CSE 291: Advanced Statistical NLP

Project 1: Machine Translation

Chen Shen A59003243

Due May 7th by midnight

Solution I—Build a seq2seq model

From Figure 1, we can see the training loss decreases as epoch increases. However, it is overfitting after 7 epoches.

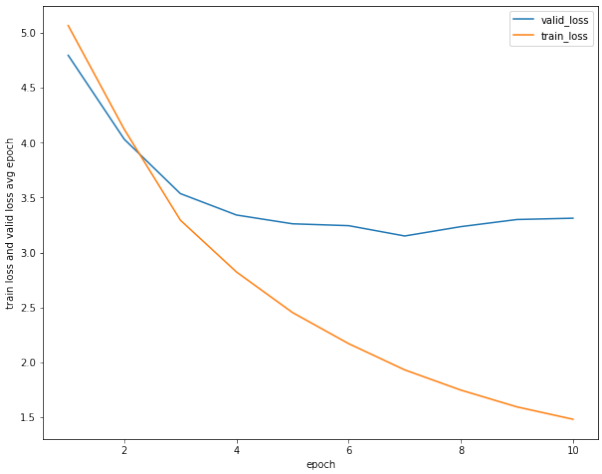

Figure 1: Training and validation loss per epoch

Then let we see some examples of translation.

Sample 1:

source	drei kleine kinder stehen um ein blau-weißes fass herum .
target	three young children stand around a blue and white barrel .
predicted	three small children stand around a blue cloth .

The attention matrices is shown in Figure 2.

Sample 2:

source	die person im gestreiften shirt klettert auf einen berg .
target	the person in the striped shirt is mountain climbing .
predicted	the person in the striped shirt is climbing a mountain .

The attention matrices is shown in Figure 3.

The BLEU score is 29.24.

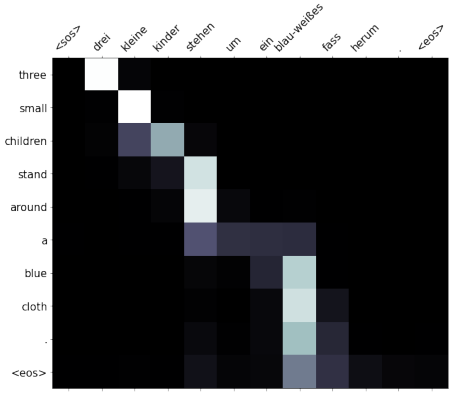

Figure 2: Sample 1 Attention Matrices

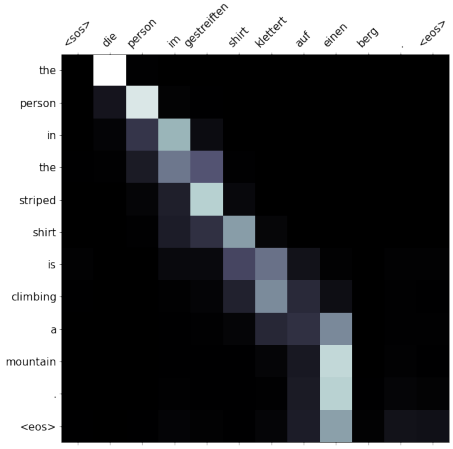

Figure 3: Sample 2 Attention Matrices

Solution II—Beam search

The following are some examples about beam search. The beam size here we use is 5.

Sample 1:

source	ein mann grillt im freien in seinem hof .
target	a man is grilling out in his backyard .
greedy search	a man is grilling in in his yard .
beam search (5)	a man is grilling outside in his yard .

Sample 2:

source	ein mann seilt sich an einer klippe über dem ozean ab .
target	a man is abseiling down a cliff over the ocean .
greedy search	a man <unk> <unk> a cliff above the ocean .
beam search (5)	a man is sunbathing on a cliff above the ocean .

Sample 3:

source	ein mann verkauft waren am rand der straße , die zu einem großen berg im canyon führt .
target	a man selling items near the roadside to the great mountain of canyon .
greedy search	a man selling selling the the the the the in a large large . .
beam search (5)	a man is selling items on the side of a large canyon .

From the above three examples, the beam search result are much better than greedy search. They are more fluent and more accurate.

Then we compare the BLEU result on test data.

Method	BLEU score on test data
greedy search	29.24
beam search (5)	29.35
beam search (10)	29.13

From the BLEU score on test data, we can see the beam search result is a little better than greedy search when I the beam size is 5.

Solution III—Copy mechanism

The following are three examples of the result after I use CopyNet mechanism against one without it. The minimum frequency of source and target vocabulary are both set to 2. The source sentence below has already been processed by source vocabulary.

Sample 1:

source	ein chevrolet der auf einer messe ausgestellt ist
original target	<unk> car on display at a convention
copynet target	chevrolet car on display at a convention
original prediction	a <unk> car that is at a a .
copynet prediction	a chevrolet car is asleep on a convention .

Sample 2:

source	ein corgi läuft einen pfad im wald herunter .
original target	a <unk> is walking down a path in a forest .
copynet target	a corgi is walking down a path in a forest .
original prediction	a <unk> is walking down a path in the woods .
copynet prediction	a corgi is walking down a trail in the forest .

Sample 3:

source	ein mädchen in einer burka lernt in einem klassenraum <unk> .
original target	a girl in a <unk> is learning <unk> in a classroom .
copynet target	a girl in a burka is learning <unk> in a classroom .
original prediction	a girl in a <unk> learning learning learning in a classroom .
copynet prediction	a girl in a burka is learning in a classroom .

Then, we compare the blue score on test data.

Method	**BLEU score on test data**
original	29.24
copynet	28.31

Benefits:
The red word in the above sample is that CopyNet directly copy from source sentence. We can easily see that CopyNet improve the prediction result to contain words that is not in the target vocabulary. And also, from sample 1 and 3, it increases the fluency of the predicted result. Maybe it is because that CopyNet contains more information than the original model.

Drawbacks:
The CopyNet model is more complicated than original model. Hence, it will be much easier to be overfitting with limited data. We can see this from the BLEU score on test data.
The CopyNet can mispredict the result by using the word in source sentence. You can see it from the following example.

source	zwei personen fahren auf fahrrädern durch eine gebirgslandschaft .
target	two people riding bikes through a mountainous region .
copynet prediction	two people riding bikes bikes through a gebirgslandschaft .