RESEARCHERID
THOMSON REUTERS
[INIT: January 2017]
ISSN 2348 – 8034
Impact Factor- 4.022
GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

SENTIMENT ANALYSIS ON TRAFFIC IN INDIAN CITIES

Aruna Devi K[*1] & Nethra M2, Shruthi C D[2]
[*1]Faculty, Department of Computer Science (PG) Kristu Jayanti College, Bengaluru, India
[2]MCA Student, Department of Computer Science (PG) Kristu Jayanti College, Bengaluru, India.

ABSTRACT

In this paper, a Social Sentiment Analysis that looked at people's feelings about traffic in India's prominent metropolitan cities includes Bangalore, New Delhi and Mumbai is presented. We analyzed Sentiments of people about the Indian cities traffic with Real-time social media data. The sentimental analysis is the process of identifying and classifying views expressed in the form of text; in order the understood the writers view towards the particular subject or a product etc., in a negative, positive or a neutral way. This research is real-time traffic detection and monitoring system from twitter tweet stream analysis. It helps to discover the hidden value of text in order to penetrate the opinions of people. This system identifies the view about a particular topic and outperforms the traditional approaches.

Keywords *- Data mining; Sentiment Analysis; Social Media; Indian Cities Traffic; Twitter.*

I. INTRODUCTION

Today, the widespread use of technological innovations across all over the globe, like ubiquitous communication networks and highly distributed mobile technology, has made cities smarter than ever before. The Indian government spends great amount of efforts in Smart city research to ensure the cities are managed and governed in an efficient way. Smart city technologies are leveraged to integrate and analyze huge volumes of data to monitor citizens, for the sake of better governance. The data can be from traffic cameras, medical records, and urban sensors. With the exponential growth of Social Media in recent years, the Governments are realizing that it can be a great medium to better understand their citizens [1]. The information generated on social media can be observed as a source of citizen voice. Twitter is a popular micro-blogging service where users post status messages to share opinions on a variety of topics and express their personal feelings, is experiencing a dramatic increase of users, more so than other social media. The audience of Twitter varies from regular users to company representatives, celebrities, and politicians; therefore it is possible to collect the text posts of users from different social and interest groups. Twitter Sentiment Analysis can be a useful vehicle to provide deep insight into how citizens feel and thus has definite use for smart city monitoring and governance [2].

Sentiment analysis is a popular study in recent trends, because of the fact that social networking sites including online users are free to express their feelings, thoughts and impressions regarding a specific topic. Sentiment analysis aims at determining the attitude of a speaker or a writer with respect to a topic or the overall contextual polarity of a document. The attitude may be his or her or evaluation, judgment, affective state or intended emotional communication. Sentimental analysis is also used in marketing Industry as it is currently immersing to the new trends of businesses. Companies also extend their customer satisfaction analysis through the web, in order to gather a large amount of data. Sentimental analysis is carried out by many scientist and researchers as it is an emerging trend packed with lots of challenges in analyzing the feeling of the individual. These studies are targeted to Twitter, for tweet updates about a specific topic or brands of products. These systems collect raw data from twitter, and use the data as a corpus to be feed upon implementing and classifying methods. The technique uses natural language processing, text analysis and computational linguistics to identify and extract subjective information from the source materials.

Natural Language Processing is an application that explores how computers can understand and manipulate natural language text or speech to do useful things. NLP researchers aim at gathering knowledge on how human beings understand and use language so that appropriate tools and techniques can be developed to make computers understand and manipulate natural languages to perform the desired tasks. The foundations of NLP lie in a number

ISSN 2348 – 8034
Impact Factor- 4.022

of disciplines, viz. computer and information sciences, linguistics, mathematics, electrical and electronic engineering, artificial intelligence and robotics, psychology, etc. Applications of NLP include a number of fields of studies, such as machine translation, natural language text processing and summarization, user interfaces, multilingual and cross language information retrieval (CLIR), speech recognition, artificial intelligence and expert systems, and so on. Text Mining is the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends through means such as Statistical pattern learning. Text mining usually involves the process of structuring the input text from deriving patterns within the structured data, and finally evaluation and interpretation of the output. Basically text mining includes text clustering, text categorization, concept extraction, document summarization, sentimental analysis, and entity relation modeling.

The paper is organized as follows. In Section II, we review the prior works on Twitter Sentiment Analysis. In Section III, we describe the data used for constructing our sentiment analysis system. The details of our Twitter sentiment methodology are presented in Section IV. We describe our experimental results in Section V. Finally, we conclude our work and illustrate potential directions for future work in Section VI.

II. METHODOLOGY

Text Analytics is an interdisciplinary field which depends on information retrieval, data mining, machine learning, parameter statistics and computational linguistics. Sentiment analysis is an application of Text mining. It is a process of determining the emotional tone behind a series of words expressed in text format. It is used to gain an understanding of the attitudes, opinions and emotions of a person. The need to extract the insights of the people on a particular subject or a product becomes vital for the organizations to meet the demand challenges. The human language is complex. Training a machine to analyze the various grammatical nuances, cultural variations, slang and misspellings that occur in online mentions is a difficult process. The humans have the ability to quickly identify that the person was being sarcastic or not. By applying the contextual understanding to the sentence, the sentiment can be identified as positive or negative. Without contextual understanding, a machine looking at the sentence above might interpret in erroneous way. The objective of our research is to perform the basic task in sentiment analysis, to classify the polarity of the tweets posted in web is positive, negative, or neutral. The existing approaches to sentiment analysis can be classified into three main categories:

1. Knowledge-based techniques
2. Statistical methods and
3. Hybrid approaches.

Knowledge-based techniques [12] classify text by affect categories based on the presence of unambiguous affect words such as happy, sad, afraid, and bored. Statistical methods [13] leverage on elements from machine learning such as latent semantic analysis, support vector machines, "bag of words" and Semantic Orientation. The hybrid approach deploy machine learning, statistics, and natural language processing techniques to automate sentiment analysis on large collections of texts, from web pages, online news, internet discussion groups, online reviews, web blogs, and social media. The methodology followed in Sentiment Analysis of Traffic data is given in Fig. 1.

Preprocessing is an important task and critical step in Sentiment analysis. The preprocessing improves the classifier performance. Preprocessing text is called text normalization. It includes Stop word removal, Stemming, Feature Extraction, Feature Reduction and Classification. The most frequent words often do not carry much meaning. Examples: “the, a, of, for, in”. A stop word list for the respective application domain can be created. Those words can be removed from the data fetched from the social media. When English words like ‘drive’ can be inflected with a morphological suffix to produce ‘drives, driving, driven’, they share the same stem ‘drive’. It is useful to map all inflected forms into the stem. This is a complex process, since there can be many exceptional cases (e.g., department vs. depart, be vs. were). The most commonly used stemmer is the Porter Stemmer. There are many other algorithms available. After removing stop words and stemming it is necessary to extract the essential features for feeding it to the classifier. Feature ranking and extraction improves the speed and accuracy of the classifier. The classifier is constructed using supervised technique.

ISSN 2348 – 8034
Impact Factor- 4.022

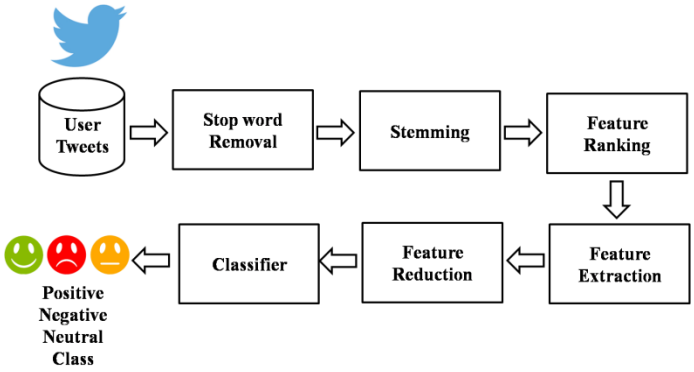

Fig. 1. ***Sentiment Analysis Methodology***

III. DATA PREPARATION

Traffic in the cities is divergent and is unpredictable. Road traffic jam has always been a challenging task for the metro cities. The traditional method for monitoring the traffic was by using sensors. Monitoring such heavy traffic using sensors or probe vehicles is also not feasible because of the high installation cost and certain environmental conditions. Twitter is a social networking and microblogging service that allows users to post real time short messages, called tweets. Tweets are microblogs, restricted to 140 characters in length. People use acronyms, post with spelling mistakes, use emoticons and other characters that express special meanings. The challenge of microblogging is the wide range of topic that is covered. Extracting sentiment from a piece of text such as a tweet, a review or an article can provide us with valuable insight about the author's emotions and perspective: whether the tone is positive, neutral or negative, and whether the text is subjective (meaning it's reflecting the author's opinion) or objective (meaning it's expressing a fact). For our research we acquired data using Twitter hashtags (e.g., #traffic, #fail, #news) to identify positive, negative, and neutral tweets to use for training three-way sentiment classifiers [11]. We collected 300 manually annotated Twitter data (tweets) using Rapidminer. A twitter connection is established through Rapidminer to access the tweets. Tweets of English language and most recent and popular are fetched. An example set consisting of 100 record set from the Twitter API which comprises the tweet text, the tweet ID, the number of re-tweets, the date of creation, the language, the geo-location, the used source of the tweet, and user information are downloaded. Out of 11 attributes from the dataset, Tweet text is taken for Sentiment Classification.

IV. RELATED WORKS

Alexander Pak and Patrick Paroubek [3] have performed linguistic analysis of the collected corpus and explain discovered phenomena. Using the corpus, they have built a sentiment classifier, which determines positive, negative and neutral sentiments for a document. Experimental evaluations showed that the proposed technique is efficient and performed better than previously proposed methods.

Varsha Sahayak et al [4] proposed the approach automatically classified the sentiments of Tweets taken from Twitter dataset. These messages or tweets are classified as positive, negative or neutral with respect to a query term. This has been useful for the companies who want to know the feedback about their product brands or the customers who had to search the opinion from others about product before purchase. They used machine learning algorithms for classifying the sentiment of Twitter messages using distant supervision. The training data consisted of Twitter messages with emoticons, acronyms which are used as noisy labels. They examined sentiment analysis on Twitter data. The authors used Parts Of Speech (POS)-specific prior polarity features and used a tree kernel to prevent the need for monotonous feature engineering.

Rajni Singh and RajdeepKaur [5] analyzed Social data such as Twitter Tweets using sentiment analysis which checks the attitude of User review on movies. They developed a combined dictionary based on social media

keywords and online review and also found hidden relationship pattern from this keyword. HarshitaRajwani et al [6] presented a system to dynamically analyze traffic and its causes, using twitter stream analysis. Twitter is a social networking site which allows people to share and read tweets. The system fetches the tweets from twitter; applies natural language processing technique on them; categorizes the tweets related to traffic; notifies the registered users about it. Natural language processing (NLP)focuses on developing efficient algorithms to process text and convert it into machine understandable language. Here, we apply NLP onthe tweets to detect the traffic.

G.Vinodhini and RM.Chandrasekaran [7] presented a survey covering the techniques and methods in sentiment analysis and challenges appear in the field. There are different problems predominating in this research community, namely, sentiment classification, feature based classification and handling negations. Hence accurate method for predicting sentiments could enable us, to extract opinions from the internet and predict online customer's preferences, which could prove valuable for economic or marketing research. Safa Ben Hamouda and Jalel Akaichi [8] explored the potential applications of text and sentiment mining techniques on statuses update in order to analyze the Tunisian's behavior during the revolution. They chose a random population having Facebook accounts. It includes males and females, students, workers, housewives, etc. The age of targeted population is varying between 21 and 54 years old. Through the application of machine learning algorithms, they aimed to identify the nature the statuses update, and to link them to behaviors and sentiments characteristics. For that purpose, they created our own dataset and then we applied on it two machine learning algorithms: Naïve Bayes and Support Vector Machine. The expected output is to classify the extracted statuses into semantic classes useful, not only for people that aim to know themselves, but also for political decision makers. Geetika Vashisht and Sangharsh Thakur [9] introduced a method to perform a sentiment analysis on text-based status updates &comments, disregarding all verbal information and using only emoticons to detect both positive and negative sentiments. They identified the most commonly and frequently used emoticons & classified them on the basis of the sentiment they strengthen which eventually decides the polarity of the sentence. Farman Ali et al [10] proposed a fuzzy ontology-based sentiment analysis and semantic web rule language (SWRL) rule-based decision-making to monitor transportation activities (accidents, vehicles, street conditions, traffic volume, etc.) and to make a city- feature polarity map for travelers. Their system retrieved reviews and tweets related to city features and transportation activities. The feature opinions were extracted from these retrieved data, and then fuzzy ontology is used to determine the transportation and city-feature polarity. A fuzzy ontology based intelligent system prototype was used. Their experimental results showed satisfactory improvement in tweet and review analysis and opinion mining.

V. RESULT & DISCUSSION

The stated methodology is implemented using Rapidminer tool. In our paper we used AYLIEN text analysis extension to analyze the sentiments from the text. The 300 tweets are filtered according to the location and given to the sentiment analysis operator. The connection for the Twitter API and Aylien API should be established to perform the analysis. The outcome of the operator for each category of data is given in the Table I, II and III. The Fig. 2 shows the polarity distribution of the sentiment of traffic data by Delhi people. Similarly the Fig. 3 shows the polarity distribution of Bangalore traffic data and Fig. 4 shows the polarity distribution of Mumbai traffic data.

Polarity Of Traffic Data For The Location Delhi

Nominal Value	Absolute Count	Percentage
Neutral	82	93%
Negative	4	5%
Positive	2	2%

ISSN 2348 – 8034
Impact Factor- 4.022

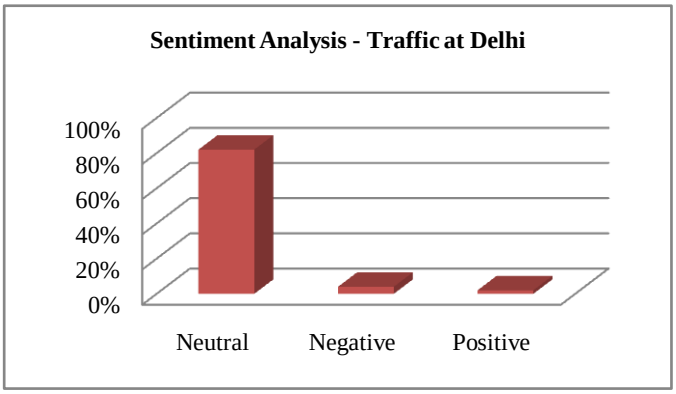

Fig. 2. Sentiment Analysis of the Traffic Data for the location Delhi

Polarity Of Traffic Data For The Location Bangalore

Nominal Value	Absolute Count	Percentage
Neutral	55	55%
Negative	45	45%
Positive	0	0%

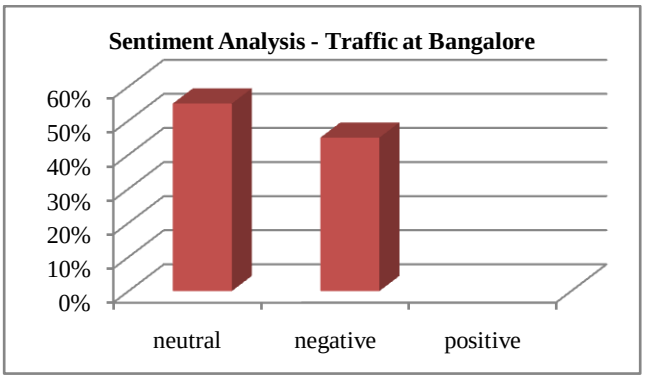

Fig. 3. Sentiment Analysis of the Traffic Data for the location Bangalore

Polarity Of Traffic Data For The Location Mumbai

Nominal Value	Absolute Count	Percentage
Neutral	81	81%
Negative	12	12%
Positive	7	7%

ISSN 2348 – 8034
Impact Factor- 4.022

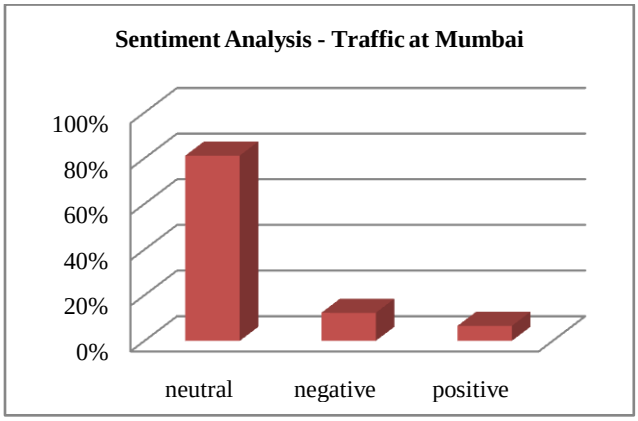

Fig. 4. ***Sentiment Analysis of the Traffic Data for the location Mumbai***

VI. CONCLUSION

In this paper, an approach to monitor traffic effectively using social media data is presented. The method is applied to real-time user generated data in the form of tweets and posts to analyze the traffic conditions in Indian metropolitan cities. It effectively filters unwanted data and gives summarized results. The efficiency of the system will increase in the future as more data traffic related tweets are generated in these metropolitan cities.

REFERENCES

1. *Mengdi Li, Eugene Ch'ng, Alain Chong, Simon, "The New Eye Of Smart City: Novel Citizen Sentiment Analysis In Twitter", Fifth International Conference on Audio, Language and Image Processing, ICALIP 2016, Shanghai China, July 2016.*
2. *Chaudhari S B, Shaikh Kamran, Shaikh Musaib, Alefiya Naseem and Priyanka Kamble, "Data Mining of Social Media for Traffic Monitoring", International Journal for Scientific Research & Development, Vol. 3, Issue 08, 2015, pp. 543 - 545.*
3. *Alexander Pak, Patrick Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, pp. 1320-1326.*
4. *Varsha Sahayak, Vijaya Shete, Apashabi Pathan, "Sentiment Analysis on Twitter Data", International Journal of Innovative Research in Advanced Engineering (IJIRAE) Issue 1, Volume 2, January 2015, pp. 178-183.*
5. *Rajni Singh and Rajdeep Kaur, "Sentiment Analysis on Social Media and Online Review ", International Journal of Computer Applications (0975 – 8887),Vol 121 – No.20, July 2015, pp. 44-48.*
6. *Harshita Rajwani, Srushti Somvanshi, Anuja Upadhye, Rutuja Vaidya, Trupti Dange, "Dynamic Traffic Analyzer Using Twitter", International Journal of Science and Research (IJSR) , 2014, pp. 984-987.*
7. *G.Vinodhini and RM.Chandrasekaran, "Sentiment Analysis and Opinion Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 6, June 2012, pp. 283-292.*
8. *Safa Ben Hamouda & Jalel Akaichi, "Social Networks' Text Mining for Sentiment Classification: The case of Facebook' statuses updates in the Arabic Spring Era", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 2, Issue 5,470-478, May 2013.*
9. *Geetika Vashisht and Sangharsh Thakur, "Facebook as a Corpus for Emoticons-Based Sentiment Analysis", International Journal of Emerging Technology and Advanced Engineering, Volume 4, Issue 5, 2014, pp. 904-908.*
10. *Farman Ali, Daehan Kwak, SM Riazul Islam, Kye Hyun Kim, Kyung Sup Kwak, "Fuzzy Domain Ontology-based Opinion Mining for Transportation Network Monitoring and City Features Map", The Journal of The Korea Institute of Intelligent Transport Systems, Volume 15, Issue 1, 2016, pp.109-118.*

11. *Kouloumpis E, Wilson T, and Moore J, "Twitter sentiment analysis: The good the bad and the OMG!", In Proceeding of AAAI conference on weblogs and social media, 2011, pp. 538–541.*
12. *Cao J, Zeng K and Wang H, "Web-based traffic sentiment analysis: Methods and Applications", IEEE transactions on Intelligent Transportation systems, vol. 15, 2014, pp.844-853.*
13. *Cambria E, Schuller B, Xia Y, Havasi C, "New avenues in opinion mining and sentiment analysis", IEEE Intelligent Systems. 28 (2), 2013, pp. 15–21*