

(1)

(a)

It is because the characters themselves don't hold much information, so a lower dimension is fine. However, if they are composed in a certain way, they do contain more information.

(b)

word-based lookup embedding:

$$\text{Number of Parameters} = V_{\text{word}} e_{\text{word}}$$

character-based embedding:

$$\text{Number of Parameters} = V_{\text{character}} e_{\text{word}} + k e_{\text{word}}$$

(c)

One advantage of using convolution nets over RNN is the ability for feature recognition. Since you are using filters to determine features in a space, it should be able to pick up on parts of words. As opposed to the rnn which would struggle with determining the features.

(d)

One advantage of Max pooling is that it is able to determine the features with the most efficiency. It is only taking the window that correlates most with the filter. One advantage of Average pooling is that all the information from the previous layer is still there, so no information is lost.

(2)

0.1 (f)

BLEU: 24.5

(3)

a

traducir occurs

traduzco does not exist

traduces does not occur

traduce does occur

traduzca doesn't exist

traduzcan also does not exist

Since all of these have similar meanings, it would make intuitive sense that information should be shared between them. A word based NMT might not have enough info on each to learn the different form of the same word. However, our character-aware NMT might take tradu and derive that it should have something to do with translate. Then, it can use rules from other places to figure out the form of translate.

(b)

(i)

closest based on cosine similarity:

financial: economic  
neuron: nerve  
Francisco: San  
naturally: occurring  
expectation: norms

(ii)

closest of char embedding

financial: vertical  
neuron: Newton  
Francisco: France  
naturally: practically  
expectation: exception

(iii)

It seems like the CharCNN embedding layer is definitely finding words that have similar letters, rather than finding words that have the same meaning. This means that instead of going from expectation to norms it favors exception because they share a lot of similar characters.

c

original spanish sentence: Hoy estoy aqu para hablarles sobre crculos y epifanas.

reference english sentence: I'm here today to talk to you about circles and epiphanies.

a4 generated sentence: I'm here today to talk to you about circles and junk.

a5 generated sentence: I'm here today to talk about circles and skines.

This is not an acceptable response. Skines means nothing, this may happen because of lack of training data involving the word epiphany.

original spanish sentence: Le encontramos un lugar, la internamos, y la cuidamos y nos encargamos de su familia, porque era necesario,

reference sentence: We found her one, we got her there, and we took care of her and watched over her family, because it was necessary.

a4 sentence: We found a place, the junk and the monitored and we get rid of his family, because it was necessary, because it was necessary, because it was necessary, because it was necessary, because it was necessary, because he was needed.

a5 sentence: We found it a place, we tried, and we take it out of their family, because it was necessary.

While not ideal, I think this is an acceptable response. There may be multiple meanings of words in spanish and the model got confused.