

Project abstract:

Jarkko Tulensalo (jarkko.tulensalo@aalto.fi)

Julius Hietala (julius.hietala@aalto.fi)

We intend to experiment with sentiment analysis (sentence classification) using different word embeddings and the Sentiment140 dataset containing labeled Twitter tweets. More specifically, we intend to train our own word embeddings using word2vec and then use those with a deep learning sentiment analysis model (TBD: LSTM/GRU/Transformer) that we will create using Pytorch. The main experiments we will perform are

- Use Pytorch's trainable embeddings with random initialization
- Use Pytorch's trainable embeddings with our trained word2vec initialization
- Only use our trained word2vec embeddings as inputs

and then compare the accuracy results for the sentiment analysis task.