

ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

COMPUTER SCIENCE, COMPUTER ENGINEERING AND CONTROL

УДК 004.023

doi: 10.21685/2072-3059-2025-3-1

Математическая модель интегрального критерия классификации и алгоритм формирования его графического представления

А. С. Бождай¹, Л. Н. Горшенин²

^{1,2}Пензенский государственный университет, Пенза, Россия

¹bozhday@yandex.ru, ²gorshenin.lev@gmail.com

Аннотация. *Актуальность и цели.* Важнейшим методом научного исследования является классификация. Особую сложность представляет задача классификации объектов с гетерогенным пространством информационных признаков, которая на данный момент не имеет унифицированного решения. При этом важно учитывать зависимость результатов классификации от выбранной точки зрения на предметную область – критерия, который определяет состав и типологию значимых признаков рассматриваемых объектов. Целью работы является математическая и алгоритмическая формализация процесса выработки критериев классификации, а также разработка единого формата представления информационных признаков объекта, независимо от их типа, состава и полноты. Это позволит свести задачу классификации к единому автоматизированному процессу – хорошо управляемому, наглядному, поддающемуся машинному обучению и инвариантному к специфике предметной области. *Материалы и методы.* Раскрываются основные проблемы традиционных подходов к решению задачи классификации в условиях динамичности точек зрения на предметную область. Методологической основой работы является теория множеств, методы машинного обучения, нейронные сети, растровая компьютерная графика. *Результаты.* Предложены концепция и формализованное теоретико-множественное представление интегрального критерия классификации, позволяющего алгоритмизировать выбор точек зрения на исследуемую предметную область. Критерии классификации определяют состав значимых признаков объекта. Для его унифицированного представления разработан единый формат в виде растрового графического изображения – графо-хроматической карты, а также алгоритм формирования таких карт. *Выводы.* Предлагаемый подход существенно упрощает процедуру распределения гетерогенных объектов по классам за счет ее приведения к хорошо изученной и отработанной задаче машинного обучения – классификации растровых графических изображений. Формализация интегрального критерия классификации позволяет быстро и удобно задавать различные точки зрения на предметную область, представляя их в виде кортежей данных. Алгоритм формирования графо-хроматических карт обеспечивает получение растровых изображений единого размера и формата для подачи на вход

предобученного нейросетевого классификатора. При этом при изменении точки зрения на предметную область (при изменении числа и состава значимых признаков объектов) не потребуется менять структуру классификатора или производить его переобучение.

Ключевые слова: классификация, методы классификации, критерий классификации, гетерогенные данные, нейронные сети, графо-хроматическая карта, растровое изображение

Для цитирования: Бождай А. С., Горшенин Л. Н. Математическая модель интегрального критерия классификации и алгоритм формирования его графического представления // Известия высших учебных заведений. Поволжский регион. Технические науки. 2025. № 3. С. 5–16. doi: 10.21685/2072-3059-2025-3-1

Mathematical model of the integral classification criterion and the algorithm for forming its graphical representation

A.S. Bozhday¹, L.N. Gorshenin²

^{1,2}Penza State University, Penza, Russia

¹bozhday@yandex.ru, ²gorshenin.lev@gmail.com

Abstract. *Background.* The most important method of scientific research is classification. The task of classifying objects with a heterogeneous space of information features, which currently has no unified solution, is particularly difficult. It is important to take into account the dependence of classification results on chosen point of view on subject area – a criterion that determines composition and typology of significant features of objects under consideration. The purpose of the work is the mathematical and algorithmic formalization of the process of developing classification criteria, as well as development of a unified format for presenting information features of an object regardless of their type, composition and completeness. This will reduce classification task to a single automated process - well-managed, visual, amenable to machine learning and invariant to specifics of subject area. *Materials and methods.* The work reveals the main problems of traditional approaches to solving the classification problem, in context of dynamic points of view on subject area. Existing works devoted to formalization of classification criteria and problems of choosing significant information features of objects are considered. The methodological basis of the work is set theory, machine learning methods, neural networks and raster computer graphics. *Results.* As a result, a concept and formalized set-theoretic representation of the integral classification criterion are proposed. The criterion allows algorithmization of choice of view's points on the studied subject area. Classification criteria determine composition of significant features of the object. For its unified representation, a single format in the form of a raster graphic image - a graphic-chromatic map, as well as an algorithm for creation of such maps have been developed. *Conclusions.* The proposed approach significantly simplifies the procedure of distributing heterogeneous objects into classes by reducing it to a well-studied and proven task of machine learning - classification of raster graphic images. Formalization of the integral classification criterion allows quickly and conveniently set different points of view on the subject area, representing them in the form of data tuples. The algorithm for forming grapho-chromatic maps provides obtaining raster images of a single size and format for feeding to the input of a pre-trained neural network classifier. At the same time, when changing the point of view on the subject area (i.e. when changing the number and composition of significant features of objects), there is no need to change the structure of the classifier or retrain it.

Keywords: classification, classification methods, classification criterion, heterogeneous data, neural networks, graph-chromatic map, raster image

For citation: Bozhday A.S., Gorshenin L.N. Mathematical model of the integral classification criterion and the algorithm for forming its graphical representation. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki* = *University proceedings. Volga region. Engineering sciences*. 2025;(3):5–16. (In Russ.). doi: 10.21685/2072-3059-2025-3-1

Введение

Современные научные исследования характеризуются постоянным усложнением и детализацией предметных областей. Это порождает потребность постоянного изменения точки зрения на исследуемые объекты, чтобы охватить и учесть все аспекты изучаемых процессов. Поскольку основным научным методом является классификация, то высокая динамика точек зрения исследователя приводит к принципиальным трудностям распределения объектов с вариативным гетерогенным набором признаков по заранее заданным классам [1–3]. К таким объектам можно отнести: профили пользователей и сообществ в социальных сетях, электронные документы, товары в интернет-магазинах, объекты наблюдения средствами машинного зрения и т.д. Например, классификация документов является одной из ключевых задач для реализации систем электронного документооборота [4]. Задача классификации профилей и сообществ в Интернете актуальна для таргетированной рекламы [5], борьбы с киберугрозами [6], предотвращения преступлений [7]. Классификация товаров требуется в таможенных целях [8], для определения наиболее выгодных рынков сбыта [9]. Классификация наблюдаемых объектов (например, летательных аппаратов) может быть необходима для распознавания целей, радарного мониторинга, противовоздушной обороны [10, 11].

Проблема классификации не имеет однозначного формализованного решения, так как сильно зависит не только от текущего набора значимых признаков объектов, который может изменяться со временем, но и в целом от точки зрения на предметную область. Например, для классификации профилей пользователей по политическим предпочтениям и по социальной активности имеют значение различные наборы признаков. То есть значимость тех или иных признаков сложного объекта зависит от целей классификации. Кроме того, исследователи часто пользуются различными эмпирическими подходами с множеством итераций, на каждой из которых могут обнаруживаться более важные критерии распределения объектов по классам. Изменение набора признаков, по которым проводится классификация, в большинстве случаев требует переобучения классификаторов.

Важной задачей здесь представляется формализация понятия «точка зрения на предметную область». Точка зрения исследователя (фокус его внимания на проблему) может складываться из множества разнородных и зачастую противоречивых факторов. При этом важность отдельных, простых и составных, факторов может постоянно меняться, особенно в сложных предметных областях с огромным количеством целевых состояний и гетерогенных информационных признаков. Для обозначения такой точки зрения, определяющей цели классификации и состоящей из набора признаков, имеющих определенную значимость, предлагается ввести термин «интегральный критерий классификации» (ИКК). Теоретико-множественное представление ИКК позволит разработать аппарат для свертки разнородных информационных признаков объектов предметной области в единое электронное представле-

ние. Однако такое математическое представление плохо подходит для нейросетевых классификаторов, когда нейросеть обучается на основе входных массивов данных. При существенном изменении объема и характера данных (а это неизбежно будет происходить при изменении точки зрения на предметную область), потребуется постоянно переобучать сеть и даже менять ее структуру.

Преодолеть эту проблему возможно путем преобразования разнородной совокупности значимых признаков к единому универсальному представлению, который был ранее описан авторами в работе [12]. В качестве такого универсального представления предложено использовать формат растрового изображения, называемый графо-хроматической картой (ГХК) и состоящий из множества графических примитивов с управляемыми характеристиками. Это позволит решать задачу классификации унифицированным способом, не зависящим от природы классифицируемых объектов, а также от типа и состава их признаков.

Материалы и методы

Существует несколько работ, затрагивающих вопросы формализации критериев классификации, а также проблемы выбора значимых информационных признаков объекта.

Работа [13] американского ученого R. Stepp посвящена «целесоориентированному» подходу к кластеризации и классификации объектов. Модель целей классификации представляется в виде дерева (GDN – Goal Dependency Network) от основных целей к подцелям. На основе этого дерева целей определяется целевой набор классов и критерии для проведения классификации.

Оригинальный подход к классификации, основанный на «точках зрения» изложен в работе [14] B. Selic и A. Pierantonio. Авторы упоминают следующие ограничения традиционной модели классификации:

- сложность учета различных «точек зрения» на классифицируемые сущности;
- сложность учета динамически изменяющегося набора признаков объектов.

В работе [14] «точка зрения» представлена в виде совокупности сведений о рассматриваемых объектах и о классах. Каждый класс имеет информацию о наборе признаков и о множестве значений этих признаков, которые должен иметь объект, чтобы принадлежать этому классу. Каждый объект имеет набор информационных признаков и может рассматриваться с разных точек зрения. UML-диаграмма предлагаемой модели классификации приведена на рис. 1.

Одним из недостатков такой модели является неустранимая проблема сложности процесса классификации при наличии у объектов гетерогенного пространства информационных признаков.

Из всего множества признаков объекта для классификации могут быть важны только некоторые признаки. Из всех признаков рационально выбрать подмножество наиболее значимых для конкретных целей классификации.

Методы выбора признаков для классификации, такие как фильтрующие (filter), обертки (wrapper) и встроенные (embedded) методы, рассмотрены в работах [15, 16].



Рис. 1. Метамоделю подхода к классификации, основанного на «точках зрения»
(перевод на русский на основе англоязычной модели из работы [14])

Особенность фильтрующих методов заключается в обработке отдельных признаков. Их значимость определяется через корреляцию с целевой переменной, но при этом не учитываются взаимосвязи между различными признаками и специфика классификатора. Кроме того, требуются доказательства гипотез зависимости целевой переменной от каждого признака в отдельности.

Методы-обертки рассматривают различные подмножества признаков, оценивая групповое влияние признаков на качество классификации. Недостатком этих методов является резкое возрастание временных и ресурсных затрат при увеличении числа признаков [15].

Embedded-методы встраиваются в структуру классификатора. Это позволяет значительно сократить временные затраты по сравнению с методами-обертками. Но при этом существенно усложняется сам классификатор из-за необходимости встраивания в его структуру дополнительного функционального аппарата выбора признаков.

Для улучшения существующих подходов к выбору признаков и самого процесса классификации предлагается разработать такую модель ИКК, которая будет представлять информационные признаки самой разной природы в одном универсальном формате, инвариантном их типу и количеству. В качестве такого представления предлагается графический растровый формат. Это позволит использовать для классификации сверточные нейронные сети, имеющие множество хорошо зарекомендованных на практике архитектур [17]. Кроме того, можно будет использовать уже обученные модели, дообучив их на предварительно сформированных обучающих наборах ГХК. При этом графический формат достаточно информативен и компактен, что позволит даже на относительно небольших изображениях отображать сведения о

значительном количестве гетерогенных признаков. Для построения модели ИКК используется теоретико-множественный подход, позволяющий группировать любые совокупности объектов различной природы с точки зрения общности их свойств. В частности, используются следующие объекты теории множеств:

- 1) кортеж – конечная упорядоченная последовательность элементов, где каждый элемент принадлежит некоторому соответствующему множеству;
- 2) вектор – частный случай кортежа, в котором все элементы принадлежат одному и тому же множеству [18].

Результаты и обсуждение

Поскольку ИКК может включать произвольное число признаков, предлагается представить информацию о значимых признаках в виде вектора элементов. Следовательно, кардинальное число этого вектора будет равно числу таких признаков. Признаки могут состоять из одного или нескольких значений одного типа, поэтому каждый элемент вектора представляется вложенным вектором значений отдельного признака.

Ввиду того, что одних значений недостаточно для проведения классификации, каждый элемент вектора должен содержать информацию и о типе признака. Так как состав признаков может меняться, то для гарантированного различения классификатором значений разных признаков каждый элемент вектора должен иметь некоторый идентификатор признака. Также идентификатор признака позволит получать служебную информацию о признаке, необходимую для генерации ГХК (например, диапазон допустимых значений или путь к файлам библиотеки графических примитивов). Поскольку в ИКК должны учитываться и цели классификации, элементы вектора должны содержать информацию о степени значимости признака для классификации. Следовательно, каждый элемент вектора признаков должен представлять собой четырехэлементный кортеж: *<Идентификатор признака, Тип признака, Степень значимости, Вложенный вектор значений>*.

В качестве множества значений идентификатора признака выбрано множество значений guid (128-битные номера, используемые для уникальной идентификации ресурсов) [19]. Для удобства дальнейшего растрово-графического представления ИКК в 256-тоновой градации в качестве множества значений степеней значимости выбран целочисленный отрезок {0, ..., 255}. Элементы в векторе значений представлены объектами произвольной природы, множество допустимых значений которых зависит от типа признака.

Так как для преобразования информации о признаке в графический формат должен существовать и быть реализован соответствующий алгоритм, множество поддерживаемых типов ограничено и конечно. Базовыми элементами этого множества являются: численный тип, категориальный тип, текстовый тип.

Таким образом, теоретико-множественный вид информации о значимых признаках объекта, подлежащего классификации, можно представить следующими выражениями:

$$I = \text{Guid} , \quad (1)$$

$$T = \{\text{Numerical}, \text{Categorical}, \text{Text}, \dots\} , \quad (2)$$

$$S = [0, 255], \quad (3)$$

$$V = (R_1 \dots R_n), n \in N, \quad (4)$$

$$Info = (E_1 \dots E_n), n \in N, E \in I \times T \times S \times V, \quad (5)$$

где I – множество идентификаторов; $Guid$ – множество значений глобальных идентификаторов; T – множество поддерживаемых типов; S – множество значений степеней значимостей; V – множество значений признака; R – множество действительных чисел; N – множество натуральных чисел; E – признак объекта (4-элементный кортеж); $Info$ – совокупная информация о признаках объекта.

Следующий шаг – свертка выражения (5) в единое универсальное представление ИКК (в растровое графическое представление). При этом механизм свертки должен зависеть от типа признака и учитывать семантику признака, его положение в векторе и важность его для классификации. Следовательно, реализация свертки подразумевает, что для каждого типа признаков нужно иметь отдельный алгоритм преобразования значения в графическое представление. Некоторые из таких алгоритмов уже описаны авторами в работе [20].

Таким образом, ГХК будет формироваться путем последовательного применения алгоритмов графического преобразования векторов признаков в соответствии с их типом. На этапе построения ГХК целесообразно использовать векторную графику и лишь на завершающем этапе выполнять преобразование в растровый формат.

На вход каждый из алгоритмов должен принимать элемент вектора с информацией о признаке. Возвращаемое значение должно содержать отображение информации о признаке на графическую форму. Из соображений универсальности отображение выполняется с помощью одного или нескольких графических примитивов для каждого значения с информацией об их относительной позиции и масштабе. Таким образом, на выходе каждого из алгоритмов будет получен вектор трехэлементных кортежей следующего вида: *<Графический примитив (в векторном формате), Относительная позиция примитива, Масштаб примитива>*.

При этом степень значимости признака будет передаваться через прозрачность примитивов, соответствующих значениям признака.

Алгоритм формирования графического представления ИКК создает пустое полотно черного цвета определенных размеров. Затем он последовательно наносит на полотно примитивы из трехэлементных кортежей с учетом указанных относительных позиций и масштабов. При этом опорная позиция на полотне для нанесения примитивов определенного признака определяется его положением в векторе с информацией о признаках. После нанесения всех графических примитивов на полотно оно преобразуется в растровый формат. Блок-схема алгоритма приведена на рис. 2.

Таким образом, результатом работы алгоритма является графо-хроматическая карта, представляющая собой универсальное графическое представление ИКК.

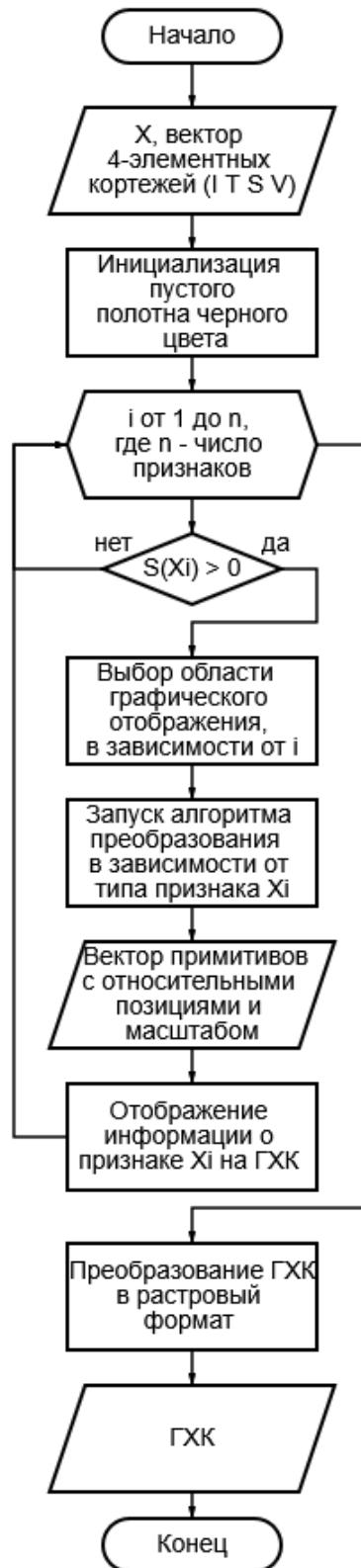


Рис. 2. Блок-схема алгоритма формирования ГХК

Математическое представление ИКК выражается формальными выражениями:

$$R = (Primitive, Pos, Scale), \quad Pos \in N^2, \quad Scale \in R, \quad (6)$$

$$Result = (R_1 \dots R_n), n \in N, \quad (7)$$

$$S(Info_n) \rightarrow Result, n \in N, \quad (8)$$

$$SpecAlg = \{S_1 \dots S_t\}, t \in T, \quad (9)$$

$$TotalAlg(Info, SpecAlg) \rightarrow GraphicalICC, \quad (10)$$

где *Primitive* – некоторый графический примитив; *Pos* и *Scale* – относительная позиция и масштаб этого примитива; *S* – алгоритм, преобразующий информацию о признаке определенного типа в графическое представление; *Info* – вектор с информацией о признаках; *SpecAlg* – множество алгоритмов преобразований признаков разных типов; *T* – множество поддерживаемых типов признаков; *GraphicalICC* – графо-хроматическая карта; *TotalAlg* – алгоритм формирования графического представления ИКК (графо-хроматической карты) [20].

В дальнейшем для проверки качества классификации с использованием ИКК планируется реализовать прототип системы классификации на языке C# в среде Visual Studio 2022. Сведения об объектах будут проходить через несколько этапов: этап предобработки данных, этап формирования растрового представления ГХК, этап классификации полученных ГХК. Для проведения классификации система будет загружать заранее обученные на наборах ГХК модели.

Использование модели графического представления ИКК для классификации позволит свести гетерогенные признаки объекта разного типа и состава в единый формат, что избавит от необходимости различных предварительных преобразований. При этом не требуется менять структуру классификатора при изменении числа признаков или при появлении признака другого типа – в любом случае классификатор принимает на вход изображение определенного размера. Предлагаемый подход позволит более гибко учитывать значимость каждого признака, а также использовать простые сверточные нейронные сети в качестве классификаторов.

Заключение

Предложены концепция и формализованное теоретико-множественное представление интегрального критерия классификации, позволяющего алгоритмизировать выбор точек зрения на исследуемую предметную область. На основе критериев классификации определяется состав значимых признаков объекта. Для его унифицированного представления предлагается использовать единый формат в виде растрового графического изображения – графо-хроматической карты.

Разработан алгоритм формирования графо-хроматических карт, на вход которого подается вектор 4-элементных кортежей (формальное представление ИКК), на выходе формируется вектор 3-элементных кортежей (формальное представление ГХК) с последующим его преобразованием в растровую графическую форму. Далее полученные ГХК объектов могут быть переданы на вход

сверточной нейронной сети для выполнения процедуры классификации, стандартным для растровых изображений образом. При этом ГХК устойчивы к пропускам в данных, что уменьшает требования к их предобработке.

Предложенное формальное представление ИКК универсально и позволяет уменьшить зависимость классификатора от количества признаков, их типа и от точки зрения на классификацию в целом. Это существенно упрощает процедуру распределения гетерогенных объектов по классам за счет ее приведения к хорошо изученной задаче классификации растровых графических изображений.

Таким образом, достигается основная цель исследования – сложная и трудноформализуемая задача классификации сводится к единому автоматизированному процессу – хорошо управляемому, наглядному, поддающемуся машинному обучению и инвариантному к специфике предметной области.

Список литературы

1. Berisha B., Mëziu E. Big Data Analytics in Cloud Computing: An overview // Journal of cloud computing. 2022. № 11 (24). doi: 10.1186/s13677-022-00301-w
2. Lim J. What is data classification? // Alation. URL: <https://www.alation.com/blog/what-is-data-classification/> (дата обращения: 12.04.2024).
3. Yang Y., Wu Y.F., Zhan D.Ch., Liu Zh.B., Jiang Y. Complex object classification: A multi-modal multi-instance multi-label deep network with optimal transport // Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018. P. 2594–2603.
4. Краснянский М. Н., Обухов А. Д., Воякина А. А., Соломатина Е. М. Сравнительный анализ методов машинного обучения для решения задачи классификации документов научно-образовательного учреждения // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2018. № 3. С. 173–182. doi: 10.17308/sait.2018.3/1245
5. Коршунов А., Белобородов И., Гомзин А., Чуприна К., Астраханцев Н., Недумов Я., Турдаков Д. Определение демографических атрибутов пользователей микроблогов // Труды Института системного программирования РАН. 2013. С. 179–194.
6. Куликова А. А. Подход к классификации пользователей в социальных сетях // Восточно-европейский журнал передовых технологий. 2011. № 2. С. 14–18.
7. Станкевич М. А., Игнатьев Н. А., Смирнов И. В., Кисельникова Н. В. Выявление личностных черт у пользователей социальной сети ВКонтакте // Вопросы кибербезопасности. 2019. № 4. С. 80–87. doi: 10.21681/2311-3456-2019-4-80-87
8. Жиряева Е. В., Наумов В. Н. Метод анализа текстов при тарифной классификации товаров в таможенном деле // Программные продукты и системы. 2020. № 3. С. 538–548.
9. Сетлак Г. Использование искусственных нейронных сетей для решения задач классификации в менеджменте // Радіоелектроніка, інформатика, управління. 2004. № 1. С. 127–135.
10. Nima M., Histon J., Dizaji R., Waslander S. L. Feature extraction and radar track classification for detecting UAVs in civilian airspace // 2014 IEEE Radar Conference. Cincinnati, OH, USA, 2014. С. 0674–0679. doi: 10.1109/RADAR.2014.6875676
11. Сенцов А. А., Поляков В. Б., Иванов С. А., Помозова Т. Г. Метод перехвата малоразмерных и малозаметных беспилотных летательных аппаратов // Труды МАИ. 2023. № 21.
12. Бождай А. С., Горшенин Л. Н. Метод классификации объектов с гетерогенным набором информационных признаков, основанный на генерации графохроматических карт // Известия высших учебных заведений. Поволжский регион. Технические науки. 2024. № 3 (71). С. 5–13. doi: 10.21685/2072-3059-2024-3-1 EDN: BGBNIC

13. Stepp R., Michalsky R. S. Conceptual clustering of structured objects: a goal-oriented approach // *AI Journal*. 1986. P. 103–110.
14. Selić B., Pierantonio A. Fixing Classification: A Viewpoint-Based Approach // *Leveraging Applications of Formal Methods, Verification and Validation. ISoLA 2021. Lecture Notes in Computer Science*. 2021. P. 346–356. doi: 10.1007/978-3-030-89159-6_22
15. Tang J., Alelyani S., Liu H. Feature Selection for Classification: A Review // *Data Classification: Algorithms and Applications*. 2014. doi: 10.1201/b17320-3
16. Dash M., Liu H. Feature Selection for Classification // *Intelligent Data Analysis*. 1997. P. 131–156.
17. Rawat W., Wang Z. Deep convolutional neural networks for image classification: A comprehensive review // *Neural Computation*. 2017. № 29. P. 2352–2449.
18. Кортежи и декартово произведение множеств // Региональный финансово-экономический институт. URL: <https://it.rfei.ru/course/~mBme/~6v6/~0pCLCk> (дата обращения: 20.05.2025).
19. Guid // Microsoft. URL: <https://learn.microsoft.com/ru-ru/dotnet/api/system.guid?view=net-8.0> (дата обращения: 10.05.2025).
20. Бождай А. С., Горшенин Л. Н. Разработка алгоритма генерации графо-хроматических карт для решения задач классификации // *Вестник Пензенского государственного университета*. 2024. № 4. P. 22–26. EDN: LOCJQ

References

1. Berisha B., Mëziu E. Big Data Analytics in Cloud Computing: An overview. *Journal of cloud computing*. 2022;11(24). doi: 10.1186/s13677-022-00301-w
2. Lim J. What is data classification? *Alation*. Available at: <https://www.alation.com/blog/what-is-data-classification/> (accessed 12.04.2024).
3. Yang Y., Wu Y.F., Zhan D.Ch., Liu Zh.B., Jiang Y. Complex object classification: A multi-modal multi-instance multi-label deep network with optimal transport. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2018:2594–2603.
4. Krasnyanskiy M.N., Obukhov A.D., Voyakina A.A., Solomatina E.M. A comparative analysis of machine learning methods for solving the problem of document classification in a scientific and educational institution. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Sistemnyy analiz i informatsionnye tekhnologii = Bulletin of Voronezh State University. Series: Systems analysis and information technology*. 2018;(3):173–182. (In Russ.). doi: 10.17308/sait.2018.3/1245
5. Korshunov A., Beloborodov I., Gomzin A., Chuprina K., Astrakhantsev N., Nedumov Ya., Turdakov D. Determining demographic attributes of microblogging users. *Trudy Instituta sistemnogo programmirovaniya RAN = Proceedings of the Institute for System Programming of the Russian Academy of Sciences*. 2013:179–194. (In Russ.)
6. Kulikova A.A. An approach to classifying users in social networks. *Vostochno-evropeyskiy zhurnal peredovykh tekhnologiy = East European Journal of Advanced Technologies*. 2011;(2):14–18. (In Russ.)
7. Stankevich M.A., Ignat'ev N.A., Smirnov I.V., Kisel'nikova N.V. Identifying personality traits in VKontakte users. *Voprosy kiberbezopasnosti = Cybersecurity issues*. 2019;(4):80–87. (In Russ.). doi: 10.21681/2311-3456-2019-4-80-87
8. Zhiryaeva E.V., Naumov V.N. Method of text analysis in tariff classification of goods in customs matters. *Programmnye produkty i sistemy = Software products and systems*. 2020;(3):538–548. (In Russ.)
9. Setlak G. Using artificial neural networks to solve classification problems in management. *Radioelektronika, informatika, upravlinnya = Radio electronics, computer science, management*. 2004;(1):127–135. (In Russ.)
10. Nima M., Histon J., Dizaji R., Waslander S.L. Feature extraction and radar track classification for detecting UAVs in civilian airspace. *2014 IEEE Radar Conference*. 2014:0674–0679. doi: 10.1109/RADAR.2014.6875676

11. Sentsov A.A., Polyakov V.B., Ivanov S.A., Pomozova T.G. Method of intercepting small-sized and stealthy unmanned aerial vehicles. *Trudy MAI = Proceedings of MAI*. 2023;(21). (In Russ.)
12. Bozhday A.S., Gorshenin L.N. A method for classifying objects with a heterogeneous set of information features based on the generation of graphochromatic maps. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki = University proceedings. Volga region. Engineering sciences*. 2024;(3):5–13. (In Russ.). doi: 10.21685/2072-3059-2024-3-1 EDN: BGBNIC
13. Stepp R., Michalsky R.S. Conceptual clustering of structured objects: a goal-oriented approach. *AI Journal*. 1986;103–110.
14. Selić B., Pierantonio A. Fixing Classification: A Viewpoint-Based Approach. *Leveraging Applications of Formal Methods, Verification and Validation. ISoLA 2021. Lecture Notes in Computer Science*. 2021:346–356. (In Russ.). doi: 10.1007/978-3-030-89159-6_22
15. Tang J., Alelyani S., Liu H. Feature Selection for Classification: A Review. *Data Classification: Algorithms and Applications*. 2014. doi: 10.1201/b17320-3
16. Dash M., Liu H. Feature Selection for Classification. *Intelligent Data Analysis*. 1997:131–156.
17. Rawat W., Wang Z. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*. 2017;(29):2352–2449.
18. Tuples and the cartesian product of sets. *Regional'nyy finansovo-ekonomicheskiiy institut = Regional Financial and Economic Institute*. (In Russ.). Available at: <https://it.rfei.ru/course/~mBme/~6v6/~0pCLCk> (accessed 20.05.2025).
19. Guid. *Microsoft*. Available at: <https://learn.microsoft.com/ru-ru/dotnet/api/system.guid?view=net-8.0> (accessed 10.05.2025).
20. Bozhday A.S., Gorshenin L.N. Development of an algorithm for generating graphic-chromatic maps for solving classification problems. *Vestnik Penzenskogo gosudarstvennogo universiteta = Bulletin of Penza State University*. 2024;(4):22–26. (In Russ.). EDN: LOCGJQ

Информация об авторах / Information about the authors

Александр Сергеевич Бождай

доктор технических наук, профессор,
профессор кафедры систем
автоматизированного проектирования,
Пензенский государственный
университет (Россия, г. Пенза,
ул. Красная, 40)

E-mail: bozhday@yandex.ru

Aleksandr S. Bozhday

Doctor of engineering sciences, professor,
professor of the sub-department of computer
aided design systems, Penza State
University (40 Krasnaya street,
Penza, Russia)

Лев Николаевич Горшенин

аспирант, Пензенский
государственный университет
(Россия, г. Пенза, ул. Красная, 40)

E-mail: gorshenin.lev@gmail.com

Lev N. Gorshenin

Postgraduate student, Penza
State University (40 Krasnaya
street, Penza, Russia)

Авторы заявляют об отсутствии конфликта интересов / The authors declare no conflicts of interests.

Поступила в редакцию / Received 30.06.2025

Поступила после рецензирования и доработки / Revised 22.07.2025

Принята к публикации / Accepted 05.09.2025