Assignment #5

1. Character-based convolutional encoder for NMT

(a) Number of unique character is way less than the number of unique words, so we need smaller embeddings to learn the patterns.

(b) Character-based:

(1) Embedding: $e_{char} \cdot V_{char}$

(2) Convolutional layer: $e_{word} \cdot e_{char} \cdot k + e_{word}$

(3) Highway: $e_{word} \cdot e_{word} \cdot 2 + e_{word} \cdot 2$

Word-based:

(1) Embedding: $e_{word} \cdot V_{word}$

If $k=5$, $V_{word} = 50000$, $V_{char} = 96$

char-based # parameters $= 50 \times 96 + 256 \times 50 \times 5 + 256 + 256^2 \cdot 2 + 256 \cdot 2 = 200640$

word-based # parameters $= 50000 \times 256 = 12800000$

factor $= 12800000 / 200640 = 63.8$

∴ word-based model has 63.8 times more parameters than char-based model

(c) 1-D CNN has way less number of parameters to learn than RNN. In our case, 1-D CNN has $256 \times 50 \times 5 + 256 = 64256$ parameters to learn. RNN would have to learn $256^2 \cdot 2 = 131072$ parameters. Thus, 1-D CNN would be easier and faster to train.

(d) Max-pooling keeps only the most important information but ignores all other less important information. By contrast, average pooling keeps the general information but not accentuate the most import one.

3. Analyzing NMT Systems

(a) Only 'traduce' is in the vocabulary. This is bad because it's not inclusive in terms of 'who' translate. The character-aware NMT model may overcome this problem since it can infer who translate from the context. It predicts the proper word char by char, so any word can be generated.

(b) (i) financial ⟶ economic
neuron ⟶ nerve
Francisco ⟶ san
naturally ⟶ occuring
expectation ⟶ norms

(ii) financial ⟶ vertical
neuron ⟶ Newton
Francisco ⟶ France
naturally ⟶ practically
expectation ⟶ exception

(iii) Word2Vec models semantic similarity and CharCNN models word structure similarity. These are reasonable because Word2Vec predicts next word based on semantic context. Words explaining similar ideas are usually put at similar places in the context. On the other hand, CharCNN tries to construct the word correctly, so it cares more about the form of the words than semantic meaning.

(c) Spanish: Un amigo mo hizo eso -- Richard Bollingbroke.
Eng-reference: A friend of mine did that -- Richard Bollingbroke.
Q4 translation: A friend of mine did that -- Richard <unk>.
Q5 translation: A friend of mine did that -- Richard Bolivia.
This is an incorrect example since the charCNN model changes the person's name. It predicts in this way since the original word 'Bollingbroke' never appears in the training dataset. The model tried to find a similar name word instead.