

(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.