

## Sentiment Analysis of Microblogging Messages for Detecting Public Safety Events

<sup>1</sup>Miss. Megha S. Masram, <sup>2</sup>Dr. Tausif Diwan

<sup>1,2</sup>Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, India

Email: masramms@rknec.edu, diwant@rknec.edu

Received: 09<sup>th</sup> July 2018, Accepted: 14<sup>th</sup> August 2018, Published: 31<sup>st</sup> August 2018

### Abstract

Microblogging sites have gained very importance nowadays. We can know sentiments of people and predict things from their sentiments. In this paper, we are detecting public safety events by analyzing the sentiments via Microblogging text messages. In this the ratings have been given i.e, positive, negative or neutral to the tweets, where there is a pre selection of topics causing riots and also some random tweets based on it. In this we will calculate the subjectivity and polarity confidence to capture the sentiments. Three classes of tweet are considered here that are positive, negative and neutral. There are predicated topics that have high chances of riots. For a particular event related with that topic we will analyze the tweets about that topic and also the tweets taken for analyzing purpose will be taken from the microblogging site users whose locations are same as that of the event. If the tweets negative comments are more than a particular threshold then the event will be said that it requires public safety.

**Keywords:** Microblogging, Classification, Sentiment

### Introduction

Microblog mining has become a very important research field. It can be used to know commercial and political view of the people. A sentiment analysis can be considered as a two type classification problem. The negative and positive classes are considered under it.

The vector space model is very much used for text representation [1], in this it takes a vector to represent a document. An important preprocessing step for many tasks in text mining, is to transform natural language message text to numerical vector which includes cluster analysis and sentiment classification. [3]

In the previous paper, the support vector machine has been taken as the classifier. It is because its excellence in solving two-class classification problem. In this, they included three dictionaries from Dalian University, self-established dictionary and CNKI-HowNet emotional. [2,4,5]

The analysis of the sentiments can help in various fields for reviews of the people to predict things from the future like events that need public safety. We can know that how the sentiments play an important role in different field like politics business and many more.

### Literature Review

A hashtag is a term that can be used by many users. Also, the retweet is used when you have to tweet some other's post in your timeline. This feature can be very important to prove that, if two tweets are connected by retweet or hashtag then that can be called as of similar type [1,6,7,10]. To map the emotional words in the text, they used the lexicon to judge. The lexicon examines whether the text contains the emotion indicative word or not. Considering an example, given a message "I am very excited today", the word excited occurs in the lexicon's excited category, and no word exists in the lexicon's other emotion categories (happy, angry and surprise). Thus, the feature is that: happy(0), excited(1), angry(0) and surprise(0) [12]. In previous work, to detect the real time intentions used the term 'usage indicative term' which means the word that show the activities associated with it. Also, the content indicative term indicates the main subject in the text. [14]. Real time public safety events are found based on the sentiment analysis, they created social sensors based on the storytelling approach from the tweets indicating safety issues [18].

### Proposed Work

We used bag-of-words model because it gives us a better understanding of the text and also the content of the text. The Aylien text api is used to analyze the sentiment influenced words. As soon as, the sentiment analysis part is over. By using Rapid Miner we extracted the tweets from the site named Twitter. Rapid miner helps in downloading the tweets from a particular location. For sentiment analysis, we took almost 14000 tweets of controversial topics in the Microblogging world.

The major work done, is we choose the riots causing topics in India, One of them is Padmavat. We consider event related to Padmavat in Gujrat which is about to happen. We will take tweets from this particular location and the tweets will be analyzed if the tweets negative comments raised up to a threshold, which we have set previously, then we will say that the event in Gujrat related to Padmavat will require public safety.

### Example

We chose 10 topics that are B\_R\_Ambedkar, Jignesh Mewani, Mayawati, NDTV, Padmavat, PNB\_Bank, Prakash Ambedkar, Rahul Gandhi, Trump, Shivaji Jayanti.

We randomly chose a location say Gujrat and the topic Padmavat. We downloaded the tweets based from Gujrat location. And then the sentiment analysis has been done on it. The Threshold set was 50. The total percent of negative tweets of padmavat went above the set threshold that means any event related to padmavat occurring in Gujrat will require public safety arrangements.

**Bag of Word (Vector Space Model)**

Bag of Word: The bag-of-words model is used in NLP and IR and is a representation of words. It is also known as the vector space model.

The following shows a text document by using BOW.

Given two text documents:

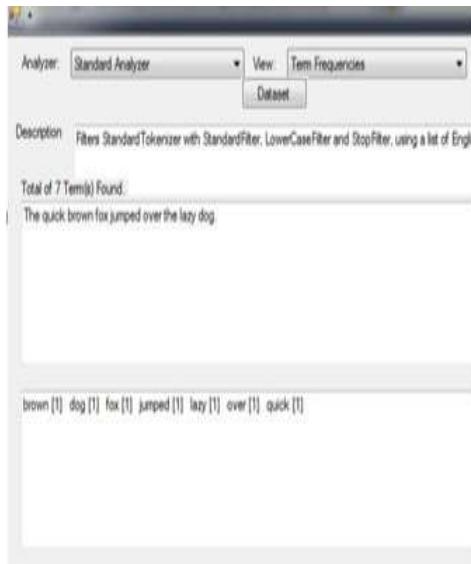
- (1) Jyoti likes to play games. Jaya likes games too.
- (2) Jyoti also likes to watch series.

The list constructed given below.

```
{
  "Jyoti",
  "likes",
  "to",
  "play",
  "games",

  "Jaya",
  "too",

  "also",
  "watch",
  "series"
}
```



**Fig.1 Data Preprocessing using Bag of Words Model**

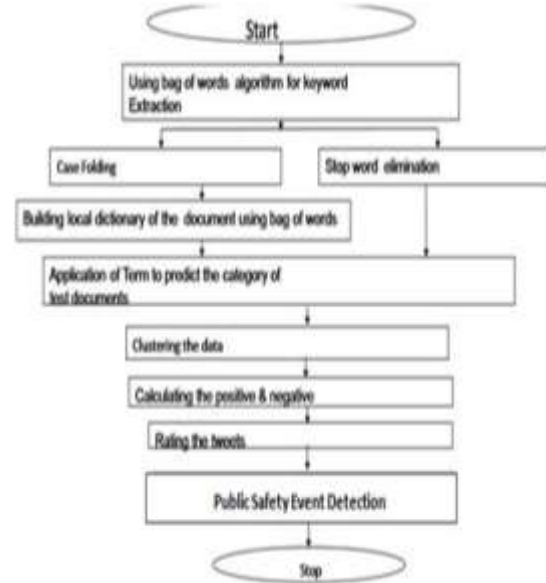
**Algorithm**

Consider negative tweets as priority If ( PN > T )

The event needs public safety arrangements If ( PN < T )  
 Event is safe  
 PN = The percentage of total number of tweets.

T= The threshold which we have taken for analysing

**Flowchart**



**Fig 2. Detecting Public Safety Events**

**Result and Discussion**

**Retrieval of offline tweets:**

To perform this task database containing tweets should be needed. So we have used a MySQL database containing 14000 tweets for offline or document classification. Whenever the tweets of a particular link or profile are retrieved, they all are imported to the database in our framework.

The Rapid Miner is used to extract tweets .It includes extraction of tweets based on various features like Geolocation, particular hashtags, time and date and many more.

**Data Preprocessing:**

It is a very important stage in our framework where it involves following step

(1)Tokenization:

Tokenization is a process in which terms in the text are split to form tokens (words, phrases or other meaningful elements). Here each tweet is broken up into words, i.e. tokens and passed as input to the next stage.

In text mining, most of the frequently used words in English are unwanted words for mining. So such words are called stop words. And the division of the natural language is the stop words. Stop words include a, an, the, and, are, as, at, be, for, from, has, he, in, is, it, its, of, on, that, the, to, was, were, will, with etc. In this step the stop words should be eliminated to make the tweets look less heavy and



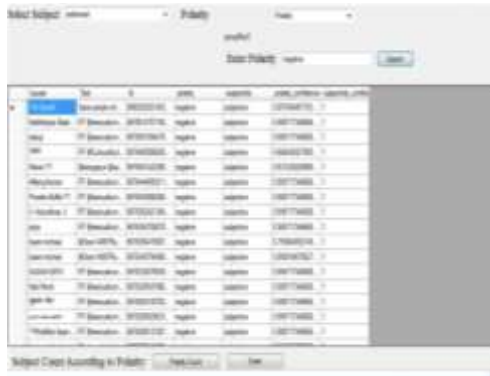
**Public Safety Event Detection:**

Here, In week 1 we calculated the total number of positive We will select the topic which we are focussing on for example we chose Padmavat here.



**Fig 6: Selection of Topic**

By entering polarity in the search box we can overlook tweets having polarity negative, positive or neutral.



**Fig 7. Overlook of Negative Tweets**

Here By clicking on the polarity count button it will show the total number of positive, negative or neutral tweets.



**Fig 8. Number of Negative, Positive and Neutral Tweets**

By clicking on the scale button it will us our final result whether the events related to Padmavat in Gujrat needs public security or not.



**Fig 9. Final Result**

The previous public safety event detection method were based on crowd sensing and a mobile crowd sensing based social media crawler is given with the spatial and temporal information. While this paper includes the hate tweets from nearby locations of an event and about that particular event .It checks whether the tweets negative comments exceed s a particular threshold, if so then the event requires public safety.

**Conclusion**

We have taken 14000 tweets for sentiment analysis of 10 different topics .The 10 topics include some riot causing topics and some random topics.By considering tweets of particular area for example Gujrat and also by choosing the topic Padmavat,we will perform sentiment analysis on tweets . The total number of negative comments will me converted into percentage. If the percent of total negative comments goes above the threshold then the Events related to Padmavat in Gujrat will need Public Safety Arrangements.

**References**

[1] Lei Xu, Chunxiao Jiang,“Microblog Dimensionality Reduction—A Deep Learning Approach,” Ieee Transactions On Knowledge And Data Engineering, Vol. 28, No. 7, July 2016.  
 [2] Zhi-Qiang Xian , “Sentiment Analysis of Chinese Micro-blog Using Vector Space Model,” APSIPA,2014.  
 [3] Amit mittal , “Social Networking text Classification in Big Data Environment,” IJIEET, 2016  
 [4] X. Yan and H. Zhao, “Chinese microblog topic detection based on the latent semantic analysis and structural property,” J. Netw., vol. 8, pp. 917–9233, no. 4, 2013.  
 [5] D. Ramage, S. T. Dumais, and D. J. Liebling, “Characterizing microblogs with topic models,” in Proc. 4th Int. Conf. Weblogs Social Media, pp. 130–137, 2010.

- [6] O. Jin, N. N. Liu, K. Zhao, Y. Yu, and Q. Yang, "Transferring topical knowledge from auxiliary long texts for short text clustering," in Proc. 20th ACM Int. Conf. Inf. Knowl. Manag., pp. 775–784, 2011.
- [6] Q. Diao, J. Jiang, F. Zhu, and E.-P. Lim, "Finding bursty topics from microblogs," in Proc. 50th Annu. Meet. Assoc. Comput. Linguistics: Long Papers-Vol. 1. , pp. 536–544, 2012.
- [8] M. A. Ranzato and M. Szummer, "Semi-supervised learning of compact document representations with deep networks," in Proc. 25th Int. Conf. Mach. Learning, pp. 792–799, 2008.
- [9] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Sci.*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.
- [10] R. Salakhutdinov and G. Hinton, "Semantic hashing," *Int. J. Approx. Reasoning*, vol. 50, no. 7, pp. 969–978, Jul. 2009.
- [11] M. A. Ranzato and M. Szummer, "Semi-supervised learning of compact document representations with deep networks," in Proc. 25th Int. Conf. Mach. Learning, pp. 792–799, 2008.
- [12] S. Zhou, Q. Chen, and X. Wang, "Active deep learning method for semi-supervised sentiment classification," *Neurocomputing*, vol. 120, pp. 536–546, 2013.
- [13] M. R. Min, L. Maaten, Z. Yuan, A. J. Bonner, and Z. Zhang, "Deep supervised t-distributed embedding," in Proc. 27th Int. Conf. Mach. Learn. , pp. 791–798, 2010.
- [14] Nilanjan Banerjee, Dipanjan Chakraborty, Anupam Joshi, Sumit Mittal, Angshu Rai , "Towards Analyzing Micro-blogs for Detection and Classification of Real-Time Intentions" IBM Research – India.
- [15] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discourse Processes*, vol. 25, pp. 259–284, 1998.
- [16] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *CoRR*, vol. abs/1301.3781, 2013.
- [17] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in Proc. 49th Annu. Meet. Assoc. Comput. Linguistics: Human Language Technol.-Volume 1., pp. 142–150, 2011.
- [18] zheng xu, neil y. Yen, hui zhang, xiao wei, zhihan lv, Kim-kwang raymond choo, lin mei, and xiangfeng luo, "Social Sensors Based Online Attention Computing of Public Safety Events", 2015.
- [19] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Sharing clusters among related groups: Hierarchical dirichlet processes," in Proc. Int. Conf. Neural Information Processing Syst, pp. 1385–1392, 2004.
- [20] C. E. Grant, C. P. George, C. Jenneisch, and J. N. Wilson, "Online topic modeling for real-time twitter search," in Proc. Text Retrieval Conf. , pp. 1–9, 2011
- [21] X. Wang, F. Zhu, J. Jiang, and S. Li, "Real time event detection in twitter," in Proc. 14th Int. Conf. Web-Age Inf. Manag., pp. 502–513, 2013.
- [22] D. Yu and L. Deng, "Deep learning and its applications to signal and information processing," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 145–154, Jan. 2011.
- [23] Y. Bengio, A. C. Courville, and P. Vincent, "Unsupervised feature learning and deep learning: A review and new perspectives," *CoRR*, vol. abs/1206.5538, 2012.
- [24] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in Proc. 25th Int. Conf. Mach. Learning, pp. 160–167, 2008.
- [25] J. P. Turian, L.-A. Ratnoff, and Y. Bengio, "Word representations: A simple and general method for semi-supervised learning," in Proc. 48th Annu. Meet. Assoc. Comput. Linguistics, pp. 384–394, 2010.