УДК 004.056

В.С. КИРЕЕВ
Национальный исследовательский ядерный университет «МИФИ», Москва

СОВРЕМЕННЫЕ МЕТОДЫ ФЕДЕРАТИВНОГО МАШИННОГО ОБУЧЕНИЯ

Федеративное обучение (Federated Learning, FL) как новая парадигма в искусственном интеллекте машинном обучении, находится на стыке задач доверенного искусственного интеллекта и конфиденциальных вычислений, и подразумевает различные формы отказа от централизации данных и переноса обучения непосредственно на границу пользовательского устройства. В данном. докладе автором представлены результаты обзора основных методов FL, обеспечивающих баланс между конфиденциальностью, производительностью, точностью, стоимостью накладных расходов федеративного обучения.

Введение

Федеративное обучение позволяет нескольким клиентам совместно обучать модель машинного обучения, используя преимущества различных наборов данных от клиентов без обмена своими локальными обучающими наборами данных. Дискуссии по применению конфиденциальных вычислений получили особенное развитие после 8 августа 2024 г., когда был подписан закон с новыми правилами обработки обезличенных персональных данных [1]. Например, при участии Ассоциации ФинТех (АФТ) обсуждаются вопросы конфиденциального обмена данными для построения моделей скоринга и антифрода. К сожалению, FL все еще имеет ряд проблем, связанных с конфиденциальностью и безопасностью, описываемые далее.

Конфиденциальное федеративное обучение

Риски атак на конфиденциальность в FL возникают, когда центральный сервер или другие участники получают доступ к параметрам модели, совместно используемым на этапах обучения или агрегирования.

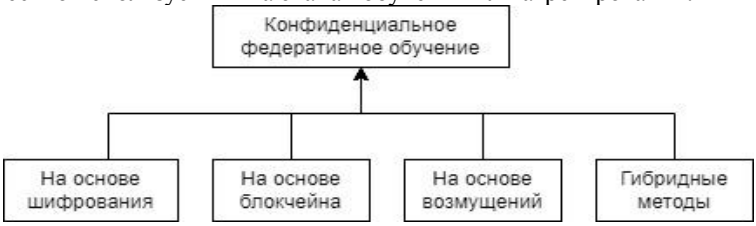

Рис. 1. Подходы к конфиденциальному федеративному обучению

Кроме того, вредоносные клиенты могут вступить в сговор друг с другом, чтобы украсть данные, модели у обычных клиентов или испортить глобальную обучающую модель. Для снижения этих рисков применяются методы, которые можно обобщенно сгруппировать, как на рис. 1:

На основе шифрования. Данные методы используют безопасные многосторонние вычисления (Secure Multi-Party Computation, SMPC) [2], например, с использованием доверенных анклавов (Trusted Execution Environments, TEE) и гомоморфное шифрование (Homomorphic Encryption, HE) [3, 4].

На основе возмущения. Эти методы в целом делятся на две категории: локальная дифференциальная конфиденциальность (LDP), реализуемая на стороне клиента, и глобальная дифференциальная конфиденциальность (GDP) [4], организуемая на стороне сервера. Эти методы вносят контролируемый шум или рандомизацию в обновления модели перед их объединением.

На основе блокчейна. Методы на основе блокчейна используют серию неизменяемых блоков данных с временными метками, управляемых распределенной сетью компьютеров, что исключает зависимость от одного контролирующего лица [3].

Гибридные методы применяют интеграцию шифрования, возмущения и технологии блокчейн, что позволяет создать комплексную защиту конфиденциальности [3].

Заключение

В докладе автором рассмотрены современные конфиденциальные методы федеративного обучения, основанные на шифровании, на основе блокчейна, на основе возмущений, и гибридные подходы. Описаны их достоинства и недостатки.

Список литературы

1. Федеральный закон № 233-ФЗ [Электронный ресурс]. URL: https://www.garant.ru/hotlaw/federal/1746724/ (дата обращения: 13.09.2024).

2. Предварительный национальный стандарт № 845-2023. Техническая структура федеративной системы машинного обучения [Электронный ресурс]. URL: https://www.garant.ru/hotlaw/federal/1746724/ (дата обращения: 13.09.2024).

3. Mo F., Haddadi H., Katevas K., Marin E., Perino D., Kourtellis N.. Privacy-preserving federated learning with trusted execution environments. Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services (2021), p. 94–108

4. Запечников С.В. Конфиденциальное машинное обучение на основе трехсторонних протоколов безопасных вычислений. Безопасность информационных технологий, 2022, № 1 (29), с. 30–43. DOI: http://dx.doi.org/10.26583/bit.2021.4.03.