

Simplex LoRA: Expand important adapters

Давыденко Григорий, Шалыгин Игорь

Научный руководитель: Безносиков А. Н.

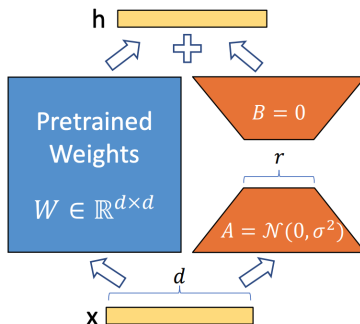
Научный консультант: Веприков А. С.

Московский физико-технический институт

18 марта 2025 г.

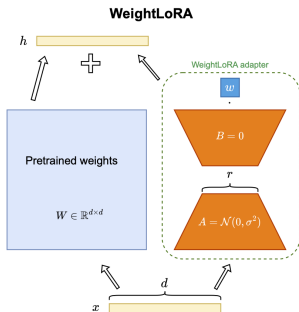
- 1 LoRA: Low-Rank Adaptation
- 2 Предыстория: WeightLoRA и WeightLoRA+
- 3 SimplexLoRA
- 4 Структура репозитория
- 5 Научный и технический продукт
- 6 Модели, датасеты и бенчмаркинг

LoRA: Low-Rank Adaptation



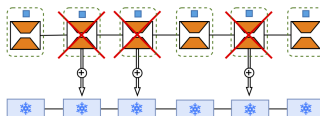
- Тюнинг линейных слоев больших моделей через низкоранговые матрицы $A \in \mathbb{R}^{d \times r}$ и $B \in \mathbb{R}^{r \times d}$
- $r \ll d$ (например, $r = 8$ для моделей с $d = 1024$)

Предыстория: WeightLoRA и WeightLoRA+



- finetuned LoRA adapters - pretrained weights

WeightLoRA: keep only necessary adapters



1. Learn WeightLoRA adapters on all layers.
2. Select most important layers.
3. Continue finetuning with classic LoRA on selected layers.

- Основная идея WeightLoRA: избавиться с помощью параметров важности ω от лишних LoRA
- Основная идея WeightLoRA+: увеличить ранги оставшихся LoRA

Теперь ω - это веса на симплексе ($\sum_{i=1}^n \omega_i = 1$).

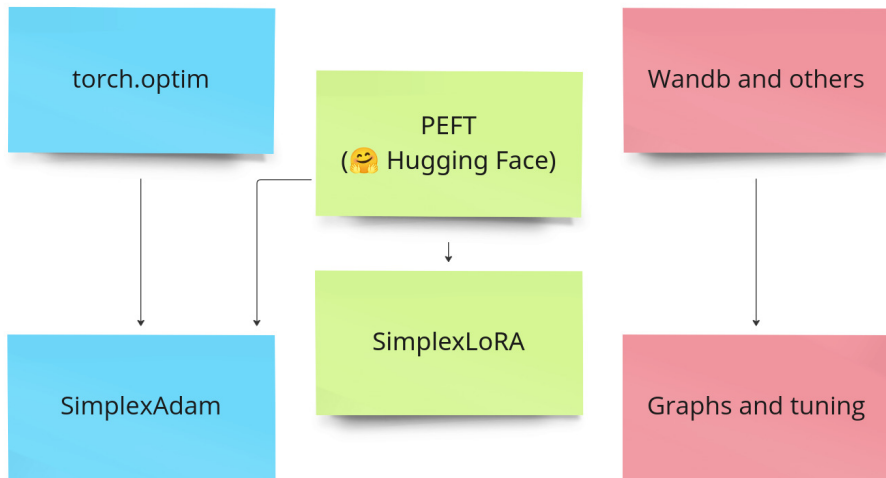
За время обучения мы совершаем раз в несколько шагов изменения рангов LoRA в соответствии с их обучаемыми весами:

- 1 Сначала все ранги равны гиперпараметру $r_i^{(0)} = r^0$, веса - $\omega_i^{(0)} = \frac{1}{n}$
- 2 После шага в оптимизаторе $\omega^{(k)} = \text{proj}_{\text{simplex}}(\omega^{(k)})$
- 3 Через фиксированное количество шагов x меняем ранги LoRA согласно новым весам:

$$r_i^{(u)} = \text{ceil}(r^0 \cdot n \cdot \omega_i^{(u \cdot x)})$$

- 4 Пункты 2 и 3 повторяем несколько раз.
- 5 Запускаем стандартное обучение LoRA с подобранными рангами.

Структура репозитория



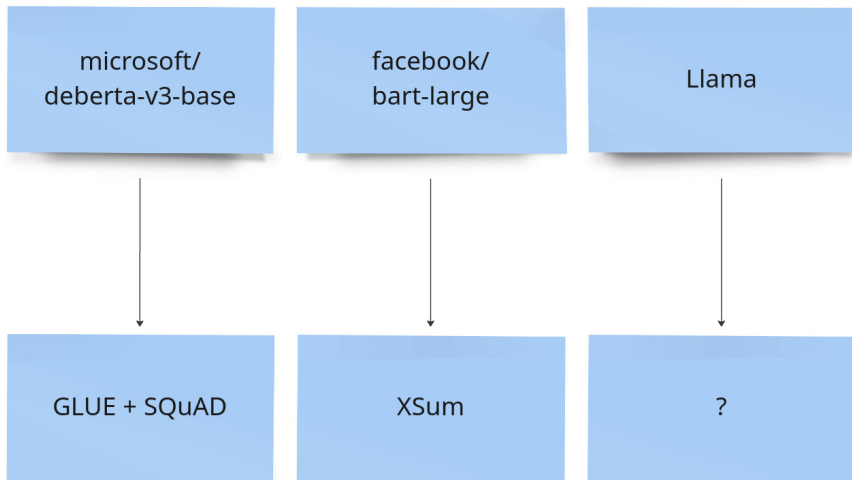
Научный продукт

- Эффективный метод выбора наиболее важных лор
- Балансировка между эффективностью алгоритма и затратами по памяти и FLOP

Технический продукт

- Реализация SimplexLoRA в библиотеке PEFT
- Оптимизатор SimplexAdam в библиотеке PEFT

Модели, датасеты и бенчмаркинг



- LoRA: Low-Rank Adaptation of Large Language Models
- DeBERTa model
- General Language Understanding Evaluation (GLUE) benchmark
- Stanford Question Answering Dataset (SQuAD)
- BART model
- The Extreme Summarization (XSum) dataset