Journal of Theoretical and Applied Information Technology
31st August 2017. Vol.95. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A FRAMEWORK FOR SEMENTIC LEVEL SOCIAL SENTIMENT ANALYSIS MODEL

[1]MADHU BALA MYNENI, [2]L V NARASIMHA PRASAD, [3]J SIRISHA DEVI

[1]Professor&Dean (Research), Department of CSE, Institute of Aeronautical Engineering, India

[2]Professor, Department of CSE, Institute of Aeronautical Engineering, India

[3]Professor, Department of CSE, Institute of Aeronautical Engineering, India

E-mail: [1]baladandamudi@gmail.com, [2]lvnprasad@yahoo.com, [3]siri.cse21@gmail.com

ABSTRACT

Social sentiment analysis is playing a vital role in analytics applications like product assessments, people opinions on sudden events and disaster assessments etc. Now the current research is focusing on dynamic big data analysis. The rich sources of dynamic data are twitter, face book, linkedln, snapchat, instagram, reddit and e-commerce web resources. In this paper the importance of semantic level social sentiment analysis with issues, tools and algorithms and machine learning algorithms role are discussed. A case study on Indian railway passenger tweets analysis is discussed and finds the sentiment of passengers on railway services.

Keywords: *Social Sentiment, Machine Learning, Text Processing*

1. INTRODUCTION

The massive growth of social media has given new thrust to the field of sentiment analysis, which is the computational action of sentiments on text data. Big data analytics is state-of-the-art research area in the sentiment analysis for addressing many promising areas. Big Data for sentiment analysis is a crowed sourcing data which can create in social media like forums, blogs, micro blogs (twitter) and so on. Now a day's, people share knowledge, thoughts and experiences in social networks to the world in the form of structured, semi-structured and unstructured data. Now the social media has become an exceptional part of everyone's life, to express their opinions about various issues. Due to all these scenarios, it opens up huge scope for sentiment analysis or opinion mining. For example to assess the public responses on products or movie reviews leads to decide changes or improvements to the same. Thus, social media big data research has become a very essential dynamic data source. But the challenge lies in this is extracting only relevant information from the big data that would help correctness in sentiment.

Dynamic sentiment analysis frameworks comprise the make use of natural language processing (NLP), text processing and extraction of subjective information from dynamic data sources [1]. Dynamic sentimental analysis is concerning to get the actual opinions of public towards specific goods, services provided, movies, current news, about organization, sudden events, and other issues. It aims to decide the opinion of a person about an issue or the appropriate polarity of the data. The major task in sentiment analysis is to categorize the sentiment of the given text towards positive, negative or neutral as final conclusion [1].

Lexicon based approach works on a hypothesis that the group polarity of a document or sentence is the sum of polarities of the individual words or phrases. The current research on this focuses on pre-process and algorithms applied on document type of data. The current research in this area focuses on algorithms for querying a search engine with major focus on text expressions for query string to categorize the outcome text [2], Text classification using lexical relations [3], A specific word list to categorize the text with an assumption those words have related polarity and orientation [4], on twitter data by using lexicon based practice [5,6], semantic sentiment on dynamic data source [7,8]. Sentiment analysis on micro blog data is more challenging because the text nature is in short length, familiar words, short words, misspelled and emoticons. Here with the experimental works with that most of existing techniques does not scale to big data sets effectively. So lexicon based techniques take more time for algorithm execution and hence these are not suitable for big data analysis.

The related research work in sentiment analysis includes the static data analysis. The tools,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

techniques, the list of resources and the current research are discussed below. The sentiment analysis is performed on different data sources like ecommerce sites, social networks and SMS etc. The various available dictionaries for positive and negative words and promising techniques to assess the sentiment are adopted by many researches.
Tools for sentimental Analysis:
The existing tools and techniques used for sentiment analysis on text data are refered as follows:

EMOTICONS:

Emoticons contained in the text based sentiment analysis are not appropriate for all the cases. In addition to this other attributes also need to be considered. Combination of all these are used for building a training dataset in supervised machine learning techniques [12]. The emoticon based analysis is not sufficient to find the sentiment on text, this is useful to train the model.

Linguistic Inquiry and Word Count (LIWCS):

According to the word count and dictionary words, the sentiment classified as emotional, cognitive, and structural components of a text [13]. For dynamic data, the limited dictionary words are not all the time sufficient to define the sentiment.

SENTI STRENGH:

LIEC dictionary with new features to find the sentiment strength and weakness by assigning scores to negative and positive phrases in text by sentiment lexicon. The drawback in this is it varies from context to context.

SentiWordNet:

Lexical dictionary and scores obtained by semi-machine learning approaches are used to find sentiment. Each synset is associated to three numerical scores positive (Pos), negative (Neg) and objective (Obj) by opinion mining, which is based on WordNet [14]. For example, a given synset s = [bad, wicked, terrible] has been extracted from a tweet. The scores in SentiWordNet is assigned for positive as 0.0, for negative as 0.8 and for objective as 0.15. These are evaluated with dictionary words. For assigning polarity scores of a word, it considers the average scores of associated synsets of the given text and the tweet is consider as positive while comparing with negative score. The scores of objective tweets is not considered for sentiment.

Sentic Net:

Natural language processing (NLP) approach is used for inferring the polarity at semantic level. Artificial intelligence and semantic Web techniques used in tools like entropy-weighted genetic algorithm (EWGA)[15], Feature relation network (FRN)[16] and parallel relations [17].

Mostly, SenticNet was used to measure the polarity level in suggestions or reviews and used to label the data with different opinions and then defined as positive, negative and neutral.The concept level automatic sentiment is assessed effectively with this tool, even though the drawback in this is noise also considered for analysis.

HAPPINESS INDEX:

Evaluate the happiness index based on English norms in the given text. Each word Frequency and the average in the given text using English words (ADEW) [18,19]. The happiness index score is used to evaluate the song or movie title,lyrics or story etc. To find the polarity for sentiment, the happiness index is considered as the range of [1..5] as negative and [6..9] as positive

PANAS-t:

Psychometric scale is used to define the mood fluctuations of the users. The scale is termed as positive affect Negative Affect scale (PANAS). This is based on eleven moods as joviality, serenity, assurance, fear and surprise as positive affect words and fatigue, sadness, guilt, shyness, hostility and attentiveness. It computes the sentiment for entire data by considering the associated words for specific sentiment. Polarity of the given time period data is used to find the change over time and which fits in the scale of [-1.0, +1.0]. For example, P(cool) of the given tweets is 0.234, then it is inferred as the sentiment related to cool is 25% increase in that period. Similarly, P(cool) = - 0.012 means that the sentiment s decreased by 1.2%.

iFeel Web System:

iFeel web system is a combination of linguistics and psychology. For example, for the sample input “we are feeling too bad :(”. In this example, according to emoticons study it detects strong negative sentiment and according to linguistics also negative sentiment. This system will be most ppropriate tool to the researchers in this

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

field to compare the result in analysis of social network data. As a extension of this system, by adding the more moods to find the sentiment of opinions by seeing varied categories of opinions beyond positive and negative.

2. SOCIAL SENTIMENT ANALYSIS CHALLENGES

With the explosion of social media data the research focus in NLP with a new band called sentiment or emotion analysis has focuses on determining people's opinions on specific event or experiences. However sentiment analysis for current or sudden events for predicting or assessing the results is challenging task. Though many models are available, building a model on twitter streams of dynamic data is empirically challenging task. Many existing opinion mining systems, frameworks and tools have been developed for concepts like user reviews on products or movies or current events with respective attributes.

In this paper, one of the major focuses is challenges included in the sentiment analysis of dynamic event data like elections and disasters. The key challenges are changes in topic of conversation at that moment and the opinions on specific social media posts or announcements.

Here some of the challenges included in the analysis during monitoring of the presidential election. In this analysis, Twitris system has used. It is used in previous US elections including the 2012 [30] and 2016 election [31]. In this analysis, a multi class supervised classifier has created to classify as positive, negative and neutral opinions about each participating candidate.

In this analysis, one observation is same tweet on one issue can be treated as positive for one candidate but the same is negative for another. Hence the sentiment of a tweet is dependent on multi attributes.

This model included multiple issues as encompassing budget, finance, education, energy, environment, healthcare, immigration, gun control, and civil liberties and multiple candidates including Bernie Sanders, Donald Trump, Hillary Clinton, John Kasich, and Ted Cruzs. The model for this analysis has built with traditional machine learning SVM model with uni to three gram and hashtags of positive and negative for each candidate.

Out of this work, the high accuracy deep learning model on this data is convolutional neural network. The challenges in multi attribute sentiment model are:

2.1 Dynamic dataset

The active or sudden events face with dynamic nature in social data due to timely changes on people opinion on new aspects or new context of unfolded events. So the dynamic model is required to overcome this issue with frequent updates in training data set and distribution of a different training and test data sets.

In general, the classification model approaches equal distribution of train/ test data sets but for many real world problems the test or target data changes over time. In this context, to overcome from fore mentioned challenges an active learning of the model is necessary and including more influential tweets in the training dataset.

2.2 Target dependent sentiment

Furthermost sentiment analysis systems work in a target independent manner, which yields to poor results on active events. In most of the tweets text, the targets are not specific because a casual conversation includes multiple candidate names.

For example, "I am getting so nervous because I want Trump to win so bad". Hillary scares me to death and with her America will be over" and "I don't really want Hillary to win but I want Trump to lose can we just do the election over".

In these cases, it may go to misclassification. So to overcome this target dependent analysis is grouped into syntax based on positive words tagging or syntax parsing as features and context based is defined on left and right of each target [21].

2.3 Relevant Data

In supervised classification the accuracy of classifier is dependent on cool supervision with weekly labeled training set. The training data automatically labeled in different approaches: emoticon like positive as :) and negative as :(are appended in tweets text: hashtags [23] for emotion identification.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The challenges included in this dynamic data set is same hashtag is strong for one group and week for another group. People can use the same hashtag as positive for one group and the same as sarcasm for other group. So due to this, it leads to incorrect labeling of tweets and decrease of accuracy of classifier.

2.4 Tweets content completeness

In general, all existing approaches are considered only tweets text and ignoring the content of URLs point to external sources. But the facts are saying in 2016 elections data 36% are referred external links and 2012 elections [22] 60% tweets contain URLs. So ignoring URLs leads to incompleteness in data and finding the sentiment is not accurate. Therefore, by incorporating the content with text, keywords and URLs title as features for analysis will cause the gain of performance.

2.5 Identification of sarcasm words

Presently, one of the major challenges in micro blogs is sarcasm. It leads to misclassification. To detect this, a refined tools and approaches are needed to be identified. More recently, [24] employ deep neural network (pre-trained convolutional neural network) for identifying sentiment, emotion, and personality features for sarcasm detection. Looking closer at these works, they mostly focus only on detecting the sarcasm in the text and not on how to cope with it in the sentiment analysis task.

This raises the interesting question about how sarcasm may or may not affect the sentiment of the tweets and how to deal with sarcastic tweets in both the training and prediction phases. An algorithm has proposed to recognize the common form of sarcasm which flips the polarity in the sentence [26]. These kinds of polarity reverser sarcastic tweets often express the positive (negative) sentiment in the context of a negative (positive) activity or situation. However, determining the scope of sarcasm in tweets is still challenging [25].

In fact, the polarity of sarcasm may apply to part of a tweet or its hashtags but not necessarily the whole. As a result, dealing with sarcasm in the task of sentiment analysis is an open research issue worth more work. Based on our observation, 7% of Trump's tweets and 6% of Clinton's tweets are sarcastic. Among these sarcastic tweets, 39% and 32% of them were classified incorrectly by our system.

In terms of the training set, our hypothesis is that excluding the sarcastic instances from the training set will remove the noise and improve the quality of our training set.

2.6 Outcome Interpretation

Emotion study is one of the advancements in sentiment analysis with fine granularity level. The outcome of sentiment is expressed as positive, negative and neutral levels where as emotions as sad, joy, anxiety, kind and like. These emotions are categorized according to sentiment.

The above discussed issues affect the quality of sentiment analysis. The role of active users is considered in 2012 election result prediction. For effective sentiment analysis, the role of active users is more [22]. They influence the people in the real world towards their opinion. So the events assessment and predictions are possible by considering user level sentiment at normalized instead of tweet level. Hence both these are required for prediction of result.

2.7 Location of users

In existing tools and applications are developed for different levels of granularities [27]. For prediction models, the location wise influence and trends are needed to be considered. The user location is estimated by considering the tweet geographical location with latitude and longitude attributes or user profile location.

The research gap with current practices are: Trusted data acquisition; Effective data storage and processing environment; Handling crowed data; Theme based data categorization; Automated positive and negative word lists; and calculation of sentiment score to assess the opinions.

3. SENTIMENT ANALYSIS FRAMEWORK

The figure 1 shows the architecture of a multi attribute semantic sentiment analysis system. It incorporates a multiple attributes source as social networks (twitter) from this crowd source of data the texts to be preprocessed, categorized according to theme by using social graph clustering and dynamic word list for scoring polarity by using machine learning algorithm.

Journal of Theoretical and Applied Information Technology
31st August 2017. Vol.95. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The functionality of all functional blocks is:

3.1 Twitter Crawler

The data extraction is performed through this for theme relevant data from twitter API. For achieving this different approaches like #tagword; longitude or latitude for location based; or specific duration for finding about sudden events.

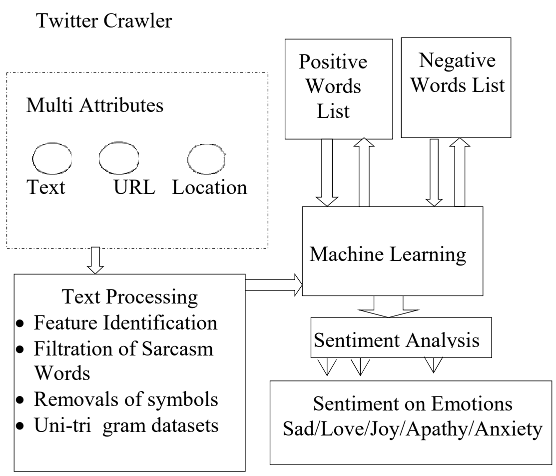

Figure 1: Framework for semantic sentiment model

3.2 Data Extraction

The Twitter API, through which, it is easy to extract huge collection of tweets are available among which as a sample data only 1% of tweets tweeted at that time are exacted. The attributes of the data for this analysis includes the text, longitudinal, latitudinal and URL.

The relevant data extraction is possible with related famous hashtags of theme, so here #railMinIndia hashtag is used for sample data. Now the extracted data contains 32 attributes, which includes user location, date, time of post, text of opinions, URLs of images or news etc.

Data Size

The twitter allows the text length of about 140 characters; the location is indicated with longitude and latitude. In this proposed model, the part of data is sampled in to multiple sets according to location and sentiment analysis is performed on text data. Twitter users post their opinions in the form of short messages or URLs on numerous topics.

3.3 Text Processing

The extracted theme based tweets are not ready to process. It contains all noise also. On this pre-process is performed to remove all noisy data, unwanted symbols and URLs. Now on this processed data the sequence of steps as tokenization, term frequency of each word, document term frequency are performed by using suitable algorithm to calculate the score of polarity.

In general, twitter messages are highly misspelled and informal speech and about multiple domains. For any analysis on text pre-process is one of the essential steps. It makes the corpus ready for analysis. The pre-process includes the removal of small length words, symbols, punctuations and white spaces etc.

After performing the cleaning operation, the data is tokenized at different levels as uni gram (single word), bi gram (Two level Associate words) and tri gram (Three word level as pre and after words for each word combination). Identification and removal of sarcasm words from the word list in the text.

3.4 Machine learning Algorithm

Machine learning algorithms are playing a vital role to automate the computing process in any system. In dynamic sentiment analysis, the preparation of positive and negative word lists related to analysis theme is critical process. According to the theme of data, the positive and negative word lists are needed to update according to new theme.

Natural Language Processing is stand at the level of semantics based context aware paradigms like convolutional networks with multiple kernel learning [24], recurrent belief networks and statistical learning theory. Recent studies of NLP focus on social network analytics and sentiment and emotional analysis.

At early stages, this is done by kNN classifier on amazon reviews data by considering high frequency words of ordered sequences. An experimental analysis on the imbalance of sentiment levels between user given star rating and reviews are evaluated on amazon data.

Textual features at linguistic level are assessed in two dimensions as relevance and representativeness. N- gram features as uni gram, bi gram and tri gram. Pragmatic features as emoticons and punctuation marks are used for classification problems. By using support vector machine (SVM)

Journal of Theoretical and Applied Information Technology
31st August 2017. Vol.95. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

classifier and logistic regression on multiple types of data like unigram feature set, pragmatic feature set as emoticons data and lexical features based on dictionary words. The most ample work with varied range of supervised and unsupervised classification methods as logistic regression, J48 classification tree, JRip rule-based, SVM, AdaBoost, Multilayer Perceptron, and Na¨ıve Bayes on dictionary words. One of the existing work relies on the LIWZ dictionary words with the expansion of new feature for Social network context [13]. In this work to strengthen the sentiment the booster words like "very" and week words as "somewhat" are added to the existing dictionary words. Along with emoticons and repeated punctuations are also strengthen the sentiment. In this work the data has been used from different sources as Twitter, Myspace, Digg and YouTube Comments.

Deep convolutional neural networks

For example, Sarcasm detection on micro blog posts has been implemented on uni gram and emoticons features with supervised [26] and unsupervised pattern mining approaches [25]. Using deep CNN very effective and optimistic feature set is generated automatically in smaller size for detection of sarcasm behavior on large corpus. The model is built with pre trained sentiment model, emotion model for identification of sarcasm text. This can be improved with machine learning to classify as sarcastic and non-sarcastic text.

3.5 Sentiment Analysis

For the given text, the positive, negative and neutral polarity is determined to assess the opinion of people on the event. Sentiment calculation is done for every tweet and a polarity score is given to it. If the score is greater than 0 then it is considered to a positive sentiment on behalf of the user, if less than 0 then negative else neutral. The polarity score is calculated by using algorithm 1 by using programming model.

The senti score algorithm is used to estimate the score of sentiment for each word.

Algorithm 1: Senti Score

```
Input: Tweets, Word_Dict
Output: Sentiment (positive, negative or neutral)
BEGIN
1) For each tweet Ti do
2) P_Score = 0;
3) For each word Xj in Ti that exists in Word_Dict.
  If polarity(Xj) = blind negation then Return
negative.
   Else
      Else If polarity[Xj] = positive
              P_Score=P_Score + 1
      Else If polarity[Xj] = negative
             P_Score =P_Score - 1
      Else If polarity[Xj] = negative
            P_Score= P_Score - 0.5
      If polarity[Xj] = negation
           P_Score = P_Score* -1.
 If P_Score of Ti >0
          Sentiment = positive.
  Else If P_Score of Ti<0
          Sentiment = negative.
  Else Sentiment = neutral
4) Return Sentiment
5) END
```

4. APPLICATIONS OF SENTIMENT ANALYSIS: RAILWAY PASSENGER TWEETS DATA ANALYSIS

For example, the sentiment analysis on railway passengers tweets for assessing their quality of service. Now a days, quality of service impacts lot of changes business world. In public transport review of passenger opinions has vital role to continue the existing services and assessing the quality of service. In this analysis, the related data is collected by using #railMinIndia hashtag, which is officially released by Ministry of railways of government of India. The Twitter data is analyzed by using linguistic techniques, and then it converts likely to forecast possible future activities. The passenger tweets contains future tense then it happens in future; if location is specified then a person can visit that location within a week to perform the mentioned activity. From these sorts of assumptions the model can build to forecast the future needs of the passengers. The same can imply for officers to plan the trips to that location.

Table 1. Sample Twitter Attributes Of Railway Passengers

Text	Date	Name	Location
Wrst Exp in Kolkata Rajdhani (runing late by 8hrs).we hv been served only dal rice w/o curd. Poor Service quality.	11-12-2016 11:39	manish jiit17	Noida
58% of rail tickets are sold online: Who says India isn't ready for... https://t.co/fVw2BaiA2p by #RailMinIndia via @c0nvey	11-12-2016 11:34	Debashis0907	Dhanbad, India
#RailMinIndia ....my ticket was cancelled on 29 Nov but still waiting for refund..	11-12-2016 11:20	PyaShr	jaipur
58% of rail tickets are sold online: Who says India isn't ready for... https://t.co/ng7Ij388fH by #RailMinIndia via @c0nvey	11-12-2016 10:07	Scribed Life	India
58% of rail tickets are sold online: Who says India isn't ready for... https://t.co/fW6Cw2crQL by #RailMinIndia via @c0nvey	11-12-2016 09:49	RandhirBharat	Pune
58% of rail tickets are sold online: Who says India isn't ready for... https://t.co/0YXow7MBEY by #RailMinIndia via @c0nvey	11-12-2016 09:17	amit42809	Silchar
58% of rail tickets are sold online: Who says India isn't ready for... ttps://t.co/AYvVwTNDDg by #RailMinIndia via @c0nvey	11-12-2016 09:11	32_mishra	Dhanbad
MR @sureshpprabhu Inaugurates Two Coaches For Cancer Treatment On Lifeline... https://t.co/GHS4rAs7f0 by #RailMinIndia via @c0nvey	09-12-2016 14:44	Sonu8676869885	
MR @sureshpprabhu Inaugurates Two Coaches For Cancer Treatment On Lifeline... https://t.co/3G9s53RJfd by #RailMinIndia via @c0nvey	09-12-2016 13:48	amit42809	Silchar

From the extracted data, the attributes considered for analysis are location related as longitudinal and latitudinal, text and URLs.

The missing data of location is imputed with proper location from correlated personal data [32]. First all the tweets are preprocessed for removal of unwanted symbols and noisy data. The text data is processed with sequential steps of removal of unwanted text, symbols, numbers, spaces and transform to unigram level data.

The pre-processing of tweets includes the following steps:

Replace all the emoticons with their sentiment polarity by looking up the emoticon dictionary
Step 1. Replace all URLs with a tag ||U||
Step 2. Replaces targets (e.g. "@John") with tag T||
Replace all negations (e.g. not, no, never, n't, cannot) by tag "NOT"
Step 3. Replace a sequence of repeated characters by three characters, for example, convert cooooooooool to coool.

Location wise data is shown in figure 2. It includes opinions on all different themes of railways as complaints suggestions on services. According to other attributes, the data is split according to other attributes. Now all data related to services theme can get as a single dataset at specified location.

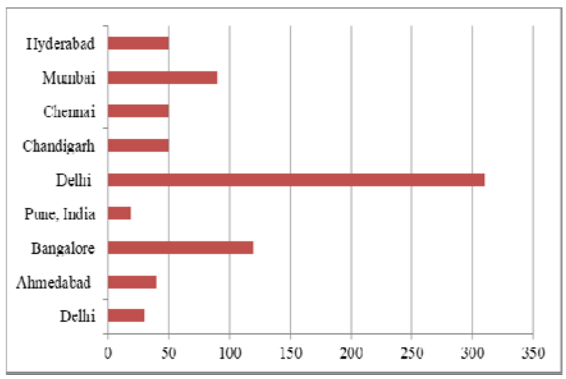

Figure 2. Location wise tweets data

The railway system needs a time to time survey on both upstream and downstream transport activity relevant to a particular geographic location. The authorities of particular sections of railways together form as closed networks.

For example, connections between intercity and local services may be posted on the website but are of interest to local providers seeking to improve connection services. Therefore it is necessary to identify those messages pertinent to the location or specific transport services for the task. Two approaches to identifying the location from social network data are either, to identify the current location of the person posting the message or to identify the message content.

It outlines the process involved, for an example of public transport messages analysis based on the fusion of information either in the message or

Journal of Theoretical and Applied Information Technology
31st August 2017. Vol.95. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

attached to it. The sample word cloud is shown in figure 3.

Figure 3. Sample Word cloud of passenger tweets

Automatic assessment of sentiment on text is performed using polarity score of each word at uni gram level. The previous works of author on floods impact assessment the location is one of the attribute considered for split of data [33]. Now from frequent word list, according to the service the positive and negative words are classified.

By using sentiment word list the analysis is performed based on the polarity score of each word. The experiment is done on total 11100 words. The table 2 shows the emotions for each category of sentiment. According to this the positive, negative and neutral scores are calculated.

Table 2: Emotions categories

Sentiment Type	Emotions
positive	Joy, love
negative	Anxiety, sadness
neutral	apathy

From railway passengers tweets, the positive and negative words are listed in table 3.

Table 3. Sample Positive and Negative words

Negative Words	Positive words
refund	Ready
delay	Sold
poor	Change
not	Coaches
worst	Selling
useless	Travels
late	Lifeline
action	Light
waiting	Response
broken	Soon
pathetic	Food
unauthorized	Help

The words in this context group are tagged as:

Neutral: Neutral words hold no value in the equation but do affect word count (n).

Negator: A character vector of terms reversing the intent of a positive or negative word.

Amplifier: A character vector of terms that increase the intensity of a positive or negative word.

De-amplifier: A character vector of terms that decrease the intensity of a positive or negative word.

A Sample Positive and Negative words identified from training data of passenger tweets. These words are identified as positive and negative based on the conversation about services provided by Indian railways.

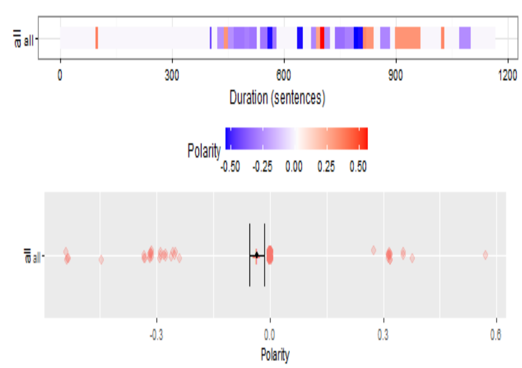

Figure 4. Polarity plot of railway passenger twitter data

The polarity plot is shown in figure 4. The average polarity score of passenger tweets is - 0.035 means a neutral sentiment are expressed about railway services. It indicates the passengers are showing very neutral opinion on existing services.

5. CONCLUSION

In this work, a semantic level sentiment analysis model has proposed for finding the sentiment and emotion on services of railways. It addressed the relevant data extraction, automated assignment of people opinions and finding the sentiment score on large corpus. This analysis helps in railway services to assess the quality of service and in decision making process to continue the same service in future.

In this work the major focus is on including multiple attributes of tweets along with the text as location, emoticons, and URLs for understanding

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the user profile and personality. Hence, it improves the insight analysis of passenger opinions.

REFRENCES:

[1] Liu, Bing (2012) Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, pp 1-167.

[2] Ting-Chun Peng and Chia-Chun Shih (2010) An Unsupervised Snippet-based Sentiment Classification Method for Chinese Unknown Phrases without using Reference Word Pairs, International Conference on Web Intelligence and intelligent Agent Technology Journal of Computing,2(8).

[3] Kamps, Maarten Marx, Robert J. Mokken and Maarten De Rijke, (2004) Using wordnet to measure semantic orientation of adjectives, Proceedings of 4th International Conference on Language Resources and Evaluation, pp 1115-1118.

[4] Andrea Esuli and Fabrizio Sebastiani, (2005) Determining the semantic orientation of terms through gloss classification, Proceedings of 14th ACM International Conference on Information and Knowledge Management, pp. 617-624.

[5] Prabu Palanisamy, Vineet Yadav, Harsha Elchuri (2013) Serendio: Simple and Practical lexicon based approach to Sentiment Analysis, Serendio Software Pvt Ltd.

[6] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll and Manfred Stede (2011) Lexicon based methods for Sentiment Analysis. Computational linguistics,37(2), pp 267–307 .

[7] Peter Turney and Michael Littman (2003) Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems 21(4) pp 315–346.

[8] Minqing Hu, Bing Liu. Mining and Summarizing Customer Reviews, Department of Computer Science, University of Illinois at Chicago, Research Track Paper.

[9] Long-Sheng Chen, Cheng-Hsiang Liu, Hui -Ju Chiu (2011) A neural network based approach for sentiment classification in the blogosphere, Journal of Informetrics pp 313–322.

[10] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll and Manfred Stede (2011) Lexicon based methods for Sentiment Analysis. Computational linguistics,37(2), pp 267–307.

[11] Michael Wiegand et. Al. (2010) A survey on the role of negation in Sentiment Analysis. Proceedings of the workshop on negation speculation in natural language processing, pp 60–68.

[12] Read, J. (2005) Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In ACL Student Research Workshop, pp 43-48.

[13] Tausczik, Y.R. & Pennebaker, J.W (2010) The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology, 29(1), pp 24-54.

[14] Miller, G.A (1995) WordNet: a lexical database for English. Communications of the ACM, 38(11), pp 39-41.

[15] Mayfield, E., & Rosé, C.P (2012) LightSIDE: Open Source Machine Learning for Text Accessible to Non-Experts. In the Handbook of Automated Essay Grading, Routledge Academic Press.

[16] Hassan, A., Abbasi, A., Zeng, D (2013) Twitter Sentiment Analysis: A Bootstrap Ensemble Framework, Proceedings of the ASE/IEEE International Conference on Social Computing, pp 357-364.

[17] Abbasi , A (2010) Intelligent Feature Selection for Opinion Classification. IEEE Intelligent Systems, 25(4), pp 75-79.

[18] Dodds, P.S. & Danforth, C.M (2009) Measuring the happiness of large-scale written expression: songs, blogs, and presidents. Journal of Happiness Studies, 11(4), pp 441-456.

[19] Bradley, M.M., Lang, P.J (1999) Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. Technical Report C-1, Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

[20] Wang, C.J., Tsai, M.F., Liu, T., Chang, C.T (2013) Financial Sentiment Analysis for Risk Prediction. In Proceedings of the Sixth International Joint Conference on Natural Language Processing pp 802-808.

[21] D. Vo and Y. Zhang, "Target-Dependent Twitter Sentiment Classification with Rich Automatic Features," IJCAI, 2015.

[22] L. Chen, et al., "Extracting Diverse Sentiment Expressions with Target-Dependent Polarity from Twitter," Sixth International AAAI Conference on Weblogs and Social Media, 2012.

[23] W. Wang, et al., "Harnessing twitter 'big data' for automatic emotion identification," Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom), IEEE, 2012.

[24] S. Poria, et al., "A deeper look into sarcastic Tweets using deep convolutional neural networks," arXiv preprint arXiv: 1610.08815, 2016.

[25] D. Maynard and M. A. Greenwood, "Who cares about Sarcastic Tweets? Investigating the Impact of Sarcasm on Sentiment Analysis," LREC, 2014.

[26] E. Riloff, et al., "Sarcasm as Contrast between a Positive Sentiment and Negative Situation," EMNLP, vol. 13, 2013.

[27] Z. Ji, et al., "Joint Recognition and Linking of Fine-Grained Locations from Tweets," Proceedings of the 25th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2016.

[28] A. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," Proceedings of the National Academy of Sciences, 111(24), 2014, pp. 8788-8790.

[29] K. Lee, B. D. Eoff, and J. Caverlee, "Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter," ICWSM, 2011.

[30] http://tcrn.ch/29DNjon, http://bit.ly/TwBrexit

[31] http://j.mp/E16Cognovi, http://bit.ly/Ele2016

[32] Myneni Madhu Bala, M. Srinivasa Rao and M Ramesh Babu. "Sentiment Trends on Natural Disasters using Location Based Twitter Opinion Mining", International Journal of Civil Engineering and Technology, 8(8), 2017, pp. 09–19.

[33] Madhu bala Myneni, Narasimha Prasad L V and G Geetha Reddy, "Automatic Assessment of Floods Impact Using Twitter Data", International Journal of Civil Engineering and Technology, 8(5), 2017, pp. 1228–1238