

1. a). Because there are much more ^{parameters in} characters than words. Therefore, to reduce the calculation, we are having smaller embeddings.

b). Character embedding: Parameter: Embedding + Convolutional network + Gates and

$$\text{projection} = V_{\text{char}} \cdot e_{\text{char}} + f \cdot e_{\text{char}} \cdot \text{act } f + e_{\text{word}} \cdot e_{\text{word}} + e_{\text{word}}$$

$$= V_{\text{char}} \cdot e_{\text{char}} + e_{\text{word}} \cdot e_{\text{char}} \cdot k + e_{\text{word}} + e_{\text{word}}^2$$

$$= 96 \cdot 50 + 256 \cdot 50 \cdot 5 + 2 \cdot 256 + 256^2$$

$$= 134848$$

Word embedding: Parameter: $V_{\text{word}} \cdot e_{\text{word}}$

$$= 50000 \cdot 256$$

$$= 2800000$$

c). A 1D convnet have kernels to help to see the relationships between characters inside a word by using n-gram. We can not do it in RNN which we could only look at characters sequentially

d). Max pooling: It helps to show the most prominent feature by choosing the strongest signal.

Average pooling: It helps to show the overall features and could give some information of weak signals.

3. a). 1. traduir: 36743

2. traduce: 8747

Others no.

It is bad, because other "traduir" would be translated to English as <UNK>.

Our character based NMT could help to reconstruct these words according to the "traduir" we have into the correct translation.

b). i). financial \rightarrow economic

neuron \rightarrow nerve

Francisco \rightarrow San

naturally \rightarrow occurring

expectation \rightarrow norm

ii). Character-based: financial \rightarrow vertical

neuron \rightarrow Newton

Francisco → France
naturally → practically
expectation → exception

The Word2Vec uses the frequency of word, so it would display words that are often used together. Therefore, the nearest words are close in meaning. The char-based model checked on characters. Therefore, the structures of words are similar.

Spanish: Vol en el avin vicepresidencial por ocho aos.

reference English translation: reference English translation

English translation generated by A4: I'm here to talk to you about circles and <unk>

character-based model: Did not run

The possible reason for that is the model would take the information of the sentence and generate a word closest to the meaning it tried to predict based on the information.