ANALYTIC FRAMEWORK FOR SOCIAL OPINION PREDICTION OF ONLINE NEWS BASED ON SENTIMENT ANALYSIS

[1]Deepa R. Gawande, [2]Dr. K. P. Wagh
[1]PG Scholar, [2]Assistant Professor
[1]Computer Science and Engineering, [2]Information Techonology
[1]Goverment College of Engineering, Amaravati, India

Abstract : ***In today's world everything available on the internet because of the digital era. With the fast development of web, peoples spent their most time on internet. Recently, in social platforms new function introduced in which user may select one emotion out of many choices, to express their opinion more precisely. For this tasks the social emotion mining technique is used. This is the most important and widely used in social platforms and businesses alike. In history, lots of work has been done of social emotion mining from author's perspectives. Recent research focuses on reader's emotions with news articles. In existing system Social Opinion Mining is developed which predict the reader's emotions of news based on the social opinion network. The construction of opinion network based on the semantic distance between two words. In existing system model for analysis of news the system do various comparisons which is time consuming. Therefore, in proposed system we are going to develop an Intellectual Sentiment Analytic Model for the social opinion prediction. Due to Sentiment analysis the Intellectual Sentiment Analytic Model reduced time to large extent and effective results constructed as compared to the state of the art model.***

IndexTerms - Social Emotion Mining, Social opinion prediction, News Article, Preprocessing, Sentiment Analysis, Intellectual Sentiment Analytical model.

I. INTRODUCTION

It is a fact that social media has tremendously changed the way people interact and carry on with their every-day lives. The internet has becoming a lifeblood in every peoples life, majority of people who are online spend most of their time on social media sites. One of the most important service of social media is News delivery. This motivate many social websites. The websites providing a way in which user will share their opinion after reading the news article. With the increased development of web, numerous documents are evoke with user-generated emotions such as happiness, sadness and surprise. On research it has been found that emotional labels widely used in social web services. The Social Emotion Mining is used to understand the emotions of users. The Social Emotion Mining understand and predict the user's emotions towards online documents. In history many models and methods have been proposed and well discussed to deal with social emotion mining. In existing work the Social Opinion Mining model used for social opinion prediction. This model based on the construction of real-time opinion network. On studies, it is clear that the performance of social opinion mining model is more stable as compared to the previous models i.e. Emotion Term, Emotion Topic Model, Sentiment Topic Model, Affective Topic Model and Contextual sentiment topic model.

In the existing system, the Social Opinion Mining Model for the analysis of News do the comparison of various vector conversions which is time-consuming. Therefore in a proposed system, we are going to develop an Intellectual Sentiment Analytic Model for the interceptive matching of the proposed news article with the existing dataset. The intellectual Sentiment Analytic Model constructed based on the sentiment analysis of news article, in which parsing, data cleaning and processing done over the news article. Due to sentimental analysis time will reduce to large extent and effective results will be constructed. The accuracy of the prediction of social opinion prediction is enriched as compared to the social opinion mining model.

II. SOCIAL EMOTION MINING

In Data/Text Mining the term "mining" is an analogy to the resource extraction process of mining for rare minerals. The Social Emotion Mining is a process of understanding and predicting the hidden emotions of the social user towards online contents known as social emotion mining. It has become increasingly important to both social platform and businesses alike to better understand their user and leverage the learned knowledge to their advantage. In 2016 Facebook introduced the function, where the user may select one emotion out of many choices to express their opinion more precisely. The similar function allowed in Chinese news portal i.e. www.sina.com. This is one of the largest news websites in China. Sina websites contain breaking news, current events and useful information on life, culture, and travel in China. Large numbers of people concerned about a hot news online. In preliminary work there are many models have been studied and proposed for the Social emotion Mining.

III. RELATED WORK

There are two different perspectives to mined emotions from the text: From the perspective of writer and reader. Past work focused on the writer's perspectives i.e. accurately model the connections between word and emotions. Recent research focuses on reader's emotions with the News article. The sentiment analysis of reader's emotions is more meaningful as compared to classical sentiment analysis from writer's perspectives. Due to this prediction of reader's emotions has become a promising research area. Many models have been proposed and well-studied in previous work to deal with social emotion mining.

C. Strapparava et al. proposed a SWAT model [1] In SWAT in which a word emotion mapping dictionary has been constructed, then used this to detect emotions of unlabeled news headings.

Shenghua Bao et al. proposed a Joint Emotion-Topic Model [2], [3] in which first developed the Emotion –Term model which is the word-level model used to find word emotion associations. In Emotion-Term model different senses can be evoked from the same word. To overcome this drawback Joint Emotion-Topic Model which is topic-level is proposed. The Joint Emotion-Topic model established additional emotion generation layer to Latent Dirichlet Allocation.

Yanghui Rao et al. proposed two sentiment topic models [4], The Multi-label Supervised Topic Model (MSTM) find the association between latent topics with readers emotions. The Sentiment Latent Topic Model (SLTM) which invent purposeful latent topics from which social emotions produced.

Yanghui Rao et al. proposed Affective Topic Model [5] (ATM) which used to develop a social emotion lexicon and to distinguish social emotions of documents.

Yanghui Rao et al. proposed Contextual Sentiment Topic Model [7] (CSTM) The CSTM model solved the problem of adaptive social emotion classification by classifying reader's emotions across different context

IV. EXISTING SYSTEM

Xintong Li et al. proposed Social Opinion Mining (SOM) Model [8] which is constructed based on the social opinion network, which is used for social opinion prediction. Compared with the previous model the SOM model treats the news content and emotions distribution jointly in opinion network.

Kim and Hovy [11] describe an opinion can represent with a quadruple i.e. Topic, Holder, Claim, and Sentiment. This kind of social opinion can be represented with a quadruple <event, f, s, t> where an event is a social event; f is the text feature set of the social event; s is the result of voting and t is the time when this social event occurred. The Social Opinion Mining model aims to predict s based on the social opinion quadruples.

The following Figure shows the flowchart of the existing model i.e. Social Opinion Mining model.

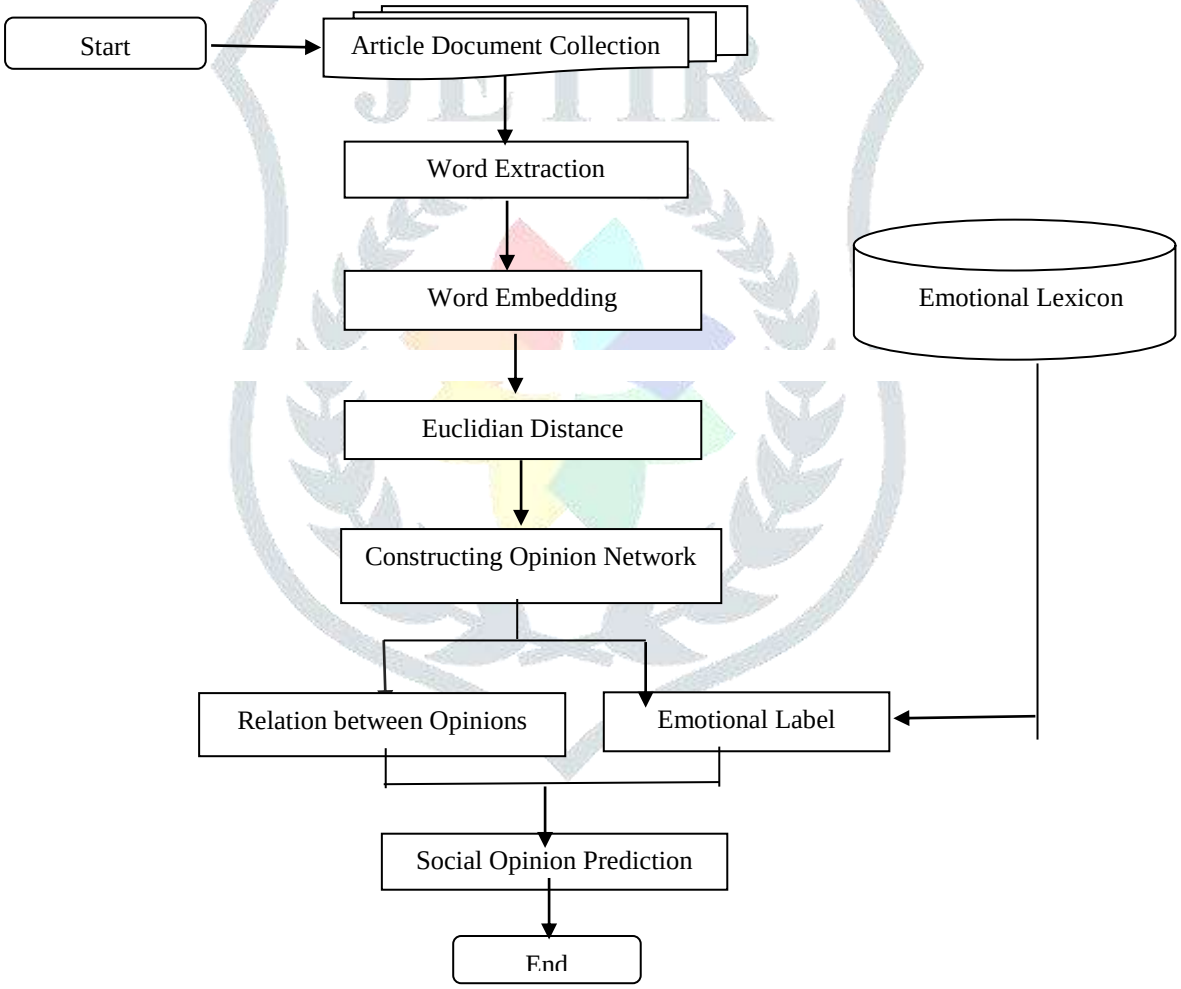

Figure 1. Social Opinion Mining Model

4.1 Working of Social Opinion Mining Model

The Social opinion mining model predicts social opinions by measuring the semantic similarity between events. In this model, the opinion network is constructed based on the semantic distance. The Social Opinion Mining model treats the news content and emotion distributing as a whole opinion structure. The following are the steps of Social Opinion Mining Model.

1. Word Extraction

The first step of social opinion model is word extraction. Here the raw feature (f) of social event is considered as the bag of words. For measuring the importance of a word in corpus the term frequency (TF) is used.

TF (t) = (Number of times term t appears in a document) / (Total number of terms in the document).

2. Word Embedding

The second step of social opinion mining model is word embedding. For embedding of words, Word2vec is used. First, train the word vectors according to most recent Wikipedia Word Corpus. Word Embedding converts the text into numbers.

3. Euclidean Distance

In this step, calculate the semantic distance between two words by using Euclidian Distance. The Euclidean distance between words in word2vec space measures the semantic similarity.

4. Construction of Opinion Network

The Social Opinion Mining model is constructed based on the social opinion network. The opinion network constructed using the semantic distance between two words. We can predict the social opinion through the network. ForceAtlas2 [12] algorithm used to arrange the layout of nodes. Threshold value 0.7 is used for pruning the network.

5. Social Opinion Prediction

From the opinion network, the social opinion is predicted. There is a relationship between communities in opinion network and emotional labels. The threshold value depends on the current network state, therefore not applicable to community structure. Thus we used the inference of community structure of opinion network and then we predict the social opinion based on the opinion growing network [8].

V. DATA COLLECTIONS

For the application of the existing system and proposed methodology to test effectiveness dataset is needed. We have collected the data of news article from popular news websites i.e. www.sina.com. This news article contains URL address, publishing date (from January to April of 2014), News title, Content, and user ratings over 8 emotion labels: “touching”, “empathy”, “boredom”, “anger”, “amusement”, “sadness”, “surprise” and “warmness”.

VI. PROPOSED METHODOLOGY

Here, we are going to develop an Intellectual Sentiment Analytic Model for the social opinion prediction. In existing system model opinion network is constructed for social opinion prediction. This opinion network is constructed based on the semantic distance between two words. To calculate semantic distance these words are mapped to the word vector. The word embedding is performed on the word vector. The word embedding transferred the text into numbers.

In Intellectual Sentiment analytic model taking input News article as an input. The News article contains ‘Title’, ‘Content’, ‘User rating’, ‘Time’, ‘URL’. We can predict the social opinion through the sentiment score.

5.1 SENTIMENT ANALYSIS

Sentiment analysis refers to a broad area of Natural Language Processing and text mining. It is concern not with the topic a document is about but with opinion it expresses that is the aim is to determine the attitude (feeling, emotion and subjectivities) of a speaker or writer with respect to some topic to determine opinion polarity. Sentiment analysis of social media data performed by machine learning approach and dictionary based or corpus based approach as a two different methods. Most popular supervised machine learning methods for sentiment analysis is naïve Bayes and support vector machine which formulates sentiment analysis problem as learning and classification problems as positive, negative and neutral. Recently several research work has been done on finding and classifying emotions like joy, sad, happy, and surprised. Unsupervised learning uses NLP techniques such as part of speech, bag of word, word frequency for learning and classification of text to improve accuracy. In dictionary based approach, dictionary with large list of word with their sentiment orientations is used to calculate and identify sentiment if text in the document. Articles are a type of content that can be easily produced, have a small size, short lifespan, and low cost, properties that makes them interesting for fast information diffusion through social media platforms or social networks. As a consequence, a significant amount of research has been focused in understanding the interest around news, including general observation on how content is generated, describing the decay of interest over time, community detection, and prediction of popularity.

5.2 FLOWCHART OF PROPOSED SYSTEM

The above figure shows the flowchart of the proposed the system.

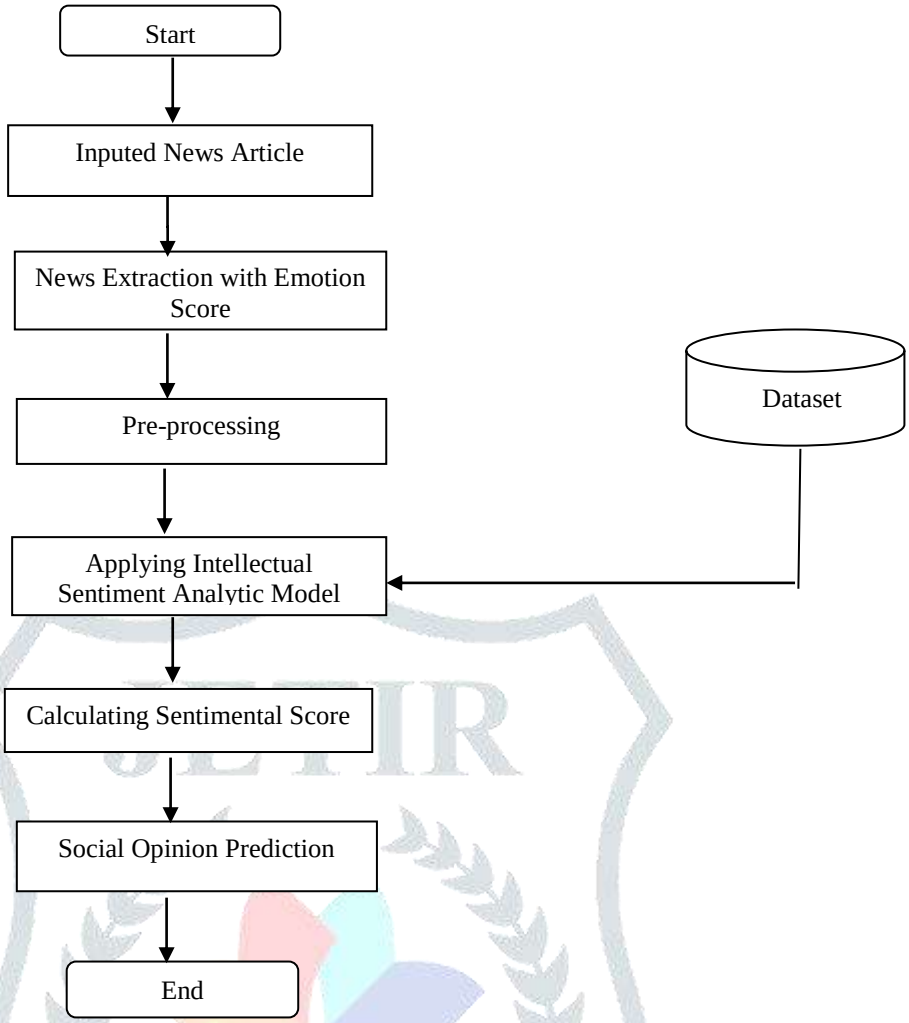

Figure 2: Flowchart of Proposed System

The following are the steps of Intellectual Sentiment analytic Model:

1. INPUT

Take a News Article as an input from to generate social opinion prediction. Here we give News Title as an input. Then we are going to match this title with existing dataset. After matching the matched news title with emotional labels extracted from each required news article.

2. DATA PRE-PROCESSING

The aim of data pre-processing is to transform each document into a sequence of features that will be used in subsequent steps. Stop word Removal is an important pre-processing technique used in a Natural Language Processing. Stop words are occur frequently and has no meaning. Therefore, we removed this words from the sentence. Apply named entity recognition to filter out person names from the documents, because we found that few of the person names occurring in news articles bear any consistent effective meanings. So we remove all the words that represent no meaning related to any of the specified emotions and by thus optimizing the content.

3. SENTIMENTAL SCORE

In this step, we calculate the sentimental score of the news article. Consider N is the number of the news article in a dataset.
M(w) is the match word of the news article.
T(w) is the total number of the word present in the title of news.
C(cat) is social emotion category

$$\text{Sentiment Score} = \sum_{i=0}^{N} \frac{\text{M(w)}}{\text{T(w)}} \qquad \ldots\ldots\ldots\ldots (1)$$

$$\text{Predicted Score} = \sum_{i=0}^{N} \frac{\text{C(cat)}}{\text{N}} \qquad \ldots\ldots\ldots\ldots (2)$$

4. SOCIAL OPINION PREDICTION

From the above sentiment score and predicted score we can predict the social opinion. The accuracy and time for the social opinion prediction of the proposed model are more as compared to the Social Opinion Mining model.
The steps required to build the system of social opinion prediction are given in this algorithm. The step by step implementation of the system architecture is described. Input News article Output: Predicting social opinion with the News article Steps:

The following algorithm used for social opinion prediction.

1: Identifying Title of the input News

2: Title word extraction and perform data preprocessing

3: Comparing Title's word with every available news in the dataset

4: If
T(w) >= Threshold
Add T(w) to newsque( )
Else
Eliminate;
5: Apply Score manipulation on newsque( )

6: Filter Results

7: Social Opinion Prediction

VI. RESULTS

Following table shows the results of the proposed model i.e. Intellectual Sentiment Analytic Model over the existing model i.e. Social Opinion Mining Model. The sentiment score of each news is calculated by using the sentimental formulae. We observed that the proposed Intellectual Sentiment Analytic Model required less time as compared to existing Social Opinion Mining model which shown in the following table.

News No.	URL	Title	Sentimental Score	Time for Existing System(MS)	Time for Proposed System(MS)
1	http://english.sina.com/china/p/2014/0101/659858.html	First photos of Liaoning battle group made public	0.11111111	0.655500000000001	0.0345
2	http://english.sina.com/china/2013/1231/659681.html	President Xi confident of New Year reforms	0.18181819	0.666600	0.0425
3	http://english.sina.com/china/2014/0103/660445.html	The duck returns in Taiwan	0.61904764	0.688440	0.0487

Table 1. Results of the Proposed Model

VII. CONCLUSIONS

With the increased use of internet, the social opinion prediction is an important perception of market analysis and political decision. In existing system model for analysis of online social opinion, the system news does the comparison of various vector conversions and the opinion network is constructed for the social opinion prediction which is time-consuming. Therefore in proposed system, we develop the Intellectual Sentiment Analytic Model based on sentiment analysis of news article in which the parsing, data cleaning, and processing done over the news article. Due to sentimental analysis time will reduce to large extent and effective results will construct.

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] 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.

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

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

[5] 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.

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

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

[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 R. Mihalcea, "Learning to Identify Emotions in Text," Proc. 23rd Ann. ACM Symp. Applied Computing (SAC '08), pp. 1556-1560, 2008.

[14] 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.