

На правах рукописи



МИХАЛЕВ ПАВЕЛ АНДРЕЕВИЧ

**УПРАВЛЕНИЕ ДЕЦЕНТРАЛИЗОВАННЫМ ОБМЕНОМ
ИНФОРМАЦИЕЙ ДЛЯ ПРИНЯТИЯ РЕШЕНИЙ
В РАСПРЕДЕЛЕННЫХ СИСТЕМАХ МНОГОЭЛЕМЕНТНОЙ
КЛАССИФИКАЦИИ С НЕПОЛНЫМИ ДАННЫМИ**

Специальность 2.3.1. Системный анализ, управление и обработка информации, статистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Москва – 2026

Работа выполнена в Федеральном государственном казенном военном образовательном учреждении высшего образования «Академия Федеральной службы охраны Российской Федерации».

Научный руководитель: Куцакин Максим Алексеевич, кандидат технических наук, старший преподаватель кафедры информатики и вычислительной техники Федерального государственного казенного военного образовательного учреждения высшего образования «Академия Федеральной службы охраны Российской Федерации», г. Орел, Россия

Официальные оппоненты:

- 1) Басов Олег Олегович, доктор технических наук, доцент, главный инженер по внедрению технологий акционерного общества «АСТ», г. Москва, Россия.
- 2) Жихарев Александр Геннадиевич, доктор технических наук, доцент, профессор кафедры информационно-телекоммуникационных систем и технологий, Федеральное государственное автономное образовательное учреждение высшего образования «Белгородский государственный национальный исследовательский университет» г. Белгород, Россия.

Ведущая организация: Федеральное государственное бюджетное образовательное учреждение высшего образования «Воронежский государственный лесотехнический университет Г.Ф. Морозова» г. Воронеж, Россия.

Защита диссертации состоится 20 мая 2026 г. в 14 часов на заседании диссертационного совета 75.1.026.01 в АО «Научно-исследовательский институт вычислительных комплексов им. М.А. Карцева» по адресу: 117437, г. Москва, ул. Профсоюзная, д. 108.

С диссертацией можно ознакомиться в библиотеке АО «Научно-исследовательский институт вычислительных комплексов им. М.А. Карцева» и на сайте <https://niivk.ru/>

Автореферат разослан « ___ » _____ 2026 г.

Ученый секретарь
диссертационного совета 75.1.026.01
канд. техн. наук, доцент



Е.И. Мутина

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования Развитие систем интеллектуальной обработки данных (ИОД), которые в настоящее время являются основой множества информационных, управляющих и контролируемых систем в различных предметных областях, неразрывно связано с совершенствованием методов, математического, программного, аппаратного обеспечения и технологий машинного обучения (МО) и моделей машинного обучения (ММО). В частности, вопросы МО имеют высокую актуальность для совершенствования систем, реализующих процесс многоклассовой (многоэлементной) классификации и имеющих распределенную архитектуру.

В традиционных системах классификации процесс МО опирается на централизованный сбор, хранение и распределение данных, являющихся обучающими и тестовыми выборками. Однако, постоянно возрастающая сложность этих моделей требует в процессе их обучения все больших объемов анализируемых данных, что существенно усложняет схемы организации центров обработки данных, на базе которых они развертываются. Другим аспектом, усложняющим использование централизованных схем, является потенциальная невозможность консолидации всей совокупности данных обучающих выборок в единой системе хранения. Это может быть связано, как с распределенным характером хранения данных, так и вопросами безопасности их использования в силу конфиденциальности некоторого их подмножества.

Одним из способов решения этой проблемы является использование моделей и методов федеративного МО (ФМО), исследование и разработка которых в последнее время получила активное развитие. В основе функционирования такого вида распределенных систем на основе ФМО лежит взаимодействие множества распределенных вычислительных узлов (worker-nodes), каждый из которых поддерживает собственную локальную модель обучения (ЛМО). Таким образом, сложный процесс обучения декомпозируется на множество простых процессов (реализация стратегии «слабый ученик»), а в дальнейшем производится их агрегирование, в результате которого формируется итоговая (глобальная) модель обучения (ГМО). Эта модель и используется в качестве финальной и применяется для решения прикладных задач. Такой подход наиболее актуален для решения задач многоэлементной классификации, когда множество ЛМО реализуют классификацию для двух и более классов, а ансамблированная ГМО поддерживает задачи классификации по всему множеству классов.

Основными исследовательскими проблемами в предметной области распределенной многоэлементной классификации являются проблемы управления процессом обмена данными выборок локальных

классификаторов в нормальных и специальных условиях, который обеспечивает формирование итогового классификатора.

Тематика диссертационной работы соответствует научному направлению ФГКВООУ ВО «Академия Федеральной службы охраны Российской Федерации» «Повышение эффективности функционирования распределенных вычислительных систем».

Степень разработанности темы Существенный вклад в развитие предметной области методов и алгоритмов распределенных систем классификации, и в частности многоклассовых классификаторов на их основе, внесли: Мак-Махан Б. (Brendan McMahan), Ремейдж Д. (Daniel Ramage), Такурта А. (Abhradeep Thakurta), Бэй С. (Stephen D. Bay), Бишоп К. (Christopher M. Bishop), Вольф Д.А., Агафонова Ю.Д, Соломин А.А.

При этом существующие исследования в основном ориентируются на архитектурные аспекты организации таких систем и разработку протоколов взаимного информационного согласования данных локальных классификаторов в контролируемых условиях, и не в полной мере рассматривают вопросы, связанные со специальными условиями, такими, как неполнота локальных элементных матриц классов, возникающих, если используются, как обучающие выборки на основе общедоступных данных, так и выборки на основе конфиденциальных (применительно к конкретному локальному классификатору) данных.

Таким образом, актуальность темы диссертационного исследования связана с необходимостью разработки специальных средств математического и программного обеспечения управления обменом данными для принятия решения в распределенных системах многоэлементной классификации, функционирующих в условиях неполноты локальных элементных матриц, с целью повышения эффективности процесса классификации.

Объектом исследования является распределенная система многоэлементной классификации.

Предметом исследования являются модели и методы формирования элементных матриц классов в процессе обучения распределенной системы многоклассовой классификации.

Цель и задачи исследования Целью диссертационного исследования является повышение эффективности процесса классификации в распределенной системе многоэлементной классификации в условиях неполных данных для принятия решения, за счет разработки модели и алгоритмов управления обменом данными об оценках ненаблюдаемых классов.

Для достижения поставленной цели необходимо решить следующие частные научные задачи:

1) Разработать модель классификатора распределенной системы многоэлементной классификации, учитывающую неполноту локальных элементных матриц и основанную на представлении меток ненаблюдаемых классов распределением плотности вероятности.

2) Создать алгоритм получения значений оценок вероятностной функции ненаблюдаемых классов локальных классификаторов, основанный на методе расчета параметров статистической вероятностной модели со смешанными распределениями.

3) Разработать алгоритм децентрализованного взаимодействия узлов распределенной системы многоэлементной классификации, функционирующей в условиях неполноты локальных элементных матриц классов, обеспечивающий получение полной элементной матрицы с учетом потенциально ненаблюдаемых классов.

4) Модифицировать существующую архитектуру распределенной системы многоэлементной классификации и реализующий ее программный комплекс, обеспечивающие формирование итоговой модели обучения классификатора для условий неполноты локальных элементных матриц классов.

Методология и методы исследования При решении поставленных в диссертации задач использовались: методы классификации данных, методы машинного обучения, статистические методы оценки параметров вероятностных моделей, методы математической статистики и планирования экспериментов.

Научная новизна В диссертации получены следующие результаты, характеризующиеся научной новизной:

– модель многоэлементного классификатора, отличающаяся от известных учетом условий неполноты локальных элементных матриц, и обеспечивающая представление значений оценки вероятностной функции ненаблюдаемых классов параметрами статистической вероятностной модели;

– алгоритм получения значений оценок вероятностной функции ненаблюдаемых классов, отличающийся от известных итерационным оцениванием параметров модели методом максимального правдоподобия, и обеспечивающий выбор таких оценок, которые наиболее полно представляют пространство признаков локальных классов;

– алгоритм децентрализованного управления обменом данными системы многоэлементной классификации, отличающийся гибридной схемой взаимодействия узлов в условиях неполноты классов локальных классификаторов, и обеспечивающий дополнение их элементных матриц оценками вероятностных функций ненаблюдаемых классов;

– архитектура распределенной системы многоэлементной классификации, отличающаяся от известных реализацией трехэтапной процедуры получения вероятностной функции ненаблюдаемых классов и обеспечивающая формирование итогового классификатора на основе динамически получаемых локальных элементных матриц.

Теоретическая значимость исследования заключается в том, что предлагаемый новый подход к формированию модели классификатора распределенной системы многоэлементной классификации в условиях неполных данных для принятия решения может быть использован

в совершенствовании теоретических и экспериментальных перспективных систем интеллектуальной обработки данных.

Практическая значимость работы заключается в повышении эффективности процесса многоэлементной классификации, примененного в различных областях человеческой деятельности. Разработано специальное программное обеспечение компонентов распределенной системы многоэлементной классификации объектов на цифровых изображениях в условиях конфиденциальности части данных обучающей выборки. Предложены рекомендации для существующих вариантов систем интеллектуальной обработки данных по реализации процесса распределенной классификации в условиях неполных данных локальных элементных матриц классов.

Достоверность результатов подтверждается использованием при разработке моделей известных математических методов и результатами имитационных экспериментов.

Положения, выносимые на защиту:

1) Модель многоэлементного классификатора обеспечивает представление значений оценки вероятностной функции ненаблюдаемых классов параметрами статистической вероятностной модели.

2) Алгоритм получения значений оценок вероятностной функции ненаблюдаемых классов обеспечивает выбор оценок вероятностной функции, наиболее полно представляющих пространство признаков локальных классов.

3) Алгоритм децентрализованного управления обменом данными системы многоэлементной классификации обеспечивает дополнение локальных элементных оценками вероятностных функций ненаблюдаемых классов.

4) Архитектура распределенной системы многоэлементной классификации обеспечивает формирование итогового классификатора на основе динамически получаемых локальных элементных матриц.

Апробация работы Основные положения диссертационной работы докладывались и обсуждались на следующих конференциях: Сборнике статей научно-исследовательского института систем связи и управления (НИИССУ), XVI Всероссийской научно-практической конференции «Территориально распределенные системы охраны» (Калининград, 2023), XXVIII-th International Open Science Conference «Modern informatization problems in the technological and telecommunication systems analysis and synthesis (MIP-2023'SCT)» (Yelm, WA, USA, 2023), XXIX-th International Open Science Conference «Modern informatization problems in simulation and social technologies (MIP 2024'SCT)» (Yelm, WA, USA, 2024), XIV Всероссийской межведомственной научной конференции «Актуальные направления развития систем обеспечения безопасности объектов государственной охраны и защиты охраняемых объектов, специальной связи для нужд органов государственной власти и специального информационного обеспечения государственных органов» (Орел, 2025), а также на научных

семинарах кафедры информатики и вычислительной техники Академии ФСО России (2022–2025 гг.).

Реализация и внедрение результатов работы Результаты диссертации внедрены в практическую деятельность ООО «Айти Интегра Системс» (г. Москва), а также в образовательный процесс Академии ФСО России (дисциплина – «Компьютерные сети»).

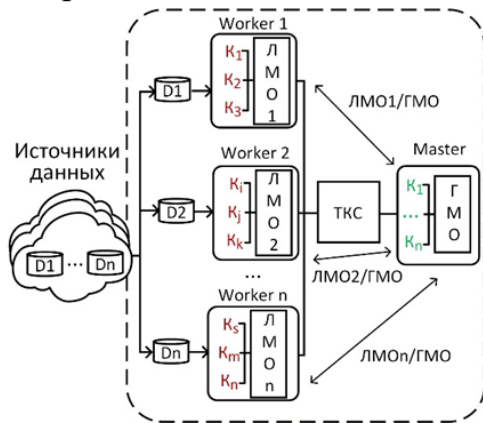
Соответствие паспорту специальности. Содержание диссертации соответствует п. 3 «Разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта», п. 5 «Разработка специального математического и алгоритмического обеспечения систем анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта», п. 15 «Теоретический анализ и экспериментальное исследование функционирования элементов систем управления в нормальных и специальных условиях с целью улучшения технико-экономических и эксплуатационных характеристик» паспорта специальности 2.3.1. Системный анализ, управление и обработка информации, статистика.

Публикации. По результатам диссертационного исследования опубликовано 11 научных работ, в том числе 3 – в изданиях, рекомендованных ВАК РФ, 1 – в издании Scopus и 1 свидетельство о регистрации программы для ЭВМ. В работах, опубликованных в соавторстве и приведенных в конце автореферата, лично автором получены следующие результаты: [1, 5, 7, 9] – модель классификатора многоэлементной системы классификации, функционирующей в условиях неполноты классов локальных классификаторов, [2, 4, 11] – алгоритм децентрализованного управления обменом данными системы многоэлементной классификации, [3, 4] – алгоритм получения значений оценки вероятностной функции ненаблюдаемых классов локальных классификаторов, [3, 7, 8, 9, 10] – архитектура системы многоэлементной классификации с федеративным машинным обучением, поддерживающая формирование итогового классификатора на основе динамически полученных матриц классов локальных классификаторов.

Структура и объем работы. Диссертация состоит из введения, четырех глав, заключения и 3 приложений. Работа изложена на 140 страницах машинописного текста, включая 40 рисунков, 14 таблиц и список литературы из 106 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В первой главе выполнена постановка задачи исследования и разработана модель классификатора распределенной системы многоэлементной классификации, учитывающая неполноту классов подмножества ее локальных классификаторов. Для этого рассмотрены системы обработки данных с ФМО, выявлены их обобщенные характеристики, в частности, применительно к решению задачи распределенной многоэлементной классификации (рисунок 1). Проведен



анализ подходов к бинаризации в задаче многоэлементной классификации, а также виды функции потерь и метрики оценивания эффективности, применяемые при решении этой задачи. Рассмотрена проблема неоднородности данных в обучающих выборках локальных классификаторов таких систем, в силу наличия, как общедоступных (ООД), так и локальных конфиденциальных (ЛКД) данных, приводящая к проблеме не

наблюдаемости классов (рисунок 2).

В этой связи было сформулировано противоречие между необходимостью снижения значения функции потерь в процессе распределенной многоэлементной классификации, и отсутствием в существующих системах с ФМО методов и алгоритмов, учитывающих наличие неоднородности данных в обучающих выборках локальных классификаторов, что порождает неполноту классов в процессе агрегации глобальной модели обучения.

Рисунок 1 – Схема системы обработки данных с ФМО для многоэлементной классификации

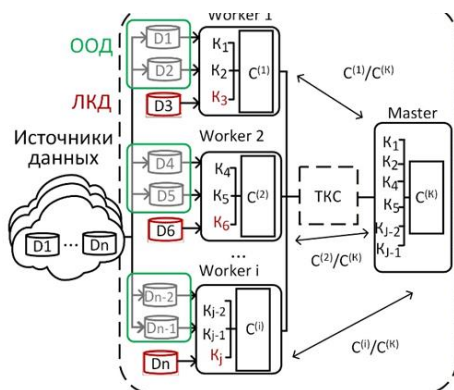


Рисунок 2 – Проблема не наблюдаемости классов в системах многоэлементной классификации с ФМО

На основании этого противоречия сделана постановка научной задачи по разработке модели многоэлементного классификатора $C^{(K)}$ и алгоритмов децентрализованного обмена данными локальных классификаторов $C^{(i)}$, устраняющих проблему не наблюдаемости части классов.

В ходе решения задачи моделирования многоэлементного классификатора был выбран метод VOTE для агрегации моделей обучения локальных классификаторов, а также обосновано использование в качестве вида их функции потерь категориальной кросс-энтропии (выражение 1).

$$CCE = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K y_{i,k} \log(p_{i,k}), \quad (1)$$

Была формализована задача, решаемая i -м классификатором $C(i)$, как задача минимизации кросс-энтропийной функции потерь в условиях полноты классов (выражение 2)

$$C^{(i)} = \arg \min_{k \in K} f_L^{CCE}(y_{ik}^{набл}), \quad (2)$$

где функция потерь f_L^{CCE} определяется как:

$$-\sum_{i=1}^{n_j} \sum_{k \in K_j} l(y_{ik}^{набл}) \log C^{(i)}_{x_{ij}}, \quad (3)$$

а $l(y_{ik}^{набл})$ - бинарная индикаторная функция

$$\begin{cases} 1 \rightarrow y_{ik} \in y_{ik}^{набл} \\ 0 \rightarrow y_{ik} \notin y_{ik}^{набл} \end{cases},$$

отражающая наличие или отсутствие принадлежности y_{ik} исхода классификации всему множеству наблюдаемых классов.

На основании этого выполнено моделирование многоэлементного классификатора, функционирующего в условиях полноты классов локальных классификаторов:

$$C^{(N)} = \sum_{j=1}^N C_j = \arg \min_{C^{(N)}} \left\{ -\sum_{j=1}^N \sum_{i=1}^{n_j} \sum_{k=0}^K l(y_{ik}^{набл}) \log C^{(i)}_{x_{ij}} \right\} \quad (4)$$

Был выполнен переход к моделированию многоэлементного классификатора для условий неполноты классов локальных классификаторов.

Определим гибридные данные, который обрабатывает j -й локальный классификатор, как $\Gamma Д_j = ОДД_j + ЛКД_j$.

При этом, в силу конфиденциальности данных $ЛКД_j$, значение конкретного элемента k множества меток классов K можно получить только путем интегрирования оценки плотности вероятности значения x_{ij} каждого j -го локального классификатора. Итоговый многоэлементный классификатор $C^{(N)}$ при этом оперирует суммой условных вероятностей:

$$P(y=k, \Gamma Д_j \in C^{(j)} | X), \quad (5)$$

Определив функцию распределения вероятностей меток классов для всей матрицы входных данных $X_{j \times d}$ j -го локального классификатора как $f_X^{(j)}$, а условную вероятность $P(y=k, ГД_j \in C^{(j)} | X)$ как $p_N^{(j)}$, было определено представление глобального классификатора для условия неполноты классов локальных классификаторов:

$$C^{(N)} = \sum_{j=1}^N v_{j,k} C^{(j)}, \quad (6)$$

где $v_{j,k}$ – весовой коэффициент, пропорциональный двум функциям:

$$v_{j,k} \propto l(y_{ik}^{набл}, y_{ik}^{набл}) \delta(f_X^{(j)}, p_N^{(j)}), \quad (7)$$

где $\delta(f_X^{(j)}, p_N^{(j)})$ – некоторая функция с неопределенными параметрами. При этом, в силу неопределенности значений параметров $f_X^{(j)}, p_N^{(j)}$, одним из видов функции $\delta(f_X^{(j)}, p_N^{(j)})$ может быть их произведение $f_X^{(j)} \cdot p_N^{(j)}$. Тогда весовой коэффициент $v_{j,k}$ определяется как:

$$v_{j,k} = \frac{f_X^{(j)} p_N^{(j)}}{\sum_{j=1}^L f_X^{(j)} p_N^{(j)}} \quad (8)$$

Тогда выражение 6 представляется как:

$$C^{(N)} = \sum_{j=1}^N C^{(j)} \times \frac{f_X^{(j)} p_N^{(j)}}{f_X} = \sum_{j=1}^N C^{(j)} \times \frac{f_X^{(j)} p_N^{(j)}}{\sum_{j=1}^N f_X^{(j)} p_N^{(j)}} \quad (9)$$

В силу того, что первый множитель можно задать индикаторной функцией $l(y_{ik}^{набл}, y_{ik}^{набл})$ а, второй – весовым коэффициентом $v_{j,k}$, было получено итоговое представление модели многоэлементного классификатора, функционирующего в условиях неполноты классов локальных классификаторов:

$$C^{(N)} = \sum_{j=1}^N l(y_{ik}^{набл}, y_{ik}^{набл}) \times v_{j,k} C^{(j)} \quad (10)$$

Для этой модели было определено условие, определяющее наличие неполноты классов хотя бы в одном из локальных классификаторов:

$$v_{j,k} \geq \frac{(C^{(j)})^{N-1}}{N} \quad (11)$$

Используя представление функции потерь j -го локального классификатора, как кросс-энтропийную функцию $\sum_{k=0}^K l(y_{ik}^{набл}, \overline{y_{ik}^{набл}}) \log C^{(j)}(x_{ij})$ и подставив его в модель многоэлементного глобального классификатора (выражение 10) было получено ее представление для множества локальных классификаторов:

$$\sum_K \log \sum_{j=1}^N \sum_{k=0}^K l(y_{ik}^{набл}, \overline{y_{ik}^{набл}}) \log C^{(j)}(x_{ij}) = \sum_K \log \sum_{j=1}^N v_{j,k} C^{(j)}(x_{ij}) \quad (12)$$

С использованием неравенства Йенсена было доказано, что множество локальных классификаторов, достигая минимума категориальных кросс-энтропийных потерь, обеспечивает минимум потерь многоэлементного классификатора:

$$\sum_K \log C^{(j)}(x_{ij}) = \sum_j v_{j,k} l(y_{ik}^{набл}, \overline{y_{ik}^{набл}}) \log C^{(j)}(x_{ij}) = 1 \quad (13)$$

При этом в силу выдвинутого предположения о том, что в условиях использования в обучающих выборках j -го локального классификатора данных ГД, метки ненаблюдаемых классов данных x_{ij} можно задать значением оценки плотности вероятности.

Для их параметрического представления был выполнен обоснованный выбор модели гауссовой смеси распределений (GMM), которая описывается тремя типами параметров: средними значениями компонентов подсовокупности (одномерное представление); ковариациями средних значений компонентов подсовокупности (многомерное представление); весовыми коэффициентами смеси компонентов подсовокупности. В общем случае она представляется взвешенной суммой R компонентов гауссовых плотностей распределения вероятностей:

$$p(x|\lambda) = \sum_{i=1}^R w_i g(x|\mu_i, \Sigma_i), \quad (14)$$

где x – это многомерный вектор элементов, $w_i, i=1, \dots, R$ – смеси весов этих элементов, а $g(x|\mu_i, \Sigma_i), i=1, \dots, R$ – многомерная функция определяющая компоненты гауссовой плотности вероятностей, которая представляется как:

$$g(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{D}{2}} \left| \Sigma_i \right|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x - \mu_i)' \Sigma_i^{-1} (x - \mu_i) \right\}, \quad (15)$$

где D – размерность векторного пространства элементов, μ_i – среднее значение вектора, а Σ_i – ковариационная матрица.

При этом параметризация суммарной модели GMM выполняется на основе средних значений векторов, ковариационных матриц и весов смеси из всех плотностей вероятностей R компонентов:

$$\lambda = \{w_i, \mu_i, \Sigma_i\}, i=1, \dots, R \quad (16)$$

Было сделано вывод, что для представления не наблюдаемых классов модели j -го локального классификатора, функцию распределения их вероятностей можно представить моделью GMM с R компонентами:

$$f_X^{(j)}(X) = \sum_{r=1}^R \pi_r^j f_{X|z_j=r}(X), \quad (17)$$

Таким образом, разработана модель многоэлементного классификатора, отличающаяся от известных учетом условий неполных данных для принятия решения, позволяющая представить значения оценки вероятностной функции меток не наблюдаемых классов параметрами модели гауссовой смеси распределений (GMM). Разработанная модель отличается от известных возможностью реализации процесса многоэлементной классификации в условиях неполноты классов локальных классификаторов, связанной с наличием в их обучающих выборках гибридных (общедоступных и конфиденциальных) данных.

Вторая глава посвящена разработке алгоритма получения значений оценок вероятностной функции меток ненаблюдаемых классов локальных классификаторов.

Разработанная в первой главе GMM-модель конкретизирована применительно к многоэлементной классификации объектов на цифровых изображениях.

Для получения значений оценок вероятностной функции ненаблюдаемых классов, представленной моделью GMM, были исследованы методы оценивания ее параметров и выбран метод максимального правдоподобия – поиска таких параметров, которые максимизируют функцию ее правдоподобия с учетом используемых элементов. Исходя из выражения 16, для последовательности R векторов (с учетом их независимости), функция правдоподобия модели GMM на некотором интервале T определяется как:

$$p(X|\lambda) = \prod_{t=1}^T p(x_t|\lambda) \quad (18)$$

Очевидно, в силу нелинейности функции параметров λ , получение аппроксимации $p(X|\lambda)$ аналитически невозможно. Поэтому, для решения этой задачи были исследованы итерационные методы, которые позволяют, начиная с функции λ (первая итерация), оценить вид новой функции $\bar{\lambda}$, для которого выполняется условие:

$$p(X|\bar{\lambda}) \geq p(X|\lambda) \quad (19)$$

При этом на следующей итерации полученный вариант $\bar{\lambda}$ становится начальным и цикл повторяется до тех пор, пока не будет достигнут заданный порог сходимости. Из рассмотренных итеративных методов был выбран EM-алгоритм, состоящий из шагов: условного ожидания (E (expectation)-шаг); максимизации (M (maximization)-шаг). На E-шаге рассчитываются условные

ожидания отсутствующих данных (ЛКД) с учетом всей совокупности наблюдаемых данных и Θ – оценок параметров модели GMM. При этом условные ожидания ненаблюдаемых данных с учетом наблюдаемых данных и оценок параметров модели GMM вычисляются на основе выражения:

$$Q(\Theta_0 | \Theta_n) = E_{Z|x, \Theta_n} [\log L(\Theta, x, z)], \quad (20)$$

где $L(\Theta, x, z)$ – функция правдоподобия, Θ – вектор параметров, Θ_n – оценка параметров модели GMM, x – наблюдаемые данные (данные ГД), z – не наблюдаемые данные (ЛКД). На M-шаге выполняется поиск оценки параметра Θ , которая максимизирует функцию правдоподобия данных E-шага:

$$\Theta^* = \arg_{\Theta} \max Q(\Theta_0 | \Theta_n) \quad (21)$$

Шаги E и M чередуются с целью обновления некоторой оценки Θ_n неизвестных на каждой итерации оценок параметров Θ . Предложенная модификация EM-алгоритма была включена в состав цикла формирования моделей локальных классификаторов, являющегося одним из проходов двухциклового алгоритма формирования модели многоэлементного классификатора (рисунок 3).



Рисунок 3 – Схема алгоритма получения значений оценок вероятностной функции меток ненаблюдаемых классов локальных классификаторов

Таким образом, разработан алгоритм получения значений оценок вероятностной функции меток ненаблюдаемых классов локальных классификаторов, обеспечивающий выбор тех из них, которые наиболее полно представляют распределение векторов признаков этих классов, и отличающийся от известных итерационным оцениванием параметров модели гауссовой смеси распределений методом максимального правдоподобия, основанном на этапах EM-алгоритма.

Третья глава посвящена разработке алгоритма децентрализованного управления обменом данными узлов системы многоэлементной классификации, функционирующей в условиях неполных данных для принятия решения, который определяет порядок получения вида многоэлементного классификатора на основе агрегирования множества локальных классификаторов, с метками классов, представленных параметрами их GMM-моделей. Его особенностью является учет наличия в пространстве признаков классов отдельного локального классификатора новых классов, которые, являются потенциально ненаблюдаемыми для остального или части подмножества локальных классификаторов. При этом было определено, что этапу формирования вида агрегированного обобщенного классификатора должны предшествовать этапы: формирования локальных классификаторов отдельными узлами Worker-node с поддержкой нахождения новых (потенциально ненаблюдаемых) классов; их взаимного информационного согласования множеством Worker-node с учетом наличия ненаблюдаемых классов, обеспечивающим решение проблемы неполноты классов.

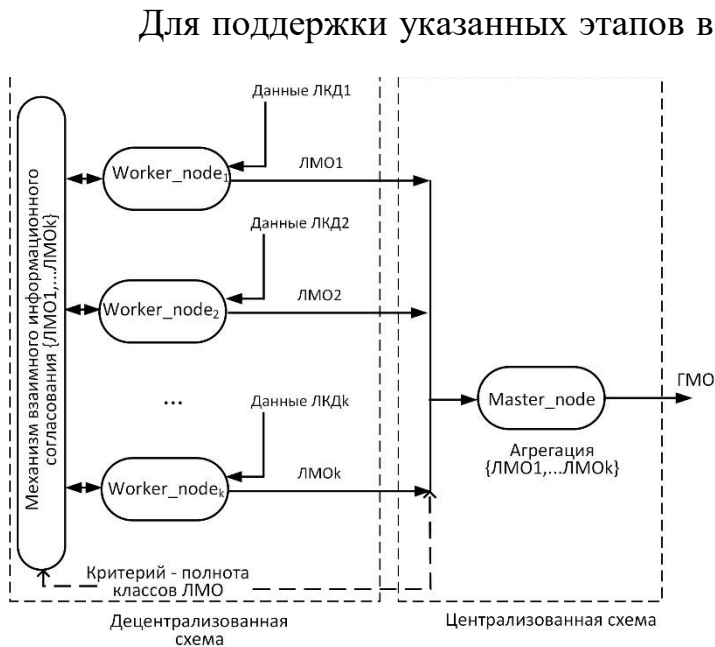


Рисунок 4 – Гибридная схема взаимодействия узлов системы многоэлементной классификации

децентрализованного Peer-to-Peer взаимодействия и обоснованно выбран класс протоколов DHT. Выбор реализации ВИС позволил разработать трехэтапный алгоритм взаимодействия узлов распределенной системы многоэлементной классификации, функционирующей в условиях неполноты классов локальных классификаторов (рисунок 5).

Для поддержки указанных этапов в исследовании предлагается применение гибридной схемы взаимодействия узлов системы, включающей децентрализованную и централизованную компоненты (рисунок 4). Основой децентрализованной схемы является механизм взаимного информационного согласования (ВИС) моделей локальных классификаторов. Поскольку в процессе ВИС узлы Worker-node выполняют широковещательную рассылку параметров GMM-модели классов-кандидатов для отнесения к новому классу, в качестве основы механизма ВИС рассматривались протоколы децен-

Этапы алгоритма выполняются последовательно. Переход к очередному этапу выполняется при полной сходимости алгоритма

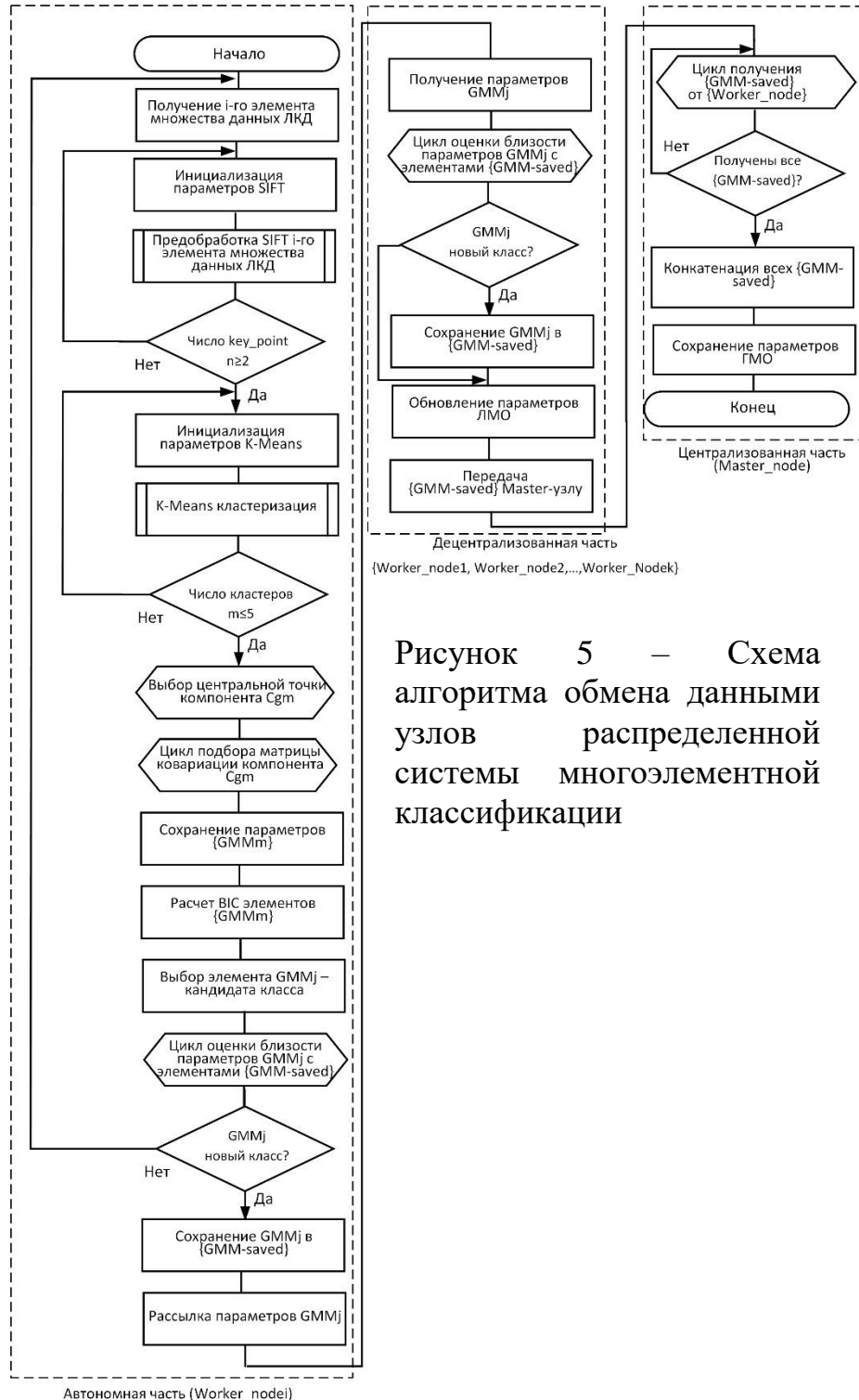


Рисунок 5 – Схема алгоритма обмена данными узлов распределенной системы многоэлементной классификации

подмножества разосланных GMM-моделей к классу ненаблюдаемых. Этап 3 является классическим централизованным этапом раунда ФМО, когда параметры множеств GMM-моделей всех Worker-узлов передаются Master-узлу, где выполняется их конкатенация.

Таким образом, разработан алгоритм децентрализованного управления обменом данными системы многоэлементной классификации, функционирующей в условиях неполных данных для принятия решения, основанный на трехэтапной гибридной (автономная-

му этапу выполняется при полной сходимости алгоритма предыдущего этапа. Этап 1 определяет автономное функционирование множества узлов Worker-node. В силу автономности его выполнения этап носит асинхронный характер, и его временная сложность зависит от особенностей функционирования отдельных узлов Worker_node при формировании модели GMM-классификатора. Этап 2 реализует децентрализованное взаимодействие подмножества (или всего множества) узлов Worker_node в процессе их ВИС по отнесению

децентрализованная-централизованная) схеме функционирования ее узлов и реализующий обобщенный цикл получения элементных матриц подмножества локальных классификаторов с учетом потенциально не наблюдаемых классов.

В четвертой главе разработана архитектура распределенной системы

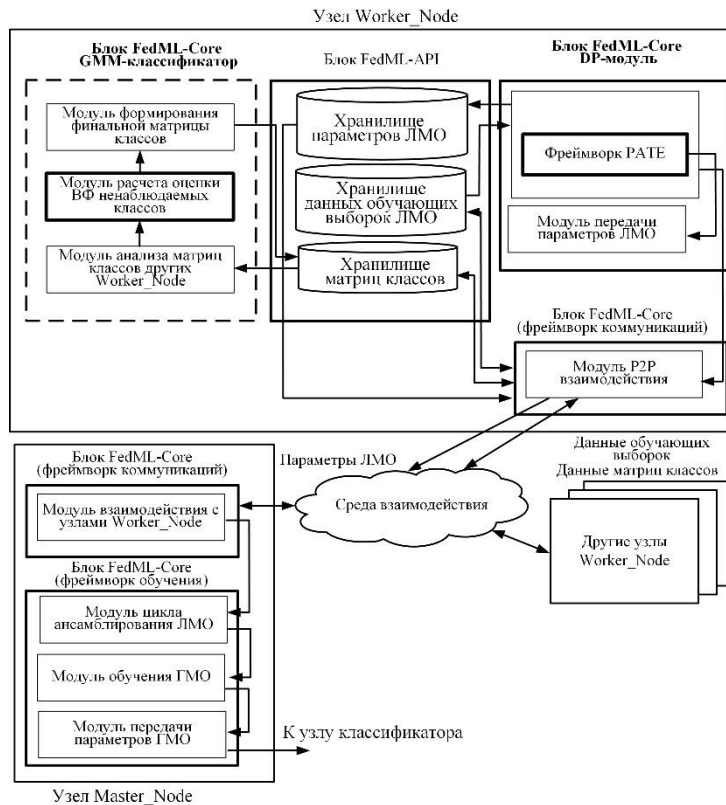


Рисунок 6 – Структура программного комплекса с предложенным блоком формирования финальных элементных матриц локальных классификаторов

мирования финальной матрицы классов, подключаемый через интерфейс библиотеки FedML-API. Он представлен тремя модулями, функционирующими последовательно: модулем анализа матриц классов остальных узлов Worker_Node работающих в составе системы, модулем расчета вероятностных функций ненаблюдаемых классов, рассматриваемых матриц классов, реализующим предложенный в главе 3 модифицированный EM-алгоритм и модулем формирования финальной матрицы классов узла Worker_Node, содержащей вероятностные оценки ненаблюдаемых классов других узлов Worker_Node. Для хранения получаемых от остальных узлов Worker_Node матриц классов, а также сохранения собственной матрицы классов в состав блока FedML-API, наряду с хранилищами обучающих выборок и параметров локальной модели обучения, включено хранилище матриц классов, использующих структуры данных фреймворка линейной алгебры NumPy. Взаимодействие узлов Worker_Node между собой производится с использованием блока P2P-связи (реализация протокола DHT).

многоэлементной классификации, функционирующей в условиях неполных данных для принятия решения, включающая разработанное специальное математическое обеспечение поддержки процесса многоэлементной классификации в условиях неполноты классов локальных классификаторов. В качестве базовой программной основы вычислительных узлов системы обоснованно выбран фреймворк FedML. Структура программного комплекса узлов Worker_Node и Master_Node предлагаемой архитектуры представлена на рисунке 6.

Из рисунка 6 видно, что в состав структуры вычислительного узла Worker_Node добавлен блок формирования финальных элементных матриц локальных классификаторов

Для оценивания степени достижения цели исследования был спланирован имитационный эксперимент на специально разработанной имитационной модели системы. В качестве среды моделирования, поддерживающей фреймворк FedML, была выбрана облачная среда на основе открытого исходного кода TensorOpera Federate Simulation, которая поддерживает, как репрезентативные алгоритмы ФМО для различных топологий взаимодействия узлов, так и тестовые наборы данных, такие как EMNIST (классификация текстов, библиотека FedNLP) и CIFAR-100 (классификация изображений, библиотека FedCV). Обобщенная схема разработанной имитационной модели представлена на рисунке 7.

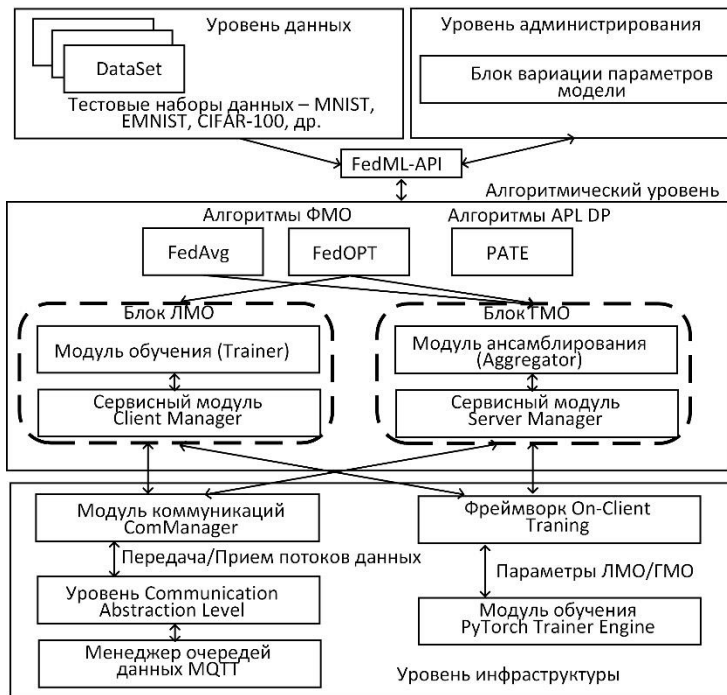


Рисунок 7 – Структура имитационной модели разработанной системы

изменялось от 10 до 40); количество классов данных в обучающих выборках локальных классификаторов; распределение общедоступных и конфиденциальных данных в обучающих выборках, обеспечивающее принцип Non-IID (отсутствие независимых, идентично распределенных данных); начальное заполнение хранилищ данных обучающей выборки локальных классификаторов. Разработанные алгоритмы подключаются к модулю обучения (Trainer) блока ЛМО и модулю ансамблирования (Aggregator) блока ГМО модели через интерфейс FedML-API (фреймворк FedML Parrot, модуль PyTorch_Dataloader).

Пример сравнительного оценивания предложенного решения и немодифицированной схемы – оценивание качества классификации изображений набора данных CIFAR-100 по показателям точности (accuracy (ACC)) и Каппа Коэна (Cohen's kappa (κ)) с варьируемым коэффициентом распределения изображений, относящихся к ООД и ЛКД $K_{ГД} = \{0,25; 0,5; 0,75\}$ и:

– фиксированным количеством меток классов ЛМО $L_C=3$;

В рамках разработанной имитационной модели был спланирован сравнительный имитационный эксперимент для оценивания качества процесса многоэлементной классификации в контролируемых условиях обучения моделей локальных классификаторов на выборке гибридных (общедоступных и конфиденциальных) данных. Входные параметры для каждого прогона имитационного эксперимента случайным образом формировались соответствующими

генераторами нагрузки имитационной модели: количество блоков ЛМО (значение

– случайным распределением количества меток классов из множества $L'_C = \{3, 4, 5\}$ представлены на рисунках 8 и 9 соответственно.

Из рисунков 8 и 9 видно, что вне зависимости от количества узлов в моделируемой системе, а также коэффициента распределения $K_{ГД}$ и значения L'_C схема с предлагаемым математическим и программным обеспечением реализует более высокое значение показателей качества многоэлементной классификации. При этом близость значений показателей качества сравниваемых схем классификаторов в условиях повышения значений $K_{ГД}$ и L'_C связана как с необходимостью оптимизации алгоритмических и программных решений, так и с исследованиями по выбору метода многоэлементной классификации.

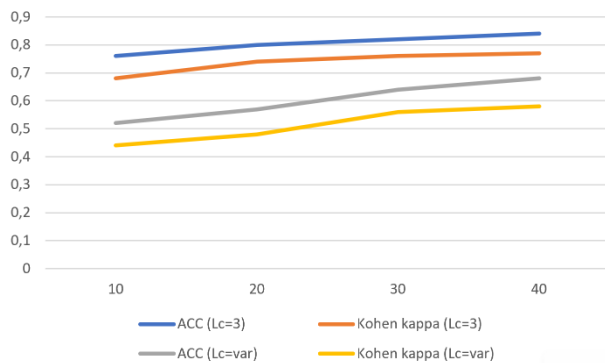


Рисунок 8 – Значения показателей качества предложенного решения

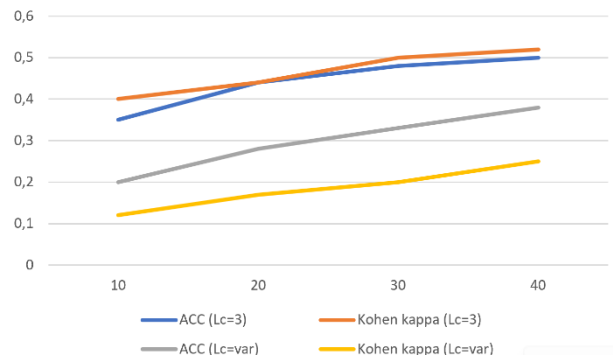


Рисунок 9 – Значения показателей качества процесса классификации не модифицированной схемы системы многоэлементной классификации

Таким образом, разработана архитектура распределенной системы многоэлементной классификации и структура ее программного комплекса, обеспечивающая формирование многоэлементного классификатора в условиях неполноты классов подмножества локальных классификаторов, входящих в ее состав. Результаты имитационного эксперимента подтверждают достижение цели исследования. На реализованные элементы разработанного ПО специальных модулей локального и глобального классификаторов, получено свидетельство о регистрации программы для ЭВМ в реестре ФИПС.

ЗАКЛЮЧЕНИЕ

На основании теоретических исследований и вычислительных экспериментов сделаны следующие выводы.

1. На основе анализа моделей, методов и алгоритмов в предметной области распределенных систем многоэлементной (многоклассовой) классификации разработана модель классификатора, базирующаяся на представлении

меток не наблюдаемых классов распределением плотности вероятности и параметрически представляемая моделью гауссовой смеси распределений. Предложенная модель отличается от известных возможностью реализации процесса классификации в условиях неполноты классов локальных классификаторов, связанной с наличием в их обучающих выборках гибридных (общедоступных и конфиденциальных) данных.

2. Разработан алгоритм получения значений оценок вероятностной функции меток ненаблюдаемых классов локальных классификаторов, представленной моделью гауссовой смеси распределений, обеспечивающий выбор тех из них, которые наиболее полно представляют пространство их признаков, и отличающийся от известных итерационным оцениванием параметров модели методом максимального правдоподобия, основанном на этапах EM-алгоритма.

3. Разработан алгоритм децентрализованного управления обменом данными узлов системы классификации, функционирующей в условиях неполных данных для принятия решения, отличающийся гибридной схемой взаимодействия и обеспечивающий формирование матриц классов подмножества локальных классификаторов с учетом ненаблюдаемых ими классов, основанный на гибридной (автономная-децентрализованная-централизованная) схеме функционирования узлов и реализующей обобщенный цикл получения матриц классов подмножества локальных классификаторов с учетом потенциально ненаблюдаемых классов.

4. Разработана архитектура распределенной системы многоэлементной классификации и структура ее программного комплекса, обеспечивающая формирование модели классификатора в условиях неполноты классов подмножества локальных классификаторов, входящих в ее состав. На программную реализацию прототипа системы получено свидетельство о регистрации программы для ЭВМ в реестре ФИПС № RU 2024662112.

5. Проведен сравнительный имитационный эксперимент по оцениванию эффективности процесса многоэлементной классификации в контролируемых условиях неполноты классов локальных классификаторов для традиционной (централизованной) схемы и предложенного решения. Результаты экспериментов показывают повышение на 0,2-0,3 (условие неполноты классов) качества классификации по показателям точности и каппа Коэна для разработанной архитектуры системы федеративного машинного обучения в сравнении с традиционным решением.

Дальнейшие разработки и исследования необходимо вести в направлении оптимизации разработанных алгоритмов и развитии программных средств перспективных систем распределенных систем интеллектуальной обработки данных, реализующих процесс многоэлементной классификации.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в изданиях списка ВАК

1. Ветров, И.И. К вопросу о структурно-параметрической оптимизации систем федеративного машинного обучения/ И.И. Ветров, М.А. Куцакин, **П.А. Михалев**// Системы управления и информационные технологии. – 2022. – № 3 (89). – С. 42-48.

2. Ветров, И.И. Подход к моделированию многоклассового классификатора системы федеративного машинного обучения, функционирующего в условиях неполноты классов локальных классификаторов / И.И. Ветров, М.А. Куцакин, **П.А. Михалев** // Системы управления и информационные технологии. – 2024. – № 4 (98). – С. 26-32.

3. Михалев П.А. Разработка алгоритма многоклассового классификатора системы федеративного обучения, функционирующей в условиях неполноты классов локальных классификаторов / **П.А. Михалев**, М.А. Куцакин, О.В. Карамыхова// Моделирование, оптимизация и информационные технологии. – 2024. – № 12 (3). Режим доступа: <https://moitvvt.ru/ru/journal/pdf?id=1731>.

Публикации в изданиях, индексируемых в Scopus

4. Mikhalev, P.A. An algorithm for obtaining an ensemble machine learning model for node federation in decentralized machine learning systems / **P.A. Mikhalev**, N.Yu. Bumazhkina, A. A. Rubtsov // AIP Conference Proceedings. Manuscript № AIPCP21-AR-MIP 2024-00155.

Свидетельство о государственной регистрации программы для ЭВМ

5. Программа на языке Python «GMM-классификатор локального машинного обучения»/ **П.А. Михалев**, М.А. Куцакин, В.В. Живодеров, М.В. Кондратенков. - Свидетельство о государственной регистрации программы для ЭВМ №2024662112 от 23.05.2024. – М.: Роспатент, 2024.

Статьи и материалы конференций

6. **Михалев П.А.**, Куцакин М.А. Подходы к моделированию распределенной автоматизированной системы информационного мониторинга в режиме реального времени в условиях нестационарной нагрузки // Сборник статей научно-исследовательского института систем связи и управления (НИИССУ). – Москва, – 2022. – С. 26-32.

7. **Mikhalev P.A.**, Kutsakin M.A., Mironov O.Yu. On the need for parametric optimization of systems with federated machine learning // Modern informatization problems in the technological and telecommunication systems analysis and synthesis (MIP-2023'SCT). Proceedings of the XXVIII-th International Open Science Conference. – Yelm, WA, USA, 2023. – p. 37-41.

8. **Mikhalev P.A.**, Kutsakin M.A., Mironov O.Yu. An approach to the ensembling of models of local classifiers under conditions of incompleteness of classes in systems with federated learning // Modern informatization problems in the technological and telecommunication systems analysis and synthesis (MIP-

2024'SCT). Proceedings of the XXIX-th International Open Science Conference. – Yelm, WA, USA, 2024. – p. 36-40.

9. **Михалев П.А.**, Куцакин М.А. Анализ подходов к построению математических моделей систем активного мониторинга, функционирующих на основе машинного обучения // XVI Всероссийская научно-практическая конференция «Территориально распределенные системы охраны». – Калининград, 2023. – С. 84-88.

10. **Михалев П.А.**, Куцакин М.А. О проблеме централизованного подхода к организации управления подсистемы анализа поведенческих профилей системы контроля и управления доступом // XVI Всероссийская научно-практическая конференция «Территориально распределенные системы охраны». – Калининград, 2023. – С. 67-70.

11. **Михалев П.А.**, Куцакин М.А., Карамыхова О.В. Алгоритм многоклассового классификатора системы федеративного обучения, функционирующей в условиях неполноты классов локальных классификаторов // XIV Всероссийская межведомственная научная конференция «Актуальные направления развития систем обеспечения безопасности объектов государственной охраны и защиты охраняемых объектов, специальной связи для нужд органов государственной власти и специального информационного обеспечения государственных органов». - Орел, 2025. – С. 133-136.