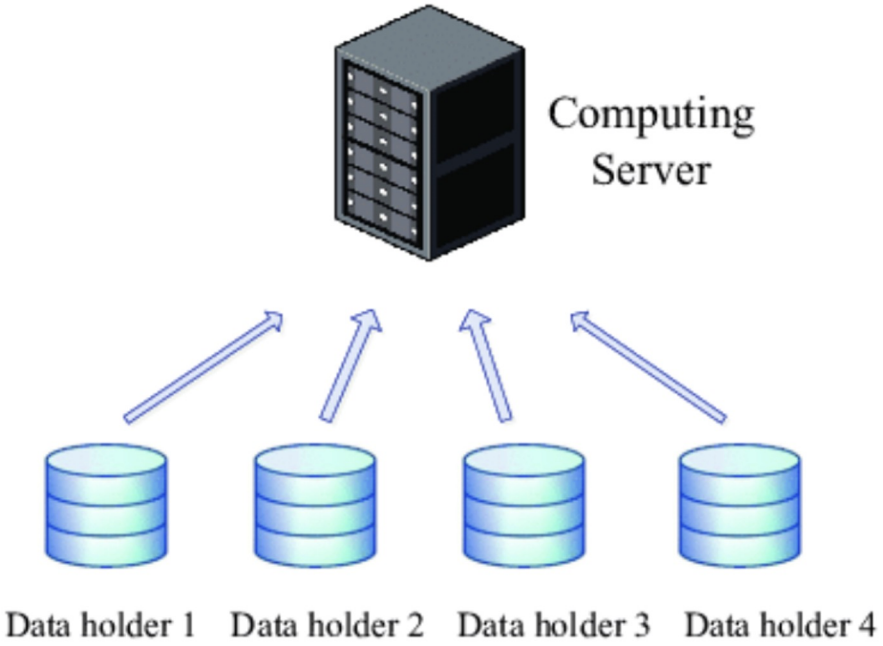
Таксономия методов федеративного обучения, обзор существующих платформ и основных игроков, существующие проблемы и тенденции

Денис Афанасьев

Проблемы централизованного ML



- **Data volume**
- **Data privacy**

Законодательные ограничения

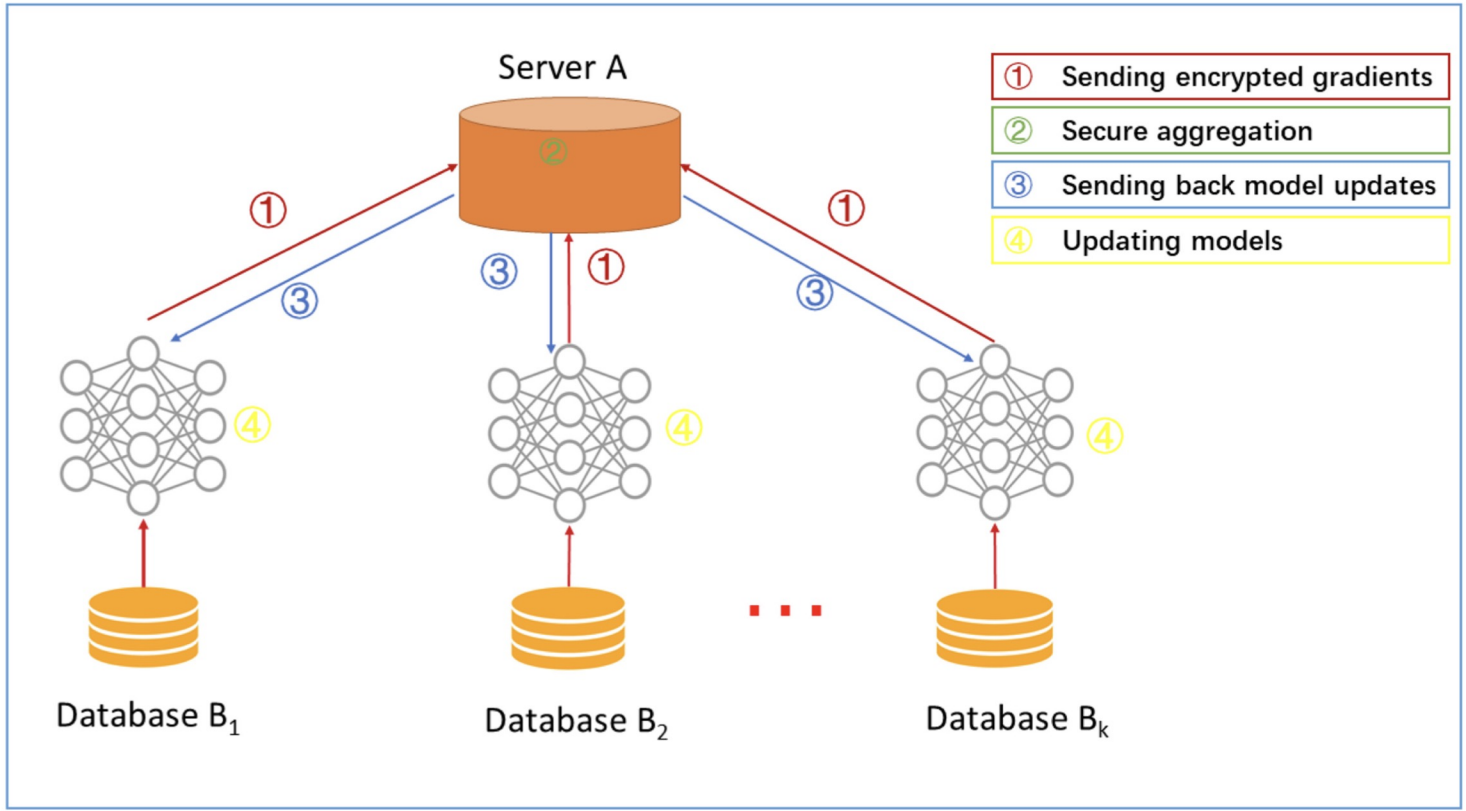
- **Цели использования:** законы о защите данных требуют, чтобы персональные данные собирались и обрабатывались для конкретной, законной цели и не использовались для каких-либо других целей без получения дополнительного согласия.

- **Право на удаление:** Законы о защите данных предоставляют физическим лицам определенные права на их личные данные, такие как право на доступ, исправление, удаление или возражение против обработки их данных.

- **Передача данных:** Законы о защите данных регулируют передачу персональных данных за пределы страны происхождения и требуют наличия соответствующих мер безопасности для защиты конфиденциальности и безопасности этих данных.

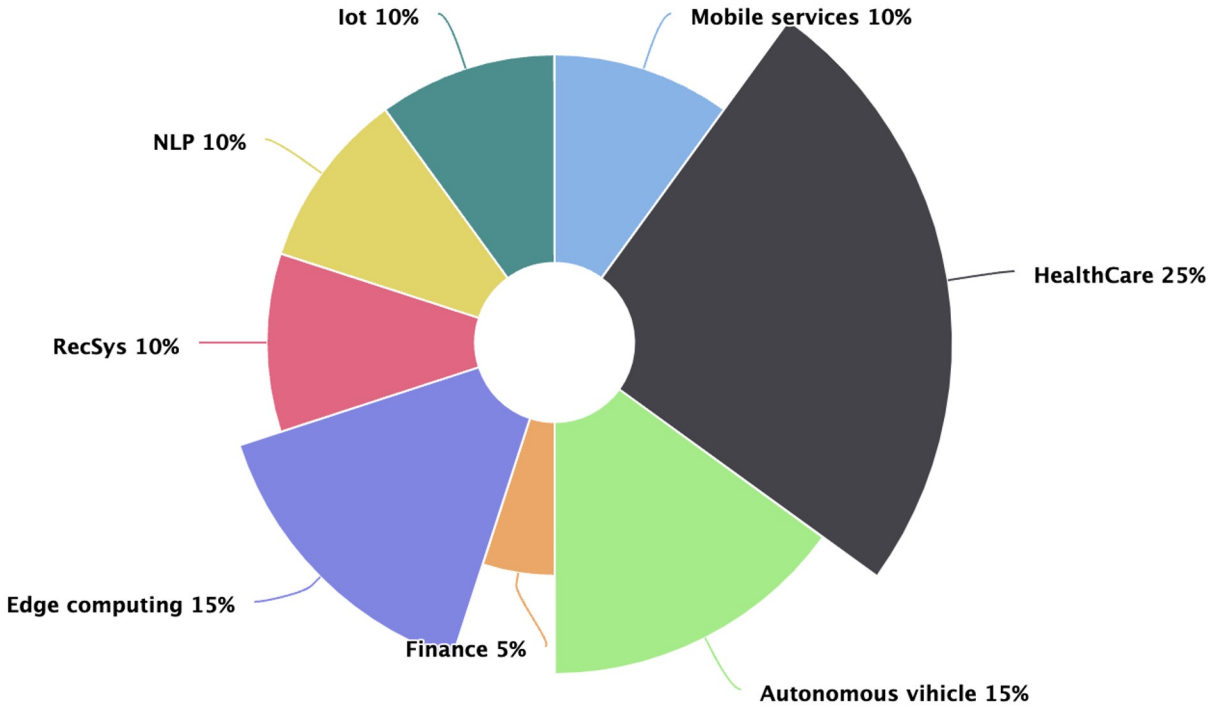
Сложно соблюдать при передаче данных третьей стороне

Federated Learning



Data minimization
Data security

Federated Learning - кейсы использования



https://www.mdpi.com/2079-9292/11/4/670

Federated Recommendation Systems (FedRec)

Data: MovieLens

Точность FedRec лучше, чем точность каждой отдельной RecSys с собственными данными.

Результаты подтверждают, что результаты моделей FCF и CF очень похожи с точки зрения показателей производительности рекомендаций набора тестов. В среднем процентная разница diff % CF и FCF по любой из пяти метрик составляет менее 0,5 %.

Анализ сходимости показал, что федеративная модель обеспечивает надежные и стабильные решения за счет адаптивной скорости обучения.

Результаты показывают, что федеративная модель может обеспечить такое же качество рекомендаций, как и широко используемый стандартный коллаборативный фильтр, при полном сохранении конфиденциальности пользователя.

	CF	FCF	diff %
	Movie-Lens		
Precision	0.3008 ± 0.0079	0.2993 ± 0.0083	0.4987
Recall	0.1342 ± 0.0044	0.134 ± 0.0046	0.149
F1	0.1552 ± 0.0047	0.1548 ± 0.0049	0.2577
MAP	0.2175 ± 0.008	0.2155 ± 0.0082	0.9195
RMSE	0.6988 ± 0.056	0.6994 ± 0.0558	0.0859
	In-House		
Precision	0.0916 ± 0.0173	0.0914 ± 0.0172	0.2183
Recall	0.1465 ± 0.0289	0.146 ± 0.0289	0.3413
F1	0.1104 ± 0.0214	0.11 ± 0.0213	0.3623
MAP	0.0669 ± 0.017	0.0664 ± 0.0171	0.7474
RMSE	0.8076 ± 0.0316	0.8083 ± 0.0323	0.0867
	Simulated		
Precision	0.2014 ± 0.0057	0.2013 ± 0.0059	0.0497
Recall	0.8867 ± 0.0196	0.8863 ± 0.0199	0.0451
F1	0.3208 ± 0.0088	0.3207 ± 0.009	0.0312
MAP	0.5805 ± 0.0341	0.5798 ± 0.0341	0.1206
RMSE	0.5387 ± 0.0193	0.5391 ± 0.0192	0.0743

https://arxiv.org/pdf/1901.09888.pdf

BlockFL

Ограничение обычного FL: он не вознаграждает участников обучения, хотя участники с большим количеством данных вносит больший вклад в глобальное обучение

- BlockFL позволяет разбивать данные на блоки, которые используются для локального обучения моделей на устройствах.
- BlockFL позволяет локально вычислять обновления модели и передавать их на центральный сервер только при необходимости, что снижает затраты на связь.
- BlockFL использует криптографические методы, такие как гомоморфное шифрование, безопасные многосторонние вычисления и дифференциальную конфиденциальность, чтобы обеспечить конфиденциальность и безопасность данных.
- Некоторые будущие направления исследований для BlockFL включают повышение эффективности обучения моделей и коммуникации, разработку новых методов сохранения конфиденциальности и изучение приложений BlockFL в реальных сценариях.

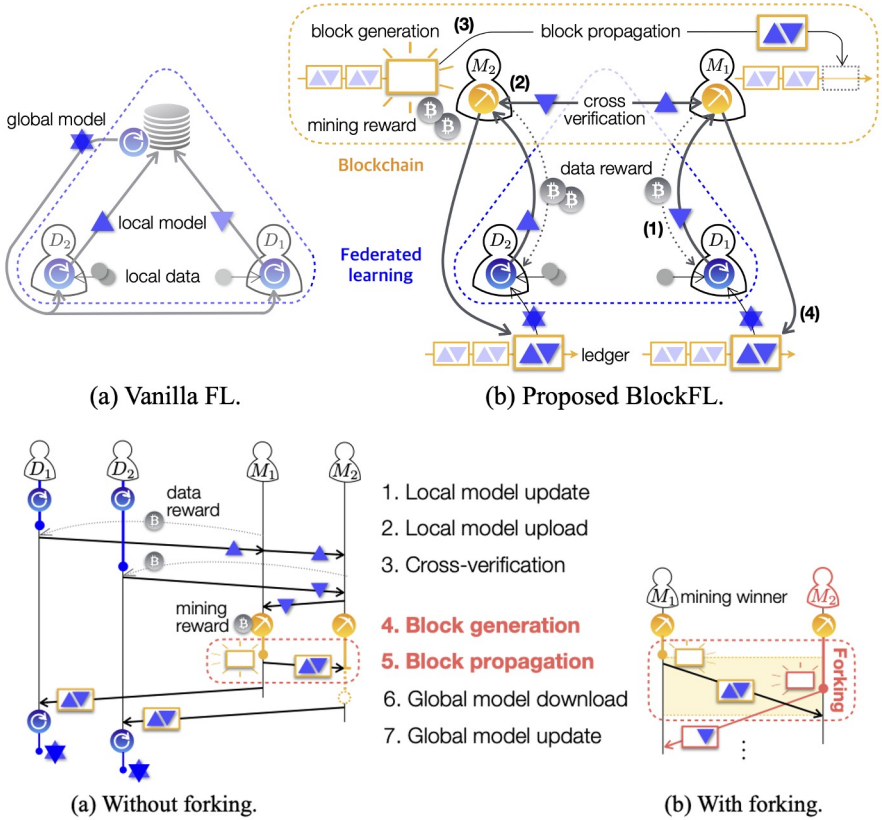


Fig. 2. The one-epoch operation of BlockFL with and without forking.

https://arxiv.org/pdf/1808.03949.pdf

Open source FL Tools and Framework

Направления будущих исследований в FL

1. **Конфиденциальность и безопасность.** Одной из основных задач, стоящих перед федеративным обучением, является обеспечение конфиденциальности и безопасности в децентрализованной среде. Будущие исследования, вероятно, будут сосредоточены на разработке более эффективных методов сохранения конфиденциальности и безопасных протоколов связи для защиты конфиденциальных данных, совместно используемых в системах федеративного обучения.

2. **Распределение ресурсов и эффективность.** Еще одной важной областью исследований в области федеративного обучения является повышение эффективности распределения ресурсов в децентрализованных системах. Будущие исследования могут быть сосредоточены на разработке более эффективных алгоритмов планирования задач и балансировки нагрузки, а также на изучении новых подходов к периферийным вычислениям и проектированию сетевой инфраструктуры.

4. **Междоменное и многозадачное обучение.** Системы федеративного обучения часто используются для обучения моделей в нескольких предметных областях или для одновременного выполнения нескольких задач. Будущие исследования могут быть сосредоточены на разработке более эффективных методов междисциплинарного и многозадачного обучения, включая методы трансфертного обучения и адаптации предметной области.

5. **Федеративное обучение в реальных приложениях:** будущие исследования в области федеративного обучения, вероятно, будут сосредоточены на применении методов федеративного обучения к реальным проблемам, таким как здравоохранение, финансы и транспорт. Это может включать разработку специализированных алгоритмов и моделей для конкретных областей, а также изучение новых приложений федеративного обучения в таких областях, как робототехника и Интернет вещей.

Спасибо!

Denis.Afanasev@gmail.com