

**Protokol 2: Iuliia Khobotova „Comparing representation learning over word-level, character-level and combination of both in NLP tasks“.**  
**Betreuer: Wenpeng Yin**

In dieser Arbeit geht es darum, unterschiedliche Inputs den Neuronalen Netzen zu gehen und zu beobachten, welche Auswirkungen das auf die Accuracy des Systems haben kann, wie schnell wird die berechnet, welche Parameterset am besten ist und was wird die Änderungen bei Wort- und Zeichenembeddings auslösen.

Die Aufgabe ist diese Prozesse für solche NLP Tasks wie sentiment classification and POS tagging zu beobachten.

Dafür wurden zwei neuronale Netzwerke benutzt: CNN (Convolutional Neural Network) und RNN (Recurrent Neural Network). Der größte Unterschied liegt daran, dass CNN hierarchisch aufgebaut ist, wobei RNN's sequenzielle Architektur besitzt.

Folgende Parameteränderungen wurden ausprobiert: Grösse von Embedding, Anzahl von Trainingsbeispielen und Größe des Hidden layers .

Die Daten, die benutzt werden kommen aus Stanford Sentiment Treebank (annotierte Filmreviews, die aus circa 215 Tausend Phrasen besteht) und Python Framework Theano (deep learning platform), das aus input layer, hidden layer and output layer besteht.

Die Evaluierung wird statistisch durchgeführt, auf Basis von Accuracy und wird graphisch dargestellt.