IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
Vol. 3, Issue 4, April 2016

A Novel Technique of Sentiment Analysis by Sandwich of Data Mining and Image Processing

Prince Patel[1], Aishwarya Parkhi[1], Poorva Hulyalkar[1], Ronak Parmar[1], Prof W. P. Rahane[1]

Department of Information Technology, NBN Sinhgad School of Engineering Ambegaon (BK), Pune, India[1]

Abstract: When a text is accessible in a digital format, it is a task in the field of text mining to accurately predict the sentiment of a text or the intention with which the text was written. In this paper, we are trying to develop a system that will analyse the intention with which a particular text is written. Sentiment analysis becomes a tedious job in the domain of text mining as the problem of polysemy (one word with different meanings) and synonymy (many words with the same meaning) arises frequently. Our intention is to take this process one step ahead by analysing and verifying the sentiment using image processing in order to identify the intention with which a sender is sending a particular text message. By processing the image, it will help to improve the accuracy of sentiment analysis as the facial expression is captured while the sender is sending the message. The facial expression will account for the mood of the sender at that instance which enhances the sentiment analysis. This system will give a report of all the activities of the sender and will eventually detect the mood of the sender just with the help of the messages that he has sent.

Keywords: Data mining, Image processing, Emotion, Sentiment analysis, Natural Language Processing (NLP), ADAboost, Region of Interest (ROI).

I. INTRODUCTION

With the advent of new technologies, communication has reached to a new level. It is just a matter of seconds to interact with people afar. With this ease of use of communication technologies, many people can communicate efficiently. Nowadays, almost everybody uses the digital form of communication to send their messages to the intended receiver. It is used for many purposes such as Business talks, chatting, advertising, entertainment and planning. Planning has two implications. It could either be constructive or destructive. If we make the computer understand each message being sent by the sender, then we can stop destructive planning to a great extent. Consider person A sends a message to person B. Person A writes "I'm going to kill person C " to person B. Since it's a destructive planning, the proposed system will escalate this message to the concerned authority so that further action can be taken. This can be achieved by data mining techniques. Data Mining is the process of discovering patterns that are present with various data sets. However, there could be two scenarios. Person A might be jesting or he might be serious about it. If he is serious, it is not a problem as it will be reported to the higher authority. But, if he is jesting, and the system reports to the concerned authority, then the person might face trouble without any bad intention. The importance of sentiment analysis increases with the addition of new social networks in the organizations.[1]

It is important to analyse the correct intention from the given text in order to minimize this type of errors. One way to do this is by detection of expression. Along with the text message, if we provide the facial expression of the sender, the accuracy will increase manifold.[10].If the sender sends the message and the expression captured is a happy one, the system will ignore the message. But, if the expression captured is angry or sad, the system will definitely report to the concerned authority.

II. RELATED WORK

Various text representations used for knowledge discovery from text. A prevalent one is the collection of words which uses keywords as components or vectors of the feature space. Here Rocchio classifieris used. Moreover, Global IDE and entropy weighing scheme is used as well.[10] Pattern mining is a key aspect in data mining.

A diverse collection of efficient algorithms such as Apriori-like algorithms PrefixSpam, FP-tree, SPADE, SLPMiner and GST have been introduced. The research mainly focuses on deriving the efficient pattern model from these proposed algorithms. Keyword spotting is a technique in which words are categorised into emotions.[9]Also the concept of close sequence mining has been introduced to improve efficiency of text mining. Visual perceptive enhances the overall efficiency.

Sentiment Analysis:

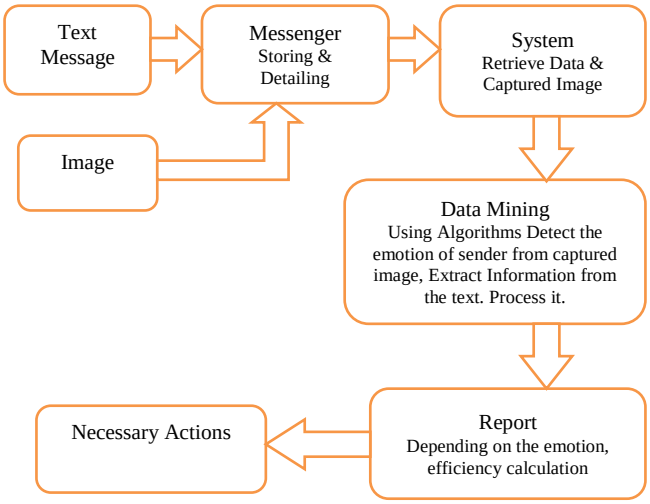
Analysis of varied sentiments is quite a tough task.[3] Researchers from Natural Language Processing(NLP) and information retrieval have developed different approaches to solve this problem, achieving promising results. Identifying emotion of person from social media applications is usually tough as it involves informal, short and vague data. Tumasjan et al in 2010 and Bollen et al in 2011 introduced pre-defined dictionaries for measuring the level of sentiment of the Tweets from twitter. [4]

Davidov et al in 2010[5] used the weak labelsfrom a enormous amount of twitter tweets. In contradiction, they manually selected hashtags with strong positive and negative sentiments and ASCII smileys are also utilized to label the sentiments of tweets. Furthermore, Hu et al.in 2013 incorporated social signals into their unsupervised sentiment analysis framework. Siersdorfer et al. (2010) proposed a machine learning algorithm to predict the sentiment of images using pixel level features. He was influenced by the fact that sentiment involves high-level

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Vol. 3, Issue 4, April 2016

abstraction, which may be easier to explain by objects in images.[6] Both Bort et al. 2013 and Yuan et al. 2013 proposed visual entities or attributes as features for visual sentiment analysis. Machine learning techniques such as Support vector machines(SVM) and naïve bayes classifier were also used to improve the accuracy of the sentiments.[7] .Gabor filters using ADAboost algorithm (to select the best features out of thousands of features)followed by SVM classifiers also giver better results.[8]

III. PROPOSED ARCHETECTURE.



Above diagram depicts the functioning of the proposed system. When the text message sent by the sender is loaded into the messenger, an image is captured of the sender. The text and image together are retrieved from the messenger by the system. Using tools such as OpenCV and Haar Cascade, the image is processed to detect the mood or emotion of the sender. Along with this, the sentiment is analysed from the text by text mining technique as well. The sentiment retrieved from the text and mood detected by image processing will be scrutinised and sent to the concerned authority. For example if sender sends a message "I will kill you" and the emotion detected by image processing is happy then the message will NOT be considered to be a threatening message. If the detected mood is serious then it will be reported to the concerned authority and further actions will be taken accordingly. At the end a report will be generated which will depict the statistics of the sender's one session activities and one can monitor the activities without the sender being aware that he is being watched.

Technical Aspects

OpenCV:

OpenCV represents Open Source Computer Visual Library. It comprises of machine learning libraries as well. It was brought to existence in order to integrate applications of computer vision and machine learning.

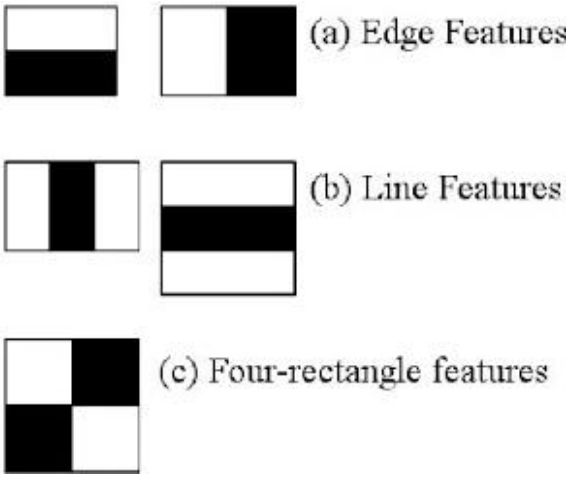
Haar Cascade:

In the proposed system HAAR cascade is used for detection purpose where it has pre-trained classifiers it is machine learning based feature which consists of cascade function. The function first undergoes learning from a large datasets of positive and negative images.

Then this training helps the function to detect the emotion of face. Once the classifier undergoes rigorous training the exact features from the images can be extracted. The exact features are determined by integration of sum of pixels lying in that particular area under consideration.

The feature can be extracted by 3 ways:-

1-Edge feature
2-Line Feature
3- Four-rectangle feature



In the proposed system we have used The Horizontal Edge feature for the extraction of eyes.

This is the depiction of edge feature. It is considered that the Eye portion is darker than the other portions of the face. The required steps for face creation and eye detection can be:-

1- Load XML classifier
2- Load input image in greyscale mode
3- Apply DetectMultiscale function
4- Write function to find location in form of rectangle (a,b,c,d)
5- Create Region of Interest (ROI) for face. (i.e features that are needed the most)
6- Then apply eye detection on the ROI.

In this way by considering a small single feature we can cover whole face part by part and detect the emotion by integrating these parts.

NLP:

Natural language Processing is an efficient technique of processing inputs that are human readable and extract meaning out of the text. It converts information from the given database to a language that can be read and summarized and given as output. NLP is principally used for mining of text wherein speech recognition also comes as a bonus. In this paper, we classify the sentiments as positive and negative using NLP.[2]

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Vol. 3, Issue 4, April 2016

Stemming:
The process of reducing each word (i.e., token) to its stem or root form, by removing its suffix. "Processing" and "processed" are forms of the same word "Process". Hence we get the root word from the existing word which makes text mining an easier task

IV. SYSTEM DISCRIPTION

Mathematical Model
Set Theory Analysis:
Let 'S' be the | Employee data mining based on text and stress analysis as the final set S

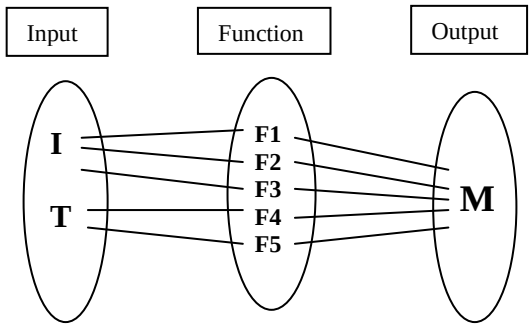
a. Identify the inputs as I,T
S = {I, T …}
I= {I1, I2, I3, I4 …| 'I' given as a image taken from camera.}
T= {T 1, T 2, T 3, T 4, …| 'T' given as text which are comment by the employee.}

b. Identify the outputs as O
S = {I, T, M…}
M= {M 1, M 2 …| 'M' is as mood detected like angry, sad and happy.}

c. Identify the functions as 'F'
S = {I, T ,M…
F = { F1(), F2(), F3(), F4(), F5(),F6(),F7(),F8()}

F1 (I) :: Process Requests on image taken by camera.
F2 (I) :: Respond as image matching with text.
F3 (T) :: Process Requests on given text taking from the employee comment
F4 (T) :: Respond as analyzing the text
F5 (T) :: Process Requests on mining the text.

Hence the functionality can be shown as,



V. FUTURE WORK

This system can be useful in companies wherein the employees can be kept a watch on by the managers and take actions if the employees are just whiling away their time or they are planning any evil activity.

VI. CONCLUSION

Many text mining techniques have been implemented till date. The most prevalent being, keyword spotting, tokenization, lexical affinity method and models like pattern taxonomy model. These techniques are not that efficient to identify the meaning of particular text. The proposed system, therefore takes this efficiency to another level by combining text mining along with emotion detection by analysing the facial expressions. The sender 's expression while sending the message will be captured and thus help the system to understand the intention of the sender in a more accurate and efficient form. Sentiments can thus be discovered in a better way.

REFERENCES

1. Liu B (2012) "Sentiment analysis and opinion mining." Synth Lect Hum Lang Technol 5(1):1–167
2. Nasukawa T Yi J (2003) "Sentiment analysis: capturing favorability using natural language processing."
3. O'onnor B, Balasubramanyan R, Routledge BR, Smith NA (2010)" From tweets to polls: linking text sentiment to public opinion time series." In: Proceedings of the international AAAI conference on weblogs and social media, pp 122–129
4. Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, Isabell M. Welpe "Predicting Elections with Twitter: What 140 Characters Revealabout Political Sentiment "
5. Dmitry Davidov, Oren Tsur, Ari Rappoport "Enhanced Sentiment Learning Using Twitter Hashtags and Smileys"(2010).
6. Stefan Siersdorfer,Enrico Minack, Fan Deng, Jonathon Hare."Analyzing and Predicting Sentiment of Images on the Social Web".
7.. Go, Lei Huang and Richa Bhayani , "Twitter Sentiment Analysis", Project Report, standford,2009
8. Bartlett MS, LittlewortG, Frank M, Lainscsek C, Fasel I, Movellan J (2005)" Recognizing facial expression: machine learning and application to spontaneous behavior." In: IEEE computer society conference on Computer vision and pattern recognition, CVPR 2005, vol 2. IEEE, pp 568–573
9. Shiv Naresh Shivhare1 and Prof. Saritha Khethawat, "EMOTION DETECTION FROM TEXT ".
10. Ning Zhong, Yuefeng Li, and Sheng-Tang Wu, "Effective Pattern Discoveryfor Text Mining."
11. Wan S, Aggarwal J (2014) Spontaneous facial expression recognition: a robust metric learning approach. Pattern Recognit. 47(5):1859–1868computer society conference on computer vision and pattern recognition, CVPR'91. IEEE, pp586–591
12. Andrea Esuli and Fabrizio Sebastiani, "Determining the semantic orientation of terms through gloss classification", Proceedings of 14th ACM International Conference on Information and Knowledge Management, pp. 617-624, Bremen, Germany, 2005.
13. Gamgarn Somprasertsri, Pattarachai Lalitrojwong , Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization, Journal of Universal Computer Science, vol. 16, no. 6 (2010), 938-955