

Assignment 5: Solutions

Shubham Gupta

March 9, 2020

1 Character-based convolutional encoder for NMT

1.1 Character embedding

- Character embeddings are lower in dim because:
 - Typically they consist of characters plus punctuation symbol. This small set of symbols are able to represent almost all words in the english language
 - Using char embeddings we can obtain embeddings for OOV words as well
 - Good for representing new words, emoticons, etc.
 - Good for representing infrequent words which may be missing from the dataset

1.2 Model size

- Character embedding
 - $e_{char} = 50$
 - $V_{char} * e_{char} + e_{word} * k * e_{char} + 2 * (e_{word} * e_{word} + e_{word}$
- Word embedding
 - $V_{word} * e_{word}$

1.3 Advantage of CNN over RNN

- We know that RNN feed the output of sequence as the input to the next sequence. Bi-directional RNN learn from both forward and backward directions thereby creating contextual embeddings. HOWEVER, these embeddings can only take a window of words around it and they will compute the embeddings sequentially. CNN can learn different features using different kernels in parallel

1.4 Max vs Avg pooling

- Max pooling
 - Identifies the most important feature
 - Discards other info that could potentially be useful
- Avg pooling
 - Gives every word equal importance. Preserve data info
 - Incase there is one strong signal and multiple weak signals, it can dilute the important signal causing loss of information