Analytic system based on prediction analysis of social emotions from users on E-commerce

[1]Lina S. Dhage, [2]Sachin A. Murab

[1]Student, Dept. of CS Engineering, J.C.O. Engineering & Technology, Yavatmal, Maharashtra, India
[2] S.G.B.A. University, Amravati, India
Email – lina.darne15@gmail.com

Abstract: *: Over social media there are lots of symbols are used as compared to text this is an unstructured type of which get considers day by day increase in such symbols is moving the towards the new data prediction determination technique.*

Due to the rapid development of Web, large numbers of documents assigned by readers' emotions have been generated through new portals. Comparing to the previous studies which focused on author's perspective, our research focuses on readers' emotions invoked by news articles. Our research provides meaningful assistance in social media application such as sentiment retrieval, opinion summarization and election prediction. In this paper, we predict the readers' emotion of news based on the social opinion network. More specifically, we construct the opinion network based on the semantic distance. The communities in the news network indicate specific events which are related to the emotions. Therefore, the opinion network serves as the lexicon between events and corresponding emotions. We leverage neighbor relationship in network to predict readers' emotions. As a result, our methods obtain better result than the state-of-the-art methods. Moreover, we developed a growing strategy to prune the network for practical application. The experiment verifies the rationality of the reduction for application.

In this paper, we implement social opinion prediction by generating a real-time social opinion network. In more details, first, we train word vectors according to the most recent Wikipedia word corpus. Second, we calculate se-mantic distance between news via word vectors.

Key Words: *Affect sensing and analysis, recognition*

1. INTRODUCTION:

Social emotion prediction is of value to market analysis and to political decision With the free and convenient communication environment of internet, people show increasing enthusiasm of online communication. Meanwhile, the internet users prefer to pro-duce and convey online information through expressing personal opinions than just obtain online information. In this way, numerous news articles and comments have been published and shared rapidly via social media ser-vices. As a result, abundant underlying positive or negative emotion information spreads and reflects the social sentiment tendency. Most intuitively, emotional label has been widely used in social web services. Fig. 1 indicates the result of voting for a news article using emotion labels from a popular news portal. Large numbers of people concerned about a hot news online. Therefore, valuable and available emotional information is continuously pro-vided for scientific research work[4].Furthermore, comparing to the traditional methods, which need to do numbers of surveys offline, data processing technology has been developed more feasible in the field of emotional extraction, analysis and prediction with its benefits of lower cost, higher efficiency and more accuracy. Under this circumstance, readers' emotions prediction shows a highly research potential.

Compared with the typical tasks of sentiment analysis, opinion mining or affect recognition which based on subjective text,social opinion prediction focuses on objective text, for example news articles, which may not contain any opinion, but can evoke readers' certain emotion. Due to the particularity of the task, social opinion prediction has potential applications which are different from those of writer-sentiment analysis [5]. Considering the effect of social media on the public sentiment, social emotion analysis engenders large benefits to social and economic problem, such as political issues and brand perception.

In this paper, we implement social opinion prediction by generating real-time social opinion network. In more details, first, we train word vectors according to the most recent Wikipedia word corpus. Second, we calculate se-mantic distance between news via word vectors. As a metric between opinions, semantic distance allows us to construct the opinions growing network to describe the dynamical social opinions. Last, we predict follow-up news' social emotion based on the network.

Fig. 1. An example of emotion labels and user ratings

2. LITERATURE REVIEW:

In existing paper it is proposed that the system can do the prediction of emotions of the users they are taken the reference of the news article which helps us to know about the user's emotions regarding to such a article .In this the experiment get proposed on datasets. Social opinion prediction is a difficult research endeavour. As the initial research work on social opinion prediction, "affective text" in SemEval-2007 Tasks [11],[13]. Intend to annotate news headlines for the evoked emotion of readers. Another research focus on readers 'emotion evoked by news sentences [15]. Existing methods of social opinion prediction can be divided into three categories: knowledge-based techniques, statistical methods and hybrid approaches. Because of the deficiency of information of news text [13], [16]. It are unmanageable to annotate the emotions consistently. Knowledge-based techniques utilize existing emotional lexicon to supplement the prior knowledge for annotating the emotions. The popular emotional lexicon includes Affective Lexicon, linguistic annotation scheme [18], Word Net-Affect [19], Sent Word Net[20], and Septic Net[21]. The drawback of knowledge-based techniques is the reliance on the coverage of the emotional lexicon. These techniques cannot process terms that do not appear in the emotional lexicon. Statistical methods predict social opinion by training a statistical model based on a large number of well-labelled corpuses. There is two principal categories of statistical methods: word-level [11], [14] and topic-level[22] methods. Word-level methods focus on exploiting the sentiment of individual words [11][14]on the idea that words are the foundation of user sentiments. In order to model the word-emotion association, a variant of Naïve Bayes model named Emotion-Term (ET) is created. The words extracted from the news articles are considered as independent features which indicate the emotion. However, word-level features in social opinion prediction are always interfered by the background noise words. In particular, the methods treat each word in-dividedly; many emotional words are usually mixed with background noise words

3. METHOD:

3.1 Social Opinion Model

Sentiment-related phenomena can be explained as the process of evaluation of events, objects or persons[23], [25]. The opinions are caused by the subjective evaluation of the "raw" stimuli. The "raw" stimuli may have no intrinsic emotional meaning, but will be appraised by personal relevance and implications. For social opinion, the "raw" stimuli are only text and its features which are difficult to expound the corresponding sentiment-related phenomena. In fact, there are less than 5% of directly emotional words of a text in daily speech, emotional writing, and affect-laden poetry. In journalism domain, a lower percentage is undisputed. It is rarely influenced by the personal relevance under the social community. To simplify the problem, we focus on implications without personal relevance.

According to the cognitive approaches, the result of voting is "the person's experience, goals and opportunities for action" [24]. It is process that evaluates an event by dimensions such as urgency, consistency with goals, etc. All the social opinions share the similar emotional experience, goals and opportunities for action with each other. From the NLP perspective, the models are inexplicable but feasible. From psychology and linguistics perspective, the models are explicable but lack of use in the service. Based on the general characteristic, similarity is one of six principles that guide human perception of the world in Gestalt theory. We can predict social opinions by measuring the semantic similarity between events.

TABLE 1
NOTATIONS OF FREQUENTLY-USED VARIABLES

Notations	Description
D	All the news
$w_{d,k}$	The kth word token in news d
$W=\{w_i\}$	The set of vocabulary
e_k	The kth emotion label
$E=\{e_k\}$	The list of emotion labels. the common instances of E are "joy", "anger", "fear", "surprise", "touching", "empathy", "boredom", "sadness", "warmness", etc.
$v_d=\{v_{d,e_k}\}$	The set of ratings over E emotion labels
v_{d,e_k}	the number of online users who have voted the kth emotion label e_k for news d
$\hat{v}_n$	The prediction of emotion conditioned on future unlabeled news d which only contain words.
$s_{i,j}$	The similarity between news i and news j
$c(i,j)$	The distance between word w_i and word w_j
x_i	The vector representation for word w_i

The social cognitive process can be modeled based on a stereotypical knowledge set consisting of social opinion.
P={ < Event, f, s, t >1, <Event, f, s, t >2, …., <Event, f, s, t >n } .

Instead of establishing appraisal criteria, the cognitive process can be regarded as the neighbor analysis in set *P*. The set *P* can be interpreted as social experience. The cognitive process can be simplified as matching the "raw" stimuli between the priori social opinions in set *P*. The social community has stabilized emotion towards specific events. It can be explained by a social psychology behaviour named "stereotype" which is a fixed view of people, groups, events, institutions, or problems. Stereotype widely exists in media. To be more accurate, the task is modeling the relationship between current event and priori social opinions based on semantic similarity.

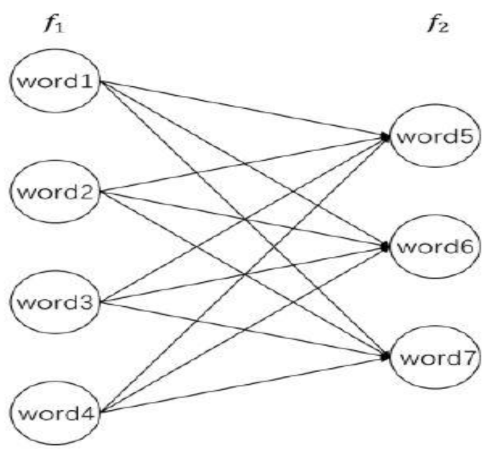


Fig. 2. Optimal transport distance between two histograms.

As we defined, the social opinion is a quadruple . In social opinion model, the first step is called words-extract in which the raw feature () of social event is processed as Bag of Words (BOW) which is widely used for document representation. Considering the online performance of the algorithm, we utilize the term frequency (TF) instead of term frequency–inverse document frequency (TF-IDF)for measuring the importance of a word in corpus.TF-IDF is not available in real-time updated data. Because we need to know the global distribution of words to calculate TF-IDF. By contrast, TF can be explained as social community ratio of attention to the words. Based on the explanation of attention, it is persuasive that the readers concern each news equally, in other words, pay equal attention to each news. Based on it, the assumption is justifiable that the BOW should be normalized. So far, the preprocessed in social opinion quadruple < event, f, s, t > i is a normalized histogram which is finite dimension vectors with nonnegative coordinates whose sum is equal to 1. That is

$$f_i \in \Sigma_w \quad \Sigma_w \triangleq \left\{ u \in \mathbb{R}_+^w \mid \sum_{i=1}^{w} u_i = 1 \right\}$$

We f_{ij} denote as the weight of *j*th word in *i*th event.

Considering the implications in social opinion feature, we leverage the recent result named word2vec[53]which shows that a log-bilinear model can learn high quality embedding of words by local co-occurrences in sentences. An embedding of a word is a finite dimensional vector which expresses the word's meaning. We can measure the semantic distance between two words by Euclidean distance. The stable word-embedding is a cornerstone of our model. We require large corpora to perform word2vec.The learning corpora can influence the quality of embedding directly. It is reasonable to choose Wikipedia as learning corpora. Com-paring with the web-based text collected from online media, Wikipedia is a free online encyclopedia consisting of various entries which ensures completeness of words so
Then stability of the model. Moreover, it ensures that the model can provide strong robustness because of Wikipedia' real time self-renewal.
The Euclidean distance between words in word2vecspace measures the semantic similarity. More precisely, $c(i,j) = \|x_i - x_j\|_2$ denotes the distance between word i and word j where x_i and x_j represent corresponding word. By measuring the similarity between social opinions $\langle event, f, s, t \rangle_1$ and $\langle event, f, s, t \rangle_2$, can we model the relationship. As f is a normalized histogram and each word in f_1

has a meaning distance to each word in f_2. To compare two histograms f_1 and f_2, we apply optimal transport distance. The task of optimal transport distance between two histogram scan be illustrated in Fig. 2. Word1-4in f_1 can be moved toword5-7 in f_2, at a cost $c(i,j)$ based on the semantic distance. Based on the semantic distance. The semantic transport task between f_1 and f_2, is to move all the words in f_1 to f_2.
We utilize the word mover's distance metric as our distance. The measurement of similarity can therefore be achieved, in principle, by the measurement of optimal transport distance. First, we assume that word *i* in f_1
Can be transformed totally or partially into any word in f_2. There exists a flow matrix $T \in \Re^{n \times n}$, in which element $T_{ij} \geq 0$ indicates the extent of switch from word *i* in f_1 to word *j* in f_2, are that:

(1) the amount of outgoing flow from word *i* equals

f_{1i}, i.e. $\sum_j T_{ij} = f_{1i}$;

(2) likewise, the sum of incoming flow to word j equals

f_{2j}, i.e. $\sum_i T_{ij} = f_{2j}$.

Now, the distance between the two histograms are defined as the minimum (weighted) cumulative cost required to transform all words from f_1 to f_2, i.e. $\sum_{i,j} T_{ij} c(i,j)$.

It can be described as following:

$$\min_{T\geq 0} \sum_{i,j=1}^{n} T_{ij} c(i,j)$$

$$Subject\ to: \sum_{j=1}^{n} T_{ij} = f_{1i} \quad \forall i \in \{1,\cdots,n\} \qquad (1)$$

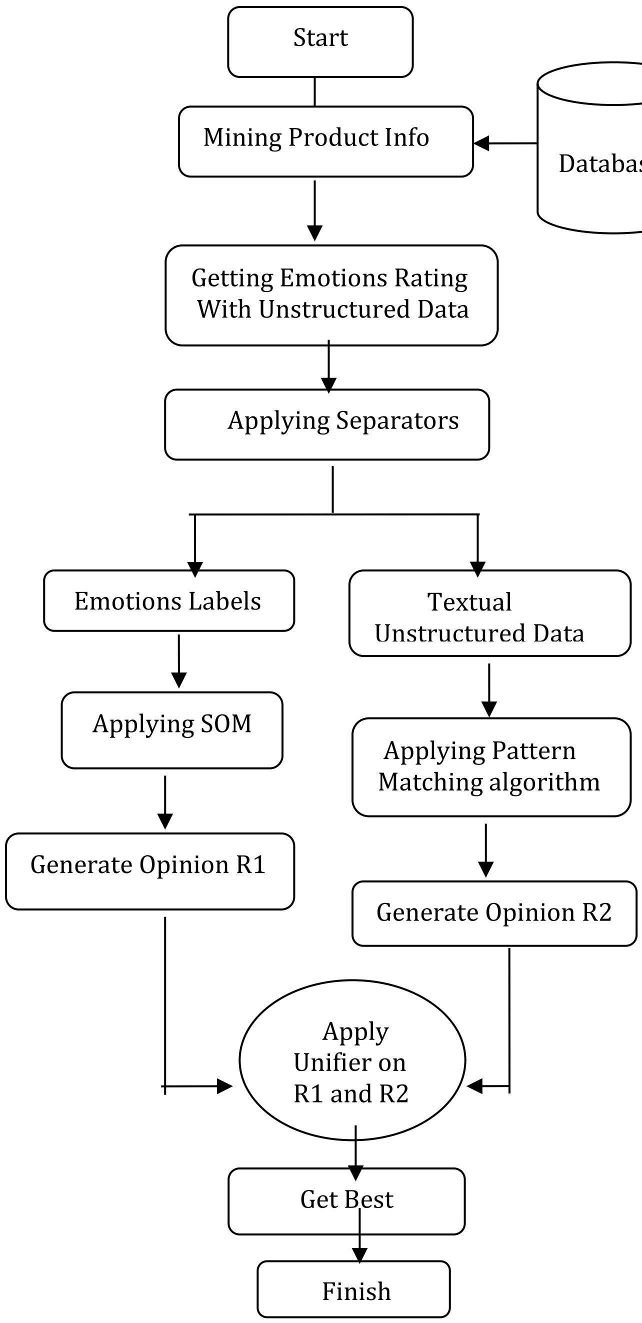
$$\sum_{i=1}^{n} T_{ij} = f_{2j} \quad \forall j \in \{1,\cdots,n\}$$

By solving the linear program above, can we model the relation between opinions as opinion distance.

4. DISCUSSION:

In Proposed System firstly upload data in two data sets, that is “sina.com’ and “times of India” th en convert unstructured data into structure data by the data mining algorithm then we getting emotions rating and structure data. Then applying separator on given data, and data are divided into two forms that is emotional labels and textual unstructured data, then apply the SOM (social emotion model) and pattern matching algorithm respectively on those given data, and outputs are generate opinion R1 and R2 respectively.
Then after apply unifier on generation R1 and R1 and get best result, get propose work and existing work result.

4.1 Proposed System Diagram



By looking towards the technique given in existing we are proposed a propose social opinion model for measuring similarity among news and business intelligence analytic module based on emotion detection regarding to the product reviews based on mining with reviews, feedback, complaints given by users this will help us the user for giving the instant and fast response and which also become very proper for business development. In proposed we can implement the opinion network and emotion opinion model on the datasets retrieved from business data. Opinion prediction system will helps to predict and decision making in business intelligence

5. ANALYSIS:

In this section, the problem about social opinion prediction is well defined, including the relevant general terms and notations. Later on, we describe the social opinion model and opinion network in detail. Finally, we propose a strategy for social opinion prediction.

5.1 Problem Formulation

We define the following notations for describing the social opinion prediction: An online news collection D consists of news

d, and the emotion ratings labels E . The list of emotion labels is denoted by $E=\{e_k\}$, and e_k indicates emotion titled “joy”, “anger”, “fear”, “surprise”, “touching”, “empathy”, “boredom”, “sadness”, “warmness” etc. In particular, a news is a set of word tokens $W=\{w_i\}$, and a set of ratings over *E* emotion labels denoted by $v_d = \{v_{d,e}\}$. The value of v_{d,e_k} is the number of online users who have voted the kth emotion label e_k for news d. Table 1 summarizes the notations of these frequently-used variables.
According to Kim and Hovy, the opinion can be split into four parts: topics, opinion holder, claim, and sen-timent.To be specific, a holder believes a claim about a topic with a sentiment. For social opinion, the opinion holder stands for users who have voted the news. The topic can be replaced by the content of the news. The sentiment can be measured by the vote around the set of predefined emotional labels. The claim is unobservable and inessential in this task. This kind of social opinion can be model with a quadruple $\langle event, f, s, t\rangle$, where event stands for the social event; f is the text feature set of social event; is the result of voting towards social event which is represented as distribution over the predefined emotional labels. *t* is the time when the social events occurred. The social opinion prediction task this paper discussed is focused on the prediction of s based on the former social opinion quadruples.

6. RESULT:

	Time taken for processing (NS)	Accuracy
SOM	8954623	92%
Proposed	8902345	95%

Result Analysis

	No of Record Scan	Time Taken
Existing Work	3	0.10350000000000001
Proposed Work	1	0.0345

7. RECOMMENDATIONS:

We can integrate the system with various news portal as well as to some data processing tools in which the proposed work will save the execution time of and improve the system to much more this system is get recommended for analytics tools for data processing.

Advantages

- This system will helps us to perform data mining on emotion based data
- It will help to develop business intelligence based on review of emotions
- helpful for the business development

Disadvantage

- Will be able workout properly in some business intelligence due improper data management system

8. CONCLUSIONS:

In this paper, we analyze the online social opinions and propose social opinion model for measuring similarity among news and users give the comments on particular news. Due to word-embedding pre-trained on Wikipedia, model's stability and robustness are guaranteed and can hardly be influenced by the size of news data. Based on the similarity, we construct an opinion network to detect user-generated social emotion by the structures of opinion network.

There is significant correlation between emotion and structures of news network as we expected. The performance of the prediction based on opinion network is more stable and accurate than existing models. In addition, we propose a threshold-based net-work growing strategy for pruning the network

In proposed the prediction of products online social opinions on news for the e-commerce is a crucial way by performing mining implementation on the feedback, reviews and comment given by them so to improve the business productivity the proposed work helps the e-commerce platform to analyze the customer data in business driven ways.

REFERENCES:

1. www.researchgate.net/publication/316259881_Predicting_Social_Emotions_from_Readers-27_Perspective, IEEE*/paper*
2. E. Cambria, B. Schuller, Y. Xia, and B. White,(Sep. 2016) "New avenues in knowledge bases for Natural language processing," Knowledge- Based Syst., vol. 108, pp. 1–4,.
3. E. Cambria,(Mar. 2016) "Affective Computing and Sentiment Analysis," *IEEE Intell. Syst.*, vol. 31, no. 2, pp. 102–107,
4. B. Zhang, X. Guan, M. J. Khan, and Y. Zhou,(2012), "A time-varying propagation model of hot Topic on BBS sites and Blog networks," Inf. Sci. (Nay), vol. 187, pp. 15–32,.
5. Q. Wang, O. Wu, W. Hu, J. Yang, and W. Li,(Sep. 2016) "Ranking social emotions by learning List wise preference," in 1st Asian Conference on Pattern Recognition, ACPR avenues in knowledge bases for Natural language processing," Knowledge- Based Syst., vol. 108, pp. 1–4,.
6. B. Zhang, X. Guan, M. J. Khan, and Y. Zhou,(2012) "A time-varying propagation model of hot Topic on BBS sites and Blog networks," Inf. Sci. (Nay)., vol. 187, pp. 15–32,.
7. Q. Wang, O. Wu, W. Hu, J. Yang, and W. Li,(2011) "Ranking social emotions by learning List wise preference," in 1st Asian Conference on Pattern Recognition, ACPR 2011, pp. 164–168,.
8. K. H.-Y. Lin, C. Yang, and H.-H. Chen,(2007) "What emotions do news articles trigger in their readers?," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '07*, pp. 733-734,.
9. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre,(2008) "Fast unfolding of Communities in large networks," J. Stat. Mech. Theory Exp., vol. 0008, no. 10, pp. 155-168,.
10. P. Katz, M. Singleton, and R. Wicentowski,(2007) "SWAT-MP: The SemEval-2007 Systems for Task 5 and Task 14," in Proceedings of the 4th International Workshop on Semantic Evaluations, pp. 308- -313,
11. Y. Kim,(2014) "Convolutional Neural Networks for Sentence Classification," in EMNLP, vol. 21, no. 9, pp. 1746–1751,.
12. T. Mikolov, K. Chen, G. Corrado, and J. Dean,(2013) "Efficient estimation of word representations in vector space," arXiv Prepr. arXiv1301.3781, pp. 1–12,.
13. C. Strapparava and R. Mihalcea, "Semeval-2007 task 14: Affective text," *Proc. of SemEval-2007*, no. June, pp. 70–74, .
14. P. K. Bhowmick, (2009) "Reader Perspective Emotion Analysis in Text through Ensemble based Multi-Label Classification Framework," in *computational intelligence and security*, vol. 2, no. 4, pp. 64-74,
15. C. Quan and F. Ren, (2010) "An Exploration of Features for Recognizing Word Emotion," in *international conference on computational linguistics*, pp. 922-930,.
16. A. Ortony, G. L. Clore, and A. Collins, (1988) "*The Cognitive Structure of Emotions*," The Quarterly Review of Biology, vol. 18, no. 6..
17. J. Wiebe, T. Wilson, and C. Cardie, (2005)"Annotating expressions of opinions and emotions in language," *Language Resources and Evaluation*, vol. 39, no. 2–3. pp. 165–210,.
18. C. Strapparava and A. Valitutti, (2004) "WordNet-Affect: an affective extension of WordNet," *Proc. 4th Int. Conf. Lang. Resour. Eval*, pp. 1083–1086,.

19. A. Esuli and F. Sebastiani,(2006) "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining," *Proc. 5th Conf. Lang. Resour. Eval*, pp. 417–422,.
20. E. Cambria, D. Olsher, and D. Rajagopal, (2014) "SenticNet 3: a common and common-sense knowledge base for cognition- driven sentiment analysis," *Twenty-eighth AAAI Conf.*, pp. 1515–1521,.
21. Y. Rao,(Jan. 2016) "Contextual Sentiment Topic Model for Adaptive Social Emotion Classification," *IEEE Intell. Syst.*, vol. 31, no. 1, pp. 41–47,
22. A. Ortony, G. L. Clore, and A. Collins,(1988) "*The Cognitive Structure of Emotions*," The Quarterly Review of Biology, vol. 18, no. 6..
23. T. Mikolov, K. Chen, G. Corrado, and J. Dean,(2013) "Efficient estimation of word representations in vector space," *arXiv Prepr. arXiv1301.3781*, pp. 1–12,
24. R. E. D. L. Lopes and O. Vian, (2007) "The Language of Evaluation: appraisal in English," in *symposium/workshop on electronic design, test and applications*, vol. 23, no. 2, pp. 371–381,.