CIS, LMU München
Fach: Kolloquium Computerlinguistisches Arbeiten SS2017
Studentin: Faridis Alberteris Azar

Protokoll zur Sitzung am 22.05.2017

Titel: Comparing representation learning over word-level, character-level and combination of both in NLP tasks

Student: Iuliia Khobotova
Supervisor: Wenpeng Yin

The student Iuliia Khobotova began the presentation explaining the objective of her thesis:

- how different input styles influence the accuracy of convolutional neural network?
- what is the best set of parameter for CNN?
- how fast the accuracy will be calculated?
- what will change the combination of both word- and character-embeddings?

The input is represented by the commonly used word-embeddings, character-level NLP and a combination of both. The main task proposed in this thesis is to compare these representations in typically NLP tasks like sentiment analysis or POS tagging.

The meaning of CNN (Convolutional Neural Network) and RNN (Recurrent Neural Network), two main types of deep neural networks, were presented very briefly.

One of the experiments that are going to be developed in this thesis is the change of some parameters for the CNN, such as embedding size, hidden size and batch size, with the aim of searching for the set of parameter that achieves a high accuracy.

Data

She uses for her purposes the Stanford Sentiment Treebank, which contains annotated data from movie reviews and a lot of unique phrases.

Implementation

The experiment of this thesis is implemented in the Python Framework Theano. This framework is devided on three parts: input layer, hidden layer and output layer. While input and output layers stay the same, hidden layers varies in neural networks.

Evaluation and Results

The evaluation of the applied methods is based on comparison of accuracy. The results will be represented in a graphic.