

RESEARCH

Open Access



A machine learning approach to analyze customer satisfaction from airline tweets

Sachin Kumar* and Mikhail Zymbler

*Correspondence:
sachinagnihotri16@gmail.com
Department of System
Programming, South Ural
State University, Chelyabinsk,
Russian Federation

Abstract

Customer's experience is one of the important concern for airline industries. Twitter is one of the popular social media platform where flight travelers share their feedbacks in the form of tweets. This study presents a machine learning approach to analyze the tweets to improve the customer's experience. Features were extracted from the tweets using word embedding with Glove dictionary approach and n-gram approach. Further, SVM (support vector machine) and several ANN (artificial neural network) architectures were considered to develop classification model that maps the tweet into positive and negative category. Additionally, convolutional neural network (CNN) were developed to classify the tweets and the results were compared with the most accurate model among SVM and several ANN architectures. It was found that CNN outperformed SVM and ANN models. In the end, association rule mining have been performed on different categories of tweets to map the relationship with sentiment categories. The results show that interesting associations were identified that certainly helps the airline industries to improve their customer's experience.

Keywords: Twitter, Machine learning, Convolutional neural network, Association analysis, Customer satisfaction

Introduction

Advancement in technology has boosted the availability and use of smart mobile phones. At present, the number of smart phone users is 2.71 billion across the world [1]. The major online social media (SM) platforms i.e. Facebook, Twitter and Instagram are available as mobile applications in the smart phones. Therefore, there is no need to visit cyber cafes to access them, as everything is available in the smart phones.

Every piece of information shared on SM carries an emotion, sentiment or feeling. These emotions can be positive, negative and neutral. All these emotions may come from a travel trip, restaurant trip, exhibitions, movies, elections, hospital visits etc. These emotions carries some hidden information related to comfort/discomfort in related areas. Hence, there is a good scope of analyzing this information to detect the patterns of the emotions. This analysis can help us to understand the emotions of the people in respective domain and the reasons behind it.

Air travel is one of the most convenient mode for long distance travel at both national and international level [2]. There are many airline service providers (ASPs) around the world. The competitive world motivates the airlines company to attract the customers.

However, a traveler considers several points before selecting any airline. These points can be airfare, travel time, number of stoppages, number of baggage allowed, and existing customer feedback etc. Therefore, all ASPs are working in all these customer service areas to improve their facility and in-flight comfort in order to attract the customers.

It is very important to understand the needs and comfort level of customers i.e. customer satisfaction during the flight. Therefore, customer feedback is very important for any airline industry. There could be several possible ways to collect the customer feedback. The most easiest and traditional way is the customer feedback form available during the journey. However, most of the passengers do not show any interest in filling feedback forms. Another shortcoming of this approach is that it may or may not have appropriate questionnaire and may be biased on certain parameters i.e. the feedback form may only have certain specific questions. Other approaches for customer feedback collection could be through online website or online mobile applications of the airlines. After the journey, an email with a link can be sent to the passenger to request for a feedback. However, there is no guarantee of its success. Another approach is to send a message on passenger's mobile phone and ask them to rate your service (1 for poor and 5 for excellent) on certain parameters. All these traditional methods opted by the industry are restricted to certain parameters only. The more convenient way for a passenger is to express their feedback, as they want. Therefore, the most convenient way for the passengers to share their opinions is the social media instead of feedback form. Social media provides a platform where a user can freely express his feedbacks on any issues they observed during flight. Twitter [3] is one of the popular platform worldwide. The information from Twitter can be utilized to develop a recommender system [4]. In addition, travelers are more comfortable in sharing their views about travel experiences on Twitter.

A variety of major issues affects the emotions of a passenger in air travel. These issues can be cabin crew behavior, food quality, loss of baggage, seat comfort, flight delay, airfare etc. All these issues may give rise to both positive and negative emotions. Also, if there is a continuous trend of negative tweets for an airline, then it may put a negative impact to the economic growth of the airline company. Therefore, it is important to understand the issues that give rise to negative tweets so that the respective airline company can take appropriate action on time. There are a large number of airlines operating every day to connect different geographical locations [5]. Therefore, we may expect a large number of people travelling every day in these flights. In addition, the number of tweets by passengers for airlines would be very large. Therefore, it is a challenging task to extract the hidden emotion behind a tweet. Therefore, we required some tools and techniques that are able to handle such a large number of tweet database and can provide insights to help airline industry.

Machine learning [6] and big data technologies [7] made it possible to analyze huge database and to develop highly accurate prediction or classification models. In this article, machine learning techniques are used to develop a binary class sentiment classification model for twitter data for some of the popular airlines worldwide. The study opted to develop a classification model for two categories of sentiments i.e. positive and

negative. The reason behind selecting only positive and negative sentiments is that neutral sentiment does not provide any information whether the service was nice or not. It may also depends on a neutral and kind passenger who takes things normally. Also, neutral sentiment just pretends that everything was normal but airline industries are more interested in more specifically “what was good or bad”. Therefore, only positive and negative sentiments classes have been opted for analysis in this study. Support Vector Machine (SVM) and Artificial Neural Networks were trained on the preprocessed tweets. Further, convolutional neural network (CNN) is trained on the data and its performance were compared with the best model among SVM and ANN models. Results shows that CNN outperformed all other models in terms of accuracy and performance. Further, association rule mining is used to map the relationship between several issues related to passenger’s comfort during flight with the nature of emotions (positive or negative).

The remaining part of the article is organized as follows: “[Related work](#)” section describes the various theoretical and methodological aspects of sentiment analysis. In “[Materials and methods](#)” section, a description about methodology and data set is given. In “[Results and discussion](#)” section, the results are presented and discussed in detail. The article in concluded in “[Conclusion](#)” section.

Related work

Sentiment analysis is an important approach to extract emotions from any textual information i.e. online articles, product review, movie reviews, Twitter data etc. Twitter data is usually contains information about a person’s opinion on any miscellaneous topic. Air travel is also one of these hot topics that are widely spread on Twitter. Air passengers usually share their travel experience on Twitter. This information can be useful if analyzed using machine learning techniques and can provide insights that helps to understand the comfort level of the passenger in the flight. The literature in the field of sentiment analysis is so deep and vast. Two major categories of opinion mining are lexicon-based techniques (LBT) and ML techniques. In LBT, the concept is based on the sentiment lexicon. Sentiment lexicon can be defined as a set of known and in-built terms, phrases and idioms. These terms or phrases are developed in respect to traditional genres i.e. Opinion finder [8], ontologies [9] and dictionaries [10, 11]. If we dive into details then we find two categories of LBT. The first one is dictionary based approach (DBA) and the other one is corpus based approach (CBA). In DBA, an initial set of seeds or terms are collected and the comment or description of each term is provided manually. Further, synonyms and antonyms of the terms are used to enlarge this dataset. SentiWordNet [12] is the name of popular thesaurus that were developed using a dictionary called WordNet [13]. The major issue with DBA is its inability to handle domain specific orientation. Therefore, one can use DBA to determine a generalize orientation of the context but for domain specific, it is not a better approach.

Then, CBA came up with a solution to the above issue by providing domain specific dictionaries. Statistical methods i.e. latent semantic analysis [14] or simply counting

mode/frequency of words in a set of documents is a good approach to develop such domain specific dictionaries. Semantic methods [15] also provide a good solution by using synonyms and antonyms of the terms and their relationships driven from popular thesaurus like WordNet. Cambria believes that sentiment analysis is a very restricted domain of natural language processing (NLP) where each sentence/topic should correlate with positive or negative sentiments [16]. However, sentiment analysis has always been a popular domain in NLP as there is a lot to achieve in this domain.

In order to perform sentiment analysis, it is important to determine the level of analysis i.e. document level, sentence level or entity level. In document level sentiment analysis, the entire document is considered as an opinion [17–20]. However, in a document, if there are sentences that represent different context, then sentence level sentiment analysis is preferred. Sentence level sentiment analysis is closely related to subjectivity analysis of a sentence [21, 22].

In order to get more insights from the text, the entity level analysis can be performed which is a finest grained level analysis [23]. At this level of analysis, the base for sentiment analysis is the main context. Therefore, this level of analysis is slightly more complex as high precision is required on output [24, 25]. Feature based sentiment analysis and summarization comes into this category of analysis. On the other part, machine learning has played an excellent role in NLP. Machine learning is broadly categorized into supervised and unsupervised learning. Supervised techniques such as Naïve Bayes (NB), support vector machines (SVM), Maximum Entropy are widely used for sentiment analysis [26, 27]. The data set with unlabeled set of documents can be analyzed using unsupervised ML techniques [28, 29]. In addition, hybrid version of ML techniques that combined both supervised and unsupervised can also be used for sentiment analysis [30, 31].

Dutta et al. [32] analyzed airline twitter data using Naïve Bayes classifier for sentiment analysis. They used R and Rapid Miner tools to develop the classification model and map the tweets into positive, negative and neutral category. They mentioned that the results achieved using Naïve Bayes classifiers were promising. Rane and Kumar [33] performed a comparative study on six US based airline companies using decision tree, random forest, Gaussian Naïve Bayes, SVM, K-nearest neighbors, logistic regression and AdaBoost. They trained the classifiers on 80% of the data and remaining data was used for testing. They classified the tweets into three categories of sentiments. They mentioned that logistic regression, AdaBoost, random forest, and SVM performed well on the model with an accuracy of more than 80%, however, the improvements can be achieved adding more number of tweets in the analysis.

Sternberg et al. [34] studies the customer engagement of Turkish airlines using the official Turkish airline's Facebook page. They suggested that although there was not enough information on the Facebook page but certain good relationships were identified after analysis. They suggested that the Facebook page could be helpful for Turkish

airlines to generate short-term revenues. Additionally, they mentioned that there is not enough research on airlines industry using social media data.

Ashi et al. [35] used two different pre-trained word embedding models i.e. fastText Arabic Wikipedia and AraVec-Web for aspect based sentiment analysis. They considered 5000 Arabic tweets for airline services and manually labeled in aspect categories. Further, they used SVM classifier to train the model for aspect detection as well as for polarity detection. They observed an improvement in the performance of SVM classifier when features were extracted using word embedding. However, results were slightly more accurate with fastText Arabic Wikipedia word embedding model than the AraVec-Web. Their research revealed that use of word embedding is good for sentiment analysis.

Association rule mining is also important in sentiment analysis [36]. Dehkharghani et al. [36] proposed an approach for casual rule discovery from Twitter text. They tested the proposed approach on political tweets related to Kurdish political issue in Turkey and found the results promising. Therefore, it is evident that association rule mining can help in revealing associations between different aspects of airline tweets as well.

Keeping in mind the usefulness of word embedding feature vectors, this study tries to investigate the usefulness of word embedding for airline tweets. In addition, SVM, ANN and CNN approaches are considered to develop a classification model to classify the tweets into positive and negative categories. Further, association rules were extracted from the tweets data to understand the association between the sentiments and related subject of tweets.

Materials and methods

This section provides a description of the data collection and its preprocessing. Further, the methodology is discussed that consists of the overview of the machine learning methods used in the study.

Data description and pre-processing

The tweets related to several major airlines with large number of followers across the world have been extracted from Twitter server. In order to download tweets, a python script was written using Twitter API. Tweepy package is used to download tweets from Twitter server. Qatar airways (QTA), Royal Dutch airlines (KLM), Air New Zealand (ANZ), Turkish airlines (TKA), JetBlue airways (JBA), American airlines (AMA), United airlines (UTA), British airways (BTA), Emirates (EMR), Delta (DLT) and Lufthansa airlines (LFT) were selected to be included in the analysis. In order to download tweets related to each of the airlines mentioned above, proper hash tags were carefully chosen in analysis. Only English language tweets were considered in this study for different airlines. The tweets were collected for the duration from 1 March 2019 to 11 March 2019 (Due to the availability on the Twitter server). Initially, extracted tweets were in JSON (JavaScript Object Notation) format that are converted into csv format. The total number of tweets for all the mentioned airlines was 146,731. AMA has the highest number of tweets that consists of 44.13% of all tweets downloaded (Fig. 1). After preprocessing and removing the retweets, 120,766 tweets were remained in the data set that are used for further processing.

The pseudocode used for tweet preprocessing is provided in pseudocode procedure1.

Procedure1: Tweet Pre-Processing

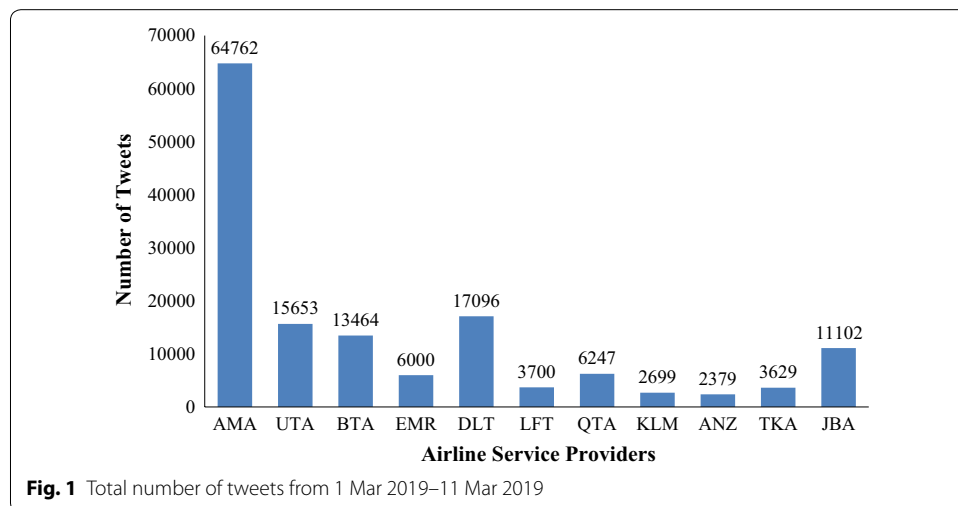
Input : Tweets data obtained from Twitter server for major airlines
Output : Pre-processed Tweets
Parameters : R_E : Regular expression python library, T_w : Tweets, RT_w : Re-Tweets, C_R : Corpus, R_{RT} : Remove Re-Tweets, R_{HT} : Remove #Tag, R_{USN} : Remove username, R_{URL} : Remove URL, P_{TF} : Parse Tweet File
Initialization: $P_{TF} \leftarrow [Null]$

```

Begin,
  for  $T_w$  in  $C_R$  do
    D = json.load ( $T_w$ )
     $P_{TF} \leftarrow D ['Text']$ 
  end
  for  $T_w$  in  $P_{TF}$  do
     $R_{RT} = R_E.findall (T_w, "RT@w+")$ 
     $R_{URL} = R_E.findall (T_w, "http (s)?://(.|[a-z]|[A-Z]|[0-9]+)")$ 
     $R_{HT} = R_E.findall (T_w, "#\w+")$ 
     $R_{USN} = R_E.findall (T_w, "@\w+")$ 
     $R_{RT} = R_{URL} = R_{HT} = R_{USN} = [NULL]$ 
  end
end
    
```

Feature extraction

Feature extraction is one of the important parts of the sentiment analysis. Raw data needs to be processed and must be transformed into numerical representation for analysis. The typical size of a tweet is 140 characters maximum on Twitter. We have used n-gram models and word embeddings (WE) using GloVe dictionary approach to prepare the data set for analysis. In tweet analysis, text must be represented as a weighted featured vector. N-gram is a potentially useful approach as it provides a sequence of words



that helps to assign the probabilities to a sequence of words in a tweet. N-gram model can be simply stated as a sequence of words within a fixed window size n .

For example, consider a tweet, “The flight was so pleasant except that food quality can be improved”. The n -gram representation of the above tweet for 3-gram representation would be, “The flight was”, “flight was so”, “was so pleasant”, “so pleasant except”, “pleasant except that”, “except that food”, “that food quality”, “food quality can”, “quality can be”, “can be improved”. It is important to understand that 1-gram model only consists of individual words of the tweet.

Another approach used in this study is word embeddings (WE). We used GloVe (Global Vectors for Word Representations) [37] algorithm to transform the tweet’s words into vectors. GloVe is a count based model and it creates a matrix of co-occurrences of words and performs some dimensionality reduction to learn the word vectors. The GloVe method is briefly explained as follows:

Procedure2: GloVe

Input : V : Vocabulary, T : Tweets Data, W : Window size, C : Minimum frequency threshold

Output : X : Sparse matrix

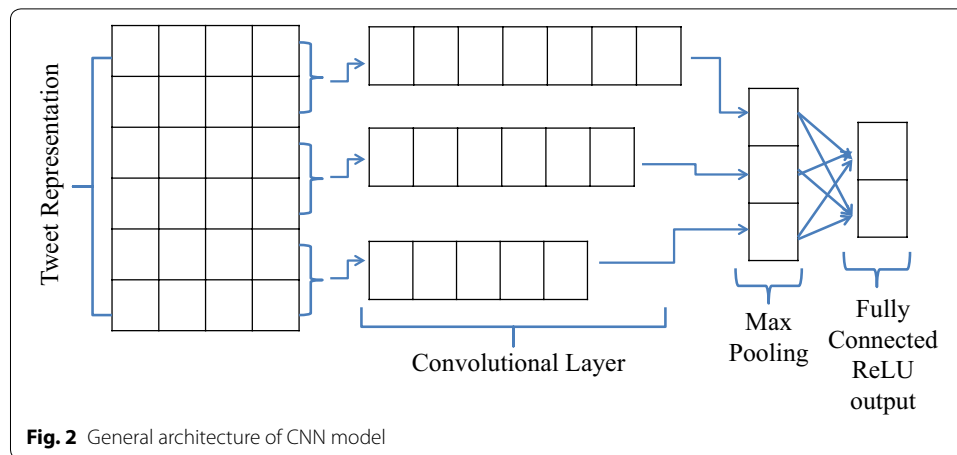
1. Initialize model parameters $W=V \times D$ where V is the size of the vocabulary and D is the specified vector dimensions
 2. Initialize bias b within range $(-0.5,+0.5)$
 3. Repeat until
 - Calculate the cost function
 - For each iteration
 - Calculate the gradient of the cost function with respect to parameters
 - Update the parameters according to learning rate
-

Sentiment classification approaches

Once the preprocessed features are extracted, machine learning methods can be used to develop the classification model for sentiment analysis. After the crucial investigation of literature, we decided to use Support Vector Machines (SVM), Artificial Neural Networks (ANNs), and Convolutional Neural Networks (CNN) in our study.

Support vector machines (SVM)

SVM is a well know machine learning method that has been widely used for sentiment analysis [20, 35]. For binary class classification problem, it distinguishes two classes with a margin threshold or hyperplane that separates data points from different classes. There could be several hyperplanes but the best one separates two classes with maximum margin. SVM is considered as one of the techniques for sentiment analysis in this study based on its popularity and effectiveness.



Artificial neural networks (ANN)

ANN is one of the popular prediction and classification technique for both numeric and categorical data. A general ANN architecture has one input layer, one output layer and one or more hidden layers. Number of neurons in each layer depends on the type of data and number of features in the data set. There are respective weights on each layer that has to be multiplied with the input values and then pass on to the next layers. BPANN is another version of ANN that back propagates the prediction error (actual output – predicted output) to the previous layers. This propagated error is used to modify the weights on each layer. The aim of this back propagation approach is to minimize the prediction error and improve the accuracy. The minimization of prediction error also depends on the number of times a network is trained on particular dataset.

Convolutional neural networks (CNN)

CNN or convolutional neural network can be seen as an advancement in deep neural networks that was initially designed for the image data. Moreover, the application of CNN is growing effectively for the data other than images such as speech recognition. A general architecture of CNN for tweet classification is illustrated in Fig. 2. Initially, each tweet is represented as a word vector in the n dimensional space. Assuming that d is the dimension of word vectors and l in the length of the tweet (the number of concatenated words in the tweet). Therefore, the dimension of the tweet matrix can be defined as $l \times d$ [38]. Collobert and Weston [39] suggested that text matrix can also be considered as an image matrix in order to perform convolution with filters with size similar to the region size. In the convolutional layer, n filters are applied on a window size of t of each tweet. Given $W_{[k:k+t]}$ are the word vectors from words w_k to w_{k+t} . We can generate a feature f_i for any filter F using the Eq. 1:

$$f_i = \sum_{a,b} (W_{[k:k+t]} a, b \cdot F_{a,b}) \quad (1)$$

All vectors in the tweet are concatenated to generate a feature vector $f \in \mathbb{R}^{l-k+1}$. All these feature vectors are then combined over all p filters into a feature map matrix $F \in \mathbb{R}^{p \times (l-k+1)}$. Further, an activation function (non-linear) is applied to the parameters

of convolution layer and then pass on to the pooling layer. The objective of max pooling layer is to reduce the size of spatial dimensions in order to improve the computational performance and to reduce the over-fitting. The output of the pooling layer is a feature map matrix of the form $F_{\max_pooled} \in \mathbb{R}^{p \times \left(\frac{l-k+1}{z}\right)}$, z is length of the interval on which the word vector elements were aggregated. Further, these features are passed to the next layer which acts as a fully connected neural network model. The output of the final hidden layer is applied to the ReLU (rectified linear unit) activation function, which categorizes the respective tweet vector into positive or negative class.

Association analysis

Another important aspect of this study is to investigate the effect of several parameters on the emotions of the passengers in flight. Association rule mining is one of the popular methods that can extract important associations between different aspects of travel comfort. We have used Apriori algorithm [40] to establish the association between different travel issues that can affect a passenger’s emotion during the flight. In Association rule mining is a market basket based approach where every data instance is a transaction with certain items purchased together. In our case, each tweet is considered as transaction and the category of words used in the tweets are the attributes. Every word present in the tweet is considered as an item. The tweets were separated based on both positive and negative sentiments. Apriori algorithm is applied on the data set to generate frequent item sets and further association rules were generated for both category of emotions i.e. positive sentiments and negative sentiments.

Performance evaluation measures

The quality and potential of association rules can be evaluated using certain performance metrics i.e. support, confidence and lift. Given a data set D with n transactions where each transaction T belongs to D ($T \in D$). Considering I be the set of items where $I = \{I_1, I_2, I_3, \dots, I_n\}$. An item A could appear in transaction T if and only if $A \subseteq T$. $A \rightarrow B$ can be considered as an association rule if $A \subset I$, $B \subset I$ and $A \cap B = \phi$. Support, confidence and lift over a rule $A \rightarrow B$ can be defined as follows: Support (S_p) of a rule $A \rightarrow B$ is the total percentage of occurrence of A and B together in the dataset and can be evaluated using Eq. 2. Support is also known as threshold value to generate rules. The strong support values can be considered to generate more strong rules.

$$S_p = \frac{P(A \cap B)}{N} \tag{2}$$

Confidence (C_f) of a rule $A \rightarrow B$ can be defined as the ratio of the occurrence of A and B together with the total occurrence of A only in the dataset and can be evaluated as Eq. 3. The high confidence value close to 1 indicates a more strong rule.

$$C_f = \frac{P(A \cap B)}{P(A)} \tag{3}$$

Lift (L_f) is another important parameter to evaluate the effectiveness of a rule $A \rightarrow B$. It measure the occurrence of A and B together than expected. In other words, lift is the ratio of actual confidence value and expected confidence value. Actual confidence value

Table 1 Employed ANN configurations

Configuration	No. of hidden layers	No. of neurons in layer 1	No. of neurons in layer 2	Batch size
ANN1	1	16	–	256
ANN2	1	32	–	256
ANN3	1	64	–	128
ANN4	2	16	4	256
ANN5	2	32	8	128
ANN6	2	64	16	64

evaluate the occurrence of A and B together with respect to the occurrence of A whereas expected confidence evaluate the occurrence of A and B with respect to the occurrence of B. The formula to evaluate the lift is given in Eq. 4. The expected values for lift may range from 0 to ∞ . In general, the lift value greater than 1 is considered as good to select the rules to be evaluated on new data.

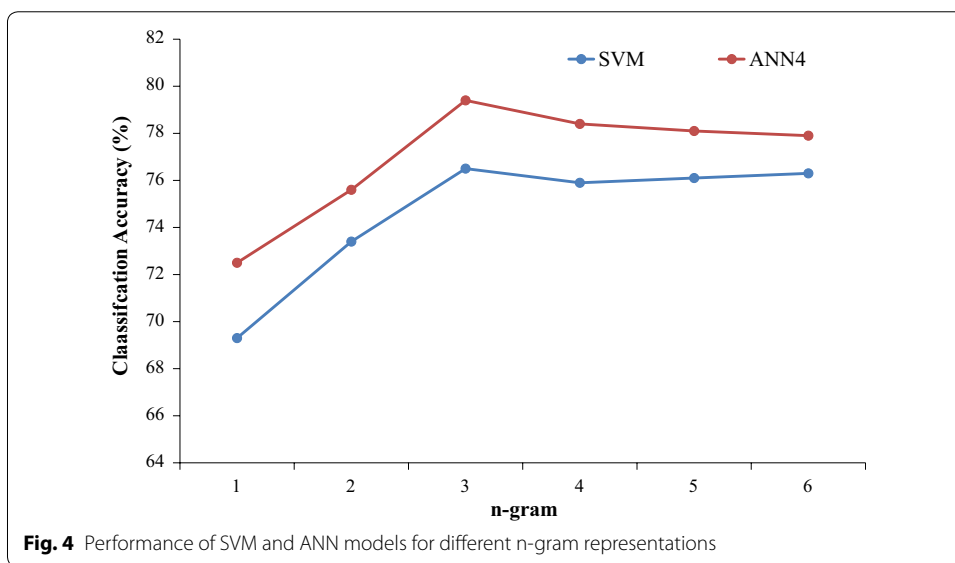
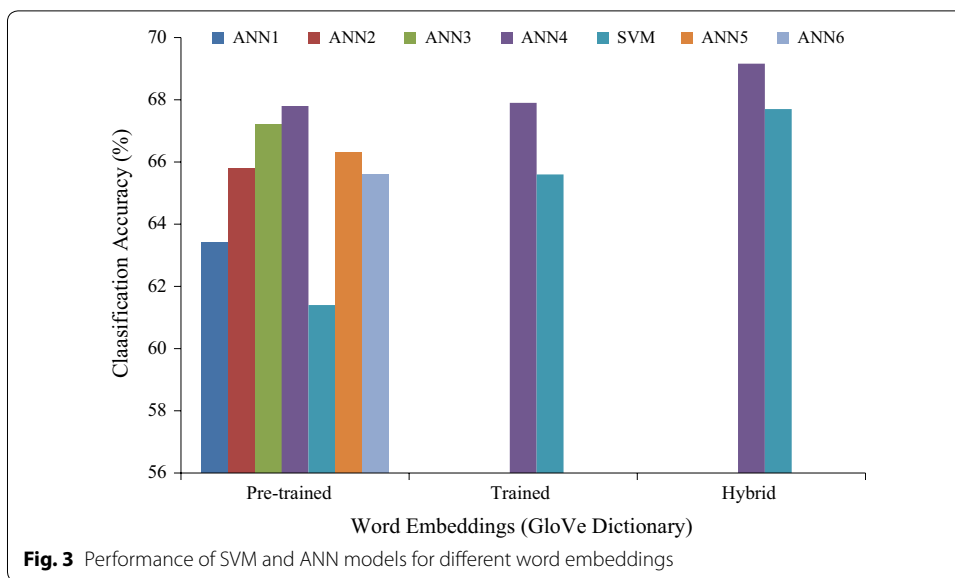
$$L_f = \frac{P(A \cap B)}{P(A) \times P(B)} \quad (4)$$

Results and discussion

Experiments are performed using Scikit learn library and keras and tensor flow modules on Python 3.6 version. The balanced data set were carefully prepared for the analysis with two category of emotions i.e. positive and negative. As we have only two sentiment classes for our dataset, linear SVM classifier was used for analysis. For SVM implementation, we set penalty parameter to 1 for training purpose. We tested the performance of the SVM on the vectored tweets using WE approach. Three categories of word embedding i.e. trained, pre-trained and hybrid were used to prepare the dataset. ANN with different configurations were tested on the data to find the best configuration. We employed different ANN configurations along with SVM and evaluated on three word embedding obtained from Glove representation by using different variety of dictionaries. Out of 10 employed ANN configurations, best 6 configurations that performed well are shown in Table 1.

Figure 3 shows that all the configurations are tested on pre-trained WE of tweets. ANN4 architecture with 2 hidden layer, 16 and 4 neurons in layer 1 and 2 respectively and trained with a batch size 256 has achieved the highest accuracy. Therefore, we only tested it along with SVM for trained and hybrid set. It can be seen that all ANN configurations with back propagation achieved better performance than SVM on pre-trained data set. Moreover, we have also evaluated the performance of SVM and ANN4 (as it performs best on word embedding representations) configurations for different n-grams representation of tweets data as shown in Fig. 4. We employed 6 n-gram models to be evaluated using ANN4 and SVM. It is clear in Fig. 4 that n-gram model with n=4 provides better performance for both SVM (76.5%) and ANN4 (79.4%) model.

Next to this, we performed the analysis using convolutional neural networks on word embedding feature vectors. Several CNN configurations were generated and tested for sentiment analysis on tweets. The best performance were achieved with a single hidden layer CNN with 128 neurons, 128 filters of size 3, 4 and 5, 128



dimensions of WE, with a drop out of 0.5 and a batch size of 128. The performance of CNN architecture on the dataset is given in Fig. 5.

Figure 5 illustrates that CNN performed well for both training and test dataset. It achieved an accuracy of 92.3% after 2700 iterations on the validation set which is a good score in comparison to ANN4 model that achieved the accuracy of 69.16% (highest among other models developed). Therefore, it is clear that CNN is more powerful than traditional ANN and is able to perform more accurately even for text data.

Next to this, we performed association analysis on the preprocessed words extracted from the tweets. We prepared a separate dataset for association analysis. The columns in the data set are the different categories of the words from the tweets

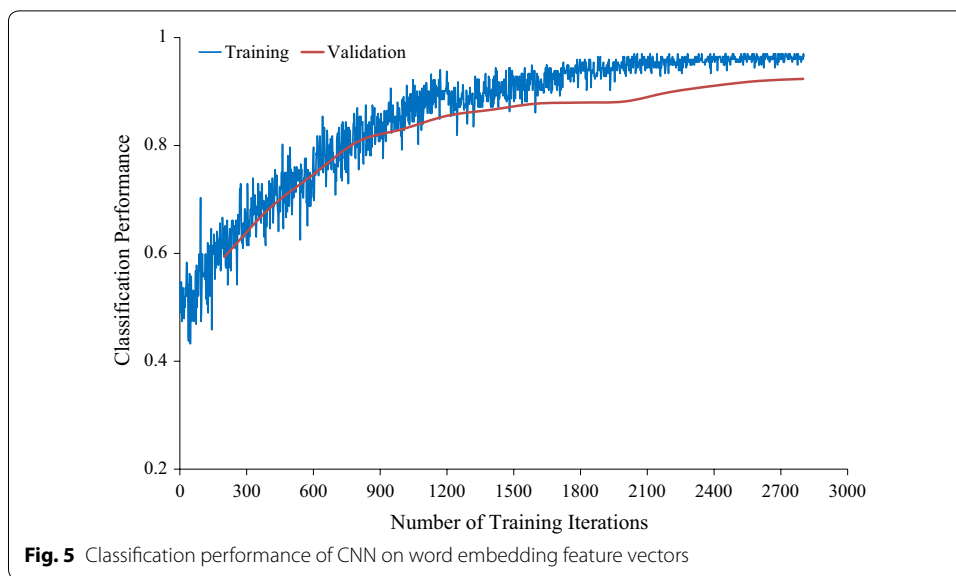


Table 2 The categories of words and number of tweets in which they occurred

S. no.	Category of words	Number of tweets in the category
1	CCB	72,165
2	FQL	66,589
3	CLN	34,568
4	IFC	55,648
5	FDC	48,457
6	LOB	37,847

and each row represents individual tweets. In order to prepare the dataset, each word is mapped to different categories of services based on domain expert knowledge. Several words within a tweet were evaluated to be mapped into word categories. For example, “it was an awesome meal” and “does not look like meal” both have the word “meal” but in different aspects. Therefore, other words from the tweet were also considered prior to a word assigned into respective category. These categories are Cabin Crew Behavior (CCB), Food Quality (FQL), Cleanliness (CLN), In-Flight Comfort (IFC), Flight Delayed/Cancelled (FDC) and Loss of Baggage (LOB). These categories and the respective number of tweets are shown in Table 2. The prepared dataset for association analysis had 7 dimensions (6 word categories and 1 additional sentiment category: 1-positive, 0-negative) as mentioned in Table 2. For every tweet, each column can have a value either 0 or 1. If the tweet consists of the words from a category then it will have a value 1 in the respective column otherwise 0. Further, association rules were generated separately for both positive and negative sentiments to understand the participation of each category on sentiment classes.

Further, strong rules were selected based on the support, confidence and lift values to illustrate the effect of all the categories on negative and positive sentiments. It can be seen in Table 3 that CCB and FQL has the highest influence in both the negative

Table 3 Participation of each category of words on sentiments

S. no.	Category of words	Influence of sentiments (%)	
		Positive	Negative
1	CCB	28	26
2	FQL	24	23
3	CLN	10	3
4	IFC	19	18
5	FDC	15	16
6	LOB	4	14

Table 4 Relevant rules illustrating the association between different categories of words

Rule no.	Association rule	Support	Confidence	Lift
R1	[CCB = 0 AND FQL = 0] → [Sentiment = 0]	0.3	0.86	1.34
R2	[FQL = 0] → [CCB = 0]	0.4	0.91	1.63
R3	[FQL = 1] → [CCB = 1]	0.4	0.93	1.69
R4	[FQL = 1 AND CCB = 1] → [Sentiment = 1]	0.3	0.86	1.35
R5	[FDC = 1] → [LOB = 1]	0.3	0.64	1.02
R6	[FDC = 0 AND CCB = 0] → [IFC = 0]	0.3	0.87	1.65
R7	[IFC = 0] → [Sentiment = 0]	0.5	0.92	2.35
R8	[CCB = 0] → [Sentiment = 0]	0.5	0.85	1.48
R9	[LOB = 1] → [Sentiment = 0]	0.2	0.69	1.06
R10	[FQL = 1] → [Sentiment = 1]	0.2	0.61	0.92

and positive sentiments. LOB and CLN has the lowest contribution in the positive and negative sentiments respectively.

In order to investigate the correlation between two or more categories of words, we analyzed the generated association rules for both positive and negative sentiments. It was found that CCB and FQL have a strong correlation with each other. Table 4 provides top 10 strong rules to illustrate the correlation between different categories that influence the sentiments.

It can be seen that CCB and FQL together affects the sentiment category. If food quality and cabin crew behaviour was good then tweet represented the positive emotions. It can also be seen in Table 4 (from R2 and R3) that food quality and cabin crew behaviour has strong correlation with each other. If food quality is poor then cabin crew behavior has higher chance to be negative (maybe due to a lot of complaints received from the passengers). If food quality is good then behavior of cabin crew was also nice. Both R2 and R3 have a good confidence and lift values to prove their significance. In R4, it is evident then good food quality and good cabin crew behaviour leads to positive sentiments. Rule 5 indicates that if the flight was delayed or cancelled then traveler complained about loss of baggage. Further, R6 stated that Inflight comfort and behaviour of cabin crew affects the passengers emotions as well. Moreover, R7 stated that if in-flight comfort was bad then it leads to negative emotions in most of the cases (0.5 support value). Besides in-flight scenario, another important category that leads to negative emotions is loss of baggage (R9). It has been observed by R5 and R6 that delay/cancellation in flight

or change in flight time is another aspect of negative emotions. Also, in flight delayed cases, people spotted the behavior of airline staff to be rude that provoked people to post negative tweets. Therefore, it is clear that besides developing a sentiment classification model using any powerful machine learning approach, it is important to understand the factors behind those sentiments. Association analysis is certainly helpful in achieving such goal and can be helpful to provide suggestions for organization to improve the customer's experience.

There are certain limitations of the present study that needs to be discussed with a future scope of the work. We have only selected the English language tweets for this study. There are several tweets in different languages for several airlines on Twitter. Therefore, the outcomes of the study does not involve the opinion of all nationalities around the world (as a lot of people do not tweet in English language in several countries). Therefore, analyzing tweets of other language will definitely explore this study and can provide better reasons for customer satisfaction to airline industry.

Conclusion

The study analyzed twitter data extracted with respect to airline industry. Several popular major airline companies are selected across the world based on their followers and number of tweets on the Twitter. Tweets were extracted, preprocessed and converted into feature vectors (numerically suitable form) using n-gram and GloVe dictionary (WE) approach for analysis. Initially, several ANN models were developed and tested along with SVM on pre-trained, trained and hybrid word embedding dataset. ANN4 model found more accurate among all developed ANN architectures and SVM models. Further, ANN4 and SVM were also evaluated for different n-gram representation of tweets data. It was found that n-gram model with $n=4$, provides better performance for both ANN4 and SVM model. At the end, we developed a CNN model to analyze the sentiments in the tweets and it provides a drastic improvement in the performance of classification model. Further, association analysis was performed on a preprocessed data set to understand the reasons behind the sentiments. Therefore, the study recommends to use association analysis on tweets data to gain insight about the tweets so that it can help the respective airlines to improve their customer experience.

Abbreviations

ASPs: airline service providers; SVM: support vector machines; ANN: artificial neural networks; CNN: convolutional neural networks; ReLU: rectified linear unit; WE: word embeddings; DBA: dictionary based approach; CBA: corpus based approach.

Acknowledgements

This work was financially supported by the Ministry of Education and Science of Russian Federation (Government Order 2.7905.2017/8.9).

Authors' contributions

SK designed the idea, prepared the draft of manuscript, performed experiments. SK and MZ wrote the article. Both authors read and approved the final manuscript.

Funding

The research received no external funding.

Availability of data and materials

The datasets used in this study would be made available after a reasonable request from the author.

Competing interests

The authors declare that they have no competing interests.

Received: 27 March 2019 Accepted: 7 July 2019

Published online: 17 July 2019

References

1. Number of smart phone users: <https://www.statista.com/statistics/330695/number-of-smartphone-users-world-wide/>. Accessed 06 May 2019.
2. Flight is best mode of travel: <https://www.flightnetwork.com/blog/10-reasons-flying-still-best-way-travel/>. Accessed 06 May 2019.
3. Twitter <https://en.wikipedia.org/wiki/Twitter>. Accessed 23 Jan 2018.
4. Abel F, Gao Q, Houben G-J, Tao K (2013) Twitter-based user modeling for news recommendations. In: Rossi F (ed) IJCAI 2013, proceedings of the 23rd international joint conference on artificial intelligence, Beijing, China, August 3–9, 2013. IJCAI/AAAI.
5. Number of flights in a day: <https://www.quora.com/How-many-airplanes-fly-each-day-in-the-world>. Accessed 25 Jan 2018.
6. Han J, Kamber M. (2006) Data mining: concept and techniques. 3rd ed.
7. Zomaya AY, Sakr S. Handbook of big data technologies. Berlin: Springer; 2017. <https://doi.org/10.1007/978-3-319-49340-4>.
8. Wilson T, Hoffmann P, Somasundaran S, Kessler J, Wiebe J, Choi Y, Cardie C, Riloff E, Patwardhan S. OpinionFinder. In: Proceedings of HLT/EMNLP on interactive demonstrations, association for computational linguistics. 2005. p. 34–5.
9. Kontopoulos E, Berberidis C, Dergiades T, Bassiliades N. Ontology-based sentiment analysis of twitter posts. Expert Syst Appl. 2013;40(10):4065–74.
10. Turney PD. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th annual meeting of the association for computational linguistics. 2002. p. 417–24.
11. Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. Lexicon-based methods for sentiment analysis. Comput Linguist. 2011;37(2):267–307.
12. Baccianella S, Esuli A, Sebastiani F. SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Calzolari N, Choukri K, Maegaard B, Mariani J, Odijk J, Piperidis S, Rosner M, Tapias D. (eds) Proceedings of the seventh international conference on language resources and evaluation (LREC'10). European Language Resources Association (ELRA), Valletta, Malta. 2010. p. 2200–04.
13. Miller GA. WordNet: a lexical database for English. Commun ACM. 1995;38:39–41.
14. Deerwester SC, Dumais ST, Landauer TK, Furnas GW, Harshman RA. Indexing by latent semantic analysis. J Am Soc Inform Sci. 1990;41:391–407.
15. Zhang W, Xu H, Wan W. Weakness finder: find product weakness from Chinese reviews by using aspects based sentiment analysis. Expert Syst Appl. 2012;39(11):10283–91.
16. Cambria E, Schuller B, Xia Y, Havasi C. New avenues in opinion mining and sentiment analysis. IEEE Intell Syst. 2013;28(2):15–21.
17. Paltoglou G, Thelwall M. A study of information retrieval weighting schemes for sentiment analysis. In: Proceedings of the 48th annual meeting of the association for computational linguistics (ACL '10). 2010. p. 1386–1395.
18. Yessenalina YY, Cardie C. Multi-level structured models for document-level sentiment classification. In: Proceedings of the 2010 conference on empirical methods in natural language processing (EMNLP '10). 2010. p. 1046–56.
19. Zhang D, Zeng J, Li F-Y, Wang W. Zuo, Sentiment analysis of Chinese documents: from sentence to document level. J Am Soc Inform Sci Technol. 2009;60(12):2474–87.
20. Zhang W, Yoshida T, Tang X. Text classification based on multi-word with support vector machine. Knowl Based Syst. 2008;21(8):879–86.
21. Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the conference on human language technology and empirical methods in natural language processing (HLT '05), Association for Computational Linguistics. Morristown, NJ, USA. 2005. p. 347–54.
22. Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis. Comput Linguist. 2009;35(3):399–433.
23. Serrano-Guerrero J, Olivás JA, Romero FP, Herrera-Viedma E. Sentiment analysis: a review and comparative analysis of web services. Inf Sci. 2015;311:18–38.
24. Ojokoh BA, Kayode O. A feature-opinion extraction approach to opinion mining. J. Web Eng. 2012;11(1):51–63.
25. Thet TT, Na J-C, Khoo CSG. Aspect-based sentiment analysis of movie reviews on discussion boards. J Inform Sci. 2010;36(6):823–48.
26. Rushdi-Saleh M, Martín-Valdivia M, Montejó-Ráez A, Ureña López L. Experiments with SVM to classify opinions in different domains. Expert Syst Appl. 2011;38(12):14799–804.
27. Ye Q, Zhang Z, Law R. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert Syst Appl. 2009;36(3):6527–35.
28. He Y, Zhou D. Self-training from labeled features for sentiment analysis. Inf Process Manag. 2011;47(4):606–16.
29. Xianghua F, Guo L, Yanyan G, Zhiqiang W. Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. Knowl Based Syst. 2013;37:186–95.
30. Kim K, Lee J. Sentiment visualization and classification via semi-supervised nonlinear dimensionality reduction. Pattern Recogn. 2014;47(2):758–68.
31. König AC, Brill E. Reducing the human overhead in text categorization. In: Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining—KDD '06. New York: ACM Press; 2006. p. 598–03.

32. Dutta DD, Sharma S, Natani S, Khare N, Singh B. Sentiment analysis for airline twitter data. *IOP Conf Ser Mater Sci Eng.* 2017;263(4).
33. Rane A, Kumar A. Sentiment classification system of twitter data for US airline service analysis. In: *Proc: 42nd IEEE computer software and applications conference, COMPSAC 2018.* Tokyo, Japan. p. 769–73.
34. Sternburg F, Pedersen KH, Ryelund NK, Mukkamala RR, Vatrappu R. Analysing customer engagement of Turkish airlines using big social data. In *Proc: 2018 IEEE international congress on big data (Big Data Congress), San Francisco, 2–7 July 2018.*
35. Ashi MM, Siddiqui MA, Nadeem F. Pre-trained word embeddings for Arabic aspect based sentiment analysis of airline tweets. *Adv Intell Syst Comput.* 2019;845:245–51.
36. Dekharghani R, Mercan H, Javeed A, Saygin Y. Sentimental casual rule discovery from Twitter. *Expert Syst Appl.* 2014;41:4950–8.
37. Pennington J, Socher R, Manning CD. 2014. GloVe: Global Vectors for Word Representation.
38. Zhang Y, Wallace BC. A sensitivity analysis of convolutional neural networks for sentence classification. *arXiv: 1510.03820v4.*
39. Collobert R, Weston J, A unified architecture for natural language processing: deep neural networks with multitask learning. In: *Proceedings of the 25th international conference on machine learning.* 2008. p. 160–167.
40. Rakesh A, Ramakrishnan S. Fast algorithms for mining association rules in large databases. In: *Proceedings of the 20th international conference on very large data bases.* September 12–15. 1994. P. 487–99.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
