

Ensemble aware optimization in federated learning

Зайнуллин А.Р.¹, Белинский Т.Д.¹,
Феоктистов Д.Д.²

¹Московский физико-технический институт (национальный исследовательский университет)

²Факультет вычислительной математики и кибернетики (ВМК МГУ)

Ансамбль глубоких моделей это мощный подход в машинном обучении, который комбинирует в себе несколько независимо обученных моделей, значительно повышая метрики. С точки зрения оптимизации, обучение ансамбля из M моделей с параметрами x_1, \dots, x_M , можно сформулировать как следующую задачу:

$$\min_{x_1, \dots, x_M \in \mathbb{R}^d} \left\{ L_{ind}(x_1, \dots, x_M) := \frac{1}{M} \sum_{m=1}^M \sum_{j=1}^D \mathcal{L}(f(a_j, x_m), y_j) \right\}, \quad (1)$$

Где (a_j, y_j) это элемент датасета (A, y) размера D , $\mathcal{L}(\cdot, \cdot)$ - функция потерь, $f(a_j, x_m)$ предсказания модели с весами x_m для объекта a_j с меткой y_j .

Другой подход - обучать модели таким же образом, каким они будут производить инференс, т.е. усреднять предсказания до подсчета функции потерь. Этот подход известен как joint обучение

$$\min_{x_1, \dots, x_M \in \mathbb{R}^d} \left\{ L_{joint}(x_1, \dots, x_M) := \sum_{j=1}^D \mathcal{L} \left(\frac{1}{M} \sum_{m=1}^M f(a_j, x_m), y_j \right) \right\}, \quad (2)$$

Теперь подведем нашу задачу к контексту федеративного обучения [2]. Каждый агент обучается на своем приватном датасете $((A^m, y^m)$ размера D_m), а joint компонента считается на другом датасете (A^{joint}, y^{joint}) размера D , например публичный датасет или собранный из кусков, которыми скинулись агенты.

$$\min_{x_1, \dots, x_M \in \mathbb{R}^d} \left\{ \frac{1}{M} \sum_{m=1}^M \sum_{j=1}^{D_m} \mathcal{L}(f(a_j^m, x_m), y_j^m) + \lambda \sum_{j=1}^D \mathcal{L} \left(\frac{1}{M} \sum_{m=1}^M f(a_j^{joint}, x_m), y_j^{joint} \right) \right\}, \quad (3)$$

Параметр λ контролирует трейд-офф между независимым обучением и коммуникациями.

Вычислительная эффективность

Многие FL алгоритмы обучения имеют стоимость коммуникации пропорциональную весу модели. Например FedAvg [3] и SCAFFOLD [1]. В нашем проекте мы предлагаем новый подход - Federated Ensemble (FEEN). Рассмотрим сэмпл $a \in A^{joint}$, который есть у каждого агента локально. Агент считает предсказание, отправляет его на сервер. Сервер усредняет и считает функцию потерь. Дальше модель делает шаг обучения. То есть стоимость коммуникации в нашем случае пропорциональна размеру shared датасета.

Для проведения экспериментов мы выбрали модель ResNet-18, которую обучали на CIFAR10. Сначала мы проводили эксперименты на несложном сетапе. Мы оставили два класса (бинарная классификация). Подтвердив нашу гипотезу, мы перешли на сетап с

10 классами. На докладе мы представим результаты исследования алгоритма FEEN на экстремально гетерогенных данных, т.к. гетерогенность естественна для федеративного обучения [4] и представим графики, где FEEN обходит бейзлайны.

References

- [1] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International conference on machine learning*, pages 5132–5143. PMLR, 2020.
- [2] Jakub Konečný, H Brendan McMahan, Daniel Ramage, and Peter Richtárik. Federated optimization: Distributed machine learning for on-device intelligence. *arXiv preprint arXiv:1610.02527*, 2016.
- [3] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.
- [4] Mang Ye, Xiuwen Fang, Bo Du, Pong C Yuen, and Dacheng Tao. Heterogeneous federated learning: State-of-the-art and research challenges. *ACM Computing Surveys*, 56(3):1–44, 2023.