УДК 004.89

ФЕДЕРАТИВНОЕ ОБУЧНИЕ С ДИФФЕРЕНИЦАЛЬНОЙ ПРИВАТНОСТЬЮ: АНАЛИЗ ЗАВИСИМОСТИ КАЧЕСТВА МОДЕЛИ ОТ УРОВНЯ ПРИВАТНОСТИ

Петрова А. А.[1]
Научный руководитель – канд. техн. наук Логинов И. П.[1]
[1]Университет ИТМО
loginov@itmo.ru

Введение

Современные системы машинного обучения требуют больших объёмов данных для обучения, однако в здравоохранении, финансах и телекоммуникациях данные носят конфиденциальный характер. Федеральный закон № 152-ФЗ и европейский регламент GDPR существенно ограничивают возможности их централизованного сбора, что делает традиционные подходы к обучению моделей неприменимыми. Таким образом, возникает научная проблема: как обучать эффективные модели без прямого доступа к исходным данным.

Существующим решением является федеративное обучение – подход, при котором обучение модели происходит непосредственно на устройствах участников, а центральный сервер получает лишь обновления параметров, но не исходные данные [5]. Тем не менее данный механизм не обеспечивает полноценной защиты: ряд исследований продемонстрировал возможность извлечения конфиденциальной информации из передаваемых обновлений, что указывает на необходимость дополнительных мер обеспечения приватности [6].

В зарубежной практике данная проблема решается путём совмещения федеративного обучения с дифференциальной приватностью. Google применяет этот подход для обучения языковых моделей на устройствах пользователей, медицинские консорциумы в Европе – для совместного обучения диагностических систем без обмена данными пациентов. Математические основы дифференциальной приватности заложены в работах Dwork и Roth [4] , практическая реализация обеспечена библиотекой Opacus. В отечественной науке исследования в данной области немногочисленны и не охватывают систематического анализа компромисса между уровнем приватности и качеством модели, что определяет актуальность настоящей работы [3].

Выводы

Совместное применение федеративного обучения и дифференциальной приватности обеспечивает математически обоснованную защиту данных при сохранении возможности обучения моделей машинного обучения в распределённых системах. Экспериментально установлено, что значения параметра ε в диапазоне от 5 до 10 позволяют достичь приемлемого компромисса между уровнем приватности и качеством модели для большинства прикладных задач. При значениях ε ниже 1 качество модели снижается до неприемлемого уровня, что ограничивает практическое применение строгой приватности [1, 2].

В качестве предложений по внедрению рекомендуется апробация разработанной методики выбора параметра ε на реальных данных финансовых организаций с последующей выработкой отраслевых стандартов применения дифференциальной приватности. Реализация предложенного подхода возможна на основе открытой библиотеки Opacus в среде PyTorch без необходимости разработки специализированной инфраструктуры, что существенно снижает порог внедрения для организаций.

Литература

1. Научный журнал “ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И КОММУНИКАЦИИ” [Электронный ресурс]. – Режим доступа: https://ijitt.ru/index.php/component/content/article/19-2025-13-2/182-2025-13-2-52-68 (Дата обращения 22.02.2026)
2. Opacus Documentation [Электронный ресурс]. – Режим доступа: https://docs.antigranular.com/references (Дата обращения 22.02.2026)
3. Серезлеев Д. С., Абаев Ю. К. Метода дифференциальной анонимизации данных на основе доверительной нейронной сети для защиты персональной информации клиентов банка: [Электронный ресурс]. – Режим доступа: http://www.ivdon.ru/uploads/article/pdf/IVD_57N12y25_serezleev_abaevy.pdf_de0c2df034.pdf (Дата обращения 22.02.2026)
4. Dwork C. The Algorithmic Foundations of Differential Privacy / C. Dwork, A. Roth // Foundations and Trends in Theoretical Computer Science. – 2014. – Vol. 9, № 3–4. – P. 211–268.C
5. S. Arora, E. Hazan, and S. Kale. The multiplicative weights update method: A meta-algorithm and applications. Theory of Computing, 8(1):121–164, 2012.
6. Y. Chen, S. Chong, I. A. Kash, T. Moran, and S. P. Vadhan. Truthful mechanisms for agents that value privacy. Association for Computing Machinery Conference on Electronic Commerce, 2013.