

Отчёт по большому домашнему заданию 2

Введение

Понятное дело, что в первую очередь нам нужно определиться, как мы будем обрабатывать входные данные, какая модель подходит для нашей задачи и т.п. Давайте обо всём по порядку.

Промежуточный чекпоинт

На удивление, эта часть у мен отняла очень много времени. Я довольно быстро разобрался с тем, как устроены данные и как можно обработать. Но вот с моделью... Там я выбрал рекуррентную Seq2Seq, она смогла выдать мне необходимые для прохождения чекпоинта $BLEU = 10$. Я проделал достаточно большое количество экспериментов с этой моделью, и лучшее качество, которое мне удалось выбить на ней оказалось где-то порядка $BLEU = 14$, в общем слабо (графики я не делал тогда, без них). Ну и перевод был корявым каким-то, типо повторялись постоянно одни и те же слова, хотя этого не должно быть, впрочем, всё как на МДЗ-3.

Основной этап

Данные

Понятное дело, что для начала нужно было создать словари, что я собственно и сделал. Не знаю, что тут ещё описать можно...

Модель

Понятное дело, что рекуррентной нейронки будет мало, чтобы выбить нормальное качество, поэтому я использовал трансформер (finally). Ну, он достаточно классический, в него входит следующее:

- Positional Encoding
- Multi-Headed Attention
- Encoder
- Decoder
- Transformer

Ход работы и лучший результат

Итак, сначала я обучил модель с меньшим количеством слоёв и небольшим dropout на 20 эпохах. Вот её параметрами:

- `d_model = 256`
- `N = 3`
- `heads = 4`
- `dropout = 0.2`

У меня сразу же получился отличный BLEU (что-то порядка 26.5). Однако там было замечено резкое и сильное переобучение. На графике изображены loss и BLEU на валидации. Я попробовал добавить побольше голов (8), но особо результат не улучшился.

Далее я попробовал усложнить модель и сделать больше количество слоёв и голов в Multi-Headed Attention:

- `d_model = 512`
- `N = 6`
- `heads = 8`
- `dropout = 0.2`

BLEU чуть улучшилось (в районе 27.5). Но я не тот джедай, каким должен быть. Я хочу больше! Поэтому на следующем шаге я увеличил dropout до 0.3. И результат сразу же скакнул до `BLEU = 27.74` (лучший результат). Потом я пробовал играть с параметрами, но особого результата это не давало. Пробовал ещё dropout увеличить, но толку это вообще никакого не давало, и я уже решил забить.

Таким образом лучший результат получился со следующими параметрами трансформера:

- `d_model = 512`
- `N = 6`
- `heads = 8`
- `dropout = 0.3`

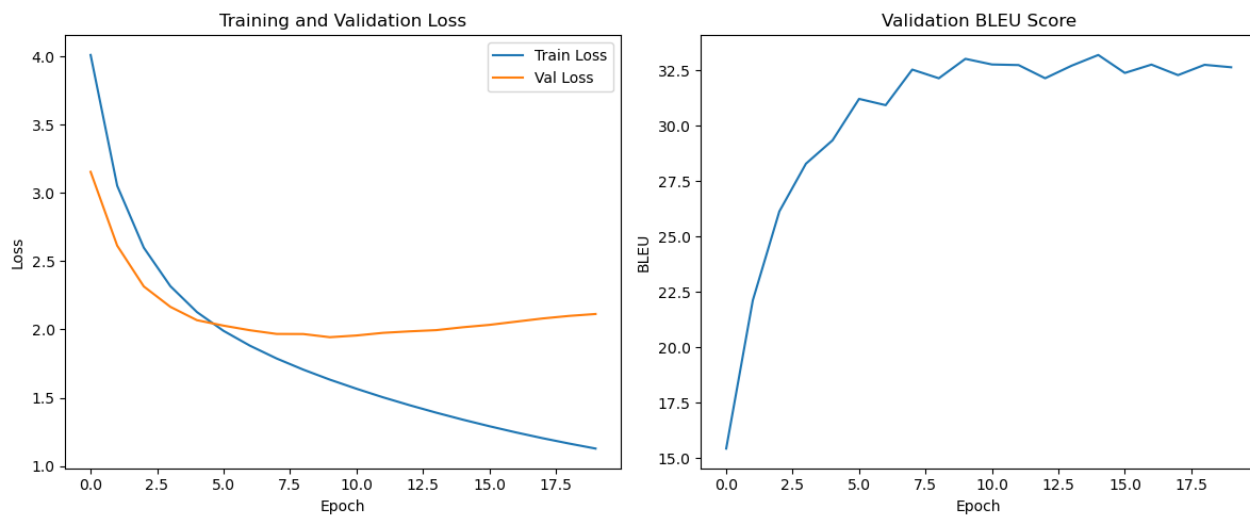


Рис. 1: Трансформер с небольшими параметрами.

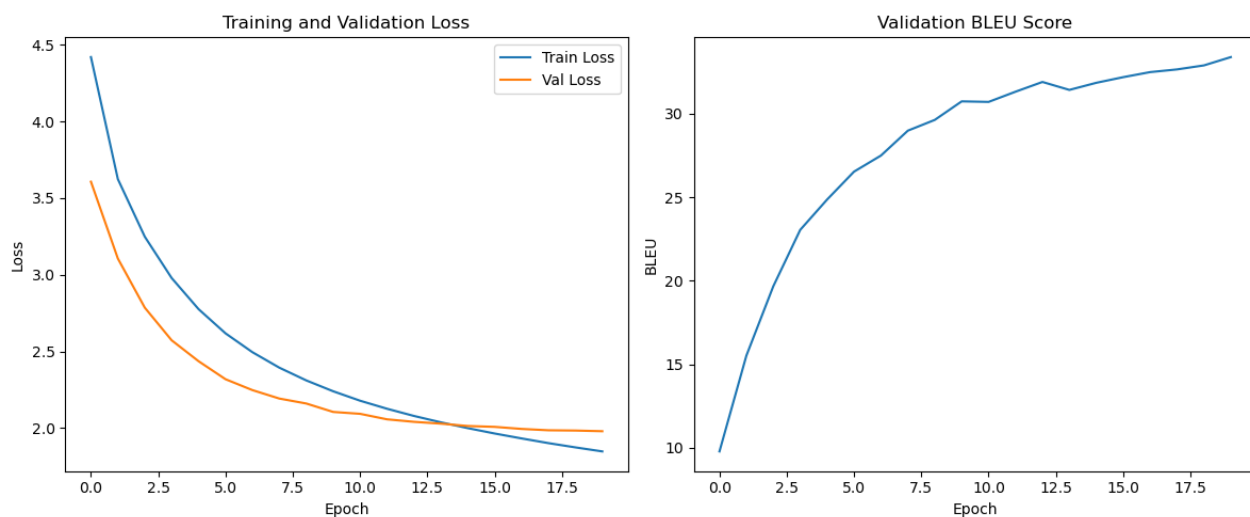


Рис. 2: Трансформер с Dropout(0.3) (лучшая модель).

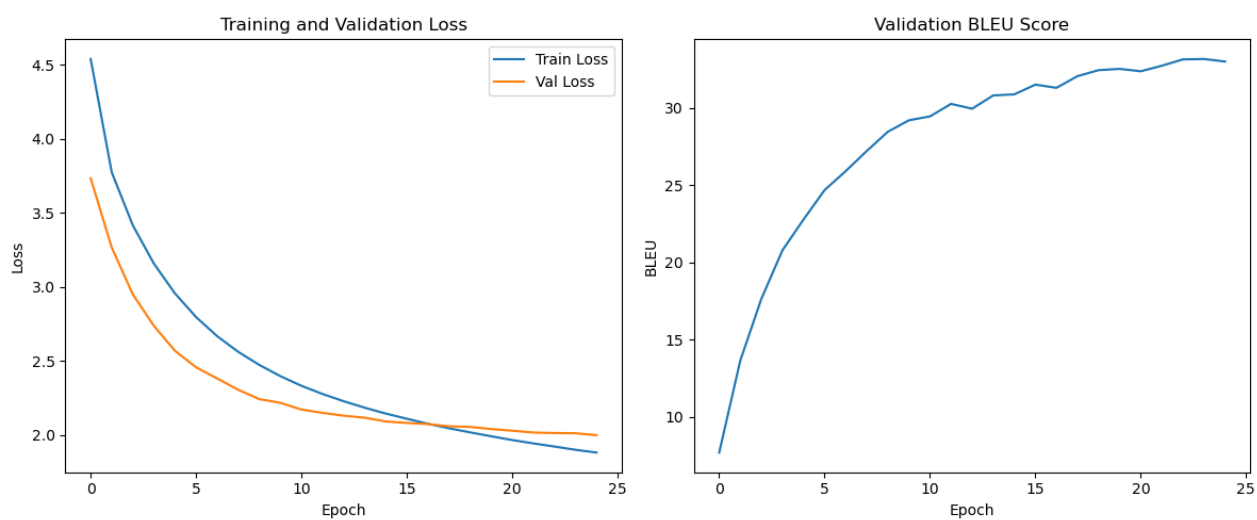


Рис. 3: Трансформер с Dropout(0.35).