EMOTION GENERATION AND SUMMARIZATION FORM AFFECTIVE TEXT

[1]SAYALEE SANDEEP RAUT, [2]KAVITA PANKAJ SHIRSAT

[1]Student, Vidyalankar Institute of Technology, Mumbai
[2]Assistant Professor, Vidyalankar Institute of Technology, Mumbai
E-mail: [1]sayalee.raut@vit.edu.in, [2]kavita.shirsat@vit.edu.in

Abstract— Mining social emotion from text deals with new aspect for categorizing the document based on the emotions such as victory, love, anger etc. In order to predict the emotion contained in content a joint emotion-topic model is proposed by enhancing Latent Dirichlet Allocation with an additional layer for emotion modeling. Using this it first generates a latent topic from emotions, followed by generating perceptual terms from each topic. First it generates an emotion from a document-specific emotional distribution, and then it generates a latent topic from a multinominal distribution conditioned on emotions. The model which we proposed will utilize the complementary advantages of both emotion-term model and topic model and also it include more websites for creating a large vocabulary. Emotion-topic model allows associating the terms i.e. words and emotions via topics which is more flexible. Also it has better modeling capability.
For each emotion, it generates a meaningful latent topic and also based on emotions, songs recommendation will be available for user. So that user can upload and enjoy their own choice of song.

Keywords— Affective Text Mining, Emotional-Term Model.

I. INTRODUCTION

Mining frequent patterns is perhaps one of the most important concepts in data mining. From this concept a lot of other data mining tasks and theories stem. It should be the beginning of any data mining technical training because, on one hand, it gives a very clear cut idea about what data mining is and, on the other, it is not extremely technical. Affective text based mining allows us to conclude a number of conditional probabilities for unseen documents, e.g., the probabilities of latent topics given an emotion, and that of terms given a topic.

There are different methods used to deal with the affective text mining those are Emotion-Term model topic, based-SVM model, term- based SVM model, and LDA model and so on.LDA model can only discover the topics (words) from document but it cannot bridge the connection between social emotions and affective text. Previous work mainly focuses on titles information, so the efficiency of these models is varying [1]. Emotion-term model treats terms individually and cannot discover the contextual information within the document. Emotion-term model unable utilize the term co-occurrence information within document and cannot distinguish the general terms from the affective terms [5]. Traditional topic model can only discover the latest topics from the document set but it cannot able to bridge the connection between social emotions and affective text.

II. EXISTING APPROACHES FOR EMOTION GENERATION

Mainly the all this work is focuses on affective text mining and topic modeling.

A. Affective Text Mining:

Large work has been done in past on affective content mining but not on connection between affective terms and emotion.To explore the connection between affective terms and social emotions, a task named SentiWordnet is considered. “A Publicly Available Lexical Resource for Opinion Mining” it basically involves Opinion about any product or political candidate from that social website user [3]. Mainly it collects information about Opinion from text and that will be based on Text SO polarity which is nothing but subjective and Objective polarity and also Text PN polarity for positive and negative categorization of a opinion, basically from that Strength of PN polarity will used for calculating score assignment to each opinion with values 0.0 to 1.0, weight age for opinion is being calculated from that total average value [3]. Existing approaches will not consider relationship across word. In previous work the emotions and terms were not linked and there will be only minimum likelihood of estimation of emotions then with the help of proposed model, now we are able to visualize the emotion assignments at the term level.

B. Topic Based Modeling:

A topic model is a type of statistical model for discovering the abstract "topics" that occur in a collection of documents.A topic model will capture this intuition in a mathematical framework, which allows examining a set of documents and discovering, based on the statistics of the words in each, what the topics might be and what each document's balance of topics is.

i. LDA: Latent Dirichlet allocation (LDA) is a generative statistical model that allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar.

ii. PLSI: Probabilistic latent semantic analysis (PLSA), also known as probabilistic latent semantic indexing (PLSI, especially in information retrieval circles) is a statistical technique for the analysis of two-mode and co-occurrence data.

In LDA, each document may be viewed as a mixture of various topics. This is similar to probabilistic latent semantic analysis (pLSA), except that in LDA the topic distribution is assumed to have a Dirichlet prior. In practice, this results in more reasonable mixtures of topics in a document. It has been noted, however, that the pLSA model is equivalent to the LDA model under a uniform Dirichlet prior distribution

LDA has been enlarging to more advanced application domains with additional sampling steps. There will be several techniques available to predict topics but the main difference lies in different sampling distributions [4]. Their author variable is chosen uniformly from a set of authors while emotion variable is sampled from multinomial distributions by the emotions contributed by web users. LDA is extended with a different set of information, i.e. social emotions which are contributed by online users, in the latent topics modeling process [4]. So to achieve our aim, proposed system present two baseline models one is Emotion Term model while other is LDA topic model and we are going to use combination of both i.e nothing but Emotion-Topic model.

- Emotion-Term model uses naïve byes to model affective terms and social emotion via their co occurrences.
- LDA Topic model utilizes terms co occurrences information within a document and discover the fundamental topics within affective text.
- Emotion-Topic model can jointly estimate the latent document topics and emotion distribution in a unified probabilistic graphical model.

III. PROPOSED SYSTEM

Explanation

I. Training Phase:

Here we will be following supervised learning algorithm to train the data set. Emotion data set containing emotion and sentences will be our input training data set in form text files. We have kept our scope limited to English language, Also we will be handling six emotion. Those are Warmness, Boredom, Amusement, Empathy, Touched and Surprise. We pre process our training data set. Following are steps that we follow for pre-processing:

- **Stop Word Removal:** Here we are removing extremely common words which would appear to be of little value in helping select documents matching a user need are excluded from the vocabulary entirely. These words are called stop words. We are maintaining a list of stop words for this.
- **Slang Words Identification and Replacement:** Slang words have the ability to interrupt and falsify Natural Language Processing tasks done on social media text. Now a day's user may use slang words, so we need to handle those with slang word identification and replacement.
- **Pos Tagging:** For Pre processing we will perform POS Tagging and Document Frequency Identification. In word separation it may use part-of-speech to separate words. Part-of-speech tagging (POS tagging or POST), also called grammatical is the process of marking up a word in a text as corresponding to a particular part of speech. A simplified form of this is commonly the identification of words as nouns, verbs, adjectives, adverbs, etc. It gives relevant information about the role of a word in its narrow context. We will be using noun for finalizing our emotion training data set.
- After preprocessing we will be treating that as bag of words for which we will be calculating probability with respective each emotion.
- After all this we calculate LDA sampler which will be having

II. Testing and Result Generation:

In testing phase user gives input data set.

For user entered dataset pre-processing is performed same as in training phase. Then pre-processed data is given to LDA model which generates probabilities for input dataset. This user entered input is compared with training dataset probability and we generate result.

As a result we are showing emotion with emoticon. Based on emotion we are playing song.

1. Application

- Emotion aware recommendation like advertisement and songs etc
- Interactive E-learning sessions with emotion regulation
- Analysis of blogs, comments, feedback, reviews and tweets based on emotion present in it.
- Predicting Personality factor with help of emotions

2. Results

The model which we used is probabilistic model so we have shown analysis in graphical format.

Graph (a) shows us emotion probability, while Graph (b) shows us emotion distribution over topics.

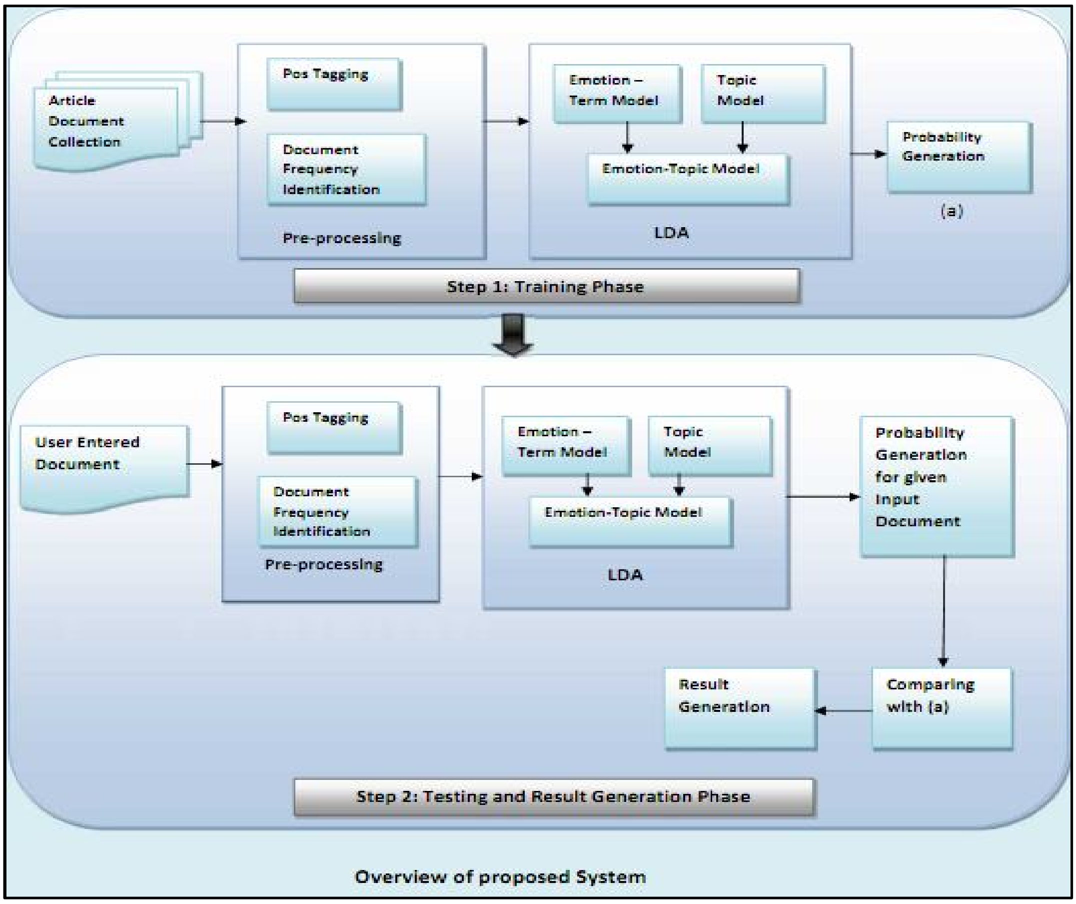

Fig.3.1 Overview of Proposed System

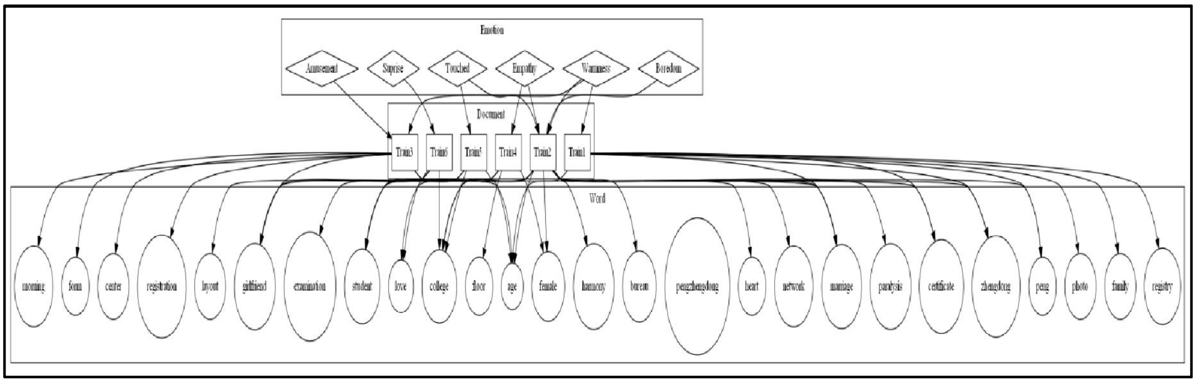

Fig.3.2 Relationship Diagram

Above image shows relationship between emotion, document and words.

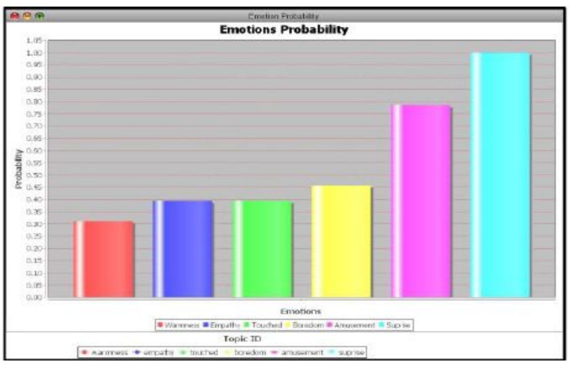

Fig.5.1Emotion Probability

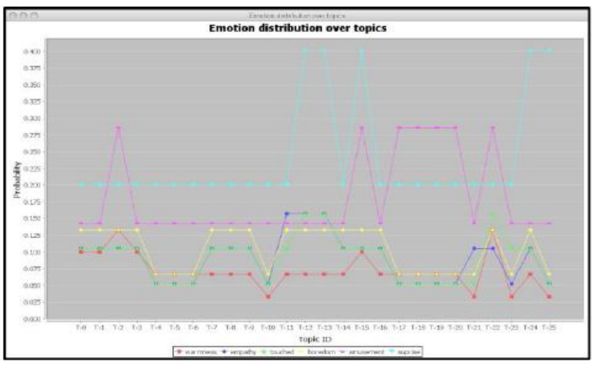

Fig.5.2 Distribution of Emotion over Topics

We observed that for LDA, as training data set increases it will take more time for processing. We deal with emotion detection but along with that we will be displaying result inform of images (emoticons) and music.

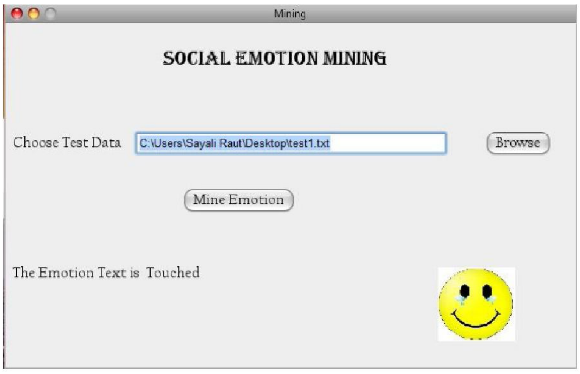

Fig.5.3 Result of Emotion Detection

CONCLUSION

We present and examine a new problem called mining emotion with affective help of affective terms, which aims to discover and model the connection between online documents and user-generated social emotions. [5] We proposed to determine our model with a larger scale of online document collections, and apply the model to other applications such as emotion-aware recommendation of advertisement, songs and categorizing online document based on emotion preference, As this is digital era we can see that everything on internet. Our proposed system plays an important role, as it is detecting emotion. We can able to provide best solution, high experience to our users based on reviews or recommend relevant information. Future Scope of this project is now we are dealing with only text. This may create ambiguity in finding emotion. If we combined it with voice (speech) then it will give us more accurate result. Also if we want more accuracy to detect emotion we can use combination of voice and facial expression. We can configure database to provide input so we can consider large data.

REFERENCES

[1] C.Yang, K.H.-Y. Lin, and H.-H. Chen, “Emotion Classification Using Web Blog Corpora,” Proc. IEEE/WIC/ACM Int’l Conf. Web Intelligence (WI ’07), pp. 275-278, 2007.

[2] R. Cai, C. Zhang, C. Wang, L. Zhang, and W.-Y. Ma., “Musicsense: Contextual Music Recommendation Using Emotional Allocation,” Proc. 15th Int’l Conf. Multimedia, pp. 553-556, 2007

[3] C. Strapparava and R. Mihalcea, “Learning to Identify Emotions in Text,” Proc. 23rd Ann. ACM Symp. Applied Computing (SAC ’08), pp. 1556-1560, 2008

[4] D.M.Blei,A.Y.Ng,and M.I.Jorden, ”Latent Dirichlet Allocation,”j.MachineLearning Research,Vol.3,pp.993-1022,2003

[5] Shenghua Bao,Shengliang Xu,Li Zhang, Rong Yan,Zhong Su, Dingyi Han, and Yong Yu “Mining Social Emotions from Affective Text” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 9, SEPTEMBER 2012

[6] A. Esuli and F. Sebastiani, “Sentiwordnet: A Pub-Licly Available Lexical Resource for Opinion Mining,” Proc. Fifth Int’l Conf. Language Resources and Evaluation (LREC ’06), 2006.

[7] T. Griffiths and M. Steyvers, “Finding Scientific Topics,” Proc. Nat’l Academy of Sciences USA, vol. 101, pp. 5228-5235, 2004.

[8] C. Strapparava and R. Mihalcea, “Learning to Identify Emotions in Text,” Proc. 23rd Ann. ACM Symp. Applied Computing (SAC ’08), pp. 1556-1560, 2008.

[9] TejasviniPatil and SachinPatil, “Automatic Generation of Emotions for Social Networking using Text Mining” IEEE 2013

[10] Mohammed Yassine and Hazem Hajj, “A Framework for Mining from Text in online Social Networks” IEEE 2010

[11] D.M.Blei,A.Y.Ng,and M.I.Jorden, ”Latent Dirichlet Allocation,” j.Machine Learning Research,Vol.3,pp.993-1022,2003

[12] Carlo Strapparava and Alessandro Valitutti,“WordNet Affect: an Affective Extension of WordNet.”, LREC, 2004 - hnk.ffzg.hr

[13] M Hu, B Liu,” Mining and summarizing customer reviews”, Proceedings of the tenth ACM SIGKDD international 2004

[14] Vairaprakash Gurusamy and Subbu Kannan, “Preprocessing Techniques for Text Mining”, Conference Paper • October 2014

[15] Stopwords.txt - Twitter sentiment analysis using machine learning techniques, Available at: https://code.google.com/p/twitter-sentiment-analysis/source/browse/trunk/files/stopwords.

★★★