

# Assignment 5

Marius Seritan

February 15, 2019

## 1 Character-based convolutional encoder for NMT

- (a) Explain one reason why the embedding size used for character-level embeddings is typically lower than that used for word embeddings

When using character-level embeddings the size of the vocabulary is smaller than the size of the word-level vocabulary and multiple character embeddings are combined together to build a word. As such the size of each embedding can be smaller.

- (b) Total number of parameters

For character-level embeddings the number of parameters is

$$N_{char} = V_{char} * e_{char} + e_{word} * e_{char} * k + e_{word} + 2 * (e_{word} * e_{word} + e_{word})$$

For word-level embeddings the number of parameters is

$$N_{word} = V_{word} * e_{word}$$

For  $k = 5$ ,  $V_{word} = 50000$  and  $V_{char} = 96$  and given that  $e_{word} \approx 10 * e_{char}$  the approximate number of parameters are

$$N_{char} = 96 * e_{char} + e_{word}^2 * e_{char} + e_{word}^2 \approx e_{word}^2 * e_{char}$$

$$N_{word} = 50000 * e_{word}$$

The ratio is

$$\frac{N_{word}}{N_{char}} = \frac{50000}{e_{word} * e_{char}}$$

For  $e_{word} = 500$  and  $e_{char} = 50$  there are **twice as many parameters in the word embeddings than in the character embeddings**.

- (c) One advantage of using a convolutional architecture rather than a recurrent architecture for the character based embedding models

When a 1d convnet computes features from character embeddings the features are given the same importance no matter where the original characters are placed in the word. This allows the convnet to focus on subwords/n-grams and learn the most relevant sections of the word. In contrast an RNNs always look at the whole word which makes it hard to discard the prefix, for example. Therefore for longer words convnets are more likely to extract the more relevant sub-word which could be specially for languages with joined words.

- (d) Compare max-pooling and average pooling

TODO

## **2 Character-based LSTM decoder for NMT**

BLEU score:

### **3 Analyzing NMT Systems**