

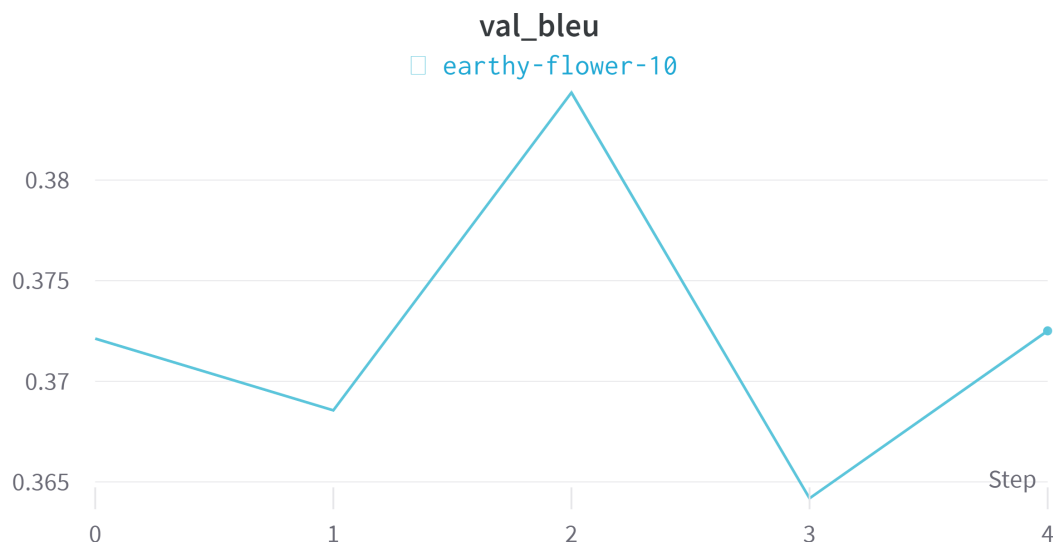
Отчет по БДЗ-2

Бабак Артём Юрьевич 201

22 марта 2023 г.

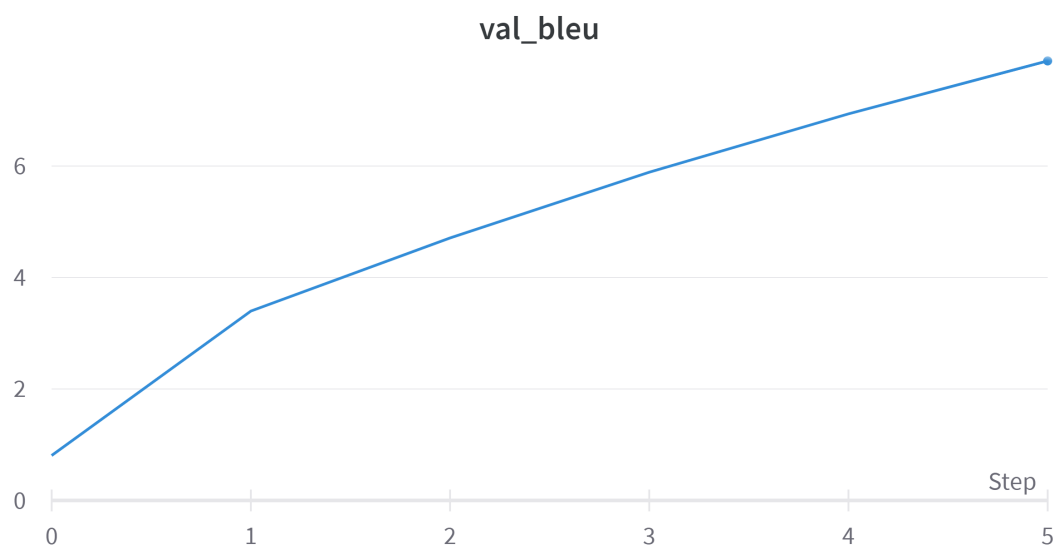
1 RNN

В первую очередь я решил попробовать использовать модель из МДЗ про генерации шуток немного адаптировав под текущую задачу. К сожалению адекватного результата я не получил. Было использовано различное количество слоев/размеры батчей/токенайзеры(bpe/word). Результат был примерно везде одинаковый.



2 Seq2Seq

Затем я решил перейти к sequence to sequence модели. Я реализовывал архитектуру из статьи Sequence to Sequence Learning with Neural Networks. С использованием данной архитектуры я смог получить BLEU > 5. Я использовал 2 слоя декодера и 2 слоя энкодера, использовал токенизацию по пробелам и включал в словарь все слова. Менять параметры я не пробовал.



3 Transformer

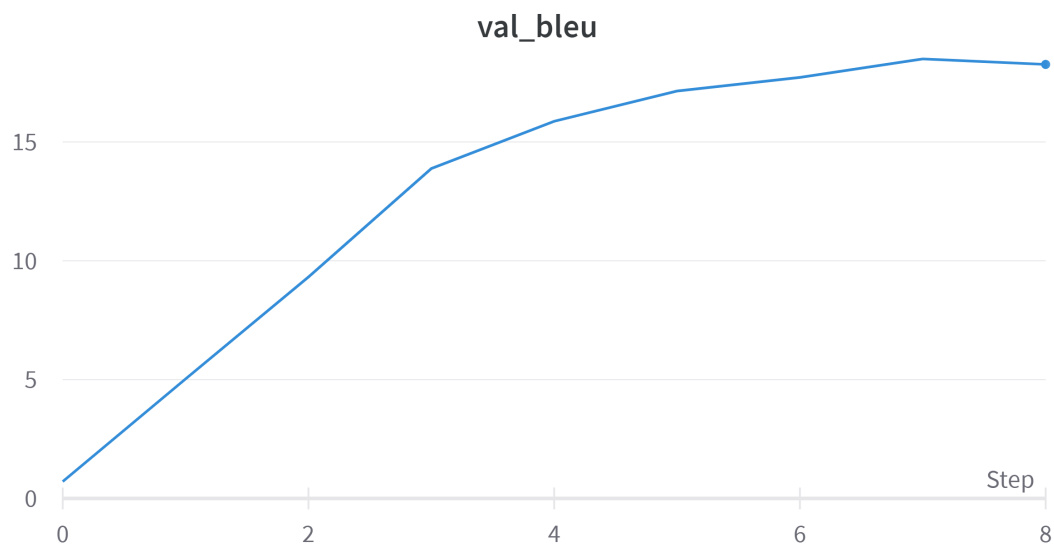
Затем я решил использовать архитектуру трансформера. Реализацию я взял из туториала с сайта pytorch. Изначально я использовал следующие значения параметров.

- Оптимайзер Adam с $lr = 1e-4$, $\beta_1 = 0.9$, $\beta_2 = 0.98$
- По 3 слоя энкодера и декодера
- Размер батча 128
- Размер feed forward слоя 512
- 8 слоев внимания
- Токенизация через split
- В словарь попадали только токены, которые встречались минимум 20 раз
- Dropout равный 0.1

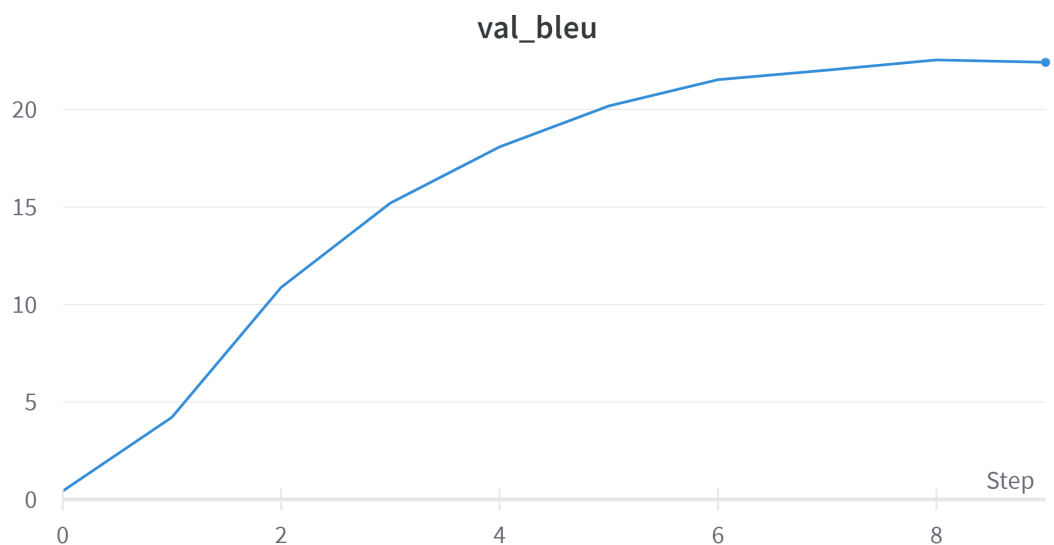
Обучив на 0.3 данных я получил следующий результат



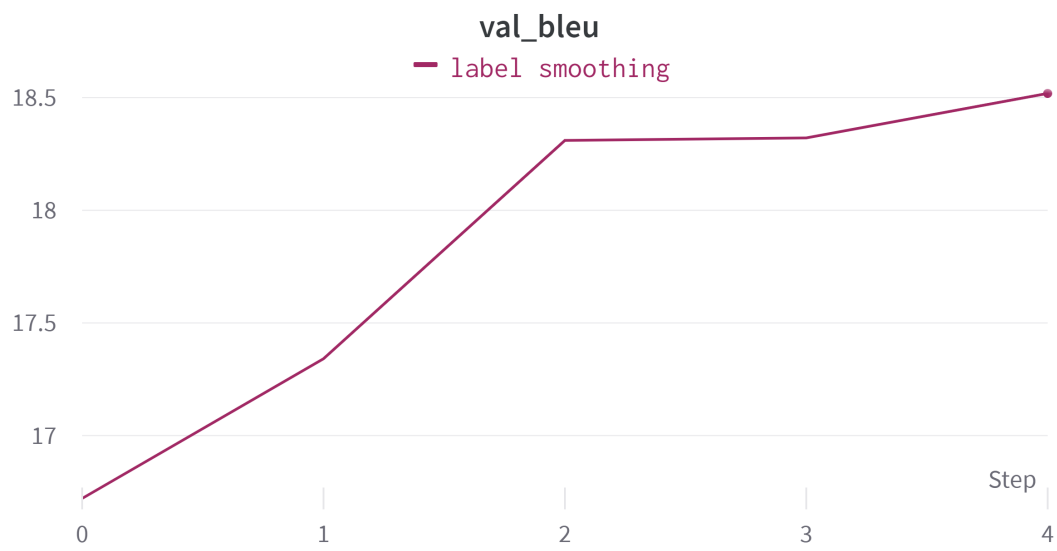
Затем я попробовал добавить шедулер. Я решил использовать OneCycleLR с $\max lr$ равным $3e-4$. Это улучшило качество на несколько пунктов.



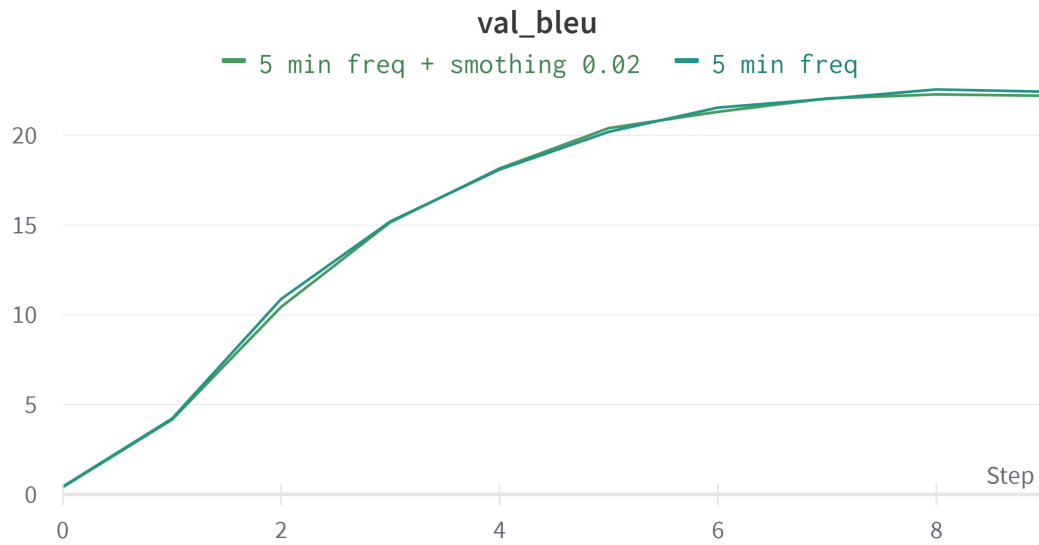
Затем я решил включить больше слов в словарь уменьшив минимальную частоту слова для его включения в словарь до 5. Это также заметно увеличило BLEU до 20+



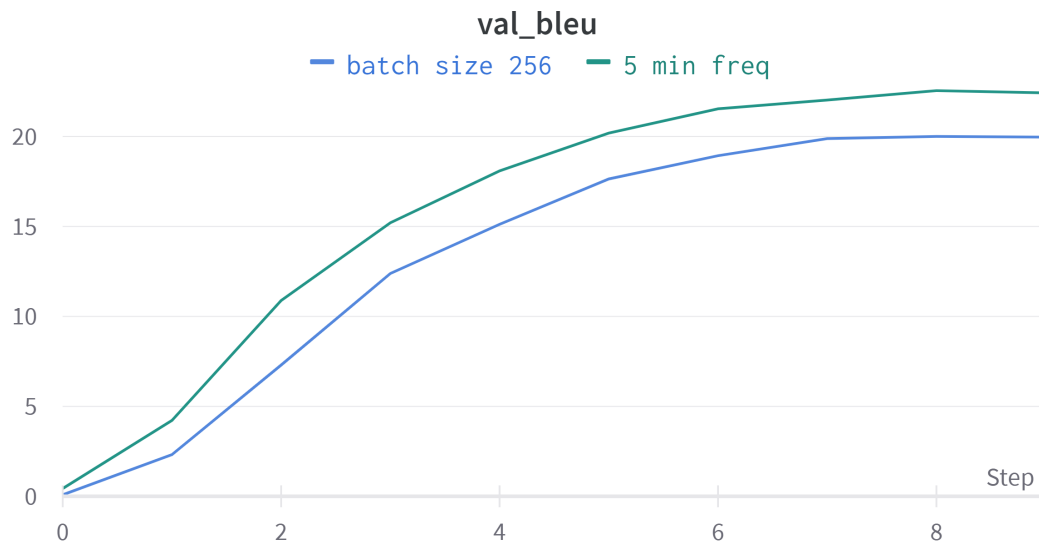
После я решил попробовать использовать label smoothing со значением 0.1. Наверное это оказалось очень большим значением так как произвело странный эффект. BLEU с первой эпохи стал довольно большим, но дальше уже модель не обучалась



Я решил уменьшить label smoothing до 0.02, однако с таким значением изменения не было совсем. В итоге я решил отказаться от label smoothing

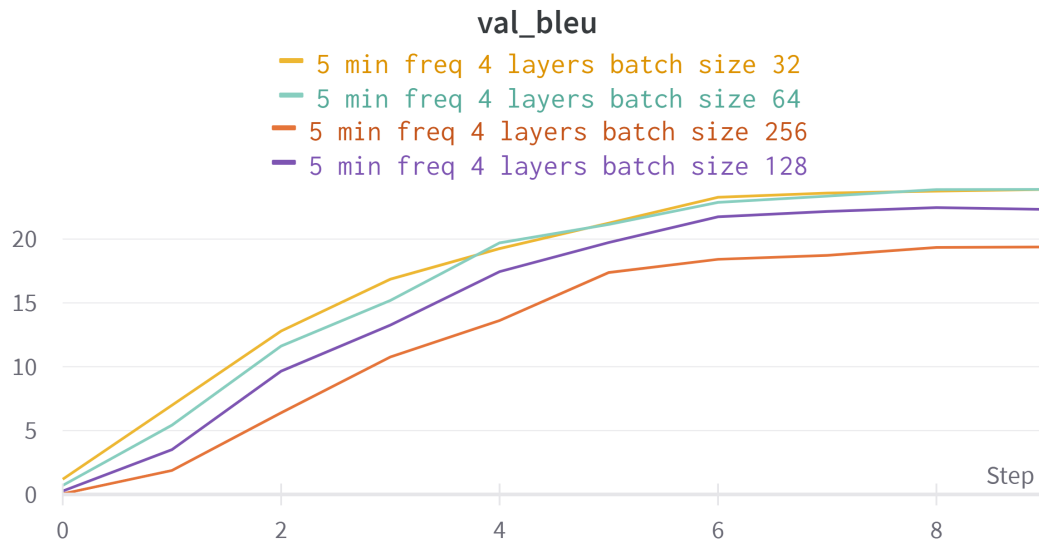


Дальше я решил попробовать увеличить размер батча до 256. Данное изменение только ухудшило BLEU при этом без видимого тренда на улучшение в будущем

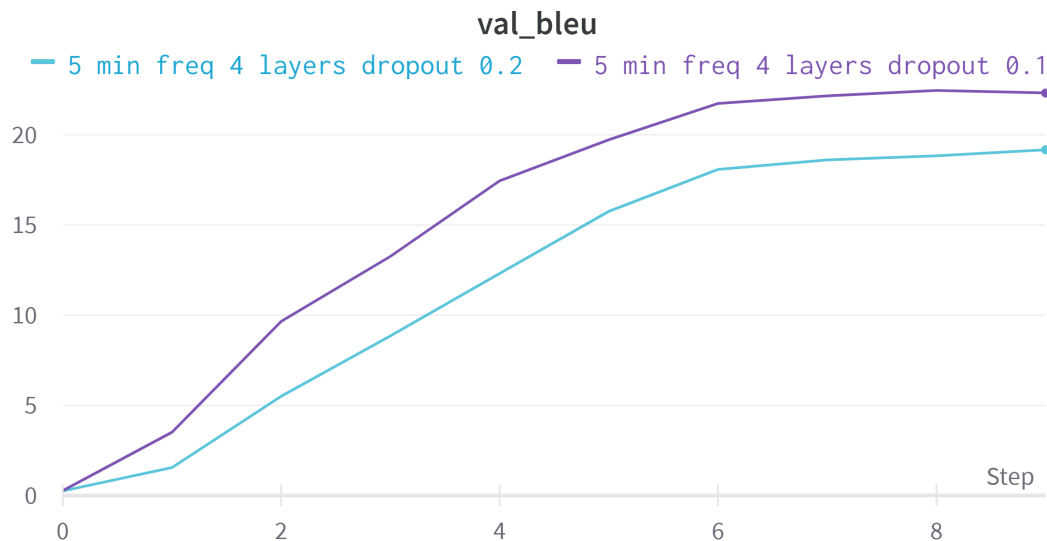


Также я решил увеличить количество слоев энкодера и декодера до 4. Изменений это не превнесло, но я решил оставить, так как решил, что в случае обучения на всех данных это все таки может помочь.

Я решил продолжить менять размер батча, но уже в меньшую сторону. До 64 и 32. Это дало положительный эффект и я решил оставить размер батча 64



Последнее, что я решил использовать это повысить dropout до 0.2, что плохо сказалось на качестве.



В итоге учитывая все изменения я обучил модель на всех данных и получил следующий результат

