17076749

Mark Baber

Sentiment Analysis

Individual Project

Table of Contents

[Chapter 1 – Introduction (100 Words?) 2](#_Toc23344980)

[1.1 – Background (100 words) 2](#_Toc23344981)

[1.2 - Justification (350 words) 2](#_Toc23344982)

[1.3 - Aims and objectives (100 words) 2](#_Toc23344983)

[1.4 – Conclusion (100 words) 3](#_Toc23344984)

[Deliverables? 3](#_Toc23344985)

[Chapter 2 – Lit Review 4](#_Toc23344986)

[2.3 – Current Software (IBM & Google) 5](#_Toc23344987)

[Chapter 3 – Design 6](#_Toc23344988)

[3.1 – Use outputs 6](#_Toc23344989)

[Chapter 4 – Implementation (Prototype/Deliverable) 7](#_Toc23344990)

[4.1 – Based on design 7](#_Toc23344991)

[Chapter 5 – Evaluation of prototype/deliverable 8](#_Toc23344992)

[Chapter 6 - Conclusions 9](#_Toc23344993)

[LSEPI – (Put in appendix in final) 10](#_Toc23344994)

[Legal 10](#_Toc23344995)

[Social 10](#_Toc23344996)

[Ethical 10](#_Toc23344997)

[Professional 10](#_Toc23344998)

[Issues 10](#_Toc23344999)

[References 12](#_Toc23345000)

# Chapter 1 – Introduction (100 Words?)

To determine whether there is a link between the sentiment of a TV show and the viewer ratings. This project will explore areas of Sentiment Analysis (SA), different ways to perform sentiment analysis, why is this relevant and how will others benefit from this. This will involve some data analysis and manipulation to find out if there are any correlations. This chapter explores the background of the research as well as a justification for it. The aims and objectives are also considered.

## 1.1 – Background (100 words)

WHAT?

Sentiment analysis has been defined as opinion mining (Ding, et al. 2008) and according to Feldman (2013), sentiment analysis is used to look at the “decision-making process of people”. The value of this is we can better understand them as consumers, voters, reviewers etc.

## 1.2 - Justification (350 words)

(why? Practitioner and academic value)

Feldman (2013) states that by using sentiment analysis it “offers these organizations the ability to monitor the different social media sites in real time and act accordingly”. This would give companies a much better understanding of their customers and can benefit from this.

## 1.3 - Aims and objectives (100 words)

The aim of this project is to create a script which will look at the sentiment of an episode of a TV show and the viewer ratings and see if there is a link between them.

* To perform a literature review of sentiment analysis.
* To investigate the sentiment of a TV show, per episode/season.
* To investigate the viewer rating of a TV show from reviewer websites.
* To investigate if there is a link between both results.

## 1.4 – Conclusion (100 words)

## Deliverables?

* Dissertation
* R-Script which will be created to do the sentiment analysis.
* User manual which will be created to guide other people through the script.

# Chapter 2 – Lit Review

According to Pang & Lee (2008) sentiment analysis has also been referring to it as ‘brand monitoring,’ ‘buzz monitoring’ and ‘online anthropology,’ to ‘market influence analytics,’ ‘conversation mining’ and ‘online consumer intelligence’.

## 2.1 - What does it do?

Sentiment analysis is a method of analysis which looks at the emotion of a word with the positivity and negativity of the said word. This style of analysis is used in marketing to measure the reviews of a service or product with the product reviews which is also what Taboada, et al (2011) states.

## 2.2 - Types of sentiment analysis

There are multiple types of sentiment analysis, which looks at different types of entities within a data set. These different types are called: Document-level sentiment analysis, Sentence-level sentiment analysis, Aspect-based sentiment analysis, Comparative sentiment analysis and Sentiment lexicon acquisition Feldman (2013).

### 2.2.1 – Document-level sentiment analysis

The first type of sentiment analysis which will be explored is Document-level. This type of sentiment analysis is known as the simplest form of as it looks at the whole document as one attribute (Feldman 2013). For an example of this, we could look at different types of reviews from Amazon and would give you an overall rating. This type can also be done with machine learning which consists of supervised and unsupervised learning. The main difference between them is one has an estimate of the output such as positive or negative and the other doesn’t.

As Bibi (2017) pointed out, these include “Naive Bayes, Maximum Entropy classification and Support Vector Machines (SVM).”

* Naive Bayes – has real time prediction, is very fast algorithm.
* Maximum Entropy Classification
* Support Vector Machines

Advantages -

Disadvantages -

### 2.2.2 - Sentence-level sentiment analysis;

The second type of sentiment analysis is Sentence-level. This type looks at each sentence as an individual entity, so will break down each sentence into an ‘opinion’. Looking at the emotion of each sentence and will show the overall sentiment at the end.

* subjectivity classification
* sentiment classification

### 2.2.3 - Aspect-based sentiment analysis

Known as feature-based sentiment & Used to identify sentiment of many attributes.

### 2.2.4 - Comparative sentiment analysis;

Comparing a product to another & Looks for words like, (More, less, lighter)

### 2.2.5 - Sentiment lexicon acquisition

Lexicon is most common, Uses Dictionaries, Can be hand coded or crowd sources (BagOfWords Method).

“This approach allows the system to handle opinion words that are context dependent, which cause major difficulties for existing algorithms” Ding, et al. (2008).

## 2.3 – Current Software (IBM & Google)

# Chapter 3 – Design

## 3.1 – Use outputs

# Chapter 4 – Implementation (Prototype/Deliverable)

## 4.1 – Based on design

# Chapter 5 – Evaluation of prototype/deliverable

# Chapter 6 - Conclusions

# LSEPI – (Put in appendix in final)

## Legal

Software licenses – The aim of this project is to use open source tools which will give anyone the ability to follow along with this project.

* R - An open source front end for the programming language R, which is great for creating and manipulating scripts and data frames.
* R-Studio -
* Git - is an open source version-control system for keeping track of changes in code.
* GitHub is a website for developers to upload their code externally, which was built on Git and allows for collaboration. (Microsoft)
* Visual Studio Code - An open source text editor for developers.

## Social

Anonymity of user input – With the use of web scrapping for this project, this could be a difficult issue for some websites. For the types of websites this project will focus on, have a developer version which allows users to download and analyse their data sets. This has been confirmed in the Terms Of Service (TOS) and usually requires the user to create an account and tell the company why/what you’ll be doing with the data.

## Ethical

To lay out the rules for ethics, we would have to take into account of how personal it can be from person to person. Whilst it can be interchangeable usually the work place would follow some general ethical concerns.

* To treat people fairly
* To respect the autonomy of individuals
* To act with integrity
* To seek the best results

## Professional

Five characteristics of a professional!?

A professional should be someone who has the following:

* A duty of care
* A responsibility for consequences

This also falls under the BCS Code of Conduct which is the body of British Computer Society, who sets out the professional standards required as part of their members.

Within the BSC Code Of Conduct they state that all members, no matter their rank or length of being a member will provide public interest.

(Look at BSC Code Of Conduct)

## Issues

# 

# References

Bibi, M. (2017) ‘*Sentiment Analysis at Document Level*’. Available at: <https://www.researchgate.net/publication/320729882_Sentiment_Analysis_at_Document_Level> (Accessed 30/10/2019)

Ding, X., Liu, B., & Yu, P, S. (2008) ‘A Holistic Lexicon-Based Approach to Opinion Mining’. Proceedings of the 2008 International Conference on Web Search and Data Mining., Pages 231-240, Palo Alto, California, USA. February 11 - 12, 2008.

Feldman, R. (2013) ‘Techniques and Applications for Sentiment Analysis’. *Communications of the ACM*, vol. 56, no. 4.

Lima, A, C, E, S., & de Castro, L, N. (2012) ‘Automatic Sentiment Analysis of Twitter Messages’. Fourth International Conference on Computational Aspects of Social Networks (CASoN). November 2012.

Nair, V. (2017) ‘*The rise of big data*’. Available at: <https://www.bcs.org/content-hub/the-rise-of-big-data/> (Accessed 28/10/2019)

Pang, B. & Lee, L. (2008) ‘Opinion mining and sentiment analysis’. *Foundations and Trends in Information Retrieval*, 2(1-2), pp. 1-135.

Silge, J. & Robinson, D. (2017) *Text Mining with R*. 1st ed. O'Reilly Media.

Taboada, M. et al. (2011) ‘Lexicon-Based Methods for Sentiment Analysis’. *Computational Linguistics*, 37(2), pp. 267-307.

<http://ataspinar.com/2016/01/21/sentiment-analysis-with-bag-of-words/>

https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d