

ОТЗЫВ

официального оппонента на диссертацию и автореферат

Михалева Павла Андреевича на тему

«Управление децентрализованным обменом информацией для принятия решений в распределенных системах многоэлементной классификации с неполными данными», представленную на соискание ученой степени кандидата технических наук по специальности

2.3.1 – Системный анализ, управление и обработка информации, статистика

Актуальность темы исследования

Непрерывный рост сложности моделей машинного обучения, разрабатываемых для актуальных практических задач, требует всё большего объёма данных для обучения, что существенно усложняет создание и эксплуатацию центров обработки данных. Дополнительным серьёзным ограничением централизованных подходов выступает невозможность объединения всех данных в едином хранилище как по техническим причинам, так и в связи с требованиями безопасности, конфиденциальности и нормативными ограничениями на передачу отдельных чувствительных данных.

Одним из эффективных путей решения указанных проблем является совершенствование моделей и методов федеративного машинного обучения, в рамках которого процесс обучения новых моделей декомпозируется на множество простых процессов, реализуемых в распределенных вычислительных узлах (каждый из которых поддерживает собственную локальную модель обучения) с последующим агрегированием результатов в итоговой (глобальной) модели обучения. Такой подход актуален для решения задач многоэлементной классификации, когда множество локальных моделей реализует классификацию для двух и более классов, а глобальная модель – по всему множеству классов.

Обработка чувствительных данных отдельными локальными моделями порождает ситуацию, в которой отдельные классы будут недоступны другим локальным классификаторам. Такая ненаблюдаемость классов, в свою очередь, порождает неполноту классов, формируемых итоговой глобальной моделью классификации.

В описанных условиях диссертационное исследование Михалёва П.А., направленное на повышение эффективности процесса классификации в распределённой системе многоэлементной классификации в условиях неполных данных для принятия решений, безусловно является актуальным.

Оценка научная новизны основных результатов

Полученные автором результаты обладают научной новизной, заключающейся в том, что:

1) предложена архитектура распределённой системы многоэлементной классификации, отличающаяся от известных реализацией трёхэтапной процедуры получения вероятностной функции ненаблюдаемых классов и обеспечивающая формирование итогового классификатора на основе динамически формируемых локальных элементных матриц;

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

3) предложена модель многоэлементного классификатора, отличающаяся от существующих учётом неполноты локальных элементных матриц и обеспечивающая представление оценок вероятностной функции для ненаблюдаемых классов параметрами статистической вероятностной модели;

4) разработан алгоритм получения оценок вероятностной функции ненаблюдаемых классов, отличающийся процедурой итерационного оценивания параметров модели методом максимального правдоподобия и

позволяющий выбирать оценки, наиболее полно отражающие пространство признаков локальных классов.

Обоснованность научных положений и достоверность полученных результатов, выводов и рекомендаций

Обоснованность полученных автором научных результатов и сформулированных выводов обеспечивается грамотным выбором и корректным использованием апробированных методов исследования, чётким математическим доказательством выдвигаемых положений и гипотез, а также совпадением расчётных теоретических показателей с результатами экспериментов. Ключевые итоги работы отражены в рецензируемых научных журналах из перечня ВАК РФ и в международных изданиях, индексируемых в базах данных Scopus. Диссертация прошла достаточно полную апробацию в ряде научно-технических и отраслевых конференций.

По теме диссертации опубликовано 11 печатных трудов, среди которых 4 статьи в изданиях, рекомендованных ВАК при Минобрнауки России, одна статья в журнале, индексируемом в Scopus, осуществлена государственная регистрация одной программы для ЭВМ.

Значимость для теории и практики

Теоретическая значимость исследования определяется развитием предметной области распределённых систем (многоэлементной) классификации за счёт предложенного научно-методического инструментария, обеспечивающего повышение качества классификации в условиях неполноты данных, используемых для принятия решений.

Практическую значимость работы определяет созданное соискателем специализированное программное обеспечение для компонентов распределённой системы, осуществляющее многоэлементную классификацию объектов на цифровых изображениях при условии конфиденциальности части данных обучающей выборки.

Общая характеристика диссертационной работы

Диссертация написана грамотным научным языком и хорошо структурирована. Главы работы содержат основные результаты диссертационного исследования. Уровень изложения результатов позволяет сделать вывод о сложившейся практике научных исследований, а также о законченности работы

Автореферат и публикации автора в полной мере передают ключевые результаты и выводы диссертационной работы.

Замечания по диссертационной работе и автореферату

1. В основу разрешаемого автором противоречия положена проблема неполноты классов глобальной модели классификации, возникающая в следствие не наблюдаемости классов локальными моделями. Между тем, проблемы не наблюдаемости и неполноты классов в системах многоэлементной классификации с федеративным машинным обучением в одноименном пункте диссертации (п. 1.5, с. 38–41 диссертации) не раскрыты.

2. Из текста работы не ясно, почему проведённое исследование ограничено только горизонтальной схемой федеративного машинного обучения, тогда как на практике злободневной является задача объединения данных об одних и тех же объектах, имеющихся у стейкхолдеров разных направлений деятельности (например, данных о клиентах в банках (кредитная история) и у телеком-компаний (данные о звонках)).

3. В диссертационной работе недостаточно детально проработана постановка задачи исследования (пункт 1.7 диссертации) – указано, что требуется разработать, но не показана взаимосвязь разрабатываемых дефиниций с целью исследования. Одновременно с этим при заявленной цели «*повышение эффективности процесса классификации...*» в пункте 1.6 диссертации рассматриваются *подходы к оцениванию производительности многоэлементной классификации*, тогда как, на мой взгляд, повышаться и

оцениваться в исследовании должно *качество многоэлементной классификации*.

4. Не понятно, почему автор, рассмотрев большинство метрик, применяемых при оценке качества классификации, остановился только на accuracy и не использовал точность (precision) и полноту (recall) классификации, традиционно используемые на практике.

5. 6. Значительно усилило бы работу демонстрация повышения качества классификации (или эффективности процесса классификации) на конкретном примере. Между тем, автор ограничил себя оценкой качества многоэлементной классификации по показателям точности (accuracy) и Каппа Коэна для централизованной и разработанной (гибридной) схемы. При этом само по себе сравнение данных схем затруднительно ввиду того, что оценка их качества визуализирована на разных рисунках (на рис. 4.15 диссертации и рис. 9 автореферата – для централизованной схемы; на рис. 4.16 диссертации и рис. 8 автореферата – для разработанной).

6. Во второй главе диссертации детально рассмотрена модификация EM-алгоритма для итерационного оценивания параметров модели гауссовой смеси распределений (GMM) при получении значений оценок вероятностной функции меток ненаблюдаемых классов локальных классификаторов. Однако в работе не представлена методика выбора начальных оценок параметров модели для реальных объектов (например, цифровых изображений), выбора (задания) порога сходимости для перехода на следующую итерацию EM-алгоритма.

7. На с. 11 автореферата утверждается, что «С использованием неравенства Йенсена было доказано, что множество локальных классификаторов, достигая минимума категориальных кросс-энтропийных потерь, обеспечивает минимум потерь многоэлементного классификатора». Между тем, доказательств того, что функции потерь локальных классификаторов являются выпуклыми функциями, в работе не представлено.

8. Не ясно, почему при обосновании фреймворка федеративного машинного обучения не рассматривалось отечественное решение с открытым кодом Stalactite, а выбор пал на американский FedML, который не поддерживается уже более двух лет.

9. Диссертация изобилует излишним обзорным материалом, например, шесть страниц четвёртой главы посвящено описанию фреймворков федеративного машинного обучения, тогда как для обоснования его выбора вполне достаточно сводной таблицы по ним (табл. 4.1 диссертации).

10. В тексте автореферата автор использует аббревиатуры некоторых терминов без указания их полного названия и не расшифровывает названия переменных, используемых в математических выражениях, что сильно осложняет прочтение автореферата без обращения к тексту диссертации.

11. Ряд замечаний касается оформления диссертации, например, на с. 32 диссертации указана ссылка на рис. 1.13, тогда как должна быть указана ссылка на рис. 1.12; в тексте диссертации расположена подрисуночная надпись для рис.1.12, однако вместо самого рисунка расположена чёрная рамка. Имеется в диссертации и автореферате ряд синтаксических, орфографических и стилистических ошибок и опечаток, таких как: «...представленная выше постановка задачи является труднореализуемой» (с. 34 диссертации) и т.д.

Однако, отмеченные замечания не снижают ценность проведённых исследований и не влияют на основные научные и практические результаты диссертации.

Заключение

Диссертационное исследование Михалева Павла Андреевича на тему «Управление децентрализованным обменом информацией для принятия решений в распределённых системах многоэлементной классификации с неполными данными» представляет собой законченную научно-квалификационную работу, в рамках которой решена актуальная научная

задача повышения эффективности распределённых систем многоэлементной классификации в условиях неполноты локальных элементных матриц классов.

Материал диссертационного исследования изложен логично и имеет чёткую структуру, полученные результаты согласуются с поставленной целью и задачами. Автореферат верно передаёт основное содержание диссертации.

Диссертация отвечает требованиям пунктов 9–11 Положения о присуждении ученых степеней, утверждённого постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842, предъявляемым к диссертациям на соискание ученой степени кандидата наук, а её автор – Михалев Павел Андреевич – заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.1 – Системный анализ, управление и обработка информации, статистика.

Главный инженер по внедрению технологий
акционерного общества «АСТ»
доктор технических наук, доцент

Басов Олег Олегович

20 апреля 2026 г.

115230 г. Москва, ул. Каширское шоссе, д.3, к.2, с.4
+7 (495) 679-86-86
hello@acti.ru

Подпись Басова О.О. удостоверяю
Ведущий специалист кадровой службы

С.В. Ульянов

22.04.2026

