

Protokoll zur Sitzung vom 22.05.2017 – Computerlinguistisches Arbeiten

2. Vortrag: Iuliia Khobotova, "Comparing representation learning over word-level, character-level and combination of both in NLP tasks"

BA-Betreuer: M. Sc. Wenpeng Yin

Den zweiten Vortrag des Tages hält Iuliia Khobotova, der von Herr Yin betreut wird. Ihre Presentation dauerte ziemlich kurz im Vergleich zu den anderen Vorträgen an dem Tag. Sie begann mit einem Überblick und stellte ihre Ideen für die Gliederung ihre Bachelorarbeit vor.

Am Anfang präsentierte sie das Objective ihrer Bachelorarbeit: Wie unterschiedliche Eingabestile die Genauigkeit von faszinierendem Neuronennetz beeinflussen? Was ist der beste Set von Parametern für CNN? Wie schnell wird die Genauigkeit berechnet? Was wird die Kombination von Wort- und Charakter-Einbindungen ändern?

Die Eingabe wird durch die häufig verwendeten Wort-Einbindungen, Zeichen-Ebene NLP und eine Kombination von beiden dargestellt.

Ihre Aufgabe besteht darin, diese Darstellungen in NLP-Aufgaben wie Sentimentanalyse oder POS-Tagging zu vergleichen. Die Bedeutung von CNN (Convolutional Neural Network) und RNN (Recurrent Neural Network), die zwei Haupttypen von tiefen neuronalen Netzwerken sind, wurden sehr kurz dargestellt.

Eines der Experimente, die in dieser Arbeit entwickelt werden, ist die Veränderung von einigen Parametern für die CNN, wie z. B. embedding size, hidden size and batch size, mit dem Ziel: Suchen nach dem Set von Parameter, der eine hohe Genauigkeit erreicht.

Sie verwendet für ihre Zwecke der Stanford Sentiment Treebank, die kommentierte Daten aus Movie Bewertungen und eine Menge von einzigartigen Phrasen enthält.

Das Experiment dieser Arbeit ist im Python Framework Theano implementiert. Dieses Framework besteht aus drei Teilen: input layer, hidden layer and output layer. Während die Eingangs- und Ausgangsschichten gleich bleiben, variiert die verborgenen Schichten in neuronalen Netzwerken.

Die Ergebnisse der angewandten Methoden basiert auf dem Vergleich der Genauigkeit. Die Ergebnisse werden in einer Grafik dargestellt.