International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 3 (2015) pp. 8287-8296

http://www.ripublication.com

Traffic Analysis Based On Web User Behavior

Boopalan [1] and Rajesh [2]

[1] Research Scholar,Department of Computer Science and Engineering, Bharath University,Selaiyur,Chennai, TamilNadu
{Boopalan, k_boopalan}@rediffmail.com
[2]Department of Computer Science and Engineering,C.Abdul Hakeem College of Engineering and Technology,Melvisharam,Vellore,Tamil Nadu
{Rajesh,amrajesh73}@gmail.com

Abstract

Sentiment analysis has developed rapidly in recent years as the internet usages have reached successful growth among people. Only few studies have focused on the field of transportation even that failed to meet strict requirements of safety, efficiency and information exchange of ITS (Intelligent Transportation System). To overcome the problem, TSA (Traffic Sentiment Analysis) system is used. The TSA treats the traffic problems in newer angle. A technique called keyword extraction is proposed in this paper. Our work will be helpful in the development of TSA.

Index Terms— Sentiment analysis, Social-networking, Web-based.

1 INTRODUCTION

Transportation systems plays a major role among people but the ITS system failed to get the opinions from the public. To complete the space, wisdom and opinions of the public are collected from blogs, wikis, online forums and social-networking groups and are analyzed.

Fig.1. Role of sensing, computing and supporting of ITS.

TSA system (Traffic Sentiment Analysis) is used for processing the information regarding traffic from the websites. TSA tackles the traffic related problems in a different view and makes the current ITS system more efficient. TSA system has three functions and that can be illustrated as follows. 1) *Investigation:* Public opinions are gathered from websites through the TSA system. 2) *Evaluation:* The traffic related information are extracted using keywords. 3) *Prediction:* Sentiment analysis is done.

TSA system can be further developed to predict some social events. For example: cancellation of the flight can be predicted by analyzing the expressions of the passengers from their words posted in facebook and twitter. TSA system will help the ITS system to process more efficiently as the public opinions are collected through the TSA system. The TSA system is independent of the current system, which is particularly useful in an emergency when other devices such as sensors or cameras were damaged.

The paper is arranged as follows: Section II reviews sentiment analysis. Section III discusses the issues in TSA. Section IV describes the architecture and process of TSA. Section V describes the ITS system. Section VI describes the methods of data collection. Finally, Section VII concludes the summary.

2 RELATED RESEARCH

Sentiment analysis is extraction of the negative informations that must be considered to solve the traffic problems. The sentiment polarity contained in the information (such as happy or sad, agree or disagre) are recognized to get the affective information. Once the traffic related informations are collected through TSA, the affective informations are mined to tackle the problem of traffic.

In paper [3] the authors describes a method for identifying an opinion with its holder and topic, given a sentence from online news media texts. An approach is introduced of exploiting the semantic structure of a sentence, anchored to an opinion bearing verb or adjective. This method uses semantic role labeling as an intermediate step to label an opinion holder and topic using data from FrameNet.

In paper [4] the authors explained that the web has become an excellent source for gathering consumer opinions. There are now numerous Web sites containing such opinions, e.g., customer reviews of products, forums, discussion groups, and blogs. It focuses on online customer reviews of products.

In paper [5] the authors describes an important part of their information-gathering behavior has always been to find out what other people think. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise as people now can, and do, actively use information technologies to seek out and understand the opinions of others. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a first-class object.

In paper [7] the authors explained that identifying the sentiments is a challenging problem. A system was presented that automatically fids the people who hold opinions about the given topic and the sentiment of each opinion. The system contains a module for determining word sentiment and another for combining sentiments within a sentence.

3 ISSUES IN TSA

The selection of the sentiment analysis approaches on web-based data is the primary problem of TSA. In this paper, the data were gathered from social-networking groups; these data has the following properties.

1) Variation in the length of the text. Some may post the information containing about hundreds of words and some may post one sentence containing few words only.
2) Words may not be complete. Some may use shorter form of words.

For example: ‘d’ to pronounce a word ‘the’.

3) Using same letter for different meanings.

For example: ‘s’ may be used to mean ‘yes’ and also can be used to mean ‘is’.

In this paper, keyword extraction is a technique used for mining affective traffic information that are gathered from websites, forums, social-networking media. Affective information are mined because, thee are the data that are needed to be processed in an ITS system to solve the traffic problems more efficiently.

4 ARCHITECTURE AND PROCESS

4.1 TSA ARCHITECTURE

The web has become an excellent source for gathering consumer opinions. There are now numerous websites containing such opinions. For example: customer reviews of products, discussion groups and blogs. An important part of our information gathering behaviour has always been to find out what other people think. Traffic information is

evaluated by senetence, text or document and the sentiment analysis is done. TSA sysem have been used to collect the information of traffic and mining only the affective information to be processed with.

TSA architecture consists of three funcions i.e. collecting messages from social networking sites, parsing and performing sentiment analysis. TSA sysem uses two components in the function of Investigate namely, News/Expert commentaries and Tweets/Facebook. These two components are used to extract the traffic information from the commentaries of experts on traffic and from the posts of public.

Parser is the component used in the function of Evaluate. Parser is a program, usually part of a compiler, that receives input in the form of sequential source program instructions, interactive online commands or some other defined interface and breaks them up into parts (for example: the nouns (objects), verb (methods), and their attributes or options) that then be managed by other programming. A parser may also check to see that all input has been provided that is necessary. The input sentence is of plain text format. Four activities are performed in this stage: Sentence segmentation, Tokenization, Removing stop word and Word stemming. Sentence segmentation is boundary detection and source text is separated into sentence. Tokenization is separating the input document into individual words. Stop words are the words which appear frequently in document but provide less meaning in identifying the important content of the document such as ‘a’, ‘an’, ‘the’, etc.. and those stop words are removed. The last step for preprocessing is Word Stemming; Word stemming is the process of removing prefixes and suffixes of each word.

Sentiment analysis is used in the function of Predict which is used to mine affective informations.

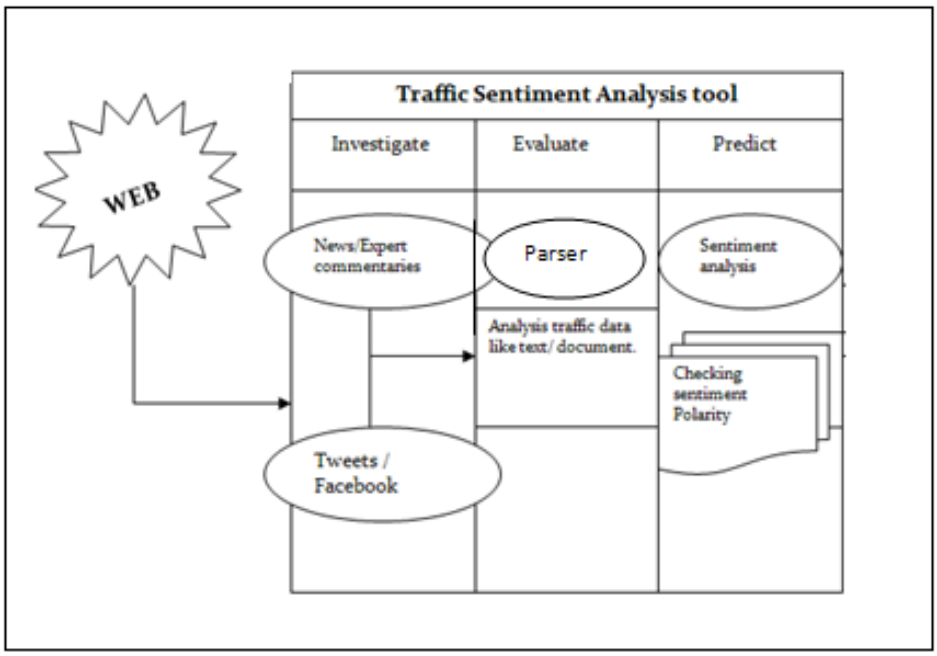

Fig.2. Architecture of TSA system

The sentiment polarity of the word assumed is determined by its morphemes. The word is positive, if it appears more frequently in positive lexicon; otherwise, the word is negative. Equations (1) and (2) are used to measure the positive and the negative tendencies of the morpheme q, positive and negative weights are assigned to the morphemes as follows:

$$wXci = (g\,xci / \sum_{i=1}^{n} g\,xci) \;/\; ((g\,xci / \sum_{i=1}^{n} g\,xci) + (g\,yci / \sum_{i=1}^{n} g\,yci)) \quad (1)$$

$$wYci = (g\,yci / \sum_{i=1}^{n} g\,yci) \;/\; ((g\,yci / \sum_{i=1}^{n} g\,yci) + (g\,xci / \sum_{i=1}^{n} g\,xci)) \quad (2)$$

$$Pci = wXci - wYci \quad (3)$$

The Pci in the formula (3) depends on morphemes Ci, and the absolute value of Pci is the degree of tendency of morphemes Ci. Sentiment polarity of words can be calculated by using the following steps: scan the positive and negative lexicon; P_w = 1, if the word *w* appears in the positive word lexicon. P_w = -1, if the word *w* appears in the negative word lexicon. The sentiment polarity is otherwise computed using morphemes by

$$P_w = 1/p \sum_{j=1}^{p} Pci \quad (4)$$

Where, P_w in equation (4) represents the sentiment polarity of the word *w*. If $P_w > 0$, the word is positive, otherwise the sentiment polarity of the word is negative.

4.2 TSA PROCESS

Communication platforms like blogs and social-networking groups have become a rich data-mining source for the detection of public opinions. Therefore, TSA have been proposed for processing those information. The situation from humanized perspective is detected as the traffic information from public opinion. The traffic informations are gathered from websites that are posted by public. Text can be categorized into three levels namely, Word, Sentence and Document levels. Public may express their opinions by words, by sentence and by using images and videos. Public opinions in images and videos are converted into text, sentence format and if necessary, it can be converted to document level.

Sentiment analysis is done on the text. Traffic related keywords are used to extract the negative information such as accident, ambulance, blocked. Traffic information can be considered only if it contains more than 15 letters. Affective informations are extracted using keywords and those affective information are sent to the ITS system (Intelligent Transportation System) to process more efficiently. Here, we ensure that conclusions are based on public opinion. Traffic Sentiment Analysis is done effectively.

4.2.1 Traffic Sentiment Analysis Algorithm:

The classification represents a supervised learning method as well as a statistical method for classification. Assumes an underlying probabilistic model and it allows us

to capture uncertainty about the model in a principled way by determining probabilities of the outcomes. It can solve diagnostic and predictive problems. In pre-processing, the unnecessary data like repeated messages, repeated letters, urls, emotion icons, WH questions, special symbols are removed.

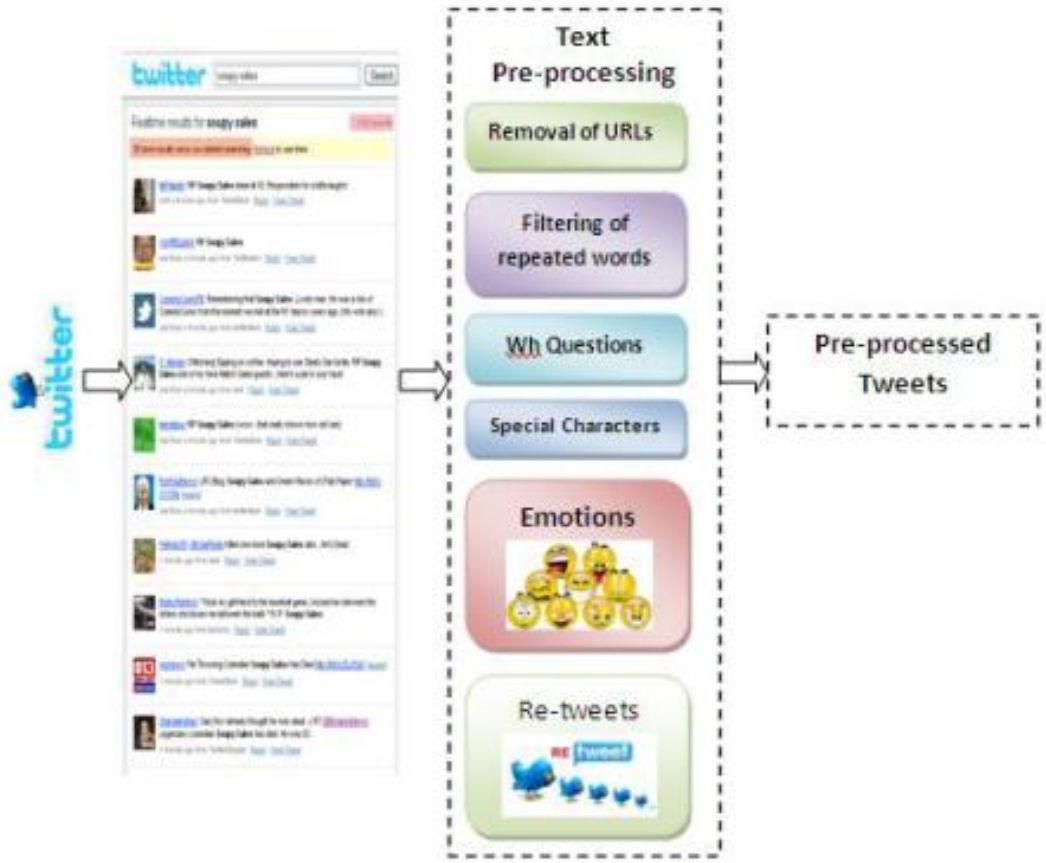

Fig.3. Overview of Data collection and Pre-processing process

Traffic Sentiment Analysis Algorithm:

Input: messages m=$\{ m_1, m_2, m_3 \ldots\ldots, m_n\}$,
Database: N_T
Output: Positive messages p= $\{ p_1, p_2, p_3 \ldots\ldots\}$
Negative messages n= $\{n_1, n_2, n_3 \ldots\ldots\}$

M= $\{ m_1, m_2, m_3 \ldots\ldots\}$
Step 1: Divide a message into words
$m_i = \{ w_1, w_2, w_3 \ldots\ldots\}$, i= 1, 2, ……, n
Step 2: if $w_i \in N_T$
Return p-polarity and n-polarity
Step 3: Calculate overall polarity of a

word= log (p-polarity) – log (n-polarity)

Step 4: Repeat step 2 until end of words

Step 5: Add the polarities of all the words of a message
(i.e) total polarity of a message

Step 6: Based on the polarity, message can be positive or negative
Negative information is considered to be processed with.

Step 7: Repeat step 1 until M € NULL

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. In practice, context mining includes conservation extraction and co-reference analysis. Conservation extraction refers to handling the text, such as "citation, @." In addition, co-reference analysis refers to mining the object represented by other words.

5 ITS SYSTEM

Existing ITS communication is between device / sensor only. If a device like sensor or camera is not working or malfunctioned, then the information will not be transferred to the Intelligent Transportation System. Public opinion is volatile in decision making and opinion mining which is absent. Sentiment analysis in traffic is not done. TSA plays the role of sensing, computing and supporting decision making in ITS.

The situation is detected from humanized perspective as we process traffic information from public opinion. Traffic information is evaluated by sentence, text or document and the sentiment analysis is done.

TSA output is transferred to ITS Admin, who has the authority to post the output in ITS system. Certificate Authority (CA) is responsible for transferring information to different vehicular devices. Non-vehiculr users can also receive informations through android mobiles. CA is also responsible for displaying traffic information in display unit. Therefore, TSA have been proposed to view the traffic problems in a new vision.

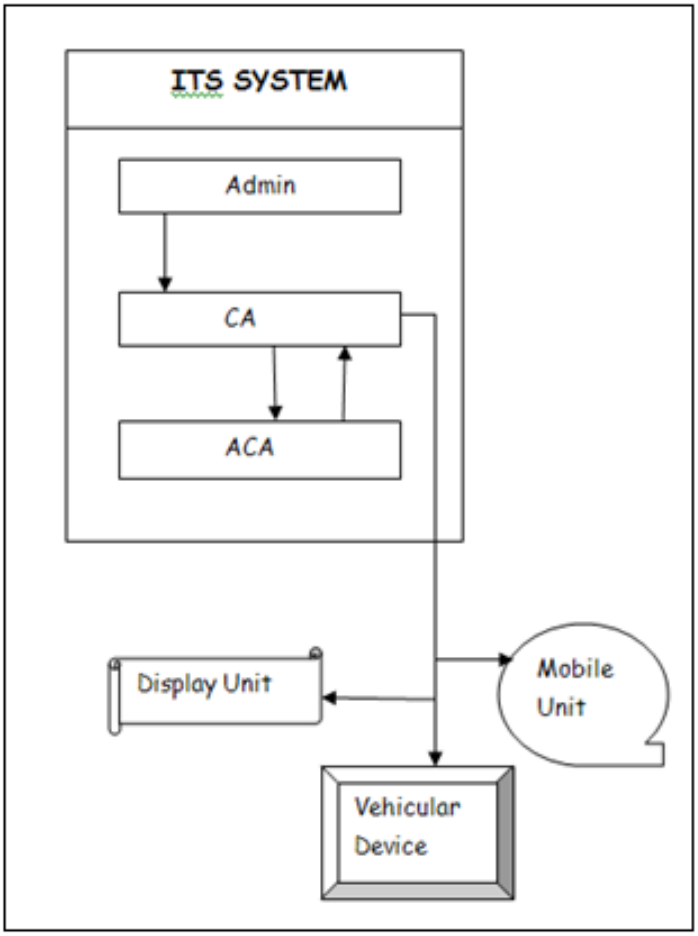

Fig.4. Architecture of ITS

6 DATA COLLECTION

Since the data gathered in TSA system are the public's opinion, they are collected from online through websites, blogs, forums and other socila-media networks. Collecting public opinion on particular topic is different. Traffic informations on the web can be classified into three categories.

1) The first category consists of news, expert commentaries, etc., from the traffic websites.
2) The second category includes the posts from the transport sector in forums. These forums provide a platform through which users can exchange information about social events and topics.
3) The third category inclues real-time informations about traffic through socila-networking groups.

The posts of first category is true and meaninful. The discussion in the second category is clear and is usually discussion on certain topics may be highly valuable for

tracking public opinion. The third category includes real-time traffic informations. Such informations bears significance for obtaining real-time information of travelers.

7 CONCLUSION

Since the sentiment analysis is done on traffic information, Web-based TSA have been proposed to solve the traffic problems in a humanizer way. The study of TSA will give us a newer vivion when facing with traffic problems.

The main content of the paper can be addressed by concluding our work as the following two folds: 1) designing the architecture of TSA; 2) using a technique to extract the affective information that are transferred to ITS system. TSA system have been proposed to get public opinions that are collected through social websites like Facebook, Twitter and News, experts commentary etc. Implementing TSA system into existing ITS is important and hence it needs further research. The keynote of implementation is jointly accommodating the traveler's best interest and reasonable workload. More techniques will be developed for the joint performance of ITS (Intelligent Transportation System) with the TSA (Traffic Sentiment Analysis) system in the future.

REFERENCES

[1] F. Y. Wang, "Social computing: Concepts, contents, and methods," *Int. J.Intell. Control Syst.*, vol. 9, no. 2, pp. 91–96, 2004.

[2] F.-Y. Wang, R. Lu, and D. Zeng, "Artificial intelligence in China," *IEEE Intell. Syst.*, vol. 23, no. 6, pp. 24–25, Nov./Dec. 2008.

[3] S.-M. Kim and E. Hovy, "Extracting opinions, opinion holders, and topics expressed in online news media text," in *Proc. Workshop Sentiment Subj.Text*, 2006, pp. 1–8.

[4] B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on theWeb," in *Proc. 14th Int. Conf. World Wide Web*, 2005, pp. 342–351.

[5] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retrieval*, vol. 2, no. 1/2, pp. 1–135, Jan. 2008.

[6] F.-Y. Wang, "Agent-based control for networked traffic management systems," *IEEE Intell. Syst.*, vol. 20, no. 5, pp. 92–96, Sep./Oct. 2005.

[7] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in *Proc. 20th Int. Conf. Comput. Linguist.*, 2004, pp. 1367–1373.

[8] V. Hatzivassiloglou and K. R. McKeown, "Predicting the semantic orientation of adjectives," in *Proc. 8th Conf. Eur. Chapter Assoc. Comput. Linguist.*, 1997, pp. 174–181.

[9] E. Riloff, S. Patwardhan, and J. Wiebe, "Feature subsumption for opin-ion analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2006, pp. 440–448.

[10] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques," in *Proc. ACL Conf. Empirical Methods Natural Lang. Process.*, 2002, vol. 10, pp. 79–86.

[11] E. Riloff, S. Patwardhan, and J. Wiebe, "Feature subsumption for opin-ion analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2006, pp. 440–448.

K.Boopalan is working as an Associate Professor in the Computer Science And Engineering Program of C. Abdul Hakeem College of Engineering & Technology, Tamilnadu, India & he is currently pursuing his Ph.D in Bharath University, Chennai. He received his M.E. Degree from Sathyabama University, Chennai, India, in April 2005 in the field of Computer Science and Engineering. and his B.E. Degree from Madras University, Tamilnadu, India in the field of Computer Science and Engineering.

A. Rajesh is a Professor & Head in the Computer Science & Engineering Program of C. Abdul Hakeem College of Engineering & Technology, Tamilnadu, India, and he received his Ph. D Degree from Dr. M. G. R. Educational and Research Institute University, Chennai, India, in March 2011 and his M.E. Degree from Sathyabama University, Chennai, India, in April 2005 in the field of Computer Science and Engineering. His area of interests includes Datamining, Natural Language Processing, Semantic Web and Intelligence Systems.