ОТЗЫВ

официального оппонента на диссертационную работу Ширяева Егора Михайловича «Математическая модель, методы и алгоритмы эффективной реализации искусственных нейронных сетей, сохраняющих конфиденциальность» представленную на соискание ученой степени кандидата физико-математических наук по специальности 2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей

Актуальность темы. Диссертационная работа Ширяева Егора Михайловича посвящена решению актуальной научной задачи исследования и разработки математической модели, методов и алгоритмов для повышения эффективности реализаций искусственных нейронных сетей, сохраняющих конфиденциальность на основе схем полностью гомоморфного шифрования (ПГШ). Бурное развитие методов, основанных на искусственных нейронных сетях, позволило расширить множество решаемых практических задач. Они достаточно активно внедряются и используются во многих областях нашей жизни, значительно ускоряя множество производственных процессов. Однако необходимо отметить ряд прецедентов утечки конфиденциальной информации при применении этих методов, что является сдерживающим фактором для внедрения во многих организациях. Для решения данной проблемы требуется разработка специализированного программного обеспечения, позволяющего эффективно реализовать функционал искусственных нейронных сетей на базе схем гомоморфного шифрования.

Полностью гомоморфное шифрование в области обеспечения конфиденциальности обрабатываемых искусственными нейронными сетями данных, позволяет обрабатывать зашифрованные данные с помощью гомоморфного сложения и умножения. В контексте искусственных нейронных сетей операции сложения и умножения являются ключевыми. Основными препятствиями для применения схем ПГШ являются высокая вычислительная сложность и ограниченность набора поддерживаемых операций, как правило сложением и умножением. Возникает дополнительная задача разработки и адаптации моделей, методов и алгоритмов, как гомоморфных операций ПГШ, так и самих искусственных нейронных сетей.

Таким образом, можно сделать вывод о том, что тема диссертационной работы, направленной на решение перечисленных выше задач, является чрезвычайно актуальной.

Структура и содержание диссертации. Диссертационная работа состоит из введения, трех глав, заключения, списка литературы,

включающего 186 использованных источников. Общий объем работы составляет 156 страниц.

Во введении показана актуальность темы диссертации, сформулированы цель и основные задачи работы, обоснованы научная новизна и практическая значимость результатов проведенного исследования, представлены основные положения, выносимые на защиту.

Первая глава посвящена вопросам конфиденциальности в распределенных вычислительных системах, в частности, облачных и туманных вычислениях. В качестве основного метода ее обеспечения выбрано гомоморфное шифрование, среди схем которого подробно рассмотрены TFHE, CKKS и BFV. Показано, что применение гомоморфных шифров позволяет выполнять вычисления над зашифрованными данными, что исключает необходимость их расшифровки и обеспечивает защиту информации на всех этапах обработки. Особое внимание уделено возможностям интеграции гомоморфного шифрования с методами искусственного интеллекта, а именно со сверточными нейронными сетями, для безопасного функционирования в публичных распределенных системах.

Во второй главе проведено исследование библиотек и алгоритмов, реализующих схемы гомоморфного шифрования. Определены и выбраны для дальнейшего анализа библиотеки TenSEAL и Concrete-ML, предоставляющие инструментарий для работы с гомоморфными шифрами и их интеграции в искусственные нейронные сети. Исследованы методы конфиденциального матричного умножения. Сформулированы и доказаны теоремы об умножении зашифрованной матрицы на открытую и об аппроксимации функции искусственной нейронной сетью. На основе полученных результатов предложена модификация алгоритма гомоморфного умножения матриц, позволившая существенно сократить требования к памяти и вычислительную сложность. Проведено исследование методов дистилляции сверточных нейронных сетей, модифицирован метод дистилляции для задач классификации и распознавания образов. Определены оптимальные параметры метода: константа контроля ошибки и размер коалиции сетей-учителей, что позволило радикально уменьшить размеры моделей при незначительной потере точности. Исследованы методы квантизации на базе схем гомоморфного шифрования CKKS и BFV. Установлено, что метод квантизации 2^N демонстрирует наилучшее соотношение скорости обработки и точности распознавания. Показано, что схема BFV уступает схеме CKKS как по точности, так и по скорости выполнения операций.

Третья глава посвящена разработке математических моделей сверточных нейронных сетей, как открытых, так и обеспечивающих

конфиденциальность, на базе нескольких приближенных функций активации. Разработанные модели исследованы в два этапа на наборах данных MNIST с изображениями размером 28×28 и 8×8: первый этап – исследование моделей без применения дистилляции, второй – с ее применением. В рамках исследования построены модели СНС-учителей и СНС-учеников, получены полиномиальные аппроксимации функций активации, а для функции с обучаемыми коэффициентами в ходе обучения определены полиномы второй степени.

Установлено, что схема TFHE обрабатывает данные в среднем на 50% быстрее, тогда как схема CKKS обеспечивает в среднем на 5% более высокую точность распознавания. Определено, что модели на базе TFHE демонстрируют наилучшее соотношение скорости и точности при работе с дистиллированными сетями и данными размера 8×8, в то время как модели на базе CKKS показывают наилучшие результаты при распознавании изображений 28×28. Наивысшая точность распознавания достигнута на основе функции активации с обучаемыми коэффициентами в моделях CKKS. Таким образом, в ходе исследования получены оптимальные модели сверточных нейронных сетей, сохраняющих конфиденциальность, для различных наборов данных, обеспечивающие требуемое соотношение точности и скорости.

Разработанный на основе полученных результатов программный комплекс позволяет расширить применение моделей СНС за счет возможности безопасной обработки конфиденциальных данных в общедоступных облаках и туманных вычислениях.

В заключении отражены основные результаты диссертационной работы и сформулированы выводы об эффективности их использования, а также представлены направления дальнейших исследований.

Автореферат в целом адекватно отражает структуру и содержание диссертации.

Соответствие паспорту научной специальности. Тема, содержание и основные результаты диссертации полностью соответствуют областям исследования паспорта специальности 2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Научная новизна определяется следующими результатами диссертационного исследования:

- предложен модифицированный метод умножения матриц для реализации слоев искусственной нейронной сети, сохраняющей

конфиденциальность, позволяющий уменьшить вычислительную сложность алгоритма с $O(n^4)$ до $O(n^2)$;

- разработаны методы дистилляции и квантизации параметров для адаптации искусственных нейронных сетей к ограничениям гомоморфных шифров, позволяющие сократить потребление памяти примерно в 1500 раз, уменьшить время обработки данных в среднем в 30 раз, при потере доли верных классификаций в диапазоне от 0.5% до 1%;
- разработаны функции активации для проектирования искусственных нейронных сетей, сохраняющих конфиденциальность;
- разработан программный комплекс для реализации искусственных нейронной сети, сохраняющих конфиденциальность.

Теоретическая и практическая значимость. Теоретическая значимость данной диссертационной работы заключается в разработанных математической модели, методах и алгоритмах, включая методы повышения эффективности операции умножения матриц для гомоморфных шифров, методы адаптации искусственных нейронных сетей, учитывающие ограничения гомоморфных шифров, такие как дистилляция знаний и квантизация параметров, алгоритмы полиномиального приближения функций активации и функции активации на основе полиномов с обучаемыми коэффициентами.

Практическая значимость заключается в возможности создания на основе разработанных моделей и алгоритмов конфиденциальных искусственных нейронных сетей, способных обрабатывать данные в зашифрованном виде. Результаты работы могут быть применены для построения нейросетевых моделей в областях с повышенными требованиями к защите информации, таких как здравоохранение, финансовый сектор и банковская деятельность.

Результаты диссертационной исследования были использованы при выполнении грантов, поддержанных Российским научным фондом. В рамках работы также получены три свидетельства о государственной регистрации программ для ЭВМ.

Достоверность и обоснованность научных результатов диссертационной работы обусловлены строгостью проведённых математических доказательств, в основе которых лежит аппарат математического анализа, теории чисел и численных методов. Кроме того, они подтверждаются сравнительным анализом, показавшим преимущества разработанных методов перед известными аналогами по скорости обработки данных, объёму потребляемой памяти и точности распознавания образов.

Апробация. Основные результаты диссертации в полной мере представлены на профильных российских и международных конференциях и опубликованы в 29 научных работах, включая 4 публикаций в рецензируемых научных изданиях, рекомендованных ВАК РФ, 8 работ в журналах, индексируемых в международных базах цитирования Web of Science, и 15 в тезисах докладов конференций. Кроме того, получены 3 свидетельства о государственной регистрации программ для ЭВМ. Указанный авторский вклад соискателя не вызывает сомнений.

Замечания. Диссертация выполнена на высоком профессиональном уровне. В тоже время к данной работе имеется ряд частных замечаний, которые не влияют на несомненную теоретическую и практическую ценность диссертационного исследования в целом.

1. В качестве критериев искусственных нейронных сетей, сохраняющих конфиденциальность, целесообразно было бы так же рассмотреть и вычислительную безопасность.
2. В работе предложены методы проектирования нейронных сетей, сохраняющих конфиденциальность, однако, не исследован вопрос о применимости этих методов при проектировании больших языковых моделей, сохраняющих конфиденциальность. Сохраняется ли эффективность предложенных подходов при переходе к реальным, более масштабным задачам? К каким реальным задачам применимы предложенные подходы?
3. Часть представленных графиков (например, рис. 3.7-3.10) сложна для восприятия из-за перегруженности данными. Для улучшения наглядности целесообразно было бы вынести детализированные данные в дополнительные таблицы, размещённые в приложениях.
4. В тексте диссертации присутствуют опечатки и пунктуационные ошибки, не искажающие смысла изложения. В частности, допущена ошибка в названии библиотеки Concrete-ML (стр. 128) и в расшифровке аббревиатуры «СНС» (стр. 130).

Заключение. Диссертационная работа Ширяева Егора Михайловича выполнена на высоком научном уровне по актуальному направлению исследований. Ее содержание полностью соответствует поставленным целям и задачам, обладает внутренним единством и правильно отражает логическую последовательность выполнения основных этапов исследования. Диссертация изложена в хорошем научном стиле, ясно и доказательно. Основные результаты, положения и выводы диссертации являются вполне достоверными и обоснованными.

В целом диссертация Ширяева Егора Михайловича является самостоятельно подготовленной, завершенной научно-квалификационной работой, в которой на основании выполненных автором исследований эффективно решается актуальная задача исследования и разработки математической модели, методов и алгоритмов для повышения эффективности реализаций искусственных нейронных сетей, сохраняющих конфиденциальность на основе схем полностью гомоморфного шифрования.

Считаю, что диссертация соответствует всем требованиям Положения ВАК РФ о присуждении ученых степеней, предъявляемым к кандидатским диссертациям, а ее автор, Ширяев Егора Михайлович, заслуживает присуждения ему искомой ученой степени кандидата физико-математических наук по специальности 2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Официальный оппонент
заведующий лабораторией комплексных информационных систем, Федеральное государственное бюджетное учреждение науки Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук, доктор технических наук — Фёдоров Роман Константинович

Согласен на обработку персональных данных, доктор технических наук — Фёдоров Роман Константинович

Контактные данные организации: Федеральное государственное бюджетное учреждение науки Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российском академии наук (ИДСТУ СО РАН), 664033, г. Иркутск, ул. Лермонтова, д.134, e-mail: idstu@icc.ru, тел: +7 (3952) 42-71-00

31.10.2025