

**ПРИМЕНЕНИЕ МЕТОДА АДАПТИВНОЙ
СТОХАСТИЧЕСКИ УСТОЙЧИВОЙ ОПТИМИЗАЦИИ В
ЗАДАЧАХ С ГЕТЕРОГЕННЫМИ ДАННЫМИ**

Боровко Никита Алексеевич

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: borovkona@my.msu.ru

Научный руководитель — Горшенин Андрей Константинович

Задачи обучения глубоких нейронных сетей чаще всего решаются минимизацией среднего эмпирического риска по обучающей выборке [1]. При гетерогенных данных и дисбалансе такой подход нередко приводит к тому, что модель показывает приемлемое среднее качество, но заметно хуже работает на редких классах или отдельных клиентах. Это требует методов, которые поддерживают устойчивое качество по группам (классам, доменам или клиентам), а не только в среднем. В частности, это важно в федеративных постановках обучения [2].

В работе экспериментально исследуется применимость робастного стохастического оптимизатора *Adaptive Loss Scaling Optimizer* [3] для обучения в условиях дисбаланса. Целью является оценка того, насколько ALSO повышает качество на наихудших группах при сохранении вычислительной эффективности и устойчивости процесса оптимизации по сравнению с распространёнными базовыми оптимизаторами, прежде всего AdamW [4].

Первая серия экспериментов выполнена в постановке федеративного обучения на паре разнородных датасетов (Food101 и Flowers102), рассматриваемых в качестве двух клиентов с различающимися распределениями данных. Обучение инициировалось на первом клиенте, после чего второй клиент подключался с задержкой, а дальнейшая оптимизация осуществлялась попеременно на обоих клиентах. Для корректного и сопоставимого сравнения оптимизаторов использовалась автоматическая настройка гиперпараметров при одинаковом числе итераций. Итоговые результаты этого сравнения, выраженные как наилучшая средняя точность по двум датасетам за весь прогон, приведены в табл. 1; ALSO занимает первое место с заметным отрывом, что является ключевым эмпирическим выводом для федеративной постановки.

Во второй серии экспериментов рассматривалась безусловная генерация изображений диффузионной моделью на MNIST, где устой-

чивость обучения существенна для корректного воспроизведения редких классов. Сравнение с AdamW при согласованных настройках показало, что ALSO обеспечивает более быструю стабилизацию процесса генерации и снижает долю некачественных образцов на промежуточных этапах обучения.

Эксперименты показали, что ALSO в сценариях с гетерогенностью данных демонстрирует устойчивое преимущество по качеству как на «слабых» частях данных, так и в среднем. Полученные результаты указывают на практическую целесообразность применения подхода в задачах, где требуется контролировать не только средние метрики, но и надёжность модели на отдельных клиентах и редких режимах данных.

Место	Оптимизатор	Лучшая средняя точность
1	ALSO	0.6393
2	RECOVER	0.6151
3	AdamW	0.6067
4	Spectral Risk	0.6066
5	FastDRO	0.6038
6	SGD with Momentum	0.4390
7	DRAGO	0.3211

Таблица 1: Наилучшая точность предсказания, усредненная по 2 датасетам, среди всех итераций настройки гиперпараметров.

Литература

1. Mohri M., Sivek G., Suresh A. T. Agnostic Federated Learning // In Proceedings of the 36th International Conference on Machine Learning (ICML 2019), Long Beach, USA, 2019, P. 4615–4625.
2. Feoktistov D., Ignashin I., Veprikov A., Borovko N., Bogdanov A., Chezhegov S., Beznosikov A. Aligning Distributionally Robust Optimization with Practical Deep Learning Needs // arXiv preprint arXiv:2508.16734, 2025. <https://arxiv.org/pdf/2508.16734>
3. Loshchilov I., Hutter F. Decoupled Weight Decay Regularization (AdamW) // arXiv preprint arXiv:1711.05101, 2017. <https://arxiv.org/abs/1711.05101>
4. Ho J., Jain A., Abbeel P. Denoising Diffusion Probabilistic Models // arXiv preprint arXiv:2006.11239, 2020. <https://arxiv.org/abs/2006.11239>