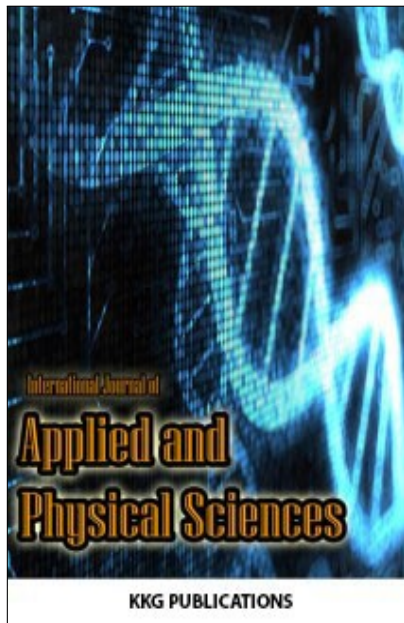


This article was downloaded by:

Publisher: KKG Publications

Registered office: 18, Jalan Kenanga SD 9/7 Bandar Sri Damansara, 52200 Malaysia

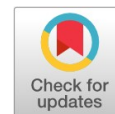


Key Knowledge Generation

Publication details, including instructions for author and subscription information:

<http://kkgpublications.com/applied-sciences/>

Sentiment Analysis Applied to Airline Feedback to Boost Customers' Endearment



AROCKIA XAVIER ANNIE R ¹, VIGNESH MOHAN ²,
SREE HARISSH VENU ³

^{1, 2, 3} College of Engineering, Guindy, India

Published online: 27 July 2016

To cite this article: A. X. Annie R, V. Mohan and S. H. Venu, "Sentiment analysis applied to airline feedback to boost customers' endearment," *International Journal of Applied and Physical Sciences*, vol. 2, no. 2, pp. 51-58, 2016.

DOI: <https://dx.doi.org/10.20469/ijaps.2.50004-2>

To link to this article: <http://kkgpublications.com/wp-content/uploads/2016/2/Volume2/IJAPS-50004-2.pdf>

PLEASE SCROLL DOWN FOR ARTICLE

KKG Publications makes every effort to ascertain the precision of all the information (the "Content") contained in the publications on our platform. However, KKG Publications, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the content. All opinions and views stated in this publication are not endorsed by KKG Publications. These are purely the opinions and views of authors. The accuracy of the content should not be relied upon and primary sources of information should be considered for any verification. KKG Publications shall not be liable for any costs, expenses, proceedings, loss, actions, demands, damages, expenses and other liabilities directly or indirectly caused in connection with given content.

This article may be utilized for research, edifying, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly verboten.

SENTIMENT ANALYSIS APPLIED TO AIRLINE FEEDBACK TO BOOST CUSTOMERS' ENDEARMENT

AROCKIA XAVIER ANNIE R ¹, VIGNESH MOHAN ², SREE HARISSH VENU ^{3*}

^{1, 2, 3} College of Engineering, Guindy, India

Keywords:

Pile
Bearing Capacity
Artificial Neural Networks
Finite Element Method

Received: 28 August 2015

Accepted: 12 January 2016

Published: 27 July 2016

Abstract. Customers differ greatly in their demographics, lifestyles, needs, perceptions, preferences, and behaviors. A business entity needs to understand the profitability of customers in a segment and their potential lifetime profitability. In this paper, our focus is on applying sentiment analysis to analyze feedback of passengers obtained from the airline forum. For this purpose, Multinomial Naive Bayes and Linear Support Vector models are used. Training data consisted of 1217 positive reviews and 955 negative reviews. Sentiments were predicted for 868 reviews. This work also aims to find a suitable data model that achieves a high accuracy and dependence on various pre-processing approaches. The sentiment analysis result is plotted as a bar graph visualization and evaluated against the overall trip rating obtained from the forum. We want to help the airline industry maximize the delivery and service to meet customer expectations and build customer loyalty.

©2016 KKG Publications. All rights reserved.

INTRODUCTION

The Web contains rich source of sentiments/opinions in the form of movie review, product review, social discussions, trip review, etc. These data are gold mines for business analyst in taking decisions. The prime difference between computer and human is that the latter can perceive emotions while the former cannot. If it is made possible for computers to computationally understand sentiments/opinions, then it can be used to provide information about the likes and dislikes of users.

The airline industry is a billion dollar industry and millions of passengers use this transportation service daily for transportation from one place to another. Due to the increasing competition, a lot of start-ups as well as established airline companies are finding it difficult to survive in the current market. Building customer loyalty is posing a great challenge these days and is also the major concern among companies. For all these reasons, it is vital to provide good service, value feedback of passengers and satisfy the passenger's needs. In the past passenger reviews were manually read, analyzed and categorized which demanded a lot of effort, labor and time.

Sentiment analysis of text is the employment of natural language processing and computational linguistics methods to classify the polarity of text at the document, sentence and aspect level. Typically, two main approaches are used for sentiment analysis of text. First being the lexical analysis approach, where dictionaries are built. The second approach being the machine learning approach where feature vectors are built.

The study by [1] classifies movie reviews as positive or negative using supervised classification techniques and compares the sentiment classification with the traditional topic level detection. The study by [2] examines sentiment analysis of tweets and uses pos-specific prior polarity features and partial tree kernel features for sentiment classification. The study by [3] classifies movie reviews by subjectivity/objectivity and negative/positive attitude using machine learning techniques - Naive Bayes, Decision Tree, Maximum Entropy, K-Means Clustering. The study by [4] focuses on using sentiment mining approach to detect sentiment polarity and sentiment topic detection from airline tweets.

Minimal research has been made on the sentiment analysis of airline review provided by passengers. Existing works have used reviews obtained from twitter where character limit is 140 hindering individual passenger to convey his/her entire experience. Also, abbreviations are commonly used in tweets making it difficult to perform sentiment analysis.

In our work, we develop a sentiment analysis engine that uses different pre-processing strategies and machine learning approach to determine the polarity of passenger's review. The main contribution of our work can be summarized as below:

- The approach uses passenger's review collected from popular airline forum rather than from twitter.
- Developed pre-processing strategy to handle removal of country names, date, baggage weight and flight identi-

*Corresponding author: Sree Harissh Venu

†Email: vharissh14@gmail.com



fication numbers using regular expression. Study relating variation of accuracy, number of features and pre-processing approach is made.

- Representation of tokens as sparse matrix and pipelining is applied to boost the performance of the system.
- Performance of machine learning algorithms through metrics such as accuracy, recall, f1-score and precision is judged. Also, runtime analysis of the algorithms is studied.
- Graphical representation of sentimentally analyzed data is plotted as bar graph and evaluated with overall rating.

The classifier was trained using 1217 positive reviews and 955 negative reviews. The trained classifier was made to predict the polarity for 868 reviews. The comparison of the two classifier methods showed that Linear Support Vector outperformed that of Multinomial Naive Bayes. It was also observed that accuracy improved upon applying various pre-processing techniques. The structure of the paper is as follows: Section 2 provides the literature survey of existing works. System architecture and system design are discussed in section 3 and 4 respectively. Section 5 provides an insight on performance analysis of the developed system. Section 6 discusses about experiments and results of the machine learning algorithms. Conclusion and future work are discussed in section 7 and 8 respectively.

LITERATURE SURVEY

An Approach to Sentiment Analysis The Case of Airline Quality Rating

This research paper classifies the airline tweets as positive, negative opinion or emotion. This re-search paper also performs Sentiment Topic Recognition to examine the reputation of airline by calculating Airline Quality Rating based on customer sentiment analysis [4].

Airline Quality Rating is evaluated based on four criteria On-Time, Denied Boarding, Mishandled Baggage and Customer Complaint. For Sentiment Detection, the tweets are subjected to preprocessing before classification. The pre-processing includes removing retweets, html entries, markups, punctuation, numbers, @, people and unnecessary spaces. The tweets are classified using Naive Bayes algorithm to predict whether the tweet is positive or negative. Sentiment Topic Recognition of tweets is performed. Sentiment Topic Recognition identifies the representative topic for each sentiment. Sentiment Topic Recognition uses Correlated Topics Models (CTM) with Variational Expectation-Maximization (VEM) algorithm which extracts terms from the airline tweets to build lexicon for the Airline Quality Rating category. This model categorizes the topic related terms into positive or negative for each Airline Quality

Rating category.

Sentiment Analysis of Movie Review Comments

Sentiment Analysis of Movie Review Comments classifies the movie review comments obtained from social networking site 'Diggs' into positive/negative/neutral and subjective/objective with the help of machine learning techniques [3]. In order to perform machine learning, features are extracted from the comments. Feature vectors are represented by $\vec{F} = (f_1, f_2, ..f_n)$. Each coordinate of a feature vector is represented by a clue, f_i of the text. The presence or absence of a feature is represented by a binary value or integer value.

In bag-of-words model, frequency of words in the documents is noted and some words which occur frequently above a threshold are also noted. Infrequent words are eliminated. Then, the synonymous words are clustered. In order to handle negation, chunking is used. Classification of comments involves supervised and unsupervised learning techniques. Supervised learning techniques use four machine learning techniques - Naive Bayes, Maximum Entropy and Decision Tree classifier. Unsupervised approach includes K-Means Clustering. Naive Bayes algorithm is the best classification algorithm with the application of negation handling and bag-of-words. K-Means clustering performs well in subjectivity analysis.

Sentiment Analysis Of Twitter Data

Sentiment Analysis of Twitter Data involves classifying the tweets into three categories - positive, negative and neutral or junk [2]. The "junk label" means that the tweet cannot be understood by a human annotator. Pre-processing steps include replacing all the emoticons with their sentiment polarity by looking up the emoticon dictionary, replacing all URLs with a tag, replacing all sets of negation characters with "not" and replacing repeated characters with a set of three characters. Then other pre-processing techniques like tokenizing, removing stop words and removing punctuation marks are also performed. Tweets are converted into Partial Tree Kernel. Partial Tree depicts the tweet features in a tree data structure. Similarities between the features are estimated by finding the correlation between the subtrees of kernels. The classification of tweets follows five models - Unigram model, Tree Kernel model, 100 Senti-features model, Unigram + Senti-features model and Kernel + Senti-features model. Support Vector classification is used for classifying tweets. It is found that Tree kernel model provides better accuracy than Unigram model and Senti-features model. Combination of Unigram and Senti-features model yields better accuracy than combination of Tree kernel and Senti-features model.

SYSTEM ARCHITECTURE

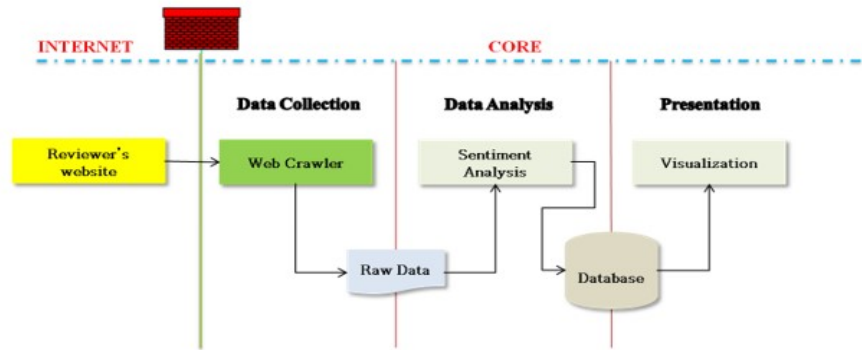


Fig. 1 . High level architecture diagram

Figure 1 depicts the high level architecture diagram of sentiment analyzer system for analyzing passenger reviews. The system is divided into three phases. The first phase is the data collection phase, here the individual passenger’s reviews along with rating are scraped from collection of airline forum links.

The second phase is the data analysis phase, here the reviews extracted are analyzed using machine learning techniques to obtain the sentiment. The third phase is the presentation phase and it involves plotting the rating and sentiment graphically.

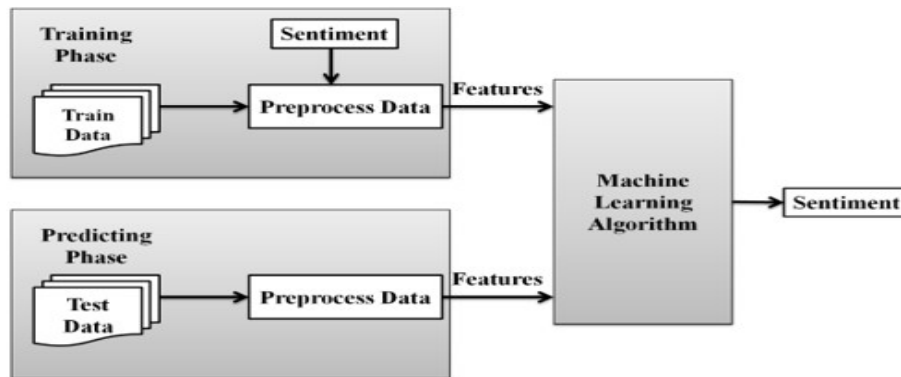


Fig. 2 . Architecture of Sentiment Analysis phase

Figure 2 depicts the block diagram of sentiment analysis phase shown in Figure 1. The sentiment analysis task is divided into two phases: training phase and predicting phase. The training phase involves loading train data (manually classified reviews) which is used to train the machine learning algorithm. The predicting phase involves loading test data (unclassified reviews) for which sentiments need to be determined. Both train and test data are pre-processed to obtain meaningful features before sentiment is obtained.

SYSTEM DESIGN

Data Extraction

The data extraction phase involves collecting data from airline forum [10] and building text corpora of reviews for use in data analysis phase. The list of URL is given as input to the

web scrapper which is used for extracting webpages. Extracted webpages are in HTML format and are parsed into textual format. Reviews and rating of 3040 passengers are extracted. Web scraper script was written in python.

Data Analysis

Pre-Processing

The motivation behind pre-processing is to reduce the number of features in the classification of reviews and also to reduce the complexity of the sentimental classification. This will render a helping hand in improving throughput. There are various pre-processing strategies available. In this paper we concentrate on the following: Normalization, Lemmatization, Stemming, Stop Words Removal and Negation Handling.

Normalization

Normalization is the process of reducing the number of features which is required for classification. The training data and test data are word tokenized. For example 'This airline is good' is tokenized to 'This', 'airline', 'is', 'good'. The words are tokenized for better pre-processing. The punctuation doesn't play any role in sentiment classification of airline reviews. So, the punctuation marks like '!', '?', '.' are removed.

Stop Words Removal

Stop words removal plays an important role in reducing the length of documents in sentiment analysis. It is the process of filtering words which are of little help in processing the documents. Some words like articles, pronouns, etc. are prevalent in all the documents. These words don't determine the sentiment of a document. For example, words like 'the', 'a', 'these', etc. are of no use in sentiment analysis and hence they can be removed. Country names, date of travel, numerical value of baggage weight, etc. are also removed in our work.

Lemmatization

Lemmatization uses vocabulary and morphological analysis of word to remove inflectional endings, thereby returning words to their dictionary form [5].

Lemmatization scans through the entire document and can be able to distinguish the meanings of the word according to the parts of speech. For instance the words 'new' and 'news' convey different meanings. WordNet is a lexical database of English words. Lemma or root word is a gateway to the WordNet [6].

The words which are conceptually same and can be used interchangeably are grouped into an unordered list. For example 'loved' and 'loves' are grouped together to a root word or a lexeme called 'love'. The lemmatizer will look up the WordNet and replace the word 'loved' by 'love' given that the part of speech is a noun.

Stemming

Stemming is the process of reducing grammatical, derivational and inflectional forms of a word to their common base form called stem [7]. It removes the suffix of a word. For example 'connection', 'connective', 'connectivity' are reduced to a common form called 'connect'. The most important distinguishing parameter between stemming and lemmatization is that stemming does not bother about interpreting the parts of speech of a word whereas lemmatization is concerned about analyzing the parts of speech of a word by scanning through the full text. There are different stemming algorithms Porter stemming, Lancaster Stemming, etc. We used Porter stemming

algorithm for the purpose of stemming. Porter stemmer strips the suffixes in a word. It strips plurals and also deals with suffixes such as 'ed', 'y', 'ness', 'ation', etc.

Term Frequency and Inverse Document Frequency

Term frequency and inverse document frequency are two distinct terms. 'Term frequency' is a measure of number of times each term appears in a document. It is computed by dividing number of times a word occurs in a document by the total number of documents. 'Inverse document frequency' scales down frequently occurring terms and scales up rarely occurring terms [8].

It is the measure of the logarithm of the division of the number of documents by the number of documents containing term t . Tf-idf is the measure of product of term frequency and inverse document frequency.

If N_t is the number of times of each term t in the document and D_n is the number of terms in documents n where $n \in N$ where N is the total number of documents then term frequency is given by:

$$tf_t = \frac{N_t}{D_n}$$

If N is the total number of documents and M_t is the number of documents with term t , then inverse document frequency is:

$$idf_t = \log\left(\frac{N}{M_t}\right)$$

Term frequency inverse document frequency is the product of term frequency and inverse document frequency and is given by:

$$fidf_t = tf_t \times idf_t$$

Machine Learning Algorithms

In this work, we have used two supervised machine learning algorithms Multinomial Naive Bayes and Linear Support Vector Machines. The machine learning algorithms are trained using 1217 positive and 955 negative reviews. 868 reviews were used as test data for classification.

Linear Support Vector Machines

Linear Support Vector Machine is a supervised machine learning technique for sentiment analysis of documents. We use support vector machines to classify airline reviews collected from airline forum. Given a training set, Linear Support Vector machines find a linear separation called hyperplane to separate the datasets. In our case, the linear decision boundary separates the datasets into class positive and negative. Let the class positive be represented as +1 and negative as -1. The hyperplane must be in such a way that the distance between the closest data points and the hyperplane must be maximum. The closest point to the hyperplane is called the margin. The points which are

closest to the separating hyperplane are also known as support vectors.

Generally a hyperplane is of the form $w^T + b$ where w^T is a weight vector and b is the bias. The term w^T is the normal vector and it is a measure of orientation of the hyperplane and $\frac{b}{\|w\|}$ is the displacement of the hyperplane from the origin. In order to find the points w_T and b , we must find the points with the smallest margin. So, we maximize the margin which can be written as

$$argmax_{w,b} \{min_n(label.(W^T + b)). \frac{1}{\|w\|}\}$$

Where label $\varepsilon \{+1,-1\}$.

After finding the support vectors, it is important to classify the test data. Test data X_i are mapped to a function $f(X_i) = w^T.X_i + b$ If X_i is the test data and label which is represented by $y(i)\varepsilon+1, -1$, then

$$f(X_i) \begin{cases} \geq 0 y_i = +1 \\ < 0 y_i = -1 \end{cases} \quad (1)$$

Equation (1) states that if $f(X_i) \geq 0$, then the class is positive and if $f(X_i) < 0$, then the class is negative.

Multinomial Naive Bayes

In Multinomial Naive Bayes, a document d can be assumed to be collection of word features f_i . It is represented as $d = f_1, f_2, \dots, f_n$. f_i can be assumed to be the frequency of word w_i in a document d . Vocabulary is the dictionary of distinctive repository of words in the total collection of documents. Documents are represented as bag-of-words model because the frequency of words in a document is listed as a feature. The probability of class c given the test document d is represented as

$$P(c/d) = \frac{P(c) \prod_{i=1}^n P(w_i/c)^{f_i}}{P(d)}$$

where $P(c)$ is the prior probability of a class c , $P(w_i/c)$ is the probability of word w_i occurring in a document d corresponding to class c , $P(d)$ is the likelihood of a document and f_i is the number of occurrences of a word w_i in a document d . The prior probability of a class is given by:

$$P(c) = \frac{N_c}{N}$$

where N_c is the number of documents corresponding to class c and N is the total number of documents. The conditional probability $P(w_i/c)$ of a word w_i occurring in a document d in a class c is given by:

$$P(w_i/c) = \frac{N_{ci} + \alpha}{N_c + \alpha.|V|}$$

where N_{ci} is the number of times w_i appearing in c and N_c is the total count of words in class c and $|V|$ is the number of unique words in the vocabulary.

In order to avoid zero probabilities, parameter α is set. If $\alpha=1$, the smoothing applied is called Laplacian smoothing.

Data Storage and Data Visualization

The sentiments of reviews are stored in the database. Database connectivity is established between Business Intelligence tool and the database. Data visualization is essential because information can be better understood through pictorial representation. It is the presentation of data in pictorial form such as graphs and charts [9]. Section 6.2 displays the various visualization plots obtained as a result of sentiment analysis.

PERFORMANCE ANALYSIS

Classifier Evaluation

TABLE 1

EVALUATION REPORT FOR LINEAR SUPPORT VECTOR AND MULTINOMIAL NAIVE BAYES

Linear Support Vector				Multinomial Naive Bayes			
Sentiment	Precision	Recall	F1-Score	Sentiment	Precision	Recall	F1-Score
Negative	0.87	0.85	0.86	Negative	0.82	0.75	0.79
Positive	0.84	0.75	0.79	Positive	0.84	0.89	0.86
Accuracy			88.59%	Accuracy			84.56%

The most commonly used metrics to evaluate the classifiers are precision, recall, f1-score and accuracy. Before looking into details of these metrics, it is important to understand four terms namely: true positive, true negative, false positive and false negative. True positive is the number of reviews that are classified as positive and are really positive. True negative is the number of reviews that are classified as negative and are really negative. False positive is the number of reviews that are actually positive but are classified negative.

False negative is the number of reviews that are negative but are classified positive.

Precision

Precision is ratio of prediction made by classifiers that are actually correct. The range of precision is between 0 and 1. Higher the value of precision better is the classification. Precision is given by the below formula:

$$P = \frac{tp}{tp+fp}$$



Where, tp is the number of true positive and fp is the number of false positive. From table 1 it can be inferred that precision for Linear Support Vector classification is 0.87 for negative sentiment and 0.84 for positive sentiment whereas for Multinomial Naive Bayes classification it is 0.82 for negative sentiment and 0.84 for positive sentiment.

Recall

Recall is a measure of sensitivity or true positive of the classified review against true positives and true negatives. The range of recall is between 0 and 1. Higher the value of recall better the classification. Recall is given by the below formula:

$$R = \frac{tp}{tp+fn}$$

Where, tp is the number of true positive and fn is the number of false negative. From table 1 it can be inferred that recall for Linear Support Vector classification is 0.85 for negative sentiment and 0.75 for positive sentiment whereas for Multinomial Naive Bayes classification it is 0.75 for negative sentiment and 0.89 for positive sentiment.

F1-Score

F1-score is a meld of precision and recall. Recall is given more importance while calculating the score. The range of f1-score is between 0 and 1. Higher the value of f1-score better is the classification. F1-score is given by the below formula:

$$F1 = \frac{2 * P * R}{P + R}$$

Where, P is the precision and R is the recall. From table 1 it can be inferred that f1-score for Linear Support Vector classification is 0.86 for negative sentiment and 0.79 for positive sentiment whereas for Multinomial Naive Bayes classification it is 0.79 for negative sentiment and 0.86 for positive sentiment.

Accuracy

Accuracy is defined as the percentage of correctly classified positive and negative reviews.

$$Accuracy = \frac{tp+tn}{tp+tn+fp+fn}$$

Where, tp denotes true positive, tn denotes true negative, fp denotes false positive and fn denotes false negative. From table 1 it can be inferred that the accuracy of Linear Support Vector is 88.59% and that of Multinomial Naive Bayes is 84.56%.

Run Time Analysis

The classifiers were implemented in python and tested on laptop running an Intel(R) core(TM) i5-3210M CPU with frequency 2.50 GHz. The total memory used for training and classifying passenger reviews is 138.3 MB.

From figure 3, we infer that Multinomial Naive Bayes is faster than Linear Support Vector Machines. Naive Bayes is computationally quicker while making decisions. Support Vector Machines take more time to build a decision surface and then perform sentiment classification.

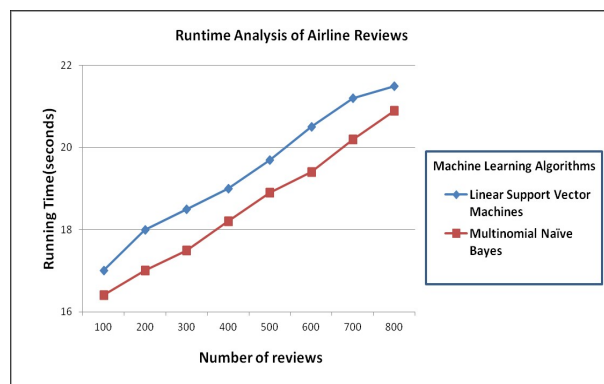


Fig. 3 . Runtime analysis of sentiment classification of machine learning algorithms

EXPERIMENTS AND RESULTS

In this section, we discuss about the experiments and results of two machine learning approaches: Linear Support Vector classification and Multinomial Naive Bayes classification. This section is split into two parts: 1) Classifier accuracy and pre-processing strategy dependence and 2) Visualization of analyzed data.

Classifier Accuracy and Pre-Processing Strategy Dependence

This section details about the correlation that exists between the accuracy, number of features and the different pre-processing techniques used. The results are summarized in table 2.

TABLE 2
VARIATION IN ACCURACY AND NUMBER OF FEATURES BASED ON PRE-PROCESSING APPROACH

Approach	Number of Features	Linear Support Vector Accuracy	Multinomial Naive Bayes Accuracy
Without Pre-processing	10112	86.05%	83.17%
Normalization	12144	86.40%	84.21%
Normalization + Stop Words Removal	10056	86.75%	84.10%
Normalization + Stop Words Removal + Stemming	6553	87.55%	84.33%
Normalization + Stop Words Removal + Stemming + Lemmatization	6444	88.59%	84.56%

Table 2 clearly suggests that, as more number of pre-processing techniques are applied, there is an increasing degree of accuracy and reduction in the number of features. We infer that there is significant reduction in number of features when stop words removal and stemming are applied. Though there is marginal decline in number of features when lemmatization is applied, we observe a steep increase in Linear Support Vector accuracy by 1.04%. In Multinomial Naive Bayes, steep increase in accuracy by 1.04% was observed upon performing normaliza-

tion. The overall increase in Linear Support Vector classification accuracy when compared to without pre-processing is 2.54% whereas it is 1.39% in Multinomial Naive Bayes classification.

Visualization of Analyzed Data

In this section, we display the graphical analysis of sentiment analysis as well as the ratings extracted from the forum.

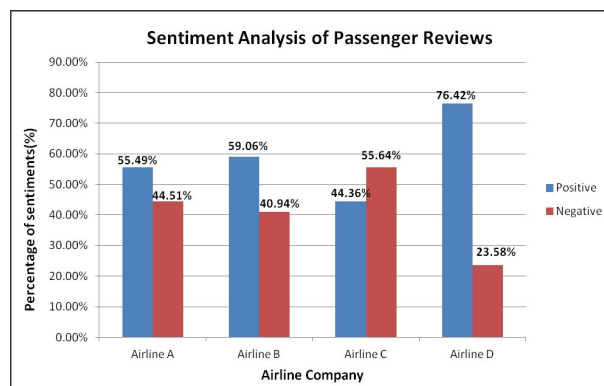


Fig. 4 . Sentiment analysis of passenger reviews

Figure 4 delineates percentage of passengers whose reviews are classified positive/negative. We infer that Airline D has been received well among the passengers compared to other three airline companies. Airline C as seen from figure

4 has failed to meet its passengers’ expectations. Airline A and Airline B have received mixed reviews from passengers indicating that they should work on their services more.

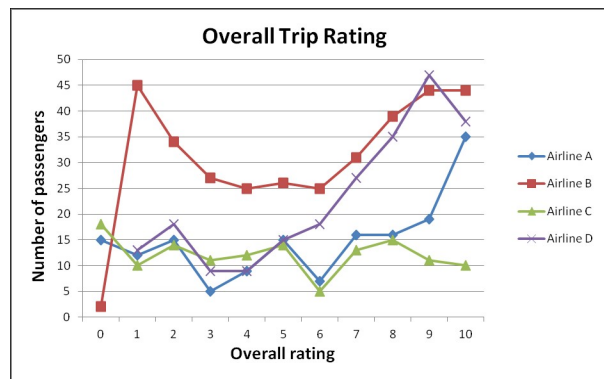


Fig. 5 . Overall trip rating

Figure 5 shows the overall trip rating rated by individual passenger on a scale of 10. Figure 4 and Figure 5 synchronize with each other. For example, figure 4 reveals that Airline C earned more negative reviews than positive reviews which is in correspondence with figure 5 that Airline C garnered poor ratings.

CONCLUSION

In this work we have focussed on analyzing the sentiments of passenger reviews obtained from popular airline forum. Our work suggests that Linear Support Vector classification yields better accuracy than Multinomial Naive Bayes. But, when it comes to run time, Multinomial Naive Bayes classifies faster than Linear Support Vector. The number of features was reduced by 36.27% on the application of different pre-processing techniques. It was also observed that classifier accuracy improved with the application of different pre-processing techniques.

The graphical plot of the sentiments using business intelligence tool makes it easy to evaluate performance of an airline. The limitation of our work is that the system does not provide any information about overall scope of the sentiment. Also, the system requires training with manually classified reviews unlike traditional lexicon-based analysis system.

FUTURE WORK

Subjective and objective sentence detection can be employed to remove opinion sentences from reviews which will make sentiment classification task easier. The data set can be increased and experiment can be repeated on different classification methods such as clustering. Pre-processing strategies like bigrams, trigrams, negation handling, feature selection, etc. can be used to improve classifier accuracy. Hybrid classifier system can be used to improve accuracy. Sentiment Topic Detection can be employed to determine the most talked about topic.

REFERENCES

- [1] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, vol. 10, Association for Computational Linguistics, Stroudsburg, PA, pp. 79-86, 2002.
- [2] K. Yessenov and S. Misailovic, "Sentiment analysis of movie review comments," Unpublished dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2009.
- [3] A. Agarwal, B. Xie, I. Vovsha, O. Rambow and R. Passonneau, "Sentiment analysis of twitter data," in *Proceedings of the Workshop on Languages in Social Media*, Association for Computational Linguistics, Stroudsburg, PA, pp. 30-38, 2011.
- [4] E. Adeborna and K. Siau, "An approach to sentiment analysis-the case of airline quality rating," in *Pacific Asia Conference on Information Systems PACIS*, Chengdu, China, 2014, p. 363.
- [5] V. Balakrishnan and E. Lloyd-Yemoh, "Stemming and lemmatization: A comparison of retrieval performances," *Lecture Notes on Software Engineering*, vol. 2, no. 3, pp. 262-267, 2014.
- [6] P. Bhattacharyya, A. Bahuguna, L. Talukdar and B. Phukan, "Facilitating multi-lingual sense annotation: Human mediated lemmatizer," in *Global WordNet Conference*, Tartu, Estonia, 2014.
- [7] S. S. Megala, A. Kavitha and A. Marimuthu, "Improvised stemming algorithm-TWIG," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 7, pp. 168-171, 2013.
- [8] C. D. Manning, P. Raghavan and H. Schutze, *Introduction to Information Retrieval*, New York, NY: Cambridge University Press, 2008.
- [9] SAS. (n.d). *Data visualization, what it is and why it matters* [Online]. Available: <http://goo.gl/4GUUov>
- [10] SKYTRAX (n.d). *Airline rating and reviews* [Online]. Available: <http://www.airlinequality.com/>

— This article does not have any appendix. —