

Convolutional Neural Networks for Sentence Classification



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Nils Reimers

Course-Website: www.deeplearning4nlp.com

Implementation

- Implementation is based on:
 - Kim, 2014, *Convolutional Neural Networks for Sentence Classification*
https://github.com/yoonkim/CNN_sentence
- This folder contains the following code:
 - preprocess.py: Transforms the train/dev/test.txt files to a format more suitable for Keras
 - cnn.py: Usage of the Keras functional API. Several convolutional with different filter lengths can be used

- Each token is mapped to its respective word index in the embedding matrix. This creates a matrix like:
 - [[4, 5, 8], #First sentence, 3 tokens with word indices 4, 5 & 8
[1, 3, 9, 1], #Second sentence
[8, 2, 2, 1, 4, 7]] #Third sentence
- In order to pass it to Keras/Theano/Tensorflow, all sentences must be padded to the same length. The function `sequence.pad_sequences()` uses zero padding. This creates a matrix like:
 - [[0, 0, 0, 4, 5, 8],
[0, 0, 1, 3, 9, 1],
[8, 2, 2, 1, 4, 7]]
- Ensure that first token in the embedding matrix is a padding token and `wordEmbeddings[0] = [0, 0, ..., 0]`

cnn.py – Functional API

- Check this guide to the functional API of Keras:
<https://keras.io/getting-started/functional-api-guide/>
- The functional API allows the definition of an arbitrary graph

