ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертацию Перминова Андрея Игоревича

«Доверенный байесовский классификатор для данных малой размерности на основе многослойного перцептрона», представленную на соискание учёной степени кандидата физико-математических наук по специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»

Актуальность работы

Диссертационное исследование Перминова А.И. посвящено актуальной проблеме создания доверенных моделей машинного обучения. Несмотря на успехи нейросетевых методов, их применение в критически важных областях (медицина, финансы) сдерживается отсутствием формальных гарантий корректности решений, механизмов оценки уверенности и определения границ применимости. В работе обоснованно отмечается разрыв между эмпирическими нейросетевыми подходами и аппаратом математической статистики. Диссертант делает важный шаг в направлении построения математической теории для нейросетевых моделей, фокусируясь на задачах классификации в пространствах малой размерности. Предложенный подход к созданию доверенного классификатора на основе многослойного перцептрона, обладающего способностью к отказу от классификации вне носителя распределения и устойчивостью к дисбалансу классов, является своевременным и востребованным.

Основное содержание работы

В первой главе проведен обзор современных методов классификации в контексте требований доверенного искусственного интеллекта. Автор приходит к выводу о фрагментарности существующих решений и необходимости создания интегрированного, статистически обоснованного подхода.

Во второй главе излагается основной теоретический результат. Автор предлагает модификацию байесовского классификатора путем добавления искусственного фонового класса, что обеспечивает отказ от классификации вне носителя обучающей выборки. Для аппроксимации предложен метод на основе многослойного перцептрона, а для интерпретации его решений – оригинальный метод построения объясняющего дерева eXBTree. Экспериментально подтверждена эффективность подхода и его устойчивость к состязательным атакам.

Третья глава посвящена решению проблемы дисбаланса классов. Разработан метод унарной классификации, при котором для каждого класса строится свой классификатор, отделяющий его от равномерного фона. Предложены специализированные метрики для оценки качества. Проведены эксперименты на данных UCI и сравнение метода с моделью XGBoost.

В четвертой главе предложен метод генерации синтетических табличных данных на основе обученной унарной модели, сохраняющий геометрическую структуру исходного распределения. Сравнительный анализ с CTGAN показывает практическую полезность подхода для данных малой и средней размерности.

Пятая глава описывает разработанный программный комплекс для решения задач машинного обучения – автономное веб-приложение, реализующее все предложенные методы и служащее инструментом для верификации результатов.

Новизна и значимость результатов

Научная новизна работы заключается в разработке метода построения доверенного классификатора на основе многослойного перцептрона, который, в отличие от известных подходов, обеспечивает статистически обоснованную оценку апостериорной вероятности и формальный механизм отказа от классификации вне обучающего распределения. Важным результатом является оригинальный метод унарной классификации, устраняющий проблему

дисбаланса классов. Также предложены новый метод генерации синтетических данных, сохраняющих геометрическую структуру оригинала, и инструмент объяснения решений классификатора –eXBTree.

Теоретическая значимость работы

Теоретическая значимость работы подтверждена сформулированными и доказанными теоремами, обосновывающими предложенные методы и вносящими вклад в развитие статистических основ доверенного искусственного интеллекта. Практическая значимость обосновывается использованием разработанных методов в Исследовательском Центре Доверенного Искусственного Интеллекта ИСП РАН, а также созданием и публикацием в программные системы с открытым исходным кодом. Предложенные методы могут быть применены специалистами в области анализа данных и машинного обучения, разработчиками программного обеспечения для критических инфраструктур, исследователями в области доверенного искусственного интеллекта, а также инженерами, решающими задачи классификации в условиях дисбаланса классов и высокой неопределённости данных.

Достоинства работы

К достоинствам работы стоит отнести глубокую методологическую проработку исследования, в котором органично сочетаются строгие теоретические результаты с их практической реализацией в виде работающего наглядного программного инструмента. Особого внимания заслуживает подход к визуализации и интерпретации решений нейросетевых моделей (дерево eXBTree), что является важным шагом к преодолению проблемы «чёрного ящика» в машинном обучении. Также, значительным достоинством является наличие открытой программной реализации.

Замечания к содержанию и оформлению диссертации

1. Большинство эмпирических подтверждений эффективности предложенных классификаторов представлено на простых двумерных наборах данных (например, структура «кольцо–круг», двухвитковые спирали, шахматная разметка 2×2, пересечение гауссовых распределений). При этом не исследована переносимость предложенных подходов на более сложные задачи.
2. В работе отсутствуют результаты апробации модифицированного байесовского классификатора на практико-ориентированных задачах, что затрудняет объективную оценку практической применимости разработанных методов.
3. Для наборов данных MNIST и CIFAR-10 не проведено сравнение с другими архитектурами нейросетевых моделей. Кроме того, достигнутая на CIFAR-10 точность (0,53) является крайне низкой на фоне большинства современных свёрточных архитектур.
4. По задаче унарной классификации отсутствует сравнение метрик качества с альтернативными реализациями одноклассовых классификаторов. Сравнение с моделями на основе градиентного бустинга не включает более актуальные реализации CatBoost и LightGBM, а также не поднимает вопрос настройки гиперпараметров.
5. По задаче генерации синтетических данных выполняется сравнение только с одним методом генерации – CTGAN. При этом, улучшение верности данных достигнуто только на наиболее простом тестовом примере.
6. В целом в диссертации не обсуждаются вопросы конкурентоспособности предложенных решений по сравнению с современными свёрточными архитектурами нейронных сетей в задачах классификации и генерации синтетических данных.
7. В тексте присутствует упоминание о публикации проекта на платформе GitHub, но не указаны ссылка на репозиторий.

Отмеченные замечания не снижают значимости полученных результатов и не влияют на общую положительную оценку диссертационного исследования.

Заключение

Таким образом, диссертация Перминова Андрея Игоревича по теме «Доверенный байесовский классификатор для данных малой размерности на основе многослойного перцептрона» является самостоятельным и завершённым исследованием, обладающим научной новизной и практической значимостью. Работа отвечает требованиям ВАК о порядке присуждения учёных степеней к кандидатским диссертациям, а соискатель заслуживает присуждения учёной степени кандидата физико-математических наук по специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Официальный оппонент
Никитин Николай Олегович
кандидат технических наук, руководитель группы научно-технического развития
Федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет ИТМО»

«29» марта 2026 г.