Volume 5, Issue 11, November 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Survey on Web based Traffic Sentiment Analysis

[1]Hemant Bankar, [2]Prof. Deepak Gupta
[1]M.E. Student, [2]Assistant Professor
[1, 2] Department of Computer Engineering, Siddhant College of Engineering, Sudumbare, Pune, Maharashtra, India

Abstract— ***Sentiment analysis or opinion mining is a machine learning approach in which classification of the human's sentiments, emotions, opinions etc in the form of positive, negative or neutral comments underlying the text. The social media is continuously growing technology which can add and using it tremendously. In this social media such as face book, twitter, online forum and other web, users continuously use it and give their response and feedback for any activities quickly. There is various application of sentiment analysis and many of researchers have research on these applications but there are no more studies on transportation system, for safety, efficient transportations. Hence to reduce the traffic related problems, in this paper we proposes the traffic sentiment analysis (TSA).This survey will try to focus on sentiment analysis approaches, related work for automated web data crawling, different levels of SA, subjectivity classification, some machine learning techniques on the basis of their usage and importance for the analysis, evaluation of Sentiment classifications and its recent advancements and the future research directions in the field of traffic Sentiment Analysis.***

Keywords— Sentimental analysis, web based data, machine learning, TSA

I. INTRODUCTION

Sentiment analysis is an information gathering task to attain user's feelings. By analyzing a large numbers of documents, these feelings can be expressed in positive or negative ways in the form of comments, questions and requests. The Web is a huge depository of ordered and amorphous data. [1].Traffic injuries and fatalities are an enormous public health problem. To reduce traffic related injuries and fatalities, it would be helpful to monitor traffic in real time in order to quickly identify regions and activities that have the potential to become a risk to public safety. Hence traffic sentiment analysis is proposed by this paper.

Sentiment analysis concerns itself about the issues of traffic in particular transportation systems, and fairly enough traffic sentiment analysis can be considered as a subset of it. It is clearly impractical to find every corner information of the transportation network, but because of social media in all forms, including blogs, online forums, face book, and twitter, it should be possible to know the updated information traffic using users opinion's [1]-[3]. This would enable us to collect timely and comprehensive information about the current status of the transportation network and traffic flow to support advanced safety enhancement. To develop a system to automatically retrieve tweets or comment or positive/negative opinion related to transportation safety, extract the potential safety topics (e.g., traffic accidents, road flooding), calculate public sentiments was the primary objective of this survey.

II. REALATED WORK

DIFFERENT LEVELS OF SENTIMENT ANALYSIS:

- **DOCUMENTLEVEL:**

In this document level, it recognizes whole polarity of the document means whether the polarity of the document is positive or negative. For example, for a product review, whether the review gives an overall positive or negative opinion about the product is determined by the system and is commonly known as document-level sentiment classification. Each document expresses opinions on a single entity are the assumption made by this level. Thus the document level sentiment classification has some advantages and some disadvantages. Advantage is that we can get an overall polarity of an entity from a document. Disadvantage is that it could not extract the different opinions about different features f an entity.

- **SENTENCE LEVEL:**

Whether each sentence expressed a positive, negative, or neutral opinion is decided by this level. Neutral usually does not give any opinion. This level of analysis contains the subjectivity classification, which distinguishes objective sentences from the subjective sentences. Objective sentence express factual information about an entity and subjective sentences that express emotions and opinions about an entity. Only single opinion contain in simple sentence hence sentiment analysis is easy for simple sentence level. But in complex sentence contains various opinions hence sentiment classification is not done [3] [4].

- **ENTITY AND ASPECT LEVEL:**

Because of some disadvantages of document level and the sentence level analyses do not recognize peoples thinking about an entity. So finer-grained analysis performed by aspect level analysis. Aspect level is also called as feature-based opinion mining. It does not consider structure of language like sentences, documents, paragraphs or phrases; instead it directly looks at the opinion. It only considers sentiment of the opinion whether it is positive or negative and a target of opinion. An opinion without its target being identified is of limited use. Realizing the importance of opinion targets also help understand the sentiment analysis problem better.[3][4]

CLASSIFICATION OF SENTIMENT ANALYSIS:

The sentiment classification is broadly divided into the machine learning approach, lexicon based approach [5]. The machine learning approach uses the algorithms and linguistic features. The lexicon-based approach is based on a sentiment lexicon, which is the collection of known sentiment terms. Maximum entropy (ME), Naive Bayes (NB), and support vector machines (SVM), Rule Based Approach have achieved great success in text categorization

- **NAIVE BAYES CLASSIFICATION :**

Naive Bayes method is one of the popular techniques for text classification. Many researchers has proved that it perform extremely well in practice. It is an approach to text classification that assigns the class c* = argmaxc P(c | d), to a given document d. It is a simple probabilistic model. Naive base classifier uses Bayes' theorem. Probability model of naive bayes can be called as "independent feature model". The Bayes' rule of the Naïve Bayes (NB) classifier is eq. (1),

$$P(c|d) = \frac{P(c)P(d|c)}{P(d)} \quad (1)$$

Where, P(d) plays no role in selecting c*. To estimate the term P(d|c), Naive Bayes decomposes it by assuming the fi ′s are conditionally independent given d's class as in eq.(2)

$$P_{NB}(c|d) = \frac{P(c)(\prod_{i=1}^{m} P(f_i|c)^{n_i(d)})}{P(d)} (2)$$

Where, m is the no of features and fiis the feature vector.

Text categorization using naive bayes perform very well although its conditional independence assumption clearly does not tend in real-world situations [9]. With highly dependent features Naive Bayes is optimal for certain problem classes [10].

- **SUPPORT VECTOR MACHINES:**

At traditional text classification Support vector machines (SVMs) have been very effective. As compared to Naive Bayes and Maximum Entropy support vector machine have large margin. In the two-category case, the basic idea behind the training procedure is to find a maximum margin hyperplane, represented by w , that vector separates the document in two classes, and also it finds which margin, is as large as possible. This corresponds to a constrained optimization problem; letting cj ∈ {1, −1} (corresponding to positive and negative) be the correct class of document dj, the solution can be written as in Eq.(3),

w =∑ j aj cjdJ , αj ≥ 0 (3)

Where, the αj 's (Lagrangian multipliers) are obtained by solving a dual optimization problem. Those dJ such that αj is greater than zero are called support vectors, since they are the only document vectors contributing to w. [6][13]

III. TRAFFIC SENTIMENT ANALYSIS

TSA is a subfield of sentiment analysis, which concerns about the issues of traffic in particular. Due to the field sensitivity of sentiment analysis, it is necessary to discuss the TSA system briefly with its working. For the completeness of ITS space, it is necessary to collect and analyze the public wisdom and opinions. With the remarkable advancement of Web 2.0 in the last decade, communication platforms, such as blogs, wikis, online forums, and social-networking groups, have become a rich data-mining source for the detection of public opinions. Therefore, we propose traffic sentiment analysis (TSA) for processing traffic information from websites. As taking consideration of human affection, TSA will enrich the performance of the current ITS space. The TSA treats the traffic problems in a new angle, and it supplements the capabilities of current ITS systems [7]

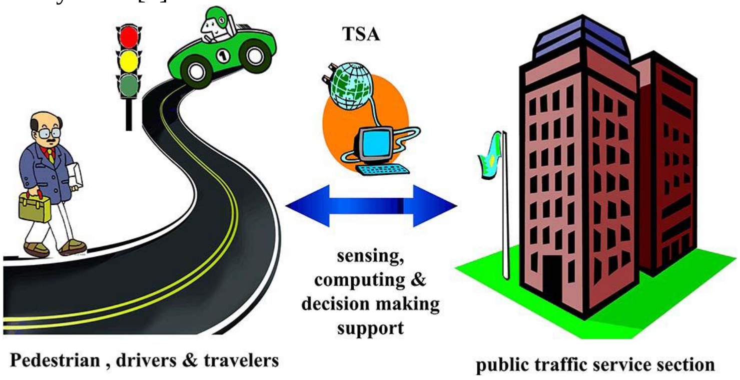

Fig. 1 TSA system

Fig. 1 illustrates the modules of ITS and exhibits that the TSA plays the role of sensing, computing, and supporting the decision making in ITSs.

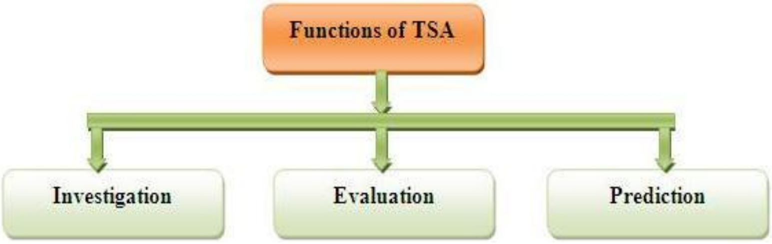

Fig.2 Functions of TSA system

The functions of the TSA system can be illustrated as follows. 1) Investigation: It is more economical and efficient than the public poll to collect the public opinion through the TSA system. 2) Evaluation: The computational production of the TSA system can be used to evaluate the performance of traffic services and policies. 3) Prediction: The TSA system can be further developed to predict the trends of some social events. [7] [15]

IV. COMPONENTS OF TRAFFIC SENTIMENT ANALYSIS

Fig.1 illustrated of TSA system. Following are the components of TSA system.

1) Web data collection- Gather data from several websites which gives the information related to the transportation system, such as face book, twitter, online forum for traffic information ensuring that the conclusions are definitely based on public opinion or, at least, represent part of the public opinion. From web like twitter collect all the data related to the traffic.

2) Preprocessing- In the preprocessing, the following steps are included: 1) the segmentation of text, 2) the labeling of words, and 3) the replacement of synonymous expressions.

Here to avoid unnecessary disturbances and improve precision, preprocessing should be conducted according to the material and the demand of the algorithms used word segmentation optimization techniques.

3) Extraction of subjects and objects- Subjects and objects are mainly extracted by context mining and document analysis. In TSA, appropriate models should be designed in context mining according to different data sets and resources. Context mining should obtain results as efficiently as possible to provide the necessary background knowledge for the subsequent steps. The second approach of extracting subjects and objects is text analysis, which is extracting the opinion-oriented information through the pure text.

4) Extraction of sentiment properties- The extraction of properties is based on the sentiment, modifier, and rule bases.

5) Sentiment calculation and classification- In this component calculate the sentiment of sentences and classify with its word.

6) Evaluation- The approach should be evaluated according to data sets.

V. ISSUES IN TRAFFIC SENTIMENT ANALYSIS

The primary problem of TSA is the selection of sentiment analysis approaches on Web-based data. Since the performance of both the rule- and learning-based approaches depends on the data to some extent, the features of web data should be identified first. The data we discussed in this paper are gathered from online forums, blogs, Twitter; the properties of these data can be described as follows.

1) The lengths of the texts vary. Some texts contain thousands of words, whereas others consist of only one sentence that can be as short as a one word.
2) The stylistic features of the texts are diverse. Given that Web 2.0 provides an equal platform for each user, words used for communication are not regulated, and texts are not under specific norms. In addition, different users express themselves in various ways, hence the different features of expressions on the Web.
3) New Internet expressions frequently emerge. With changes through time, the same sentiment may be expressed in different ways. In extreme cases, the same word may carry a different sentiment polarity after certain public events.[7]

VI. CONCLUSION

In this literature we have observed and underlined that sentiment analysis or opinion mining and its approaches, techniques. Concepts of text mining but also the concepts of information retrieval are encompassed within opinion mining. This paper gives their contribution to the real-world application. We have proposed Web-based TSA to recognize the traffic related problems in a humanizer way. Here TSA system which performing the processing of web data and gives predicted results. This paper presents the real-time web monitoring system for the detection of safety related patterns from web data. Mine the big unstructured data has become an important research problem. The task to implement the TSA system into existing ITSs is also critically important, and it does need further research. We suggested that take the policy evaluation part to support decision making of managers and view the evaluation results related to specific location as sensor information

ACKNOWLEDGMENT

I am very thankful to my guide Prof. Deepak Gupta, PG Coordinator Prof. Shayam Gupta and HOD Prof. Kumbharkar Sir to make me aware and guide me time to time. I hope in future the treat us in the same way they make us understand how to work and think.

REFERENCES

[1] Mukherjee S, Bhattacharyya P 2013 Sentiment Analysis: A Literature Survey, IIT Bombay, Mumbai
[2] Pang B, Lee L, Vaithyanathan S 2002 Thumbs up? Sentiment classification using machine learning techniques. Proc. ACL-02 Conf. on Empirical methods in natural language processing, 10: 79-86.
[3] Turney P D 2002 Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. Proc. 40th Annual Meeting on Association for Computational Linguistics (ACL), Philadelphia, 417-424.
[4] B. B. Khairullah Khan, Aurangzeb Khan, "Sentence based sentiment classification from online customer reviews," ACM,2010.
[5] Diana Maynard, Adam Funk," Automatic detection of political opinions in tweets."In: Proceedings of the 8th international conference on the semantic web, ESWC'11; p. 88-99, 2011.
[6] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Found. Trends Inf. Retrieval, vol. 2, no. 1/2, pp. 1–135, Jan. 2008.
[7] Jianping Cao, Ke Zeng, Hui Wang, "Web-Based Traffic Sentiment Analysis: Methods and Applications", IEEE Transactions on Intelligent Transportation Systems, Vol. 15, No. 2, April 2014.
[8] F. Y. Wang, "Social computing: Concepts, contents, and methods," Int. J. Intell. Control Syst., vol. 9, no. 2, pp. 91–96, 2004.
[9] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," Proceedings of the42nd Annual Meeting on Association for Computational Linguistics, ACL, 2004.
[10] Diana Maynard, Adam Funk," Automatic detection of political opinions in tweets."In: Proceedings of the 8th international conference on the semantic web, ESWC'11; p. 88-99, 2011.
[11] Mining Hu and Bing Liu, "Mining and Summarizing Customer Reviews", Proceedings of the tenth ACM SIGKDD International conference on knowledge discovery in data mining (KDD-2004), August 22-25.
[12] Read J, carol J.," Weakly supervised techniques for domain independent sentiment classification", In: Proceeding of the 1st international CIKM workshop on topic-sentiment analysis for mass opinion; P.45-52, 2009.
[13] P. D. Turney, "Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews," in Proc. 40th Annu. Meet. Assoc. Comput. Linguist., 2002, pp. 417–424.
[14] Kang Hanhoon, YooSeongJoon, Han Donglil,"Senti-lexicon and improved Naive Bayes algorithms for sentiment analysis of restaurant reviews" Expert SystAppl ,39:6000-10,2012.
[15] Chin Chen Chien, Tseng You-De,"Quality evaluation of product reviews using an information quality framework",Decis Support Syst;50:755-68,2011.
[16] Li Yung-Ming, Li Tsung-Ying,"Deriving market intelligence from microblogs",Decis Support Syst,2013. 13. Nigam K., Lafferty J., and McCallum A, (1990), "using maximum entropy for Text Classification". In Proc of the IJCAI-99 Workshop on Machine Learning for Information Filtering.
[17] Berger A., A Brief Maximum Entropy Tutorial. 15. Gongde Guo, Hui Wang, David Bell, Yaxin Bi and Kieran Greer, "KNN Model-Based Approach in Classification", Proc. ODBASE pp- 986 – 996, 2003.
[18] Ms Kranti Ghag and Dr. Ketan Shah,Comparative Analysis of the Techniques for Sentiment Analysis, ICATE 2013 Paper Identification Number-124

© 2015, IJARCSSE All Rights Reserved