International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.03, April 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Social Emotion Mining: A Survey

Deepa R. Gawande[1], Prof. K. P.Wagh[2] and Dr. P.N.Chatur[3]

[1] PG Scholar, [2]Assistant Professor, [3]Associate Professor

[1,3]Department of Computer Science and Engineering,

[2]Department of Information Technology,

[1,2,3] Government College of Engineering Amravati. (India)

ABSTRACT

Recent years with increasing richness of Web 2.0, people are embolden to have various social interactions on the websites. This Social websites has given new way of communication technology for people to share their opinions, interest and sentiments. Many online documents are assigned by social users with emotion labels such as happiness, sadness, and surprise etc. In history work they focused on author's perspective i.e. accurately model the connections between words and emotions. Recent research focuses on reader's emotions with news articles. The sentiment analysis of reader's is sometimes more meaningful in social media as compared to classical sentiment analysis from writers perspective. Due to this situation, reader's emotion prediction shows a promising research area. Many models have been proposed and well discussed in past work to deal with the predictions of social emotions. There are different methods used to deal with the problem of mining social emotions from affective text. In this paper we discuss the Social Opinion Mining Model which is based on the social opinion network. This model used for social opinion predictions. Compared with the previous model i.e. Emotion-Term model, Emotion Topic model, MSTM, SLTM, ATM model the SOM model gives better results. This paper focuses on providing an in-depth survey of different work done in Social Emotion Mining from reader's perspective and social opinion prediction. This paper reviews the analysis different models which used in Social Emotion mining and social opinion prediction.

Keywords— Reader's emotion, Social affective text mining,Social emotion mining, social opinion prediction, social opinion mining model.

I. INTRODUCTION

The world has basically changed as the Internet has become a universal means of communication. As this is digital era we can see that everything on internet. Social Media are at the heart of our communications and are among the most visited places on the Web. Day by day people show increasing eagerness to online communication due to free and convenient communication environment of internet. At the same time many internet users prefer to produce and convey online information through expressing personal opinions than just obtain online information. Usually, when we make a decision, opinions and emotions of others have always been important information for reference. It is necessary for general people, marketers, public relations officials, politicians and managers to cognize the answer of "What others people think and feel" [15].

Recently, with the paradigm shift in the usage of the Web from information consumption to information production and sharing numerous social media services have emerged. Users can express their opinions and emotions conveniently through news portals, blogs and micro blogs, where they become both the consumers and producers. There are different websites i.e. sina.com.cn and people.com.cn which provide a new service that allows users to share their emotions after browsing news articles. A lot of studies have been carried out to automatically predict the most probable emotions for documents. In prediction of emotion studies, the way of how text documents affect online user's social emotions is yet not covered. To over-come this we refer to the problem of discovering and mining connections between social emotions and online documents as social affective text mining. This include the prediction of emotions from online documents, associating emotions with latent topics, and so on.

The affective text based mining allows us to assume a number of conditional probabilities for unseen documents. In previous work there are different methods used to deal with the affective text mining and following process such as, Emotion-Term model, Emotion Topic model, MSTM, SLTM, ATM model and so on.

The task of automatically detecting public emotions evoked by online documents is become prominent due to the vast amount of data. This task is treated as a classification problem according to the polarity (positive, neutral or negative) or multiple emotion categories such as joy, sadness, anger, fear, disgust and surprise. However, the limited information, annotating news headlines for emotions is a hard task. Even for human being it is usually intractable to annotate headlines consistently. Reader emotion prediction is always treated as a text classification problem in which a piece of text are classified into the most likely triggered emotion category.

Previous works on social emotion detection have focused mainly on using the sentiments of individual words. Several studies have been carried out to automatically predict the most probable emotions for documents. In today's online world Emotion detection technology can find several applications. In era of text messaging, users are constantly texting and may send inappropriately angry messages to others. If emotion detection is implemented, in such cases, the application can take appropriate action such as popping up a warning to the user before sending a message. This paper gives a general overview of the different approaches of mining social emotions and detecting this emotions. Therefore Social emotion mining has attracted a large amount of attentions from researchers of natural language processing and machine learning. This paper highlights different models which used in social affective text mining.

II.FUNDAMENTALS

From the overall study we have seen that there are two basic fundamental concepts of mining social emotions. Which are affective text mining and social affective text mining followed by discussing their connections and differences with the different model.

2.1 Social Emotion Mining:

Understanding and predicting latent emotions of users toward online contents is known as social emotion mining. Due to rapid development of internet the Social Emotion Mining has became increasingly important to both social platforms and businesses alike. The Social Emotion Mining is one of the most popular method for mining users digital footprints to unearth users "emotions" toward particular products or services on social platforms. The Social Emotion Mining is widely used for mining emotions from reader's perspectives. It has become increasingly important for businesses to better understand their users and leverage the learned knowledge to their advantage. User's latent emotions can be indirectly peeked via various channels –e.g. Chinese news portal, Sina, where users may select one emoticon, out of many choices, to more precisely express their emotions.

2.2 Affective Text Mining:

Research into social emotion detection began with the "Affective Text" in SemEval-2007 tasks. Affective Text is a supervised system that annotates headlines using a predefined list of emotions. The Affective Text focused on exploiting reader emotions with individual words. Affective Text uses a corpus of headlines hand-annotated by non-experts. Affective text based mining of social emotion deals with new aspect for categorizing the document based on the emotions such as victory, sympathy, love etc. The affective text based mining allows us to assume a number of conditional probabilities for unseen documents. There are different methods used to deal with the affective text mining. Existing approaches will not consider relationship across word. So the emotions and terms were not linked in previous work. The limitation of early word-level studies was that different senses could be evoked from the same word.

2.3 Social Affective Text Mining

In data mining process mining of frequent patterns is probably one of the most important concepts. Many data mining tasks and theories tail from this concept. Here Social affective text mining aims to discover and model the connections between online documents and user-generated social emotions. The Main objective of social affective text mining is to accurately model the connections between words and emotions. This will improve the performance of its related tasks such as emotion prediction. Here we are comparing the extracted and optimized content with the already founded latent topics that relating to each emotion. Based on the result we are finding which emotion the particular content represents. Based on the user emotion request the categorized content will display. Two baseline models are used in social affective text mining which are emotion-term model and LDA topic model.

III.RELATED WORK

This section reviews some of the related work on Social Emotion Mining. The problem of Social Emotion Detection is mostly studied as a classification problem by past researchers. A lot of work have been carried out to prediction and mining of social emotions. Following are the different techniques which we used in prediction and mining of social emotions from reader's perspective.

C.Strapparava and R. Mihalcea [1] began research into social emotion detection with the "Affective Text" in SemEval-2007 tasks. Affective Text is a supervised system that annotates headlines using a predefined list of emotions. The Affective Text focused on exploiting reader emotions with individual words. Affective Text uses a corpus of headlines hand-annotated by non-experts.

C.Strapparava and Mihalcea [13] evaluated several knowledge-based and corpus-based methods for the automatic identification of six emotions. A straightforward method to model the word-emotion associations is called the emotion-term model. The Emotion-Term model follows the Naive Baye's method by assuming that words are independently generated from social emotion labels. The Naïve Bayes classifier is used to calculate the probability of each word with respect to each class.

S. Bao et al. [4] proposed a joint Emotion-Topic model (ET) by augmenting an intermediate layer into LDA, in which a topic acts as an important component of an emotion. It first generates a set of latent topics from emotions, followed by generating affective terms from each topic. Emotion-Topic model can jointly find the emotion from emotion term and topic model. The Emotion-Topic model allows us to assume a number of conditional probabilities for unseen documents.

Yanghui Rao et al. [3] proposed two sentiment topic models called Multi-label Supervised Topic Model (MSTM) and Sentiment Latent Topic Model (SLTM). The Multi-label Supervised Topic Model (MSTM) is a multi-labeled sentiment topic model which correlate latent topics with invoke emotions of readers. Sentiment Latent Topic Model (SLTM) is a multi-labeled sentiment topic model which discover meaningful latent topics which produce social emotions, where both explicit and implicit emotive words are identified. The performance of the SLTM and MSTM is more stable than the baseline ETM with varied topic numbers. In terms of the averaged accuracy, SLTM and MSTM outperform the baselines ETM, SWAT and ET.

Yanghui Rao, Qing Li, Liu Wenyin, Qingyuan Wu, Xiaojun Quand [6] proposed Affective Topic Model (ATM) is a multi-labeled topic model for social emotion detection. The Affective Topic Model is developed to differentiate between affective and background topics. This model is evolved to bridge the gap between social media materials and a reader's emotions by introducing an intermediate layer. The Affective topic model can be used to classify the social emotions of unlabeled documents and to generate a social emotion lexicon. The performance of the ATM is more stable than the baseline Emotion Topic Model when the number of topics used was varied.

Yanghui Rao [7] proposed a Contextual sentiment topic model (CSTM) is a context level topic model which is developed to classify reader's emotions across different contexts. Most works on social emotion detection, including ETM, ATM, MSTM, and SLTM, have focused mainly on associating emotions with topics specific to one context. The contextual sentiment topic model is developed to classify reader emotions across different contexts. We targeted the problem of adaptive social emotion classification and developed the contextual sentiment topic model to solve it.

Xintong Li, Qinke Peng, Zhi Sun, Ling Chai, and Ying Wang [8] presented a new model which is related with social opinion predictions called as Social Opinion Model (SOM). The SOM is constructed based on the social opinion network. The constructions of opinion network depends on the semantic distance. To predict reader's emotions the neighbor relationship in network is used. This methods obtain better results as compared to the previous methods. We will discuss this model in detail.

IV.MODEL DESCRIPTION

4.1 Social Opinion Model

The Social Opinion model is developed to social opinion prediction. The social opinion model is used for measuring similarity among news. Compared with other previous topic-level and word-level models, this model treats the news content and emotion distributing as a whole opinion structure. The SOM model finds the correlation between words and emotions. The performance of the prediction based on opinion network is more stable and accurate than existing models. In SOM for pruning the network we developed a threshold-based network growing strategy is used. Social Opinion Mining model outperforms others models i.e. Emotion term, ET, MSTM, SLTM, ATM, and CSTM.

With increasing use of internet the Social emotion prediction pay more attention to market analysis and political decision. From the NLP perspective, SOM models are inexplicable but feasible. From psychology and linguistics perspective, the models are explicable but lack of use in the service [8]. We can predict social opinions by measuring the semantic similarity between events. Based on semantic similarity the SOM is casting the relationship between current event and priori social opinions [8].

4.2 Model Architecture

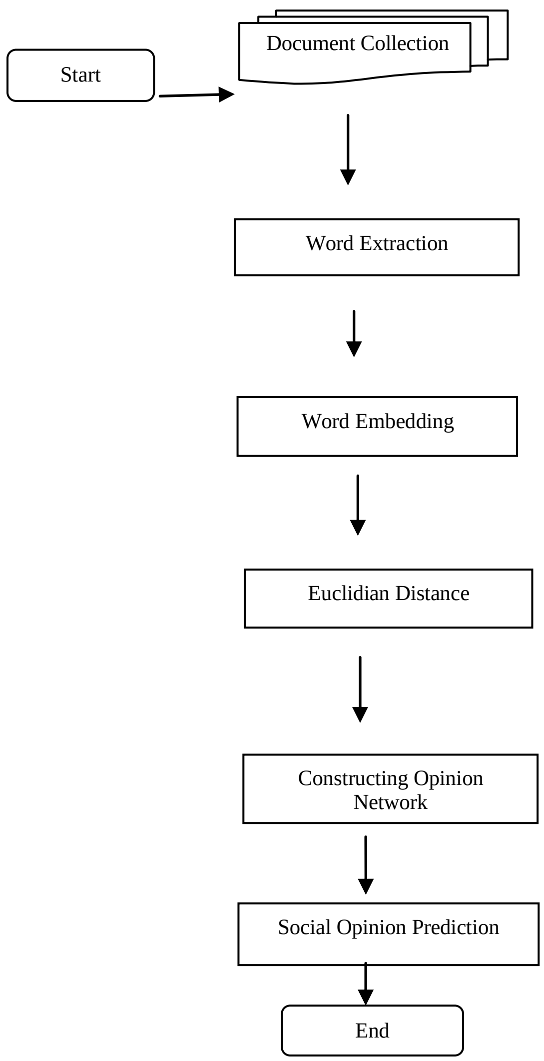

Fig. 1: Social Opinion Mining Model Architecture

The above Figure 1 shows the architecture of the social opinion mining model. The following are the steps of Social opinion mining model.

4.2.1 Word Extraction

The first step of social opinion model is word extraction from the news collections of documents. Each document is represented as a Bag of Words. Bag of words (BoW) is an algorithm that counts how many times a word appears in a document. Bag of Words is a simplifying representation used in a natural language processing and information retrieval. For measuring the importance of a word in corpus we utilize the term frequency (TF) instead of term frequency–inverse document frequency (TF-IDF). The term frequency measures the numbers of times that word occurs in documents.

4.2.2 Word Embedding

Word2vec is a group of related models that are used to produce word embedding. Word2vec can utilize either of two model architectures to produce a distributed representation of words i.e. continuous bag-of-words (CBOW) and continuous skip-gram. In the continuous bag-of-words architecture, the model predicts the current word from a window of surrounding context words. In the continuous skip-gram architecture, the model uses the current word to predict the surrounding window of context words. The skip-gram architecture weighs nearby context words more heavily than more distant context words where words or phrases from the vocabulary are mapped to vectors of real numbers. Word embedding turns text into numbers.

4.2.3 Euclidian Distance

In this step we calculate the semantic distance between two words by using Euclidian Distance formulae. The Euclidean distance between words in word2vec space measures the semantic similarity. More precisely, C (i, j) $=\| X_i - X_j \|$denotes the distance between word i and word j where X_i and X_j represent corresponding word.

4.2.4 Construction of Opinion Network

The SOM is constructed based on the social opinion network. We construct an opinion network based on the relation between opinions. Based on the semantic distance we construct the opinion network. In this network nodes indicate opinions and edges indicate relation between opinions. The opinion network can serve as the dictionary between events and corresponding emotions. We can predict the social opinion through the network. To construct the opinion network, we add the edge between nodes to denote distance. Then we analyse the relationship between network structure and social emotions. Since the opinion network here is fully connected networks, we filter the edge to visualize the network structure. Label the nodes of 8 emotions by color [8]. The network structure is different in thresholds. Here we choose the threshold for the visualization of the network. We proposed a threshold-based network growing strategy for pruning the network. The threshold is chosen as 0.7 manually for visualization of network. We utilize ForceAtlas2 [12] algorithm to arrange the layout of nodes.

4.2.5 Social Opinion Prediction

From the above opinion network we can predict the social opinion. There is the correlation between emotion and community in opinions network. The threshold for pruning edges depends on the current network state. It is unable to determine the appropriate threshold of filter for the community structure. To predict emotion conditioned on future unlabeled opinion which only contain words, we leverage the inference of community structure of opinions network and simplify it as neighbor analysis. So that, we can predict the social opinion based on the opinions growing network [8].

V.DATASETS

For testing the effectiveness of this proposed model, we utilize two datasets. (The dataset is available in public: github.com/lixintong1992/Social Emotion Data).

1. First dataset used here is Yanghui Rao's corpus [3] which collected 4570 news articles from the Society channel of Sina. The attributes of each article include the URL address, publishing date (from January to April of 2012), news title, content, and user ratings over 8 emotion labels: "touching", "empathy", "boredom", "anger", "amusement", "sadness", "surprise" and "warmness". This Dataset called as Dataset2012.

2. The other dataset is collected from the Society channel of Sina from January to December of 2016, a total of 5258 hot news data. User ratings over 6 emotion labels: "touching", "anger", "amusement", "sadness", "surprise" and "curiosity". This Dataset called as Dataset2016.

   Both of the above datasets obeys normal distribution. The expectation and variance of the normal distribution of Dataset2016 are greater than Dataset2012. From this we considered that the semantic correlation is relatively small. Dataset2016 has large time interval, less semantic correlation as compared to Dataset2012. We split the above datasets into training and testing sets by chronological order.

VI.PERFORMANCE ANALYSIS

We used two coarse-grained and fine-grained evaluation metrics as indicators of performance:

- Acc@1 i.e. the accuracy at top 1
- The averaged Pearson's correlation coefficient over all documents (AP)

The Acc@1 is a coarse-grained metric and stands for the accuracy at top 1. Acc@1 is computed by dividing the number of correctly predicted documents by the total number of documents.

Second is Averaged Pearson (AP) which is fine-grained metric. Here we used AP because the Acc@1 misheed emotional distributions. But in Social emotion mining the emotional distributions is more important. For each document, AP measures the correlation between the predicted distributions and the actual votes over all emotion labels. The value of AP ranges from −1 to 1, where 1 indicates a perfect positive correlation.

VIII.CONCLUSIONS

In this paper we have done a survey of different models which we used in Social Emotion Mining from writer's and reader's perspectives. There are many potential application of social emotion mining, which include emotion based document retrieval, and emotion classification for online news articles. There are varieties of techniques and methods have been developed for document retrieval and classification and contingent on different models have been proposed. The Social Emotion Mining can also help to understand the preferences and perspectives of online users. We have studied different models which used in social affective text mining along with their basic fundamentals. On our investigations and research it has been that the performance of SOM is significantly more stable than that of Emotion term, ET, MSTM, SLTM, ATM, and CSTM.

REFERENCES

[1] C. Strapparava and and R. Mihalcea, "Semeval-2007 task 14: Affective text," Proc. of SemEval-2007, no. June, pp. 70–74, 2007.

[2] D. jennifer et al. "Affective Text based Emotion Mining in Social Media," in International Journal of Advance Research in Computer Science and Management Studies, vol 2 , no. 3, pp. 86-94, 2014.

[3] Yanghui Rao, Qing Li , Xudong Mao , Liu Wenyin, "Sentiment topic models for social emotion mining," Information Sciences, vol 266, pp. 90-100, May 2014.

[4] S. Bao et al., "Mining Social Emotions from Affective Text," IEEE Trans. Knowl. Data Eng., vol. 24, no. 9, pp. 1658–1670, Sep. 2012.

[5] C. Strapparava and A. Valitutti, "Wordnet-Affect: An Affective Extension of Wordnet," Proc. Fourth Int'l Conf. Language Resources and Evaluation (LREC '04), 2004.

[6] Yanghui Rao et al. "Affective topic model for social emotion detection," Neural Networks, vol. 58, no. 2012, pp. 29-37, Oct 2014.

[7] Y. Rao, "Contextual Sentiment Topic Model for Adaptive Social Emotion Classification," IEEE Intell. Syst., vol. 31, no. 1, pp. 41– 47, Jan. 2016.

[8] Xintong Li et al., "Predicting Social Emotions from Readers' Perspective," IEEE Transactions on Affective Computing, 2017.

[9] S. Bao et al., "Joint emotion-topic modeling for social affective text mining," in Proceedings - IEEE International Conference on Data Mining, ICDM, pp. 699–704, 2009.

[10] X. Li, J. Ouyang, and X. Zhou, "Supervised topic models for multi-label classification," Neurocomputing, vol. 149, no. PB, pp. 811–819, 2015.

[11] S. S.-M. S. Kim, E. Hovy, S. S.-M. S. Kim, E. Hovy, and E. Hovy, "Determining the sentiment of opinions," Proc. 20th Int. Conf. p. 1367–es, 2004.

[12] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, "ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software.," PLoS One, vol. 9, no. 6, 2014.

[13] C. Strapparava and A. Valitutti, “Wordnet-Affect: An Affective Extension of Wordnet,” Proc. Fourth Int’l Conf. Language Resources and Evaluation (LREC ’04), 2004.

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

[15] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, “New Avenues in Opinion Mining and Sentiment Analysis,” IEEE Intell. Syst., vol. 28, no. 2, pp. 15–21, 2013.

[16] Sneha V. Dehankar, K. P. Wagh and P. N. Chatur, "Web Page Classification Using Apriori Algorithm and Naive Bayes Classifier" IJARCSMS volume 3, issue 4, April 2015, pg-527- 533.

[17] Hetal C. Chaudhari, K. P. Wagh and P. N. Chatur, “Search Engine Results Clustering Using TF-IDF Based Apriori Approach” IJECS vol. 4, Issue 5, May 2015, Pg-11956- 1196.

[18] Sanjay D. Sawaitul, Prof. K. P. Wagh, Dr. P. N. Chatur,” Classification and Prediction of Future Weather by usingBack Propagation Algorithm-An Approach”,IJETAE,ISSN 2250-2459, Volume 2, Issue 1, January 2012.

[19] 10. P. H. Govardhan, K. P. Wagh, P. N. Chatur,” Web Document Clustering using Proposed Similarity Measure”, International Journal of Computer Applications (0975 – 8887) National Conference on Emerging Trends in Computer Technology (NCETCT-2014).