



Gender based Emotion Mining from the Text on Social Networking Sites (SNS)

¹Tajinder Kaur, ²Dr. Kiranjyoti

¹M.tech. Student, ²Assistant Professor

^{1,2}Department of Computer Science & Engineering

Guru Nanak Dev Engineering College, Ludhiana, Punjab, India

Abstract- Social Networking Sites (SNS's) has provided a platform for the online users to share their emotions, opinions and ideas with each other on some significant matter either positively or negatively. The online user can be identified as a Male or Female. Predicting the gender identity and the emotional state of the online user on the basis of text/ comment written by him/her is a most challenging task. The main purpose of this study is to review various emotion detection techniques to predict the exact behavior of the gender on the basis of emotions expressed on social networking sites like Tweeter, Facebook etc.

Keywords- Emotions, Gender, API, SNS, EM (Emotion Mining), Text.

I. INTRODUCTION

Emotion mining deals with the automated discovery and extraction of knowledge about emotions of the people. It recognizes emotions from textual data such as personal blogs, review websites, customer forms, and comments on SNS. The process of extracting emotion from the text is known as emotion mining. Text can be in the form of comments, tweets and posts etc. Mining people's feelings and predicting their reactions to events and generating suitable emotions are typical task in Emotion Mining.

Need of Emotion Mining based on Gender differences:

This is a widespread belief that women are emotionally expressive whereas the men are emotionally impassive. The research has confirmed that tendency for women to report feeling stronger and longer emotion and to express emotions more clearly. Gender differences in emotion provided vary by social context and by sort of emotion. Women seem more ready to express negative emotions whereas men seem more ready to let somebody see anger. In research, it is found that women use positive emotions amongst friends whereas men don't. These gender differences may be intensified for emotions expressed on SNS's, because women seem to be more ready to share emotions in public than men.

Gender and Language use:

Gender variations in language use are relevant to the research question especially when they relate to emotional expression. Gender variations have two main factors based on the basis of languages used are:

1. Affiliative Language: - Affiliative language contains positive words or emotions in a sentence. It positively engages the person. Women tend to use affiliative language more by showing support or expressive agreement.
2. Assertive Language: - Assertive language contains negative words or emotions in sentence. It contains criticism and directive statements. Men tend to use assertive language more because male mostly uses words through commands and challenging statements.

II. PROCESS OF EMOTION MINING

1. Emotion Extraction: - Emotion extraction is one of the main techniques to understand emotions. It allows extraction of emotion at word, sentence and document levels from the articles. These articles are available online on web in huge large data. SNS's allow to extract emotions online using APIs. For example Graph API is used to extract the data from facebook.

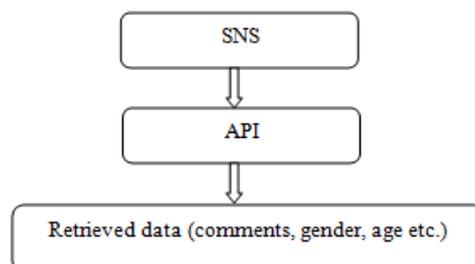


Fig.1: Online text extraction using API's from SNS's

2. Emotion Analysis: - In order to mine the emotions from the text, the text should be analyzed. It involves the following tasks:-
 - 2.1 Symbolic-Cue analysis: - In symbolic cue analysis Sentences are analyzed for sensations, acronyms, outbursts etc.
 - 2.2 Word-Level analysis: - In word level analysis, each word found in the database is analyzed to find the characteristics of word that represents a vector of emotional state.
 - 2.3 Syntactical-Structure analysis: - In syntactic structure analysis, the syntactic structure of the sentence is analyzed.
 - 2.4 Phrase-Level analysis: - The main aim of performing phrase level analysis is to detect the emotions involved in phrases and then in subject, verb and object form.
 - 2.5 Sentence-Level analysis: - In this kind of analysis the emotional vector of sentence or clause is generated from verb, subject and object formation.
3. Emotion Detection: The process of detecting the emotional state of person by analyzing text written by him/her is known as emotion detection. Recognizing the emotion of the text plays a key role in the human computer interaction. Emotions may be articulated by speech, facial and text based emotion respectively. Various methods used for text based emotion detection are:
 - 3.1 Keyword spotting technique: it can be described as the problem of finding occurrences of keyword (emotion words) from a text document. This method is based on certain predefined keywords. Further these words are classified into categories such as disgust, sad, happy, angry, fearful, surprised etc. Rate of these emotion words discover and on the basis of that an emotion class is assigned to the text document.
 - 3.2 Lexical affinity methods: Detection of emotion based on related keywords is an easy to use and basic method. Lexical affinity approach is an expansion of keyword spotting technique. It consigns a probabilistic 'affinity' for a particular emotion to arbitrary words part from picking up emotional keywords. These probabilities are often a part of linguistic corpora but have disadvantages: first disadvantage is the assigned probabilities are biased toward corpus generated class of text, second it fails to spot out emotional content that resides deeper than the word level on which technique operates. For example the word "Accident" signifies negative emotion would not contribute correctly to the emotional assessment of the phrases 'I avoided an accident' and 'I met my friend by accident'.
 - 3.3 Learning based methods: Learning based methods are used to formulate the problem differently. First determine the emotions from the input text and then classify the input text into different emotions. disparate keyword based method, learning based methods try to detect emotions based on a previously trained classifier, which pertain various theories of machine learning such as super vector machine(SVM) and conditional random words to determine which emotion category should the input text belongs.
 - 3.4 Hybrid methods: Keyword based methods and learning based methods could not acquire satisfactory results some systems use hybrid approach by combining both keyword spotting techniques and learning based methods which helps to improve the accuracy.
4. Emotion Classification: It is the process of classifying the text on the basis of emotional polarity into three major classes namely as:
 - 4.1 Positive: Positive emotional text represents positive feelings of the individual.
 - 4.2 Negative: negatively expressed emotional text shows negative feelings of the human beings
 - 4.3 Neutral: Text that does not slouch in positive and negative category is known as Neutral emotional text.

III. SURVEY OF THE EMOTION MINING BASED ON GENDER DIFFERENCES AND ITS METHODS

[1] Ahmad et al. (2015) , proposed model , they divided different emotions like happiness, sadness, anger, fear, disgusted and surprise into different 11 levels. In this case one is the lowest value for the sad emotion while 11 is the highest state of the emotion. In the proposed model they had calculated cognition skills in different individuals under different conditions. The 50 voluntaries were used from different age groups with equal representation of both genders. In this paper, there experimented how to measure state of different emotions and how to test their cognitive abilities.

[2] Patel and Mistry (2015) had discussed that social networking sites are one of the most popular media among the internet users to express their emotions. They had conducted a review on various data classification techniques namely as Super Vector Machine (SVM), Naïve Bayes, Decision Tree, K-Nearest Neighbor and Neural Network. Each classification techniques have own advantages and disadvantages. The authors had compared classification techniques with each other. They had also discussed document frequency selection methods. It includes document frequency threshold, information gain, mutual information and chi-square statics.

[3] Dhawan et al. (2014) proposed a new framework for characterizing emotion based interaction in social media. They had performed text mining techniques on comments/ messages to determine the types of emotions like happy, sad, angry, disgust, fear and surprise. They had presented a new prospect for studying emotional expressions based on number of interjections, emoticons, social acronyms, punctuation marks and number of affective words. The technique adopted was unsupervised based on k-means clustering algorithm and nearest neighbor algorithm. In their research, experiments were performed in both determining subjectivity of text and predicting emotions.

[4] Jennifer et al (2014) presented affective text based mining of social emotions deal with new aspect for categorizing the document based on the emotions such as victory, sympathy and love etc. To predict the emotion contained in content a joint emotion-topic model was proposed. It first generates a set of latent topics from emotions followed by generating effective terms from each topic. Then generates an emotion from a document specific emotional distribution and then generates a latent topic from a multinomial distribution conditioned on emotions. The proposed model utilizes both emotion term model and topic model. Emotion topic model allows associating the terms and emotions via topics. For each emotion, a meaningful latent topic can be generated based on emotions.

[5] Shelke (2014) had described different approaches of emotion detection from text. Emotions can be extracted from text (tweets/blogs) using keyword based method, vector space model and point wise mutual exclusion. He concluded that the opinions are important to make decisions for practical applications such as product, purchasing, business and organization. Keyword selection, sentiments in domain specific, multiple opinions in a sentence, Negation handling, Sarcasm, comparative sentences and opinion spam are the major issues raised during emotion detection from text available on social media.

[6] Murugaboopathy et al. (2013) described an investigation of authorship gender distribution mining from e-mail text documents. The set of topic, context free email document features such as structural characteristics, style makers and gender preferential language features are used. In this paper gender linked features are divided into five subsets named as character based features, word based features, syntactic features, and structural features and function features. Decision Tree and SVM algorithm is used for data classification. Experiments are performed on a corpus of e-mail documents generated from a large number of authors of both genders performed for author gender categorization.

[7] K and Devi (2012) proposed a detailed view of extraction of opinions from knowledge nuggets. This provides to capture emotions from web documents and unstructured text documents. This process refers to extracting non-trivial previously unknown patterns from unstructured text. The huge amount of online resources available on web and famous and rich resources lead to the requirement of extracting the emotion from human author text document from online review sites and personal blogs. This opinion and views makes sentiment analysis has become an interesting area of research human perception and user views has greater potential for the knowledge, discovery and decision-based system. The data sets representing a particular corpus were classified using classification algorithms such as NB or SVM.

[8] Patil et al (2012) had classified emotions from text using super vector machine approach. They presented a novel method to classify news sentences into 6 categories of emotions are anger, disgust, fear, happiness, sadness and surprise. The corpus consists of 1000 news sentences WordNet affect dictionary was used to check the presence of word and classification of news sentences. The dataset of affective text was used for training and testing. They proposed SVM classification using quadratic programming for automatic emotion classification and analysis.

[9] Balabantaray et al (2012) presented new approach to classify the emotions based on the tweets tweeted on tweeter. They extracted 1000 random tweets of different users from the tweeter. they prepared an automated corpus for automatic emotion analysis. Syntactic, semantic and contextual based approaches are used for automatic emotion classification. Emotions were classified into 6 basic classes i.e. positive, negative, fear, joy, surprise, hate and disgust. SVM was used for emotion classification.

[10] Parkins (2012) had checked the emotional expression of Men and Women in Social Networking Sites. Total 50 participants had taken part in this study. In which 25 of the participants were Males and the other 25 were Females. The participants age was between 17-25. Through examining the participants Facebook and twitter accounts, It was observed that there are 6 methods that can be used to express emotions are use of Emotions, Full-stops, Capitalized text, different means of expressing Happiness and Same letter within a word to express emotions in sentence.

[11] Zhang et al. (2011) explored the increasing importance of the internet in people daily lives and use of internet between females and males. In paper they have used feature based online media text classification technique to investigate the online gender differences between females and males participates in web-forums by examine their writing styles and topic of interest. They have built five different feature sets by adding content specific features to content free features and conducting feature set selection. There experimental study was based on Islamic women, political forum, and the features set combing both content specific and content free features significantly better than those consisting of only content free features.

[12] Mazman et al (2011) had considered facebook to determine individual's usage purposes of social networks with a focus on possible difference between females and males. Their study consisted of 870 facebook users who responded online survey designated by researchers. Their analysis of the results showed that usage purposes can be categorized under four categories named as maintaining existing relationship, making new relationship using for academic purpose and following specific agenda. Significant differences were found between genders in all of the purposes mentioned while the differences on making new contacts was in favor of males, the differences on the other user were in the favor of females.

[13] Thelwall et al (2010) had mined mining emotions present in Myspace comments. They used both data mining and content analysis technique for mining emotions, exploring age and gender. A random sample of 819 public comments to or from US users was classified for strength of positive and negative emotions. They concluded that Female are likely to give and receive more positive comments than males and Females are more successful social network site users partly because of their greater ability to textually harness positive effect.

[14] Bollen et al (2010) had performed sentiment analysis of all tweets published on the tweeter. They used psychometric instrument as POSM to retrieve the moods states. POMS measures six individual dimensions of mood namely Depression, Fatigue, Vigor, Anger and Confusion. They computed six dimensional mood vectors for each day in

timeline. A corpus consists of 9,604,952 tweets. Each individual tweets from the corpus is normalized and passed before processing to remove commas, dashes, stop words etc.

[15] Yang and Lee (2009) proposed a machine learning technique for mining lyrics from allmusic.com. They developed a model that covers 23 specific emotions. This model is used to classify 168 different music moods used to describe thousands of songs available online at allmusic.com. A training set of 1032 songs was chosen from the site and lyrics from these songs were located from popular music lyrics websites. Each song's lyrics were transformed into a feature vector of 182 psychological features using a content analysis. Decision tree was used to understand classification models.

IV. CONCLUSION

In this paper, we have presented a survey of all the techniques that has been used to extract the text (comments/tweets) from the social networking sites to identify the gender of the online users of the social media. The papers employed in this study describe the importance of emotions expressed in social media in different domains. Various phases of emotion mining such as Emotion Extraction, Emotion Analysis, Emotion Detection and Emotion Classification have been presented with their methods. Still the optimized framework to evaluate the text to predict the emotions on the basis of exact behavior of the gender has to be set up.

REFERENCES

- [1] Ahmad, S., Adnan, A., Khan, G. & Mahmood, N., "Emotion, age and gender based cognitive skill calculations", *International Journal of Computer Theory and Engineering*, vol.7, no.1, 2015.
- [2] Patel, P. & Mistry, K. , "A Review: Text classification on social media data", *IOSR Journal of Computer Engineering (IOSR-JCR)*, vol.17, pp.80-84, 2015.
- [3] Dhawan, S., Singh, K. & Sehrawat, D. , "Emotion detection techniques in social networking sites", *International Journal of Information and Computational Technology*, vol.4, no.12, pp.1145-1153, 2014.
- [4] Jennifer, D. & G, S. , "Affective text based emotion mining in social media", *International Journal of Advance Research in Computer Science and Management Studies (IJARCSMS)*, vol.2, 2014.
- [5] Shelke, M., "Approaches of emotion detection from text", *International Journal of Computer Science and Information Technology*, vol.2, pp.123-128, 2014.
- [6] Murugaboopathy, G., Hariharasitaraman, S., Sankarram, N. & Babu, R. , "Appropriate gender identification from text", *International Journal of Research in Management and Technology*, vol.2, pp.1660-1664, 2013.
- [7] K, A. and Devi, T., "Study on mining emotions from literature: a method using computer to extract opinionated knowledge nuggets", *International Journal of Emerging Technology and Advanced Engineering (IJETA)*, vol.2, 2012.
- [8] Patil, S. & Chaudhari, A., "Classification of emotions from text using SVM based opinion mining", *International Journal of Computer Engineering and Technology (IJCET)*, vol.3, pp. 330-338, 2012.
- [9] Balabantaray, R C., Mohammad, M. & Sharma, N., , "Multiclass twitter emotion classification: A new approach", *International Journal of Applied Information Systems (IJASIS)*, vol.4, no.1, 2012.
- [10] Parkins, R. (2012), "Gender and Emotional expressiveness: an analysis of prosody feature in emotional expression", *Griffin working papers in pragmatics and Intercultural Communication*, vol.5, no., pp.46-54.
- [11] Zhang, Y., Dung, Y. & Chen, H. , "Gender classification for web forums", *IEEE transactions on system, man and cybernetics*, vol.4, no.4, 2011.
- [12] Mazman, G. & Kocak Usluel, Y. , "Gender differences in using social networks", *The urkish Online Journal of Educational Technology (TOJET)*, vol.10, 2011.
- [13] Thelwall, M., Wilkinson, D. & Uppal, U. , "Data mining in social network communication: Gender differences in MySpace", *Journal of the American Society for Information science and technology*, vol.61, no.1, pp.190-199, 2010.
- [14] Bollen, J., Mao, H. & Pepe, A. , "Modeling public mood and emotion: Twitter sentiment and Socio-Economic phenomena", *Proceedings of the 5th International AAAI conference on weblogs and social media*, vol.11, no.4, pp. 441-456, 2010.
- [15] Yang, D. & Lee, W., "Music Emotion identification from Lyrics", *IEEE International symposium on Multimedia*, 2009.