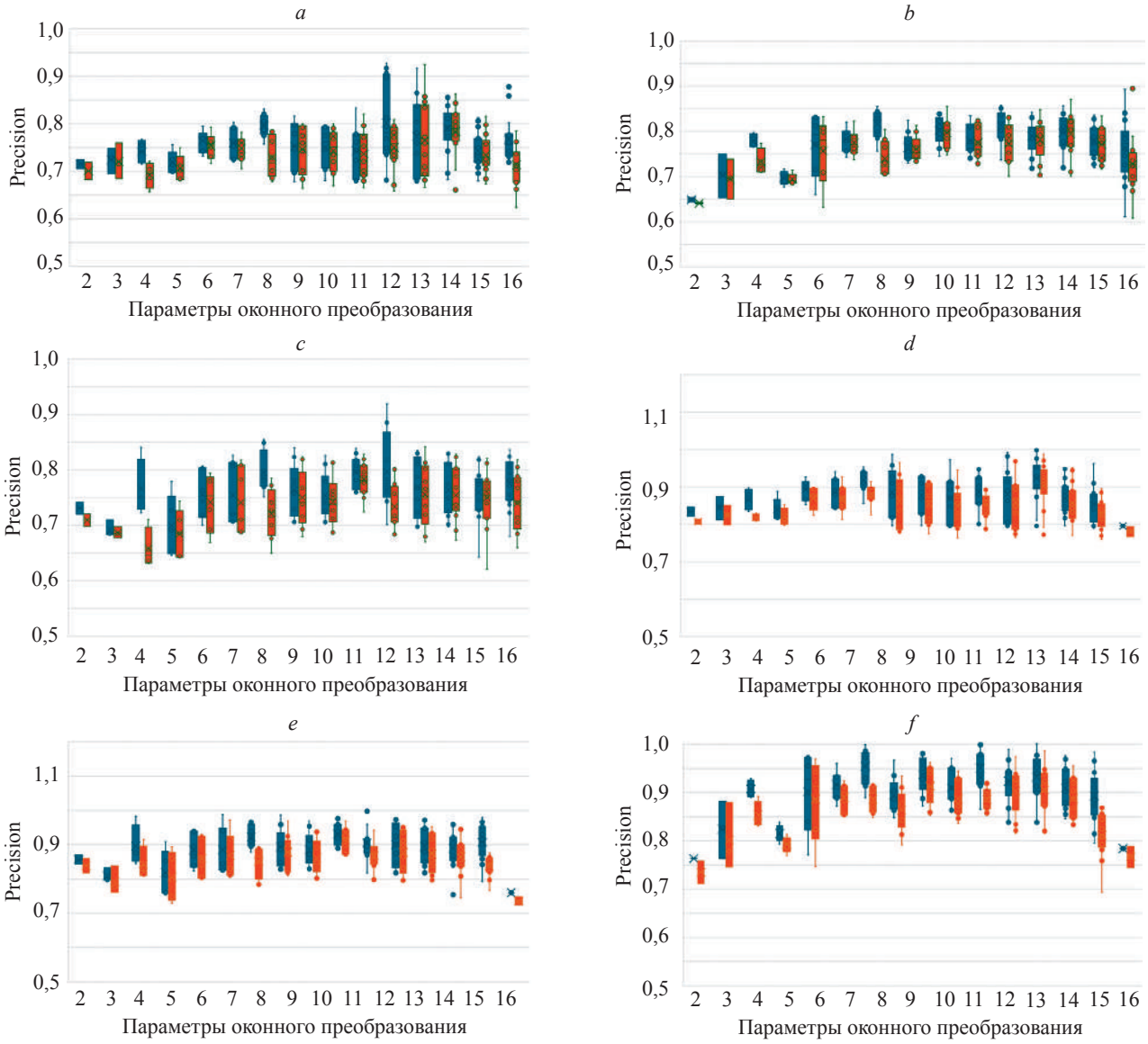


Рис. 4. Сравнение значений Precision для различных длин окна и после добавления шума на каждый элемент: NB (*a*); SVM (*b*); LR (*c*); LSTM (*d*); RNN (*e*); XGBoost (*f*)

Fig. 4. Comparison of Precision values for different window lengths and after adding noise to each element: NB (*a*); SVM (*b*); LR (*c*); LSTM (*d*); RNN (*e*); XGBoost (*f*)

n = TP + FP — число объектов, предсказанных классификатором как принадлежащих положительному классу; TP (True Positive) — число истинно положительных результатов; FP (False Positive) — число ложноположительных результатов.

На рис. 7 представлены изменения доверительных интервалов (2) выбранных классификаторов.

Доверительный интервал позволяет судить о достоверности результатов. Для части сдвигов при одинаковой длине патча доверительные интервалы не перекрываются, что позволяет говорить о влиянии длины и сдвига при формировании патча на достигаемые значения показателя качества.

Как показывают результаты эксперимента на рассмотренных данных, целенаправленное формирование обучающих выборок способствует росту показателей качества. Изменение размера патча и манипуляции данными путем сдвига дают возможность сформировать обучающие выборки, объекты наблюдения которых содержат информативные признаки.

Проверка на выборках данных

В качестве объектов анализа были выбраны временные ряды, описывающие разнообразные процессы. В зависимости от задачи классификации состояний требуется выделить из подпоследовательностей те объекты, которые обладают наиболее информативными характеристиками для достижения требуемых показателей качества.

Для экспериментальной проверки были отобраны три набора данных временных рядов: PowerCons, Sony и DodgerLoopGame. Перед анализом наборы данных прошли предварительную обработку: нормировку, уда-

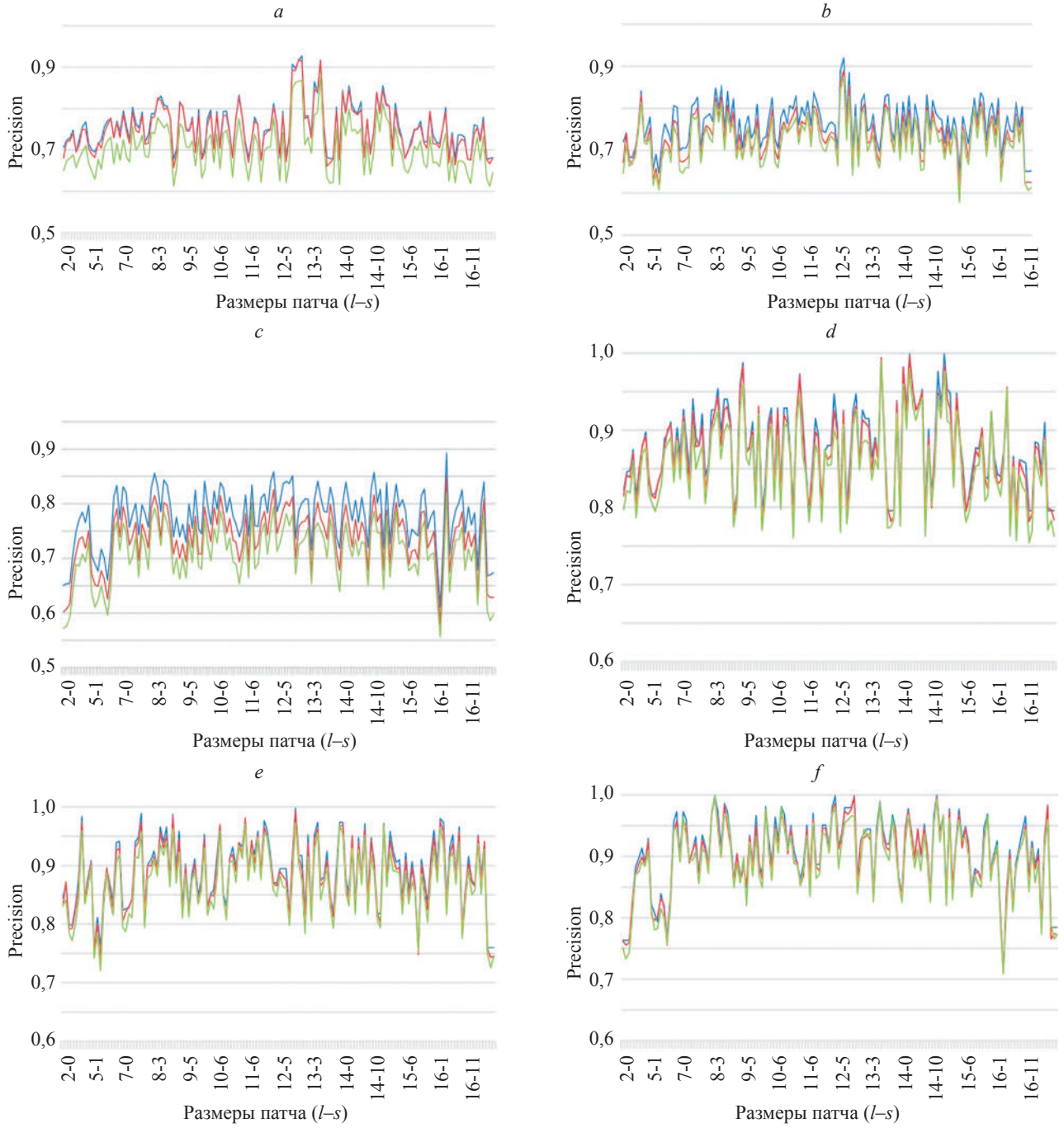


Рис. 5. Значения показателя Precision для классификаторов NB (*a*), LR (*b*), SVM (*c*), LSTM (*d*), RNN (*e*) и XGBoost (*f*) при различных размерах обучающей выборки: 90 % (синяя линия), 80 % (оранжевая линия), 70 % (зеленая линия)

Fig. 5. Precision values for classifiers NB (*a*), LR (*b*), SVM (*c*), LSTM (*d*), RNN (*e*), and XGBoost (*f*) for different training sample sizes: 90 % (blue line), 80 % (orange line), 70 % (green line)

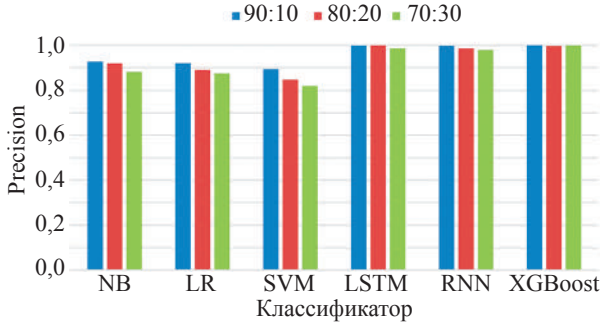


Рис. 6. Влияние разбиения выборки на обучающую и тестовую части и на значения показателя Precision классификаторов

Fig. 6. Effect of the training/test split on the Precision values of classifiers

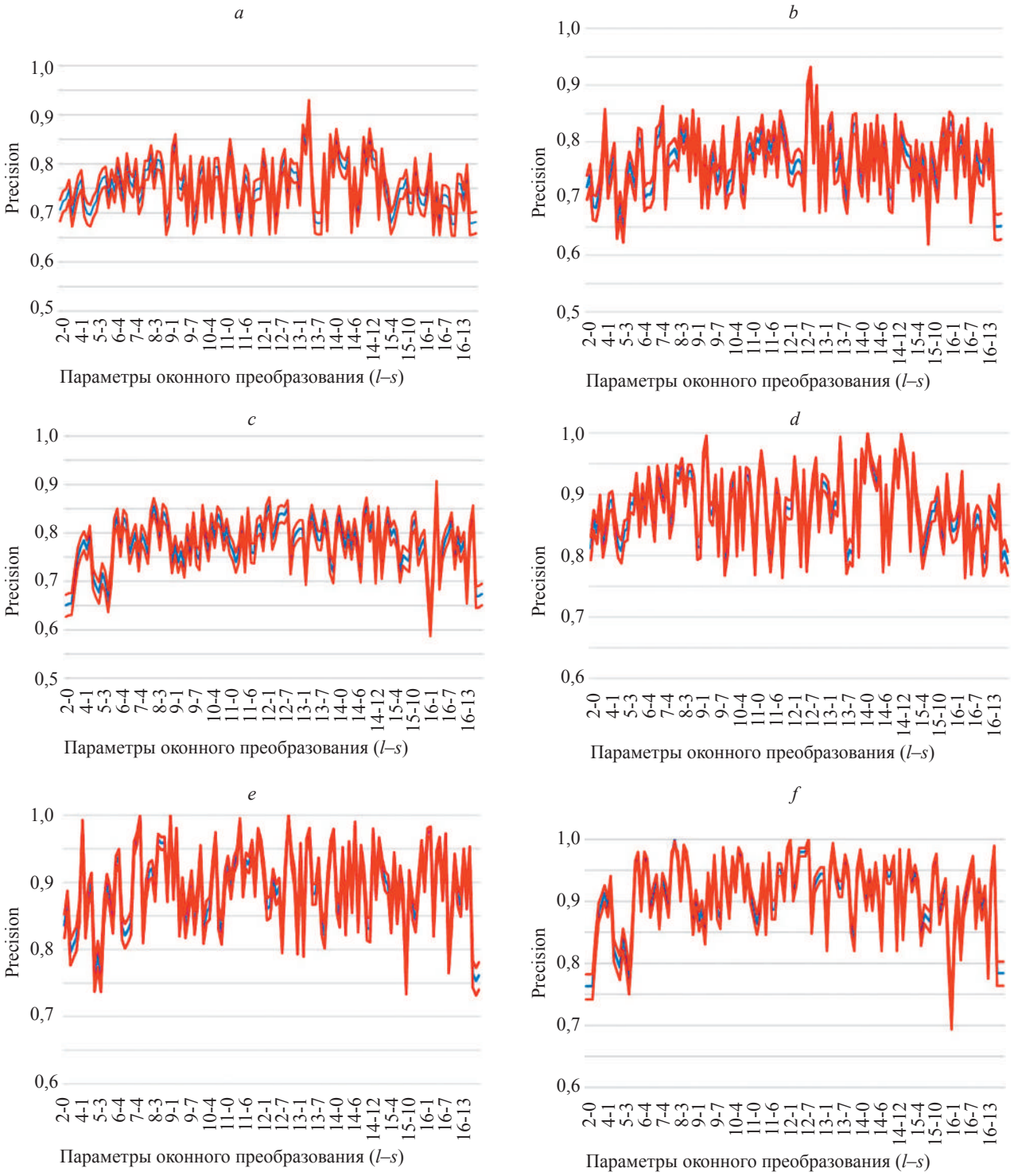


Рис. 7. Значения Precision и границы доверительных интервалов для классификаторов: NB (*a*); LR (*b*); SVM (*c*); LSTM (*d*); RNN (*e*); XGBoost (*f*)

Fig. 7. Precision values and confidence interval bounds for classifiers: NB (*a*); LR (*b*); SVM (*c*); LSTM (*d*); RNN (*e*); XGBoost (*f*)

ление дубликатов и некорректных значений, исключение выбросов, а также формирование последовательностей с присвоением меток класса для последующей проверки метода. Для поиска лучших значений по выборкам временных рядов, обладающих разными свойствами, размер объектов наблюдения был увеличен. Длина объектов наблюдения при переборе изменялась в диапазоне от 2 до 50 отсчетов. Под лучшим значением в таблице понимается максимальное достигнутое значение показателя качества при переборе параметров формирования объектов наблюдения, а под средним ($\overline{P_w}$) — взвешенное условное среднее значение показателя, вычисленное по всем исследованным комбинациям длины патча и сдвига для рассматриваемого набора данных и классификатора:

Таблица. Лучшие значения на наборах данных
Table. Best values obtained for the datasets

Классификатор	Параметр или показатель качества	Наборы данных					
		PowerCons		Sony		DodgerLoopGame	
		лучшее	среднее	лучшее	среднее	лучшее	среднее
NB	Длина патча, количество отсчетов	30	—	28	—	27	—
	Сдвиг, количество отсчетов	6	—	8	—	7	—
	Precision	0,81	0,72	0,84	0,8	0,79	0,71
	Recall	0,78	0,69	0,81	0,77	0,87	0,82
	F-мера	0,79	0,71	0,83	0,79	0,83	0,76
LR	Длина патча, количество отсчетов	30	—	22	—	23	—
	Сдвиг, количество отсчетов	26	—	14	—	3	—
	Precision	0,79	0,70	0,82	0,73	0,86	0,76
	Recall	0,76	0,67	0,96	0,89	0,84	0,73
	F-мера	0,77	0,69	0,88	0,80	0,85	0,74
SVM	Длина патча, количество отсчетов	30	—	17	—	23	—
	Сдвиг, количество отсчетов	26	—	10	—	3	—
	Precision	0,78	0,72	0,86	0,79	0,87	0,75
	Recall	0,71	0,65	0,97	0,88	0,82	0,71
	F-мера	0,74	0,68	0,91	0,83	0,84	0,73
XGBoost	Длина патча, количество отсчетов	28	—	26	—	24	—
	Сдвиг, количество отсчетов	9	—	11	—	5	—
	Precision	0,85	0,78	0,91	0,86	0,90	0,83
	Recall	0,83	0,76	0,93	0,89	0,88	0,81
	F-мера	0,84	0,77	0,92	0,87	0,89	0,82
RNN	Длина патча, количество отсчетов	29	—	27	—	25	—
	Сдвиг, количество отсчетов	8	—	10	—	6	—
	Precision	0,86	0,80	0,92	0,87	0,91	0,84
	Recall	0,84	0,78	0,94	0,90	0,89	0,82
	F-мера	0,85	0,79	0,93	0,89	0,9	0,83
LSTM	Длина патча, количество отсчетов	30	—	28	—	25	—
	Сдвиг, количество отсчетов	8	—	9	—	6	—
	Precision	0,88	0,82	0,94	0,90	0,92	0,86
	Recall	0,86	0,80	0,96	0,92	0,91	0,84
	F-мера	0,87	0,81	0,95	0,91	0,92	0,85

$$\overline{P}_w = \frac{\sum_{i=1}^{N} p_i N_{\text{pred},i}}{N_{\text{pred},i}}, \qquad (3)$$

где P_i — значение Precision для i-й комбинации параметров оконного преобразования; $N_{pred,i} = TP_i + FP_i$ — число объектов, отнесенных моделью к положительному классу при данной комбинации параметров.

В таблице представлены лучшие и средние значения, полученные с использованием выражения (3).

Данные таблицы показывают, что результаты работы классифицирующих алгоритмов зависят от выбора длины патча и величины сдвига при обработке последовательностей. Предложенный способ формирования обучающих выборок позволяет повысить значения качественных показателей, в частности показателя Precision.

Сравнение ошибок классификации указывает на то, что формирование выборки объектов наблюдения предложенным методом обычно способствует улучшению показателей качества обработки. Как видно из таблицы, подбор размера патча и сдвига в большинстве случаев обеспечивает повышение показателей до 3–8 % по сравнению со случайным выбором.

Обсуждение

Представленный метод отличается низкой вычислительной сложностью, высокой интерпретируемостью параметров и минимальной ресурсоемкостью, что де-

лает его подходящим для устройств с ограниченными возможностями.

В работе [17] рассмотрено адаптивное применение моделей машинного обучения на отдельных сегментах выборки в задачах регрессии и классификации, что также связано с адаптацией процесса обработки данных к свойствам выборки.

Разработанный метод использует стратегии перебора, например Grid Search или Random Search на ограниченном пространстве, что позволяет снизить его вычислительную сложность. Метод ориентирован на предобработку временных рядов и подходит для относительно простых моделей. В отличие от предложенного, метод информационных последовательностей ищет баланс между качеством модели и ее сложностью, применяя аналитические выражения. Применение метода наиболее эффективно для статистических моделей. Эволюционные алгоритмы применяют стохастический подход к нахождению глобального оптимума гиперпараметров, что обусловливает высокую вычислительную сложность из-за множества итераций. Это дает универсальность, но требует значительных ресурсов. Байесовская оптимизация использует вероятностные модели для поиска оптимума, имеет среднюю или высокую вычислительную сложность. Она подходит для широкого спектра задач, но эффективна только если оптимизируемая функция обладает определенной гладкостью.

Предложенный метод в большей степени является подходом к предобработке данных. В отличие от сложных методов оптимизации гиперпараметров, он решает более узкую задачу формирования данных таким образом, чтобы даже простые классификаторы показывали высокое качество. Это позволяет использовать его для предварительной обработки данных перед применением более сложных моделей.

Заключение

Предложен метод автоматического формирования объектов наблюдения из временных рядов, обеспечивающий выбор параметров оконного преобразования на основе оптимизации функционала качества. Метод позволяет рассматривать параметры представления данных как настраиваемые гиперпараметры процесса обучения и направлен на повышение качества оценки состояния системы без усложнения архитектуры используемой модели.

Особенность метода состоит в переносе процедуры оптимизации с пространства гиперпараметров модели на пространство формирования объектов наблюдения из временных рядов. Оптимизируется не модель, а представление данных, что позволяет повысить качество классификации без усложнения используемых алгоритмов машинного обучения. Полученные экспериментальные результаты показали, что выбор длины патча и величины сдвига позволяет повысить значения показателей качества на 3–8 % по сравнению со средними значениями по исследованным комбинациям параметров.

Литература

1. Salehi A., Balasubramanian M. DDCNet: Deep dilated convolutional neural network for dense prediction // Neurocomputing. 2023. V. 523. P. 116–129. doi: 10.1016/j.neucom.2022.12.024
2. Zeng A., Chen M., Zhang L., Xu Q. Are transformers effective for time series forecasting? // Proc. of the 37th AAAI Conference on Artificial Intelligence and 23th Conference on Innovative Applications of Artificial Intelligence and 13th Symposium on Educational Advances in Artificial Intelligence. 2023. V. 37. P. 11121–11128. doi: 10.1609/aaai.v37i9.26317
3. Allahham M.S., Al-Sa'd M.F., Al-Ali A., Mohamed A., Khattab T., Erbad A. DroneRF dataset: A dataset of drones for RF-based detection, classification and identification // Data in Brief. 2019. V. 26. P. 104313. doi: 10.1016/j.dib.2019.104313
4. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L, Gomez A.N., et al. Attention is all you need // Proc. of the 31st Conference on Neural Information Processing System. 2017.
5. Li D., Chen D., Jin B., Shi L., Goh J., Ng S.-K. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks // Lecture Notes in Computer Science. 2019. V. 11730. P. 703–716. doi: 10.1007/978-3-030-30490-4_56
6. Wu H., Xu J., Wang J., Long M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting // Proc. of the 35th Conference on Neural Information Processing Systems. 2021.
7. Zhou H., Zhang S., Peng J., Zhang S., Li J., Xiong H., et al. Informer: Beyond efficient transformer for long sequence time-series forecasting // Proc. of the AAAI Conference on Artificial Intelligence. 2021. V. 35. N 12. P. 11106–11115. doi: 10.1609/aaai.v35i12.17325
8. Wang J., Jiang J., Jiang W., Li C., Zhao W.X. Libcity: An open library for traffic prediction // Proc. of the 29th International Conference on Advances in Geographic Information Systems. 2021. P. 145–148. doi: 10.1145/3474717.3483923

References

1. Salehi A., Balasubramanian M. DDCNet: Deep dilated convolutional neural network for dense prediction. *Neurocomputing*, 2023, vol. 523, pp. 116–129. doi: 10.1016/j.neucom.2022.12.024
2. Zeng A., Chen M., Zhang L., Xu Q. Are transformers effective for time series forecasting? *Proc. of the 37th AAAI Conference on Artificial Intelligence and 23th Conference on Innovative Applications of Artificial Intelligence and 13th Symposium on Educational Advances in Artificial Intelligence*, 2023, vol. 37, pp. 11121–11128. doi: 10.1609/aaai.v37i9.26317
3. Allahham M.S., Al-Sa'd M.F., Al-Ali A., Mohamed A., Khattab T., Erbad A. DroneRF dataset: A dataset of drones for RF-based detection, classification and identification. *Data in Brief*, 2019, vol. 26, pp. 104313. doi: 10.1016/j.dib.2019.104313
4. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L, Gomez A.N., et al. Attention is all you need. *Proc. of the 31st Conference on Neural Information Processing System*, 2017.
5. Li D., Chen D., Jin B., Shi L., Goh J., Ng S.-K. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks. *Lecture Notes in Computer Science*, 2019, vol. 11730, pp. 703–716. doi: 10.1007/978-3-030-30490-4_56
6. Wu H., Xu J., Wang J., Long M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Proc. of the 35th Conference on Neural Information Processing Systems*, 2021.
7. Zhou H., Zhang S., Peng J., Zhang S., Li J., Xiong H., et al. Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proc. of the AAAI Conference on Artificial Intelligence*, 2021, vol. 35, no. 12, pp. 11106–11115. doi: 10.1609/aaai.v35i12.17325
8. Wang J., Jiang J., Jiang W., Li C., Zhao W.X. Libcity: An open library for traffic prediction. *Proc. of the 29th International Conference on Advances in Geographic Information Systems*, 2021, pp. 145–148. doi: 10.1145/3474717.3483923

9. Lebedev I.S., Sukhoparov M.E. Adaptive learning and integrated use of information flow forecasting methods // Emerging Science Journal. 2023. V. 7. N 3. P. 704-723. doi: 10.28991/esj-2023-07-03-03
10. Lebedev I.S., Sukhoparov M.E. Improving the quality indicators of multilevel data sampling processing models based on unsupervised clustering // Emerging Science Journal. 2024. V. 8. N 1. P. 355–371. doi: 10.28991/esj-2024-08-01-025
11. Middlehurst M., Ismail-Fawaz A., Guillaume A., Holder C., Guijo-Rubio D., Bulatova G., et al. aeon: a Python toolkit for learning from time series // Journal of Machine Learning Research. 2024. V. 25. N 289. P. 1–10.
12. Canizo M., Triguero I., Conde A., Onieva E. Multihead CNN-RNN for multi-time series anomaly detection: An industrial case study // Neurocomputing. 2019. V. 363. P. 246–260. doi: 10.1016/j.neucom.2019.07.034
13. Xiao Q., Wu B., Zhang Yu., Liu S., Pechenizkiy M., Mocanu E., et al. Dynamic sparse network for time series classification: Learning what to "see" // Advances in Neural Information Processing Systems. 2022. V. 35. P. 16849–16862. doi: 10.52202/068431-1226
14. Tallman E., West M. Bayesian predictive decision synthesis // Journal of the Royal Statistical Society Series B Statistical Methodology. 2024. V. 86. N 2. P. 340-363. doi: 10.1093/jrsssb/qkad109
15. Ekambaram V., Jati A., Nguyen N., Sinthong P., Kalagnanam J. Tsmixer: Lightweight mlp-mixer model for multivariate time series forecasting // Proc. of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023. P. 459–469. doi: 10.1145/3580305.3599533
16. Chen Y., Ren K., Wang Y., Fang Y., Sun W., Li D. Contiformer: Continuous-time transformer for irregular time series modeling // Advances in Neural Information Processing Systems. 2024. V. 36. P. 47143–47175. doi: 10.52202/075280-2042
17. Лебедев И.С. Адаптивное применение моделей машинного обучения на отдельных сегментах выборки в задачах регрессии и классификации // Информационно-управляющие системы. 2022. № 3 (118). С. 20–30. doi: 10.31799/1684-8853-2022-3-20-30

9. Lebedev I.S., Sukhoparov M.E. Adaptive learning and integrated use of information flow forecasting methods. *Emerging Science Journal*, 2023, vol. 7, no. 3, pp. 704-723. doi: 10.28991/esj-2023-07-03-03
10. Lebedev I.S., Sukhoparov M.E. Improving the quality indicators of multilevel data sampling processing models based on unsupervised clustering. *Emerging Science Journal*, 2024, vol. 8, no. 1, pp. 355–371. doi: 10.28991/esj-2024-08-01-025
11. Middlehurst M., Ismail-Fawaz A., Guillaume A., Holder C., Guijo-Rubio D., Bulatova G., et al. aeon: a Python toolkit for learning from time series. *Journal of Machine Learning Research*, 2024, vol. 25, no. 289, pp. 1–10.
12. Canizo M., Triguero I., Conde A., Onieva E. Multihead CNN-RNN for multi-time series anomaly detection: An industrial case study. *Neurocomputing*, 2019, vol. 363, pp. 246–260. doi: 10.1016/j.neucom.2019.07.034
13. Xiao Q., Wu B., Zhang Yu., Liu S., Pechenizkiy M., Mocanu E., et al. Dynamic sparse network for time series classification: Learning what to "see". *Advances in Neural Information Processing Systems*, 2022, vol. 35, pp. 16849–16862. doi: 10.52202/068431-1226
14. Tallman E., West M. Bayesian predictive decision synthesis. *Journal of the Royal Statistical Society Series B Statistical Methodology*, 2024, vol. 86, no. 2, pp. 340-363. doi: 10.1093/jrsssb/qkad109
15. Ekambaram V., Jati A., Nguyen N., Sinthong P., Kalagnanam J. Tsmixer: Lightweight mlp-mixer model for multivariate time series forecasting. *Proc. of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 459–469. doi: 10.1145/3580305.3599533
16. Chen Y., Ren K., Wang Y., Fang Y., Sun W., Li D. Contiformer: Continuous-time transformer for irregular time series modeling. *Advances in Neural Information Processing Systems*, 2024, vol. 36, pp. 47143–47175. doi: 10.52202/075280-2042
17. Lebedev I. Adaptive application of machine learning models on separate segments of a data sample in regression and classification problems. *Information and Control Systems*, 2022, no. 3, pp. 20–30. (in Russian). doi: 10.31799/1684-8853-2022-3-20-30

Авторы

Лебедев Илья Сергеевич — доктор технических наук, профессор, главный научный сотрудник, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, sc 56321781100, https://orcid.org/0000-0001-6753-2181, isl_box@mail.ru

Макеенко Александр Алексеевич — аспирант, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, https://orcid.org/0009-0002-4453-9887, aleks.mackeencko@yandex.ru

Authors

Ilya S. Lebedev — D.Sc., Professor, Senior Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), Saint Petersburg, 199178, Russian Federation, sc 56321781100, https://orcid.org/0000-0001-6753-2181, isl_box@mail.ru

Alexander A. Makeenko — PhD Student, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), Saint Petersburg, 199178, Russian Federation, https://orcid.org/0009-0002-4453-9887, aleks.mackeencko@yandex.ru

Статья поступила в редакцию 29.03.2026
Одобрена после рецензирования 25.08.2026
Принята к печати 24.09.2026

Received 29.03.2026
Approved after reviewing 25.08.2026
Accepted 24.09.2026

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»