International Journal of Engineering & Technology, 7 (2.20) (2018) 384-388

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Sensing the Mood on Crimes Against Women by Exploring Social Media Using Dimensional Model

R. Anto Arockia Rosaline[1*], R. Parvathi[2]

[1]Research Scholar, VIT University,
Assistant Professor, Rajalakshmi Engineering College, Chennai, India.
[2]Associate Professor, VIT University, Chennai, India.
**Corresponding author E-mail:antoarockia.rosaline2013@vit.ac.in*

Abstract

Nowadays social media plays a vital role in sharing the information, sharing the views on a particular topic, expressing the sentiments and opinions. Mood is a sub form of sentiment analysis and opinion mining which generally describes the state of mind of a person whether happy, sad, fear and anger etc. Twitter is one of the most popular sources of public opinions. Sensing of mood is generally required to analyze the impact of marketing campaigns, launching of new products etc. This paper focuses on sensing the mood on crimes against women in social media using the quantitative measures of valence and arousal. This paper gives a comparative analysis on mood sensing among different time periods over the same topic.

***Keywords**: Arousal, mood, social media, twitter, valence.*

1. Introduction

Microblogging is so trendy because it has a stream of public opinion. Twitter is one of the most popular microblogging services. Emotional knowledge can be obtained from this opinion rich Twitter. This attitude or the emotional tone is termed to be Mood. It is a strong way of expressing the sentiment. Using this microblogs as source it is possible to analyze the public mood. It is quite often necessary to sense the public mood so as to know the outcome of launching of a new product, marketing campaigns etc. After collecting the tweets through proper API the collected tweets are analyzed and divided into various mood categories.
The contribution to the paper is threefold. First the tweets extracted were analyzed using the categorical model where the number of tweets remained unidentified. The mood classification was not possible for most of the tweets and it was a great challenge to do the classification task for the unknown tweets. Secondly the tweets were subjected to analysis using Dimensional model by which the performance improved a lot. Thirdly the correlation between valence and arousal has been analyzed. The variation in the valence and arousal according to the mood exhibited by the people in social media through tweets has been analyzed.
The remaining part of the paper is organized as follows: In the section two the related work has been discussed. The categorical model used for sensing the mood has been discussed in section three. Section four includes the description of Dimensional model. The discussion about the analysis of crimes against women is included in section five. Section six includes the implementation results, ANOVA analysis and discussion.

2. Related work

Sentiment

Soo-Min et al. presented a system which finds out the opinion holder and the sentiment of the opinion from the given topic [15]. The algorithm can be described in four steps. The sentences which contains both the topic and the holder were extracted, the people holding each sentiment needs to be identified, the polarity of the sentiment bearing words are calculated using sentence sentiment classifier. The system then combines them to produce the holder's sentiment.
Bing et al. focused on the online customer reviews and makes two contributions [2].
They provided a framework to analyze and compare consumer sentiments. It is possible to view the strengths and weaknesses of a product. The different features of the products can be observed and analysed. They proposed a technique to extract the features of a product from the pros and cons in the reviews.
Jaap et al. showed how the distance measure on Wordnet can be used to find out the semantic orientation of adjectives [10]. The lexical database named wordnet has been used and it is used to group the various English words into set of synonyms. The graph theoretic model has been investigated and the words have been analyzed using distance metric. Sunghwan has contributed his work in emotion detection and sentiment analysis [16].The various measures namely precision, recall and F-measure have been used for sentiment identification and the sentiment analysis has been used to identify the student experiences in the learning process.

Emotion Classification

Diman et al. proposed a method to organize neutrality, polarity and emotions hierarchically [6]. Also the task of classifying the

texts automatically by the emotions is explored. The flat and hierarchical approaches were used for comparison to analyze and classify the emotions. The novel method has been tested on two datasets and has been observed that the method outperforms the flat approach which does not include any hierarchical information.
Munmun et al. proposed a method to classify the various affective states of individuals [11]. The human affect on Twitter has been characterized into 11 classes. The explicit mood words were used as the hashtags.

Mood Categories

Diana et al. classified the text by emotions and mood by presenting supervised machine learning methods [5]. The various features used were Bag-of-words, length related features, and sentiment orientation and special symbols such as emoticons. A total of 132 mood categories were taken into consideration. Gilad classified the blog text based on the mood reported by the author at the time of writing [9]. They were capable of identifying the state of mind with which the post has been written. An enormous collection of blog posts, number of features were considered thus evaluating the classification accuracy. Fazel et al. has proved how the novel approach which uses the hierarchy of moods yields better result than the machine learning approach [7]. Also the performance of classification can be very well improved using sentiment orientation features.
Thin Nguyen investigated a novel problem on how to find out patterns on emotion and also the association of lexicon usage and the moods [17]. Also they presented a method in order to discover the association of affective lexicon usage and the moods. Thin et al. provided a comparative analysis of various machine learning-based text feature selection for the problem of mood classification [18]. They also addressed the problem of pattern discovery and mood classification in weblogs.

Valence and Arousal

Georgios et al. work differs from other approaches in so many aspects [8]. They focused on a five class prediction task. It also focuses on heterogeneous domain of blog pots. Instead of exclusively predicting on positiveness and negativeness they explored on ordinal classification of valence and arousal. The affective states can be mapped onto the ordinal scales in the dimensions of valence and arousal which is based on Russell's circumplex model. Munmun and Scott gave a popular representation called Circumplex model [12] which is used to characterize the affective experience through the two dimensions namely valence and activation.

3. Categorical Model

In categorical model the various mood categories are labelled. It is an uncomplicated approach for identifying the mood from the tweets. The major gain of this model is that it is used to comprehend the mood of people with the easy to recognize human mood labels. The major shortcomings of this model are the availability of limited number of mood labels. These categorical models vary by the number of categories they list.
After the authorization process is ended the tweets are collected through proper Twitter API. It is requisite to load all the Twitter authorization libraries and install all the necessary packages. Once the tweets are collected cleaning up of them is required. Remove all the entities such as special characters, punctuation symbols, numbers, html links and all unnecessary spaces.
The tweets related to search term 'Elections' have been collected. Using Naive Bayes method the text is analyzed and classified into different types of mood categories: anger, disgust, fear, joy, sadness, and surprise. The tweets are classified into the well-known labels as depicted in Figure. 1. As depicted in the Figure 1. the tweets that will not fall into one of the known six labels will be in unknown category. The major enhancement required in this case to find out the mood expressed in all the tweets falling into unknown category.

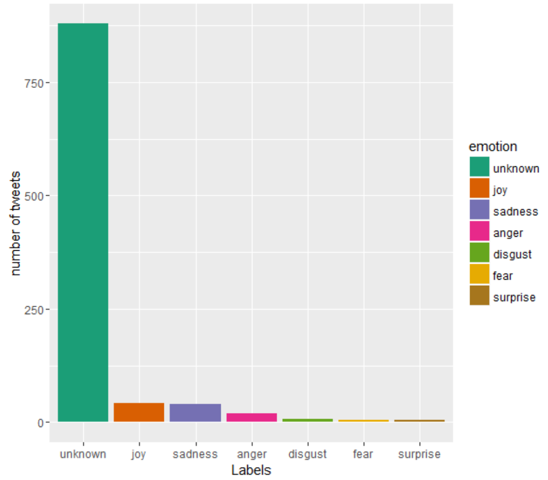

Fig. 1: Classification based on labels

4. Dimensional Model

In dimensional model the representation is purely based on quantitative measures that are used to support multidimensional scaling. The various affect terms are represented in various dimensions. It includes factors such as Valence and arousal. Valence describes the positive affectivity or the negative affectivity. For example the word hate has a mean valence of 2.12 and the word love has a mean valence of 8.72.The arousal factor describes whether the information is so calming or exciting. For example the word quiet has a mean arousal of 2.82 and the word horror has a mean arousal of 7.21.

Affective Norms for English Words (Anew)

The Affective Norms of English words (ANEW) [3] as prescribed by Bradley et al. provides emotional ratings for large number of words in English language. It provides values for various factors such as Valence, Arousal, dominance and Word frequency. Daniel et al defines the valence factor as representing the polarity of the affective content and the arousal factor as representing the intensity of the affective content [4]. ANEW dataset is so helpful to quantify the emotional content in a numerical way as prescribed by Peter et al [14].The ANEW dictionary consists of the values for about 1040 words in English. Those words contain the values for valence and arousal in a nine point scale ranging from 1 to 9.

Steps to Estimate Valence and Arousal

Step1: Extract the tweets from twitter through proper Twitter API
Step 2: Record the values of mean valence (V) and mean arousal (A) for each of the words in a tweet by referring the ANEW dictionary.
Step 3: Find out the frequency of each ANEW word that is occurring in each tweet.
Step 4: Calculate the tweets overall mean valence *valtweet* using the equation 1. [12]

$$valtweet = \frac{\sum_k val_k \, fre_k}{\sum_k fre_k} \qquad (1)$$

Where
val_k represents the valence of tweet k
fre_k is the number of times the ANEW word is present in the tweet k
Step 5: Calculate the tweets overall mean Arousal *arotweet* using the equation 2.

$$arotweet = \frac{\sum_k aro_k \, fre_k}{\sum_k fre_k} \tag{2}$$

where
aro_k represents the arousal of tweet k
fre_k is the number of times the ANEW word is present in the tweet k
Calculate the tweets overall mean valence and arousal values.

Categorical Model vs Dimensional Model

When the 1200 tweets related to the search term 'Elections' are subjected for analysis using the categorical model under the various categories such as joy, sadness, anger, disgust, fear and surprise nearly 24 percent of the tweets only were identified to express the mood. Other tweets though contain the mood was not identified because of the limited labels.
The dimensional model was applied on those tweets. After recording the mean valence and mean arousal of all the words occurring in tweets by referring the ANEW dictionary it has been found out that only 15 percent of the tweets did not express any kind of mood. So those tweets can be discarded as it is not eligible to estimate the mood score. Moreover it is possible to find out the positive affectivity and negative affectivity and also information pertaining to whether it is exciting or calming.

5. Analysis of Crimes Against Women Using Dimensional Model

An investigation has been carried out on the crimes against women happening across the country. In India different types of violence are taking place and the four major types of committed crimes[1] against women include rape, kidnapping and abduction, dowry deaths and torture.
Thomas et al. discovered the Links correspondence Method[19] to analyze how the public reacts to crime news. They considered the spatial dependency for analysis. The method is based on the analysis of the association between the actual location of the crime incident and the crime related tweets.
The tweets associated to the crimes have been composed through twitter API. The pre-processing of tweets is done using Porter's stemming algorithm. After the preprocessing the ANEW words are searched in each of the tweet. For each of the word in the tweet the valence and arousal values are computed. If more than one ANEW words are present the average of the valence and arousal values are computed. Then the collective average of the valence and arousal values is computed for the four months and for the various regions such as Chennai, Mumbai, New Delhi and Kolkata.
The correlation and regression analysis are used to quantify the association between the two continuous variables namely valence and arousal. One way analysis of variance procedures are used to produce an analysis for the quantitative dependent variable affected by independent variable. Analysis of variance is generally used to test the hypothesis. The mood of the public can very well be found out and compared.
In order to sense the mood on crimes against women the tweets have been extracted. The tweets are subjected to preprocessing and nearly 170 stop words are removed from the tweets. The stemming algorithm is applied on those tweets. Once the preprocessing is done the overall mean valence and mean arousal for the tweets are computed using the weighted average using equations 1 and 2.The tweets extracted during the month of September 2017 and December 2017 are taken into consideration, compared and the mood has been analyzed.

Table 1: Valence vs Arousal Values

Category	September 2017	December 2017
Minimum Valence value of the tweets	1.25	1.25
Maximum Valence value of the tweets	8.26	8.72
Minimum Arousal value of the tweets	2.65	2.82
Maximum Arousal value of the tweets	7.51	7.49
Average valence value of the tweets	3.595	5.057
Average Arousal value of the tweets	6.112	6.076

It is observed from Table 1 that the average valence value of the tweets extracted during the month of December 2017 is considerably higher when compared to the month of September 2017.
The Table 2 shows the top 10 words having the highest frequency of occurrence.

Table 2: Word Frequency

S.No	September 2017		December 2017	
	Word	Frequency	Word	Frequency
1.	Violent	975	Hate	221
2.	Social	235	People	122
3.	Taxi	134	Fight	84
4.	Hit	128	Sex	65
5.	Time	120	Justice	64
6.	Abuse	90	Family	63
7.	Rape	37	Violent	58
8.	Misery	25	Rape	35
9.	News	24	Innocent	33
10.	Sex	22	Save	30

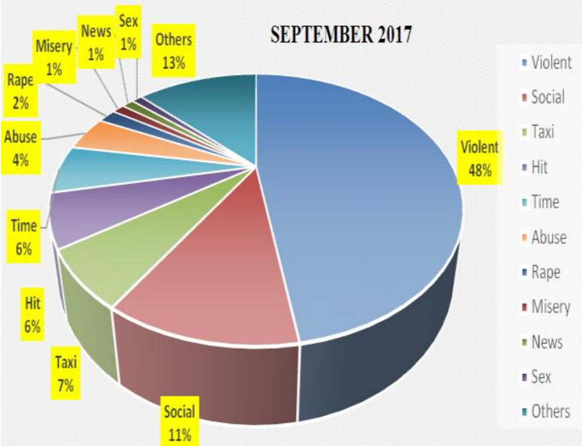

Fig. 2: High frequency word list for the month of september 2017

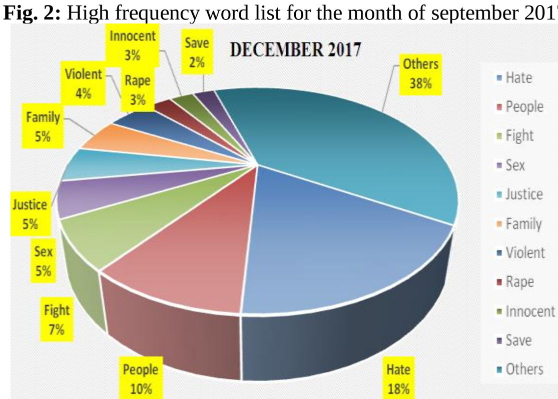

Fig. 3: High frequency word list for the month of december 2017

Figure 2 and 3 illustrates ANEW words in the tweets during the months of September and December 2017.Some of the words such as violent, rape and sex occurs in both the months.

6. Results and Discussion

Peter et al considered the various possible relations[13] between valence and arousal as independence, positive linear relation, negative linear relation, symmetric V-shaped relation, asymmetric V-shaped relation including both positivity offset and negativity bias, and an inverted V-shaped relation when valence is a function of arousal.

On observing the valence and arousal values of the most often used words, the arousal values are more of dependent on valence values. The arousal value representing the intensity of the tweet depends on the polarity of the tweet. The exciting or calm state depends on how pleasure the mood is.

The Pearson's correlation coefficient as described in the equation 3 has been applied on valence and arousal factors to find the association between them.

$$r = \frac{\sum (x-\overline{x})(y-\overline{y})}{\sqrt{\sum\left(x-\overline{x}\right)^2}\sqrt{\sum\left(y-\overline{y}\right)^2}} \tag{3}$$

On applying the correlation analysis the value of r is -0.64128 for September 2017 and -0.48156 for the month of December 2017.It has been observed that the valence(X) and arousal(Y) are negatively correlated to each other.

The mood patterns could be analyzed by grouping the words in the various categories such as High valence and Low Arousal, High valence and Low Arousal, Low Valence and High Arousal and Low valence and Low Arousal. The number of tweets grouped into the above categories is in Table 3 and Table 4.

Table 3: Valence vs Arousal for the Month of September 2017

	Arousal(0-4.4)	Arousal(4.5-9)
Valence(0-4.4)	3	1132
Valence(4.5-9)	86	307

Table 4: Valence vs Arousal for the Month of December 2017

	Arousal(0-4.4)	Arousal(4.5-9)
Valence(0-4.4)	2	326
Valence(4.5-9)	40	492

More number of tweets occurs in the group having low valence and high arousal during the month of September and more number of tweets occurs in the group having high valence and high arousal during the month of December. It is hereby concluded that over a month period there could be a change in the mood they are exhibiting. Thus it is clear that over a period the valence and arousal factors can lead to positivity or negativity.

ANOVA Analysis

The valence and arousal values computed on crimes against women for the various regions across India for four different time periods are illustrated in Table.5.

Table 5: Valence and Arousal Values

Month	Valence/Arousal	Chennai	Mumbai	New Delhi	Kolkata
Month1	Valence	3.845	3.23	3.773	2.677
	Arousal	5.797	5.744	5.874	6.392
Month2	Valence	4.275	3.733	4.099	3.522
	Arousal	6.470	6.369	6.266	6.395
Month3	Valence	3.667	5.287	4.973	5.863
	Arousal	5.594	5.136	5.196	5.729
Month4	Valence	5.020	5.394	5.44	4.375
	Arousal	5.565	4.892	5.06	6.466

ne way analysis of variance procedures produce an analysis for the quantitative dependent variable affected by independent variable. Analysis of variance is generally used to test the hypothesis.

H0: The four regions Chennai, Mumbai, New Delhi and Kolkata exhibit same mood and are not different from one another.

H1: The four regions Chennai, Mumbai, New Delhi and Kolkata exhibit different mood patterns and are different from one another.

Table 6: Region Wise Mean and Standard Deviation Values

Region	Mean	Standard Deviation
1	5.89	1.17
2	5.60	1.17
3	5.71	1.145
4	6.24	1.10

The Table 6 lists the means, standard deviations of the samples of each group where 1, 2, 3 and 4 representing Chennai, Mumbai, New Delhi and Kolkata respectively.

Table 7: Between groups and within groups variations

	Sum of squares	df	Mean square	F	Sig.
Between Groups	7.921	3	2.640	1.977	0.117
Within Groups	413.983	310	1.335		
Total	421.904	313			

The above Table.7 lists the sum of squares of the differences between means of different regions and their mean squares. In the Table.7, the between groups variation 7.921 is due to interaction in samples between groups. The sample means are close to each other and so the value is small. The within groups variation 413.983 is due to the differences within each sample. The degrees of freedom df1 is one less than the number of sample regions (4-1=3), and df2 is the difference between the total sample size and the number of sample regions (314-4=310).Calculate the mean square values by dividing each sum of squares value by its degrees of freedom (df).The F statistic 1.977 is calculated by dividing the Between Groups mean square by the Within Groups mean square. There were no statistically significant differences between group means as determined by one-way ANOVA

7. Conclusion

The present mood of the society on crimes against women could be analyzed from tweets using the dimensional model. Using the quantitative measures and the affect terms placed along the various dimensions it is quite good to sense the mood in social media. Thus the mood in social media could be sensed through the various dimensions of emotion such as valence and arousal. It could be used to analyze the increase or decrease in valence and arousal among public. This paper performs the ANOVA analysis and finds out the variations of mood exhibited over various time periods. In future it is planned to do further analysis on mood prediction based on the valence and arousal values and to carry out the analysis on the extracted features.

References

[1] Sinha A & Singha P, "CRIME AGAINST WOMEN IN INDIA: TREND AND REGIONAL VARIATION", *International Journal*, Vol.3, No.10,(2015), pp.1403-1406.

[2] Liu B, Hu M & Cheng J, "Opinion observer: analyzing and comparing opinions on the web", *Proceedings of the 14th international conference on World Wide Web*, (2005), pp.342-351.

[3] Bradley MM & Lang PJ, "Affective norms for English words (ANEW): Instruction manual and affective ratings", *Technical report C-1, the center for research in psychophysiology, University of Florida*, (1999).

[4] Preoţiuc-Pietro D, Schwartz HA, Park G, Eichstaedt J, Kern M, Ungar L & Shulman E, "Modelling valence and arousal in facebook posts", *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, (2016), pp.9-15.

[5] Inkpen D, Keshtkar F & Ghazi D, "Analysis and generation of emotion in texts", *KEPT*, (2009), pp.3-13.

[6] Ghazi D, Inkpen D & Szpakowicz S, "Hierarchical versus flat classification of emotions in text", *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, (2010), pp.140-146.

[7] Keshtkar F & Inkpen D, "Using sentiment orientation features for mood classification in blogs", *International Conference on Natural Language Processing and Knowledge Engineering*, (2009), pp.1-6.

[8] Paltoglou G & Thelwall M, "Seeing stars of valence and arousal in blog posts", *IEEE Transactions on Affective Computing*, Vol.4, No.1,(2013), pp.116-123.

[9] Mishne G, "Experiments with mood classification in blog posts", *Proceedings of ACM SIGIR workshop on stylistic analysis of text for information access*, Vol.19, (2005), pp.321-327.

[10] Kamps, J., Marx, M., Mokken, R.J. and De Rijke, M., 2004, May. Using WordNet to Measure Semantic Orientations of Adjectives. In *LREC* (Vol. 4, pp. 1115-1118).

[11] De Choudhury M, Gamon M & Counts S, "Happy, nervous or surprised? classification of human affective states in social media", *Sixth International AAAI Conference on Weblogs and Social Media*, (2012).

[12] De Choudhury M & Counts S, "The nature of emotional expression in social media: measurement, inference and utility", *Human Computer Interaction Consortium (HCIC)*, (2012).

[13] Kuppens P, Tuerlinckx F, Russell JA & Barrett LF, "The relation between valence and arousal in subjective experience", *Psychological Bulletin*, Vol.139, No.4,(2013).

[14] Dodds PS & Danforth CM, "Measuring the happiness of large-scale written expression: Songs, blogs, and presidents", *Journal of happiness studies*, Vol.11, No.4, (2010), pp.441-456.

[15] Kim SM & Hovy E, "Determining the sentiment of opinions", *Proceedings of the 20th international conference on Computational Linguistics*, (2004).

[16] Mac Kim S, *Recognising emotions and sentiments in text*, University of Sydney, (2011).

[17] Nguyen T, "Mood patterns and affective lexicon access in weblogs", *Proceedings of the ACL Student Research Workshop*, (0210), pp.43-48.

[18] Nguyen T, Phung D, Adams B, Tran T & Venkatesh S "Classification and pattern discovery of mood in weblogs", *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, (2010), pp.283-290.

[19] Lampoltshammer TJ, Kounadi O, Sitko I & Hawelka B, "Sensing the public's reaction to crime news using the 'Links Correspondence Method'", *Applied geography*, Vol.52, (2014), pp.57-66.