

Protokoll 2 von 22.05.17

Ivana Daskalovska

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

Studentin: *Iulia Khobotova*

Betreuer: *Wenpeng Yin*

Frau Khobotova hat zuerst das Thema ihrer Bachelorarbeit und Ihrer Agenda vorgestellt. Das Ziel ihrer Bachelorarbeit ist es folgende Aufgaben zu untersuchen:

1. Wie beeinflussen unterschiedliche input styles die Accuracy von CNNs (Convolutional Neural Network)?
2. Was ist das beste Set von Parametern für CNN?
3. Wie schnell wird die Accuracy berechnet?
4. Welche Änderungen treten durch die Kombination von Word- und Character-Embeddings auf?

In ihrer Bachelorarbeit verwendet sie allgemeine Word-Embeddings, Character-Level-Embeddings und Kombination von beiden. Ihrer Aufgabe besteht darin, diese in NLP- Tasks wie Sentiment Klassifikation und POS-Tagging einzusetzen und zu vergleichen.

Frau Khobotova hat danach anhand eines Beispiels die zwei Haupttypen von tiefen neuronalen Netzen CNN(Convolutional Neural Network) und RNN(Recurent Neural Network) gezeigt. Beide werden in verschiedenen NLP-Tasks verwendet. Sie konzentriert sich in ihrer Arbeit auf CNNs.

In einem Experiment versucht sie CNN-Parameter (embedding size, hidden size und batch size) zu ändern. Ihr Ziel ist es das beste Set von Parameter zu finden, bei dem die maximale Accuracy erzielt wird.

Sie verwendet als Daten die Stanford Sentiment Treebank. Das ist eine Sammlung von annotierten Daten von Filmkritiken, die circa 215k einzigartige Phrasen enthält. Die Implementierung wird durch das Python Framework Theano realisiert.

Evaluation und Ergebnisse:

Die Evaluation basiert auf Accuracy-Vergleich. Die Ergebnisse werden graphisch repräsentiert.

