CSC321: Assignment #3

Xiangyu Kong
kongxi16

March 19, 2018

Problem 1

1. The model will not perform well on long sequences. This is because after the encoder compresses the long sequence of inputs into the fixed length vector h_v, h_v will have to convey much more information. This makes the vector harder to interpret for the decoder layer, and the decoder will less likely give a correct result.

2. The added words are 'team', 'problematic', 'ink', 'obviously', 'shy', 'philosophical', and 'supercalifragilisticexpialidocious'

   The predicted results are shown in below.

   As we can see, for shorter words like team and ink, the translator does a good job. However, for longer words like 'problematic', 'philosophical' and 'supercalifragilisticexpialidocious', the translator does a very poor job.

   Listing 1: Translated Results

```
team --> eamtay
problematic --> opserarcepray
ink --> inkway
obviously --> odloy-eylway
shy --> ytsay
philosophical --> isorecarcalclay
supercalifragilisticexpialidocious --> afessfsesesssipicici
```

Problem 2

1. The problem with teacher forcing is that during training, the decoder are given in ground truth token as input to next time unit. However, when training, the decoder has to predict the next letter given the previous prediction. If the previous prediction is incorrect, then the error will be enlarged through time because every time step, the decoder is predicting based on a false previous input.

2. A solution would be to generate a token that contains necessary and important information for predicting the next result to use as the next time step's input instead of simply using the previous result.

Problem 3

Problem 4

Problem 5