CS 224n Assignment #4

Yoshihiro Kumazawa

June 18, 2020

1. Neural Machine Translation with RNNs

(a) See `utils.py`.

(b) See `model_embeddings.py`.

(c) See `nmt_model.py`.

(d) See `nmt_model.py`.

(e) See `nmt_model.py`.

(f) See `nmt_model.py`.

(g) The masked logits are made $-\infty$ and hence do not affect the softmax calculation of the other logits. Those masks are put on the hidden states from the padded words, which no attention should be paid to.

(h) (Missing)

(i) The model's corpus BLEU Score was 35.83.

(j) One advantage of dot product attention compared to multiplicative attention is that it is less prone to overfitting since it does not have weight parameters. One disadvantage is its lower expressivity. One advantage of additive attention compared to multiplicative attention is that it can learn not to be affected by the hidden state very much by letting W_1 small. One disadvantage is that it is more prone to overfitting.

2. Analyzing NMT Systems

(a) i. 1. "Aquí" is translated to "Here's" instead of "So". "favorite" is redundant.

2. "Aquí" could be "Here", but the verb "is" is unnecessary anyway. A possible reason is that sentences with no verb were seldom fed during training time. The reason the NMT translated "otro" into "another favorite" instead of "another one" might be that NMT thought it could improve translation accuracy by understanding what such demonstrative words actually mean.

3. For the first error, feeding more data that have a similar characteristic might work. For the second error, penalizing redundancy might be a good idea.

ii. 1.

2.

3.

iii. 1.
2.
3.
iv. 1.
2.
3.
v. 1.
2.
3.
vi. 1.
2.
3.

(b)

(c) i.
ii.
iii.
iv.