Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 5, Issue. 12, December 2016, pg.01 – 06

Rule Based Approach for Sentiment Analysis of Traffic Data

Hemant M. Bankar[1], Prof. Deepak Gupta[2]

[1, 2] Department of Computer Engineering,
Siddhant College of Engineering, Sudumbare, Pune
[1] hemantbankar@gmail.com; [2] deepak_gpt@yahoo.com

Abstract— Now a day's rapid growth in internet usage of people's. So every person sharing his or her opinion about any interested topic in social network sites like facebook, twitter, blogs and web forums. It contains contain huge amount of unused data shared by the people. Sentiment analysis is process for extracting information from user's opinions shared on social sites. However there is rarely focus on the traffic related problems like (injuries, fatalities, heavy area traffic and working zone) which shared by peoples on social sites. The focus of this paper is to collect traffic related information's shared by all the people's on social networking sites, wiki's, online forms, news, announcements, etc. Also implement Traffic Sentiment Analysis (TSA) system using the rule based approach. Specifically, the rule based works the sentiment pattern matching with our proposed effective rules capable of extracting traffic related opinions. It finds sentiment polarity after processing and give result in format of positive, negative and neutral form.

Keywords— Sentiment analysis, TSA, rule based, sentiment pattern.

I. INTRODUCTION

Sentiment analysis is also termed as Opinion mining. It is a computational exploration of people's opinion, attitude and feelings toward an entity or things [1]. Sentiment analysis is identify the sentiment expressed in terms of in text then analyses it. Sentiment analysis finds the opinions also identify the sentiment which is expressed and find out the polarity in terms of positive, negative or neutral [2]. For the most part, sentiment analysis finds the tendency of people's about any subject or the general feeling content. Because of the exponential upgrade in the internet utilization and submission of opinion for an entity on the internet, sentiment analysis turns into a vital procedure in today's life. Sentiment analysis can be classified with three levels as per calculating sentiment as document level, sentence level and aspect or entity level.

In the document level is whole document gives the sentiment as positive, negative or neutral as information in determined area. For instance, given an item review, the system figures out if the audit communicates a general positive, negative or neutral assessment about the item [4][6]. In the sentence level goes to the sentences and figures out if every sentence communicated a positive, negative, or unbiased assessment. Both the document level and the sentence level examinations don't find what precisely individuals preferred and did not like. For example, document-level opinion mining identifies the overall subjectivity or sentiment expressed on an entity in a review document, but it does not associate opinions with specific aspects of the entity [10]. In sentiment analysis indicates an entity or an attribute of an entity on which users express their opinions. In this level the identification of such aspect level opinion and find sentiment of it [9] [10].

II. RELATED WORK

M. Hu and B. Liu [15] developed an approach called class sequential rules to extracting option features from product reviews based on semantic patterns. In this work, M. Hu and B. Liu proposed a more principled mining method based on sequential pattern mining called Class Sequential Rules (CSR). As its name indicates, the sequence of words is considered automatically in the mining process. Here, mine sequential rules can be applied which can be mined from a set of labelled training sequences of words and part-of-speech tags. In learning-based sentiment classification requires sufficiently large training data sets with positive and negative examples manually labelled, which are often very costly and time consuming [9].

L. Ku's et al. designed an approach to opinion summarization summarizes opinions of articles by telling sentiment polarities, degree and the correlated events in news and blog corpora. Ku's designed algorithms for Opinion extraction mines opinions at word, sentence and document levels from articles. As per algorithms find out the sentiment words in sentences which is preceded by negation operator then reverse sentiment polarity. [14]

As above existing work of Ku's algorithm with sentiment analysis, here sentiment analysis of traffic web based data using rule based approach. In our system providing set of rules for calculating polarity which will be given accuracy as above said algorithm.

III.SYSTEM ARCHITECTURE

Sentiment analysis of web based traffic data system includes three phases of system.

1) Traffic Data Sets 2) Pre-processing of Data 3) Calculation and Evaluation.

Firstly, collect the all the public opinion regarding traffic from online forums, News, Announcement's, traffic websites and social networking (facebook, twitter and etc.) sites. The data should contains all traffic related data extracted from websites and forums.

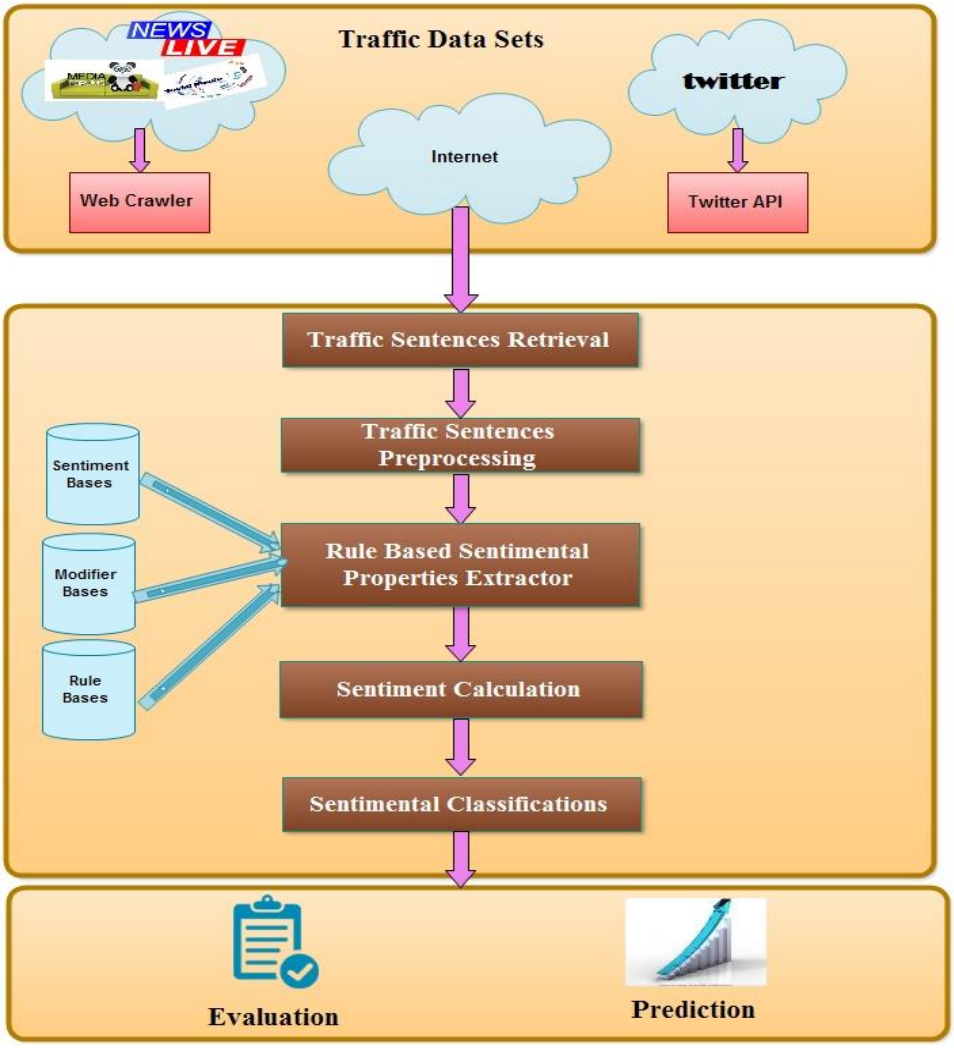

Fig. 1 Architecture of TSA system

Secondly, by computational production of the TSA system can be used to perform pre-processing of collected data, stemming and predicts the polarity of a sentence or document by analysing the occurring patterns of such words in text by applying rules on text. It's calculate polarity on basis of rule applied and calculate weight of word in sentences.

Finally the TSA system can be further developed to predict the polarity of particular data and evaluate the result in the graphical format as per user's opinion.

Rule Based Approach:

Here the extracted text from sites are with stemming and pre-processing, POS tagging concept applied. Here in sentences identify S (subject word), D (degree word) and N(negation word).

For sentiment analysis of traffic firstly required the words from traffic domain. To solve this, our main task is to maintain the sentiment lexicon of traffic. Also required the to maintain traffic word's sentiment polarity. In this source differentiate the positive and negative words with different sets.

Sr. No.	Rules	Example	Formula
1	S + D	good | convenient | more | great	$P = P_s + P_d$
2	D + S	Forever | support | absolute | safe	
3	N + S	not | safe not | beautiful	$P = - P_s$
4	N + D +S	not | very | safe not | very | dangerous	$P = -(1/3)* P_s * P_d$
5	D + N + S	very | safe	$P = - P_d * P_s$

Table 1. Defined Rules

For rule based approach the construction of rule very essential part because the sentiment of traffic sentiment is depends on the location and comparison of sentiment words [19]. The semantic rule of sentiment is based on the pattern of sentiment word S, negation/modifiers N and degree words D. The SND pattern is checked with extracted text and calculates its polarity of sentences. Modifiers are placed near to sentiment word. Here in above Table 1 defined rules which applied on the sentences.
Here calculating the polarity of each sentences which matched with our defined rules. Sentiment analysis also checks the weight of special traffic sentiment words those are used in these sentences. Here TSA sentiment words phase contains the dictionary in which all the words including traffic also have assigned value. When calculate sentiment the words are checked into this dictionary.

Algorithm for finding polarity of sentences and document using Rule Based.

Word, sentence, and document levels are the levels for finding sentiment. Firstly we decompose a document into sentences and determine the sentiment polarity of each sentence. The polarity scores of all the sentences are combined to compute for the overall polarity of the entire document. [13]

Steps:

1. A set of texts as input collected from various social sites and web as a text ‘t’
2. Do pre-processing on texts t.
3. Perform checking whether t is on document or sentence level.
4. If t is sentence level perform,
    a. Perform SND extraction operation on it
       //Apply rule for pattern checking
    b. Calculate polarity P_t of texts t.
5. End
6. Else if it found that t is document level then

7. Apply fragmentation of t and in to number of sentences{S_1….S_n}
8. Again For each sentence S_i
9. Do 4 to 7 and calculate polarity of each sentence, P_{si}
10. Do the extraction of sentences features
11. For each sentence S_i, calculating weighting W_i,
12. End for
13. For calculating polarity of whole document
14. $P_t = \sum P_{si}W_i$
15. End if
16. Do the sentiment classification on P_t
17. Output will be generated as Sentiment polarity of text t

The P_s is sentiment polarity of a sentence. P_s is calculated according matching pattern of SND.

IV.RESULT

By using the rule based approach algorithm in which matching the pattern of SND words with extracted texts. Our proposed system processes on traffic data set and calculate polarity of each sentences for traffic data. Here as per proposed approach, the expected result will be compared with Ku's algorithm. Result is compared on the basis of three parameters. Accuracy, Recall and Precision.[7] [8]

$$\text{Precision(P)} = \frac{tp}{tp+fp};$$

$$\text{Recall(R)} = \frac{tp}{tp+fn};$$

$$\text{Accuracy(A)} = \frac{tp+tn}{tp+tn+fp+fn};$$

Where,

tp - Document being classified correctly as relating to a topic.

fp - Document that is said to be related to the topic incorrectly.

tn - Documents that should not be marked as being in a particular topic and are not.

fn - Document that is not marked as related to a topic but should be.

Accuracy is the portion of all true predicted instances against all predicted instances. An accuracy of 100% means that the predicted instances are exactly the same as the actual instances. Precision is the portion of true positive predicted instances against all positive predicted instances. Recall is the portion of true positive predicted instances against all actual positive instances. Result will be increased by applying rule on the data. So it will give more accurate result as compared to Ku's algorithm.

V. CONCLUSIONS

In this paper we have observed and underlined that sentiment analysis or opinion mining plays a vital role in decision making. 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. This paper presents the real-time web monitoring system which collect the live for the detection of safety related patterns from web data. Also applied rules on web collected data simply give correct result as per users opinions.

VI. FUTURE WORK

Here we are developing new rules which are applied on the web traffic data, gives correct accurate result. Also as per location we are updating user result and by using map provide new way for travelling has less traffic problems.

ACKNOWLEDGEMENT

Our heartfelt thank goes to Siddhant College of Engineering, Sudumbare, Pune for providing strong platform to develop our skills and capabilities. I would like to thank to HOD, my guide and respective teachers for their constant and valuable support and motivation. Last but not least, I would like to thank all who directly or indirectly helped me in processing the paper.

REFERENCES

[1] Walla Medhat, Ahamad Hassan, Hoda Korashy, "Sentiment Analysis Algorithm and Applications: A Survey, 2014" Production and hosting by Elsevier B.V. on behalf of Ain Shams University.

[2] Abdullah Dar, Anurag Jain Computer Science & RGPV India "Survey paper on Sentiment Analysis: In General Term" International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-3, Issue-11)

[3] Zhen Hai, Kuiyu Chang, Jung-Jae Kim, and Christopher C. Yang "Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 3, MARCH 2014

[4] Richa Sharma, Shweta Nigam and Rekha Jain "OPINION MINING OF MOVIE REVIEWS AT DOCUMENT LEVEL" , International Journal on Information Theory (IJIT), Vol.3, No.3, July 2014

[5] 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.

[6] Peter D. Turney "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp. 417-424

[7] Q. Ye, W. Shi, and Y. Li, "Sentiment classification for movie reviews in Chinese by improved semantic oriented approach," in Proc. 39th Annu. HICSS, 2006, pp. 1–5.

[8] Pang B, Lee L 2008 Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2):1-135. A. Lindgren, A. Doria, and O. Schelen. Probabilistic routing in ntermittently connected networks. SIGMOBILE Mobile Comput. Commun. Rev., 7(3), 2003.

[9] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in Natural Language Processing and Text Mining. New York, NY, USA: Springer-Verlag, 2007, pp. 9–28.

[10] Peiquan Jin, Yongo Yu, Jie Zhao, Lihua Yeq, " Extracting Appraisal Expressions from Short Text", @ Springer International Publishing Switzerland 2015

[11] Zhen Hai, Kuiyu Chang, Jung-Jae Kim, and Christopher C. Yang "Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 3, MARCH 2014

[12] J. Wiebe, T. Wilson, R. Bruce, M. Bell, and M. Martin, "Learning subjective language," Comput. Linguist., vol. 30, no. 3, pp. 277–308, Sep. 2004.

[13] R. W. Yuen, T. Y. Chan, T. B. Lai, O. Kwong, and B. K. T'sou, "Morpheme-based derivation of bipolar semantic orientation of Chinese words," in Proc. 20th Int. Conf. Comput. Linguist., 2004, pp. 1008–1014.

[14] J. Kamps, M. Marx, R. J. Mokken, and M. De Rijke, "Using WordNet to measure semantic orientations of adjectives," in Proc. Int. Conf. Lang. Resourc. Eval., 2004, pp. 1115–1118.

[15] M. Hu and B. Liu, "Opinion feature extraction using class sequential rules," presented at the AAAI Spring Symposium Computational Approaches Analyzing Weblogs, Palo Alto, CA, USA, 2006, Paper AAAI-CAAW-06.

[16] Y. Guo and Y. Zhou, "Chinese text orientation analysis based on phrase," in Proc. Int. Conf. NLP-KE, 2009, pp. 1–6.

[17] Cambria E, Schuller B, Xia Y, Havasi C 2013 New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, 28(2):15-21.

[18] J. Wiebe, T. Wilson, and M. Bell, "Identifying collocations for recognizing opinions," in Proc. ACL Workshop Colloc., Comput. Extract., Anal., Exp., 2001, pp. 24–31

[19] Gebremeskel G 2011 Sentiment Analysis of Twitter posts about news. Master's Thesis, University of Malta

[20] Mark Dredze, Michael J. Paul, Shane Bergsma, Hieu Tran, "Carmen: A Twitter Geolocation System with Applications to Public Health," AAAI Workshop on Expanding the Boundaries of Health Informatics Using AI (HIAI), Bellevue, WA, 2013

