

Practical No. 3

Karol Pustelnik
kp446518@students.mimuw.edu.pl

April 10, 2022

1 Neural Machine Translation with Attention

g)

Explain (in around three sentences) what effect the masks have on the entire attention computation. Explain (in one or two sentences) why it is necessary to use the masks in this way.

- The Masks are used to prevent the decoder from outputting `<pad>` token. Pre-softmax value of negative infinity sets their probability to zero. We can say that masks hide padding tokens from attention mechanism so it does not pay attention to them.

i)

Please report the model's corpus BLEU Score.

- Corpus BLEU: 11.303918966764183

j)

We learned about dot product attention, multiplicative attention, and additive attention. Please provide one possible advantage and disadvantage of each attention mechanism, with respect to either of the other two attention mechanisms.

	Advantages	Disadvantages
Dot product attention	the fastest	no learnable parameters - gives worse results than the rest
Multiplicative attention	compared to additive attention - much faster and more space-efficient in practice, one matrix of learnable parameters	slower than dot product attention, adds more weights to network - increases number computations
Additive attention	the most complex form of attention - can give better results than multiplicative attention	slower than dot product attention, adds more weights to network - increases number computations, slower than multiplicative attention

2 Neural Machine Translation with Attention

a)

Please identify 5 examples of errors that your model produced. For each example you should provide a reason why the model may have made the error (either due to a specific linguistic construct or specific model limitations) and describe one possible way we might alter the NMT system to fix the observed error.

Error	Reason	Modification proposal
"Albo w pewnym sensie <unk>" - unknown word	limited model vocabulary due to the computationally intensive nature of the softmax	use different function or approximate softmax
"Nie można po prostu podstawić w nieskończoność i zobaczmy co się stanie." - wrong use of time	too small memory or not enough training examples	use transformers or add more training data
"Możemy zrobić to w każdym przypadku." - bad conjugation of the word (syntactic error)	the complexity of the conjugations of words in the Polish language	add another model that tries to correct syntax errors
"albo kiedy porównujemy liste liczby lub liste innych wartości." - wrong use of word (lista)	the corpus was not "mathematical enough" - NMT is domain specific	use specific corpus for specific domain of translation
" - lack of translation	original sentence was too long	use transformers

b)

1. Compare the results to the one from the PolEval track.
 - Our model scored third from the end in EN-PL translation. Not bad.
2. Please run `sh run.sh test full` and report the BLEU score. Check again `outputs/test outputs.txt`.
 - Corpus BLEU: 6.921245056954286
3. Why is the difference so large? What is the property of Polish language that is the main cause of the drop? What could be done to avoid this issue? Explain in less than 5 sentences.
 - Polish grammar is very complex with its verb conjugations, noun declinations, and grammatical cases. The word order in a sentence mostly does not matter and this makes it difficult for the model to learn the structure of sentences. However, the most important issue is that different conjugations of the same word can alter the BLEU score negatively. We can for example add something like syntax checker. If the output sentence is syntactically correct we can try to adjust the BLEU score calculation so it ignores conjugations of words.

c)

Question 1)

1. Please compute the BLEU scores for c_1 and c_2 .

S: So this means a strict subset

r_1 : Czyli to oznacza podzbiór właściwy

r_2 : W takim razie to oznacza podzbiór właściwy

c_1 : czyli to podzbiór właściwy

c_2 : w takim razie to oznacza jest zbiór właściwy

for c_1 :

	Count_{c_1}	$\max(\text{Count}_{r_i})$
czyli	1	1
to	1	1
podzbiór	1	1
właściwy	1	1

$\Rightarrow p_1 = \frac{4}{4} = 1$

P_2	Count_{c_1}	$\max(\text{Count}_{r_i})$
czyli to	1	1
to podzbiór	1	0
podzbiór właściwy	1	1

$\Rightarrow p_2 = \frac{2}{3}$

Now let's calculate BP

$$c_1 = 4, r^* = 5$$

$$BP = \exp\left(1 - \frac{5}{4}\right) = \exp\left(-\frac{1}{4}\right)$$

So we have:

$$p_1 = 1, p_2 = \frac{2}{3}, c_1 = 4, r^* = 5, BP = \exp(-\frac{1}{4})$$

$$\Rightarrow BLEU_c = \exp(-\frac{1}{4}) \cdot \exp(0.5 \cdot \log(1) + 0.5 \cdot \log(\frac{2}{3}))$$

$$\approx 0.64$$

for c_2

	count _{c_2}	max(count _{r_i})
w	1	1
takim	1	1
razie	1	1
to	1	1
oznacza	1	1
jest	1	0
zbidr	1	0
wtedy	1	1

$$\Rightarrow p_1 = \frac{6}{8} = \frac{3}{4}$$

bigrams for c_2

	count _{c_2}	max(count _{r_i})
w takim	1	1
takim razie	1	1
razie to	1	1
to oznacza	1	1
oznacza jest	1	0
jest zbidr	1	0
zbidr wtedy	1	0

$$\Rightarrow p_2 = \frac{4}{7}$$

Now let's calculate BP:

$$c_2 = 8 \quad r^* = 4 \quad \rightarrow \quad BP = 1$$

So we have:

$$p_1 = \frac{3}{4}, \quad p_2 = \frac{4}{4}, \quad c_2 = 8, \quad r^* = 4, \quad BP = 1$$

$$BLEU = 1 \cdot \exp(0.5 \cdot \log(\frac{3}{4}) + 0.5 \log(\frac{4}{4})) \approx 0.65$$

2. Which of the two NMT translations is considered the better translation according to the BLEU score? Do you agree that it is the better translation?
- According to the BLEU score the translation c_2 is better. However, I do not agree with that because the translation is not grammatically correct.

Question 2)

Our hard drive was corrupted and we lost Reference Translation r_2 .

1. Please recompute BLEU scores for c_1 and c_2 , this time with respect to r_1 only.

for c_1 :

	Count c_1	Count r_1
czyli	1	1
to	1	1
podzbiór	1	1
własny	1	1

$\Rightarrow p_1 = \frac{4}{4} = 1$

P_2	Count c_1	Count r_1
czyli to	1	1
to podzbiór	1	0
podzbiór własny	1	1

$\Rightarrow p_2 = \frac{2}{3}$

Now let's calculate BP

$$c_1 = 4, r^* = 5$$

$$BP = \exp\left(1 - \frac{5}{4}\right) = \exp\left(-\frac{1}{4}\right)$$

So we have:

$$p_1 = 1, p_2 = \frac{2}{3}, c_1 = 4, r^* = 5, BP = \exp(-\frac{1}{4})$$

$$\Rightarrow BLEU_{c_1} = \exp(-\frac{1}{4}) \cdot \exp(0.5 \cdot \log(1) + 0.5 \cdot \log(\frac{2}{3}))$$

$$\approx 0.64$$

for c_2

	Count _{c_2}	Count _{r_1}
w	1	0
takim	1	0
razie	1	0
to	1	1
oznacza	1	1
jest	1	0
zbiór	1	0
wtedy	1	1

$\Rightarrow p_1 = \frac{3}{8}$

bigrams for c_2

	Count _{c_2}	Count _{r_1}
w takim	1	0
takim razie	1	0
razie to	1	0
to oznacza	1	1
oznacza jest	1	0
jest zbiór	1	0
Zbiór wtedy	1	0

$\Rightarrow p_2 = \frac{1}{4}$

Now let's calculate BP:

$$c_2 = 8 \quad r^* = 5 \quad \rightarrow \quad BP = 1$$

So we have:

$$p_1 = \frac{3}{8}, \quad p_2 = \frac{1}{4}, \quad c_2 = 8, \quad r^* = 5, \quad BP = 1$$

$$BLEU = 1 \cdot \exp(0.5 \cdot \log(\frac{3}{8}) + 0.5 \log(\frac{1}{4})) \approx 0.23$$

2. Which of the two NMT translations now receives the higher BLEU score? Do you agree that it is the better translation?

- The translation c_1 now receives a higher BLEU score. I agree that it is the better translation.

Question 3)

Due to data availability, NMT systems are often evaluated with respect to only a single reference translation.

1. Please explain (in a few sentences) why this may be problematic.

- You can express the meaning of the same sentence in many different ways, including using completely different words. NMT can give a really good translation but will receive a low BLEU score because it did not use words specified in the reference translation.

Question 4)

List two advantages and two disadvantages of BLEU, compared to human evaluation, as an evaluation metric for Machine Translation.

	Advantages	Disadvantages
BLEU	automated process, fast	can be inaccurate, cannot evaluate importance of errors