Fall 2023

SENTIMENT STRENGTH AND TOPIC RECOGNITION IN SENTIMENT ANALYSIS

Esi A.R. Adeborna
Missouri University of Science and Technology

SENTIMENT STRENGTH AND TOPIC RECOGNITION IN SENTIMENT ANALYSIS

by

ESI ADEBORNA

A THESIS

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE IN INFORMATION SCIENCE AND TECHNOLOGY

2015

Approved by

Keng Siau, Advisor
Michael Hilgers
Fiona Nah

ABSTRACT

Current sentiment analysis methods focus on determining the sentiment polarities (negative, neutral or positive) in users' sentiments. However, in order to correctly classify users' sentiments into their right polarities, the strengths of these sentiments must be considered. In addition to classifying users' sentiments into their correct polarities, it is important to determine the sources and topics under which users' sentiments fall. Sentiment strength helps as to understand the levels of customer satisfaction toward products and services. Sentiment topics on the other hand, helps to determine the specific product/service areas associated with user sentiments. This paper proposes two sentiment analysis approaches. First an approach which determines the sentiment strength expressed by consumers in terms of a scale (highly positive, +5 to highly negative, -5) is proposed. The approach includes a novel algorithm to compute the strength of sentiment polarity for each text by including the weights of the words used in the texts. Second, a sentiment mining approach which detects sentiment topic from text is proposed. The approach includes a sentiment topic recognition model that is based on Correlated Topics Models (CTM) with Variational Expectation-Maximization (VEM) algorithm. Finally, the effectiveness and efficiency of these models is validated using airline data from Twitter and customer review dataset from amazon.com.

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisor Dr. Keng Siau for the continuous support during my study and research, for his patience, motivation, enthusiasm, and immense knowledge. I cannot imagine having a better advisor and mentor for my study.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Michael Hilgers and Dr. Fiona Nah, for their encouragement, insightful comments, and questions. My sincere thanks also goes to Sheri Light for all the support and help scheduling appointments and meetings with my thesis committee.

Last but not the least, I would like to thank my family: my parents and my husband, Kenneth Fletcher, for the sleepless nights we worked proof-reading and editing my work.

TABLE OF CONTENTS

Page

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

Traditionally, companies rely on telephone surveys to measure client satisfaction. Telephone surveys are very costly and may not always yield the desired data needed if those contacted are unwilling to take the survey or fully answer all the questions. However, the advent of microblogging sites such as Twitter, Tumblr, FriendFeed, Google Buzz, and web content in general has increasingly changed and shaped the corporate environment and competitive landscape. Consumers, non-profit organizations, and other interested parties are able to send messages through a variety of means to express their opinions and perceptions on companies and their brands on the web. These consumer opinions on the web are mostly free and can be used by companies as a source of survey data. Companies can leverage this data to provide insight about overall customer satisfaction and identify areas for more in-depth investigation. Analyzing individual postings manually is a daunting task and it is almost impossible. Specific methods and algorithms are required to process these opinions to extract useful information and patterns. One such method is sentiment mining.

Sentiment mining (SM) involves the analysis of a text string to determine whether a corpus is of a negative or positive opinion or emotion (e.g., happy, frustrated, bored, excited or sad). It also addresses such problems as distinguishing objective from subjective propositions, and determining the sources of different opinions expressed in a document and summarizing writers' judgment over a large corpus of texts (Pang & Lee 2008). With each opinion expressed on web, sentiment mining can be used to determine if the consumer is generally positive about the topic in discussion or negative about it. User opinions do not only reflect the user's viewpoint and sentiment polarity but also they reflect a customer's emotional intensity toward a product or service. The different emotion intensity expressed by users toward products or services reveal their satisfaction levels which can have a great impact on purchasing behaviors and other customers. The process of analyzing the emotional intensity or sentiment strength in text is termed Sentiment Strength Detection.

Sentiment strength is reflected by the strength of subjective words (words expressing user's positive, neutral or negative opinion) used in a text to depict a user's emotional intensity (Lu and Kong, 2010). In spite of lots of work done in the area of positive and negative sentiment detection there are very few done in exploring sentiment strength of texts. We propose an approach to determine the strength of each sentiment. We calculate this by including the weights of words used in each text in our analysis to determine the strength of customer sentiment based on a scale of +5 to -5 (highly positive to highly negative). Though the polarity of two different texts may be the same (both texts have a positive or negative polarity), the strength of positive or negative sentiment of one text may differ from the other based on the words used.

Consider the following two reviews, review 1 and review 2, on music and word processing software respectively. Although both reviews have negative sentiments, the strength of negative words (straight underline) used in the first review outweigh those in the second, which has some positive words (squiggle underlinefyel) as well, making the strength of negative sentiment in the first review higher than that of the second. Based on a scale of +5 to -5 (highly positive to highly negative), review 1 ranks -5 (highly negative) while review 2 ranks -3 (somewhat negative). We determined the sentiment strength based on the weights of the positive and negative words used in the reviews.

Assigning weights to words for polarity classification, helps us identify the strength or the actual overall sentiment of a text based on the weights of positive or negative words used. In this way, we are able to determine the strength of customer sentiments toward a particular product or a specific topic in the customer reviews.

Review 1 Where do people come up with this garbage? It is so obvious she's about as real an artist as her sister (rich & dumb Jessica Simpson-Lachey). The music is so vapid, phoney and devoid of any real content; the lyrics remind me of love notes scribbled on the back of a trapper keeper. I enjoy great, entertaining music like the next person - but this stuff just insults the intelligence.

Review 2 I use Word and Entourage, every single day. I use PowerPoint at least once a week. I use Excel about once every month. I must use this suite for the interaction I need with PC users, but I am not a happy camper. This program is pretty, and it has some nice features, but, overall, it's mediocre. At times, it's downright bad.

Besides the aforementioned benefits in using sentiment strength detection in SM, SM can be extended to determine the source and topic of different opinions expressed in textual data. With this aspect of sentiment mining (Sentiment Topic Recognition), opinions expressed are not only classified into positive or negative sentiments, but also provides a deep understanding of specific drivers and the overall scope of each sentiment. The insights gained from Sentiment Topic Recognition can be useful in setting performance goals, establishing performance metrics, setting service standard, attaining better brand image, and enhancing competitiveness.

Sentiment Topic Recognition (STR) seeks to identify the most representative topics discussed for each sentiment. Through STR analysis, it is possible to acquire a high level view regarding the underlying causes of positive and negative sentiments (Mostafa 2013). The research field of sentiment mining, also known as sentiment analysis or opinion mining, has developed algorithms to identify the sentiment orientation (positive or negative) of online text and to determine if a text is subjective or objective (Pang et al. 2002; Pang & Lee 2004; Thelwal et al. 2010; Pak & Paroubek 2010). Many of such algorithms have been applied to sentiment-related problems on a large-scale across multiple domains. The study by Pang et al. (2002) focuses on determining the sentiment orientation of movie reviews. Another study focuses on the average level of sentiment expressed in blogs (Dodds & Danforth 2010). The goal is to identify overall trends in levels of happiness with respect to age and geographical differences. Nonetheless, very few studies have investigated STR (Cai et al. 2010; Lin et al. 2012; Zhao et al. 2012).

The main contributions of this research are summarized below:

(1) An STR model that uses Correlated Topics Models (CTM) (Blei and Lafferty 2006) with Variational Expectation-Maximization (VEM) algorithm.

(2) The STR model is used as a means to obtain information to examine the reputation of airlines by computing their Airline Quality Rating (AQR) (Bowen & Headley 2013). We propose to assess the AQR based on customer sentiments towards three major airlines (AirTran Airways, Frontier and SkyWest Airlines) from tweets. The AQR is subsequently computed based on subjective data from microblogs instead of the usual consumer surveys.

(3) A prototype is developed and an example of how the STR approach is applied is illustrated using a case study with real-world tweets.

(4) The STR model is also used as a means to obtain information to examine the sentiment strength in customer reviews by computing the weights of each positive or negative word used in the review to determine the degree of positive or negative sentiment expressed by the consumer.

(5) A prototype is developed and an example of how the sentiment strength is illustrated using a case study with Amazon.com customer review dataset.

(6) An algorithm to match opinionated tweets to a topic lexicon is developed. We performed an evaluation on both of the prototypes developed.

This paper is organized as follows. Section 2 provides a literature review on sentiment mining. Section 3 presents an overview on sentiment mining. The case study domain and datasets are discussed in section 4. Following that, we present and discuss the STR approach and case study in section 5. In section 6, we discuss the Sentiment Strength detection framework with an experiment on big data and discuss its results. We provide the conclusion, and highlight future research in section 7.

2. LITERATURE REVIEW

In this section, we discuss related work regarding sentiment strength detection and sentiment topic recognition.

2.1. SENTIMENT STRENGTH DETECTION

Lu et al. (2010) proposed an approach for estimating the sentiment strength of user reviews according to the strength of adverbs and adjectives expressed by users in their opinion phrases. The overall sentiment strength of a review is calculated by averaging the strength of opinion phrases. They calculate the strength of each opinion phrase in a review by considering using opinion words (adverbs and adjectives), words expressing users' positive, neutral or negative opinion, because the sentiment strength is reflected by the "strength" of these words (Lu et al. 2010). They manually mark the strengths of a few frequently used adverbs in the opinion phrases with values ranging from -1 to +1 based on their intuitions. The strength of adjectives is calculated by a link analysis method based on a generated progressive relation graph of adjectives, in which each node is an adjective and the directed edge between nodes is a kind of progressive relationship obtained through the search engine based on heuristic rules.

Thelwall et al (2010) propose a new algorithm, SentiStrength, which was designed to identify opinions about user behaviors. The algorithm employs several methods to simultaneously extract positive and negative sentiment strength from short informal electronic text. SentiStrength uses a dictionary of sentiment words with associated strength measures and exploits a range of recognized nonstandard spellings and other common textual methods of expressing sentiment. The main contributions were a machine learning approach to optimize sentiment term weightings, methods for extracting sentiment from repeated-letter nonstandard spelling in informal text, and a related spelling correction method.

Another work on positive–negative sentiment strength detection by Pang and Lee (2010) used modified sentiment analysis techniques to predict the strength of human ratings on a scale of 1 to 5 for movie reviews. This is a kind of sentiment strength evaluation with a combined scale for positive and negative sentiment. Experiments with human judgments led the authors to merge two of the categories and so the final task was a four-category classification, with a three-category version also constructed for testing purposes.

The existing works in sentiment strength detection are few and none of these works uses the lexicon-based method in sentiment strength detection the way our proposed approach does. Our proposed approach combines two lexicon with known word strengths in our sentiment strength algorithm to determine the overall strength of text. Previous research works focus on short text and machine learning algorithm to determine sentiment orientation for words in a text and subsequently calculate the strength of the text. Sometimes the determination of words strength is done manually by intuition which can lead to error and misguided results. Unlike these works, our approach provides an automated way of determining sentiment strength of text.

2.2. SENTIMENT TOPIC RECOGNITION

Existing work on sentiment classification techniques focuses heavily on classifying opinionated text mostly from social media and consumer reviews into positive, negative or neutral categories. There is also an emphasis in recent research work on differentiating subjective and objective text.

The cut based classification, employed by Pang and Lee (2004), combines individual preference and relationship-based methods of classification. They proposed a text classification process that labels the sentences in a document as either subjective or objective. Discarding the latter, they then applied a standard machine-learning classifier to the resulting extract. This process prevents the polarity classifier from considering irrelevant or potentially misleading text. The Naïve Bayes and the support vector machine (SVM) are then trained on the subjective dataset and then used as a basic subjectivity detector. The former yield more accurate classification results. Pak and Paroubek (2010) built a sentiment classifier using the multinomial Naïve Bayes classifier. The classifier is based on Bayes' theorem. The classifier uses the part-of-speech (POS) distribution to estimate probability of POS-tags present within different sets of texts and uses it to calculate posterior probability. To increase the accuracy of their classifier, the authors discarded phrases or expressions that do not strongly indicate any sentiment or phrases or expressions that indicate objectivity of a sentence.

Thelwall et al. (2010) assess whether popular events are typically associated with increase in sentiment strength, which seems intuitively likely. Their results provide strong evidence that popular events are normally associated with increase in negative sentiment strength. They also provide evidence that peaks of interest in events have stronger positive sentiment than the time before the peak. Whitelaw et al. (2005) differentiate their sentiment classification method from fine-grained semantic distinction in features used for classification by employing appraisal groups such as “very good” or “not terribly funny”. An appraisal group is represented as a set of attribute values in several task-independent semantic taxonomies based on Appraisal Theory.

In a recent study on sentiment topic detection, Lin et al. (2012) propose a novel probabilistic modeling framework called joint sentiment-topic (JST) model based on latent Dirichlet allocation (LDA), which detects sentiment and determines topic simultaneously from text. The LDA model is based on the assumption that documents contain a mixture of topics, where a topic is a probability distribution over words. The JST is a weakly supervised model that adds an additional layer between the document and the topic layer. This makes JST a four layered model where sentiment labels are associated with documents under which topics are associated with sentiment labels and words are associated to both sentiment labels and topics. Similarly, Cai et al. (2010) develop a holistic sentiment mining system that consists of sentiment and topic detection method. Their sentiment detection uses a statistical-based approach while their topic detection method is based on point-wise mutual information and term frequency distribution. They conducted their experiment around an Australian brand called “Vegemite” and InsuranCo. Zhao et al. (2012) present a hierarchical generative model, called user-sentiment topic model (USTM), which captures users' topics with sentiment information. USTM refines users' topics with different sentiment trends including positive, negative and neutral. USTM is an unsupervised generative model that captures user’s sentiment on topic level by considering topic and sentiment simultaneously. Each topic extracted by USTM has a sentiment label. USTM aims at obtaining the sentiment-refined topics for investigating user-level sentiment analysis. The authors conducted experiments on one Chinese dataset (IT products) and two English datasets (movie reviews and Enron emails). The authors found that USTM performs better on modeling user's interests when the sentiment and topics extracted by USTM are informative and clear towards the sentiment label.

Different from these works, our proposed approach captures users' sentiments and topics intrinsic to such sentiments concurrently. In this way, each sentiment extracted by the model has some underlying topic(s) and provides an overall knowledge and scope of the different consumer sentiments. The approach aims at answering questions regarding the drivers of each labeled sentiment in a dataset and examining the overall breadth of the sentiment. Previous research works on sentiment analysis first extract topic-related text from documents or social media and then use classifiers to determine sentiment orientation for the text under each specific topic. Unlike these works, our model provides an underlying reason for sentiment orientation based on correlated topics in each sentiment.

3. OVERVIEW ON SENTIMENT MINING

The growing interest in sentiment mining is partially due to its application in government intelligence, review-oriented websites, product or service recommendation systems, and other business domains. Sentiment mining (SM) differs from 'classic' text mining in that it seeks to identify the point of view in corpus of text. SM techniques are employed in tracking attitudes and feelings on the web, blog posts, comments, reviews and tweets about different topics. SM is also very useful in the area of customer relationship management to track products, brands and determine whether they are viewed positively or negatively by customers. This allows business to be able to manage their reputation and accurately track their new product perception or brand perception.

Generally, language is ambiguous which makes it hard to build models that can accurately analyze the polarity of a text corpus based on the individual words in the text data. Some sentiment prediction models work by isolating words, giving positive points for positive words and negative points for negative points and then summing up the points. That way, the word order is ignored and context meaning of words and phrases in the sentence is lost.

3.1. SUBJECTIVE VS. OBJECTIVE TEXT

A subjective expression is any word or phrase used to express an opinion, emotion, evaluation, stance, speculation etc. A general covering term for such states is private state (Quirk et al 1985). Subjectivity is used to express private states in the context of a text or conversation. Private state is general term for opinions, evaluation, beliefs, perceptions, emotions etc (Padmaja and Fatima 2013). An opinion according to the Webster dictionary is a belief, judgment, or way of thinking about something; it is what someone thinks about a particular thing. In other words, opinion is a subjective belief, and it is the result of emotions or appraisal formed in the mind from interpretation of facts. Subjectivity analysis involves a range of techniques and processes that originate from Information retrieval, Artificial Intelligence and Natural Language Processing. Sentiment mining is used to detect and extract subjective information in text documents. It tries to determine the opinion of a writer about a specific topic or the overall contextual polarity of a document. On the other hand, if a user feedback has no judgment or opinion on the source content then it is called objective.

3.2. POLARITY CLASSIFICATION

The binary classification task of labeling an opinionated document as expressing either an overall positive or an overall negative opinion is called sentiment polarity classification or polarity classification (Pang et al 200). Syntactic features are used with knowledge of linguistic terms to classify contents of document into positive and negative, and subjective and objective terms. Sentiment polarity allows the use of a single dimension therefore simplifying the representation and management of the sentiment information. Existing research work offers various techniques and ideas for extracting sentimental terms or expressions from text. Most of these works use the part-of-speech (POS), stop word removal (Pak and Paroubek 2010), fuzzy pattern matching, stemming, punctuation, link based patterns, document citation, and stylistic measures for extraction of sentiments. Classification of sentimental terms or expression according to their meaning and background knowledge is called Semantic Orientation: positive or negative (good or bad) (Padmaja and Fatima, 2013; Lin et al 2013). Some researchers suppose that a single term can be used in a different sense and therefore presents a different opinion. Various methods such as WordNet and Synset are used to evaluate different senses of the same expression or term. These methods are used to measure the semantic orientation of words and to determine the similarity between words (Lin et al 2013).

3.3. THE SENTIMENT MINING PROCESS

The general steps involved in sentiment mining are depicted in the Figure 3.1. The first step in the sentiment mining process is data gathering and preparation (Pak and Paroubek 2010). In data gathering for example, if you are trying to determine the sentiment of tweets for a particular topic, you will need to obtain large volumes of tweet and all documents related to the context or the topic being studied. The format of the file, its features, message length or time composed, constitutes various ways to prepare a text data for processing. The text preparation depends on data input criteria of the system or model being used. Information Extraction concerns locating specific pieces of data in natural-language documents, thereby extracting structured information from free text (Kanya and Getha 2007). One of the processes used in Information Extraction is part-of-speech (POS) tagging. In POS each word in the text (or a sentence) is assigned a label such as subject (S), verb phrase (VP), verb (V), noun phrase (NP), preposition (PP) or determiner (Det). We then look at sentiment orientation of the patterns extracted from the text. At this stage, the machine situates the words on an emotive scale and the average sentiment orientation of all phrases is computed. They are many methods used in classifying sentiments such as Naives Bayes, Maximum Entropy and Support Vector Machine. Many sentiment classifications that deal with exact groups (Sad, happy, excited, excellent etc…) have more granularities and require more accuracy which mostly necessitates a deeper understanding of human language. The next step is interpretation of results and evaluation. The accuracy and precision of a model can be done based on some baseline (Wang and Manning 2012) or comparing the model with existing models. The final results represent the performance results of the sentiment mining system.

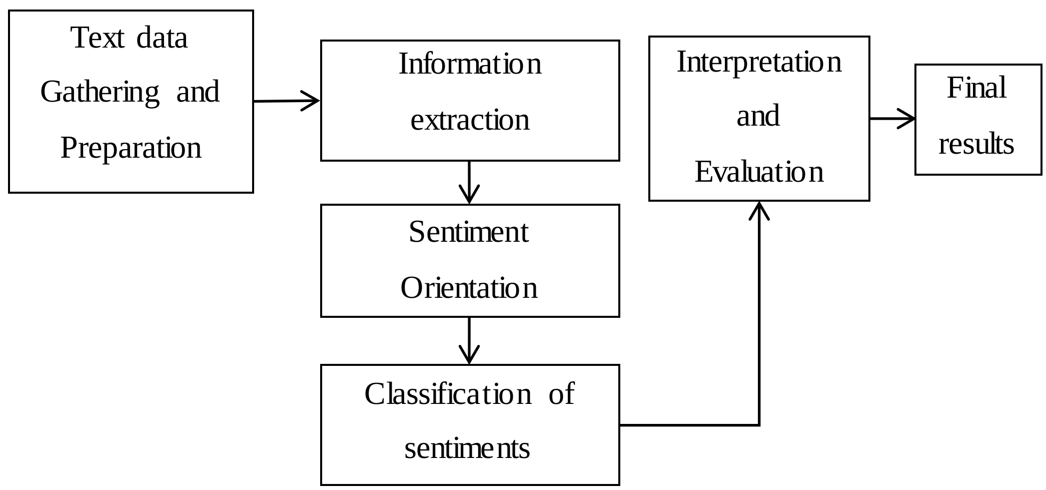


Figure 3.1 General Steps Involved in Sentiment Mining.

4. CASE STUDIES DOMAINS AND DATASET DESCRIPTION

This section describes the domains and the two datasets that have been used to validate our approaches in this paper. We also provide reasons for our choice of datasets.

4.1. AIRLINE DATASET FROM TWITTER

Twitter is a microblogging site and its central activity is posting short status update messages (tweets) via the web or a mobile device. It is used to share information and to describe minor daily activities (Java et al. 2007), although it can also be used for information dissemination. Since Twitter is the most well-known microblogging site, it was selected as a source to gather data to conduct the analysis for our study. Our data will represent a random set of tweets for three airlines (AirTran Airways, Frontier and SkyWest Airlines). The airline industry was chosen to examine the reputation of these major airlines by computing their Airline Quality Rating (AQR) based on the output from our STR approach.

Airline Quality Rating (AQR) was developed and first announced in early 1991as an objective method for assessing airline quality on combined multiple performance criteria. Dr. Bowen’s and Dr. Headley’s research on the development of the national AQR is viewed by more than 75 million people each year and is annually featured by national news outlets (Bowen & Headley 2013). Therefore, there is an increased awareness among consumers in choosing airlines based on their reputation, thus the use of the airline dataset.

4.1.1. Airline Quality Rating (AQR). The AQR is an objective method for assessing airline quality by combining multiple performance criteria (Bowen & Headley).

The formula for calculating the AQR score is:

$$\text{AQR} = \frac{(+8.63 \times OT) + (-8.03 \times DB) + (-7.29 \times MB) + (-7.17 \times CC)}{(8.63 + 8.03 + 7.29 + 7.17)} \quad (1)$$

Where OT (on-time), DB (denied boarding), MB (mishandled baggage), and CC (customer complaints) are variables considered as shown in Table 4.1. Data for all criteria is drawn from the U.S. Department of Transportation's monthly Air Travel Consumer Report (Bowen & Headley 2013). Weights reflect the importance of the criteria in consumer decision-making, while signs reflect the direction of impact that the criteria should have on the consumer's rating of airline quality. Weights were originally established by surveying 65 airline industry experts regarding their opinion as to what consumers would rate as important (on a scale of 0 to 10) in judging airline quality (Bowen & Headley 2013). The AQR values used in this research are based on April 2013 reported values. Higher AQR values indicate higher reputation (as shown in table 4.2). Virgin America (VX) for example, had the best rating in 2012 with an AQR value of -0.35 (see table 4.2).

4.1.2. Data Collection and Preparation. The airline dataset contains 452 tweets on AirTran, 499 on Frontier Airlines and 195 on SkyWest Airlines collected from twitter. The airline dataset contains 452 tweets on AirTran, 499 on Frontier Airlines and 195 on SkyWest Airlines collected from twitter. Each tweet contains some comment made about

Table 4.1 AQR, Criteria, Weight and Impact.

CRITERIA		WEIGHT	IMPACT (+/-)
OT	On-Time	8.63	+
DB	Denied Boarding	8.03	-
MB	Mishandled Baggage	7.92	-
CC	Customer Complains	7.17	-

Each of these airlines; positive, negative or neutral. Based on Table 4.2, we selected each airline such that there is a significant distinction between the different ratings of each airline during our analysis. AirTran Airways is among the top 3 airlines; Frontier Airlines is ranked 7 out of 14 for AQR, and SkyWest Airlines is ranked 12. Following Thelwal et al. (2011), only English tweets will be used to avoid complications in analyzing multilingual tweets.

After the tweets are gathered, we prepared the data for sentiment analysis. The data preparation process follows the steps listed below:

(1) Collect the related web comments discussing a particular subject (e.g., AirTran) from tweets.

(2) Remove retweet entries, html links and markups.

(3) For each given set of tweet, remove punctuations, numbers, @, people, and unnecessary spaces.

4.2. AMAZON DATASET

Amazon.com is an American international commerce website and the world's largest online retailer. Amazon.com is noted for its diversified online market that gives

Table 4.2 Airline Quality Rating Scores.

	2012 AQR		2011 AQR		2010 AQR		2009 AQR		2008 AQR		2007 AQR	
	Score	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**
AirTran	-0.51	3	-0.48	1	-0.48	1	-0.49	2	-0.84	2	-1.03	1
Alaska	-0.77	6	-0.79	5	-0.94	4	-1.39	11	-1.16	5	-1.75	7
American	-1.11	10	-1.24	10	-1.28	11	-1.25	9	-1.71	9	-2.19	9
American Eagle	-1.78	11	-2.51	15	-2.82	16	-2.83	11	-3.12	16	-3.80	15
Delta	-0.58	4	-0.80	6	-1.22	7	N/A	-	N/A	-	N/A	-
Express Jet	-1.95	13	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-
Frontier	-0.78	7	-0.75	4	-1.27	9	-1.09	7	-1.31	7	-1.71	5
Hawaiian	-0.71	5	-0.59	2	-0.58	2	-0.40	1	-0.69	1	N/A	-
JetBlue	-0.43	2	-0.60	3	-0.70	3	-0.62	3	-0.90	3	-1.30	2
SkyWest	-1.88	12	-1.15	9	-1.28	10	-1.57	14	-2.13	13	-3.09	13
Southwest	-0.81	8	-0.93	7	-1.01	5	-1.00	5	-1.23	6	-1.59	3
United	-2.18	14	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-
US Airways	-0.87	9	-1.13	8	-1.17	6	-1.19	8	-1.77	10	-2.94	11
Virgin America	-0.35	1	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-
Industry	-1.11		-1.08		-1.20		-1.27		-1.63		-2.16	

consumers and vendors the opportunity to buy and sell respectively. Any registered amazon.com customer is at liberty to write customer reviews. The items or products that reviewers write on do not have to be purchased from amazon.com. Anyone who feels to write a review about a particular product can do so at anytime as long as they are registered customers at amazon.com. Customer reviews on amazon.com stay on the site indefinitely. Since amazon.com is the most well-known online retailing site and has volumes of customer reviews for different products, it was selected as a source to gather data to conduct the analysis for our big dataset. Our data will represent a random set of customer reviews for different products. This review dataset was used by Jindal and Liu (2008). The data contains over 5.8 million reviews, 2.14 reviewers and 6.7 million products. Out of this dataset, a random number of reviews of 866,056 were used for our big dataset. In this paper, we used four main categories to classify the products in our dataset namely; books, music, DVD and electronics. Each amazon.com review consists of 8 fields;

<Product ID> <Reviewer ID> <Date> <Number of Helpful Feedbacks> <Number of Feedbacks> <Rating> <Review Title> <Review Body>

We prepared the data for sentiment analysis using the following process;

(1) Extract the Review Body field from the reviews

(2) For each review, remove tabs and unnecessary spaces

(3) Remove punctuations, numbers, and convert all review text to lower case.

5. SENTIMENT TOPIC RECOGNITION APPROACH

Sentiment Topic Recognition (STR) seeks to identify the most representative topics discussed for each sentiment. Through STR analysis, it is possible to acquire a high level view regarding the underlying causes of positive and negative sentiments (Mostafa 2013). This paper proposes an approach using STR in sentiment mining. The approach was tested on a small dataset from twitter on airlines and a big data from amazon.com on customer reviews.

5.1. FRAMEWORK

Figure 5.1 shows the framework for our proposed approach. The different steps involved in this framework are explained in the subsections below.

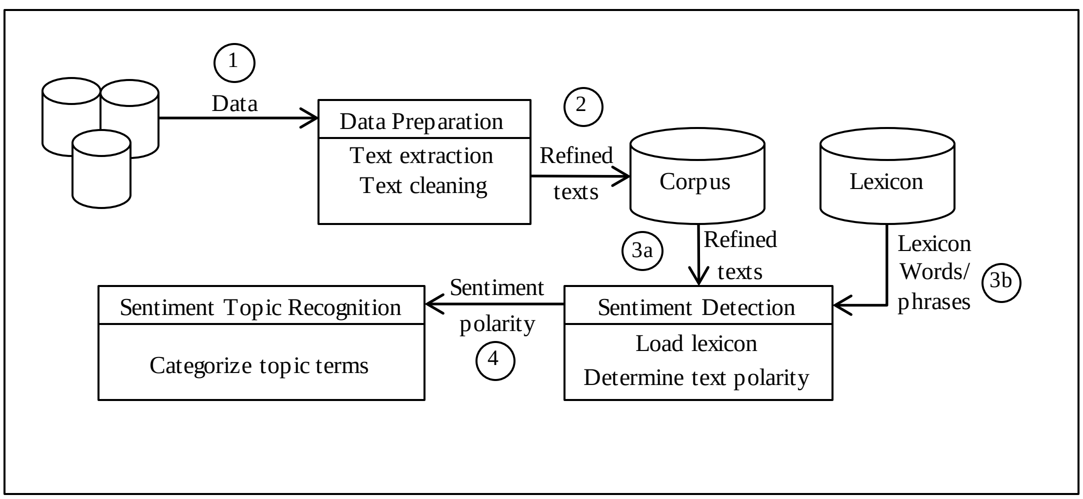


Figure 5.1 Framework for Sentiment Detection and STR.

In this framework, we collect data from source and prepare it. Some examples of data preparation are described in sections 3.1 and 3.2. The prepared data is then stored as Corpus to be used in the sentiment detection process. The sentiment detection process provides the polarity of the corpus, positive, negative or neutral. Finally, the sentiment topic recognition process categorizes the subjective (positive and negative) sentiments under topics based on the VEM and CTM algorithm.

5.1.1. Sentiment Detection. One approach to sentiment detection starts with labeled texts and uses supervised machine learning trained on the labeled text data to classify the polarity of new texts (Pang and Lee 2008). There have been many algorithms that have been applied to sentiment classification. These algorithms include decision trees (Lewis & Ringuette 1994), k-nearest neighbors (Yang & Chute 1994; Yang and Pedersen 1997; Tan 2005), neural networks (Wiener et al. 1995) and support vector machine (SVM) (Joachims 1999). Another approach creates a sentiment lexicon and scores the text based on some function that describes how the words and phrases of the text matches the lexicon. Determining sentiment polarity is done by comparing the corpus against a predefined lexicon of subjective words. In this study, we used Hu and Liu's (2004) lexicon to conduct the analysis. Successful use of this lexicon was demonstrated by (Mostafa 2013; Miner et al. 2012). This lexicon includes around 6800 seed adjectives with known orientation of 2006 positive and 4783 negative words. The Hu and Liu's (2004) lexicon was used in the sentiment detection portion of the study. However, we employed a slight modification of this lexicon in our STR model by adding words through a thorough search in WordNet based on the AQR criteria (Bowen & Headley, 2013). Our lexicon modification is based on (Neviarouskaya et al. 2011; Taboada et al. 2011). The process extracts words or compound words from WordNet that are related to the predefined topics (AQR criteria; On-time, Denied boarding, Mishandled baggage, and Customer complaints) that will be used in categorizing the topic words in STR model.

Sentiment Matching Algorithm the SentimentTopicMatching algorithm (see Algorithm) is used to match terms relating to specific topics. The idea behind this algorithm is to find those words/terms that relates to a topic, t with respect to its lexicon. The input and out to the algorithm are m tweets, T and the topic lexicons, L, respectively.

Algorithm *SentimentMatching (T, L)*

Input: tweets, $T = \{t_1, t_2 \ldots t_m\}$, topic lexicons, $L = \{l_1, l_2 \ldots l_n\}$

Output: criteria with terms, C

```
 1: Initialize word list, W, match list, M
 2: for each tweet t in T do
 3: W = split_tweet (t)
 4:     for i from 1 to n do
 5:         M ← match_lexicon(W, L_i)
 6:             for j from 1 to size(M) do
 7:                 Initialize match index, k
 8:                 if M_j > 0 then
 9:                     terms_{i,k} ← L_{i,W_{M_j}}
10:                     k ← k+1
11:                 end if
12:             end for
13:     end for
14:     Append terms at the end of C
15: end for each
16: Output C
```

The algorithm iterates through T and for each tweet t∈T performs certain actions. Line 3 splits the tweet, t, into its individual words and store the list of words in the word list, W, by calling the split_tweet function. We then match W against each given lexicon, l, where l∈L. The match_lexicon function, responsible for this action, returns the location of the word in l if a match is found otherwise -1. Lines 6-12 iterate though the match list, M, and uses the returned locations to retrieve the match words into terms. In line 14, terms for each t is appended to the criteria with terms matrix, C. The algorithm scales linearly, therefore, it’s running time increases with the size of tweets and the number as well as the size of each lexicon.

For our analysis, we employed the R software package version 3.0.2. The software was selected because it is free and open source. It also provides comprehensive packages for quantitatively analyzing and visualization data. R also permits the integration of different algorithms and provides the flexibility to customize codes to produce desired results.

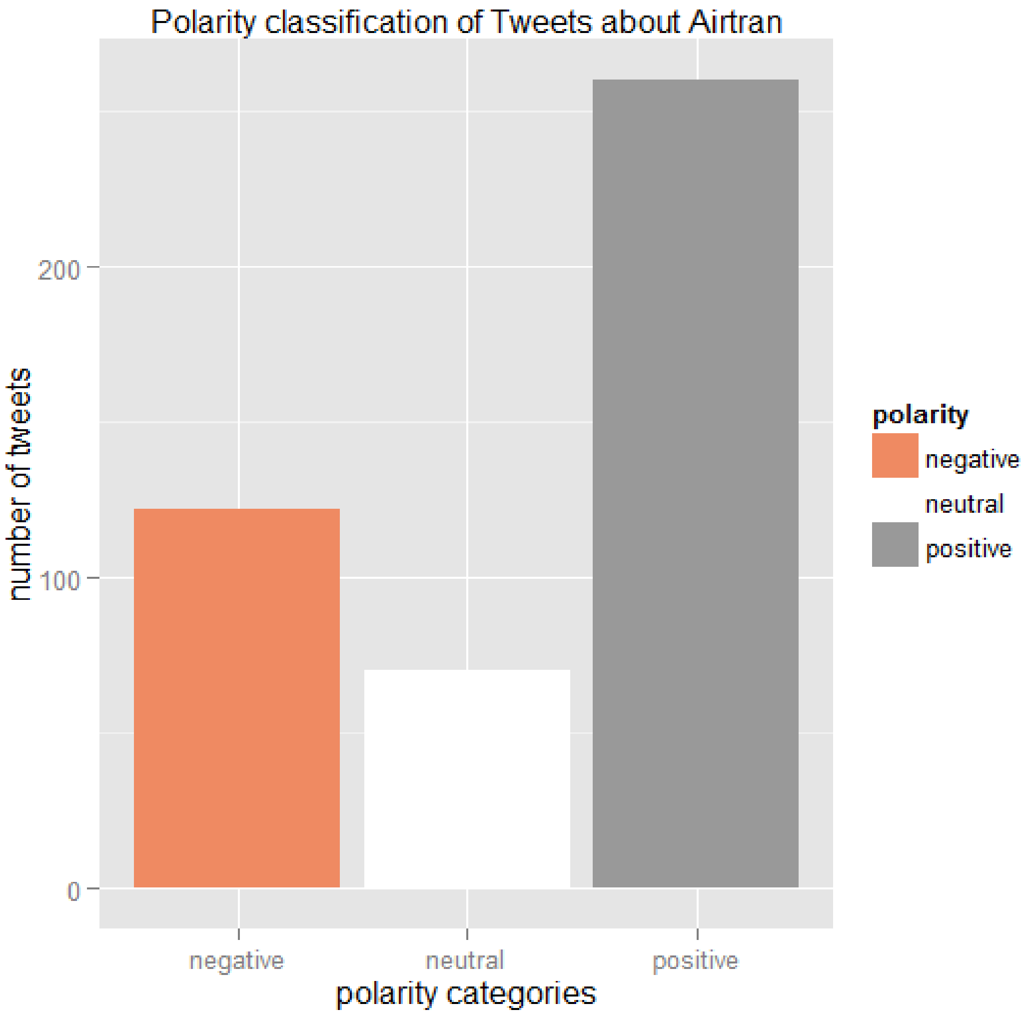
5.1.2. Sentiment Topic Recognition. The STR model used in conjunction with sentiment detection intends to reveal the underlying reason for each sentiment based on topics associated with the sentiment. In our STR model, topic words are extracted using CTM with the VEM algorithm and categorized under predefined or custom topics through semi-supervision. The CTM is an extension of the latent Dirichlet allocation (LDA) model where correlations between topics are allowed. In CTM, topic proportion exhibits correlation via the logistic normal distribution. The CTM uses an alternative and a more flexible distribution for the topic proportions that allows for covariance structure among the components. CTM offers a more realistic model of latent topic structure where the presence of one latent topic may be correlated with another to provide a better fit. CTM supports more topics and provides a natural way of exploring data. The method used for fitting the model is the VEM algorithm. Our STR model employs the R package topicmodels which currently provides an interface for fitting a CTM with the VEM algorithm as implemented by Blei and Lafferty (2006). For topicmodels a VEM algorithm is used instead of an ordinary EM algorithm because the expected complete likelihood in the E-step is still computationally intractable (Hornick and Grun 2011). Wainwright and Jordan (2008) provide a good introduction to variational inference.

5.2. CASE STUDY 1

Our first case study on STR involves classifying tweets for three airlines (AirTran, Frontier, and SkyWest) as positive, neutral or negative. We then use our proposed STR model to generate topics for each airline. The topics are then classified under the four predefined AQR categories (OT, DB, MB and CC). This case study uses the airline dataset from twitter described in section 3.1. To conduct our experiment, we used R to develop a prototype that supports the framework described in section 4.1. The results obtained are discussed under section 4.1. The twitter dataset discussed in section 3.1 was used in this case study.

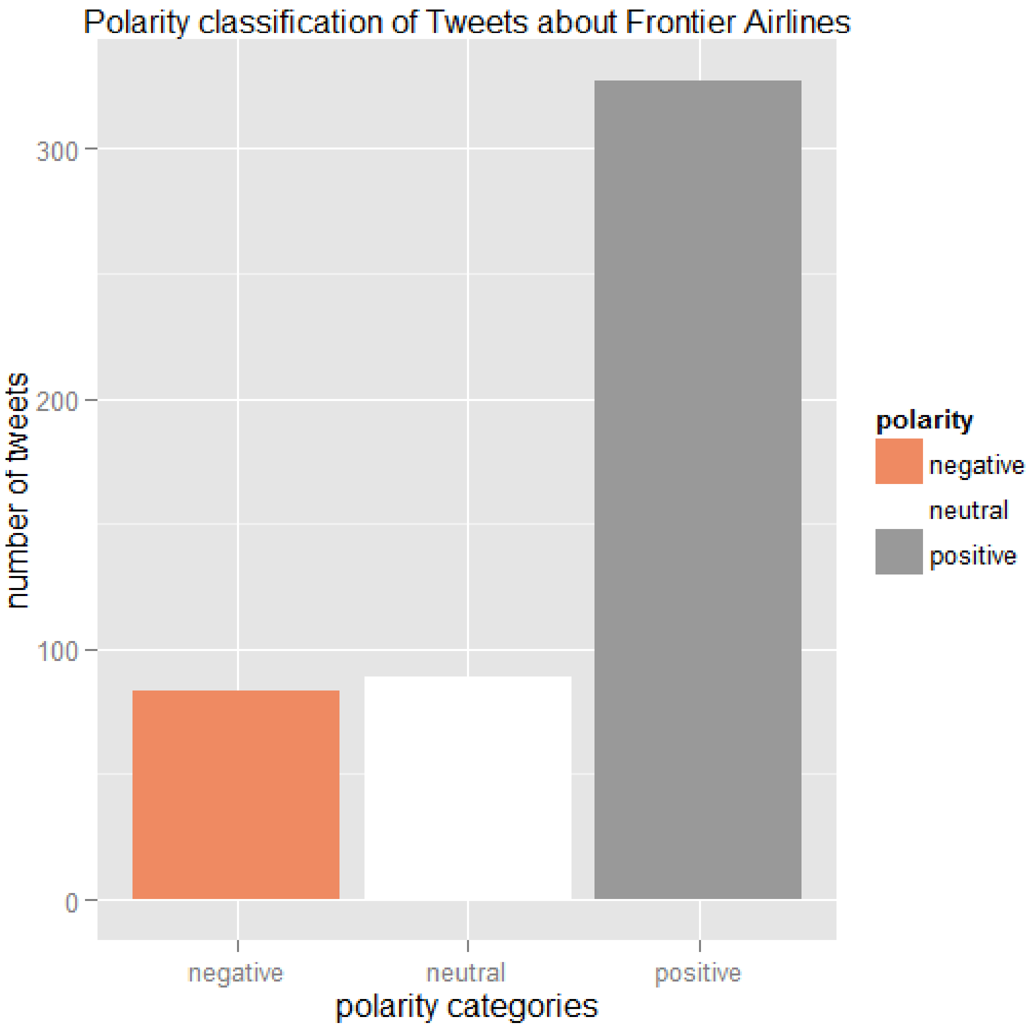
5.2.1. Sentiment Detection Results. In order to analyze consumers' sentiments toward the three airlines (AirTran, Frontier, SkyWest), we employed the lexicon-based algorithm. As noted in section 4.1.1, this algorithm performed better on subjective dataset and provided a polarity classification accuracy of 86.4%. Figure 5.2 shows the polarity classification results for each of the airlines.

From figure 5.2(a) it can be seen that there are more positive tweets for AirTran than negative tweets, approximately 57.5% and 27.6% respectively, the remaining being neutral. Frontier has approximately 64.1% positive, 18.0% negative and the rest neutral (see figure 5.2(b)). The overall sentiment score for SkyWest airline was highly positive at approximately 82% positive tweets. SkyWest has approximately 19.4% negative tweets and remaining tweets are neutral (see figure 5.2(c)).



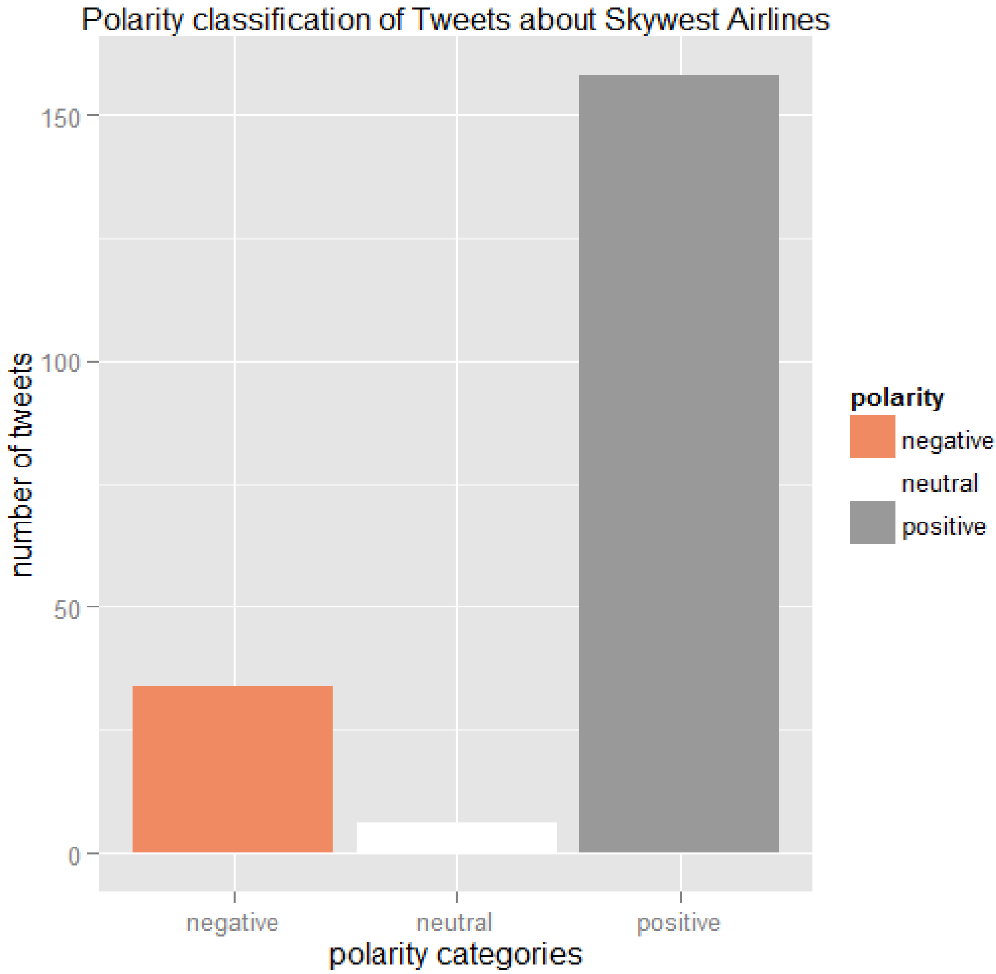
(a) AirTran

Figure 5.2 Polarity classification results.



(b) Frontier

Figure 5.2 Polarity classification results cont.



(c) SkyWest

Figure 5.2 Polarity classification results cont.

5.2.2. Sentiment Topic Recognition Results. As explained in section 4.1.2, our STR model employs the CTM with VEM algorithm. Our model produces a better comparative performance to other STR models because the dependency and correlation between sentiment topics are taken into consideration serving as an important function in sentiment analysis and STR (Lin et al, 2011). The STR model helps us to rightly categorize topic related terms used in the tweets data under each AQR criteria for positive and negative polarity. Figure 5.3 shows a snapshot of words generated for undefined topics using the STR model. The figure shows the five most frequent terms for each topic. These terms can be increased to include more term s and topics can be varied from 2 to 200 using 10-fold cross-validation.

```
     Topic 1   Topic 2   Topic 3   Topic 4
[1,] "can"     "airtran" "airtran" "airtran"
[2,] "airtran" "thing"   "ever"    "avgeek"
[3,] "flight"  "good"    "atl"     "boe"
[4,] "thank"   "time"    "pleas"   "airtranairway"
[5,] "find"    "class"   "flown"   "avigeek"
```

Figure 5.3 A snapshot of words in tweets used in building the lexicons.

Using the subjective tweets obtained from the sentiment detection process in section 5.1.1., we employed our sentiment topic matching algorithm to match tweets into the four AQR criteria. The results are shown in figure 5.4 and figure 5.5. The use of our STR algorithm helps us to rightly categorize topic related terms used in the tweets (shown in figure 5.3). The matching algorithm helps us to match those terms with the appropriate AQR criteria (shown in figure 5.4 and 5.5). The results shown in 5.4 and 5.5

shows an example of the STR model results however these results are not all meaningful because of the limitation in our dataset. For example, figure 5.5 $CustomerComplain produced a NULL result because there were not enough words used in the tweet dataset that match the terms generated by the STR algorithm. To resolve the limitation with our dataset, we employed the STR model on a big data described in case study 2, section 4.3.

```
$OnTime
 [1] "your"    "planes"  "early"   "upgrade" "upgrade" "planes"  "upgrade" "your"    "your"
[10] "upgrade" "your"

$DelayedBoarding
[1] "accommodate"

$MishandledBagagge
[1] "thing"

$CustomerComplaint
NULL
```

Figure 5.4 A snapshot of terms under each AQR criteria for positive polarity.

```
$OnTime
 [1] "your"    "planes"  "your"    "your"    "your"    "your"    "your"    "planes"  "early"
[10] "your"    "upgrade" "your"    "planes"  "gum"     "stuck"

$DelayedBoarding
NULL

$MishandledBagagge
[1] "abused" "bag"     "bag"     "thing"   "things"

$CustomerComplaint
NULL
```

Figure 5.5 A snapshot of terms under each AQR criteria for negative polarity.

5.2.3. Evaluation. The sentiment topic output lists from the STR model are used to compute the AQRs for the three airlines (AirTran Airways, Frontier and SkyWest Airlines) and the results are compared to the baseline AQR in Table 4.1. Table 5.1 shows the results of our AQR calculation per 1000 tweets.

As seen in Table 5.1, our approach produces results that mimic existing AQR results for these three airlines. The result shows that AirTran ranks first, followed by Frontier, and then SkyWest. This result demonstrates the effectiveness of our overall sentiment analysis approach in knowing the underlying reason for each sentiment based on topics associated with the sentiment. The performance of this approach is on par with the current AQR (Bowen & Headley 2013) commonly used to determine the reputations of U.S airlines.

Table 5.1 AQR Calculation per 1000 tweets.

Criteria	Number of terms per AQR criteria normalized 1000 tweets		
	AirTran	Frontier	SkyWest
On-Time	0.74	0.24	0.08
Denied Boarding	1.10	2.20	1.12
Mishandled Baggage	2.16	1.25	2.16
Customer Complaint	0.05	0.30	0.40
AQR	-0.63	-0.88	-0.90

5.3. CASE STUDY 2 AND RESULTS

While the results from the first case study in the previous section were encouraging, they had limitations because of small amount of data we had for the airline dataset from twitter. This second case study employs the same sentiment mining approach which detects sentiment polarity and sentiment topic on a big data in the retail domain. We used the Amazon data described in section 3.2. We employ the STR model on a customer review dataset from amazon.com to determine customer sentiments toward four product categories. The lexicon employed in case study 1, section 4.2.1 was employed in this case study as well.

In order to analyze consumers' sentiments toward the four product category, we employed the Naïve Bayes Algorithm. As noted in the prior section, this algorithm performed better on subjectivity dataset and provided a polarity classification accuracy of 86.4%. Figure 5.6 shows the polarity classification results for each of the airlines.

From figure 5.6, it can be seen that the customer review dataset has a little below 200, 000 negative reviews, about 610,000 positive reviews and below 100,000 reviews were neutral.

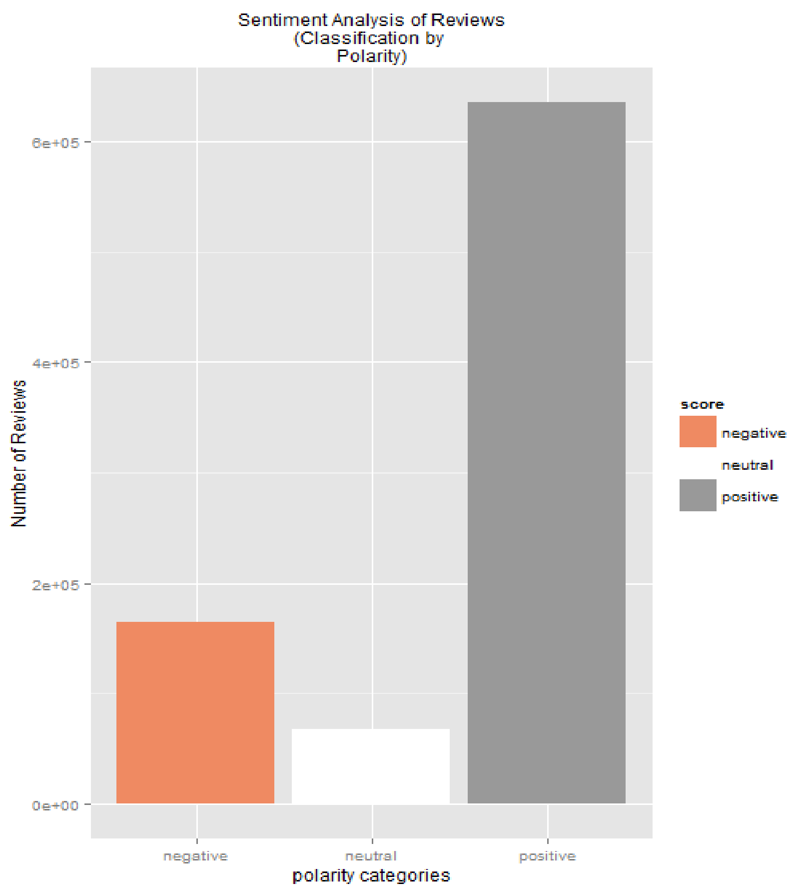


Figure 5.6 Polarity classification results for customer reviews on amazon.com.

6. SENTIMENT STRENGTH DETECTION

We propose a framework for calculating the sentiment strength of text. Sentiment strength here is the measure of satisfaction level a user has towards a particular product or service as stated in their reviews. The sentiment strength is measured on a scale of -5 to +5 (where -5 represent highly negative/dissatisfied and +5 represent highly positive/satisfied). Semantic orientation in sentiment mining refers to the polarity (positive, negative or neutral) and strength of words, phrases, or texts. Our concern is primarily with the strength of sentiment of text, and we extract the sentiment of words and phrases towards that goal.

6.1. FRAMEWORK

Figure 6.1 shows the framework for our proposed polarity strength algorithm. The different steps involved in this framework are explained in the subsections below.

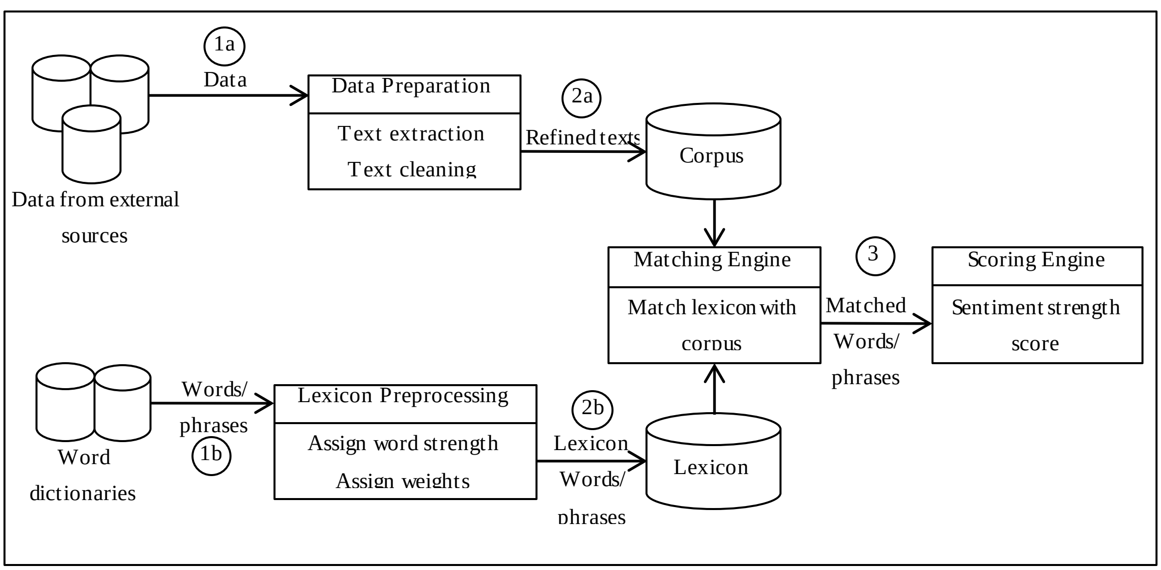


Figure 6.1 Framework for Sentiment Strength Detection.

In this framework, data is collected from source and prepared. Some examples of data preparation are described in sections 4.1 and 4.2. The prepared data is then stored as Corpus to be used in the polarity strength calculator.

6.1.1. Lexicon Preprocessing. The AFINN lexicon was employed to determine the weights of positive and negative words in each review. AFINN is a list of English words rated for valence with an integer between -5 and +5. The words were manually labeled by Finn Årup Nielsen in 2009-2011. AFINN contains 2477 words and phrases. The AFINN lexicon was combined with Hu and Liu's (2004) lexicon to conduct the analysis. The combined lexicon includes around 9277 seed adjectives with known orientation. The positive words/phrases in the Hu and Liu's (2004) lexicon were rated for valence with an integer of 2 while the negative words/phrases were assigned an integer of -2. In order to make use of the words of phrases provided in our text in our calculation of the strength of polarity, we made modification to the two lexicons mentioned above resulting in the combined lexicon. The numerical values were chosen to reflect both the prior polarity and the strength of the word/phrase. An example of the combined lexicon is shown in table 6.1. The integers assigned to the lexicon words/phrases were converted into weights from a scale of 0 to 1 (see table 6.1 for example). 0 was assigned to any word/phrase in the lexicon with integer polarity of -5. 0.1 was assigned to any word/phrase in the lexicon with integer polarity of -4, then 0.2 for -3 … and 1 for integer polarity of 5.

6.1.2. Matching Engine. An algorithm similar to the sentiment topic matching algorithm was employed to match words in the text to the combined lexicon. Anytime

Table 6.1 Example of word and their polarity value in the combined lexicon.

Words	**Value**	**Weighted Value**
Bastard	-5	0
Prick	-5	0
Bullshit	-4	0.1
Catastrophic	-4	0.1
Apathetic	-3	0.2
Abducted	-2	0.3
Abduction	-2	0.3
Absentee	-1	0.4
Absentees	-1	0.4
Aboard	1	0.6
Abilities	2	0.7
Absolve	2	0.7
Award	3	0.8
Classy	3	0.8
Awesome	4	0.9
Rolf	4	0.9
Breathtaking	5	1

there is a word match between the combined lexicon and the text, the corresponding valence integer in the combined lexicon is assigned to that word. The procedure is repeated for all words in the review and thereafter the valence integers are used to determine the degree of polarity for each review on the scale of -5 to +5.

6.1.3. Scoring Engine. This component in the framework calculates the final sentiment strength score of each text based on the weight of the words used in the text. Here we define the sentiment strength score calculation for each text as follows: Let R be a review text, $\boldsymbol{L_{pos}}$ be a positive lexicon and $\boldsymbol{L_{neg}}$ be a negative lexicon. Assume $\boldsymbol{P} = \{\boldsymbol{p_1}, \boldsymbol{p_2}, \dots, \boldsymbol{p_m}\}$ is the set of positive words in R that matches $\boldsymbol{L_{pos}}$ and $\boldsymbol{N} = \{\boldsymbol{n_1}, \boldsymbol{n_2}, \dots, \boldsymbol{n_q}\}$ is the set of negative words in R that matches $\boldsymbol{L_{neg}}$, then the “true” polarity of R will be

$$R^t = \left(\sum_{i=1}^{m} W(P_i) - \sum_{j=1}^{q} W(N_j) \right) / \left(size(P) + size(N) \right) \qquad (2)$$

Where $W(P_i)$ and $W(N_j)$ are the weights of P_i and N_j respectively.

The score assigned in the scoring engine to a text are in a weighted range of 0 to 1. These values are then converted into the -5 to +5 scale (see table 6.2) as their strength of sentiment polarity.

Table 6.2 Table for converting scoring engine values into strength of polarity of text.

Scoring Engine Value	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Strength of polarity	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5

6.2. CASE STUDY 3

This section provides a case study that demonstrates our approach by utilizing the framework described in section 6.1. The case study involves calculating the strength of polarity of customer review dataset. We use the amazon.com dataset for retail products described in section 4.2. Calculating the strength of sentiment polarity of these reviews by customers will help us determine the “true” customer feedback: how much satisfied are customers with a particular product or what is their degree of dissatisfaction about another product? We use our proposed framework in calculating the strength of sentiment polarity of the reviews on a scale of -5 to +5 (where -5 is highly negative and +5 is highly positive). To conduct our experiment, we used R to develop a prototype that supports the framework described in section 5.1. The results obtained are discussed under section 6.2.1.

6.2.1. Sentiment Strength Results. The results in the sentiment strength detection framework are explained in this section. The total count of reviews used in the case study was 866,056 (section 3.2). Figures 6.3 and 6.4 show a detail plot distribution of sentiment strength scores for negative and positive reviews respectively. These scores are the initial weight score obtained from the scoring engine plotted against the number of count of reviews. The graph in figure 6.2 represents the plot distribution of the sentiment strength scores from the scoring after converting the sentiment strength scores to a scale of -5 to +5 (see table 4). Out of this total count, table 6.3 shows the count of reviews for each sentiment strength (-5 to 5). The result shows that, the number of people highly dissatisfied (-5 score) about some particular product, 44540, are more than the number of people that are highly satisfied in the review, 35.

6.2.2. Evaluation. In our experiment we calculate the sentiment strength score of each review using formula 1. According to the calculated score, reviews that achieved a score of 0 to 0.4 are considered as negative and review that had a score of 0.6 to 1 are considered positive reviews (see table 6.2). The sentiment strength detection framework was tested on a set of data with known polarity orientation (positive, negative or neutral). Intuitively, results obtained from the sentiment strength framework as compared to the

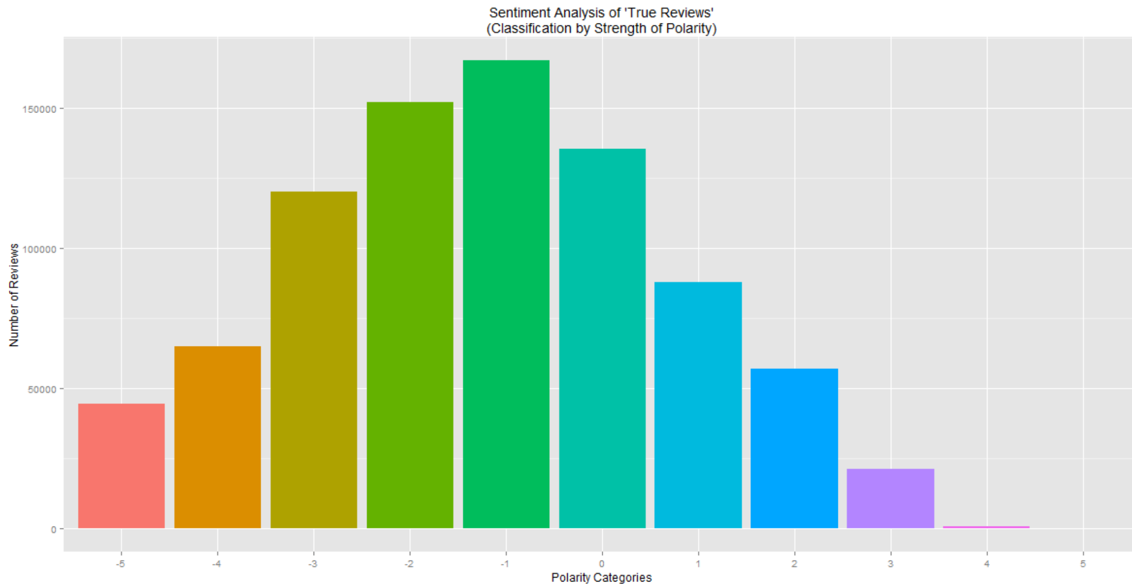


Figure 6.2 Plot distribution after converting the sentiment strength scores scale.

Table 6.3 Number of count of reviews for each sentiment strength.

Score	Count
-5	44540
-4	64797
-3	120069
-2	151871
-1	166981
0	135459
1	87677
2	56801
3	21272
4	634
5	35

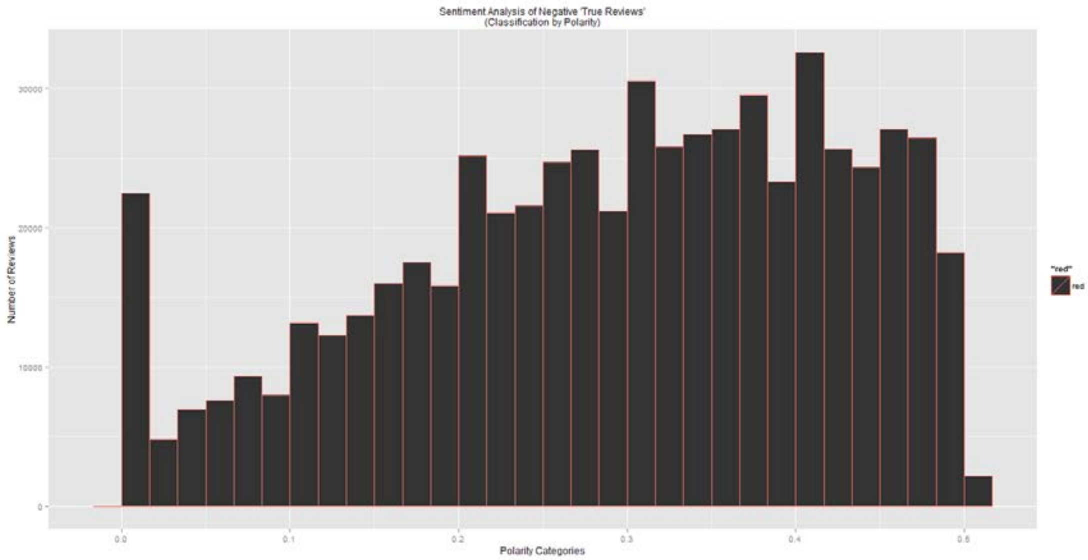


Figure 6.3 Plot distribution of the sentiment strength scores for negative reviews.

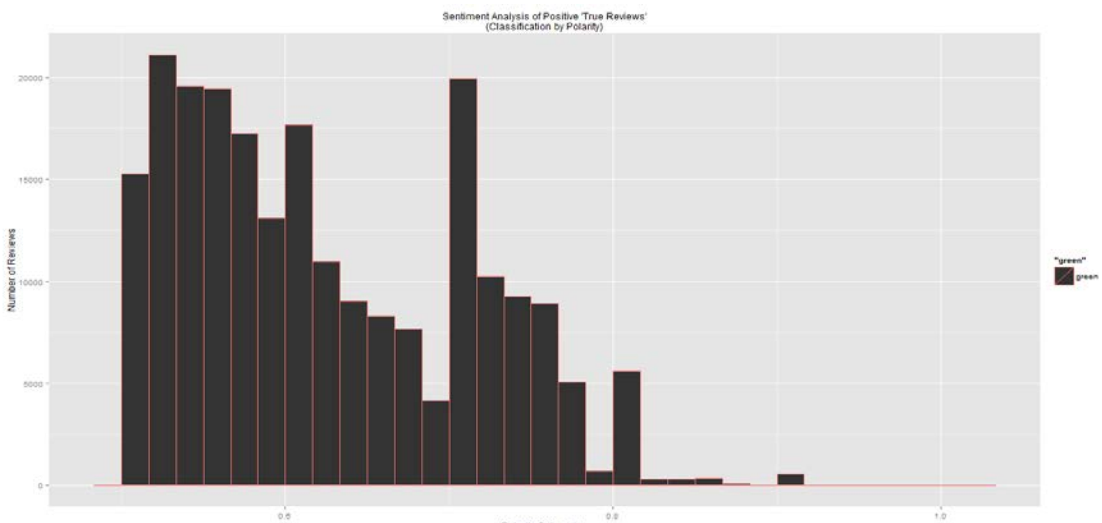


Figure 6.4 Plot distribution of the sentiment strength scores for positive reviews.

polarity results, are better in determining the "true" sentiment strength of a text and hence its semantic orientation (see table 6.4). Our approach of using a list of word/phrase strengths and identifying the strongest positive and negative words in any text enables us to correctly determine the sentiment strength of the text. For example in the first review below (table 6.4), the words matching the lexicon as shown in Table 6.5 contains both positive and negative words. Our lexicon based polarity detection algorithm determines the polarity of the review as neutral because the number of negative and positive words. However our "true" review approach examines the weight of the positive and negative to determine the actual degree of polarity for the review.

Table 6.4 Performance of sentiment strength framework against sentiment polarity.

	Review	**Polarity**	**strength (polarity)**
1	I love all things Scottish, but not this book. The descriptions were bland and the diction was horrid - at one point Carlow's hero describes his feelings about his beloved's appearance - Her figure was wonderful. What does that really mean? I have read much better books set in this time - try Diana Gabaldon's Outlander series, but avoid this book!	Neutral	-3 (Negative)
2	Fun, light fantasy. I really enjoyed this book. OK, so I am still not clear about how Kurland's time travel theory works -- I guess I like to see something more substantial than just wishing. I loved Richard myself. This book is sensual without being smutty -- sex is not described in graphic detail. I would feel comfortable recommending this book to my high school students. The medieval details were authentic and accurate. This book rates above most romance novels I have read.	Positive	5 (Positive)

Table 6.4 Performance of sentiment strength framework against sentiment polarity cont.

	Review	Polarity	strength (polarity)
3	I bought this book based on an Amazon customer's review and I wasn't disappointed. This book is very thorough and well-researched. Any writers seeking a reference for ships in the middle ages will not be disappointed by this book. There is a wealth of information, delivered in a readable manner.	Positive	5 (Positive)
4	This book was okay, but I have definitely read better novels about Scotland and the middle ages. The characters did not seem realistic, and I didn't feel there was enough description. I couldn't picture the Scottish countryside, though the descriptions of the interior of Castle Rock were fairly good.	Negative	1 (Positive)
5	In all, this book was very, very good, but not quite on a par with the rest of the series. It seemed to me that the characters were acting against their former natures - doing things they wouldn't do. In all, I do recommend it, and not just because you won't be able to follow the next book. Some of the scenes are breathtaking, and the writing is superb.	Positive	4 (Positive)

Table 6.5 Review 1 word match in lexicon, word strength and their weights.

Lexicon Match	***Word Strength***	***Weights***
Love	*2*	*0.7*
Wonderful	*4*	*0.9*
Better	*2*	*0.7*
Horrid	*-2*	*0.3*
Avoid	*-1*	*0.4*
Bland	*-2*	*0.3*

7. CONCLUSIONS

Sentiment mining has evolved from mere sentiment polarity detection into recognizing topics related to these sentiments. Our proposed approach on sentiment topic recognition captures user's sentiments and topics associated with such sentiments. In this way, each sentiment extracted by the approach has some underlying topic(s) and provides an overall knowledge and scope of the different consumer sentiment. The proposed approach aims at answering questions regarding the drivers of each labeled sentiment in a dataset and examines the overall breadth of the sentiment. We show how the proposed STR model can be used to compute airline reputation (AQR) for three major airlines (AirTran Airways, Frontier and SkyWest Airlines). Experimental results show that our proposed approach compared to the current method of computing AQR yields equivalent results for airline ranking and is less expensive - AQRs are currently computed through surveys using the U.S. Department of Transportation's monthly Air Travel Consumer Report (Bowen and Headley, 2013). Although the approach yielded an encouraging evaluation for the calculation of AQRs some of the initial results in topic term categorization were less meaningful. We therefore, applied our approach to another domain of retail products on customer review data to evaluate the approach. Our approach achieves a good performance when evaluated on the big dataset.

We also proposed a framework for analyzing the sentiment strength of text. We first extracted the words which reflect sentiment strength in texts and match the words with a predefined lexicon. We then calculated the sentiment strength of the text based on the weight of word-match in the lexicon. Experimental results show that our proposed approach is efficient and performs intuitively better than that without considering the sentiment strength of words/phrases. The results can help companies to determine the different satisfaction levels that users have in regards to a product. When this framework is combined with STR, organization can further verify not only satisfaction/ dissatisfaction levels but also the different areas or specific topics toward which uses expressed such sentiment strengths. We realized that both the STR approach and the sentiment strength detection framework performed better on a larger data set and produced more meaningful results.

In terms of future work, the next step will be to evaluate the effectiveness and scalability of this framework using other evaluation methods. It should also be noted that, while the lexicon-based approach used in sentiment detection can detect basic sentiments, it may sometimes be inadequate in detecting figurative expression such as irony or provocation. Future research would attempt to provide solutions to these limitations.

BIBLIOGRAPHY

Aggarwal, C. C., & Zhai, C. (2012). A survey of text clustering algorithms. In Mining Text Data (pp. 77-128). Springer US.

Bai, X. (2011). Predicting consumer sentiments from online text. Decision Support Systems, 50(4), 732-742.

Blaschke, C., & Valencia, A. (2002). The frame-based module of the Suiseki information extraction system. Intelligent Systems, IEEE, 17(2), 14-20.

Blei, D., & Lafferty, J. (2006). Correlated topic models. Advances in neural information processing systems, 18, 147.

Bowen, B. D. and Headley, D. E., (2012, April). Airline Quality Rating 2012. (22nd ed.) (Online). Available: http://www.airlinequalityrating.com/reports/2012aqr.pdf.

Cai, K., Spangler, S., Chen, Y., & Zhang, L. (2010). Leveraging sentiment analysis for topic detection. Web Intelligence and Agent Systems, 8(3), 291-302.

Califf, M. E., & Mooney, R. J. (2003). Bottom-up relational learning of pattern matching rules for information extraction. The Journal of Machine Learning Research, 4, 177-210.

Chen, J., Huang, H., Tian, S., & Qu, Y. (2009). Feature selection for text classification with Naïve Bayes. Expert Systems with Applications, 36(3), 5432-5435.

Conrad, J. G., & Schilder, F. (2007, June). Opinion mining in legal blogs. In Proceedings of the 11th international conference on Artificial intelligence and law (pp. 231-236). ACM.

De Carvalho, F. D. A., & Lechevallier, Y. (2009). Partitional clustering algorithms for symbolic interval data based on single adaptive distances. Pattern Recognition, 42(7), 1223-1236.

Delen, D., & Crossland, M. D. (2008). Seeding the survey and analysis of research literature with text mining. Expert Systems with Applications, 34(3), 1707-1720.

Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. Communications of the ACM, 49(9), 76-82.

Forman, G., Kirshenbaum, E., & Suermondt, J. (2006, August). Pragmatic text mining: minimizing human effort to quantify many issues in call logs. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 852-861). ACM.

Frank, E., & Bouckaert, R. R. (2006). Naive bayes for text classification with unbalanced classes. In Knowledge Discovery in Databases: PKDD 2006 (pp. 503-510). Springer Berlin Heidelberg.

Grune, D., & Jacobs, C. J. (2008). Parsing Techniques 2: A Practical Guide. Springer.

Hornik, K., & Grün, B. (2011). topicmodels: An R package for fitting topic models. Journal of Statistical Software, 40(13), 1-30.

Hotho, A., Nürnberger, A., & Paaß, G. (2005, May). A Brief Survey of Text Mining. In Ldv Forum (Vol. 20, No. 1, pp. 19-62).

Hotho, A., Nürnberger, A., & Paaß, G. (2005, May). A Brief Survey of Text Mining. In Ldv Forum (Vol. 20, No. 1, pp. 19-62).

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

Hung, S. S., Kuo, T. C., & Liu, D. S. M. (2004, January). An Efficient Clustering Algorithm for Patterns Placement in Walkthrough System. In Knowledge-Based Intelligent Information and Engineering Systems (pp. 1237-1244). Springer Berlin Heidelberg.

Java, A., Song, X., Finin, T., & Tseng, B. (2007, August). Why we twitter: understanding microblogging usage and communities. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis (pp. 56-65). ACM.

Jiang, J., & Zhai, C. (2007). An empirical study of tokenization strategies for biomedical information retrieval. Information Retrieval, 10(4-5), 341-363.

Joachims, T. (1999). Making large scale SVM learning practical.

Kalakota, R. (2011, November). Big data infographic and Gartner 2012 top 10 strategic tech trends. http://practicalanalytics.wordpress.com/2011/11/11/big-data-infographic-and-gartner-2012-top-10-strategic-tech-trends/.

Kanya, N., & Geetha, S. (2007). Information extraction-a text mining approach. International Conference on Information and Communication Technology in Electrical Science 1111-1118.

Kim, S. B., Han, K. S., Rim, H. C., & Myaeng, S. H. (2006). Some effective techniques for naive bayes text classification. Knowledge and Data Engineering, IEEE Transactions on, 18(11), 1457-1466.

Kogan, J., Nicholas, C. K., & Teboulle, M. (Eds.). (2006). Grouping multidimensional data. Springer.

Lewis, D. D., & Ringuette, M. (1994, April). A comparison of two learning algorithms for text categorization. In Third annual symposium on document analysis and information retrieval (Vol. 33, pp. 81-93).

Liddy, E. D. (2000). Text mining. Bulletin of the American Society for Information Science and Technology, 27(1), 13-14.

Lin, C., He, Y., Everson, R., & Ruger, S. (2012). Weakly supervised joint sentiment-topic detection from text. Knowledge and Data Engineering, IEEE Transactions on, 24(6), 1134-1145.

Lin, X., Wang, W., & Wu, B. (2011, July). A complementary method to determine semantic orientations of words based on WordNet. In Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on (Vol. 3, pp. 1738-1740). IEEE.

Liu, L., Shi, J., & Liu, X. (2010, December). Web Information Extraction Algorithm Based on Ontology and DOM Tree. In Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on (pp. 1-4). IEEE.

Liu, Y., Loh, H. T., & Sun, A. (2009). Imbalanced text classification: A term weighting approach. Expert systems with Applications, 36(1), 690-701.

Manjula, D., Aghila, G., & Geetha, T. V. (2003, April). Document knowledge representation using description logics for information extraction and querying. In Information Technology: Coding and Computing (Computers and Communications), 2003. Proceedings. ITCC 2003. International Conference on (pp. 189-193). IEEE.

Miao, Q., Li, Q., & Dai, R. (2009). AMAZING: A sentiment mining and retrieval system. Expert Systems with Applications, 36(3), 7192-7198.

Miner, G., Delen, D., Elder, J., Fast, A., Hill, T., & Nisbet, B. (2012). Practical text mining and statistical analysis for non-structured text data applications. Amsterdam, The Netherlands: Academic Press.

Mitchell, T. M. (1999). Machine learning and data mining. Communications of the ACM, 42(11), 30-36.

Mooney, R. (1999). Relational learning of pattern-match rules for information extraction. In Proceedings of the Sixteenth National Conference on Artificial Intelligence (pp. 328-334).

Mooney, R. J., & Bunescu, R. (2005). Mining knowledge from text using information extraction. ACM SIGKDD explorations newsletter, 7(1), 3-10.

Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. Expert Systems with Applications, 40(10), 4241-4251.

Mullen, T., & Malouf, R. (2006). A Preliminary Investigation into Sentiment Analysis of Informal Political Discourse. In AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs (pp. 159-162).

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2011). SentiFul: A lexicon for sentiment analysis. Affective Computing, IEEE Transactions on, 2(1), 22-36.

Padmaja, S., & Fatima, S. S. (2013). Opinion Mining and Sentiment Analysis-An Assessment of Peoples' Belief: A Survey. International Journal of Ad Hoc, Sensor & Ubiquitous Computing, 4(1).

Pak, A., & Paroubek, P. (2010). Twitter based system: Using Twitter for disambiguating sentiment ambiguous adjectives. In Proceedings of the 5th International Workshop on Semantic Evaluation (pp. 436-439). Association for Computational Linguistics.

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd annual meeting on Association for Computational Linguistics (p. 271). Association for Computational Linguistics.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2), 1-135.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10 (pp. 79-86). Association for Computational Linguistics.

Quirk, R., Crystal, D., & Education P. (1985). A comprehensive grammar of the English language. Vol. 397. London: Longman.

Sarawagi, S. (2008). Information extraction. Foundations and trends in databases, 1(3), 261-377.

Sebastiani, F. (2002). Machine learning in automated text categorization. ACM computing surveys (CSUR), 34(1), 1-47.

Sebastiani, F. (2005). Text Categorization.

Sebastiani, F. (2006). Classification of text, automatic. The Encyclopedia of Language and Linguistics, 14, 457-462.

Song, Y., Kołcz, A., & Giles, C. L. (2009). Better Naive Bayes classification for high-precision spam detection. Software: Practice and Experience, 39(11), 1003-1024.

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

Tan, A. H. (1999, April). Text mining: The state of the art and the challenges. In Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases (pp. 65-70).

Tan, A. H. (1999, April). Text mining: The state of the art and the challenges. In Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases (pp. 65-70).

Tan, S. (2005). Neighbor-weighted k-nearest neighbor for unbalanced text corpus. Expert Systems with Applications, 28(4), 667-671.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 61(12), 2544-2558.

Thomas, M., Pang, B., & Lee, L. (2006, July). Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In Proceedings of the 2006 conference on empirical methods in natural language processing (pp. 327-335). Association for Computational Linguistics.

Wainwright, M. J., & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. Foundations and Trends® in Machine Learning, 1(1-2), 1-305.

Wang, S., & Manning, C. D. (2012, July). Baselines and bigrams: Simple, good sentiment and topic classification. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2 (pp. 90-94). Association for Computational Linguistics.

Whitehead, M., & Yaeger, L. (2009, March). Building a general purpose cross-domain sentiment mining model. In Computer Science and Information Engineering, 2009 WRI World Congress on (Vol. 4, pp. 472-476). IEEE.

Whitelaw, C., Garg, N., & Argamon, S. (2005, October). Using appraisal groups for sentiment analysis. In Proceedings of the 14th ACM international conference on Information and knowledge management (pp. 625-631). ACM.

Wiener, E., Pedersen, J. O., & Weigend, A. S. (1995, April). A neural network approach to topic spotting. In Proceedings of SDAIR-95, 4th annual symposium on document analysis and information retrieval (pp. 317-332).

Yang, Y., & Chute, C. G. (1994). An example-based mapping method for text categorization and retrieval. ACM Transactions on Information Systems (TOIS), 12(3), 252-277.

Yang, Y., & Pedersen, J. O. (1997, July). A comparative study on feature selection in text categorization. In ICML (Vol. 97, pp. 412-420).

Yi, J., & Niblack, W. (2005). Sentiment mining in WebFountain. In Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on (pp. 1073-1083). IEEE.

Zhang, M. L., Peña, J. M., & Robles, V. (2009). Feature selection for multi-label naive Bayes classification. Information Sciences, 179(19), 3218-3229.

Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. Expert Systems with Applications, 38(6), 7674-7682.

Zhao, T., Li, C., Ding, Q., & Li, L. (2012). User-sentiment topic model: refining user's topics with sentiment information. In Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics (p. 10). ACM.

VITA

Esi Adeborna was born on February 9, 1986 at Aflao, Ghana. She received her Bachelor of Science degree in Computer Science from the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana in 2007.

She was a graduate student in the Business Department at University of Missouri, St. Louis beginning August 2008. She worked as a Graduate Assistant under Dr. Michael Elliot from January 2009 to December 2010 where she received her Master of Business Administration (MBA) in December 2010.

She also received her MS in Information Science and Technology at the Missouri University of Science and Technology in December 2015 after working for Dr. Keng Siau, advisor, as a graduate research assistant from August 2013 – August 2014.