

Computer Science Tripos – Part II – Project Proposal

# Sentiment Analysis of British Newspaper Articles on the Iraq War

Louis Slater, Pembroke College

Originator: Louis Slater

12th October 2017

**Project Supervisor:** Dr Tamara Polajnar

**Director of Studies:** Dr Anil Madhavapeddy

**Project Overseers:** Dr Timothy Griffin & Professor Anuj Dawar

## Introduction

While there have been lots of studies involving sentiment analysis of political texts to determine their bias, none of these have uniquely involved British newspaper articles. Furthermore, after extensive research, I have not found any sentiment analysis of articles to determine their stance on a war. The purpose of this project is to develop a program that can reasonably determine the stance of any British newspaper article on the Iraq war. The core part of this project will be developing a program that achieves this using a bag-of-words method.

## Starting point

In the past few years, there has been a lot of research into determining political biases of shorter segments of text, such as Tweets in the 2010 paper by Pak and Paroubek on Twitter as a Corpus for Sentiment Analysis and Opinion Mining. On the other hand, Political Ideology Detection Using Recursive Neural Networks by Iyyer, Enns, Boyd-Graber and Resnik uses a corpus containing longer texts US Congressional floor debate transcripts. Although there are clear differences between these transcripts and the newspaper articles that I plan to use (for example, the fact that the transcripts were initially spoken, whereas the articles were not), there are also many similarities (for example the length and inherently political nature of the corpora). Since this study showed that a bag-of-words method can successfully determine the bias of these transcripts with a 65% accuracy, it is justified to use a similar model to determine the bias of the newspaper articles I shall analyse.

The corpus I will use will be articles on the Iraq war from up to seven of the UKs most popular national daily newspapers and their Sunday equivalents published between 16th March 2003 and 18th April 2003. I will use a database of these articles compiled by Robinson, Goddard, Brown and Taylor in their 2003 study, Content and Framing Study

of United Kingdom Media Coverage of the Iraq War, in which they manually determine the stance of 4,893 news articles from seven British daily newspapers and their Sunday equivalents (Daily Telegraph, The Times, The Guardian/The Observer, The Independent, The Daily Mail, The Mirror, The Sun/News of the World). This database does not include the articles texts, so the first part of my implementation will be to scrape this data from as many of these articles as possible. I will be able to get these texts from existing online Newspaper archives. I have already found searchable archives for The Guardian, The Observer, The Daily Telegraph, The Sunday Telegraph, The Independent, Indy on Sunday, The Times and the Sunday Times, all of which I will be able to use. Scraping textual data from the other newspapers in the database may be prove more difficult, but I will as many possibilities as I feasibly can within the scope of the project.

## **Resources required**

In addition to the database and archives mentioned above, I will also require the use of a computer. I intend to mainly use my own computer, which has an Intel Core i7 processor and runs Windows 10. I will use the computing facilities in my college if my laptop is lost, broken or stolen. I will back up my work using both Google Drive and GitHub, which I will also use as a version control repository. I may also require the use of a server or external hard drive to store the corpus I use; however, this will be dependent on the amount of data that I scrape in the initial part of my project.

## **Work to be done**

The project breaks down into the following sub-projects:

1. Gaining access to as many of the relevant searchable newspaper archives as possible.
2. Scraping data from as many articles as possible in the database compiled by Robinson, Goddard, Brown and Taylor.
3. Implementing a program to determine the biases of texts on the Iraq war, using the corpus I gather, along with corresponding the Reporters Tones from the database compiled by Robinson, Goddard, Brown and Taylor.
4. Running the program on the texts and comparing the results with the manually determined biases to judge the effectiveness of the program.

## **Success criteria**

The project will be a success if I develop a program that can determine the stance of an article on the Iraq war with a greater than 50% accuracy.

## Possible extensions

If I meet my success criteria early, I shall attempt one, or both, of the following extensions:

- Implementing a program that performs the same function as the initial program I develop, but using a different method, such as a recursive neural network. If I complete this extension, I will be able to compare the effectiveness of the two methods.
- Extrapolating the results using new datasets and analysing these results. Possible datasets I could use are newspaper articles from different countries, publications or times or transcripts of parliamentary debates.

## Timetable

Planned starting date is the beginning of Michaelmas Week 3 (Thursday 19th October 2017).

1. **Michaelmas week 3** Gain access to as many of the relevant searchable newspaper archives as possible.
2. **Michaelmas weeks 4–5** Scrape data from as many articles as possible in the database compiled by Robinson, Goddard, Brown and Taylor, creating a database of the texts, their manually determined bias and other relevant information on them. If necessary, I will also get access to a server and store the database I compile on this server.
3. **Michaelmas weeks 6–8** Implement a program to determine the biases of texts on the Iraq war, using the corpus I gather, along with corresponding the Reporters Tones from the database compiled by Robinson, Goddard, Brown and Taylor.
4. **Michaelmas vacation** Finish the implementation, then run the program on the texts and compare the results with the manually determined biases to judge the effectiveness of the program.
5. **Lent weeks 1–2** Write the progress report and start work on possible extensions of the project.
6. **Lent weeks 3–4** Finish the extensions to the project.
7. **Lent weeks 5–6** Write the first draft of the dissertation.
8. **Lent weeks 7–8** Revise the dissertation in accordance with feedback I receive from my supervisor.
9. **Easter vacation** Finish revising the dissertation and submit the final project.