Assignment 4
Zeyad Ezzat
4492

1. Neural Machine Translation with RNNs
   a. Code
   b. Code
   c. Code
   d. Code
   e. Code
   f. Code
   g. We don't include Masking in attention score because It's not related to the context of the sentence.
      We need to use masking because we are going to find a words in the test set doesn't appear in the train set so we can mask it.

   h. Bleu score after training =35
      Training time =3:36 hour
      I didn't edit the training parameters provided
   i. Dot product doesn't require expensive computation like matrix multiplication but in matrix multiplication we can multiply two matrices to get a new shaped matrix and that's not available in dot product matrix multiplication and
      multiplication can be used in different shapes but you can't do that in additive

      advantage :doesn't disadvantage: can't map from different matrix shapes

2. Neural analysis
   a. Find the error
      i. The error: "favorite of my favorites"
         The correct :"one of my favorites"
         here the network has realized that we are taking about a one thing of the favorites but didn't get the word 'one' so we may increase the brunching factor in the beam search so we get a better performance
      ii. The error : not in translated the sentence but in the mapping the sentence from the Spanish to English ,
         so we could increase the number of RNN layers , this possibly will relate the sentence in the end to the sentence at the start.
      iii. The error : we find a word not in the corpus
         we can make a POS Detection to the words in this case , if we find that it's a name like this case , just copy it
      iv. The error in translating manzana
         because manzana has two meanings apple and block , so you

              will need to have examples with "manzana" have both meaning to know the difference in the context

v. The error : network mapped teacher to women
this bias because most of the teachers are women , so you will need to preprocess the corpus and make sure that , It's balanced

vi. The error : in transferring from hectare to acres
the network copied the number and mapped hectare to acres this is a hard one , but you will need more examples with the same case , or increase the number of neurons in each layer so we can extract more complex patterns like this

b.

    i.

      correct: You know, what I do is write for children, and I'm probably America's most widely read children's author, in fact.
produced: You know what I do is write for kids , and in fact , I 'm probably the author for kids , read in the United States .
There is two mistakes
1- changed children with kids , this can be excluded because it have the same meaning
2-the arrangement of the sentence is different

    ii. Correct: About three or four years ago, I got a phone call in the middle of the night from that teacher, Mrs. Posten, who said, "I need to see you.
Produced: About three or four years ago , I got a phone call from that teacher , the <unk> lady who said , `` I need to see you .
the mistake that Posten doesn't appear in the training set

c. Laksd

    i. Len(c1)=5   len(c2)=5   len(r1)=6   len(r2)=4
BP1=exp(1/6) BP2=exp(1)
for c1 P1=3/5   P2=2/4
for c2 P1=4/5   P2=2/4
Bleu1 = exp(1/6)*exp(.5log(3/5)+.5log(2/4))=0.90957
Bleu2 = exp(1)*exp(.5log(4/5)+.5log(2/4)=)2.2
second is better, in my opinion both is almost identical the difference is not that huge

    ii. Len(c1)=5   len(c2)=5   len(r1)=6
BP1=exp(1/6) BP2=exp(1/6)
for c1 P1=3/5   P2=2/4
for c2 P1=2/5   P2=1/4
Bleu1 = exp(1/6)*exp(.5log(3/5)+.5log(2/4))=0.90
Bleu2 = exp(1/6)*exp(.5log(2/5)+.5log(1/4))=0.45
first is better , I agree because It's more logical to map the first to it's refrence than mapping the second to the first refrence

iii. It's not accurate because we can map one sentence to more than one sentence with different words but with the same meaning
iv. A
  1. Advanteges
     easy to understand
     we can apply it to any language
  2. Disadvantages
     Doesn't really represent the meaning of the sentence
     It's just count the common words
     will be complex to apply it for high number of n-grams