

МИНОБРНАУКИ РОССИИ

Федеральный исследовательский центр
«Информатика и управление» Российской академии наук

Национальный комитет при Президиуме Российской академии наук
по распознаванию образов и анализу изображений

Институт информационных технологий Министерства науки и образования
Азербайджанской Республики

Институт проблем передачи информации им. А.А. Харкевича
Российской академии наук

Национальный исследовательский Томский государственный университет

Федеральный исследовательский центр «Карельский научный центр РАН»

Ошский государственный университет

Daqing Normal University

ОПТИКО-ЭЛЕКТРОННЫЕ ПРИБОРЫ И УСТРОЙСТВА В СИСТЕМАХ РАСПОЗНАВАНИЯ ОБРАЗОВ И ОБРАБОТКИ ИЗОБРАЖЕНИЙ

РАСПОЗНАВАНИЕ – 2025

Сборник материалов XVIII Международной
научно-технической конференции

9 – 12 сентября 2025 года

Редакционная коллегия:

С. Г. Емельянов, В. С. Титов (отв. ред.),

Э. И. Ватутин, В. С. Панищев,

Т. А. Ширабакина, С. Н. Епишева

Курск 2025

УДК 621.383.68.3:681.785(063)

ББК В 338.4я431

О 66

Рецензент

Доктор технических наук, профессор *А.В. Кониченко*

Редакционная коллегия:

С. Г. Емельянов, доктор технических наук, профессор

В. С. Титов, доктор технических наук, профессор (отв. ред.)

Т. А. Ширабакина, кандидат технических наук, профессор

Э. И. Ватутин, доктор технических наук, доцент

В. С. Панищев, кандидат технических наук

С. Н. Епишева, кандидат технических наук

О 66 **Оптико-электронные приборы и устройства в системах распознавания образов и обработки изображений. Распознавание – 2025:** сборник материалов XVIII Международной научно-технической конференции, 9–12 сентября 2025 года / ред. кол.: С. Г. Емельянов [и др.]; Минобрнауки России, Юго-Западный гос. ун-т. – Курск: ЮЗГУ, 2025. – 302 с.

ISBN 978-5-7681-1736-8

Сборник содержит материалы XVIII Международной научно-технической конференции «Оптико-электронные приборы и устройства в системах распознавания образов и обработки изображений», целью которой является ознакомление с имеющимися достижениями по созданию оптико-электронных приборов, систем и внедрению информационных технологий в научные исследования, учебный процесс и промышленность, а также координация по эффективному их применению в системах распознавания образов и обработки изображений.

Сборник предназначен для научных сотрудников, преподавателей, аспирантов и студентов вузов.

Издание осуществлено с авторских оригиналов. Редакция не несет ответственности за ошибки авторов.

Материалы для публикации одобрены программным комитетом Международной научно-технической конференции.

УДК 621.383.68.3:681.785(063)

ББК В 338.4я431

ISBN 978-5-7681-1736-8

© Юго-Западный государственный университет, 2025

СОДЕРЖАНИЕ

ВСТУПИТЕЛЬНОЕ СЛОВО	12
Bistability and hysteresis transitions in a system with unipolar pulse-width modulated control <i>Abdirasulov A. Z., Sopuev U. A., Kolomiets E. A., Tsukanov D. Yu.</i>	13
Sentiment analysis of WhatsApp data for mobile forensics	15
<i>Aghababayev R. R.</i>	
Developing a virtual machine based programming language for intelligent microcontroller systems	17
<i>Dashdamirli N. F.</i>	
Analysis of the migration of azerbaijani researchers and students based on ORCID profile data	19
<i>Hajirahimova M. Sh., Aliyeva A. S.</i>	
Algorithm for predicting hepatocellular carcinoma based on a national database	21
<i>Jabrayilova Z. G., Garayeva L. A.</i>	
Detection of incidents that occur during the operation of the software system	23
<i>Mahmudova Sh. J.</i>	
Ontological model of patient-centered medical digital twin	26
<i>Mammadova M. H., Ahmadova A. A.</i>	
Sentiment analysis of information from social media resources for determining public opinion on the effectiveness of drugs	28
<i>Mammadova M. H., Jabrayilova Z. G., Shikhaliyeva N. R.</i>	
Random bond ising model based CNN FPGA accelerator	30
<i>Usatyuk V. S., Egorov S. I.</i>	
Construction of the spectra of intercalates number in the Latin squares of orders 18–28 using volunteer distributed computing in the boinc platform	33
<i>Vatutin E. I., Jia Wei-Ting, Jun Chi Ma, Qiang Miao, Manzyk M. O., Kukushkina N. N., Kurochkin I. I., Albertyan A. M.</i>	
Причинно-следственный анализ событий кибербезопасности в сетях облачных систем	35
<i>Абдуллаева Ф. Д., Сулейманзаде С. Н.</i>	
Подход для обнаружения киберугроз в контексте обработки естественного языка для защиты персональных данных	37
<i>Алгулиев Р. М., Алыгулиев Р. М., Сухостат Л. В.</i>	
Масштабирование данных как ключевой этап предварительной обработки	39
<i>Алекперова И. Я.</i>	
Разработка программного обеспечения для распознавания объектов на спутниковых снимках с использованием сверточных нейронных сетей	42
<i>Александров И. К., Григорьевых Е. А.</i>	
Коллиматорное устройство для контроля систем технического зрения	43
<i>Алимов А. Е.</i>	

Список литературы

1. Rahman M. R., Hezaveh R. M., Williams L. What are the attackers doing now? Automating cyberthreat intelligence extraction from text on pace with the changing threat landscape: A survey // ACM Computing Surveys. 2023. Vol. 55, no. 12. P. 1–36.
2. A review on cyber security named entity recognition / C. Gao, X. Zhang, M. Han, H. Liu // Frontiers of Information Technology & Electronic Engineering. 2021. Vol. 22, no. 9. P. 1153–1168.
3. XLNet: Generalized autoregressive pretraining for language understanding / Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, Q. V. Le // ArXiv.org. 2019. URL: <https://arxiv.org/1906.08237> (accessed 07.05.2025). DOI 10/48550/arXiv.1906.08237.

УДК 004.9:314/316

Масштабирование данных как ключевой этап предварительной обработки

И. Я. Алекперова

Институт информационных технологий, Баку, Азербайджан

airada.09@gmail.com

Аннотация. Рассматривается роль масштабирования данных в процессе предварительной обработки. Анализируются основные методы масштабирования, их влияние на эффективность алгоритмов машинного обучения.

Ключевые слова: масштабирование; машинное обучение; обработка данных; нормализация данных.

Предварительная обработка данных является неотъемлемой частью процесса анализа данных и построения моделей машинного обучения. Масштабирование данных – один из ключевых этапов предварительной обработки, направленный на приведение признаков к единому масштабу, что позволяет улучшить качество обучения моделей и повысить их точность. Существует множество методов масштабирования и нормализации данных. Выбор метода зависит от конкретной задачи и характеристик данных. В машинном обучении масштабирование – это метод, используемый для преобразования признаков набора данных в общий масштаб [1].

Для масштабирования данных часто используемые скалеры: L1 Scaler, L2 Scaler, StandardScaler и Min-MaxScaler. Выбор метода масштабирования зависит от конкретных характеристик набора данных и используемого алгоритма. L1 Scaler делит каждую функцию на сумму ее абсолютных значений:

$$X_{scaled} = \frac{X}{\sum(|X|)},$$

где X_{scaled} – масштабированное значение признака; X – исходное значение признака; $|X|$ – абсолютные значения элементов в X .

Такой подход гарантирует, что сумма абсолютных значений всех функций в масштабированном наборе данных равна 1. Метод применяется, когда наличие или отсутствие признаков важнее их величины. Например, в классификации текстов, где встречаемость слова важнее его частоты.

L2 Scaler также называют «евклидова норма» (EuclideanNorm). Он делит каждую функцию на квадратный корень суммы ее квадратов значений. Это гарантирует, что сумма квадратов значений всех функций в масштабированном наборе данных равна 1, применяется, когда важны как наличие, так и величина признаков (при классификации изображений и т. д.):

$$X_{scaled} = \frac{X}{\sqrt{\sum(X^2)}}.$$

StandardScaler масштабирует признаки так, чтобы среднее значение было равно 0, а дисперсия равна 1. Обычно используется во многих алгоритмах машинного обучения, особенно тех, которые предполагают нормально распределенные признаки:

$$X_{scaled} = \frac{X - X_{mean}}{d_{std}},$$

где X_{mean} – среднее значение признака в наборе данных; d_{std} – стандартное отклонение признака в наборе данных.

Min-MaxScaler масштабирует объекты до определенного диапазона, обычно от 0 до 1. Подходит для нейронных сетей и других алгоритмов, требующих, чтобы входные характеристики находились в определенном диапазоне [2]:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}},$$

где X_{min} – минимальное значение признака в наборе данных; X_{max} – максимальное значение признака в наборе данных.

Чтобы ограничивать значения признаков определенным диапазоном, отдаем предпочтение методу нормализации по диапазону, т. к. при данном методе не требуется нормализовать целевую переменную, если она является категориальной:

$$X_{scaled} = \frac{a + (X - X_{min}) \cdot (b - a)}{(X_{max} - X_{min})},$$

где a – нижняя граница нового диапазона; b – верхняя граница нового диапазона.

Метод нормализации по диапазону применяется, когда необходимо ограничить значения признаков определенным диапазоном. Метод подобен минимаксной нормализации, но позволяет задать произвольный диапазон.

Многие алгоритмы машинного обучения, такие как логистическая регрессия и нейронные сети, предполагают, что признаки распределены нормально. Стандартизация помогает реализовать это предположение [3]. Признаки с разным масштабом могут оказывать непропорциональное влияние на модель. Стандартизация гарантирует, что все функции в равной степени способствуют процессу обучения. Это важный инструмент в статистике и машинном обучении. Стандартизация позволяет нам сравнивать данные из разных источников, улучшать работу алгоритмов и делать более обоснованные выводы. Формула стандартизации

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

где X – исходное значение признака; μ – среднее значение независимых признаков; σ – стандартное отклонение независимых признаков.

Многие алгоритмы машинного обучения, особенно те, которые используют метрические расстояния (например, k -средние, SVM), чувствительны к масштабу данных. Если один признак имеет значительно больший диапазон значений чем другие, то он будет доминировать в расчетах расстояний. Масштабирование и нормализация данных позволяют улучшить производительность алгоритмов и повысить точность моделей.

Список литературы

1. Integrating machine learning and sentiment analysis in movie recommendation systems / A. M. Sarhan, H. Ayman, M. Wagdi, B. Ali, A. Adel, R. Osama // Journal of Electrical Systems and Information Technology. 2024. Vol. 11, no. 1. P. 53.
2. An aseptic approach towards skin lesion localization and grading using deep learning and harrishawks optimization / H. M. Balaha, A. E. S. Hassan, E. M. El-Gendy, H. ZainEldin, M. M. Saafan // Multimedia Tools and Applications. Springer: Berlin, 2023. P. 1–29.
3. Maharana K., Mondal S., Nemade B. A review: Data pre-processing and data augmentation techniques // Global Transitions Proceedings. 2022. Vol. 3, no. 1. P. 91–99.