



An intelligent textual corpus big data computing approach for lexicons construction and sentiment classification of public emergency events

Wei Zhang¹ · Yan-chun Zhu² · Jia-peng Wang¹

Received: 7 September 2018 / Revised: 11 November 2018 / Accepted: 30 November 2018 /

Published online: 8 December 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

Considering the deficiencies in the existing emotional lexicons like too many manual interventions, lack of scalability and ignorance of dependency parsing in emotional computing, this paper first uses Word2Vec, cosine word vector similarity calculation and SO-PMI algorithms to build a public event-oriented Weibo emotional lexicon; then, it proposes a Weibo emotion computing method based on dependency parsing and designs an emotion binary tree based on dependency parsing, and dependency-based emotion calculation rules; and at last, through an experiment, it shows that this emotional lexicon has a wider coverage and higher accuracy than the existing ones, and it also performs a public opinion evolution analysis on an actual public event and the empirical results show that the algorithm is feasible and effective.

Keywords Textual corpus · Big data · Lexicon construction · Sentiment computing · Public emergency events

1 Introduction

The rapid rise and popularization of social media has a profound impact on the society today. The emergence of the social media broadens the channels of communication and the people's vision. Sina Weibo, as an emerging technical platform for social communication and contact,

✉ Yan-chun Zhu
kddzw@163.com

Wei Zhang
weizhang@cufe.edu.cn

Jia-peng Wang
wangjiapeng.8888@163.com

¹ School of Information, Central University of Finance and Economics, Beijing 100081, China

² Business School, Beijing Normal University, Beijing 100875, China

has characteristics that are distinctively different from the past communication tools and platforms. As an emerging multimedia platform, with its advantages, for example, being instant, user-friendly to the grassroots, zero-access restriction, high interactivity, weak control and fission-style mode of dissemination, Weibo has gradually become cybercitizens' first choice to obtain information and express their opinions [10, 14, 26]. Whenever a public event occurs, Internet users would respond by commenting, reposting and giving likes to the news, which would quickly form Internet public opinions. Internet public opinions are a collection of feelings, views and suggestions expressed by Internet users in various cyberspaces on social events or problems [18]. After a public event occurs, if the Internet users' emotional tendency, emotion types and public opinion evolution trend about the event can be analyzed and identified in a timely manner, the government and enterprises will be able to choose the correct way to handle the event, put forward correct emergency measures and improve the control efficiency of Internet public opinions [42].

In recent years, emotions expressed in social media messages have become a vivid research topic due to their influence on the spread of misinformation and online radicalization over online social networks [2, 4, 7]. Thus, it is important to correctly identify and classify emotions in order to make inferences from social media messages [11, 15, 16, 19, 33, 34, 37]. Social platforms such as Weibo provide platforms for emotion research, but they also pose great challenges for the research on natural language processing [5, 22]. In the sentiment analysis of Weibo's corpus, the emotional lexicon used often cannot cover the commonly used Weibo vocabulary, which makes it difficult to obtain good results in sentiment analysis. In addition, when the existing supervised methods are used to analyze the highly sparse features extracted by Weibo, the classification accuracy of emotion computing is quite low [31], and the relationships between emotion sources and emotional words are not considered [29].

In order to solve the above drawbacks, this paper firstly constructs a Weibo emotional lexicon with a wide coverage; then, by analyzing the dependency of Weibo text, it proposes a Weibo emotion computing method based on dependency parsing and computes the emotional tendency and intensity of a single post; finally, with an actual event as an example, it performs emotional tendency computations and analyzes the evolutionary trend of public opinions.

The remainder of the paper is organized as follows: Section 2 introduces related works on lexicons construction and sentiment classification. Section 3 introduces the main process of lexicons construction. The conducted experiments and the results are discussed and analyzed in in Section 4. Section 5 concludes the paper.

2 Literature review

2.1 Social emotional lexicons building

Intelligent sentiment analysis in texts has attracted considerable attention in recent years [2, 4, 7, 16, 34]. Most of the approaches developed to classify texts or sentences as positive or negative rest on a very specific kind of language resource: emotional lexicons. These lexicons contain words tagged with their affective valence (also called affective polarity or semantic orientation) that indicates whether a word conveys a positive or a negative content. Traditional construction of sentiment and emotion lexicons are based on machine learning approaches where each term is represented with a binary label/polarity [23]. To build these resources, several automatic techniques have been proposed. Some of them are based on dictionaries and lexical databases [12, 13, 17, 33, 35, 39, 41].

Kalamatianos et al., proposed and investigated the use of emotion lexicon-based methods as a means of extracting emotion/sentiment information from social media. They examined the topic of emotion analysis using a emotion lexicon, providing a benchmark dataset (manually rated by two humans) together with baseline performance of several simple and efficient algorithms and a New Greek Emotion Lexicon. But automated emotion results of research seem correlate to real two rater's emotion. At this point, their approach depend on humans manual emotion rating, which impact the effectiveness and expansibility of emotion lexicon [13].

Kušen et al. used two methods to build an emotional lexicon: one was to use the marked corpus to calculate the emotional polarity and intensity of words, and the other was to extent an existing lexicon based on the corpus in the relevant field and by learning a classifier [17]. Wang et al. extended the commendatory and derogatory terms in the benchmark lexicon using the pointwise mutual information method and used the semantic polarity algorithm to analyze textual emotions [39].

Yang used a neural network language model to perform statistical training on large-scale Chinese corpora, proposed a multi-dimensional emotional lexicon construction method based on switch constraint sets and constructed a multi-dimensional Chinese emotional lexicon – SentiRuc, which contained 10 emotion annotations and has achieved great effects [41]. Sun et al. constructed a deep conditional random field model and used it in combination with the web dictionary to identify the new emotional words and their emotional tendency. Experiments showed that the model had an emotional recognition accuracy rate of 70% [35]. Jiang et al. used the Word2Vec tool to extend the benchmark lexicon for the Weibo corpus through the incremental learning method, and then generated the final emotional lexicon by applying the HowNet lexicon and manual screening. Experiments showed that the lexicon-based sentiment analysis method is better than the SVM-based one [12].

One of the main advantages of the corpora techniques is that they can build lexicons that are tailored to a specific application simply by using a specific corpus [6]. However, these approaches often produce over-specified and incomplete features and are both time-consuming to define and require extensive domain knowledge [3, 33].

2.2 Emotion computing

While studying emotions in social networks platform is an important research topic, it has also proven to be a challenging task due to the complexity and the ambiguity of natural language expressions [36, 40]. In recent years, there has been a growing number of studies that utilize sentiment analysis tools to study the opinions of OSN users by analyzing written cues (texts) and multimedia content that people share online. While the scientific literature includes plenty of works focusing on polarizing positive and negative sentiments, identifying individual emotions has generally been understudied so far [40].

There are two main categories of emotion classification calculations for Weibo: supervised classifier learning and unsupervised emotion-lexicon-based methods [21]. Most supervised methods use machine learning technologies, and by extracting text features, construct classifiers such as neural networks [9, 33], support vector machines [28], and Naive Bayes [25] to classify different emotions. However, such classification methods based on machine learning can be easily affected by training corpora, and some algorithms involve complicated parameter settings, so they are not suitable for modelling [24]. For this reason, scholars proposed unsupervised methods based on emotional lexicon. Such methods determine the emotion categories according to the emotional tendency of the terms in the text in combination with

the elements like meaning, semantics and syntactic structure in the text. For example, Jiang et al. established a Weibo-oriented social emotional lexicon based on the Weibo corpus of social hot issues on that platform. Through the comparison of the results of the emotion-lexicon-based sentiment analysis and SVM-based one, they validated the effectiveness of the emotional lexicon and sentiment analysis method proposed [12, 33].

Montejo-Ráez et al., by using the lexicon-based WordNet graph structure, proposed a sentiment analysis method combining emotion computing with random walk. The experimental results show that this method has higher accuracy and recall rate than the traditional SVM method [24]. However, lexicon-based methods have problems like their reliance on lexicon coverage, need for large-scale corpora and large computing volume [38]. Tsakalidis et al. used a sentiment lexicon-based analysis on tweets, among other methods, to predict the results of the 2014 European Union elections, by assigning a polarity value (positive or negative sentiment) to every tweet. Then, they combined these results with a fusion of various classification approaches, based on different features of tweets [38].

Previous studies have pointed to the importance of identifying actual emotions rather than sentiment polarities, stating that two emotions belonging to the same affective valence might induce different reactions and lead to different decisions [17]. However, in order to improve the validity of emotion classification, these techniques need too many manual interventions to construct emotion lexicon, which results inefficient in emotion computing [1, 3, 32]. And the above emotion classification methods often experience feature sparseness when extracting textual features from large-scale short and random posts on Weibo. In addition, the existing emotion classification methods focus on only three polarities – positive, negative and neutral emotions, whose emotional granularity is not fine enough, making it difficult to fully portray the whole evolutionary process of the emotions on Weibo [20, 30]. Therefore, this paper attempts to propose a Weibo-motor computing method based on dependency parsing to improve the accuracy and computational efficiency of emotion classification.

3 Construction of the Weibo emotional lexicon

3.1 Construction of the benchmark emotional lexicon

From the Chinese emotional vocabulary ontology library developed by Dalian University of Technology (covering 27,466 emotional words), the author chose emotional words that have appeared in the Weibo corpus for over 500 times and established a benchmark emotional lexicon.

3.2 Extension of the benchmark emotional lexicon

The author used the Word2Vec tool to convert the original Weibo posts into word vectors and mapped them into the vector space, calculated the cosine similarity of two word vectors to obtain the correlation between the two words and then extended them to the benchmark lexicon.

$$\text{Cos}(W_1, W_2) = \frac{\sum_{i=1}^n W_1 * W_2}{\sqrt{\sum_{i=1}^n (W_1)_2 * \sum_{i=1}^n (W_2)_2}} \quad (1)$$

where, W_1 and W_2 are two word vectors, in the n -th dimension.

Then, the author used the SO-PMI algorithm to calculate the pointwise mutual information value between a new word and a benchmark one, so as to obtain the emotional tendency of the new word.

$$SO-PMI(W) = \sum_{pword \in Pwords} PMI(W, pword) - \sum_{Nword \in Nwords} PMI(W, nword) \quad (2)$$

where, W denotes the emotional word requiring calculation of its emotional polarity; $pword$ represents the word with positive emotions among the seed words $Pwords$; and $nword$ represents the word with negative emotions among the seed words $Nwords$.

3.3 Construction of the Weibo emotional lexicon

The author used an incremental iterative approach to continuously extend the benchmark lexicon and finally obtained a complete Weibo emotional lexicon. The entire construction process of the Weibo emotional lexicon is as following:

Step 1. Data preparation: acquire the data required for research from the Internet platform, including the comment data from the stock forum and stock market performance data, of which, the latter is obtained from the Internet through a crawler program. This study intends to write a crawler program for its own use in the Python language to obtain the comment data on a specific company within a certain period of time in the stock forum. The stock market performance data can be directly acquired from various types of financial services. All these data form the database required for this study.

Step 2. Sentiment computation

- Generate word vectors: use the word segmentation toolkit to segment textual data and remove the stop words, and generate word vectors in a large-scale corpus through incremental training.
- Construct the sentiment dictionary. By referring to the method proposed by Jiang et al. (2015), this paper adopts the method based on the emotional lexicon ontology and the deep learning algorithm Word2Vec. The specific process flow is shown in Fig. 1.

The specific process to construct the sentiment dictionary is as follows:

- (i) Data acquisition and pre-processing: capture the post and comment data published during a certain period of time at the stock forum, remove the @ flags and user names, and store them in the database. For the obtained text data, use the word segmentation program to segment the words and remove the stop words such as punctuation marks, numbers, names of people, place names and modal particles, thereby obtaining a Chinese sentiment dictionary corpus for the financial field.
- (ii) Selection of candidate sentiment words: based on the sentiment vocabulary ontology, the text is divided into seven categories of sentiments such as anger, aversion, fear, joy, happiness, sadness, and shock; after the sentiment categories are determined, count the frequencies of the words in each sentiment category in the sentiment lexical ontology and

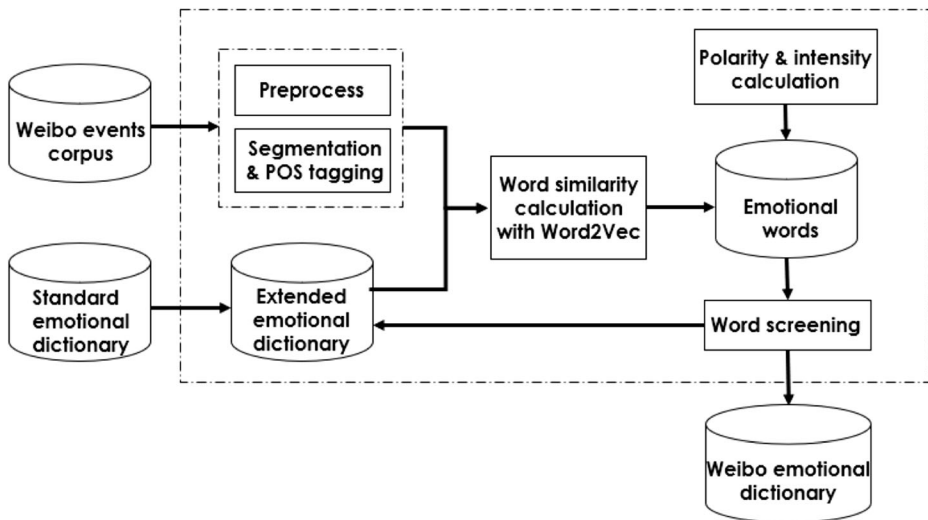


Fig. 1 Construction process of the Weibo emotional lexicon

then select the benchmark sentiment word for each sentiment category according to frequency so as to form a benchmark sentiment dictionary.

- (iii) Calculation of word similarity: use Word2Vec to vectorize the words, use the similarity of the text vector space to represent the semantic similarity of the text, and adopt the incremental iterative process to achieve the extension of sentiment words.
- (iv) Filtering of the sentiment words: use the HowNet dictionary to calculate the similarity between the newly added sentiment words and the benchmark ones and filter out the words with high similarity, and then conduct manual screening, so as to obtain the benchmark sentiment dictionary for the next step.
- (v) Generation of the sentiment dictionary: after continuous calculation of the word similarity and filtering of sentiment words, when the difference between the sentiment dictionaries in adjacent two times is less than a certain threshold, the repeated process is terminated, and the final financial sentiment dictionary is obtained.

Step 3. Calculation of the textual sentiment value

By reference to the method proposed by Wan et al. [38], use the “Language Technology Platform (LTP)” [8] to calculate the sentiment value of the text. The specific process of textual sentiment calculation is described as follows (in Fig. 2):

- Generate syntactic dependencies: use the collected forum comment data as the input to the LTP platform.

The LTP platform automatically performs operations such as word segmentation, part-of-speech tagging and syntax analysis on the input text, and gives corresponding results. The dependency relationship is represented by a directed arc, whose direction is always from the core word to the modifier, marked with a dependency name above it.

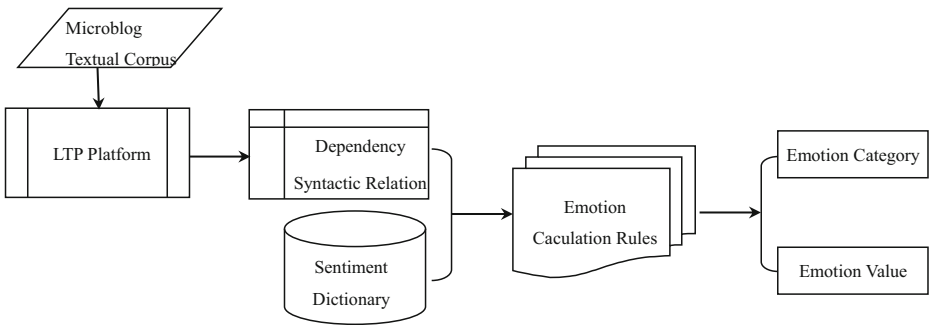


Fig. 2 Flow chart of the textual sentiment value calculation

- Calculate the textual sentiment value: the sentiment value of a word comes from the financial sentiment dictionary. According to the result of the LTP dependency syntax analysis, the sentiment value is calculated according to different rules. If there is an emoticon, it will be treated as an imperative sentence, and assigned with a weight to express its degree of modification with respect to the sentiment of the entire sentence.
- Calculate the category of textual sentiment: replace the sentiment value of the word in the sentiment calculation rule with the sentiment category, so that the textual sentiment category can be calculated. For each piece of textual data, there will be scores of seven sentiment categories, and the category with the highest score should be the sentiment category of the corresponding text.

Step 4. **Classification of fine-grained sentiments**

Based on the calculation of the textual sentiment value, this paper uses the classical algorithm LSTM of RNN to classify the sentiments of the text. The details of our proposed framework are illustrated in Fig. 3.

The improved RNN-ANN model is further fine-tuned using a specific multi-label training dataset. Each Microblog in the training dataset is represented by the semantic compositionality over the word embeddings. The loss function is based on the assumption that the labels belonging to an instance should be ranked higher than those not belonging to that instance. Finally, the trained RNN-ANN model is employed to classify multiple emotion in the test dataset. The specific process of fine-grained sentiment classification is as follows:

- Generation of word vectors: segment the words and remove the stop words in the comment data at the stock forum, and then call the module used to generate word vectors in

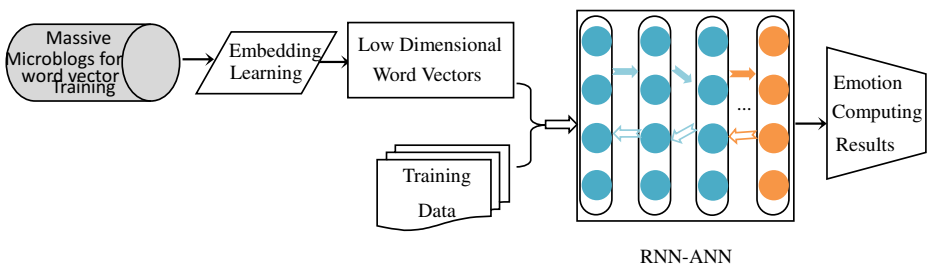


Fig. 3 Flow chart of fine-grained sentiments classification

word2vec to set the required word vector dimensions, and then the required word vectors can be obtained.

- RNN classification: RNN is a kind of neural network with fixed weights, external input and internal state.

As shown in Fig. 4, the basic model includes embedding layer X , hidden layer H , output layer Y and classification layer P four layers. The embedding layer $X = \{x_1, x_2, \dots, x_n\}$ represents the word embedding of each token in a microblog sentence. The hidden layer $H = \{h_1, h_2, \dots, h_n\}$ is a recurrent layer, where each node in H is a recurrent unit, and the j th node h_j is determined by the embedding vector x_j and last recurrent unit h_{j-1} . The output layer $Y = \{y_1, y_2, \dots, y_n\}$ is the output of RNN, the last layer is the classification layer P , which is a logistic classification. RNN applies the same set of weights recursively as follow:

$$h_t = f(Wx_t + Uh_{t-1} + b) \quad (3)$$

where f is the nonlinear active function (e.g., \tanh and sigmoid function); W , U , V are weight matrices, input vector $x_t \in R^n$ at time step t is calculated based on a hidden state, and b is bias vectors.

- ANN classification: In order to use sentiment value as an input value for the classification, a new three-layer artificial neural network (ANN) classifier is constructed. The input layer includes the RNN output (a 100-dimensional vector) and the sentiment value of the text (a real number). To avoid confusion brought by dimension, this paper adds 0 after the textual sentiment value to make it a 100-dimensional vector, too; output value of the model is the sentiment category number of the text. The intermediate layer of the ANN classifier is a hidden layer, which is set to be 50-dimensional. The process of ANN training is to make the precision of the classifier reach the global optimum by error back propagation through continuous iteration and optimization of the weight matrix of different layers.

The ANN used in this experiment uses a fully connected layer to connect adjacent layers. The output of the output layers (including the hidden layer and the final output layer of the model) is shown in Eq. (1).

$$o_j = \sum_{i=1}^I x_i \times w_{ij} \quad i = 1..100, j = 1..100 \quad (4)$$

In Eq. (4), o_j is the output of the model, x_i is the input of the model, and w_{ij} is the weight between the i -th node of the input layer and the j -th node of the output layer. In this experiment, x_i is a vector composed of the sentiment value of certain text and the RNN output, and o_j represents the corresponding sentiment category of the text. For the input training set data, the model will give its own output value, and at the same time, calculate the error between the output value and the real value to reversely optimize the parameters to be trained in the model, including the weight matrix and the excursion matrix.

- Comparison of classification results: the classification results of the two classifiers are compared in terms of classification precision, iteration times and utilization rate of computing resources to determine whether the textual sentiment value calculated above is helpful to the sentiment classification in this paper.

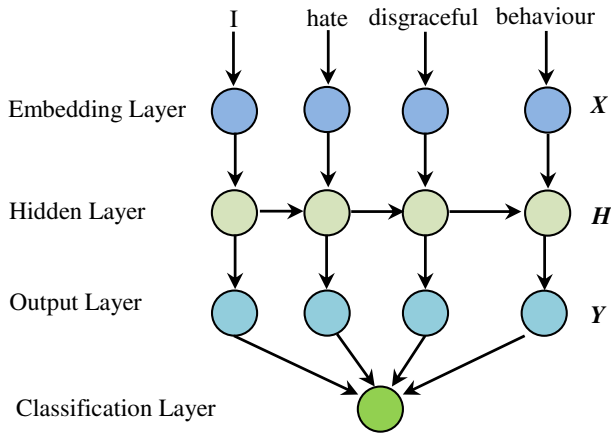


Fig 4 Basic model of RNN-based emotion classification

Step 5. Netizen sentiment index

After obtaining the textual sentiment value of each comment, calculate the daily netizen sentiment index for subsequent econometric modelling so as to study the interactions between netizens' sentiments and stock market performance.

The netizen sentiment index is divided into two categories – the weighted sum index of netizen sentiment value and the average mean index of netizen sentiment value, with a prefix of sum and avg., respectively. In order to better study the sentiments, the netizen sentiment index is subjected to three-day moving average and five-day moving average processing, and 6 proxy variables of netizens' sentiments are obtained.

Step 6. Modelling

Establish multiple models to determine the causal relationship between the netizen sentiment index constructed in this paper and the stock market performance of the listed company, and analyze the specific form of impact of the netizen sentiment index on the stock market performance of the listed company.

In the first iteration, the candidate emotional lexicon is the benchmark emotional lexicon. After inputting the pre-processed Weibo corpus, the author used the Word2Vec tool to calculate the similarity between words. The number of word vector dimensions in this paper was set to 200, and the number of word neighbouring windows was set to 6. The CBOW algorithm was used to calculate the word vectors, and the negative sampling algorithm for training. The benchmark lexicon can be extended by calculating the five words that are most similar to the benchmark emotional words. The extended words constitute the extended emotional lexicon. According to the scale of the Weibo corpus, the author carried out 8 iterations and finally obtained an emotional lexicon consisting of 4872 emotional words. Table 1 shows the number of emotional words in each emotion category in the resulting emotional lexicon:

As can be seen, the distribution of word count in the extended emotional lexicon is almost the same as that in the benchmark emotional lexicon.

Table 1 Distribution of emotional word count in the extended emotional lexicon

Emotion category	Word count	Emotion category	Word count
Happy	440	Good	1509
Sad	440	Angry	68
Fearful	195	Astonished	51
Disgusted	2169		

The author used the NLP&CC2013 Chinese Weibo sentiment analysis test set [27] to test the Chinese emotional vocabulary ontology library and the emotional lexicon constructed in this paper, respectively. The NLP&CC2013 dataset contains 10,000 pieces of corpus data with emotion annotations, and the emotion annotation categories are the same as those of the Chinese emotional vocabulary ontology library. Table 2 shows the performance of the two emotional lexicons in the data set:

It can be seen that the emotional lexicon constructed in this paper is better than the Chinese emotional vocabulary ontology library in all three indices, showing it has obvious performance advantages.

4 Emotion computing and public opinion evolution analysis

4.1 Weibo emotion computing based on dependency parsing

4.1.1 Data cleaning and pre-processing

The method uses the Python crawler to search and collect the corresponding Weibo posts and comment data using keywords, and imports the data into the database for storage; deletes the replies, reposts and the strings with reference to others, the format of web links, and the format of Weibo topics; and based on the LTP platform [8], calls its word segmentation interface, part-of-speech tagging interface, and dependency parsing analysis interface through the JAVA program to pre-process the posts.

4.1.2 Emotion binary tree based on dependency parsing

By referring to the dependency relationship designed by Wan et al. [38], the author selects 6 kinds of dependency relationships (ATT, ADV, COO, CMP, VOB and SBV) affecting the textural emotional tendency from the sequence of dependency relationships. For ATT, the dependency relationship where the modifiers are emotional words is retained; for ADV and CMP, the dependency relationship where the modifiers are emotional words, degree adverbs or negative adverbs; for COO, the dependency relationship where the adjectives are emotional

Table 2 Emotion classification results of the two emotional lexicons

	Accuracy	Recall	F value
Chinese emotional vocabulary ontology library	0.591	0.288	0.475
Our emotional lexicon	0.705	0.414	0.639

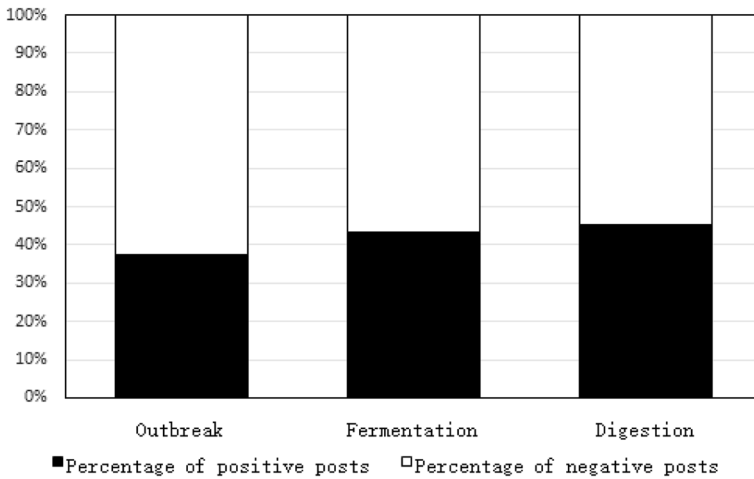


Fig. 5 Proportions of posts with positive and negative emotions in each stage

words is retained. In the filtered sequence of dependencies, the method checks whether the emotional words and modifiers in each dependency relationship are emotional words in the emotional lexicon. If so, such words are assigned with the emotional intensity corresponding to the same words in the emotional lexicon.

4.1.3 Emotion computing module based on dependency parsing

Postorder traversal is performed to the left subtree and right subtree of the emotion binary tree T. The bottom left child and right child are both the entrances for calculation of the word node. According to the emotional intensity of the left child and right child and the matching emotion computing rule, the emotion value of this node is obtained; and then according to the emotion value of this node and the emotional intensity of its brother nodes and the matching emotion computing rule, the emotion value of the parent node is obtained; and the calculation goes on upwardly like this until reaching the root node of T.

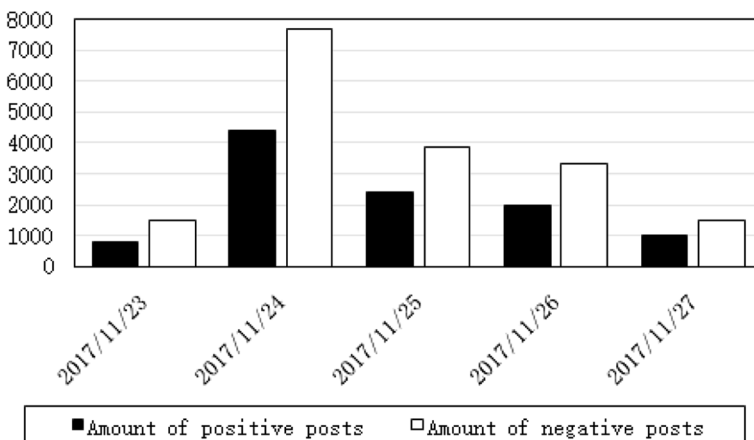


Fig. 6 Time series diagram for the numbers of positive and negative posts at the fermentation stage

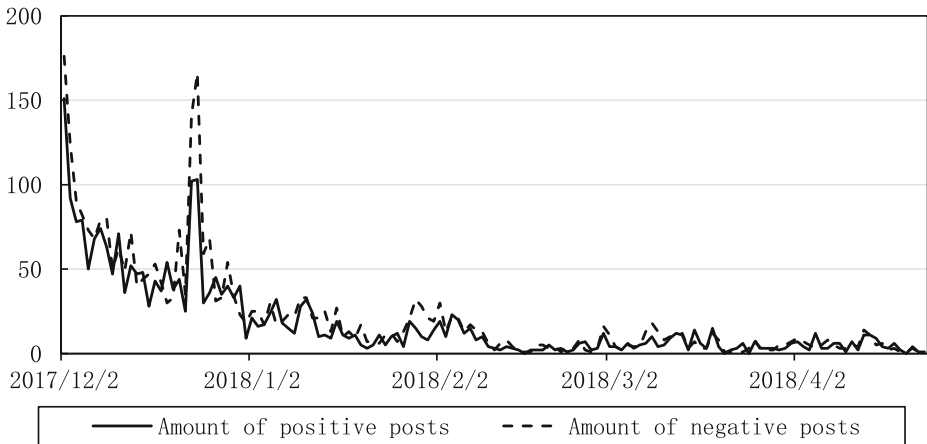


Fig. 7 Time series diagram for the numbers of positive and negative posts at the digestion stage

4.2 Public opinion evolution analysis based on textual emotion computing

Taking the “RYB Kindergarten Child Abuse Case in Beijing” as an example, the author collected a total of 97,040 posts from November 22, 2017 to April 23, 2018. Then the author used the Weibo emotion computing method based on dependency parsing to count the positive and negative emotional posts on Weibo during the three stages of Internet public opinion evolution – outbreak, fermentation and digestion and obtained the percentages of two types of emotional posts at each stage, as shown in Fig. 5.

It can be seen that there were more negative posts than positive ones at each stage, but the percentage of positive ones gradually increased with time. During the outbreak stage, due to the frequent negative reports on “RYB kindergarten child abuse” and the few measures taken by the government and the RYB organization, the percentage of negative posts was the highest (62.73%). At the fermentation and digestion stages, the government and RYB organization took more countermeasures. For example, the public security department made notifications and refuted rumours, relevant authority issued the child protection policy and the RYB organization made an apologetic statement. As a result, the percentage of negative posts gradually fell to 56.86% at the fermentation stage and 54.74% at the digestion stage. On the last day of the fermentation stage, the gap between positive and negative posts was already very small (46) (as shown in Fig. 6). At the digestion stage (as shown in Fig. 7), the gap became even smaller, and at a time there were even more positive posts than negative ones.

The author performed word segmentation, frequency statistics and sorting and extracted key words from the posts and their comments on “RYB Kindergarten child abuse” from December 24th to 25th and drew 30 keyword clouds to explore the trending topics among Internet users on these 2 days. It was found that entertainment news about the romantic disputes of celebrities and network anchors were once again centre of public attention, and that the child abuse incidents in RYB Kindergarten and Ctrip Kindergarten were being forgotten by the public. After the 25th, the number of posts about “RYB Kindergarten child abuse” dropped rapidly and no more new stimulus information popped up. At this time, the “RYB Kindergarten child abuse” was only mentioned in the news about other public events or the reviews and reflections on that incident. The RYB crisis temporarily came to an end.

5 Conclusions

Considering the deficiencies in emotional lexicon construction and emotion computing, this paper, based on a general emotional lexicon and the Weibo corpus, establishes a benchmark emotional lexicon, applies the neural network framework Word2Vec to train the corpus data and calculate the word vectors, and obtains the correlations between words by calculating the spatial distances between vectors, in order to extend the emotional lexicon. For the extended emotional words, this paper uses the SO-PMI algorithm to calculate the emotional polarity and emotional intensity, filters out unqualified emotional words according to the emotional polarity and through manual screening, and finally performs an experiment and constructs a Weibo emotional lexicon for public events. Then, this paper constructs a Weibo emotion computing method based on dependency parsing and calculates the emotional tendency and intensity of Weibo posts using the dependency-parsing-based emotion binary tree and the dependency-based emotion computing rule. At last, this paper takes a real event as an example and analyzes the public opinions in each stage and their abnormal peaks according to the emotional tendency of the Weibo posts and the high-frequency keywords in specific periods.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 71874215), Beijing Natural Science Foundation (9182016), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (17YJAZH120), and Beijing's Philosophical and Social Science Foundation (Grant No. 13JGC128, 13JGB058). We wish to thank the anonymous reviewers who helped to improve the quality of the paper. The authors gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Badaro G, Jundi H, Hajj H, El-Hajj W (2018) EmoWordNet: Automatic Expansion of Emotion Lexicon Using English WordNet. In: the Seventh Joint Conference on Lexical and Computational Semantics, p 86–93
2. Bandhakavi A, Wiratung N, Massie S, Deepak P (2016) Emotion-corpus guided lexicons for sentiment analysis on Twitter. In: Proceedings of the International Conference on Innovative Techniques and Applications of Artificial Intelligence. Springer, Cham, p 71–85
3. Bandhakavi A, Wiratunga N, Padmanabhan D, Massie S (2017) Lexicon based feature extraction for emotion text classification. *Pattern Recogn Lett* 93:133–142
4. Bandhakavi A, Wiratunga N, Massie S (2018) Emotion-aware polarity lexicons for Twitter sentiment analysis. *Expert Syst*:e12332
5. Bestgen Y (2008) Building affective lexicons from specific corpora for automatic sentiment analysis. In: Proceedings of LREC, Trento, Italy, p 496–500
6. Bestgen Y (2008) Building affective lexicons from specific corpora for automatic sentiment analysis. In: International Conference on Language Resources and Evaluation, Lrec 2008, Marrakech, Morocco 24(1):496–500
7. Buechel S, Hahn U (2018) Representation mapping: a novel approach to generate high-quality multi-lingual emotion lexicons. *arXiv preprint arXiv:1807.00775*
8. Che W, Li Z, Liu T (2010) LTP: a Chinese language technology platform. In: The 23rd International Conference on Computational Linguistics: Demonstrations. Association for Computational Linguistics, p 13–16
9. Ghiassi M, Skinner J, Zimbra D (2013) Twitter brand sentiment analysis: a hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Syst Appl* 40(16):6266–6282
10. Guan T (2018) Framing the Boundary of Sino-Japanese Conflicts in China's Communication Sphere: a Content Analysis of the News Coverage of Japan and Sino-Japanese Controversies by the People's Daily between 2001 and 2015. *J Chin Polit Sci*:1–16

11. Guo SJ (2017) The 2013 Boston marathon bombing: publics' emotions, coping, and organizational engagement. *Public Relat Rev* 43(4):755–767
12. Jiang S, Huang W, Cai M, Wang L (2015) Building social emotional lexicons for emotional analysis on microblog. *Journal of Chinese Information Processing* 29(06):166–171
13. Kalamatianos G, Symeonidis S, Mallis D, Arampatzis A (2018) Towards the creation of an emotion lexicon for microblogging. *Journal of Systems and Information Technology*, (just-accepted)
14. Kuang W (2018) Empirical studies on new-media public opinion. In: *Social Media in China*. Palgrave Macmillan, Singapore, p 257–261
15. Kušen E, Strembeck M (2018) Politics, sentiments, and misinformation: an analysis of the twitter discussion on the 2016 Austrian presidential elections. *Online Soc Netw Med* 5:37–50
16. Kušen E, Strembeck M, Cascavilla G, Conti M (2017) On the influence of emotional valence shifts on the spread of information in social networks. In: *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. ACM, p 321–324
17. Kušen E, Cascavilla G, Figl K, Conti M, Strembeck M (2017) Identifying emotions in social media: Comparison of word-emotional lexicons Identifying emotions in social media: comparison of word-emotion lexicons. In: *5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, IEEE, p 132–137
18. Lecheler S, Bos L, Vliegthart R (2015) The mediating role of emotions: news framing effects on opinions about immigration. *J Mass Commun Q* 92(4):812–838
19. Lee J, Choi Y (2018) Understanding social viewing through discussion network and emotion: a focus on South Korean presidential debates. *Telematics Inform* 35(5):1382–1391
20. Li X, Li J, Wu Y (2015) A global optimization approach to multi-polarity sentiment analysis. *PLoS One* 10(4):e0124672
21. Liang HF, Shi F, Ling WD, Ge YU (2017) Mining topic sentiment in microblogging based on multi-feature fusion. *Chin J of Comput* 40(4):872–888
22. Lin D, Li L, Cao D, Lv Y, Ke X (2018) Multi-modality weakly labeled sentiment learning based on explicit emotion signal for Chinese microblog. *Neurocomputing* 272:258–269
23. Liu B, Zhang L (2012) *A survey of opinion mining and sentiment analysis*. In: *Mining text data*. Springer, Boston, p 415–463
24. Montejo-Ráez A, Martínez-Cámara E, Martín-Valdivia MT, Ureña-López LA (2014) Ranked wordnet graph for sentiment polarity classification in twitter. *Comput Speech Lang* 28(1):93–107
25. Narendra B, Sai KU, Rajesh G, Hemanth K, Teja MC, Kumar KD (2016) Sentiment analysis on movie reviews: a comparative study of machine learning algorithms and open source technologies. *Int J Intell Syst Technol Appl* 8(8):66–70
26. Nip JY, Fu KW (2016) Challenging official propaganda? Public opinion leaders on Sina Weibo. *China Q* 225:122–144
27. NLP&CC2013. http://tcci.ccf.org.cn/conference/2013/pages/page04_tdata.html
28. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up?: sentiment classification using machine learning techniques. In: *the ACL-02 Conference on Empirical Methods in Natural Language Processing*. Philadelphia, PA, USA: Association for Computational Linguistics, 10:79–86
29. Peng H, Cambria E, Hussain A (2017) A review of sentiment analysis research in Chinese language. *Cogn Comput* 9(4):423–435
30. Peng H, Ma Y, Li Y, Cambria E (2018) Learning multi-grained aspect target sequence for Chinese sentiment analysis. *Knowl-Based Syst* 148:167–176
31. Poria S, Peng H, Hussain A, Howard N, Cambria E (2017) Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis. *Neurocomputing* 261:217–230
32. Reka M, Srividhya V (2018) Emotion classification of twitter data using lexicon based approach. *Softw Eng Technol* 10(4):69–71
33. Stojanovski D, Strezoski G, Madjarov G, Dimitrovski I, Chorbev I (2018) Deep neural network architecture for sentiment analysis and emotion identification of Twitter messages. *Multimed Tools Appl*:1–30
34. Sun X, Zhang C, Li G, Sun D, Ren F, Zomaya A, Ranjan R (2017) Detecting users' anomalous emotion using social media for business intelligence. *J Comput Sci* 25:193–200
35. Sun X, Peng X, Hu M, Ren FJ (2017) Extended multi-modality features and deep learning based microblog short text sentiment analysis. *J Electron Inf Technol* 39(9):2048–2055
36. Tubishat M, Idris N, Abushariah MAM (2018) Implicit aspect extraction in sentiment analysis: review, taxonomy, opportunities, and open challenges. *Inf Process Manag* 54(4):545–563
37. Vermeulen A, Vandebosch H, Heirman W (2018) #smiling, #venting, or both? Adolescents' social sharing of emotions on social media. *Comput Hum Behav* 84:211–219
38. Wan C, Jiang T, Zhong M, Bian HR (2013) Sentiment computing of web financial information based on the part-of-speech tagging and dependency parsing. *J Comput Res Dev* 50(12):2554–2569

39. Wang Q, Lin Z, Jin Y, Cheng S, Yang T (2015) ESIS: emotion-based spreader-ignorant-stifler model for information diffusion. *Knowl-Based Syst* 81:46–55
40. Yadollahi A, Shahraki AG, Zaiane OR (2017) Current state of text sentiment analysis from opinion to emotion mining. *ACM Comput Surv* 50(2):25
41. Yang XP, Zhang ZX, Wang L (2017) Automatic construction and optimization of sentiment lexicon based on Word2Vec. *Comput Sci* 44(01):42–47
42. Yu L, Li L, Tang L (2017) What can mass media do to control public panic in accidents of hazardous chemical leakage into rivers? A multi-agent-based online opinion dissemination model. *J Clean Prod* 143:1203–1214



Wei Zhang is an associate professor of School of Information at Central University of Finance and Economics. He received B.S. in computer sciences in 1998 and M.S. in management sciences in 2001 from Shandong University of Science and Technology, Ph.D. in system engineering in 2005 from Beihang University, respectively. He worked at Research Institute of Chinese Internet economics as a researcher from 2008 to present and at National University of Singapore in 2013 as a visiting scholar. As the principal investigator of the intelligent emotion computing project, he enhanced the main idea and did experiment for the proposed scheme. His research areas are big data, sentiment analysis, and intelligent textual corpus computing.



Yan-chun Zhu is an associate professor of Business School at Beijing Normal University. She got the B.S. Degree in management sciences and engineering in 2002, the M.S. Degree in management sciences and engineering in 2004, Ph.D. in management sciences and engineering in 2007 from Beihang University, respectively. She took part in the intelligent emotion computing project as a researcher. As the research result, she proposed the idea and wrote the first version of this paper. Her research areas are textual corpus computing, machine learning and big data.



Jia-peng Wang got the B.A. in E-commerce from Central University of Finance and Economics in 2018. As a undergraduate student, he took part in the intelligent emotion computing project. As the research result, he wrote the big data computing experimental part of this paper. His research areas are machine learning and intelligent e-commerce transaction systems.