CIL 2018: Text Sentiment Classification

Pirmin Schmid, Philip Junker, Nikolas Göbel, Aryaman Fasciati
The Optimists
Department of Computer Science, ETH Zurich, Switzerland

***Abstract*—Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.**

I. Introduction

The goal of this project is to build a sentiment classifier that predicts whether a tweet text used to include a positive smiley :) or a negative smiley :(, based on the remaining text.

Our first baseline uses random forests and achieved an accuracy of TODO: %. Our second baseline uses a Recurrent Neural Network (RNN) model with an accuracy of TODO%.

In a third model, we refined our second baseline, incorporating TODO: describe novel approach, in a RNN-based approach. This model achieved TODO: % accuracy.

II. Related Work

Write about related work [1]

III. Models

A. First Baseline (B1)

- We use pretrained GloVe [2] embeddings from Stanford. - We use the preprocessor script provided by Stanford (slightly adapted) - Words that are not in the vocabulary are ignored

Baseline1 - A tweet is represented as the mean of all of its word embedding vectors (to solve the problem that the word embedding can only be caluclated per word and the classifier can only classify one vector) - Random Forest classifier with unlimited max_depth and 20 estimators - achieved accuracy (online): 0.72

B. Second Baseline (B2)

Acknowledgements

The authors wish to express their gratitude to the Euler and Leonhard clusters at ETH, whose unwavering computational effort this project could not have done without.

References

[1] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[2] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation." in *EMNLP*, vol. 14, 2014, pp. 1532–43.