Fake News Detection

Project Outline

Author - Ioan Rowlands (ior2)
Supervisor - Richard Jensen (rkj)

Module - CS3944
GG4R - Computer Science and Artificial Intelligence

9th February 2022
Version 0.1
(Draft)

Project Description

The project of Fake news Detection will use text preprocessing methods and machine Learning techniques to identify if an article is ‘fake’ or ‘true’. I will explore multiple methods for pre-processing the data from the selected dataset. And the impact of these methods on the Classifiers.

Fake news is a large issue ranging from spreading misinformation to identifying what it is as just because you do not agree with it it is not ‘fake’. Deciding what is ‘fake’ or ‘real’ is outside of the scope of this project. However I will discuss in my final report the ethics behind classifying text. This project will however show how it is possible to train a classifier for text input where the dataset used to train it could be modified.

The main substance of this project will consist of pre-procesing the data, the classification algorithms used on this data and analysis of these results and tuning algorithm parameters to test the classifier.

The pre-processing will compare two methods TF-IDF which weights the use of words in the document and word embedding which accounts for how words may be related.

The classification algorithms used will be supported with evidence as it stands currently I am looking at using Naive Bayes, a Convolution Neural Network and an Artificail Neural Netwok. Depending on the work produced in the pre-processing step there may be scope for researching Ensemble methods which combine results of several models.

This project could be useful for anyone reading articles in order to identify for them if the text should be trusted.

Proposed Tasks

Project management and blogs - This will be the the way the project is run and I am preferring a Plan based approach due to the work being broke up into stages like research, implementation and then analysis. The blogs allow me to track my progress of the week which can be used later in the project in the report.

Set up version control system - I have set up a GitHub repository for this work this will be added to continuously throughout with any development and documentation.

Pre-Processing of data -

Investigate datasets which are suitable for use in order to classify ‘fake’ and ‘real’ news - This task will include looking at datasets used by other researchers and orgnisations in order to detect fake news. A good dataset will improve the real world succes of the algorithm. And the dataset will influence what the classifier believes in ‘fake’ or ‘real’.

Research reading into suitable Machine Learning algorithms - These algorithms chosen should be good performers on the pre-processed data. The algorithms I am currently considering are Naive Bayes, CNN's and ANN's. Look into Ensemble classifier. Also what language and libraries to use for this.

Optimizing learners - This is where I could adjust the parameters in order to improve the performance of the learner.

Analysis of Learners and the Pre-processing - Looking at

Project hand in - Mid way demonstration, final report and demonstration.

Project Deliverables

Mid project report

Reviws of Dataset chosen, pre-processing methods and ML algorithms

Working pre-processing two options to compare

Working algorithms

Algorithm results and comparison

Final report

Final Demonstration

Initial Annotated Bibliography