A Framework for Polarity Classification and Emotion Mining from Text

[1]*Sanjeev Dhawan*, [2]*Kulvinder Singh*, [3]*Vandana Khanchi*

[1,2]Faculty of Computer Science & Engineering, University Institute of Engineering and Technology,

Kurukshetra University, Kurukshetra-136119, Haryana, India.

[3]M.Tech. (Software Engineering) Research Scholar, University Institute of Engineering and Technology, Kurukshetra University,

Kurukshetra-136119, Haryana, India.

E-mail (s): [1]rsdhawan@rediffmail.com, [2]kshanda@rediffmail.com, [3]vandanakhanchi777@gmail.com

Abstract- Nowadays Online Social Networks are so popular that they are become a major component of an individual's social interaction. They are also emotionally-rich environments where users share their emotions, feelings, ideas and thoughts. In this paper, a novel framework is proposed for characterizing emotional interactions in social networks. The aim is to extract the emotional content of texts in online social networks. The interest is in to determine whether the text is an expression of the writer's emotions or not if yes then what type of emotion likes happy, sad, angry, disgust, fear, surprise. For this purpose, text mining techniques are performed on comments/messages from a social network. The framework provides a model for data collection, feature generation, data preprocessing and data mining steps. In general, the paper presents a new perspective for studying emotions' expression in online social networks. The technique adopted is unsupervised; it mainly uses the k-means clustering algorithm and nearest neighbor algorithm. Experiments show high accuracy for the model in both determining subjectivity of texts and predicting emotions.

Keywords - Online Social Networks, emotion mining, text mining, emotions, sentiment database.

Introduction

Today social networking websites are new medium for associating people related to different communities. Social networks enable users to interact with people exhibiting different social and moral values. The websites provide a very powerful medium for communication among individuals that leads to mutual learning and sharing of valuable knowledge. The social networking websites for example Facebook, LinkedIn, Twitter and MySpace where people can communicate with each other by including different communities and discussion groups. The potential growth in popularity of online social networks has significantly affected the medium people interact with friends and acquaintances or non-familiar nowadays. In fact, interacting through online social networking sites like Facebook and online chatting systems has become a major part of a person's life. Social Web-based applications, such as online social networking websites provides opportunities

to set up interaction among people leading to mutual learning and sharing of valuable information, such as comments, messages, chat, and discussion boards. Friendships and social relationships can be inferred and observed on these sites as they reflect the continuous interaction among the subjects and at last, are evidently become emotionally rich environments. In this paper, we are mining emotions from texts shared in online social networks in the form of messages and comments. The purpose is not only to determine specific emotions but also to show if the text contains emotions or not; in other words, if the text/message is subjective or moderate subjective reflecting the writer's affect and emotional state or if it is objective where the writer does not show any emotional feelings. An application of this approach is to predict relationship strength messages based on the affective content of messages. In this case, the main interest is whether the messages communicate emotions in the text or not. The identification of emotions in the text is greatly influenced by

the strength of messages that can be subjective, moderate subjective or objective. Thus this technique can be used to separate the messages based on subjectivity then after that based on emotions category on online social networks. Some part of texts in these environments is the lack of sentence structure and the use of informal language particular to these settings and different from the formal written language. These factors should be taken into account when performing any kind of text mining. This paper uses sentiWordNet and sentiSense database as a case study, and performs an unsupervised technique to categorize text messages based on subjectivity and after that find out what specific kind of emotions or which emotion category these belongs. The paper will be divided as follows: section I summarizes the introduction related to social network mining. Section II presents the literature review about emotion mining from texts. Section III includes the proposed framework. The evaluation of the model is the subject matter of section IV and the conclusion is included in Section V.

Literature Review

Emotion is that which provides communication life. A conversation between emotionally involvements of partners is lively and bright; but that association without feeling is fatal dull [1]. Emotions act as a sensitive catalyst that make lively interactions among human beings and helpful in the regulation and development of interpersonal relationships. The expression of emotions make a fine social interactions by providing observers a rich channel of information regarding the conversation partner [2] or his social intentions or feelings [3], by imposing positive or negative responses in others [4], and through predicting other's social behavior. Keltner et al. [5] expressed that "facial signs are more than just markers of internal states or feelings," that also act as a unique functions in a social environment. By determining the functional role of emotions, Frijda [6, 7] argued that emotions preserve and enhance, improve life and Lutz [8] has strong effect on their communicative, social culture and moral purposes. Social networks always have of some positive and negative feature. The positive aspects of social network are such as easily accessible, globally available, faster communication over world cost efficiently and the negative aspects of social network like promoting fake identities, lack of privacy etc. [9]. Micro-blog is a rapidly increasing multimedia mini-big. Emotion evaluation has an important role in the study of readers' feedback and readers' comments and authors' emotions for a particular affair and it has become a considerable promoting effect on the support of emotional and psychological evaluation mechanism based on the World Wide Web [10]. E-learning has growingly much interest in corporations, educational institutions and individuals alike. It commonly means to teaching efforts waved through the use of computers in an effort to differently knowledge in a non traditional classroom environment. Recently research have expressed that emotion can affect the e-learning experience. In general, emotion means how one feels. The positive feeling have a productive effect on the individual. However, a feeling of negative nature has impact negatively on the individual's learning experience [11]. The problem of social affective

text mining that aims to discover the connections between social emotions and affective terms based on user generated emotion labels. It first generated a set of hidden topics from emotions, followed by generating affective terms from each topic [12]. More psychologists and e-Learning system researchers and developers reached an opinion that emotion plays an important role in cognitive activities, especial in learning [13]. Today Computers are still quite "emotionally challenged."They neither identify the user's emotions nor possess emotions of their own. Mean while, there are several powerful data mining software which can attain good classification results. Recently WEKA might be the most popular data mining tools while GeneXproTools 4.0 is considered as "the most powerful data analysis tool in the market" by its company. Emotions are mental states created through physiological changes. Paul Ekman classified emotions into six basic categories: happiness, anger, sadness, fear, disgust and surprise [14]. Emotions can be identified based on two scales: the valence of the emotion that determine whether feeling is positive or negative and the arousal level presents the energy level associated with the emotion. The lwall et al. performed emotion mining from texts fetched from the online social network site MySpace. They showed that studying emotions or feelings based on a 2-dimensional scale (i.e. arousal and valence and) is more reliable and gives more precise results than studying emotions on a finer grain rather than based of emotion category. Even though emotions are universal, there are huge differences between social cultures and between individuals how and the extent in which these emotions or feelings are expressed. Furthermore, personality and social factors have also an effect on emotions where the expression of emotions is not bounded to a individual internal feelings but influenced by the society, personality, a person's previous experience and strategic goals [15]. Agarwal et al. stated [16] that sentiment analysis is a much-researched area that used with recognition of negative, positive and neutral sentiment in text. [17] In this paper Esuli et al. described SentiWordNet, a lexical resource produced by asking an automated classifier to relate to each synset of WordNet (version 2.0) a triplet of scores Positive, Negative, Objective describing how strongly the terms contained each of the three properties. The method used to develop SentiWordNet is based on the quantitative analysis of the glosses of synsets, and used the resulting vectorial term representations for semi-supervised synset classification. [18] The identification of emotional connotations in texts is a current task in computational linguistics. It is economical for many tasks for example, a company analyzing the blogosphere, people's opinion on its products. [19] This paper presented a new way for evaluating the affective qualities of natural language and a scenario/context for its use. Earlier approaches to textual affect sensing have used keyword spot-ting, statistical methods, lexical affinity and hand-crafted models. [20] The paper about Emotion fetching from text based on Affect Analysis Model. There were five levels of the Affect Analysis Model: Symbolic Cue Analysis (Analysis of the sentence for emotions, abbreviations, interjections etc.), Syntactical Structure Analysis (Analysis of syntactical structure of the sentence.), Word-Level Analysis (For each word found in our database, the affective characteristics of a word were represented as a vector of emotional state.), Phrase-Level Analysis (The aim of this stage is to detect emotions involved in phrases, and then in Subject, Verb, or Object form.), Sentence-Level Analysis

(The emotional vector of a simple sentence (or a clause) is produced from Subject, Verb, and Object formation vectors resulting from phrase level analysis.). [21]Web mining techniques can be used for real on-line social networking websites, such as on-line comments and blogs. This paper critically analyzing the users' behavior on social networks. Ishizuka *et al.* presented that [22] the 3D virtual world of "Second Life" copies a form of real life by enabling a space for social events and high interactions. Second life motivates people to create or strengthen interpersonal relations, to communicate ideas, to get new experiences, and to feel effectual emotions accompanying all adventures of virtual reality. Aman *et al.* [23] described the problem of classifying documents not only by topic, but also by overall sentiment, e.g. identifying whether a review is positive or

negative. Using movie reviews as data, they found that standard machine learning techniques certainly outperform human-produced baselines. However, the three machine learning methods employed (maximum entropy classification, Naive bayes and support vector machines) did not performed as well on sentiment classification as on earlier topic-based categorization. They concluded this result by examining factors that made the sentiment classification problem more challenging. [24]Emotion research has recently been attracted increasingly toward attention of the NLP community. They discussed the methodology and emotion annotation task results. The goal was to investigate the expression of emotion in language using a corpus annotation.
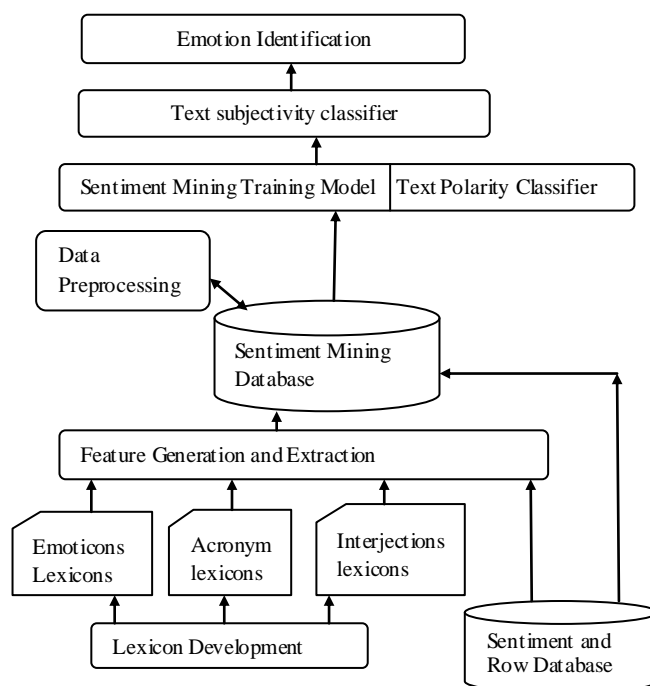
Proposed work

Figure 1: Framework for emotion mining

The framework shown in Figure 1 is framed into seven steps:

1. Raw data collection: This step is related to collecting exchanged texts between users. It is prepared through gathering data from the social network manually and stores them in a custom database.

2. Lexicons development: This step concerns with the informal language of online social networking sites. In this step, three types of lexicons have to be developed: lexicons for social acronyms, for emoticons, for interjections. The following table presents some of the popular acronyms:

Table 1: Social Acronyms

Acronyms	Description
FB	facebook
GR8	great
CU	See you

Similarly, an emoticons and interjections lexicon has been developed. The lexicons are not universally available but they cover a large percentage of these lexicons that are globally used

in online social networks. Tables 2 and 3 show examples of emoticons and interjections respectively.

Table 2: Emoticons	
Emoticons	**Description**
:-)	smile
:-(	sad
>:o	anger

Table 3: Interjections
Interjections
Haha
Hmm
Wow

3. Feature generation: This step computes new features from presented raw data gathered in step 1 to assess polarity or subjectivity of text. It uses word-matches with existing affective lexicons means from sentiment database and provides new lexicons developed in step 2 to handle social acronyms, emoticons and interjections from data. Accordingly, text messages collected from step 1 together with the features computed in this step will be stored in the Sentiment Mining Database that is the database used for analysis for emotion mining.

4. Data preprocessing: This step is applied to extract desired features according to which we are going to find out polarity of messages.

Table 4: List of Attributes

Features
Number of Interjections
Number of Emoticons
Number of social Acronyms
Number of repeated characters
Number of Punctuations marks
Number of affective words
Average rating measure of affective words

5. Creating a training model for text polarity: This step creates a model by k-means clustering algorithm with k=3 to categorize text messages into three polarity levels: subjective, moderate subjective and objective. The result of the model is the three centroids of the clusters.

6. Text subjectivity classification: In this step we uses the centroids computed in the previous steps and applies the k-nearest neighbor algorithm with k=1 to classify all messages into one of three subjectivity levels.

7. Emotions Identification: This step creates an emotion determining training model and then applies it to identify the emotions on the subjectivity of the texts messages. The identification is done by first classifying the subjectivity of the text messages exchanged that is performed in step 6 and after that find out the emotions of messages.

Table 5: Messages emotions

Messages id	Messages	Emotions
1	Carooooooooo im going to kiiiilllll uuuuuuuuuuu…..n know why! But I still like u (a little bit :p) don't worry….	Like, calmness
2	I love ur profile pic its much better like this: best CU	Like

MODEL EVALUATION

This section includes the results and describes the findings of the proposed method. The training data consisted of few messages. The evaluation of the model was performed at three different levels. First, we tested how precise the clustering

.

algorithm was in determining three different texts classes regarding text subjectivity. Second, the accuracy of the model in

determining, regarding the subjectivity of the text and third emotions identification from messages

Figure 2: Model evaluation graph

data/class Accuracy: 80.00
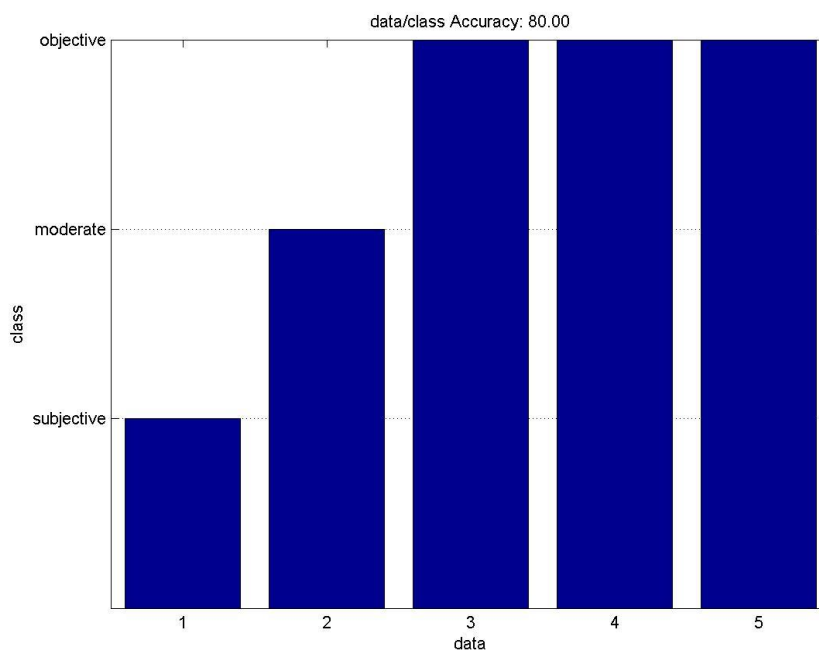
Conclusion

This paper describes a new sentiment mining technique from texts in online social networks. It performs a new perspective for studying expression of emotions in online social networks. The purpose was to identify whether the text messages conveys emotions. The processed data was first used to identify text polarity to determine the subjectivity of the texts. We developed new set of lexicons which cover common expressions used in messages, including interjections, emoticons, social acronyms, etc. After polarity classification using K-means clustering algorithm, messages emotions are identified from a set of 14 emotional categories from SentiSense database. Experiments demonstrated on the data showed high efficiency of the proposed method.

References

[1]   S. Planalp, Communicating Emotion: Social, Moral, and Cultural Processes, Cambridge University Press, Cambridge, UK, 1999.

[2]   P. Ekman, "Facial expression and emotion," American Psychologist, vol. 48, no. 4, pp. 384–392, 1993.

[3]   A. J. Fridlund, "The behavioral ecology and sociality of human faces," Review of Personality and Social Psychology, vol. 13, pp. 90–121, 1992.

[4]   U. Dimberg and A. Ohman, "Behold the wrath: psycho physiological responses to facial stimuli," Motivation and Emotion, vol. 20, no. 2, pp. 149–182, 1996.

[5]   D. Keltner, P. Ekman, G. C. Gonzaga, and J. Beer, "Facial expression of emotion," in Handbook of Affective Science, R. J. Davidson, K. R. Scherer, and H. H. Goldsmith, Eds., pp. 415–432, Oxford University Press, New York, NY, USA, 2003.

[6]   N. Frijda, "The Emotions, Studies in Emotion and Social Interaction", Cambridge University Press, Cambridge, UK, 1986.

[7]   N. Frijda, "Emotions are functional, most of the time," in The Nature of Emotion: Fundamental Questions, P. Ekman and R. J. Davidson, Eds., Oxford University Press, New York, NY, USA, 1994.

[8]   C. Lutz, Unnatural Emotions, University of Chicago Press, Chicago, III, USA, 1988.

[9]   Sanjeev Dhawan, Kulvinder Singh, Vandana Khanchi, "Review of Social Networks and On-Line Web Communities", pp. 191-196, International Journal of Computer Application, Issue 4, Volume 3 (May-June 2014), Available online on http://www.rspublication.com/ijca/ijca_index.htm, ISSN: 2250-1797.

[10]   Yang shen, Shuchen li, Ling zheng ,Xiaodong ren, Xiaolong cheng, "Emotion Mining Research on Micro-blog", 1st IEEE Symposium on web society,

pp. 71-75, 23-24 Aug., 1st IEEE Symposium on web society, Lanzhou, 2009.

[11] Haji H. BINALI, Chen WU, Vidyasagar POTDAR, "A New Significant Area: Emotion Detection in E-learning Using Opinion Mining Techniques", in Proceedings of 3rd IEEE International Conference on Digital Ecosystems and Technologies, 1-3 June, pp. 259-264, Istanbul, 2009.

[12] Shenghua Bao, Shengliang Xu, Li Zhang, Rong Yan, Zhong Su, Dingyi Han, Yong Yu, "Joint Emotion-Topic Modeling for Social Affective Text Mining", pp. 699-704, 6-9 Dec., in Proceedings of Ninth IEEE International Conference on Data Mining, Miami, FL, 2009.

[13] Feng Tian, Qinghua Zheng, Deli Zheng, "Mining Patterns of e-Learner Emotion Communication in Turn Level of Chinese Interactive Texts: Experiments and Findings", pp. 664-670, 14-16 April, in Proceedings of the 14th International Conference on Computer Supported Cooperative Work in Design, Shanghai, China, 2010.

[14] P. Ekman, "An argument for basic emotions," Cognition & Emotion, Vol. 6, No. 3, pp. 169-200, 1992.

[15] M. Thelwall, D. Wilkinson, S. Uppal, "Data mining emotion in social network communication: Gender differences in MySpace", In Journal of the American Society for Information Science and Technology, pp. 190-199, 2010.

[16] Apoorv Agarwal, Fadi Biadsy, Kathleen R. Mckeown, "Contextual Phrase-Level Polarity Analysis using Lexical Affect Scoring and Syntactic N-grams", in Proceedings of the 12th Conference of the European Chapter of the ACL, pages 24–32, Athens, Greece, 30 March – 3 April 2009. 2009 Association for Computational Linguistics

[17] A. Esuli, F. Sebastiani, "SentiWordNet: a publicly available lexical resource for opinion mining," in Proceedings of the 5th International Conference on Language Resources and Evaluation, pp. 417–422, Genoa, Italy, May 2006.

[18] F. Chaumartin. Upar7: A knowledge-based system for headline sentiment tagging. In Proceedings of SemEval-2007, pp. 422-425, Prague, Czech Republic, June 2007.

[19] H. Liu, H. Lieberman, and T. Selker, "A Model of Textual Affect Sensing Using Real-World Knowledge," in Proceedings of Eighth International Conference on Intelligent User Interfaces, pp.369-374, 2003.

[20] M. Corney, O. de Vel, A. Anderson, and G. Mohay, "Gender preferential text mining of e-mail discourse," in Proceedings of 18th Annual Computer Security Applications Conference, Vol. 13, pp. 21-27, 2002.

[21] Sanjeev Dhawan, Kulvinder Singh, Vandana Khanchi, "Critical Analysis of Social Networks with Web Data Mining", pp. 107-111, 3-4 May, in Proceedings of 2nd International Conference on Futuristic Trends in Engineering & Management 2014 (ICFTEM-2014) & in online Journal International Journal of IT and Knowledge Management (ISSN : 0973-4414), Bilaspur, Haryana.

[22] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "EmoHeart: Conveying Emotions in Second Life Based on Affect Sensing from Text", Advances in Human-Computer Interaction, vol. 2010, Article ID 209801, 13 pages, 2010.

[23] S. Aman, S. Szpakowicz, "Identifying Expressions of Emotion in Text", In Proceedings of 10th International Conference on Text, Speech and Dialogue, Lecture Notes in Computer Science 4629, pp. 196-205, 2007.

[24] B. Pang, L. Lee, S. Vaithyanathan, "Thumbs up? Sentiment Classification Using Machine Learning Techniques", In Proceedings of 2002 Conference on Empirical Methods in Natural Language Processing, pp.79-86, 2002.