Design Of Emotion Mining using Data Mining Techniques

Harsimranjit Singh[1],Jatinder Singh Bal ,Vijay Rana[2]
Research Scholar 1.Department of Computer Science and Application,
Sant Baba Bagh Singh University , Jalandhar, Punjab, India.
Email id:ruby98786@yahoo.co.in

Abstract **-**As commonnetworksite are trendy, so they turn intomain part of person's societal interaction. These communalnetwork sites are rich in emotions where people share their feelings, opinions, emotions. Extracting emotions from these social networking sites play an important role in various fields. Many techniques are proposed by various authors to removeemotion from these common network sites. This study work plannedaagenda to remove emotions from social media text by applying ensemble classification technique.

Keywords **-**emotion mining, data mining, opinions, social media, ensemble classification.

I. INTRODUCTION

Emotion are a obligatoryelement of being nature that can be consider as hereditary. Also it havebeestablish that appearance of a exacting emotion by dissimilar human being is the same. Several persistent emotions that last a great deal longer result in mood. Mood can be aoutcome of a mixture of certain emotions of a person. On the complete emotions can be categorised into two: fundamental and difficult. Basic emotions are delight, unhappiness, irritation, horror, repulsion and revelation as discussed by Ekman [1]. The complex emotions are a grouping of two or more basic emotions that are experienced by a person at an instance [2]. The civilization, the world we live in, our relatives had once seen anperiod when people converse through letter posts, telegram etc. A very recentmethod people use is telephone, wired and regularlybecome wireless. Nowadays for interact with friends, relatives, social contacts and among people of dissimilarcommunity, section and countries.Common Networking becomes a very dominant and essential tool. This technique affects and transformed the mode people relate with each other. Certainly, this becomes a very essential part of everyday life. Common networking websites like Facebook, LinkedIn, Twitterprovide a very authoritative medium and a platform for communication amongst individuals leading to common learning and contribution of knowledge. Societal Web based applications also provideopportunity to individuals or to a grouping of individuals for contribution of information in form of Wall posts, comments, chat and discussions. This provide an emotionally rich environment to the subjects to Interact [5].

Data mining is a authoritative tool that can assist to find patterns and relationships within our data. Data mining discoverunseen information from huge database. The overall objective of the data mining process is to remove information from a data set and convert it into anreasonable structure for extra use. Common networks can be used in numerous business activities like increasing word-of-mouth marketing, marketing research, General marketing, Idea generation & new product development, Co-advance, Customer service, civic relations, worker communications and in Reputation organization [7].

Emotion mining from text comprises of following phases:

- Raw Data Collection
- Data pre-processing utilizing natural language processing methods
- Feature Selection
- Identification of Emotions
- Classification of Emotions
- Evaluation stage

II. BACKGROUND WORK

As internet becomes universal mean to communicate, world has changed at a very fast pace. Social networking sites are a common means to interact with each other. These networking sites are most searched pages on the internet. Dhawan et al. [2] reviewfeeling mining techniques in common networking sites. He stated that latest research shows that it is advisable to deal with various kinds of sentences by various strategies. Also author describe the requirement of some tools to mine certain emotions from different sources of data that gives accurate outcomes.

Mohamed et al. [3] surveyed enhancement done in the field of emotion mining along with comparative study for various techniques. Also, this paper presented investigation on the technology utilized in this area and also on the implementation of these techniques. These techniques are categorized into two areas: lexical based techniques and ML based techniques. A descriptive process cycle to build such type of systems is also presented in this paper.

A novel technique to study friendship emotions and relations is presented by Yassine et al. [4]. The main aim of this work is to extract information about emotions from social networking sites. Author shows interest to depict if writer shows his/her emotions through their writings. This study considers a case study called Lebanese Facebook users. It is based on k-means clustering technique and is unsupervised technique.

Kumari et al. [5] discussdissimilar methods and technique to mine emotion from text. Emotion study algorithms are utilize to approximate the public emotion on dissimilarexertiondiscuss in the common network. Based on text comments author categorizedfriendship type.

Dhawan et al. [6] presented a novel perspective to studyexpressions of emotions' in online common networks. This paper adopts unsupervised technique; mainly it utilizes the nearest neighbor algorithm and k-means clustering algorithm. Results demonstrate enhanced accuracy for model in predicting emotions and determining subjectivity of texts.

Sharma et al. [7] discussregarding data mining techniques. This paper offered a survey of the work perform in the region of common network investigation and also focus on the future span in study on common network analysis. This paper presented study related to social networks utilizing Web mining methods.

Qamar et al. [8] proposed to utilize fuzzy logic to detect emotional content from text. Fuzzy logic was developed to handle ill-defined concepts. The transfer fromone physiological state of emotion to another is gradual and it is easy to model by fuzzy logic technique.

Dutta et al. [9] presented the chargebaseon fuzzy logicimplementation to recognize emotions from text. This technique is based on execution of fuzzy logic to recognize emotions from text in MATLAB environment.

Colneric et al. [11] explorethe utilise of deep learning for emotion recognition and produced three large collection of tweets labelled with Ekman's, Plutchik's and POMS's classifications of emotions. Persistent neural networks indeed better the baseline set by the common bag-of-words model. Experimentpropose that it is superior to train RNNs on sequence of characters than on sequences of words.

Stojanovski et al. [12] exploit an convolutional neural network architecture for emotion analysis in Twitter messages related to sporting events on 2014 FIFA world Cup. In this paper, seven different kinds of emotions were evaluated using hashtag labeled tweets that were collected from Twitter Streaming API. The training of the network is performed on two samples containing 1000 and 10000 tweets on which this approach achieves 50.12% and 55.77% accuracy respectively. Moreover, they have presented the analysis of this approach on three different games that have great impact on Twitter users.

Mishne et al. [13] addressed the task of classifying blog posts on the basis of mood of the writers. They obtainaenormousquantity of blog posts from one of the biggest online blogging communities Livejournal. The author took the advantage of the Livejournal that allows writers to update their current mood from the 132 given categories. Yahoo API was utilised to obtain a list of 1000 web pages contain a Livejournal blog post with each kind of mood.

Roberts et al. [14] presented in public available quantity of tweets annotate with seven different kinds of emotions comprising Ekman's six emotions and love. The main goal of this research is to present emotion distributions in different emotion evoking topics. Tweets were collected from twitter using Twitter API on 14 different topics that evoke emotions. Moreover, they describe a baseline scheme for mechanically annotating of emotions for tweets in which they used previously annotated tweets by professional annotators for training the data. They used series of binary SVM classifiers to detect each of the seven emotions annotated in the corpus they created. Additionally, linguistic style features are also presented in this paper for topic "U.S Elections 2012".

Wang et al. [15] exploit a technique to automatically annotate a large amount of data. They extracted large amount of tweets (2.5 million) from twitter instead of using already annotated corpus which consists of just thousands of tweets. The main focus of this research is to think about the viability of different element mixes and in addition the impact of the measure of the preparation information on the feeling analysis work.

III. PROPOSED TECHNIQUE

This research work implement a framework to extract emotions from social media text by applying ensemble classification technique. Methodology followed by proposed technique consists of following steps:

A. Data Acquisition and Annotation: The first and foremost step is to collect data for emotion analysis. Tweets are collected from twitter using Twitter API on Crimes. Annotation of tweets is done manually on the basis of Ekman's six basic emotion.

B. Preprocessing and Filtration: Every word contained in a tweet is important in decision making, therefore effective pre-processing of these tweets is an important task because these tweets are full of slang, misspellings and words from other languages. Therefore, in order to deal with this kind of noisy data, normalization of tweets is performed by intelligent text pre-processing techniques like tokenization, stop-word removal, stemming, lemmatization dimensionality reduction. Pre-processing of data is performed by the following techniques:

i). Data Cleaning: Cleaning of data involves handling of missing values by ignoring that particular tuple. If any tuple or cell is empty then that will be filled with some specific value. Inconsistency of data may be handled manually. It also handles noisy data by implementing machine inspection, clustering, binning methods and regression. All the quotes("") from the sentences are removed, URL's are removed and other characters that are not considered to be in the category of texts are removed.

ii). Data integration: Data is always collected from various sources like data warehouse, internet etc. so the collected data in particular so not have any use. It has to be added altogether for further analysis. So this step will integrate the data collected from various sources.

iii). Data transformation: Transformation of data means to change the data from one form to another. For this purpose, various methods like smoothing, normalization, aggregation and generalization are available for the transformation. Transformation steps are as follows:

- Sentence Splitting: The first step involved is sentence splitting i.e the splitting of string into words. Identifying sentence boundaries in a document is not a smaller task.
- Tokenization: Tokenization of words means to split a sentence into tokens or smallest unit of a sentence. Tokenization is an important task because many succeeding components need tokens clearly identified for analysis.
- Stop Word Filtering: There are a lot of words that do not have any meaning and can be removed from the input file. Words like "the", "and", "for", "or", "if", "that"; are referred to as stop words because they don't signify any meaning or sentiment. Therefore, removal of such words means stop word filtering and it also improves the performance of the system
- Stemming: A stemming calculation is a procedure of semantic normalization. In this process, the variations of a word are lessened to a typical frame. For instance, consider a simple example below:

Communication
Communications
Communicative
------→
Communicate
Communicated
Communicating

Feature Groups: Feature sets are defined for automatic classification of emotions in tweets, it is essential to consider emotional words which distinctly characterize emotions in tweets with hashtags.

i). Hashtag Emotion Lexicon: which provides association of words with eight emotions (irritation, horror, eagerness, faith, revelation, unhappiness, delight, and repulsion) generated automatically from tweets with emotion-word hashtags such as #happy and #anger.

C. Classification: Classification is done to automatically classify tweets into an emotion category. Ensemble machine learning algorithm is used to train the system.

With a specific end goal to enhance the execution of individual classifiers used in the paper is to the use of ensemble learning. While the benefit of utilizing outfit, strategies is the change of the execution, the weakness is about the time it takes to complete the training phase. However, the main concern was to build a model which has a better performance compared to the individual classifiers. This model is novel because not only it uses ensemble method (Voting), but also it applies a meta classifier as one of its classifier components. Also, parameter optimization approach was used on its individual classifier. Each component in our model learns some parts of the classification problem and we combine these hypotheses to decide the probability level.

D. Evaluation: Evaluation of performance has to be done using metrics like accuracy, precision, recall, F-measure, TP Rate, FP Rate and then the results of Support Vector Machines, Naïve Bayes and Ensemble classifiers has to be compared on the basis of these parameters.

IV. EXPERIMENTAL RESULTS

This section presents experimental results of the proposed technique. The proposed technique is implemented in weka tool. Various parameters are used to evaluate the results of the proposed technique. These parameters are:

- Recall
- F-Measure
- Mean Absolute Error
- Root Mean Squared Error
- Relative Absolute Error
- Root Relative Squared Error

Recall: Recall is measurement that is generally used to evaluate execution in content mining, and in content examination field like data recovery. This parameters is utilized for estimating completeness.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

F-Measure: F-Measure is harmonic mean of precision & recall. The esteem calculated using F-measure is a balance between precision & recall.

$$\text{F measure} = \frac{2 * \text{recall} * \text{precision}}{\text{precision} + \text{recall}}$$

Mean Absolute Error: The MAE measures the common size of the errors in a blueprint of checks, without considering their bearing.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_j - \hat{y}_j|$$

Root Mean Squared Error: The RMSE is a quadratic scoring rule which measures the typical size of the mix-up.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n} (y_j - \hat{y}_j)^2}$$

Table 1: Error Comparison

Parameter	SVM	Naïve Bayes	Proposed
Mean absolute error	0.0378	0.024	0.0291
Root mean squared error	0.1945	0.1345	0.108
Relative absolute error	0.1449	0. 9131	0.1115
Root relative squared error	0.5389	0. 3717	0.2991

Above table shows comparison of proposed technique with existing techniques on the basis of various error parameters.

It is clear from below figure that error of proposed technique is much less than that of existing techniques.

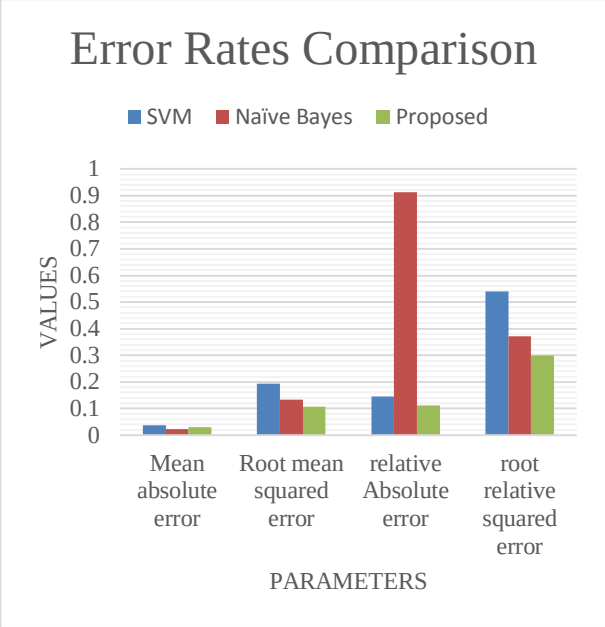

Figure 1: Showing comparison of MAE, RMSE, RAE, RRSE

Table 2: Class Parameters Comparison

Parameter	SVM	Naïve Bayes	Proposed
Precision	0.903	0.940	0.962
Recall	0.886	0.941	0.962
F-measure	0.879	0.938	0.962

Above table shows comparison of proposed technique with existing techniques on the basis of class parameters. It is clear from below figure that proposed technique gives better precision, recall, and f-measure than that of existing techniques.

Figure 2: Showing comparison of class parameters

Table 3: Classification Accuracy Comparison

Parameter	SVM	Naïve Bayes	Proposed
Accuracy	88.6486	94.0541	96.2162

Above table shows comparison of proposed technique with existing techniques on the basis of accuracy. It is clear from below figure that accuracy of proposed technique is much better than that of existing techniques.

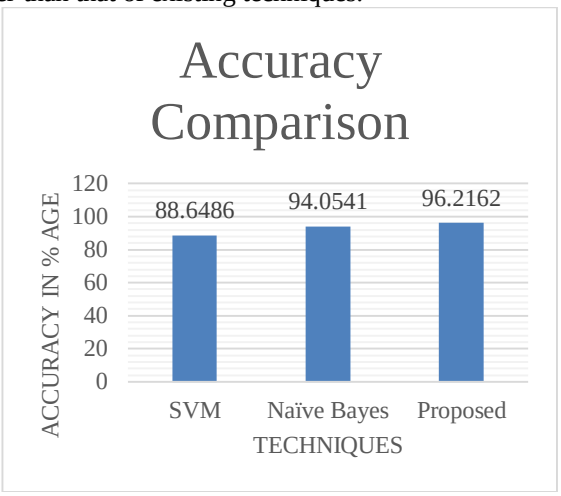

Figure 3: Showing comparison of Accuracy

Table 4: Kappa Statistics Comparison

Parameter	SVM	Naïve Bayes	Proposed
Kappa Statistics	0.8509	0.924	0.9515

Above table shows comparison of proposed technique with existing techniques on the basis of kappa statistics. It is clear from below figure that kappa statistics of proposed technique is much better than that of existing techniques.

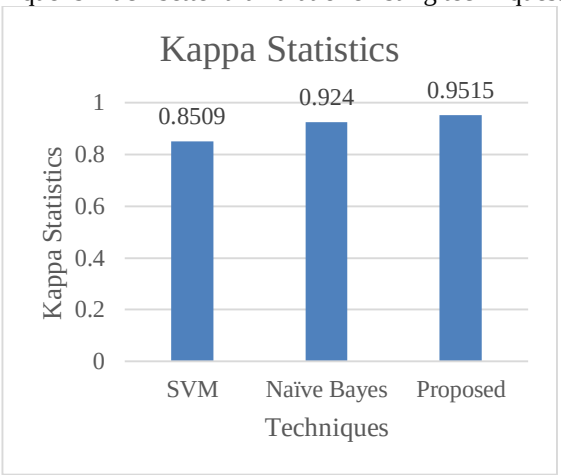

Figure 4: Showing comparison of Kappa statistic

V. CONCLUSION AND FUTURE SCOPE

Public emotions present in Social media data offers unique challenges and opportunities for decision-making in

different domains.. Data is collected from Twitter specifically tweets on crime. Natural language processing and computational techniques are applied for the pre-processing and filtering of the noisy data and to normalize the data. Moreover, popular machine learning algorithms Naïve Bayes and Support Vector Machines are applied for automatic classification and detection of emotions. These are the combined in ensemble learning proposed technique. An ensemble schema based on three statistical classifiers combined by stacking has to be used for classification. Support Vector Machines and Naïve Bayes are combined on the base-level and Random Forest is integrated at the meta-level forming an ensemble schema of classifiers. It is concluded the accuracy achieved for classification by proposed ensemble learning is better than that of Naïve Bayes and SVM. Naïve Bayes obtained an accuracy of 94.0541%, however accuracy obtained with SVM is 88.6486 %. and the proposed technique accuracy obtained is 96.2162 %. The work presented in this research can be pursued further in several domains. One of the tasks is to consider emotion intensity for classification. Explore the relation between emotion classes and emotion intensity. Content-based analysis of emotion data is yet another possible line of research. Data sets containing stickers, emoticons, & other images with texts representing emotions can also be taken into consideration in future.

VI. REFERENCES

[1]. P. Ekman, "An argument for basic emotions", Cognition and Emotion, 1992, pp. 169-200.

[2]. Sanjeev Dhawan, Kulvinder Singh, DeepikaSehrawat, "Emotion Mining Techniques in Social Networking Sites", International Journal of Information & Computation Technology, Vol. 4, No. 12, 2014, pp. 1145-1153.

[3]. Hany Mohamed, Ayman Ezzat, Mostafa Sami, "The Road to Emotion Mining in Social Network", International Journal of Computer Applications, Vol. 123, No. 18, 2015, pp. 41-47.

[4]. Mohamed Yassine, Hazem Hajj, "A Framework for Emotion Mining from Text in Online Social Networks", IEEE, International Conference on Data Mining Workshops, 2010, pp. 1136-1142.

[5]. Rashmi Kumari, MayuraSasane, "Emotion analysis using text mining on social networks", International Journal of Innovative Research in Technology, 2015, pp. 20-27.

[6]. Sanjeev Dhawan, Kulvinder Singh, VandanaKhanchi, "A Framework for Polarity Classification and Emotion Mining from Text", International Journal Of Engineering And Computer Science, 2014, pp. 7431-7436.

[7]. Anu Sharma, Dr. M.K Sharma, Dr. R.K Dwivedi, "Literature Review and Challenges of Data Mining Techniques for Social Network Analysis", Advances in Computational Sciences and Technology, Vol. 10, No. 5, 2017, pp. 1337-1354.

[8]. Saqib Qamar, Parvez Ahmad, "Emotion Detection from Text using Fuzzy Logic", International Journal of Computer Applications, Vol. 121, No.3, July 2015, pp. 29-32.

[9]. Er. Vibha Dutta, Dr. R. C. Gangwar, Er. MohitMarwaha, "An Effectual Emotion Recognition System Based on Fuzzy Logic", International Journal of Latest Trends in Engineering and Technology, Vol. 8, Issue. 1, pp.122-131.

[10]. Mohan Arava, TalasilaVamsidhar, "Driver Emotional Status Recognition Using Fuzzy Logic in ADAS: An Indian Perspective", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue. 10, October 2014, pp. 328-334.

[11]. Niko Colneric, JanezDemsar, "Emotion Recognition on Twitter: Comparative Study and Training a Unison Model", IEEE Transactions on Affective Computing, 2018, pp. 1-14.

[12]. Stojanovski, Dario, et al. "Emotion identification in FIFA world cup tweets using convolutional neural network." Innovations in Information Technology (IIT), 2015 11th International Conference on.IEEE, 2015.

[13]. Mishne, Gilad. "Experiments with mood classification in blog posts."Proceedings of ACM SIGIR 2005 workshop on stylistic analysis of text for information access.Vol. 19. 2005.

[14]. Roberts, Kirk, Michael A. Roach, Joseph Johnson, Josh Guthrie, and Sanda M. Harabagiu. "EmpaTweet: Annotating and Detecting Emotions on Twitter." In LREC, vol. 12, pp. 3806-3813. 2012.

[15]. Wang, Wenbo, Lu Chen, KrishnaprasadThirunarayan, and Amit P. Sheth. "Harnessing twitter" big data" for automatic emotion identification." In Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom), pp. 587-592. IEEE, 2012.