Aspect-Based Sentiment Analysis

MSc Research Project
Data Analytics

Mayank Gulaty
x15031705

School of Computing
National College of Ireland

Supervisor: Mr. Vikas Sahni
Industry Supervisor: Mr. Dermot Mackey

National College of Ireland
Project Submission Sheet – 2015/2016
School of Computing

Student Name:	Mayank Gulaty
Student ID:	x15031705
Programme:	Data Analytics
Year:	2016
Module:	MSc Research Project
Lecturer:	Mr. Vikas Sahni
Submission Due Date:	21/12/2016
Project Title:	Aspect-Based Sentiment Analysis
Word Count:	4208

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	
Date:	21st December 2016

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
3. Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Aspect-Based Sentiment Analysis

Mayank Gulaty
x15031705
MSc Research Project in Data Analytics

21st December 2016

Abstract

In this fast paced and social media frenzy world, decision making has been revolutionized as there are lots of opinions floating on the internet in the form of blogs, social media updates, forums etc. This paper focuses on how banks can utilize these reviews to improve their services. The reviews are scraped from the internet and the sentiment analysis is done on an aspect level to have a much clear idea of where the banks are performing bad according to the customers or users.

1 Introduction

With the advent of Social media and the freedom of people by which they can put anything on it, Internet is a place consisting of myriad of sources in which people can post their reviews or opinion on anything. Companies dont need to go anywhere to find the reviews of people, its just one click away. People using the Internet are increasing everyday and are invited to share the opinions by giving promotions and discounts to them if they do so. This shows how important it is for the companies to know their opinions. Due to this, Internet has been over-populated with the reviews of products and services of all kind. But these reviews are not useful if the companies don't use them to improve their services. The problem is that, it is not possible to read all the reviews and therefore there is a need for a system to segregate the reviews under different categories so that it becomes easy to scan through thousands of reviews and still get an insight.

Sentiment analysis or opinion mining is a process to identify the polarity of the opinion by applying NLP and text analysis. This is done by filtering the sentences that do not contribute to the polarity and then extracting the subjective information within the remaining text. It is a field of computational linguistics which has been getting a lot of attention in the past few years both in industry and academic side majorly due to the active increase in the social media engagement of the users. Chen and Zimbra (2010)

We can see from the picture below how the popularity of sentiment analysis has boosted up over the past 10 years. Figure1

The aim of this project is to first classify the review under a main category and then do the sentiment analysis of the aspects under that category. The task involves 3 main subtasks -

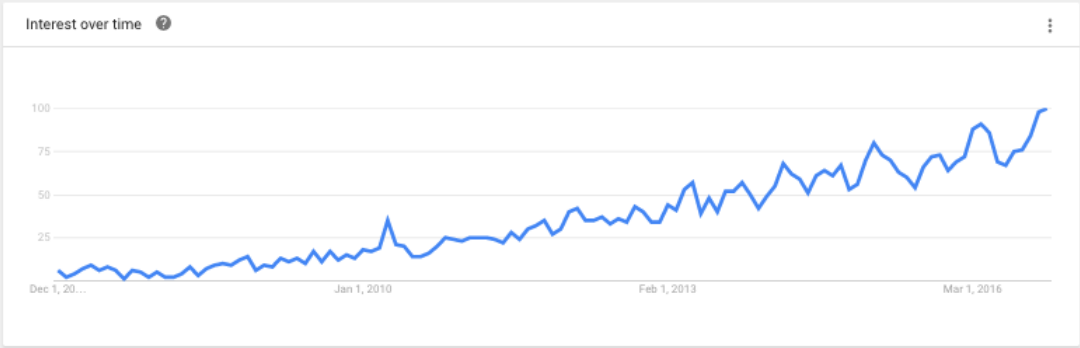

Figure 1: Sentiment Analysis trend over the last 10 years "Data source: Google Trends (www.google.com/trends)."

1.) Identify the main category of the reviews.
2.) Extract the aspects of that category.
3.) Calculate the sentiment of that aspects.

Hypothesis:

Let H be the hypothesis and C is a set of identified categories, I is the input set of reviews and O is the output.

C=Bank Account/Service, Consumer Loan, Credit Card, Credit Reporting, Money Transfers, Mortgage, Payday loan, Prepaid Card, Student Loan

A= Aspects

S=Positive, Negative, Neutral

H_0: $I \Rightarrow$O where $O \in$ {C,A,S}

2 Related Work

Sentiment Analysis can broadly be defined into 3 categories (Collomb et al.; 2007)-

1. Document Level
2. Sentence Level
3. Aspect Level

In document level, we see the document as a whole if it is positive or negative but there is a downside to it as we don't know what parts are positive or what parts are negative. In sentence and aspect level, opinions are extracted at a finer level to see which aspects user like and which ones he/she doesn't.

2.1 What are Opinions?

Opinions can be expressed as positive, negative or neutral and every category has a strength associated with it which tells us how positive or how negative the feeling of the writer is. It basically contains of 2 parts Sentiment and target where this target is the aspect or entity that we are interested in and sentiment is the polarity towards this target. The other parts of opinions that are important are the opinion holders who post the opinion and the opinion date on which the opinion was posted. These 2 parts are important to get the trends of the opinions over time or in a particular region of the world.(Liu; 2012) According to (Jindal and Liu; 2006), there are 2 main types of opinions:

1. Regular Opinions
2. Comparative opinions

Regular opinions - These are the type of opinions where people express their opinions about the products feature directly.

Comparative opinions - These types of opinions are more complex than the regular ones as there are 2 or more entities involved in the opinion and there is usually the comparison between these entities.

2.2 What is an aspect?

As mentioned above, sentiment analysis reveals about the polarity of the text but doesn't give any more information. E.g. if a person writes a bank review as "the customer service sucks and the website is horrible", the polarity of the text is negative but in aspect level we can see what are the things the customer doesn't like - customer service and website in this case. So these two are the aspects. And this becomes very important in the second type of opinion i.e. comparative opinion because there are entities involved and to make connections between them, sentiment analysis on a document level would be ineffective and wont show the real picture. This is because some person can compare a good bank with a bad one. So even if the targeted bank is praised, the overall sentiment would be mixed. However, aspect level on the other hand will give a real picture.

There have been many works done in the field of sentiment analysis on different levels. Some researches are focused on aspect detection, some on sentiment analysis and some are focused on both aspect detection and opinion analysis. Every method has its pros and cons. The approaches are shown in the diagram below.Figure3

There hasn't been much work done on aspect detection in banking domain as aspect based sentiment analysis has mainly been focused on restaurant and consumer product reviews.

2.3 What is an ontology?

Ontology is needed for the several reasons like -

1. Analyzing the domain knowledge and separate it from the operational knowledge
2. Reuse that knowledge (Noy et al.; 2001)

According to (Guarino; 1998), there are 4 types of ontologies

1. Top Level - This describes very general concepts.

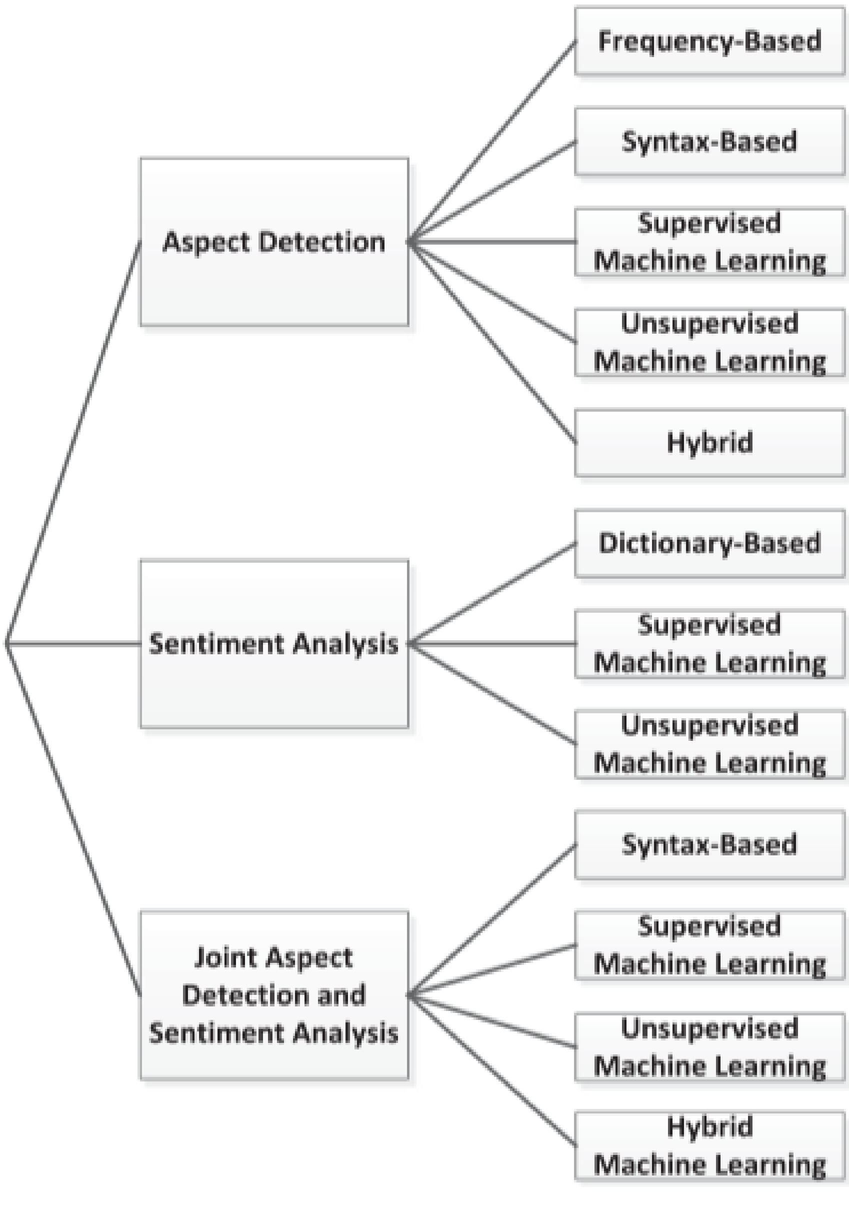

Figure 2: Current Techniques
(Schouten and Frasincar; 2016)

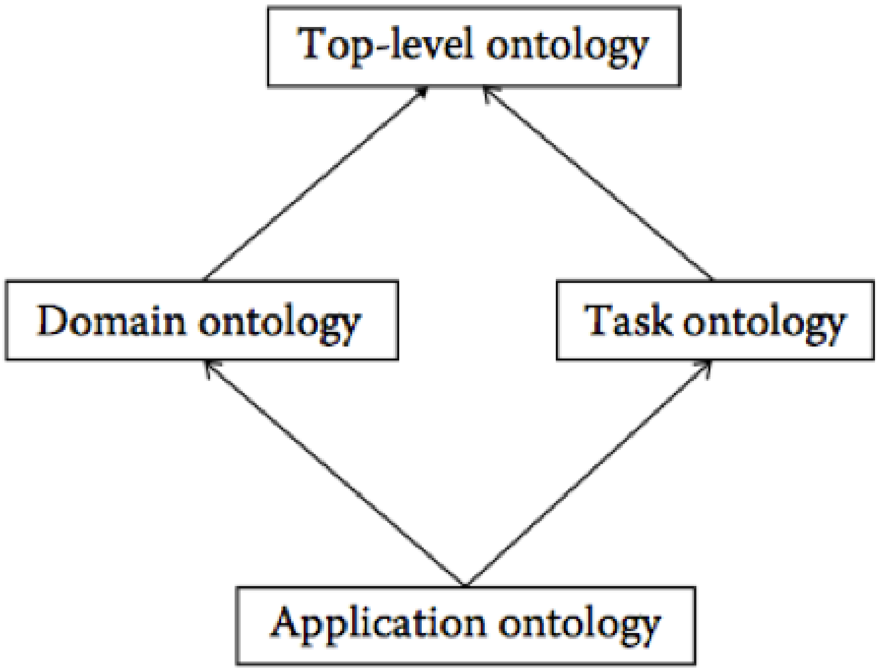

Figure 3: Types of Ontologies (Guarino; 1998)

2. Domain - This describes a very specific vocabulary specific to a domain
3. Task - This goes to a more granular level than domain ontology and specifies the vocabulary of a specific task in that domain.
4. Application - This specifies the use case involving both domain and task ontology.

Sentiment analysis uses this ontology for a more specific and accurate result. There are various approaches for sentiment mining and these are mentioned in (Ding et al.; 2008) (Khan et al.; 2011) . Most of them deal with the document level rather than the aspect level. They can be majorly categorized into the following 3 approaches -

1. With the help of lexicon (Taboada et al.; 2011)
2. Machine learning approach
3. Statistical based approach

However, there are some papers in which other techniques have been used and shown that they are also effective. A rule-based approach independent of the domain as mentioned in (Khan et al.; 2011) is a new approach which is divided into 3 main stages. The sentences are first split and tagged with a POS tagger as nouns, verbs etc. Product features used were the nouns tagged in the above step. Second step was to determine the polarity of the sentences using a publicly available lexical resource (Esuli and Sebastiani; 2006). And finally, the sentences were classified whether they are subjective or objective.

The accuracy presented in the paper is around 87 percent at the sentence level and 98 percent at the feedback level. The results may seem good but there are few points to consider. Firstly, the dataset was only from 3 domains and with an average of 1000 entries, it is too small. Also, there is no comparison with the other approaches like lexicon or machine-learning approach. According to (Ding et al.; 2008), there are 4 steps in lexicon based approach -

1. Identify the sentiment phrases and words

2. Handle the effect of valence shifters (Polanyi and Zaenen; 2006)

3. Handle the But effect

4. Add up all the opinions This can be done in various ways and have been implemented in many ways in the past. E.g. instead of addition, multiplication was used by (Kim and Hovy; 2004).

Some people used the combined approaches in their models. In (Mudinas et al.; 2012), they combined lexicon and learning based approach and concluded by showing the results that their model performed better than the lexicon-based approach and very close to the pure machine-learning based approach. Their method had four steps in which firstly they did the pre-processing of the data to remove the noise. After that, aspects were generated following which sentiment scores were calculated using the lexicon based approach and finally a feature vector was generated for each aspect. They applied their model on 2 datasets movie and software reviews. However, in the noisier reviews, the performance was not good as the sentiment model failed to detect anything but in the professional reviews, it was much better.

While everyone was making their models based on supervised learning and already made dictionaries and lexicons, (Ding et al.; 2008) proposed a model combining feature extraction and sentiment analysis using unsupervised learning approach. In contrast to (Hu and Liu; 2004), this model is not dependent on any kind of previous knowledge rather it learns by itself from the customer reviews. After the feature extraction, WordNet (Esuli and Sebastiani; 2006) database is used for opinion scores. The main part of the paper was the last stage where they predicted the rating of the product using Vector feature intensity (VFI) graph. However, this model also works on a document level but its different from the other approaches because of its unsupervised learning approach and the fact that in addition to the opinion strength, it also considers the relevance of it which is not the case in the other approaches. Dirty data seemed to be a problem in this approach as well.

When considering all the approaches, only one approach is not enough nowadays because of the complexity of the natural language. Only using lexicon based approach is not efficient because the same word can mean different in different contexts. However, a lexicon is necessary for identifying the features as only using machine learning to identify features can be a bit problematic due to the noise in the data and some infrequent mentioned features could be missed.

This paper combines the lexicon based approach (Taboada et al.; 2011) and deep learning based approach (Wang and Liu; n.d.) for the actual sentiment analysis by seeing the connections between words and not just see them as words.

3 Methodology

The process in Figure4 followed for the project and following steps were done.

1. Select the review.
2. Categorize the review.
3. Pass the review to Stanford parser.
4. Find the aspects using ontology within that sentence.
5. Calculate the sentiment using Stanford sentiment module.
6. Aggregate the results
7. Visualize

3.1 Data collection

First part is the data collection. The data was collected from several websites by web scraping. The main aim of the project is to identify the areas in which the bank is performing bad, that's why the data for categorization should mainly be complaint data. But for the sentiment analysis part, the data has to be equally distributed over negative and positive reviews and therefore a star rating is also required to extract with the review text.

3.2 Data Loading

After collecting all the data, the data was then loaded to a database. NoSQL database had to be used because of the unstructured data. So MongoDB was used because of its flexible schema and also because it uses it uses BSON language which is very similar to JSON and is very flexible.

3.3 Categorization

Data was segregated mainly into 9 categories by looking at the wordcloud, unigrams, bigrams and trigrams, read the reviews and other inspections. After that, many classification models were trained, tested and implemented using the annotation or categorization done manually.

3.4 Build the aspect ontology

The ontology was made from the reviews collected. Firstly, frequent noun approach was used and the aspects were divided in different categories as classified in the previous step. This was done because every department had certain specific terms. E.g. online transaction would lie in the area of internet banking whereas call waiting would generally lie in phone banking or customer support. The ontology would start from very general and continues going on to be specific to the department.

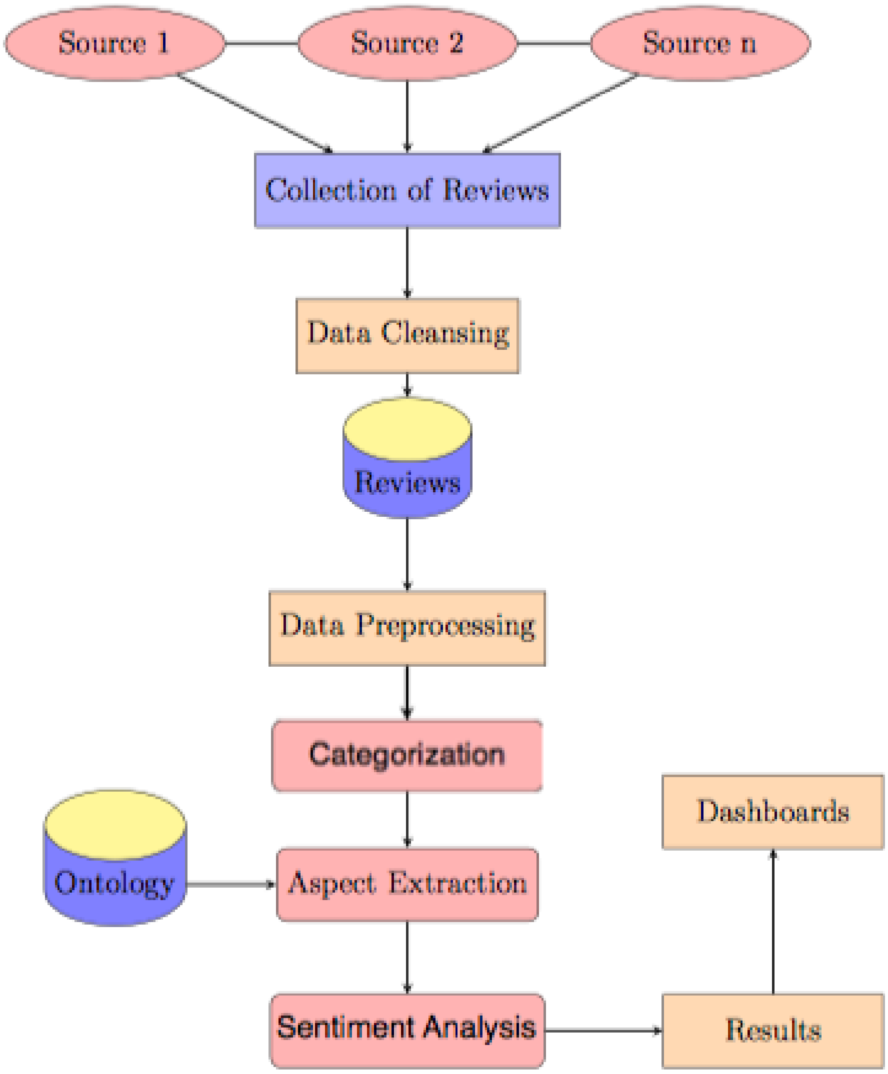

Figure 4: Flow Chart

3.5 Aspect extraction and Sentiment analysis

Sentiment analysis was done on the filtered text with the lexicon based approach combined with deep learning. Word2vec (Goldberg and Levy; 2014) will be used to convert the words to vectors. This helps representing words as numbers. If two words are related to each other, they normally have a same distance. The main parts in this are first it applies the sentence Iterator, the these sentences are tokenized, then the Vocab Cache is applied that will look up the ontology and at the end that meta data is stored as an inverted index that can be used to understand the data.

3.6 Analysis/Evaluation

Precision and recall have been used as evaluation measures for category detection and for sentiment analysis, the results were compared with the star ratings with the help of confusion matrix.

3.7 Visualization

The final data was displayed as an interactive visualization made with tableau and d3.js. The end users can see and interact with the visualization to see the more details. E.g. its reported that the bank is not performing well in category A. When they click the category A, they were presented with the reviews mentioning that aspect.

4 Implementation

Data was manually scraped through many websites including mouthshut[1], mybank-tracker[2] and consumeraffairs[3]. For this purpose, Rvest package was used in R (Wickham; 2015). The data consisted of the date of the review, the text and the location from where the review was posted.

After the dataset was gathered, a corpus was made on which various text preprocessing methods were performed including the removal of stop words, whitespace removal , stemming, converting the text to lowercase and removal of accents.This was followed by a manual inspection of the frequencies of n-grams (1-gram, 2-gram and 3-gram) by making the wordcloud to see the frequent terms appearing in all the reviews. This was done using Orange.[4]

After that, the training data set was made by marking the categories manually. This was the most time consuming part of the project as the labelling had to be done manually as there was no appropriate dataset available.

A classifier was made for classifiying the categories of the review under 9 main categories - (Bank Account/Service, Consumer Loan, Credit Card, Credit Reporting, Money

[1]http://www.mouthshut.com/
[2]http://www.mybanktracker.com/
[3]http://www.consumeraffairs.com/
[4]http://orange.biolab.si/

Transfers, Mortgage, Payday loan, Prepaid Card, Student Loan). As there were no features available in the dataset, the features were generated from the text by generating the document term matrix with unigrams and bigrams. For this purpose various algorithms were tested to see which one performs the best. The algorithms tested were Support Vector Machines (Dimitriadou et al.; 2005), glmnet (Friedman et al.; 2010), boosting (Tuszynski; 2012), random forest (Liaw and Wiener; 2002), maxent (Jurka; 2012), slda (Peters and Hothorn; n.d.), decision tree and Neural Networks. The models were implemented using RTextTools package in R (Jurka et al.; 2013) due to its ease of use. The data was separated into training and test data in 80:20 ratio. The language in the text was not gramatically correct and thats why unigrams and bigrams were also used as features which can tolerate these errors.(Cavnar et al.; 1994) Trigrams were also used in conjunction with unigrams and bigrams but the accuracy was almost the same and the time required to train the model was increased. Thats why trigrams were not included in the feature set.

Once the reviews were categorized, Aspect ontology was generated for each category. This was done by passing the reviews through the Stanford POS tagger (Manning; n.d.) and seeing the frequently mentioned nouns under that category. There are some limitations of frequency based approach but it turns out to be quite powerful as seen in (Hu and Liu; 2004).

Once the aspects are extracted and the ontology is made, Word2Vec (Mikolov et al.; 2013) was used to get the set of words similar to the aspects. This helps in getting the aspects that are not explictly mentioned in the review or are not frequent. This way helps in some way to overcome the limitation of frequency-based approach mentioned above.

Deterministic Co-Reference Resolution was also used for mapping the similar terms. e.g. Some people use acct. instead of account. So with this method, acct. and account would be mapped as same terms. (Lee et al.; 2013)

The data was passed through the sentiment analyzer to get the sentiment level of aspects in that category. StanfordCoreNLP Annotator (Manning et al.; 2014) was used in this case. Modules used were - Tokenization, Sentence Splitting, Constituency Parsing and Sentiment. Firstly the text is tokenized, which means that the sentence boundary is detected. After tokenization, text was splitted into individual sentences. Following which, syntactic analysis of the sentences is done using the parsing module which generates a tree based output giving the relations between the words of the sentence. Finally, sentiment analysis is done with the data provided.

Final Visualization is done using Tableau [5] and d3.js [6] . The visualization made using d3js is to explore different aspects in whole corpus of reviews to see exactly the problem revolving around those aspects by reading the review in that area. This is also helpful when those aspects can be seen in other Bank's reviews to see what customers are liking in the other bank.

[5] `http://tableau.com`

[6] `https://d3js.org/`

5 Evaluation

There are many evaluation measures that can be used for classification algorithms as proposed in (McLaughlin and Herlocker; 2004). But the most popular ones and easy to use are Precision, Recall and F-Score. Precision gives the information about the percentage of cases identified correctly over all the cases identified in that category. Whereas, Recall tells the percentage of sample data that was classified correctly over the samples which were actually in that category.

e.g. if there are 100 samples of Consumer Loan category and 100 of Credit card. If out of 200, 90 of them are predicted as Consumer Loan and 80 of them were correctly identified, recall in this case is 80 percent and precision would be 89 percent.

F-Score gives a score between 0 and 1 (1 being the best) by calculating the weighted average between the above 2 measures. (Sokolova et al.; 2006)

F-Score in the above case will be around 0.85.

There were many classification models used with precision and recall as below.Table 1

Algorithm	**Precision**	**Recall**	**Accuracy**
SVM	0.82	0.90	0.86
Random Forest	0.79	0.92	0.85
GLMNET	0.81	0.85	0.83
Ensemble(SVM, Random Forest, Boosting, GLMNET)	0.78	0.75	0.76
Boosting	0.72	0.78	0.75
Maximum Entropy	0.68	0.76	0.72
SLDA	0.68	0.65	0.66
Decision Tree	0.51	0.61	0.56
Neural Network	0.33	0.12	0.18

Table 1: Classification Models' Scores Comparison

As we can see that SVM, Random Forest and GLM performed equally good with F-score of above 0.80 with SVM being the best. Boosting and Maximum Entropy also performed good with the F-Score of around 0.75. However, the other algorithms didn't perform well and gave really low scores.

Even ensemble agreement was tried as mentioned in (Collingwood and Wilkerson; 2012). Ensemble agreement basically refers to when more than one algorithm predict the same label. In this case, it is shown in the following table.Table 2

It means that for a 5 ensemble agreement, 75 percent of the data is classified with a 89 percent accuracy while in 4 ensemble agreement, 90 percent of the data is classified with 84 percent accuracy. After that cross validation was done for all the algorithms to select the top four algorithms. So the best 4 algorithms identified were - SVM, Random Forest, Boosting and GLMNET but the f-score was 0.76 which was lower than SVM and random forest when applied single.

So, after testing all the algorithms SVM and Random forest turned out to be the best but finally support vector machines was used as a classifying model for categorization

n	n-Ensemble Coverage	n-Ensemble Recall
$n \geq 1$	1.00	0.78
$n \geq 2$	1.00	0.78
$n \geq 3$	0.97	0.80
$n \geq 4$	0.90	0.84
$n \geq 5$	0.75	0.89
$n \geq 6$	0.61	0.93
$n \geq 7$	0.39	0.94
$n \geq 8$	0.09	0.96

Table 2: Ensemble Summary

due to its good overall precision, recall score and speed.

5.1 Experiment / Case Study 1

The first case study is implemented in d3js. When reviews are uploaded, the visualization as shown in Figure5 is generated. Top 40 words and phrases are shown. These does not include any stop words or the words that are not useful. User can select any word or phrase and can dig deep into it. e.g. A user selects the word "customer". So, this word will be in the middle and the frequent words before and after are shown to the user as shown in Figure6. Now when he selects service, the word becomes "customer service" and all the reviews containing that word are shown with the word highlighted. There is also a bar on the left hand side depicting the sentiment of the review. Green depicts positive whereas red depicts negative. Strong shade means the sentiment is strong. e.g. in the first review, it says "the customer service is very fast and good" and the bar is green which means positive. In the second review, it says "the customer service is so poor" and the bar is red which means negative and is expected.

5.2 Experiment / Case Study 2

The second case study is when the visualization is shown as in Figure7. User can see the name of the bank on the left hand side and the number of reviews and sentiment score of that reviews on right hand side. The average stars are also shown along with the boxplot for a better understanding of the distribution.

5.3 Experiment / Case Study 3

The third case study is implemented in Tableau and the usecase taken is 'Citibank'. Figure8 shows the main categories of the reviews in a bar graph. Figure9 goes a bit deeper and shoes a stacked bar chart showing the aspects of the categories. Lastly, Figure10 shows a clear view of aspects under a prepaid card.

5.4 Discussion

We see in Figure8 that most of the Citi's complaints were of credit card. When looked deep in Figure10, we can see General Purpose card, Gift card and special card had most of the complaints. Even in Figure9, we can see in the first column that checking account

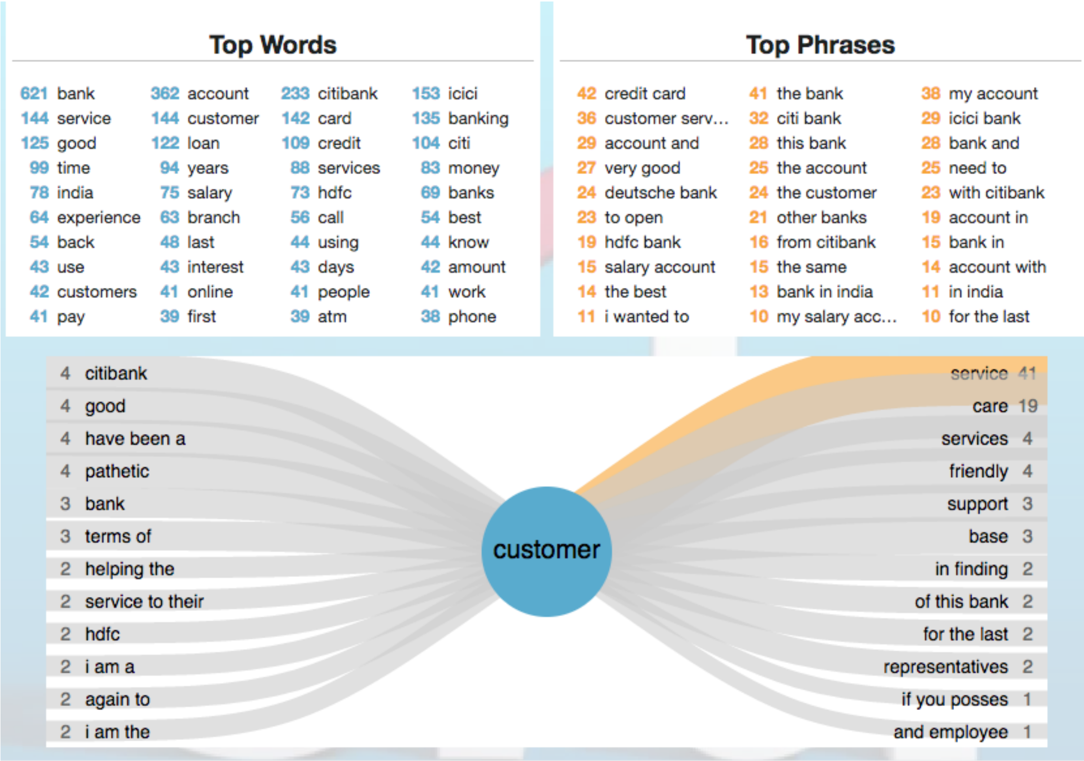

Figure 5: Top Words and phrases

Figure 6: Reviews of selected phrase

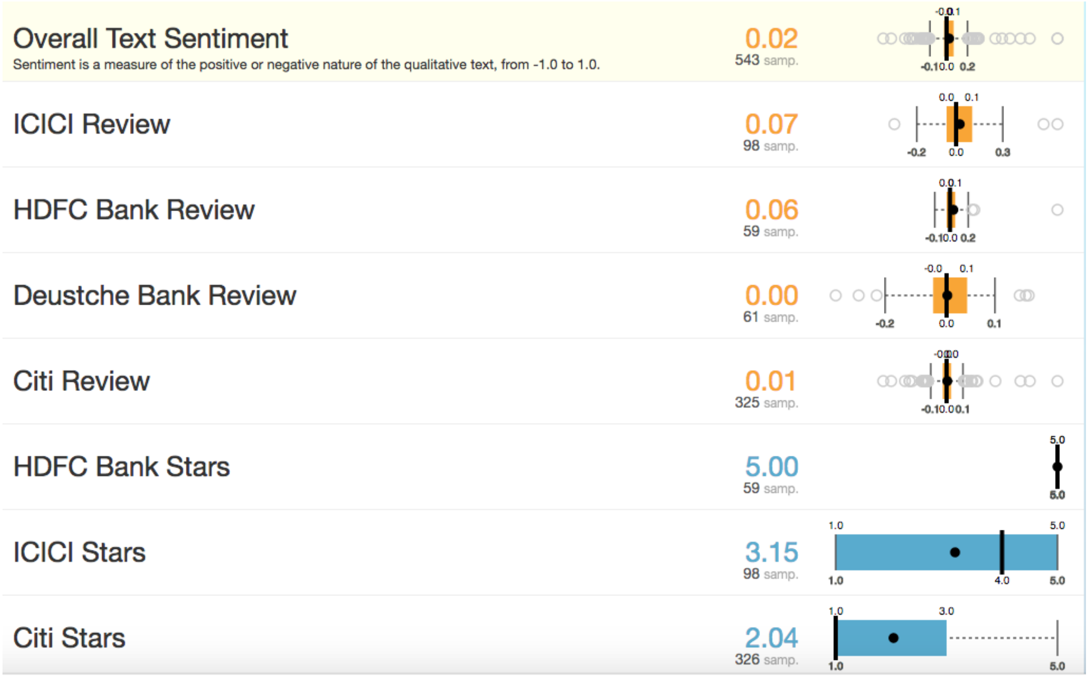

Figure 7: Sentiment and Box Plot

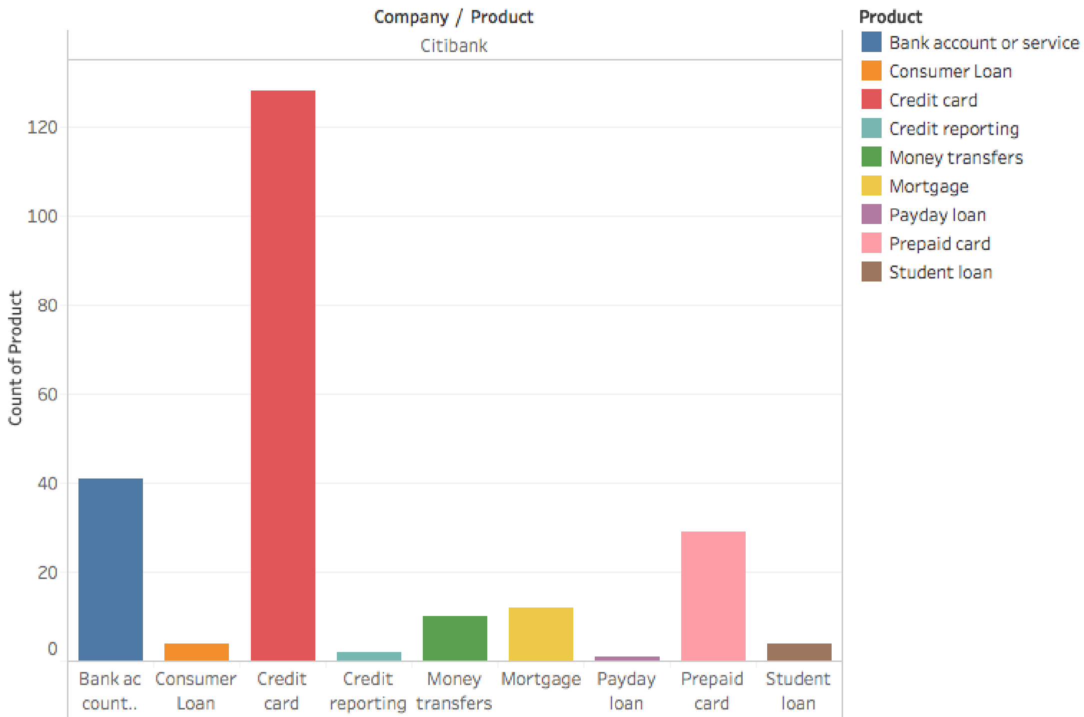

Figure 8: Bar Plot of complaints by categories

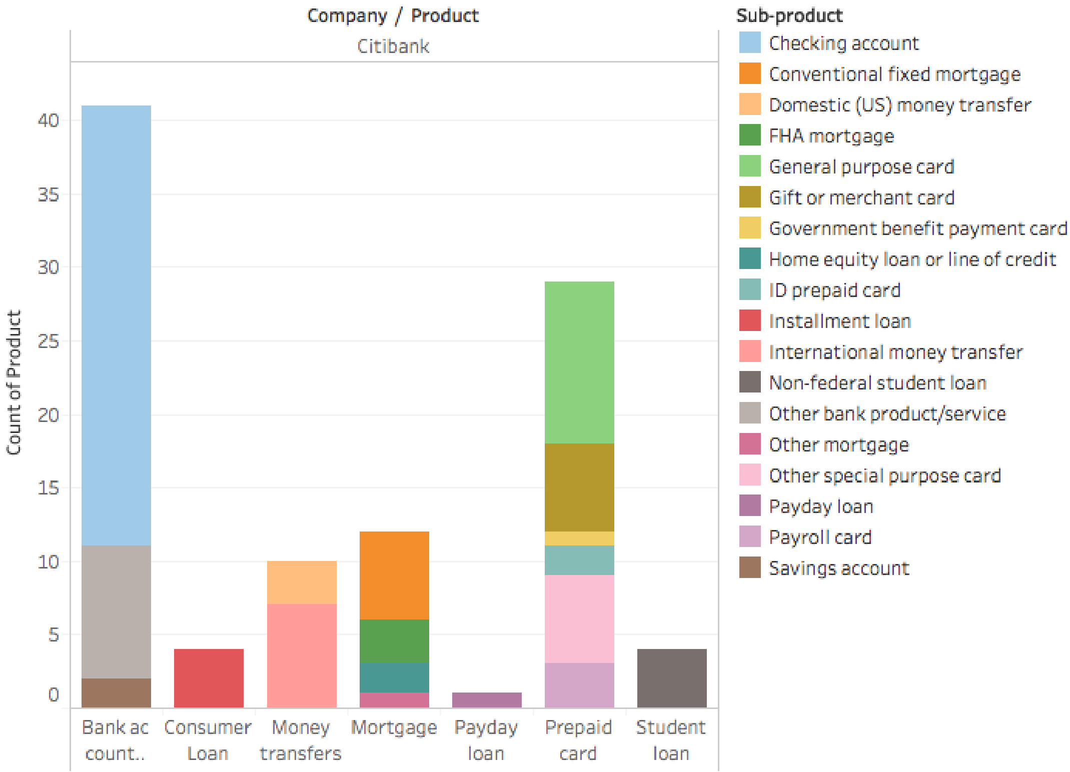

Figure 9: Stacked bar chart containing aspects of the categories

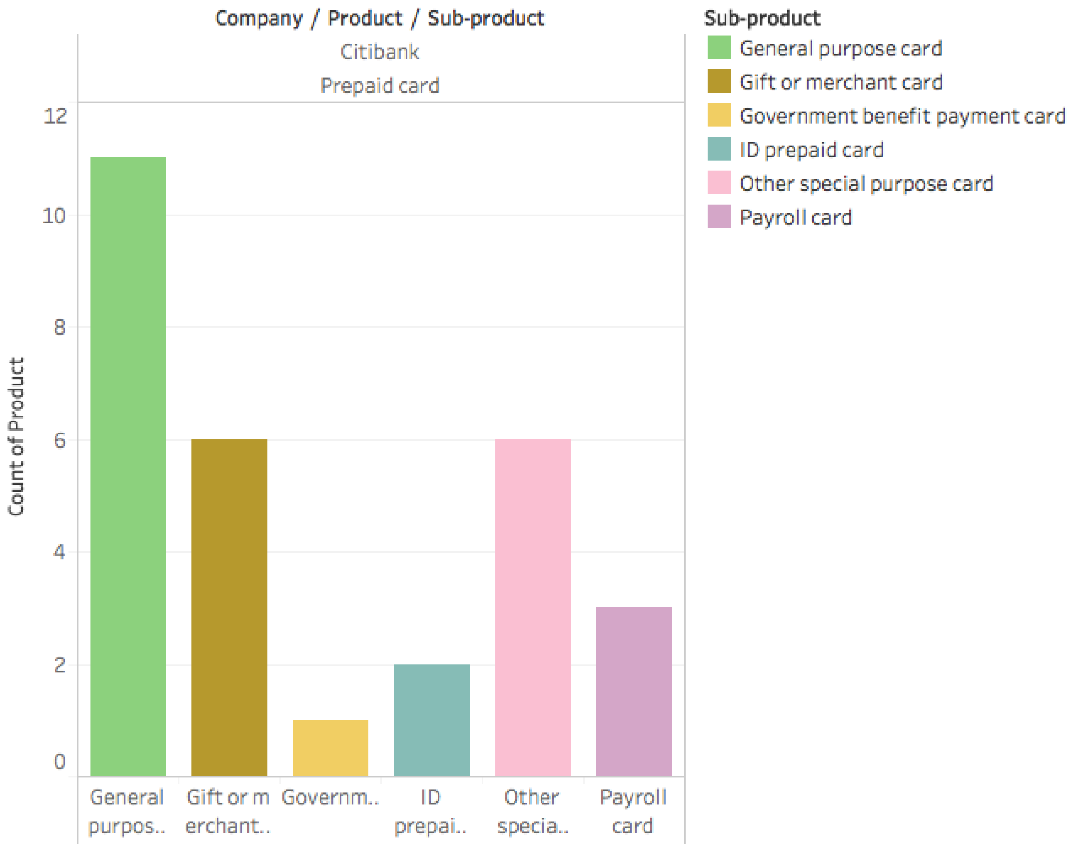

Figure 10: Bar Plot of aspects of credit card

has most of the complaints under Bank account category. Now that the scope is so narrowed down, company can just see the reviews of that category. This can help look at the customer's complaints and act on it fast.Without this system, it might not been possible to come down to a conclusion this fast. Companies can even compare themselves with other companies as shown in Figure6. In the figure, customer service is selected. So instead, company can select credit card and can see what other companies are doing good in providing the service. In this way, the company using this system would be one step ahead of the competition.

6 Conclusion and Future Work

The project implemented aspect based sentiment analysis for banking reviews and used Stanford's library for sentiment analysis. Different models were tested for categorization, however n-gram based categorization with SVM gave the best results with the accuracy of over 80 percent.

This project had 2 main stages. In the first step main category of the review was detected and in the second stage, aspects of that category were extracted and sentiments were assigned.

It is very useful for banking institutions to see where they are performing bad so that they can improve and provide better service. Also its very useful to compare with their competitors to see what users like about other banks and why. It can be useful to the consumers also when selecting their bank. They can choose between different banks by just having a glance at the visualization.

However, there are limitations of this research. The model made only applies to English language and doesn't work on any other language as every language has a different grammatical structure and a generalized model cant be made.e.g. in English, we put adjectives first as in red car but in Spanish it comes after as in carro rojo. Its just one basic example and there is a lot to complexity in different languages. Also, the model is based on supervised machine learning technique which means that it requires user input if something has to be changed or if new features are to be added.

This project has a lot of potential to improve in the future by applying unsupervised machine learning techniques. Manual text annotations can be done on the set of reviews assigning positive and negative sentiment manually. This would be useful as currently there is no annotated dataset for banking reviews.

Acknowledgements

I would like to thank my Supervisor (Mr. Vikas Sahni) who has been a constant support to me during the whole duration of the thesis. I would also like to thank my professor Dr. Simon Caton who had been a huge help throughout the whole semester giving me valuable advice for the project. A special thanks to my industry mentor Mr. Dermot Mackey who showed me how I can shape my thesis to be of use in real world. I would also like to thank my parents for always supporting me. My last acknowledgement is to

my friend Sara Guzman who has been there with me throughout the whole semester and supporting me through the hard times.

References

Cavnar, W. B., Trenkle, J. M. et al. (1994). N-gram-based text categorization, *Ann Arbor MI* **48113**(2): 161–175.

Chen, H. and Zimbra, D. (2010). Ai and opinion mining, *IEEE Intelligent Systems* **25**(3): 74–80.

Collingwood, L. and Wilkerson, J. (2012). Tradeoffs in accuracy and efficiency in supervised learning methods, *Journal of Information Technology & Politics* **9**(3): 298–318.

Collomb, A., Costea, C., Joyeux, D., Hasan, O. and Brunie, L. (2007). A study and comparison of sentiment analysis methods for reputation evaluation.

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D. and Weingessel, A. (2005). Misc functions of the department of statistics (e1071), tu wien.

Ding, X., Liu, B. and Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining, *Proceedings of the 2008 international conference on web search and data mining*, ACM, pp. 231–240.

Esuli, A. and Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining, *Proceedings of LREC*, Vol. 6, Citeseer, pp. 417–422.

Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent, *Journal of statistical software* **33**(1): 1.

Goldberg, Y. and Levy, O. (2014). word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method, *arXiv preprint arXiv:1402.3722* .

Guarino, N. (1998). Formal ontology and information systems, *Proceedings of FOIS*, Vol. 98, pp. 81–97.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews, *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 168–177.

Jindal, N. and Liu, B. (2006). Mining comparative sentences and relations, *AAAI*, Vol. 22, pp. 1331–1336.

Jurka, T. P. (2012). Maxent: An r package for low-memory multinomial logistic regression with support for semi-automated text classification, *The R Journal* **4**(1): 56–59.

Jurka, T. P., Collingwood, L., Boydstun, A. E., Grossman, E. and van Atteveldt, W. (2013). Rtexttools: A supervised learning package for text classification, *The R Journal* **5**(1): 6–12.

Khan, A., Baharudin, B. and Khan, K. (2011). Sentiment classification from online customer reviews using lexical contextual sentence structure, *International Conference on Software Engineering and Computer Systems*, Springer, pp. 317–331.

Kim, S.-M. and Hovy, E. (2004). Determining the sentiment of opinions, *Proceedings of the 20th international conference on Computational Linguistics*, Association for Computational Linguistics, p. 1367.

Lee, H., Chang, A., Peirsman, Y., Chambers, N., Surdeanu, M. and Jurafsky, D. (2013). Deterministic coreference resolution based on entity-centric, precision-ranked rules, *Computational Linguistics* **39**(4): 885–916.

Liaw, A. and Wiener, M. (2002). Classification and regression by randomforest, *R news* **2**(3): 18–22.

Liu, B. (2012). Sentiment analysis and opinion mining, *Synthesis lectures on human language technologies* **5**(1): 1–167.

Manning, C. (n.d.). Feature-rich part-of-speech tagging with a cyclic dependency network.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S. and McClosky, D. (2014). The stanford corenlp natural language processing toolkit., *ACL (System Demonstrations)*, pp. 55–60.

McLaughlin, M. R. and Herlocker, J. L. (2004). A collaborative filtering algorithm and evaluation metric that accurately model the user experience, *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, pp. 329–336.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J. (2013). Distributed representations of words and phrases and their compositionality, *Advances in neural information processing systems*, pp. 3111–3119.

Mudinas, A., Zhang, D. and Levene, M. (2012). Combining lexicon and learning based approaches for concept-level sentiment analysis, *Proceedings of the first international workshop on issues of sentiment discovery and opinion mining*, ACM, p. 5.

Noy, N., McGuinness, D. L. et al. (2001). Ontology development 101, *Knowledge Systems Laboratory, Stanford University* .

Peters, A. and Hothorn, T. (n.d.). ipred: Improved predictors, 2009, *URL http://CRAN. R-project. org/package= ipred. R package version 0.8-8* .

Polanyi, L. and Zaenen, A. (2006). Contextual valence shifters, *Computing attitude and affect in text: Theory and applications*, Springer, pp. 1–10.

Schouten, K. and Frasincar, F. (2016). Survey on aspect-level sentiment analysis, *IEEE Transactions on Knowledge and Data Engineering* **28**(3): 813–830.

Sokolova, M., Japkowicz, N. and Szpakowicz, S. (2006). Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation, *Australasian Joint Conference on Artificial Intelligence*, Springer, pp. 1015–1021.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M. (2011). Lexicon-based methods for sentiment analysis, *Computational linguistics* **37**(2): 267–307.

Tuszynski, J. (2012). catools: Tools: moving window statistics, gif, base64, roc auc, etc. r packag version 114, *URL http://CRAN. R-project. org/package= caTools.[accessed 01 April 2014]* .

Wang, B. and Liu, M. (n.d.). Deep learning for aspect-based sentiment analysis.

Wickham, H. (2015). rvest: Easily harvest (scrape) web pages, *R Package Version 0.2. 0* .