

Project Brief

Student's Name: Maciej Szpakowski (ms3u14 - 28170911)

Project Title: Fake News Recognition

Supervisor's Name: Dr. Jonathon Hare

Problem

Nowadays, with the increased ease of online communication and social media, people have immediate access to many different types of information. Some of this information comes from reliable portals. However, there's plenty of news coming from gossipy magazines and click baits that share scientifically dubious claims, hatred and biased speech. The problem becomes even more prominent as people often share articles that have controversial titles, increasing their popularity. Unfortunately, with more sources appearing on the internet, manual validation is becoming impossible.

Goals

The goal of this project is to explore the issue of fake news and create a deep learning based algorithm for automatic labelling of texts that are potentially deceptive. As this is a far-reaching problem that affects many areas and different types of media, the first step is to narrow it down based on available data and previous research.

The second step is to, using this data, create/choose the most efficient algorithm that will be able to determine the veracity of the news. Some of the possible approaches include context-based (e.g., using the links network), knowledge-based (e.g., using information from Wikipedia), style-based (e.g., using text classification) or an amalgamation of those.

Additionally, if time allows, the last step is to create a simple browser extension, which will automatically detect fake news and inform the user about it (e.g., by showing an alert on the feed or actual news page). It will also allow for gathering feedback from users to assess the real performance of created approach.

Scope

One of the most significant limitations of this work is the data that is openly available. The focus of this work is on using novel deep learning based approaches for natural language processing. Some of them (like LSTMs based on word embeddings) usually require large quantities of data. Therefore additional website crawling may be necessary.

Finally, the term "fake news" sometimes is associated with political views some people do not agree with. While we will do our best to make sure the datasets used for training contain actual information and are accurately classified, the purpose of this work is not to fact check all the input data to make sure initial labels are 100% correct.