Data Fusion

Оптимизация в Федеративном обучении

Александр Безносиков, заведующий лабораторией фундаментальных исследований МФТИ-Яндекс

Задача обучения – задача оптимизации

Задача обучения

Дана обучающая выборка $\{x_i, y_i\}_{i=1}^n$, где x_i – объекты (картинки, тексты и т.д.), y_i – ответы/лейблы (на картинке изображена машинка и т.д.).

Цель

Выбирается модель $g(w, x_i)$, которая принимает на вход объект и какие-то свои внутренние параметры w (их можно менять и настраивать). Необходимо настроить w (обучить модель) так, чтобы $g(w, x_i) \approx y_i$ для всех i.

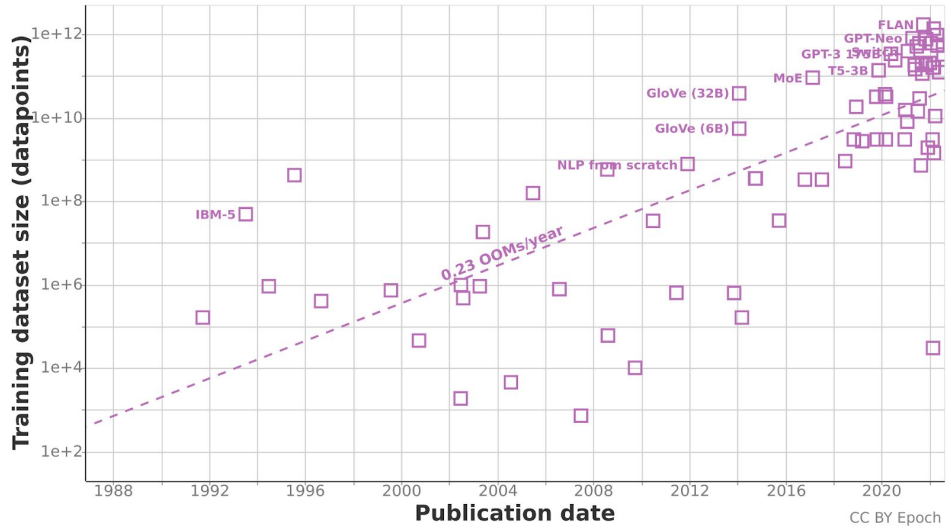
Формальная постановка

$$\min_{w \in \mathbb{R}^d} f(w) := \frac{1}{n} \sum_{i=1}^{n} l(g(w, x_i), y_i)$$

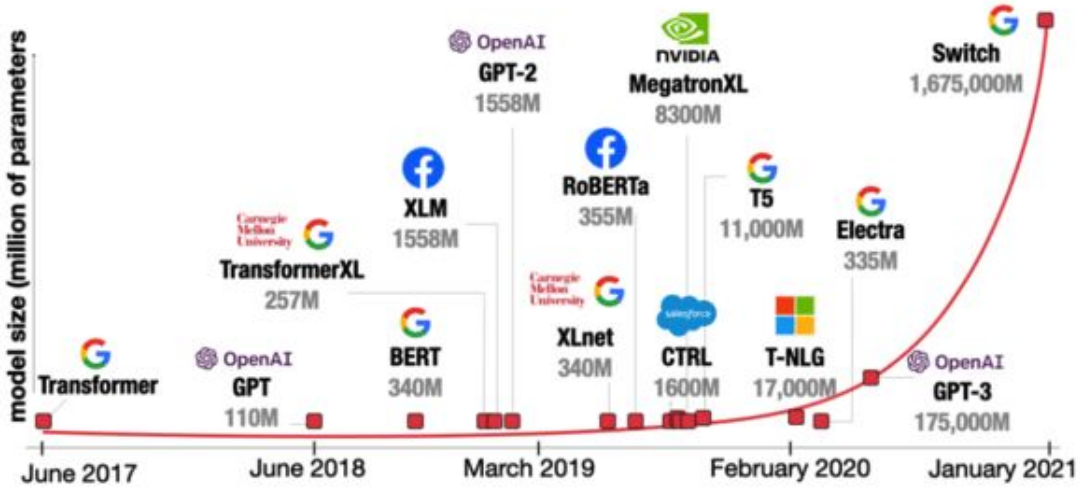
Data Fusion

Тенденции машинного обучения

Экспоненциальный рост размеров данных



Экспоненциальный рост размеров моделей



К таким задачам не имеет смысла подходить без параллелизации/распределения процесса обучения!

Data Fusion

Распределенная задача оптимизации

Делим задачу

Задача обучения делится между устройствами. Каждое устройство хранит только свой кусок данных.

Формальная постановка

$$\min_{w \in \mathbb{R}^d} f(w) := \frac{1}{M} \sum_{m=1}^{M} f_m(w) := \frac{1}{M} \sum_{m=1}^{M} \frac{1}{n_m} \sum_{i=1}^{n_m} l(g(w, x_i), y_i).$$

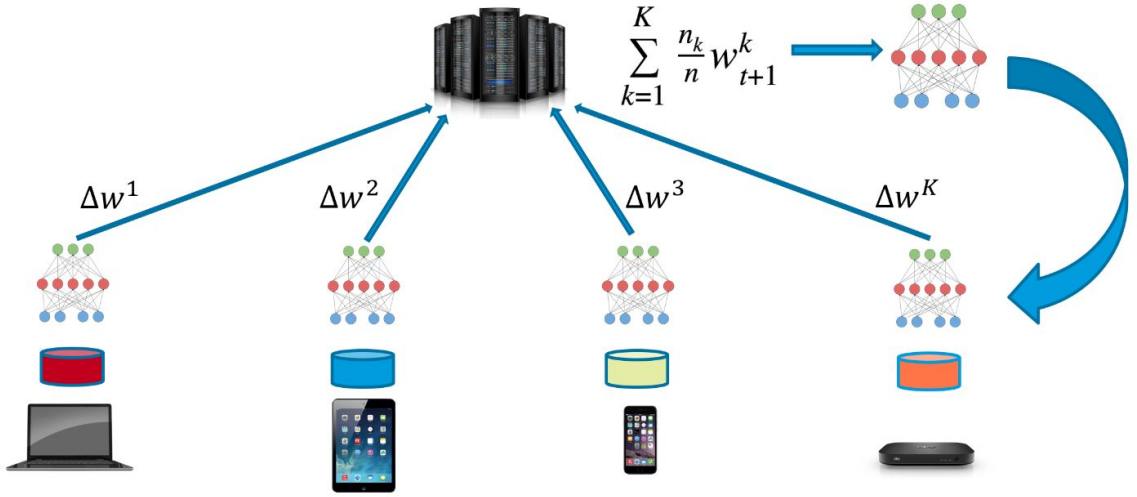
Что может помешать полному ускорению? Коммуникации в данном случае трата времени

Data Fusion

Федеративного обучение - ветвь распределенного обучения

- Классика распределенного обучения – использование вычислительных кластеров.

- Федеративное обучение – новая парадигма.



- Теперь данные хранятся на локальных устройствах пользователей, что порождает массу новых вопросов к созданию алгоритмов обучения, в первую очередь к коммуникациям.

Data Fusion

Особенности федеративного обучения

01 Данные

- локальные приватные данные пользователей

02 Мощности

- локальные вычислительные мощности пользовательских устройств

03 Коммуникации

- беспроводные сети, мобильный Интернет

Data Fusion

Какие это проблемы может повлечь за собой?

Проблема федеративного обучения

Приватность

Данные нельзя пересылать, нельзя пересылать и некоторые производные от данных.

Персонализация

Пользователю интересны только его конкретная задача.

Загрузка устройств

Нельзя перегружать устройство, когда пользователь это может заметить.

Коммуникации

Коммуникации очень медленные, а иногда и связь с отдельными пользователями может прерваться.

Data Fusion

Приватность

Аддитивный шум

$$\left(\frac{1}{M}\sum_{m=1}^{M}(\nabla f_m(x)+\xi_m)\right)+\xi$$

Сжатие

$$\text{Сжатие}\left(\frac{1}{M}\sum_{m=1}^{M}\text{Сжатие}(\nabla f_m(x))\right)$$

Шифрование

$$\text{Декодирование}\left(\sum_{m=1}^{M}\text{Кодирование}(\nabla f_m(x))\right)=\sum_{m=1}^{M}\nabla f_m(x)$$

Data Fusion

Персонализация

Регуляризация к средней модели

$$\min_{x_1,\dots,x_M \in \mathbb{R}^d} \left[\frac{1}{M} \sum_{m=1}^{M} f_m(x_m) + \frac{\lambda}{2} \|x_m - \bar{x}\|^2 \right]$$

Регуляризация к локальным моделям

$$\min_{x \in \mathbb{R}^d} \left[\frac{1}{M} \sum_{m=1}^{M} f_m(\alpha_m x_m^* + (1-\alpha)x) \right]$$

Обучение графа коммуникаций

$$\min_{x_1,\dots,x_M \in \mathbb{R}^d} \min_{W \in \mathcal{C}} \left[\frac{1}{M} \sum_{m=1}^{M} f_m(x_m) + \frac{\lambda}{2} X^T W X \right]$$

Data Fusion

Коммуникации

Сжатие информации

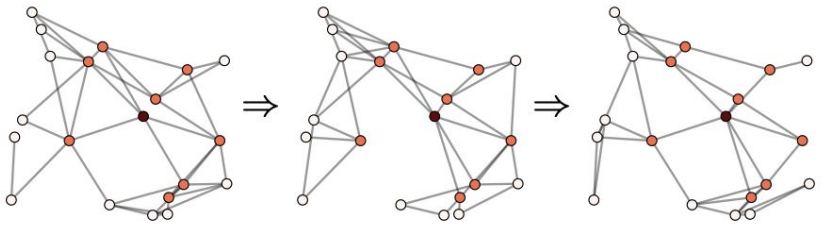
$$\text{Сжатие}\left(\frac{1}{M}\sum_{m=1}^{M}\text{Сжатие}(\nabla f_m(x))\right)$$

Локальные вычисления и похожесть

$$\|\nabla f(x) - \nabla f_m(x)\| \le \zeta$$

$$\|\nabla^2 f(x) - \nabla^2 f_m(x)\| \le \delta$$

Диффузия в децентрализованной сети



Асинхронное/частичное участие

Data Fusion