2014

AN APPROACH TO SENTIMENT ANALYSIS –THE CASE OF AIRLINE QUALITY RATING

Esi Adeborna
Missouri University of Science and Technology

Keng Siau
Missouri University of Science and Technology

AN APPROACH TO SENTIMENT ANALYSIS – THE CASE OF AIRLINE QUALITY RATING

Esi Adeborna, Missouri University of Science and Technology, USA

Keng Siau, Missouri University of Science and Technology, USA

Abstract

Sentiment mining has been commonly associated with the analysis of a text string to determine whether a corpus is of a negative or positive opinion. Recently, sentiment mining has been extended to address problems such as distinguishing objective from subjective propositions, and determining the sources and topics of different opinions expressed in textual data sets such as web blogs, tweets, message board reviews, and news. Companies can leverage opinion polarity and sentiment topic recognition to gain a deeper understanding of the drivers and the overall scope of sentiments. These insights can advance competitive intelligence, improve customer service, attain better brand image, and enhance competitiveness. This research paper proposes a sentiment mining approach which detects sentiment polarity and sentiment topic from text. The approach includes a sentiment topic recognition model that is based on Correlated Topics Models (CTM) with Variational Expectation-Maximization (VEM) algorithm. We validate the effectiveness and efficiency of this model using airline data from Twitter. We also examine the reputation of three major airlines by computing their Airline Quality Rating (AQR) based on the output from our approach.

Keywords: Sentiment mining, Sentiment Topic Recognition, Business Intelligence, Data Science, Airline Quality Rating.

1. INTRODUCTION

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. 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).

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).

In this study, we propose a sentiment mining approach that combines sentiment polarity detection with STR. The proposed approach is to serve as an intelligent tool to answer questions regarding the drivers and overall scope of sentiments.

The main contributions of this research are summarized below:

a) The approach includes an STR model that uses Correlated Topics Models (CTM) (Blei and Lafferty 2006) with Variational Expectation-Maximization (VEM) algorithm.
b) 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.
c) A prototype is developed and an example of how the approach is applied is illustrated using a case study with real-world tweets.
d) An algorithm to match opinionated tweets to a topic lexicon is developed.
e) We performed an evaluation on the prototype developed. We compare our AQR results with those obtained from the Airline Quality Rating Report (Bowen & Headley 2013).

This paper is organized as follows: Section 2 provides a literature review on STR in sentiment mining. Following that, we present and discuss the proposed approach in section 3 In section 4, we report an experiment and discuss its results. The evaluation phase of this research is presented in section 5. Finally, we conclude the paper and highlight future research in section 6.

2. LITERATURE REVIEW

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) build 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. APPROACH

Figure 1 shows the framework for our proposed approach. The different steps involved in this framework are explained in the subsections below.

3.1 Data Preparation

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). Based on Table 1, 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.

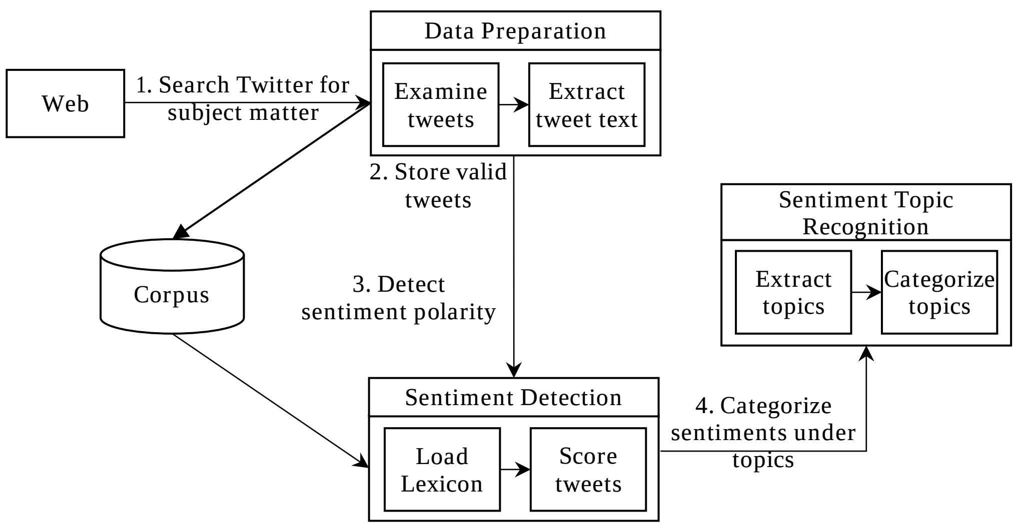


Figure 1. Framework for Sentiment Detection and STR System

Airline Quality Rating Scores														
	2012 AQR		**2011 AQR**		**2010 AQR**		**2009 AQR**		**2008 AQR**		**2007 AQR**		**2006 AQR**	
	Score	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**
Air Tran	**-0.51**	**3**	-0.48	1	-0.48	1	-0.49	2	-0.84	2	-1.03	1	-1.13	3
Alaska	**-0.77**	**6**	-0.79	5	-0.94	4	-1.39	11	-1.16	5	-1.75	7	-1.66	9
American	**-1.11**	**10**	-1.24	10	-1.28	11	-1.25	9	-1.71	9	-2.19	9	-1.83	10
American Eagle	**-1.78**	**11**	-2.51	15	-2.82	16	-2.83	18	-3.12	16	-3.80	15	-3.97	17
Delta	**-0.58**	**4**	-0.80	6	-1.22	7	N/A	-	N/A	-	N/A	-	N/A	-
ExpressJet	**-1.95**	**13**	N/A	-	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	-1.30	4
Hawaiian	**-0.71**	**5**	-0.59	2	-0.58	2	-0.40	1	-0.69	1	N/A	-	N/A	-
JetBlue	**-0.43**	**2**	-0.60	3	-0.70	3	-0.62	3	-0.90	3	-1.30	2	-0.93	2
SkyWest	**-1.88**	**12**	-1.15	9	-1.28	10	-1.57	14	-2.13	13	-3.09	13	-2.76	14
Southwest	**-0.81**	**8**	-0.93	7	-1.01	5	-1.00	5	-1.23	6	-1.59	3	-1.38	6
United	**-2.18**	**14**	N/A	-	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	-2.32	13
Virgin America	**-0.35**	**1**	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-	N/A	-
Industry	**-1.11**		**-1.08**		**-1.20**		**-1.27**		**-1.63**		**-2.16**		**-1.87**	

Table 1. AQR Ranking Table: Air Quality Rating Scores from 2006 to 2012 (Bowen & Headley 2013)

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.

3.2 Lexicon

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.

We employed a modification of this lexicon in our STR model by adding words through a thorough search in WordNet based on AQR criteria (Bowen & Headley, 2013). Our lexicon modification is based on (Neviarouskaya et al. 2011; Taboada et al. 2011). The process automatically extracts words or compound words from WordNet that are related to the following AQR criteria; On-time, Denied boarding, Mishandled baggage, and Customer complaints. In addition, to ensure maximum results from the algorithms employed in our model, we extracted terms from the tweets and included them in the lexicon using the CTM with VEM algorithm. The terms here refer to topic related words for each AQR criterion, which were obtained after stemming and removing stop-words from the tweets.

3.3 Sentiment Detection

Determining sentiment polarity is done by comparing the tweets against a predefined corpus of subjective words. 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). Among these algorithms, Naïve Bayes is popular in sentiment classification due to its computational efficiency and relatively good predictive performance (Chen et al. 2009). There are many prior literatures about classifiers that use the Naïve Bayes algorithm (Zhang et al. 2009; Song et al 2009; Frank & Bouchaert 2006; Kim et al. 2006). Naïve Bayes is a simple classification method based on

Bayes rule and on the representation documents as a bag of word using a subset of words. It predicts by reading a set of examples and then uses the Bayes theorem to evaluate the posterior probability of all qualifications. For each instance, the highest posterior probability is chosen.

Naïve Bayes assumes that the feature values for each variable on a given class or category are conditionally independent. This significantly reduces the computational cost (Mitchell 1999). Therefore, we employ Naïve Bayes algorithm for our sentiment analysis using the R[1] 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.

3.4 Sentiment Topic Recognition Model

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 the four main criteria used in AQR 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 & Grun 2011). Wainwright and Jordan (2008) provide a good introduction to variational inference.

3.5 Sentiment Topic Matching Algorithm

The *Sentiment Topic Matching* algorithm is used to match terms relating to some specific sentiment topics based on tweets. The idea behind this algorithm is to find those terms that relate to a topic sentiment with respect to the topic sentiment lexicon.

3.6 Airline Quality Rating (AQR)

The AQR is an objective method for assessing airline quality by combining multiple performance criteria (Bowen & Headley 2013). The formula for calculating the AQR score is:

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

Where OT (on-time), DB (denied boarding), MB (mishandled baggage), and CC (customer complaints) are variables considered as shown in Table 2. Data for all criteria is drawn from the U.S. Department of Transportation's monthly Air Travel Consumer Report[2] (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

---

[1] http://www.r-project.org/

[2] http://dot.gov/airconsumer/

used in this research are based on April 2013 reported values. Higher AQR values indicate higher reputation (as shown in Table 1). Virgin America (VX) for example, had the best rating in 2012 with an AQR value of -0.35 (see Table 1).

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

Table 2. AQR, Criteria, Weight and Impact (Bowen and Headley, 2012)

4. CASE STUDY AND RESULTS

Our case study involves classifying tweets for three airlines (AirTran, Frontier, 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 AQR categories (OT, DB, MB and CC). To conduct our experiment, we used R to develop a prototype that supports the approach described in section 3. The results obtained are discussed under section 4.1. The tweets used in this experiment contain 452 tweets on AirTran, 499 on Frontier Airlines and 195 on SkyWest Airlines. Each tweet contains some comment made about each of these airlines; positive, negative or neutral. The tweets collected are prepared using the data preparation process explained in section 3.1.

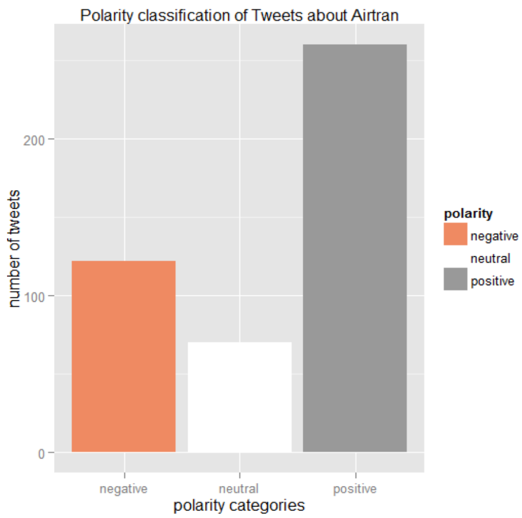
4.1 Sentiment Detection Results

In order to analyze consumers' sentiments toward the three airlines (AirTran, Frontier, SkyWest), 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 2 shows the polarity classification results for each of the airlines.

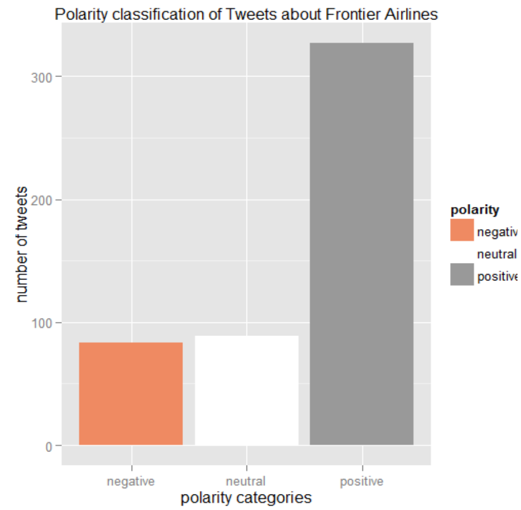
From figure 2(a) it can be seen that there are more positive tweets for AirTran than negative tweets, approximately 57.5% and 27.6% respectively, the remaining being neutral. Frontier has approximately 64.1% positive, 18.0% negative and the rest neutral (see figure 2(b)). The overall sentiment score 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 2(c)).

4.2 Sentiment Topic Recognition Results

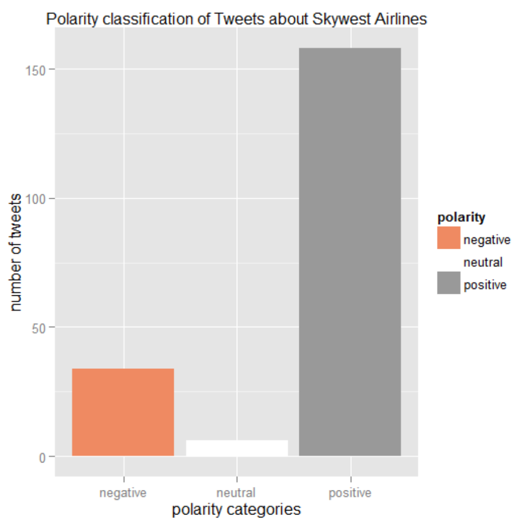
As explained in section 3.4, our STR model employs the CTM with VEM algorithm. The model generates terms from the airline tweets which are used in building the lexicon for each AQR category. In total, we came up with four lexicons; on-time, denied boarding, mishandled baggage and customer complaint lexicons. 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.



(a)



(b)



(c)

Figure 2. *Polarity classification results for AirTran (top left), Frontier (top right) and SkyWest (bottom). Y-axis provides the number of tweets and the X-axis represent the polarity categories (negative, neutral, positive).*

5. 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 1. Table 3 shows the results of our AQR calculation per 1000 tweets.

As seen in Table 3, 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.

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

Table 3. *AQR Calculation: The table shows the AQR calculation for each airline using the rates of each AQR criteria per 1000 tweets.*

6. CONCLUSIONS

Sentiment mining has evolved from mere sentiment polarity detection into recognizing topics intrinsic to these sentiments. Our proposed approach concurrently captures user's sentiments and topics intrinsic to 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). The AQR are currently computed using the U.S. Department of Transportation's monthly Air Travel Consumer Report (Bowen and Headley, 2013).

We tested our approach on a limited number of tweets and the results are very encouraging. We are continuously collecting tweets on the three airlines and we will re-test our approach with a larger data set in the near future. We are also applying our approach to other domains to evaluate the proposed approach. It should 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 should attempt to provide solutions to these limitations.

7. REFERENCES

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

Bowen, B. D. and Headley, D. E., (2013, 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.

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.

Dodds, P. S., & Danforth, C. M. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and Presidents. Journal of Happiness Studies, 11(4), 441–456.

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.

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

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.

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.

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

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

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).

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

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.

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

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

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. (2008). Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2), 1-135.

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., & 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.

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, 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.

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.

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).

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

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.