1 Character-based convolutional encoder for NMT

Model description and written questions

1. Characters can only capture limited information. To represent the information of a word, we need to concatenate several character embeddings.

Character-based embedding model:

Vchar\*echar + f\*k\*e\_char + 2\*e\_word\*e\_word

= 96\*50 + 256\*5\*50 + 2\*256\*256

= 2 \* 10^5

Word-based lookup embedding model:

Vword\*eword

= 50000\*256

= 1.28 \* 10^7

The word embedding model’s parameters are hundred times of the character embedding model.

1. CNN is good at capturing long-distance dependencies, while RNN will pay more attention to later information in the propagation process.
2. Max-pooling is like attention mechanism. We focus on an important phrase and propagate it to the higher layers. It is usually used in natural language processing.

Average-pooling is used to retain the completeness of the information and reduce the dimensionality. Global Average Pooling is a good way to perform classification.