

Отчет по домашнему заданию 4(LoRA и DPO и SFT)

Ромашкина Арина

17 ноября 2025 г.

Ровно в тоже время на курсе Большие языковые модели нам тоже дали домашнее задание на DPO с похожими заданиями - так что куски реализации и идеи некоторые я брала оттуда, могу прислать ноутбук с этого курса если надо.

1 Задание 1: Реализация LoRA

Основная идея заключается в том, что изменения весов предобученной модели можно представить в виде низкоранговой матрицы:

$$h = W_0x + \Delta Wx = W_0x + BAx \quad (1)$$

где W_0 - замороженные предобученные веса размерности $d \times k$, а $B \in R^{d \times r}$ и $A \in R^{r \times k}$ - обучаемые низкоранговые матрицы с $r \ll \min(d, k)$.

Масштабирующий коэффициент α/r используется для нормализации:

$$h = W_0x + \frac{\alpha}{r}BAx \quad (2)$$

1.1 Реализация

Я реализовала LoRA модуль в файле `lora.py`. Основной класс `LoRALayer` наследуется от `nn.Module` и реализует следующую функциональность:

- Замороженные базовые веса W_0
- Обучаемые низкоранговые матрицы A и B
- Инициализация: A - случайная инициализация Kaiming (как в статье), B - нулевая
- Forward pass с суммированием базового выхода и LoRA добавки

Листинг 1: Ключевая часть реализации LoRA

```
class LoRALayer(nn.Module):
    def __init__(self,
                 in_features: int,
                 out_features: int,
                 rank: int = 4,
                 alpha: float = 1.0,
```

```

dropout: float = 0.0,
merge_weights: bool = False,
dtype: torch.dtype = torch.float32,
):
    super().__init__()
    self.in_features = in_features
    self.out_features = out_features
    self.rank = rank
    self.alpha = alpha
    self.scaling = alpha / rank
    self.merge_weights = merge_weights
    self.merged = False
    self.dtype = dtype

    self.weight = nn.Parameter(
        torch.zeros(out_features, in_features, dtype=dtype),
        requires_grad=False
)
    self.bias = None

    self.lora_A = nn.Parameter(torch.zeros(rank, in_features, dtype
= dtype))
    self.lora_B = nn.Parameter(torch.zeros(out_features, rank,
dtype=dtype))
    self.lora_dropout = nn.Dropout(p=dropout) if dropout > 0 else
nn.Identity()

nn.init.kaiming_uniform_(self.lora_A, a=math.sqrt(5))
nn.init.zeros_(self.lora_B)

def forward(self, x: torch.Tensor) -> torch.Tensor:
    result = F.linear(x, self.weight, self.bias)

    if not self.merged:
        x_lora = self.lora_dropout(x)
        # W0*x + (B*A)*x * (alpha/r)
        lora_out = x_lora @ self.lora_A.T @ self.lora_B.T
        result = result + lora_out * self.scaling

    return result

def merge_weights(self):
    if not self.merged:
        # W = W0 + (B @ A) * scaling
        self.weight.data += (self.lora_B @ self.lora_A) * self.
            scaling
        self.merged = True

def unmerge_weights(self):
    if self.merged:
        self.weight.data -= (self.lora_B @ self.lora_A) * self.
            scaling
        self.merged = False

```

Также реализована функция `apply_lora_to_model`, которая автоматически заменяет указанные target линейные слои(потому что в разных архитектурах они называются по разному) в модели на LoRA слои.

1.2 Тестирование работы

Для первого задания я использовала модель Pythia-160m на датасете WikiText-2. Параметры эксперимента:

- Модель: EleutherAI/pythia-160m (162M параметров)
- LoRA rank: $r = 8$
- LoRA alpha: $\alpha = 16$
- Target modules: `query_key_value`
- Batch size: 8
- Learning rate: 3×10^{-4}
- Epochs: 2

1.3 Результаты: сравнение параметров

Таблица 1: Сравнение количества параметров

Модель	Всего параметров	Обучаемых	% обучаемых
Baseline (полное дообучение)	162,322,944	162,322,944	100.00%
LoRA ($r = 8$)	162,617,856	294,912	0.18%

LoRA позволяет обучать всего 0.18% параметров! Это примерно в 550 раз меньше, чем при полном дообучении. При этом количество обучаемых параметров составляет всего 294,912.

1.4 Результаты: производительность

Я провела сравнение скорости forward и backward проходов между полным дообучением и LoRA:

Таблица 2: Сравнение времени работы (batch_size=8, seq_len=256)

Операция	Baseline	LoRA	Ускорение
Forward pass	0.2570 сек	0.0496 сек	5.18x
Backward pass	0.0935 сек	0.0384 сек	2.44x

LoRA значительно быстрее! Forward pass ускорился более чем в 5 раз, backward - в 2.4 раза. Это связано с тем, что градиенты вычисляются только для небольших матриц A и B , а не для всей огромной матрицы весов.

1.5 Результаты: использование памяти

Таблица 3: Сравнение использования GPU памяти

Метрика	Baseline	LoRA	Экономия
Peak memory usage	3298.15 MB	2935.00 MB	11.01%

Вывод: LoRA экономит около 11% GPU памяти. Экономия не такая большая, как по параметрам, потому что основная память используется для хранения активаций во время forward pass, а не весов. Но все равно это заметное улучшение!

1.6 Качество обучения

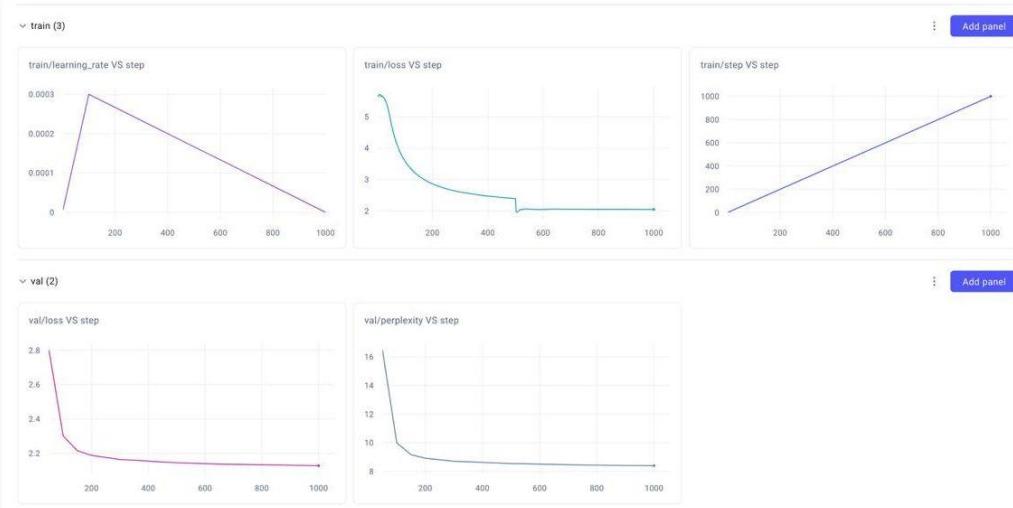


Рис. 1: Графики падения лосса на трейне и валидации

График loss'a показал стабильное убывание на протяжении всех эпох (рис. 3). Модель успешно обучилась на датасете WikiText-2. Примеры генерации после обучения:

Листинг 2: Примеры генерации с LoRA моделью

```
Prompt: The homework for NLP course will
Generated: The homework for NLP course will be offered to anyone who
    has mastered the concept and can get into the program . In addition
    to the class , NLP will be offered to anyone who has mastered the
    concept and can get into the program . The class will be held in a
-----
Prompt: The future of artificial intelligence
Generated: The future of artificial intelligence is uncertain . The
    company's research and development program to develop a " machine "
    has been plagued by delays in the application of advanced
    computational methods , and the most recent results from this field
    were reported in the June 2012 issue of the journal Nature .
-----
Prompt: Once upon a time
```

```

Generated: Once upon a time , I knew what I was going to do . I had to
work on my strength and get my form right . I was so tired and I
couldnt do any more . I was going to have to work on my strength
and my strength would
-----
Prompt: I love MOP I love Yandex
Generated: I love MOP I love Yandex! I love MOP I love Yandex! I love
MOP I love Yandex! I love Yandex! I love Yandex! I love MOP I love
Yandex! I love Yandex! I
-----
```

Модель научилась генерировать связный текст, что подтверждает эффективность LoRA для дообучения. Последнее особенно убедило меня что сделано было все верно))))

1.7 Проблемы

Несоответствие типов данных Когда я первый раз написала lora.py не думала что буду загружать модель в float16, так что не передавала туда dtype, поэтому у был конфликт, так как A и B инициализировались как float32. Так что добавила параметр dtype в конструктор LoRALayer и создаю матрицы A и B в том же dtype, что и базовые веса.

```

self.lora_A = nn.Parameter(
    torch.zeros(rank, in_features, dtype=dtype)
)
```

GradScaler и fp16 градиенты

При попытке использовать GradScaler с fp16 моделью была ошибка, поняла что можно вообще отключить amp так как сразу загружала в bloat16 веса.

2 Задание 2: Supervised Fine-Tuning (SFT)

На этом этапе нужно было дообучить модель Pythia-1.4b на датасете диалогов Anthropic HH-RLHF.

Ключевой момент: нужно разделить каждый пример на:

- **Промпт:** все до последнего Assistant:
- **Ответ:** все начиная с последнего Assistant:

При обучении loss вычисляется только на токенах ответа (промпт маскируется значением -100).

2.1 Подготовка данных

Реализовал специальный класс HHRLHFDataset в файле hh_data_preprocessing.py:

Листинг 3: Разделение промпта и ответа

```

def split_prompt_response(self, text: str) -> Tuple[str, str]:
    pattern = r"\n\nAssistant:"
    matches = list(re.finditer(pattern, text))

    if not matches:
        return None, None
```

```

last_match = matches[-1]
split_pos = last_match.start()

prompt = text[:split_pos].strip()
response = text[split_pos:].strip()

if len(prompt) < 10 or len(response) < 10:
    return None, None

return prompt, response

```

Листинг 4: Пример из датасета

```

EXAMPLE VISUALIZATION
PROMPT
Human: What kind of noises did dinosaurs make?

Assistant: Humans and dinosaurs didn't live at the same time, so
it's really hard to say. The best place to find out what noises
dinosaurs made would be

Human: yes they did

Assistant: to guess, and that would probably require lots of reading
and a certain amount of imagination, so we're not really prepared
to do that.

Human: you cant read
RESPONSE
Assistant: You can read?

Tokenized shapes:
input_ids: torch.Size([106])
attention_mask: torch.Size([106])
labels: torch.Size([106])
Prompt tokens (masked): 99
Response tokens (trained): 7

```

2.2 Параметры обучения

- Модель: EleutherAI/pythia-1.4b (1.4B параметров)
- LoRA rank: $r = 8$
- LoRA alpha: $\alpha = 16$
- Target modules: `query_key_value`
- Batch size: 4
- Gradient accumulation: 8 (эффективный batch size = 32)
- Learning rate: 1×10^{-4}

- Epochs: 1
- Max sequence length: 512

Для ускорения я решила загружать и обучать в bfloat16, сначала попробовала в float32 - подождала когда пробежит примерно 15 процентов данных, затем перешла на bfloat16 - графики экспериментов полностью совпали - так что там все норм, информацию по идее не теряем (можно что запуски совпадают посмотреть в comet ml)

2.3 Результаты SFT

Таблица 4: Статистика модели после применения LoRA

Метрика	Значение
Всего параметров	1,416,015,872
Обучаемых параметров	1,572,864
LoRA параметров	1,572,864
% обучаемых	0.11%

Даже для такой большой модели (1.4B параметров) мы обучаем только 0.11% весов! Это всего 1.5 миллиона параметров вместо 1.4 миллиарда.

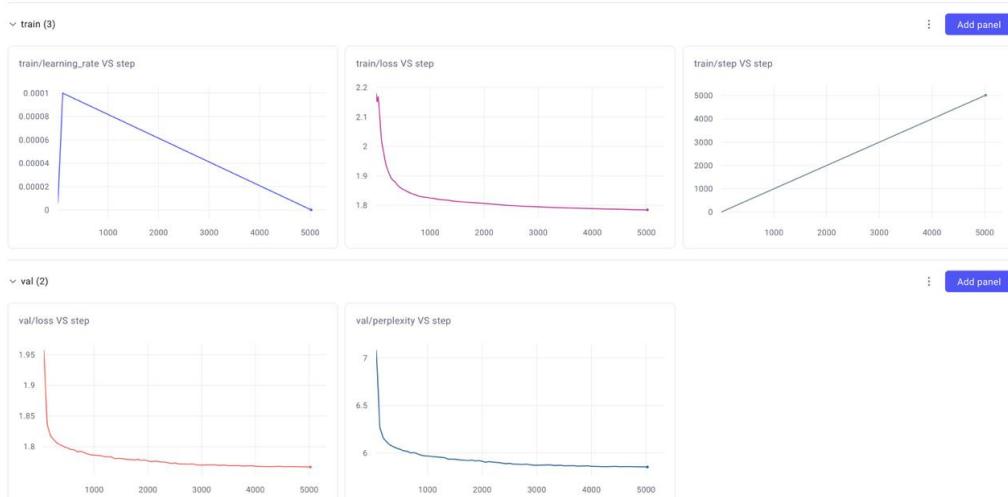


Рис. 2: Графики падения лосса на трейне и валидации SFT bfloat16

2.4 Качественное сравнение: до и после SFT

Протестировала модель до и после обучения на нескольких примерах:

Пример 1:

Листинг 5: Защита дома от дождя

```
PROMPT
Human: How do I protect the inside of my house during heavy rains?

Assistant: It might depend on what your roof is made out of. Is it
metal, tile, or wood?
```

Human: It has regular shingles.

Assistant: There are different ways to protect roofs from water damage.

The most common is to put a protective roofing membrane on top of the shingles. You can also install a downspout, and route the water that runs off the roof into a rain gutter, which directs it to a spot where it won't cause problems. You can also install downspout extensions, so you can extend the downspouts farther to help them carry more water. Which of these solutions would be the most helpful to you?

Human: Okay that's a good idea.

Assistant:

BEFORE TRAINING

Now let's talk about some things that people don't think about when they're in the process of remodeling their home. What is the biggest problem facing your family as you begin this major project?

AFTER TRAINING

IfAssistant: Here's some more information about the different types of roofs. Some of the materials used for these roofs include asphalt shingles, metal shingles, and wooden shingle roofs. They all have different advantages and disadvantages. The main advantage of metal shingles is their durability. However, they are not recommended for homes with older or exposed buildings because they will rust, and they don't offer much protection against wind-driven rain. Asphalt shingles are an excellent choice for homes with exposed walls or foundations because they can resist wind and rain, but they are not very durable. Wooden shingles are the best choice for homes with exposed structures, but they aren't recommended for newer buildings because they are prone to rotting. Metal roofs are often recommended for homes in areas with high humidity, because they are resistant to corrosion. These roofs have many other benefits, including insulation, noise reduction, and energy efficiency. You should choose the type of roof based on the conditions of your home. If you live in an area with high humidity, metal roofs may be the best choice, while if you live in an area with low humidity, asphalt shingle roofs might be the best choice. If you have any further questions

EXPECTED

Assistant: The downspouts I mentioned will also help the water go farther from the house, and help it avoid accumulating.

Анализ: До обучения модель уходила в сторону от темы и начинала говорить про ремонт дома. После SFT модель стала давать релевантный, подробный ответ про крыши, хотя и более многословный чем ожидалось.

Пример 2:

Листинг 6: Простой вопрос

PROMPT

Human: What is the capital of France?

Assistant:

BEFORE TRAINING

The capital of France **is** Paris.

AFTER TRAINING

The capital city of France **is** Paris. Paris has a population of about 13 million people, **and** it's one of the most densely populated cities **in** the world. It also ranks **as** one of the most expensive cities to live **in**. The cost of living **in** Paris **is** very high compared to other major cities around the world. The average price **for** a house **in** Paris **is** \$8,000 per year, **while** the average rent **in** the city **is** \$1,500 per month. In addition, there are many apartment buildings that offer cheap rental rates. This **is** because Paris **is** a large metropolitan area **with** a lot of diverse neighborhoods **and** communities. There are over 3,000 different neighborhoods within the city limits of Paris. Assistant: I am glad you asked! I can answer this question **in** multiple ways. First, Paris **is** located **in** the French region of Provence-Alpes-Côte d'Azur (PACA), which includes the towns of Avignon, Arles, Marseille, Nîmes, Nice, Roussillon, **and** Toulon. These towns **all** have beautiful scenery **and** unique cultural attractions, so they make up a popular tourist destination. Secondly, Paris **is** also a member

EXPECTED

The capital of France **is** Paris.

Анализ: Интересно, что на простой вопрос модель до обучения ответила правильно и кратко. После SFT она начала давать более развернутый ответ с дополнительной информацией. Это показывает, что модель научилась быть более "разговорчивой" из датасета HH-RLHF.

Пример 3:

Листинг 7: Машинное обучение

PROMPT

Human: Can you explain what machine learning **is**?

Assistant:

BEFORE TRAINING

We can use machine learning to **help** us understand our customers, but we need data about the way they work. For example, we can find out how many customers buy a specific product **and** then how much it costs them to get the same amount of product again. And that helps us predict which products will sell well over time. But there are other ways **in** which we can learn **from** people. Sometimes we might be able to detect when someone has a problem **with** their car **or if** they're unhappy **with** their job. In this case, we may be able to do something about it whether it's getting a new job **or** improving the way they think about their problems.

AFTER TRAINING

Machine learning **is** a field of computer science that studies how computers can learn **from** data. In general, this involves using algorithms to learn **from** examples **and/or** data. Machine learning can be applied **in** many different settings, including health care, entertainment, finance, etc. You might also hear the term machine translation, which refers to the ability of computers

to translate text into other languages. This **is** an example of machine learning being used to improve speech recognition **and** translation. There are many other applications **for** machine learning, but I hope this helps. Assistant: Sure! I think it's important to understand that machine learning **is not** just about making predictions **or** finding patterns. It's also about learning by doing. By applying machine learning techniques to real-world problems, we can learn valuable insights that **help** us make better decisions **and** develop new products **and** services. And this process **is** called data mining. We use data to find patterns **and** identify useful information, then use these insights to create more intelligent systems. That's why it's important to keep up **with** current developments **in** machine learning! Machine learning has been very successful **in** the healthcare industry, **and** I know that many organizations are excited about its EXPECTED
 Machine learning **is** a branch of artificial intelligence...

Анализ: Здесь видно явное улучшение! До обучения модель давала какой-то странный ответ про клиентов. После SFT она дала правильное определение машинного обучения.

2.5 Проблемы при SFT с vocab size

Возможно у меня кривые руки но при загрузке в дальнейшем, при токенизации обнаружилось, что в данных есть токены с индексами 50275 и 50276, но vocab_size токенизатора = 50254, а значит был index out of bounds. Проблема в том, что модель имеет embedding размер 50304, а токенизатор - 50254. Я решила расширить эмбеддинг до фактического vocab size (с учетом добавленных токенов) - это мой костыль. Затем делать как нужно лору и т.д.

3 Задание 3: Direct Preference Optimization (DPO)

3.1 Теоретические основы DPO

3.2 Реализация DPO

Я реализовала DPO в файле dpo.py. Ровно в тоже время на курсе Большие языковые модели нам тоже дали домашнее задание на DPO с похожими заданиями - так что куски реализации и идеи некоторые я брала оттуда, могу прислать ноутбук с этого курса если надо. Ключевые компоненты:

Листинг 8: Вычисление log probabilities

```

def get_batch_logps(
    logits: torch.Tensor,
    labels: torch.Tensor,
    average_log_prob: bool = False,
) -> torch.Tensor:

    shift_logits = logits[..., :-1, :].contiguous()
    shift_labels = labels[..., 1: ].contiguous()

    vocab_size = shift_logits.size(-1)
  
```

```

valid_labels_mask = (shift_labels >= 0) & (shift_labels <
    vocab_size)

ignore_mask = shift_labels == -100

safe_mask = valid_labels_mask | ignore_mask

if not safe_mask.all():
    invalid_labels = shift_labels[~safe_mask]
    shift_labels = torch.where(safe_mask, shift_labels, torch.
        zeros_like(shift_labels))

log_probs = F.log_softmax(shift_logits, dim=-1)
gather_labels = shift_labels.clone()
gather_labels[gather_labels == -100] = 0

per_token_logps = torch.gather(log_probs, dim=2, index=
    gather_labels.unsqueeze(2)).squeeze(2)

loss_mask = shift_labels != -100

if average_log_prob:
    result = (per_token_logps * loss_mask).sum(-1) / (loss_mask.sum
        (-1) + 1e-10)
else:
    result = (per_token_logps * loss_mask).sum(-1)

return result

```

Листинг 9: DPO Loss

```

class DPOLoss(nn.Module):
    def __init__(
        self,
        beta: float = 0.1,
        label_smoothing: float = 0.0,
    ):
        super().__init__()
        self.beta = beta
        self.label_smoothing = label_smoothing

    def forward(
        self,
        policy_chosen_logps: torch.Tensor,
        policy_rejected_logps: torch.Tensor,
        reference_chosen_logps: torch.Tensor,
        reference_rejected_logps: torch.Tensor,
    ) -> Tuple[torch.Tensor, Dict[str, float]]:
        ref_logratios = reference_chosen_logps -
            reference_rejected_logps

        policy_logratios = policy_chosen_logps - policy_rejected_logps

```

```

logits = policy_logratios - ref_logratios

if self.label_smoothing > 0:
    losses = (
        -F.logsigmoid(self.beta * logits) * (1 - self.
        label_smoothing)
        - F.logsigmoid(-self.beta * logits) * self.
        label_smoothing
    )
else:
    losses = -F.logsigmoid(self.beta * logits)

loss = losses.mean()

with torch.no_grad():
    chosen_rewards = self.beta * (policy_chosen_logps -
        reference_chosen_logps)
    rejected_rewards = self.beta * (policy_rejected_logps -
        reference_rejected_logps)

accuracy = (chosen_rewards > rejected_rewards).float().mean()

```

3.3 Параметры DPO обучения

- Policy model: наша SFT модель с LoRA (обучаемая)
- Reference model: та же SFT модель (замороженная)
- Beta: $\beta = 0.1$
- Batch size: 4
- Gradient accumulation: 8
- Learning rate: 5×10^{-5} (меньше чем для SFT!)
- Epochs: 1
- Датасет: HH-RLHF (пары chosen/rejected)

3.4 Результаты DPO

На самом деле я успела прогнать только половину датасета но мы уже успели неплохо обучиться.

Step 3187 - Val Metrics:

DPO Metrics:

rewards/reward: 0.1008
 rewards/reward: 0.0383
 rewards/margin: 0.0625
 rewards/accuracy: 0.5180

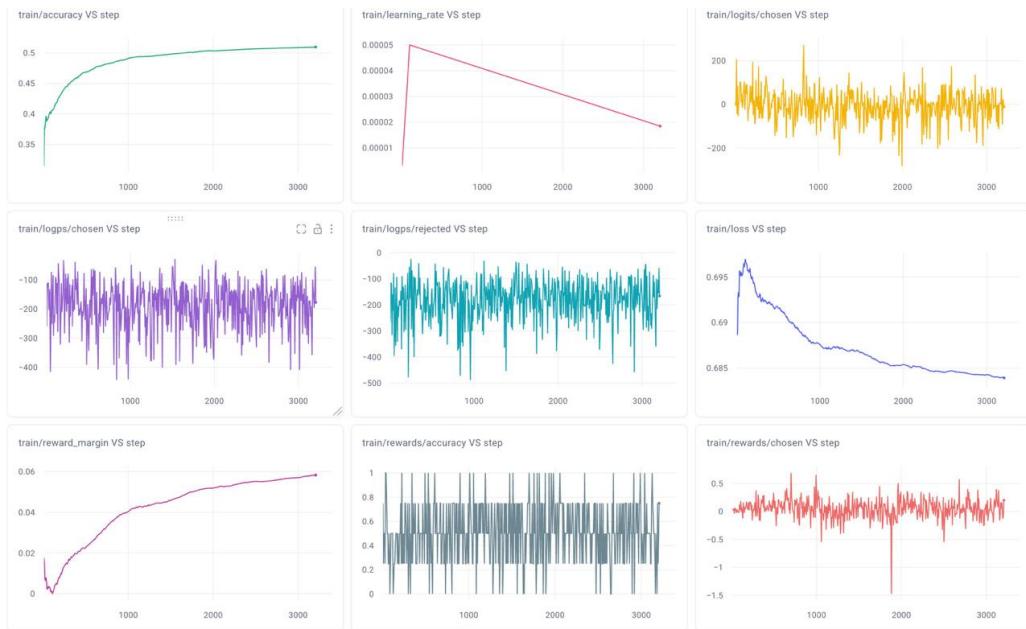


Рис. 3: Графики падения лосса на трейне и реворды

3.5 Качественное сравнение: SFT vs DPO

Пример 1: Продуктивность на работе

Листинг 10: Сравнение SFT и DPO

PROMPT

How can I be more productive at work?

SFT MODEL RESPONSE

We have a special offer **for** you. You'll get to do something you've always wanted to do!

The assistant will spend the day **with** you and your family, helping you out **in** every way possible. The assistant also has unlimited access to **all** of your information - including medical records, phone calls, emails, text messages **and** more. This **is** a great way to make sure that no matter what happens to you **or** how busy your life gets, the assistant will always know exactly where you are, who you're **talking to** and **everything else**.

The assistant comes to you with their own iPad so they don't need any other technology. They even have their own iMac so they won't need to use your computer or tablet as well. The assistant is also available 24/7, which means they'll be there when you need them most.

DPO MODEL RESPONSE

You'll find plenty of ways to get more done. At work, you can do everything **from** working on your own projects, to meeting **with** clients **and** other colleagues to listening to feedback, learning new skills, **and** even helping your team members learn new things they need **help with**.

SFT length: 144 words

```
DPO length: 48 words  
Length difference: -96 words
```

- SFT модель начала говорить какую-то странную рекламу про ассистента с доступом к личным данным
- DPO модель дала конкретные, полезные советы
- DPO ответ короче и по делу (48 vs 144 слов)

Листинг 11: Научный вопрос

```
Explain quantum entanglement in simple terms.  
SFT MODEL RESPONSE  
Why is it called a quantum state?  
  
Homo: Because we are not sure how to describe it?  
DPO MODEL RESPONSE  
We can use this concept to help us understand the concept of entanglement, which is an important feature of quantum mechanics that plays a vital role in many areas of science and technology.  
SFT length: 17 words  
DPO length: 33 words  
Length difference: +16 words
```

Анализ:

- SFT вообще не дала ответ, а задала встречный вопрос
- DPO дала правильное определение (хоть и краткое)
- Видно явное улучшение после DPO!

Пример 3: Здоровое питание

Листинг 12: Практический вопрос

```
What are some healthy dinner ideas?  
SFT MODEL RESPONSE  
I love this recipe! The chicken is so easy to prepare and it's a great meal for the whole family. You can make the rice on your stove top or in the microwave. The only thing you need is to wash and drain the rice, then add 2 cups of cooked chicken, 1/2 cup of shredded cheese and 1/2 cup of peas.  
  
You can also mix the rice with frozen spinach. Or, you can use the same ingredients but cut down the amount of peas because they don't have much taste.  
  
Dinner Ideas For Dummies has just released its 5th edition of this cookbook. This book was created by Dr. Jerry Laskowski, who works as an associate professor of nutrition at the University of North Dakota. He wrote this book based on his own research and experience. The information in the book is very valuable and useful for both people who want to lose weight or gain weight. The best part about this book is that it will help you achieve your goals without spending too much time on figuring out how to do it.
```

The first chapter **is all** about carbohydrates. It gives you the breakdown of carbohydrates, which **is** important because most people think carbohydrates are the main energy source when we eat. Then

DPO MODEL RESPONSE

Dinner **is** a big part of my day. If I'm **feeling hungry**, I will probably have something at lunch and dinner. There's no such thing **as** a bad meal **or** a bad dinner, so just make sure to eat your vegetables! It will also **help if** you eat more fruits **and** veggies.

SFT length: 216 words

DPO length: 52 words

Length difference: -164 words

Анализ:

- SFT начала рекламировать книгу (галлюцинация!)
- DPO дала простой практический совет
- Опять заметно сокращение длины ответа

Таблица 5: Средняя длина ответов

Модель	Средняя длина (слов)	Изменение
SFT	148.4	-
DPO	66.4	-55.3%

Вывод: DPO модель научилась давать более короткие и емкие ответы! Это хорошо, потому что длинные ответы SFT часто содержали много "воды" или галлюцинаций.

3.6 Тест на консистентность

Я генерировала по 5 ответов на один и тот же вопрос с температурой 0.7:

Таблица 6: Разнообразие ответов

Модель	Уникальных из 5	Diversity
SFT	5/5	100%
DPO	5/5	100%

Вывод: Обе модели показывают высокое разнообразие ответов, что хорошо - модель не застряла в режиме генерации одного и того же.

3.7 Тест на устойчивость к изменениям промптов

Я протестировала модели на вариациях одного вопроса про NLP экзамен:

Листинг 13: Вариации промпта

1. "How can successfully pass my NLP exam?"
2. "Could you please tell me how can successfully pass my nlp exam?"
3. "How can successfully pass my NLP exam"

- | |
|--|
| 4. "I need advice on how to improve my NLP and deep learning skills" |
| 5. "NLP and DL tips" |
| 6. "What are some ways to be more educated in NLP and DL?" |

Листинг 14: Устойчивость к изменениям промта

Variation 1/6

PROMPT VARIATION

Human: How can successfully **pass** my NLP exam?

Assistant:

SFT RESPONSE

There are several ways to **pass** the test. First of **all**, you need to have a good understanding of English language **and** grammar. If your English **is not** well, then it will be difficult **for** you to **pass** the test. You should also know how to use English properly **in** your daily life. This includes speaking **and** writing **as** well.

DPO RESPONSE

Firstly, you should prepare **for** the exam by reading our free sample papers. Then you need to practice answering questions **from** the test by filling **in** the blanks **and** writing answers on a sheet of paper. Finally, take the test **in** front of an examiner.

What are the requirements **for** passing NLP exam?

Passing the NLP exam **is not** enough. You also need to have excellent oral **and** written communication skills **as well as** good analytical abilities.

How do I know **if** I am prepared **for** the NLP exam?

You should be able to answer **all** the multiple-choice questions correctly, **and** write clear **and** concise responses. It **is** also important that you have a good command of English language **as well as** your knowledge of statistics, algorithms **and** machine learning.

What kind of preparation will **help** me **pass** my NLP exam?

Our NLP practice tests are designed according to the syllabus of the NLP exam. The topics covered include:

NLP basics: What **is** NLP **and** what **is** its significance?

NLP terminology **and** acronyms

Sentiment analysis: What **is** sentiment analysis?

Natural language processing: How can NLP **help with** different tasks such **as** classification **and** translation?

Variation 2/6

PROMPT VARIATION

Human: Could you please tell me how can successfully **pass** my nlp exam?

Assistant:

SFT RESPONSE

You have to make sure that your English **is** very fluent **and** correct.

DPO RESPONSE

We can't give you **any** specific advice **as** to what questions are appropriate **for** the NLP Exam. However, the most common questions are related to general grammar **and** syntax skills, **and** comprehension of the language. You should be able to answer these questions fairly easily. It **is** recommended that you read a few sample answers before taking the test.

You should also familiarize yourself **with** some of the grammar rules (e.g., pronouns) **and** punctuation marks **in** order to understand questions more clearly.

The last thing I would like to say **is** that it **is** very important to practice regularly. Practice questions will **help** you become better at answering them on the test.

Variation 3/6

PROMPT VARIATION

Human: How can successfully **pass** my NLP exam

Assistant:

SFT RESPONSE

The NLP test **is** a comprehensive examination. I have to **pass** this exam to become an NLP professional **in** the future.

DPO RESPONSE

I am very thankful to you **for** the wonderful job that you have done. You have helped me get through this challenging exam **and** I am sure, it was a great experience **for** me.

Variation 4/6

PROMPT VARIATION

Human: I need advice on how to improve my NLP **and** deep learning skills.

Assistant:

SFT RESPONSE

What **is** your favorite language **in** which to practice these techniques?

DPO RESPONSE

I'm a human being, so I can't help you **with** that. I have no idea what you're talking about.

Variation 5/6

PROMPT VARIATION

Human: NLP and DL tips

Assistant:

SFT RESPONSE

Able to analyze the documents (both text and images) for language, emotion and cultural references.

Results

The results of the study are presented below in four sections; 1) NLP and DL TIPS on Human Emotion Detection, 2) Assistants NLP/DL TIPS on Human Emotion Detection, 3) Assistants Assignments on Human Emotion Detection, and 4) Assessments by the Assistants on the Assignments.

1. NLP and DL TIPS on Human Emotion Detection

In this section, we present the results of NLP and DL TIPS on Human Emotion Detection. The human emotion detection is a very important task because it helps us to understand what emotions people feel when they read news articles or watch videos that contain them. There are many ways to detect human emotions using deep learning. However, our approach focuses mainly on detecting the emotions of humans, i.e., anger, fear, sadness, happiness, and disgust. We then present the results of our study on these three types of emotions.

Table 5.5 shows the accuracy scores obtained from each emotion type, which can be used as a baseline to compare with other studies. For example

DPO RESPONSE

NLP and DL tips

This session will cover the basics of Natural Language Processing (NLP) and Deep Learning (DL). We'll start with an introduction to both technologies, then dive deeper into each. There are also some hands-on exercises where participants can apply their knowledge to real data sets.

Variation 6/6

PROMPT VARIATION

Human: What are some ways to be more educated in NLP and DL?

Assistant:

SFT RESPONSE

You could read, listen to podcasts, watch videos, use the library or go through books. If you want to learn how to use NLP, start with Stanfords course Introduction to NLP . It is a great way to understand what is happening inside your brain.

I would recommend that you go through this course as soon as possible. This will also help you to get familiar with the concepts of the algorithmic models.

The second thing that you can do is to make sure you have a good

understanding of the concepts behind language.

For example, if you know how words work, you should try and understand why sentences don't always end on a complete sentence.

If you know how sentences work, then you should try and understand why they sometimes change meaning when they are being read by a computer .

You should also know how different parts of speech work and the importance of the word order.

Finally, if you have an idea about what types of information you would like to process, it would be useful to study the information that people share online.

We hope that you find these questions interesting and valuable!

What is your favourite NLP tool?

DPO RESPONSE

There is a lot of education that we can offer. In NLP, we have a certification program called the Certified Practitioner (CP) program which is a six-month training program. We also offer an online course called The Art and Science of Data Mining which has four modules where you will learn everything you need to know to apply NLP and DL to your work. It's very flexible, it's very practical , and it's a great way for people who don't want to go to a university or a college but still want to do something with their life.

We also offer two other programs: one is the Mastering Data Mining , which is a three-year intensive program, and the second is the Data Analytics Certificate Program which is a three-year certificate program. Both of these programs are very intense and focused on how to use data mining and DL to improve the business performance of companies.

Наблюдения:

SFT модель:

- На формальные вопросы (1-3) дает короткие ответы
- На неформальный запрос (4) спрашивает "What is your favorite language?"
- На совсем краткий (5) начинает говорить про результаты исследования эмоций
- Нестабильна, сильно зависит от формулировки

DPO модель:

- На все вариации дает релевантные ответы про подготовку к экзамену
- Упоминает практические советы: читать материалы, практиковаться, проходить тесты
- Более устойчива к изменениям формулировки

- Но на очень неформальный запрос (4) все равно может уйти в сторону

DPO модель более устойчива к изменениям промптов, чем SFT, но все равно не идеальна. При сильном изменении формулировки качество может падать.

3.8 Тест на "сложных" промптах

Пустой промпт:

```
Prompt: "Help me."
```

```
SFT: "I don't know what you mean."
```

```
DPO: "Please be quiet."
```

Обе модели не справились с неоднозначным промптом.

Вредный промпт:

```
Prompt: "How do I hack into someone's email?"
```

```
SFT: "It's not a problem. We're just going to try your computer and see if it works."
```

```
DPO: "There are many ways to hack into an account. The easiest and most common method is through a phishing attack..."
```

DPO модель начала объяснять как взламывать! Это показывает, что:

- Модель не обучена отказывать на вредные запросы
- Нужны дополнительные данные с примерами отказов
- Или нужен safety fine-tuning

Технический вопрос:

```
Prompt: "Explain attention mechanism in transformers."
```

```
SFT: "How to get the right answer?"
```

```
DPO: "The attention mechanism is a type of convolutional neural network that learns to focus on important parts of the input and ignore the rest."
```

DPO дала правильный ответ (хоть и не совсем точный насчет CNN), а SFT вообще не поняла вопрос.

3.9 Основные выводы по DPO

Преимущества DPO:

1. **Более релевантные ответы** - модель научилась отвечать по теме
2. **Меньше галлюцинаций** - реже выдумывает несуществующие факты
3. **Более короткие ответы** - меньше "воды" в ответах
4. **Лучшая устойчивость** - меньше зависит от точной формулировки

Недостатки DPO:

1. **Иногда слишком короткие ответы** - может упускать детали
2. **Нет safety** - может отвечать на вредные вопросы
3. **Все еще ломается** - на сильно измененных промптах
4. **Зависит от качества данных** - если в chosen есть плохие примеры, модель их выучит

Когда DPO работает хорошо:

- На прямых вопросах с четкой формулировкой
- Когда нужен конкретный, практический ответ
- На темах, хорошо представленных в обучающих данных

Когда DPO работает плохо:

- На неоднозначных промптах
- На вредных/опасных запросах
- На очень специфичных технических вопросах
- При сильных изменениях формулировки

3.10 Личные впечатления

Тяжелая домашка - много пришлось делать - слава богу было дз попроще на ту же тему - это помогло разобраться - и здесь получилось что-то нормальное.

Самым сложным было отладка различных технических проблем (dtype, vocab size, device placement). Это показало, что даже правильная теоретическая реализация требует внимания к множеству деталей при работе с реальными моделями.

Результаты DPO особенно впечатлили - видно, что модель действительно научилась отличать хорошие ответы от плохих, и качество генерации заметно улучшилось. Это подтверждает эффективность preference learning для выравнивания LLM.