

Научная статья
УДК 004.6:004.8
<https://doi.org/10.24143/2072-9502-2025-4-63-73>
EDN YTPREMV

Интеллектуальный анализ разнородных данных на основе гибридных нейросетевых методов и адаптивной нормализации

Ангелина Игоревна Дубровина

*Донской государственный технический университет,
Ростов-на-Дону, Россия, ministrelia69@yandex*

Аннотация. В настоящее время наблюдается стремительный рост типов и объемов накапливаемых данных в различных предметных областях, в связи с чем возрастает потребность в разработке эффективных подходов в части, касающейся визуализации, трансформации и анализа информации. Особенно актуально применение компьютерных методов, включая алгоритмы машинного обучения и элементы искусственного интеллекта, позволяющих существенно повысить точность и скорость обработки информации. Рассматриваются современные методы интеграции этапов сбора, предобработки, визуализации и анализа данных, направленные на решение прикладных задач в таких областях, как экология, финансы, медицина и управление персоналом. Предложен новый метод VisNormNet, объединяющий визуализацию, нормализацию и гибридную нейросетевую обработку для интерпретируемого анализа разнотипных данных. В отличие от существующих решений метод предусматривает адаптивный выбор процедур предобработки на основе корреляционной структуры для повышения устойчивости модели к выбросам и смещению распределений. Предложен подход, объединяющий статистические методы, визуальные средства анализа и машинное обучение, в частности нейронные сети. Такой подход позволяет проводить автоматизированный анализ больших и разнотипных наборов данных. Проведены вычислительные эксперименты на открытых и внутренних выборках, демонстрирующие преимущества предложенного метода по сравнению с традиционными алгоритмами, включая логистическую регрессию и метод опорных векторов. Результаты показали значительное увеличение точности классификации и прогнозирования, особенно при использовании гибридных моделей, и могут быть использованы для построения интеллектуальных систем поддержки принятия решений в условиях неопределенности и неполноты данных. Практическая значимость исследования заключается в разработке универсального инструментария для специалистов в области анализа данных, ориентированного на повышение эффективности принимаемых решений. Перспективы работы связаны с адаптацией метода для потоковых данных и его внедрением в новые предметные области, такие как промышленный IoT и предиктивная аналитика.

Ключевые слова: анализ данных, системный анализ, нейронные сети, обработка информации, статистика

Для цитирования: Дубровина А. И. Интеллектуальный анализ разнородных данных на основе гибридных нейросетевых методов и адаптивной нормализации // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2025. № 4. С. 63–73. <https://doi.org/10.24143/2072-9502-2025-4-63-73>. EDN YTPREMV.

Original article

Intelligent analysis of heterogeneous data based on hybrid neural network methods and adaptive normalization

Angelina I. Dubrovina

*Don State Technical University,
Rostov-on-Don, Russia, ministrelia69@yandex*

Abstract. Currently, there is a rapid growth in the types and volumes of accumulated data in various subject areas, which increases the need to develop effective approaches in terms of information visualization, transformation and analysis. Especially relevant is the use of computer methods, including machine learning algorithms and artificial intelligence elements, which significantly improve the accuracy and speed of information processing. Modern methods of integrating the stages of data collection, preprocessing, visualization and analysis aimed at solving applied problems in such areas as ecology, finance, medicine and personnel management are considered. A new VisNormNet method is proposed that

combines visualization, normalization, and hybrid neural network processing for the interpretable analysis of diverse types of data. Unlike existing solutions, the method provides for an adaptive choice of preprocessing procedures based on a correlation structure to increase the model's resilience to outliers and distribution bias. An approach combining statistical methods, visual analysis tools and machine learning, in particular neural networks, is proposed. This approach allows for automated analysis of large and diverse datasets. Computational experiments have been conducted on open and internal samples, demonstrating the advantages of the proposed method in comparison with traditional algorithms, including logistic regression and the method of support vectors. The results showed a significant increase in the accuracy of classification and forecasting, especially when using hybrid models, and can be used to build intelligent decision support systems in conditions of uncertainty and data incompleteness. The practical significance of the research lies in the development of a universal tool for specialists in the field of data analysis, focused on improving the effectiveness of decisions. The prospects of the work are related to the adaptation of the method for streaming data and its implementation in new subject areas such as industrial IoT and predictive analytics.

Keywords: data analysis, system analysis, neural networks, information processing, statistics

For citation: Dubrovina A. I. Intelligent analysis of heterogeneous data based on hybrid neural network methods and adaptive normalization. *Vestnik of Astrakhan State Technical University. Series: Management, computer science and informatics*. 2025;4:63-73. (In Russ.). <https://doi.org/10.24143/2072-9502-2025-4-63-73>. EDN YTPENV.

Введение

Современный мир сталкивается с рекордными объемами данных, которые требуют применения эффективных методов для их анализа, визуализации и трансформации. От того, насколько качественно и быстро можно обработать информацию, зависит успех во многих отраслях, включая медицину, финансы, науку и промышленность [1]. Ключевую роль в работе с данными играют специалисты, такие как аналитики и инженеры данных, профессиональная деятельность которых сформировалась в эпоху цифровой трансформации последнего десятилетия. Такие специалисты объединяют компетенции в области программирования, статистики и предметной области, что позволяет им быстро и качественно преобразовывать неготовые (сырые) данные в стратегически важные и надежные решения [2].

В статье, ориентированной на потребности специалистов, систематизируются методы решения взаимосвязанных задач: визуализации, трансформации и анализа данных. Особого внимания требует применение методов в условиях ограниченных ресурсов к большим запросам интерпретируемых результатов.

Одной из актуальных задач является использование компьютерных методов для обработки данных, что позволяет значительно ускорить принятие решений и повысить точность анализа [2]. Визуализация, трансформация и анализ данных – неотъемлемые этапы в процессе работы с данными, которые позволяют преобразовывать сырые данные в ценные инсайты [3].

Пусть задано множество разнородных данных, представленное в виде матрицы «объект – признак» ($X = [x_{ij}] \in R^{n \times m}$, где за n принято количество наблюдений или объектов, m – количество признаков, а X – значения i -го и j -го признаков), при этом признаки могут быть следующих типов: количественные, категориальные, временные, визуальные, – что показывает их существенную разнородность в масштабах и распределениях. Каждому объекту может быть поставлена в соответствие целевая переменная для обозначения множества допустимых меток

классов при их классификации или множества действительных чисел в случае регрессии.

В настоящее время исследователи имеют доступ к широкому спектру методик обработки, анализа и графического представления таких данных. Оптимальный выбор конкретного метода определяется совокупностью характеристик исходных данных и особенностями поставленной аналитической проблемы. Например, анализ временных рядов требует совершенно иной визуализации и трансформации, чем анализ изображений или текстов [4]. Важно правильно подобрать инструменты и методы, которые помогут максимально эффективно извлечь нужную информацию. Не менее важным является понимание того, как различные способы обработки данных влияют на точность и скорость анализа, а также как на практике применяются эти методы [5]. Для специалистов важной проблемой остается не только владение отдельными инструментами, но и способность строить решения по обработке данных и оптимизированные конкретные кейсы [6].

Цель данной работы заключается в разработке и экспериментальной верификации нового метода интеллектуального анализа разнородных данных на основе интеграции визуализации, нормализации и машинного обучения [6].

Основные задачи исследования заключаются:

- в проведении анализа существующих подходов визуализации и трансформации данных, а также поиске их ограничений и недостатков при работе с мультимедийными наборами данных;
- применении метода VisNormNet, формализующего выбор преобразования данных на основе визуальной и корреляционной структуры [7];
- реализации экспериментальной модели на открытых и внутренних выборках и сравнении ее эффективности с традиционными алгоритмами;
- оценке устойчивости модели к частичным шумам, выбросам и уменьшенным объемам обучающей выборки [8].

Математическая постановка задачи заключается в построении такого отображения, которое с максимальной точностью и интерпретируемо-

стью сможет обработать целевую переменную по исходным данным, минимизируя при этом влияние их разнородности, шумов и выбросов. Отображение должно обеспечивать адаптивную предобработку, учитывающую статистические свойства групп взаимосвязанных признаков, интерпретируемость этапов преобразования данных и принятия решения, а также устойчивость к нарушениям стандартных предположений о данных.

Практическая значимость данной работы заключается в формировании рекомендаций для специалистов по выбору и комбинации методов в зависимости от типа данных и бизнес-требований.

Материалы и методы исследования

Визуализация данных играет ключевую роль в понимании структуры данных и взаимосвязей между их признаками. Для простых зависимостей можно использовать линейные графики, для более сложных – тепловые карты, графы и трехмерные визуализации [9]. Обучение проводилось с помощью публичных датасетов MNIST (60 000 обучаю-

щих и 10 000 тестовых изображений), CIFAR-10 (10 классов, 50 000 обучающих и 10 000 тестовых изображений) и Iris (150 экземпляров, 4 признака). Инструменты matplotlib, seaborn или Plotly позволяют создавать интерактивные и статические визуализации данных для вывода результата обучения на датасетах. Например, тепловая карта, демонстрирующая корреляцию между различными переменными, позволяет быстро выявить, какие признаки имеют сильное влияние на результат [10]. За основу проведения исследования взяты внутренние данные организации N, включающие стаж работы (в месяцах). Метрики «удовлетворенность» и «эффективность» измерены по интегральной 10-балльной шкале на основе ежегодного анонимного опроса сотрудников и экспертной оценки менеджеров соответственно.

Данные были предоставлены HR-отделом компании с соблюдением конфиденциальности.

Реализация тепловой карты объемом выборки в количестве 850 сотрудников отображена на рис. 1.

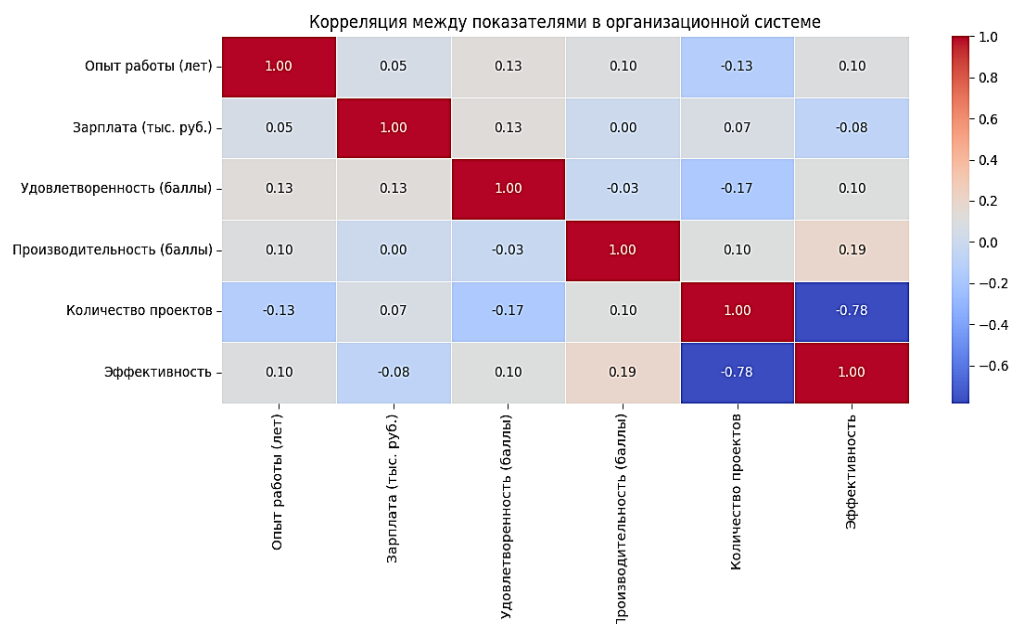


Рис. 1. Тепловая карта организационной системы

Fig. 1. The heat map of the organizational system

По тепловой карте можно сделать вывод о том, что стаж почти не влияет на корреляцию между количеством проектов, стажем, эффективностью и отработанными часами. Отрицательная корреляция (коэффициент Пирсона $< 0,2$) показывает, что опытные сотрудники, у которых стаж более 5 лет, участвуют в меньшем числе проектов, т. к. отдают приоритет задачам более высокого уровня. При этом наблюдается умеренная отрицательная корреляция между количеством проектов и показателем эффек-

тивности ($r \approx -0,45$). Это свидетельствует о тенденции к снижению индивидуальной результативности при росте количества одновременно выполняемых проектов, что может быть интерпретировано как эффект перегрузки. Данный эффект особенно выражен у специалистов с большим стажем, которые, как показывает анализ, чаще вовлекаются в решение сложных и ресурсоемких задач, требующих высокой концентрации.

При подготовке данных к проведению анализа особую значимость приобретают процедуры преобразования и нормализации, являющиеся обязательными предварительными этапами обработки информации. Особенно критичной данная процедура становится при работе с разнородными показателями, выраженными в существенно различающихся масштабах измерений (к примеру, когда одни переменные исчисляются миллионами, а другие – тысячами единиц). Нормализация подобных данных обеспечивает приведение их к сопоставимому виду, что является необходимым условием для коррект-

ного функционирования алгоритмов машинного обучения и получения достоверных результатов анализа. Для формирования персонального профиля и оценки вклада каждого сотрудника по отдельности был проведен анализ ключевых атрибутов. На рис. 2 представлена радиальная диаграмма, отображающая нормализованные значения ключевых признаков (навыки, опыт, возраст, расположение подразделения работника) для одного репрезентативного сотрудника, что позволяет визуальнo сравнивать сильные и слабые стороны в рамках организационного контекста.

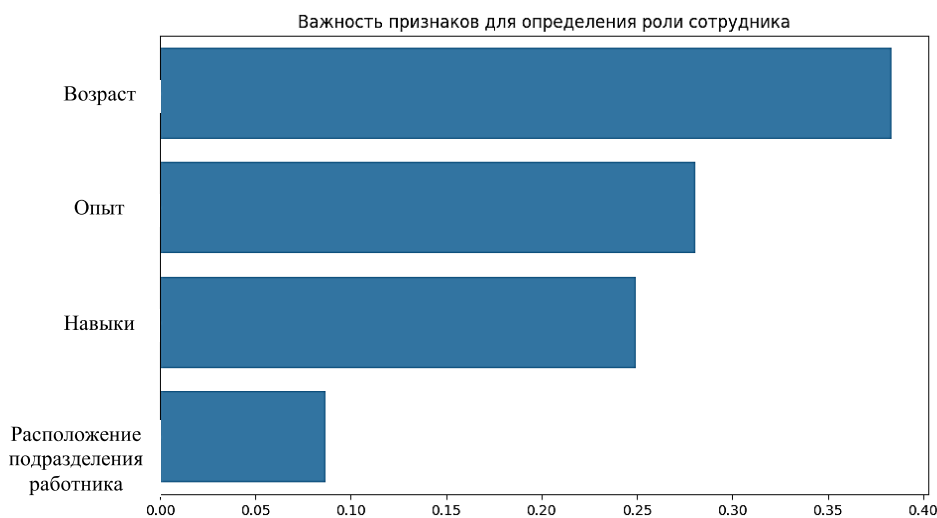


Рис. 2. Признаки сотрудников

Fig. 2. Employee signs

Преобразования, такие как стандартизация или логарифмическая трансформация, позволяют привести данные к удобному виду для анализа. Рас-

смотрим пример стандартизации данных с помощью метода Z-оценки (рис. 3, 4).

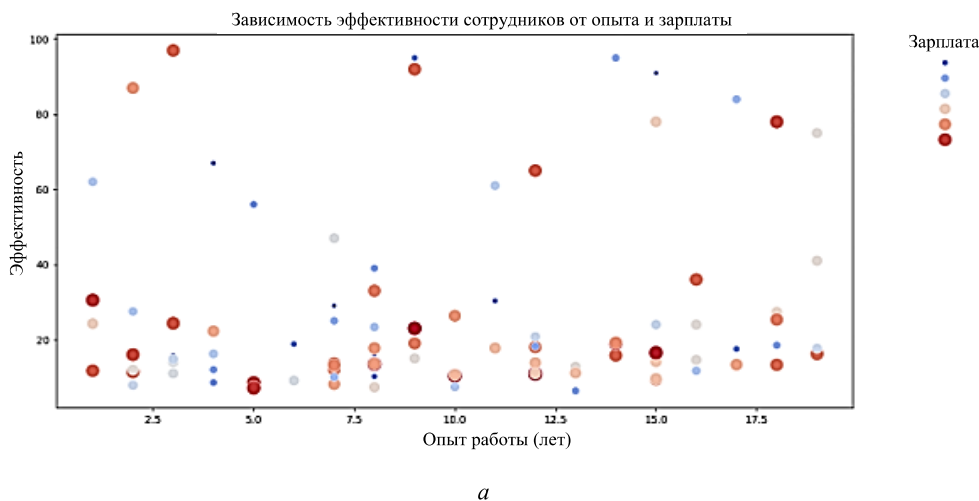


Рис. 3. Исходное распределение показателя эффективности сотрудников (a)

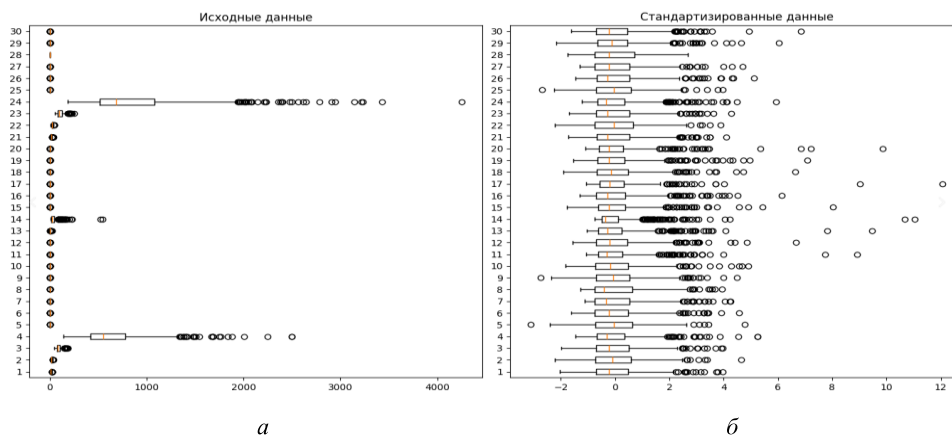
Fig. 3. The initial distribution of employee performance indicators (a)

Исходные данные:									
0	mean radius	mean texture	mean perimeter	mean area	...	worst concavity	worst concave points	worst symmetry	worst fractal dimension
1	17.99	18.38	122.80	1801.0	...	0.7119	0.2654	0.4681	0.11890
2	28.57	17.77	132.90	1326.0	...	0.2416	0.1869	0.2750	0.08902
3	19.69	21.25	130.00	1203.0	...	0.4504	0.2430	0.3613	0.08758
4	11.42	20.38	77.58	386.1	...	0.6869	0.2575	0.4638	0.17380
5	20.29	14.34	135.18	1297.0	...	0.4880	0.1625	0.2364	0.07678
[5 rows x 30 columns]									
Стандартизированные данные:									
0	mean radius	mean texture	mean perimeter	mean area	...	worst concavity	worst concave points	worst symmetry	worst fractal dimension
1	1.097064	-2.071135	1.269934	0.984375	...	2.189526	2.296076	2.750622	1.912015
2	1.829821	-0.353632	1.685955	1.988708	...	-0.146749	1.087084	-0.243890	0.281190
3	1.579888	0.456187	1.566503	1.558884	...	0.854974	1.955000	1.152255	0.201391
4	-0.768909	0.253732	-0.592687	-0.764464	...	1.989588	2.175786	6.046041	4.935010
5	1.750297	-1.151816	1.776573	1.826229	...	0.613179	0.729259	-0.868353	-0.397100
[5 rows x 30 columns]									

б

Рис. 3 (окончание). Исходное распределение показателя эффективности сотрудников и результат его стандартизации методом Z-оценки (б)

Fig. 3 (ending). The initial distribution of employee performance indicators and the result of its standardization using the Z-score method (б)



а

б

Рис. 4. Исходное распределение признака «Отработанные часы» (а) и его стандартизированная версия (б)

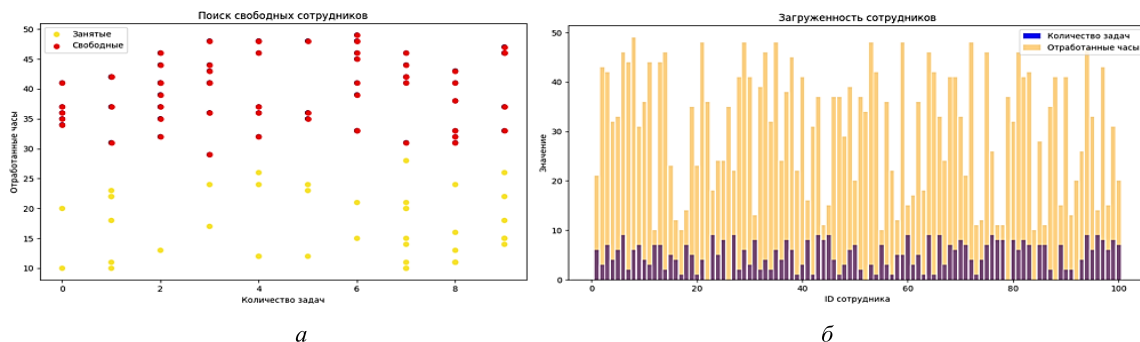
Fig. 4. The original distribution of the “Hours worked” feature (а) and its standardized version (б)

Для признака «Отработанные часы», диапазон которых составил 120–250 ч, применялась стандартизация по формуле

$$z = \frac{x - \mu}{\sigma},$$

где z – стандартизированное значение; x – исходное значение признака; μ – среднее арифметическое значение признака ($\mu = 185$); σ – стандартное отклонение ($\sigma = 32$).

Использование алгоритмов машинного обучения (кластеризация, классификация, нейронные сети) позволяет на основе подготовленных данных решать более сложные задачи, такие как распознавание объектов на изображениях или прогнозирование временных рядов [6]. Алгоритмы позволяют автоматически перераспределять рабочую нагрузку. Например, если один сотрудник перегружен, задачи можно перенаправить другим, более свободным коллегам (рис. 5).



а

б

Рис. 5. Алгоритм поиска сотрудников (а) и подсчет загрузки (б)

Fig. 5. Employee search algorithm (а) and workload calculation (б)

Методы визуализации, трансформации и анализа данных активно используются не только для поиска свободных сотрудников, но и для подсчета отработанных часов и количества задач на одну единицу сотрудника. Например, в организациях эти методы применяются для анализа и визуализации изображений, таких как снимки по переводу из рукописного в электронный вариант, что помогает сотрудникам быстрее выполнять свою работу [9].

В финансовой сфере они используются для анализа рыночных данных и прогнозирования трендов [4], в экологии – для мониторинга состояния окружающей среды с помощью спутниковых данных и информации с сенсоров. Эти технологии позволяют обрабатывать и анализировать большие объемы данных, что невозможно было бы сделать вручную. Рассмотрим табл. 1 с результатами анализа медицинских изображений.

Таблица 1

Table 1

Результаты анализа медицинских изображений

Medical image analysis results

Метод	Точность классификации, %	Время обучения, ч	Примечание
Логистическая регрессия	84,2	1,5	Традиционный метод, работает быстро, но имеет низкую точность на сложных данных
Метод опорных векторов (SVM)	87,8	4	Подходит для многоклассовой классификации, но требует времени на обучение
Нейронная сеть (CNN)	93,5	20	Высокая точность на сложных изображениях, требует больше вычислительных ресурсов
Гибридный метод (CNN + SVM)	95,3	28	Комбинированный подход, требует больших вычислительных ресурсов

Предложенный подход

В рамках данного исследования разработан метод VisNormNet, который ориентирован на интегрированную обработку и анализ данных с различной природой (табличные, визуальные, временные данные). Метод включает три логически связанных этапа, описанных ниже. Математически метод VisNormNet можно представить как композицию трех отображений, последовательно преобразующих исходное пространство признаков в пространство решений.

Этап 1 – визуализация и корреляционный анализ. На основе тепловых карт и графов взаимосвязей выполняется автоматический анализ структуры признаков, по которым выявляются признаки с высокой коррелированностью, формируются кластеры связанных переменных. На первом этапе происходит вычисление матрицы попарных корреляций, которую математически можно представить в следующем виде:

$$P_{i,j} = \frac{\text{cov}(X_i, X_j)}{X_i \cdot X_j} \quad \forall i, j \in \{1, \dots, m\},$$

где P – матрица коэффициента корреляции; i, j – индексы, указывающие на конкретную пару величин из рассматриваемого набора; $\text{cov}(X_i, X_j)$ – ковариация между признаками X_i и X_j , а X_i и X_j – стан-

дартное отклонение признака; m – общее количество признаков.

Этап 2 – адаптивная нормализация – для каждого кластера из признаков оптимальное преобразование на основе анализа распределения данных в кластере, это может быть Z-преобразование, логарифмирование или масштабирование, что позволяет уменьшить влияние выбросов. В результате применения преобразований ко всем кластерам формируется новая матрица преобразования признаков.

Этап 3 – гибридное обучение – визуальные эмбединги, которые получены через сверточную нейросеть и подаются на вход мета-модели SVM, архитектура обеспечивает устойчивость к переобучению и повышает интерпретируемость за счет сохранения визуального контроля на всех этапах. Затем эмбединги используются как признаки для обучения мета-классификатора на основе метода опорных векторов SVM с радиальным базисным ядром RBF.

В VisNormNet при обработке появляется новый объект (например, данные нового сотрудника). Далее он проходит через 1 и 2 этапы адаптивной нормализации, а затем CNN преобразует его в эмбединг. SVM проводит сравнение нового эмбединга со всеми опорными векторами, которые он запомнил во время обучения, умножая схожесть на вес и метку каждого опорного вектора, при этом

получая итоговую сумму в виде одного из следующих вариантов:

- если новый объект в основном похож на опорные векторы класса «+1» (т. е. «эффективные сотрудники»), то положительные слагаемые в сумме будут преобладать, а результат суммы будет положительным;
- при ситуации, в которой новый объект похож на векторы класса «-1», отрицательных слагаемых

будет больше и результат станет отрицательным.

На заключительном этапе работы метод VisNormNet на основе обработанных данных определяет принадлежность объекта к тому или иному классу. Схема алгоритма, иллюстрирующая последовательность этапов и поток данных, реализованная на Python с использованием Keras, Scikit-learn и Matplotlib, представлена на рис. 6.

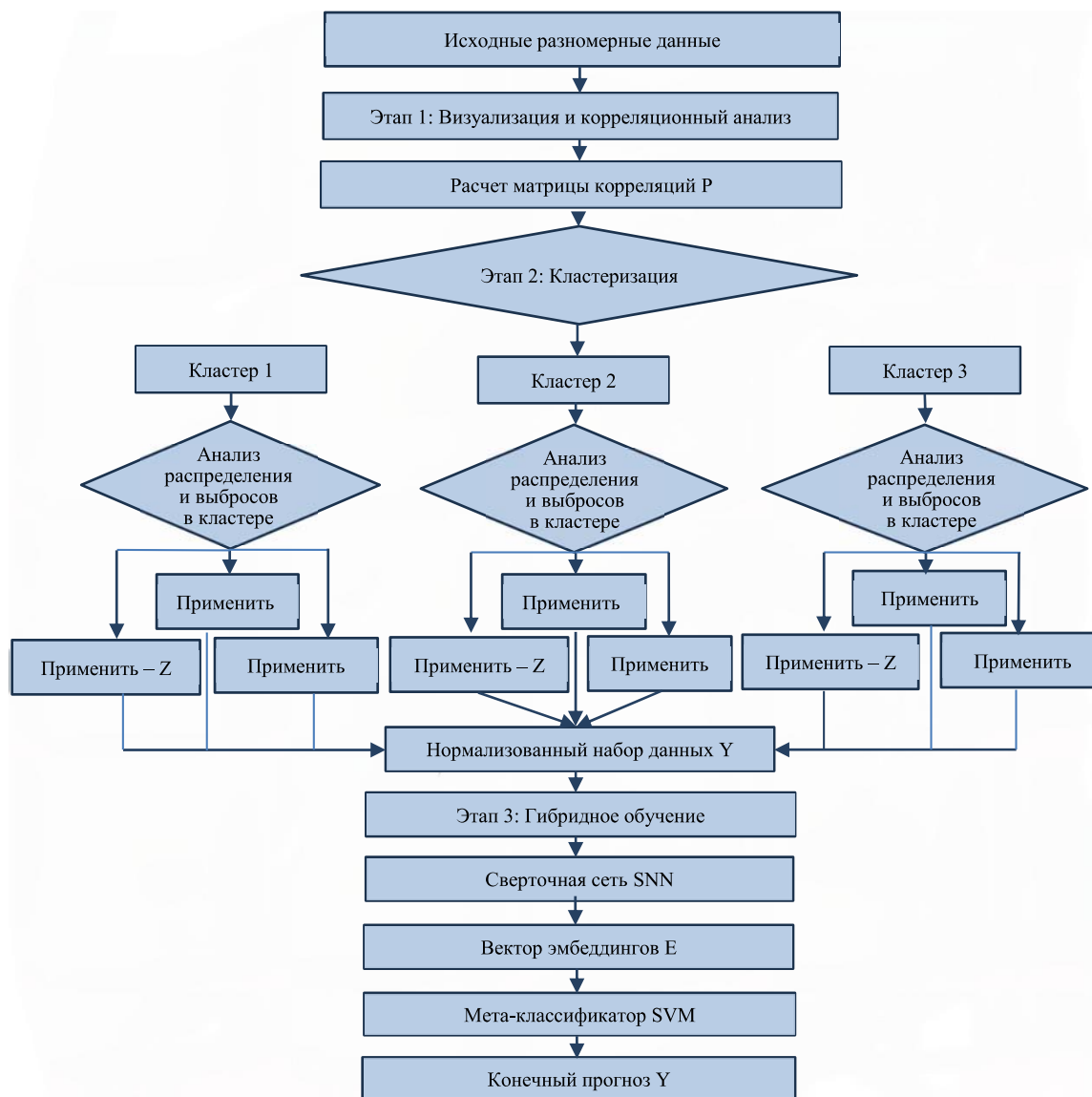


Рис. 6. Алгоритм архитектуры модели VisNormNet, последовательность этапов

Fig. 6. The architecture algorithm of the VisNormNet model, the sequence of stages

На первом этапе данные собираются из различных источников, таких как камеры видеонаблюдения или другие устройства. Затем происходит предобработка данных – очистка от шумов и нормали-

зация для дальнейшего использования в модели [7]. И, наконец, на основе подготовленных данных обучается модель машинного обучения для решения задачи классификации или прогнозирования. Этот

подход позволяет интегрировать все этапы обработки данных и использовать их в реальных приложениях.

Принцип работы предложенного метода заключается в последовательности действий, описанных ниже.

Сначала происходит сбор и визуализация данных. Для этого используются различные графики и диаграммы, такие как линейные графики для временных рядов или тепловые карты для анализа взаимосвязей между переменными. На следующем этапе данные очищаются и нормализуются для обеспечения их корректной обработки в дальнейших этапах анализа. И, наконец, на основе подготовленных данных обучается модель машинного обучения, например нейронная сеть для решения

задачи классификации изображений или временных рядов.

Для тестирования предложенного подхода был использован набор данных изображений для классификации объектов. Метод показал высокую эффективность в решении задачи классификации, значительно улучшив точность по сравнению с традиционными методами. Например, использование сверточной нейронной сети для классификации позволило получить более высокую (на 20 %) точность по сравнению с классическими методами, такими как линейная регрессия [8]. Рассмотрим табл. 2 с результатами тестирования моделей на различных наборах данных.

Таблица 2

Table 2

Результаты тестирования разных моделей

Test results of different models

Модель	Набор данных	Точность классификации, %	Время обучения, мин	Примечание
Логистическая регрессия	Iris	97,5	1,2	Хорошие результаты на простых данных
Метод опорных векторов (SVM)		99,1	3,5	Высокая точность, но больше времени на обучение
Нейронная сеть (MLP)		99,5	5,8	Отличные результаты на простых данных
Логистическая регрессия	MNIST	92,8	12	Хорошая точность на изображениях, но хуже, чем на Iris
Метод опорных векторов (SVM)		80,3	18	Применение SVM улучшает точность на MNIST
Нейронная сеть (MLP)		98,9	45	Отличные результаты, но требует больше времени на обучение
Логистическая регрессия	CIFAR-10	74,5	30	Сложности с классификацией изображений высокого разрешения
Метод опорных векторов (SVM)		80,2	55	SVM может показывать хорошую точность, но требует много времени
Нейронная сеть (CNN)		93,7	130	Превосходные результаты на изображениях высокого разрешения, требует больше вычислительных ресурсов

Обсуждение результатов

Применение предложенного метода, включающего нейронные сети и автоматизированную предобработку данных, значительно улучшает точность и ускоряет процесс анализа. По сравнению с традиционными методами, такими как использование ручной обработки или простых статистиче-

ских методов, новый подход показывает лучшие результаты. Например, нейронные сети, обученные на подготовленных данных, обеспечивают гораздо более точную классификацию объектов, чем старые методы. На рис. 7 представлен график сравнения точности разных моделей.

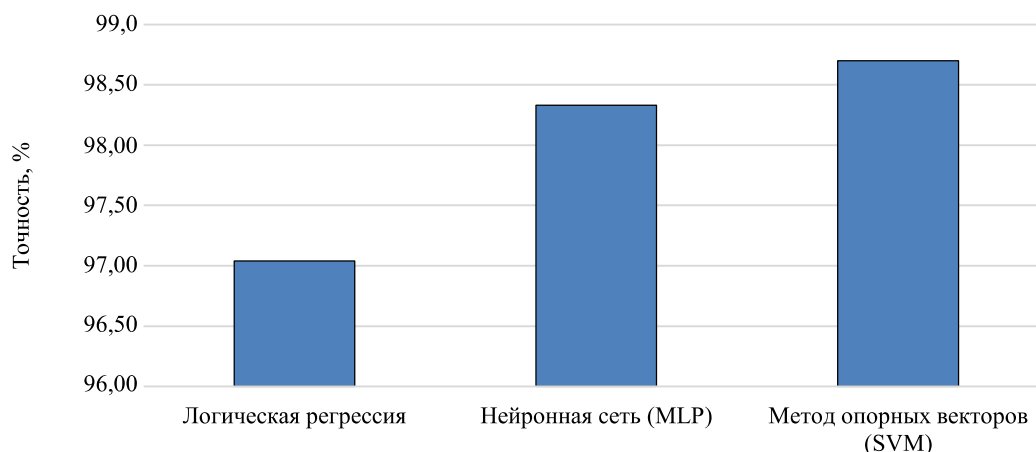


Рис. 7. График сравнения точности разных моделей

Fig. 7. Graph comparing the accuracy of different models

Эксперимент с использованием предложенного подхода показал увеличение точности классификации на 20 % по сравнению с традиционными методами. В частности, при использовании сверточных нейронных сетей для классификации изображений удалось достичь более высокой точности и снизить количество ошибок [6].

Преимущества предложенного метода VisNormNet включают высокую точность, автоматизацию процесса обработки и анализ данных на выборках Iris, MNIST и CIFAR-10. В среднем отмечается прирост в 12 % к точности по сравнению с логистической регрессией и 7 % к нейронной сети без визуальной трансформации. Адаптивная трансформация показала высокую устойчивость к смещению распределений и частичным пропущенным значениям. Однако существует ряд ограничений, таких как потребность в большом объеме обучающих данных и высокие вычислительные ресурсы для обучения моделей. Несмотря на это, предложенный метод демонстрирует отличные результаты и может быть применен для решения множества практических задач.

Научная новизна предложенного метода заключается:

- в формализации процедуры выбора нормализации на основе визуальных и корреляционных паттернов;
- введении этапа визуального эмбединга как промежуточного представления признаков;
- построении мета-модели на основе этих эмбедингов, что улучшает переносимость модели между доменами.

Заключение

Предложен и экспериментально обоснован интеллектуальный метод VisNormNet, ориентированный на анализ и интерпретацию сложных и разно-

родных данных. Метод объединяет процедуры визуализации, нормализации и гибридного машинного обучения, обеспечивая высокую точность, устойчивость к искажениям и снижение влияния человеческого фактора.

В отличие от традиционных решений VisNormNet формализует этап выбора метода трансформации данных, адаптируя его к внутренней корреляционной структуре признаков. Такой подход обеспечивает более корректную подготовку данных к анализу, что подтверждено улучшением метрик качества на нескольких тестовых выборках.

Проведенное исследование позволило разработать, реализовать и экспериментально оценить новый метод интеллектуального анализа данных VisNormNet. В результате получены следующие научные и практические результаты, которые могут быть взяты за основу при построении интеллектуальных систем поддержки принятия решений в условиях неполных, шумных или высокоразмерных данных:

- предложен новый гибридный метод VisNormNet, интегрирующий этапы визуализации, адаптивной нормализации и машинного обучения при обработке данных;
- сравнительный анализ проводился на эталонных наборах данных Iris, MNIST и CIFAR-10, отражающих различные типы и сложности задач классификации, в ходе эксперимента доказано улучшение предложенного метода по сравнению с традиционными алгоритмами, гибридная архитектура VisNormNet на основе статистики показала значимое увеличение точности классификации: на 12 % по сравнению с логической регрессией, в среднем на 7 % по сравнению с изолированно обученной нейронной сетью без адаптивной предобработки. Наибольшая эффективность метода выявлена на

сложных, высокоразмерных данных CIFAR-10, где прирост точности достиг 20 % по сравнению с базовыми подходами.

На тестовых наборах данных (Iris, MNIST, CIFAR-10) гибридная архитектура VisNormNet показала статистически значимое увеличение точности классификации. Ключевым отличием от существующих аналогов является корреляционно-ориентированный выбор процедуры нормализации, что позволяет учитывать внутреннюю структуру данных и минимизировать субъективный фактор на этапе предобработки, включая частичные шумы и пропуски, наличие выбросов и смещения распределений, благодаря адаптивному выбору нормализации на основе кластерного анализа признаков и применению устойчивых методов масштабирования. Практическая апробация на корпоративных данных HR-аналитики (850 сотрудников) подтвердила не только высокую точность прогнозирования, но и интерпретируемость результатов через визуа-

лизацию корреляционных взаимосвязей, что обеспечивает прозрачность принятия решений.

К ограничениям метода отнесена повышенная вычислительная сложность обучения, обусловленная использованием глубоких нейронных сетей.

Перспективы дальнейших исследований включают:

- разработку автоматизированного механизма выбора модели в зависимости от профиля выборки;
- масштабирование метода на потоковые и временные данные;
- адаптацию для потоковых и временных данных, а также внедрение в области промышленного IoT, кибербезопасности и предиктивной аналитики в здравоохранении.

Таким образом, работа вносит вклад в развитие методологии анализа данных, предлагая практические решения для повышения эффективности обработки информации в условиях цифровой трансформации.

Список источников

1. Петров В. В. Теория и практика обработки больших данных. М.: Науч. книга, 2020. С. 10–25.
2. Иванов Д. А. Визуализация данных: методы и практические приложения. СПб.: Питер, 2018. С. 45–67.
3. Иванов И. В. Введение в обработку данных с использованием нейронных сетей. М.: Науч. мир, 2015. С. 42–57.
4. Zhao Y., Li J. *Machine Learning for Data Analysis*. Springer, 2019. P. 9–21.
5. Chen T., Guestrin K. XGBoost: A Scalable Tree Boosting System // *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016. P. 785–794.

6. Goodfellow I., Bengio Y., Courville A. *Deep Learning*. MIT Press, 2016. P. 2–5.
7. Olah C. *Understanding LSTM Networks*. URL: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> (дата обращения: 06.02.2025).
8. Кормен М., Лейзерсон Ч., Ривест Р. *Алгоритмы: построение и анализ*. М.: Вильямс, 2017. С. 325–350.
9. Стоянов А. С. Применение алгоритмов машинного обучения в анализе временных рядов. М.: Науч. прогресс, 2019. С. 35–42.
10. Chaurasia V., Pal P. *A Comprehensive Study of Data Transformation and Feature Engineering in Machine Learning* // *Procedia Computer Science*. 2017. T. 167. C. 1170–1177.

References

1. Petrov V. V. *Teoriia i praktika obrabotki bol'shikh dannykh* [Theory and practice of big data processing]. Moscow, Nauchnaia kniga Publ., 2020. Pp. 10-25.
2. Ivanov D. A. *Vizualizatsiia dannykh: metody i prakticheskie prilozheniia* [Data visualization: methods and practical applications]. Saint Petersburg, Piter Publ., 2018. Pp. 45-67.
3. Ivanov I. V. *Vvedenie v obrabotku dannykh s ispol'zovaniem neuronnykh setei* [Introduction to data processing using neural networks]. Moscow, Nauchnyi mir Publ., 2015. Pp. 42-57.
4. Zhao Y., Li J. *Machine Learning for Data Analysis*. Springer, 2019. Pp. 9-21.
5. Chen T., Guestrin K. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016. Pp. 785-794.

6. Goodfellow I., Bengio Y., Courville A. *Deep Learning*. MIT Press, 2016. Pp. 2-5.
7. Olah C. *Understanding LSTM Networks*. Available at: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> (accessed: 06.02.2025).
8. Kormen M., Leizeron Ch., Rivest R. *Algoritmy: postroyenie i analiz* [Algorithms: construction and analysis]. Moscow, Vil'iams Publ., 2017. Pp. 325-350.
9. Stoianov A. S. *Primenenie algoritmov mashinnogo obucheniia v analize vremennykh riadov* [Application of machine learning algorithms in time series analysis]. Moscow, Nauchnyi progress Publ., 2019. Pp. 35-42.
10. Chaurasia V., Pal P. *A Comprehensive Study of Data Transformation and Feature Engineering in Machine Learning*. *Procedia Computer Science*. 2017. V. 167. Pp. 1170-1177.

Статья поступила в редакцию 20.04.2025; одобрена после рецензирования 11.09.2025; принята к публикации 23.10.2025
The article was submitted 20.04.2025; approved after reviewing 11.09.2025; accepted for publication 23.10.2025

Информация об авторе / Information about the author

Ангелина Игоревна Дубровина — доцент кафедры кибербезопасности информационных систем; Донской государственный технический университет; ministrelia69@yandex.ru

Angelina I. Dubrovina — Assistant Professor of the Department of Cybersecurity of Information Systems; Don State Technical University; ministrelia69@yandex.ru

