Practical No. 3

Maciej Domaradzki
md406127@students.mimuw.edu.pl

10 kwietnia 2022

1 Part 1

g)

Mask sets e_t to $-\infty$ for padded tokens. Thanks to this probability of padded token is zero, because $\exp(-\infty) = 0$ and attention doesn't relay information from padded tokens.

i)

BLEU Score: 10.391320795176298

j)

	Advantage	Disadvantage
Dot product attention	Doesn't need extra matrix with weights	gives worse results than additive attention
Multiplicative attention	Less computationally expensive than other two types	Needs extra matrix with weight comparing to dot product attention
Additive attention	Gives the best results	More complex than other two types

2 Part 2

a)

- Original: So what do we know about this?
  Correct: Więc - co wiemy?
  Model's output: Co wiemy o to?

In Polish there are many ways to translate word 'this' and our model didn't concluse correct one from the context. We could add more similar examples to training set to fix this error.

- Original: So we have 45 minus - - what's 5 times 4?
  Correct: więc mamy 45 odjąć - ile to jest 5 razy 4?
  Model's output: Czyli mamy 45 minus minus ile to jest 5 razy 4?

  Model translated '-' as minus. This may be couse by the fact, that there are many examples in training data, where '-' is used with numbers as a minus not a hyphen. We could use more sentences not related to math expressions, but with numers and hyphens.

- Original: 'Anyway, this is an important thing in probability and '
  Correct: Niemniej jednak jest to ważna rzecz w prawdopodobieństwie
  Model's 'W każdym razie, to jest ważne rzecz w prawdopodobieństwo'

  Model didn't used correct grammatical form of word 'prawdopodobieństwo'. Maybe in training data, there isn't used form 'prawdopodobieństwie' or is used rarely, so model couldn't conclude correct form. We could add more examples, where the same English word is translated to Polish word in diffrent forms

- Original: 'this huge, the largest economy in the world doing everything in its power '
  Correct: 'ten wielki kraj, największa gospodarka na świecie robi wszystko co w jej mocy, '
  Model's output: <unk >to największe gospodarka w świecie robią wszystko w jego <unk >'

  We have here direct translation of 'in the world' to 'w świecie'. Model didn't conclude that in Polish, there is used different preposition in this context. We could add more examples with different translation of prepositions to training set.

- Original: But the Greek people could actually then move on with their lives.
  Correct: Ale Grecy mogliby właściwie żyć normalnie dalej.
  Model's output: Ale grecka ludzie mogą być w rzeczywistości <unk >

  Model tried translate 'Greek people' as two words. It could be caused by the fact, that model seen many examples of translating 'Greek' as 'grecki' or 'greacka', but didn't see it in context with word 'people' or seen it too rarely. We could add more examples in form nationality + 'people' to training set.

b)

The best model got 28.23 BLUE Score and our model got 6.316. Difference may be caused by using bigger training set and/or using better model. The property of Polish language that causes the drop is probably it's complex grammar. To overcome, at least partially, this issue, we could use more data during training with diverse examples.

c)

1. For c_1:
$$p_1 = \frac{4}{4} = 1, \quad p_2 = \frac{2}{3}, \quad c = 4, \quad r_* = 5, \quad BP = exp(-\frac{1}{5})$$

$$BLEU = exp(-\frac{1}{5}) \times exp(\sum_{n=1}^{2} \frac{1}{2} \log p_n) \approx 0.636$$

   For c_2:
$$p_1 = \frac{6}{8}, \quad p_2 = \frac{4}{7}, \quad c = 8, \quad r_* = 7, \quad BP = 1$$

$$BLEU = 1 \times exp(\sum_{n=1}^{2} \frac{1}{2} \log p_n) \approx 0.655$$

   The second translation got bigger BLEU score. I don't agree that it's better translation, because there is 'zbiór' instead of 'podzbiór' and there should be only one of those words: 'oznacza' and 'jest'.

2. For c_1:
$$p_1 = \frac{4}{4} = 1, \quad p_2 = \frac{2}{3}, \quad c = 4, \quad r_* = 5, \quad BP = exp(-\frac{1}{5})$$

$$BLEU = exp(-\frac{1}{5}) \times exp(\sum_{n=1}^{2} \frac{1}{2} \log p_n) \approx 0.636$$

   For c_2:
$$p_1 = \frac{3}{8}, \quad p_2 = \frac{1}{7}, \quad c = 8, \quad r_* = 5, \quad BP = 1$$

$$BLEU = 1 \times exp(\sum_{n=1}^{2} \frac{1}{2} \log p_n) \approx 0.231$$

   Now the first translation got bigger BLEU score and I agree that it's better translation.

3. Usually there are more than one way we can translate given sentence. Using only one reference may cause the overfitting of our model.

4.
   - Adventages:
     - fast and easy to computate
     - thanks to it commonness, models are easy to compare
   - Disadventages:
     - doesn't consider meaning
     - doesn't work well for morphologically rich languages