УДК 004.852+336.7

Федеративное обучение в финансовом секторе: обнаружение мошенничества и прогнозирование финансовых трудностей

Н.М. Кошелев, А.В. Тарлыков, А.П. Преображенский

Воронежский институт высоких технологий, Воронеж, Россия

В данной работе рассматривается федеративное обучение – метод совместного обучения моделей машинного обучения на распределённых данных без их физической централизации. Показано, каким образом ключевым фактором распространения данной технологии послужило регуляторное давление GDPR, а не абстрактная забота о конфиденциальности. Подробным образом анализируются теоретические основы федеративного обучения и алгоритм FedAvg, а также два прикладных применения – обнаружение мошенничества с кредитными картами (F1 = 77%) и прогнозирование финансовых трудностей потребителей США по данным NFCS (F1 = 42,2%, при потере относительно централизованной модели менее 0,5 п.п.). Вместе с тем рассматриваются атаки восстановления данных из градиентов, проблема неоднородности участников и правовая неопределённость внедрения. Вывод: федеративное обучение представляет собой полезный, однако условно применимый инструмент с конкретными техническими и институциональными ограничениями.

Ключевые слова: федеративное обучение, машинное обучение, финансы, обнаружение мошенничества, кредитный риск, конфиденциальность данных.

Federated Learning in the Financial Sector: Fraud Detection and Financial Distress Prediction

N.M. Koshelev, A.V. Tarlykov, A.P. Preobrazhenskiy

Voronezh Institute of High Technologies, Voronezh, Russia

This paper considers the method of federated learning – a technique for training machine learning models on distributed data without centralizing it physically. It is shown how the key driver of this technology's adoption in banking was regulatory pressure from GDPR rather than abstract privacy concerns. Detailed consideration is given to the theoretical foundations of federated learning and the FedAvg algorithm, as well as two applied cases – credit card fraud detection (F1 = 77%) and U.S. consumer financial distress prediction using NFCS data (F1 = 42.2%, less than 0.5 p.p. below a centralized baseline). Additionally, gradient inversion attacks, participant data heterogeneity, and legal uncertainty around deployment are examined. The conclusion is that federated learning constitutes a useful yet conditionally applicable tool subject to specific technical and institutional constraints.

Keywords: federated learning, machine learning, finance, fraud detection, credit risk, data privacy.

Введение

Введение регуляторных требований GDPR в мае 2018 года поставило перед европейскими банковскими институтами конкретную задачу: обеспечить совместное обучение моделей машинного обучения при одновременном соблюдении законодательных ограничений на централизацию персональных данных. Каждый крупный банк располагал тысячами размеченных примеров мошеннических транзакций – объёмом, недостаточным для надёжного обучения модели, – однако любая попытка объединить эти данные с отраслевыми партнёрами немедленно порождала правовые

проблемы, связанные с вопросами персональных данных, трансграничной передачи и ответственности за утечки. В данной связи федеративное обучение (ФО) сформировалось как технический ответ на данное регуляторное ограничение.

Следует отметить принципиально важное обстоятельство: распространение технологии в банковском секторе определяется не абстрактной заботой о правах клиентов на приватность, а конкретными требованиями комплаенс-подразделений к выполнению норм GDPR при сохранении возможности обучать модели на достаточном объёме данных. Внедрение идёт снизу вверх – из юридических и комплаенс-отделов, которым требуется решить конкретную задачу. Вопросы применения машинного обучения в финансовой сфере детально рассмотрены в работе [1].

В рамках ФО централизации подвергается не массив данных, а сам процесс обучения. Каждый участник системы обучает копию глобальной модели на локальных данных и передаёт на агрегирующий сервер обновления параметров – градиенты. Сервер выполняет их взвешенное усреднение по алгоритму FedAvg [2] и рассылает обновлённую модель участникам. Процесс повторяется итерационно, и транзакционные записи при этом не покидают периметр банка-участника.

Вместе с тем необходимо принимать во внимание ограничение, которое нередко остаётся вне поля рассмотрения: передача градиентов не является передачей «ничего». Существует направление атак восстановления данных из градиентов, в рамках которых при определённых условиях из переданных обновлений возможна частичная реконструкция исходных обучающих примеров. Тезис о полной конфиденциальности в базовом варианте ФО является упрощением, а не установленным фактом. В данной работе проводится рассмотрение двух прикладных применений ФО: обнаружения мошенничества с кредитными картами [3] и прогнозирования финансовых трудностей потребителей США [4].

Применение искусственного интеллекта в финансах: что работает, а что преувеличено

До перехода к рассмотрению ФО целесообразно дать оценку применению машинного обучения в финансах в общем [1]. Алгоритмы обнаружения аномалий демонстрируют превосходство над системами на основе жёстких правил при работе с многомерными данными. Правило вида «заблокировать транзакцию выше порога X из страны Y» перестаёт обеспечивать надлежащую защиту, как только мошенники определяют его параметры методом проб и ошибок. Модель, одновременно обрабатывающая сотни признаков, значительно сложнее систематически обойти – это обусловлено более высокой размерностью признакового пространства, а не архитектурными преимуществами конкретных алгоритмов.

Кредитный скоринг на основе альтернативных данных также демонстрирует реальное улучшение показателей, в особенности для клиентов с ограниченной кредитной историей. Выигрыш достигается за счёт расширения информационной базы принятия решений. Вместе с тем ряд возможностей машинного обучения в финансах систематически преувеличивается: точность моделей существенно снижается при смене рыночного режима – модель, обученная на данных 2015–2019 годов, демонстрировала неудовлетворительную эффективность в марте 2020 года. Данное ограничение является принципиальным, и масштабирование данных эту проблему не устраняет [1].

Математические основы федеративного обучения

Формально задача ФО состоит в минимизации взвешенной суммы локальных функций потерь [2]. Задача представляется следующим образом:

$$J(W) = \sum_{k=1}^{K} \frac{n_k}{n} L_k(W), \quad (1)$$

где L_k – функция потерь k-го участника, n_k – размер его обучающего набора, $n = \sum n_k$ – суммарный объём данных, W – параметры глобальной модели. Алгоритм FedAvg [2] обеспечивает итерационное решение задачи (1): сервер рассылает текущие параметры случайному подмножеству участников; каждый участник выполняет несколько эпох локального SGD на своих данных; участники возвращают обновлённые параметры; сервер проводит их взвешенное усреднение с коэффициентами, пропорциональнымиn_k.

По существу, речь идёт о распределённом SGD с разреженными коммуникациями. Практические трудности возникают в условиях неоднородного распределения данных у различных участников, что в финансовой сфере является нормой, а не исключением. Небольшой региональный банк и крупный международный институт обслуживают принципиально различную клиентуру. В подобной ситуации FedAvg сходится к компромиссному решению, приемлемому для среднего участника, однако потенциально уступающему специализированной локальной модели каждого конкретного банка.

Относительно гарантий конфиденциальности необходимо отметить следующее. ФО в базовом варианте представляет собой технику распределённого обучения, а не средство защиты данных в полном смысле. Для обеспечения реальных гарантий требуются дополнительные механизмы: дифференциальная конфиденциальность, безопасная агрегация и гомоморфное шифрование. Задача определения оптимального баланса между защитой и качеством модели остаётся открытым исследовательским вопросом.

Задача 1: Обнаружение мошенничества с кредитными картами

В первом рассматриваемом применении [3] используется набор данных транзакций европейских держателей кредитных карт за сентябрь 2013 года: около 285 тысяч транзакций, из которых мошеннических – 492 (0,172%). Данные предварительно обработаны методом PCA, что одновременно обеспечивает снижение размерности и анонимизацию исходных признаков. Архитектура классификатора – многослойный перцептрон (MLP); тип задачи – бинарная классификация.

Отдельного рассмотрения требует проблема несбалансированности классов. Следует отметить, что модель, классифицирующая все транзакции как немошеннические, достигает accuracy = 99,83%, что представляет собой вводящую в заблуждение метрику. Корректной метрикой оценки в данной задаче служит F1-score по классу мошенничества, и в особенности recall: пропуск реального мошенничества, как правило, обходится банку дороже, чем ложная блокировка легитимной транзакции.

В федеративной постановке каждый условный участник обучает MLP на своём подмножестве данных; агрегация осуществляется по алгоритму FedAvg в соответствии с выражением (1). Получены следующие результаты: precision = 82%, recall = 89%, F1 = 77%. Для сравнения: логистическая регрессия в тех же условиях даёт F1 ≈ 68%, дерево решений – F1 ≈ 72%. Преимущество MLP обусловлено способностью к захвату нелинейных взаимодействий между признаками.

Заслуживает внимания один контринтуитивный результат. При реалистичном соотношении классов 1:100 модель на начальных итерациях показывала результат

хуже, чем при искусственно сбалансированных данных. Однако с ростом числа итераций ситуация менялась: модель, обученная на реальном несбалансированном распределении, в итоге превзошла сбалансированную версию. Из этого следует практический вывод о нецелесообразности применения oversampling при наличии достаточного числа наблюдений. Необходимо дать объективную оценку полученным результатам: F1 = 77% является хорошим, но не предельным показателем; централизованные модели на аналогичных данных достигают F1 > 85%, и разрыв в 8–10 п.п. представляет собой реальную цену конфиденциальности [3].

Задача 2: Прогнозирование финансовых трудностей потребителей США

Вторая рассматриваемая задача [4] характеризуется более богатой методологией. Исходные данные – Национальное исследование финансовых возможностей США (NFCS): более 25 тысяч респондентов, 126 вопросов об их финансовой жизни. Целевая переменная – факт контакта с коллекторским агентством в течение предшествующих 12 месяцев. В рамках федеративной постановки каждый из 50 штатов и округ Колумбия выступают как отдельный участник, итого 51 клиент.

Выбор целевой переменной требует отдельного рассмотрения. Контакт с коллектором является поздним индикатором финансовых трудностей: к данному моменту заёмщик уже несколько месяцев не выполняет платёжные обязательства, ситуация приобрела острый характер. Для задач проактивной поддержки необходимы более ранние предикторы. Авторы работы [4] сами указывают на данное ограничение, объясняя выбор переменной доступностью данных NFCS. Из этого следует, что построенная модель обнаруживает уже сложившийся финансовый кризис, а не прогнозирует его заблаговременно.

Применяемая архитектура – восьмислойная highway network [5], разработанная для преодоления проблемы исчезающего градиента в глубоких сетях. В каждом слое реализованы «ворота» (transform gate и carry gate), управляющие степенью трансформации информации на данном уровне или её сквозной передачей. Набор признаков включает 12 переменных: демографические характеристики, доход, статус занятости и жилищные условия. Информация о конкретных долгах намеренно исключена в целях конфиденциальности.

Ключевые результаты: глобальная федеративная модель – F1 = 42,2%; централизованная модель – F1 = 42,4%; средняя локальная модель (данные одного штата) – F1 = 33,3%. Прирост от федеративного подхода относительно локального обучения составил около 9 п.п.; потеря относительно централизованного подхода – менее 0,5 п.п. Данный результат убедительно демонстрирует практическую ценность метода. В каждом раунде случайным образом отбирались 12 из 51 штата; суммарный объём переданных данных сократился с 13 ГБ до менее 5 ГБ за 200 раундов без существенных потерь качества.

Анализ важности признаков посредством Owen values [6] позволил выявить наиболее содержательный результат работы: финансовое образование оказалось среди наименее предсказательных переменных. Данный вывод противоречит распространённому представлению о том, что финансовая грамотность сама по себе снижает риск финансовых трудностей. Данные свидетельствуют об ином: знание финансовых инструментов слабо предсказывает реальное поведение при известных уровне дохода, занятости и демографических характеристиках.

Вызовы реального внедрения

Стандартные барьеры внедрения ФО достаточно подробно задокументированы в литературе. В данном разделе проводится рассмотрение трёх менее очевидных проблем.

Координация в конкурентной среде. Банки, которым предлагается совместное обучение антифрод-модели, как правило, являются конкурентами на одном рынке. Даже при безупречных технических гарантиях конфиденциальности сохраняется стратегический вопрос о мотивации банка косвенно улучшать систему конкурента через агрегированные обновления. Данный вопрос не имеет технического решения; его разрешение обеспечивается правовыми соглашениями, прозрачными экономическими стимулами и нейтральным координатором.

Неравенство участников. Алгоритм FedAvg взвешивает обновления пропорционально объёму данных в соответствии с формулой (1). Крупный банк с сотнями миллионов транзакций будет доминировать в агрегации над региональным банком. В результате глобальная модель оптимизируется преимущественно под клиентуру крупных участников. Существуют модификации алгоритма (FedProx, q-FedAvg), смягчающие данную проблему, однако не устраняющие её полностью.

Правовая неопределённость. В большинстве европейских юрисдикций по-прежнему не имеется однозначного ответа на вопрос о том, является ли передача градиентов «обработкой персональных данных» в смысле GDPR. Финансовые институты принимают решения о внедрении ФО в условиях, когда стоимость юридического заключения о соответствии регулятивным нормам сопоставима или превышает стоимость самого внедрения.

Заключение

Федеративное обучение решает реальную, практически значимую задачу: обеспечение совместного обучения моделей в условиях регуляторных ограничений на передачу данных. Рассмотренные эксперименты [3, 4] подтверждают практическую ценность подхода: F1 = 77% на задаче обнаружения мошенничества без централизации транзакционных данных; потеря менее 0,5 п.п. по F1 относительно централизованной модели при выигрыше в 9 п. п. над локальными моделями. Таким образом, проведённое рассмотрение показало, что ФО представляет собой инструмент с конкретными условиями применимости, а не универсальное решение проблемы конфиденциальности. Реальное внедрение сдерживается не только техническими барьерами, но и организационными и правовыми ограничениями, которые являются значительно более трудноустранимыми. Наиболее перспективным направлением следует считать персонализированные модификации ФО (pFedMe, DITTO), позволяющие каждому участнику получать индивидуализированную версию глобальной модели. Именно в данном направлении сосредоточена наиболее активная исследовательская деятельность, и именно здесь ФО может получить реальное конкурентное преимущество перед централизованными подходами.

СПИСОК ИСТОЧНИКОВ

1. Kumar A. Redefining Finance: The Influence of Artificial Intelligence (AI) and Machine Learning (ML) / A. Kumar // arXiv [Электронный ресурс]. – URL: https://arxiv.org/abs/2410.15951 (дата обращения: 16.01.2026).

2. Communication-Efficient Learning of Deep Networks from Decentralized Data / B. McMahan, E. Moore, D. Ramage [et al.] // Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20–22 April 2017, Fort Lauderdale, FL, USA. – PMLR, 2017. – P. 1273–1282.

3. Sha X. Research on financial fraud algorithm based on federal learning and big data technology / X. Sha // arXiv [Электронный ресурс]. – URL: https://arxiv.org/abs/2405.03992 (дата обращения: 09.02.2026).

4. Carta L. Explainable Federated Learning for U.S. State-Level Financial Distress Modeling / L. Carta, F. Spadea, O. Seneviratne // arXiv [Электронный ресурс]. – URL: https://arxiv.org/abs/2511.08588 (дата обращения: 20.01.2026).

5. Srivastava R.K. Highway Networks / R.K. Srivastava, K. Greff, J. Schmidhuber // arXiv [Электронный ресурс]. – URL: https://arxiv.org/abs/1505.00387 (дата обращения: 13.01.2026).

6. Lundberg S.M. A Unified Approach to Interpreting Model Predictions / S.M. Lundberg, S.-I. Lee // Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 04–09 December 2017, Long Beach, CA, USA. – 2017. – P. 4765–4774.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Кошелев Никита Михайлович, аспирант, Воронежский институт высоких технологий, Воронеж, Россия.

Тарлыков Александр Вячеславович, аспирант, Воронежский институт высоких технологий, Воронеж, Россия.

Преображенский Андрей Петрович, доктор технических наук, профессор, Воронежский институт высоких технологий, Воронеж, Россия.