A Proposed Framework for Acessing Bias on English Newspapers

Curado, Antonio & Dahl, Morten

Masters in Advanced Analytics @ Nova IMS

Abstract

This project contributes to the detection of fake news by analysing bias in english newspapers. These Articles are grouped by topic and been undertaken a sentiment analysis to detect the bias and tendencies within each article. It results in a visualization, which shows the media bias per newspaper, topic and keyword.

s

Motivation 💪

The motivation from this project lays in the recent doubts on media neutrality. People tend to distrust the media and call serious newspapers 'fake news', whereas dubious newspapers gain popularity. So this project aims to provide the following:

- provide a measure for biasness and tendentious media coverage
- create transparency which newspapers tend to publish more tendentious articles
- initial notions towards an automated biasness detection in media coverage

s

Methods

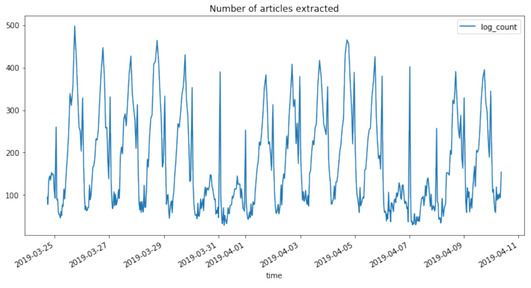
The methodology followed in this project is as followed: As a first step data was aquiered from a **web scrapping** aproach from which news articles are downloaded from the web representations of various well-known newspapers. The data was limited to the politics section of each newspaper. It was tried to pick a balanced selection of different politically located newspapers through a qualitative assessment. These articles were then filtered on keywords [e.g. "Trump", "Syria", "Brexit"] and been used to train a **LDA** model for topic detection. Simultaniously a **sentiment analysis** was carried out to measure the tendentious nature of each article. "A sentence on how it was done" Finally the results of both analysis were mapped and visualised in an interactive scatterplot.

s

Data Extraction

As a way to make the analysis relevant and up to date with the most current news topics it has been developed a new news dataset with the following porperties

- Built a dataset with over **70.000** news articles
- Scraped over **19** newspapers for over **2** weeks
- With an average of **3.600** news articles per newspaper



The dataset was build only using the newspapers3k python package

s

LDA Analysis

Nam quis odio enim, in molestie libero. Vivamus cursus mi at nulla elementum sollicitudin. Nam quis odio enim, in molestie libero. Vivamus cursus mi at nulla elementum sollicitudin.

$$E = mc^2 \tag{1}$$

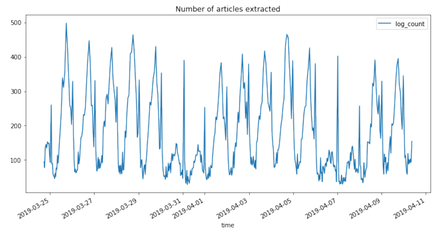
s

Sentiment Analysis

Another way to search for is to analyse the sentiment which each newspaper looks at a topic. With a help of a word dictionary quantifiyng the polatiy of each word, the polarity for each article has been calculated.[**?**]

s

Trump

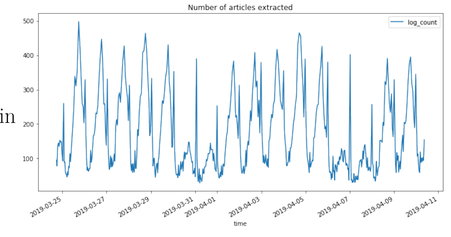


Fusce quis massa dictum tortor **tincidunt mattis**. Donec quam est, lobortis quis pretium at, laoreet scelerisque lacus. Nam quis odio enim, in molestie libero. Vivamus cursus mi at *nulla elementum sollicitudin.*

s

Brexit windows/1F1E7.png

Fusce quis massa dictum tortor **tincidunt mattis**. Donec quam est, lobortis quis pretium at, laoreet scelerisque lacus. Nam quis odio enim, in molestie libero. Vivamus cursus mi at *nulla elementum sollicitudin.*



s

Syria

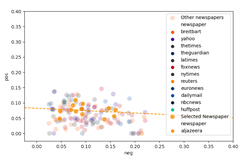


Figure: Al Jazeera

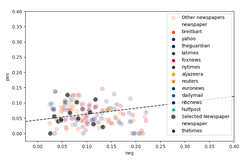


Figure: The Times

The two plots show how articles from the Al Jazeera and The Times newspapers compare to each other, on a positivity/negativity scale. Both are measured on the topic 5 which contains as most frequent keywords: From this one could possibly infer that when Al Jazeera publishes on the topic, it takes more serious and possible more negative approach to it. Never the less, these are only possible assumptions that the framework helps to clarify. Each article would be needed to be read in details to derive further conclusions

s

Conclusion 🙏

Measuring bias is a very subjective task. Usually computers have a hard time executing subjective tasks. The framework developed does not aim to automate bias detection but rather to empower humans with summarization capabilites otherwise not possible. From the three example analysis above shown, it is possible to validate the usefullness of the framework proposed on a hard topic as bias detection. As news generation gets automated, as is the case of fakenews(needs reference), the methods of flaging them also require modern digital frameworks. This analysis and framework takes another step in this direction.

s

Future Work 🙏

As previously mentioned this project aims to set a staring framework for discusion on automated detection of bias. As so there is much further work to be done, mainly on 5 different areas:

- Increase the number of newspapers and find a method to select the same number of articles for each one
- Extrapolate the work to different regions and languages
- Automate topic detection
- Build a ML/Rule based algorithm to flag possible high bias on articles
- Apply sliding window mechanimns to track changes in sentiment towards selected topics

s

Hutto, C.J. and Gilbert, Eric, *A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*, Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014, 2015.

The project was possible with the great work done in the newspaper3k, nltk and gensim python libraries