1 Character-based convolutional encoder for NMT

(a)

  Vocabulary size for characters is typically less than 100, which is 2 or even 3 magnitudes less than vocabulary size for works, that's why, meanwhile dense vectors of this 50 can hold much less information, than dense vectors of size 256, there is no need for capacity to be that big.

(b)

  Character-based embedding model size:

  $V_{char} * e_{char}$ parameters for embeddings.

  $e_{word} * e_{char} * k + e_{word}$ parameters for convolutional layer.

  $2 * e_{word} * e_{word} + 2 * e_{word}$ parameters for highway layer.

  In total: $V_{char} * e_{char} + e_{word} * e_{char} * k + e_{word} + 2 * e_{word} * e_{word} + 2 * e_{word}$.

  Word-based lookup embedding model size:

  In total: $V_{word} * e_{word}$

  Lets say that $e_{word} = 256$, $e_{char} = 50$, $k = 5$, $V_{word} = 50000$, $V_{char} = 96$ and compare models sizes:

  Character-based embedding model size $= 96 * 50 + 256 * 50 * 5 + 256 + 2 * 256 * 256 + 2 * 256 = 200640$

  Word-based lookup embedding model size $= 50000 * 256 = 12800000$

  So, word-based lookup embedding model has more parameters, in 67 times.

(c)

  As an output, convnet has an vector of dimension f, which is arbitrary and doesn't depend on e_{char}, by contrast, a RNN would has output of shape either e_{char} or $2 * e_{char}$, depending whether we use a bidirectional model or not, but the output size cannot be arbitrary. The convnet model, unlike a RNN, doesn't have such restriction.

(d)

  Max-pooling has an advantage of capturing the information whether or not there was part of the sentence that fires the neuron, by contrast, average-pooling looks at all words in the sentence, including articles and other words that doesn't have much meaning.

  Average-pooling has an advantage of capturing information the whole sentences, two completely different sentences can have same max-word and will be represented the same by max-pooling, by contrast, average-pooling will distinguish this sentences and contain more information about them.