Rumour detection and analysis of Tweets

COMP90042 - Natural Language Processing Project Report

Abstract

This document contains the report for the COMP90042 Natural Language Processing project. The following report explores the topic of automatic rumour detection in Twitter, it is also accompanied by an analysis section. The automatic rumour detection section explores multiple state of the art techniques for text classification, one model explored is a simpler model which does not take the tree structure into consideration whereas the more complex one does.

1 Introduction

Automatic Rumour detection is a difficult topic.

2 Strategy

The strategy used to obtain the best model was simply to do some quick, lost cost experiments to establish a baseline for accuracy. There were 4 models that were explored briefly at this stage of the project. The four models explored were a simple feed forward neural network, a convolutional neural network, a random forests classifier and a naive bayes classifier. The text preprocessing done was handled by Spacy alone. In addition to these 4 models, I also used FastText, Facebook's text classification system to quickly determine the best accuracy I would be able to obtain. The accuracy scores for these simple models are listed below.

Table 1: Accuracies for some ML models

Accuracies				
ANN	CNN	RF	NB	FT
FAIL	%67.67	%70.08	-	%80.31

Please note that the simple feed forward neural network was not able to be trained due to a OOM exception, even when using a 8 node GPU cluster, it was clearly evident that the model needed to be simplified and this lead to convolutional neural networks being explored.

References

Ipek Baris, Lukas Schmelzeisen, and Steffen Staab. 2019. Clearumor at semeval-2019 task 7: Convolving elmo against rumors. Proceedings of the 13th International Workshop on Semantic Evaluation.

Jodie Burchell. Using vader to handle sentiment analysis with social media text.

Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson F. Liu, Matthew Peters, Michael Schmitz, and Luke S. Zettlemoyer. 2017. Allennlp: A deep semantic natural language processing platform.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Rumor detection on twitter with tree-structured recursive neural networks. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).