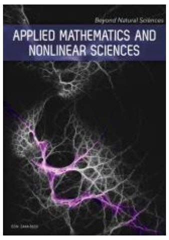
Applied Mathematics and Nonlinear Sciences, 9(1) (2024) 1-15



Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

Optimization Model Construction of Online Public Opinion Dissemination Based on Behavioral Data Mining

Sijia Yao[1,†]

1. Hunan University of Information Technology, Changsha, Hunan, 410151, China.

Submission Info

Communicated by Z. Sabir
Received January 12, 2023
Accepted May 10, 2023
Available online October 30, 2023

Abstract

In this paper, we propose an optimization method based on behavioral data mining for the emergence of negative emotions in network public opinion in unexpected situations leading to problems that endanger social stability and use data mining to confirm the correspondence of the kNN algorithm. In the opinion propagation model, the radius of opinion radiation is assumed to use the network node density and distribution density as features of node exchange information. Then, the kNN algorithm is used to train the comment set for the analysis of user sentiment evolution of network opinion in the social network environment, and the web crawler technology is used to obtain the output interface data APIs of microblogs and WeChat, and the social media user comment data is used as the data for the empirical analysis of network opinion. snowNLP and kNN algorithms are used to analyze the sentiment score of network opinion and the sentiment score less than 0.5, i.e., the sentiment polarity. There were 15 days when the sentiment score tended to be negative and 65 days when the sentiment score was greater than 0.5, i.e., the sentiment polarity tended to be positive.

Keywords: Online public opinion; Behavioral data mining; kNN algorithm; Web crawler technology; Sentiment evolution.
AMS 2010 codes: 97P33

†Corresponding author.
Email address: scarlettyao2020@163.com

ISSN 2444-8656
https://doi.org/10.2478/amns.2023.2.00892

1 Introduction

The 21st century is the era of rapid development of modern technology, and in daily life, the influence of the Internet is everywhere and has long become the mainstream media nowadays. With the advent of the era of self-media, every individual can publish information, express their opinions, disseminate information, etc., on microblogs, WeChat, short videos and other APP platforms [1-2]. The collection of public attitudes, emotions, and behavioral tendencies toward various social affairs is known as public opinion, and it possesses the characteristics of clarity, difference, intertwining, dependence, neglect, and extensiveness [3]. Social networks have become the primary carrier of public opinion due to the development of mobile Internet technology. In a broad sense, online public opinion is a variety of emotions, opinions and attitudes of various social groups existing on online platforms; in a narrow sense, online public opinion is "the socio-political attitudes generated and held by the public, as the subject of public opinion, toward state administrators around the occurrence, development and changes of social matters in a certain Internet space" [4-6]. In addition to the general characteristics of public opinion, the special nature of its reliance platform also makes online public opinion have other characteristics [7].

Online public opinion is another expression of social opinion on the Internet platform reflecting public sentiment and public opinion, and it has the characteristics of fast dissemination, wide dissemination, and a large dissemination base [8-9]. When a hot event occurs in society, the amount of data generated by online public opinion may grow exponentially and cannot be ignored [10-11]. If the government fails to control it promptly, it may result in an uncontrollable situation and even lead to public instability and social unrest [12]. For this reason, the regulation and disposal of online public opinions have become a topical social issue with practical significance today.

The literature [13] investigated the influence of influencing factors in different dimensions on disseminating public opinion information in social media. By selecting multidimensional factors, including information characteristics, communication network structure, and user-level attributes, an econometric model was applied to analyze them. The key factor in information dissemination is not influencing traditional information characteristics and source attributes but possibly information channels. The issue of similarity between the epidemic dissemination model and the sentiment analysis and dissemination model on social media was investigated in depth in the literature [14]. The method was verified to predict the development trend of public opinion effectively. The use of deep learning techniques allows for insight into the core issues of topic and sentiment analysis from the perspective of public opinion analysis. The literature [15] examined the use of two new node states to enhance the effect of information diffusion in a coupled social network environment for information dissemination. An improved SIR model is used to participate in independent extenders and cross-network extenders, whose participation in information expansion, using independent booths as sources of information, yields better diffusion probability and wider diffusion range. The improved SIR model provides a good fit and a more convenient way to interpret the data to visualize information diffusion in the case of cross-network dissemination of public opinion.

The literature [16] investigated a sentiment tendency analysis model by incorporating a lexicon of commonly used sentiment words and expressions into a lexicon using the positive and negative sentiment lexicon of the HowNet Sentiment Dictionary. By creating a sentiment lexicon and proposing a sentiment tendency analysis model, online opinion participants' sentiment characteristics can be explored effectively to gain insight into their behavior. The process of user opinion fusion under interest and trust thresholds was explored in the literature [17]. A new HKSEIR model of opinion evolution was constructed using a combination of opinion fusion HK and epidemic communication SEIR models. The model considers topic interest degree and analyzes the interaction behaviors among users under interest and confidence thresholds to calculate the probability of topic propagation caused by inter-user opinion interaction under group pressure and the probability of users changing from infection to deletion status under topic epidemic. Online users' information data is collected and processed based on the behavioral characteristics of online opinion participants. When users' opinion fusion interactions reach a certain level, users' interests become the primary factor in shaping public opinion.

The literature [18] studied the topics of public opinion in the Macromedia network and revealed the negative topics being discussed on the Macromedia network. The multidimensional network model of the topology of online public opinion for self-media was constructed using the characteristics of the opinion elements in the "micro media" network as a reference for classifying the opinion elements. The multidimensional network model can effectively describe the communication of multiple topics on the micro-media network. The key nodes of public opinion evolution on social media were examined in the literature [19]. A new approach is proposed to construct a framework for analyzing public opinion in terms of the temporal pattern of online information, the evolution of frames, and the evolution of users through an innovative combination of content analysis and some curve-fitting methods. Social events have a double diffusion cycle, and a completely different mechanism forms each cycle, while the characteristics of online opinion diffusion can be depicted by the key parameters of the bimodal Gaussian model.

In this paper, firstly, we put forward hypotheses on the public opinion dissemination model and study the density distribution of network nodes in a certain communication range where communication nodes exchange information. The optimization method of behavioral data mining is used to change the problem of negative emotions appearing in network public opinion under unexpected situations and to avoid leading to situations that endanger social stability. Secondly, an analysis model of the emotional characteristics and evolutionary laws of users of online public opinion in social networks is constructed based on the kNN algorithm. Web crawler technology is used as the data acquisition technique, the BosonNLP tool is used for word separation to construct a text matrix and text corpus, and social media user comment data is used as the data for empirical analysis of online public opinion. Then, the model parameter weights are adjusted and cross-validated characteristics according to the convolutional neural network, and the correct rate, recall rate and F1 value are tested in the test set by the convolutional layer, pooling layer and fully connected layer, and the optimal model is determined by debugging the parameters to obtain relevant data. Finally, the above data are used as kNN algorithm and snowNLP algorithm network opinion sentiment analysis models, which are used to score the sentiment of microblog comments and analyze the sentiment score of network opinion under unexpected situations.

2 Network opinion dissemination model

2.1 Modeling Construction of Public Opinion Communication

2.1.1 Assumptions of public opinion dissemination model

In this paper, we assume that within a certain area, the number of network nodes is N in a plane of area A, and assume that these N nodes are uniformly distributed in the area, thus obtaining the network node density distribution density in this finite area $p = N / A$. Assuming that the signal radiation radius of each node is r, all other nodes within a circle of radius r with the node as the center are able to receive the node's signal and can communicate with them. Assuming that node i exchanges information with at most k DTN nodes at a certain time, the maximum degree of node i is k. However, because node i is constantly moving with a certain speed v in an area of A, resulting in random or predetermined connectivity with other nodes, the degree of node i is not always the same value. At different times, there will be a different number of nodes $k^{'}\left(0 \le k^{'} \le k\right)$, $k^{'}$ being the number of nodes met by node i that happen to fall into the same communication field of node i.

In other words, only $k^{'}\left(k^{'} = 0,1,2,3,\ldots,10\right)$ nodes happen to be in the communication field of node i in a given time t period. And only the nodes that are in the communication field of node i can connect with node i and receive information. Since the actual degree of node i varies with time, and at a given time t the degree is uncertain, the degree of node i may take on a value of $1,2,\ldots,k$. Therefore, when formalizing DTNs and designing information transmission models, time is the key factor to consider, and at different times, $k^{'}\left(0 \le k^{'} \le k\right)$ has different values:

$$k^{'} = k(t) = \lfloor k \times P(t) \rfloor \tag{1}$$

Where $p(t)$ is a probability density function, $p(t)$ takes a time-dependent value, and $\lfloor k \times P(t) \rfloor$ denotes the integer part of $k \times P(t)$. Thus, if for security reasons a given node u can exchange information with up to k nodes, but due to the mobility and restricted communication range of the DTN node, at the moment t he can only have the opportunity to communicate with $k^{'}\left(0 \le k^{'} \le k\right)$ of them, the node communication is schematically shown in Figure 1. Assuming that node i moves with speed v in an area A and can detect other nodes in a radius r with a total of N nodes in A, it is possible to indicate how many nodes the node can exchange information with during its movement.

$$p = N / A \tag{2}$$

$$S = 2r \times vt + \pi r^2 \tag{3}$$

$$k = S \times \rho \tag{4}$$

Where ρ denotes the density of nodes in area A, S denotes the area covered by nodes moving at speed v from a to b in time t, and $S \times \rho$ denotes the number of nodes k that can communicate in time t.

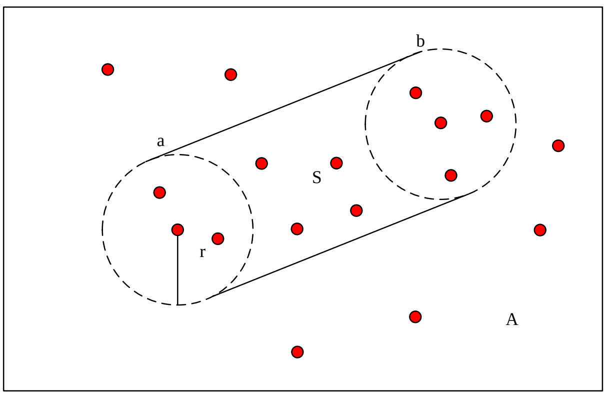


Figure 1. Schematic of node communication

2.1.2 State transition of the public opinion dissemination model

Based on the new terminology defined above and the original epidemic communication theory, the process of online opinion state transformation is shown in Figure 2.

Each of the N nodes of the network can have three states: unknown state, propagated state, and uninvited state. A node that has never received a message in the random network and is ready to receive it is in the unknown state I. A node that has received a message and is ready to disseminate it is in the propagation state S. A node that has received a message but is no longer interested in forwarding it for other reasons is in the unreceptive state R.

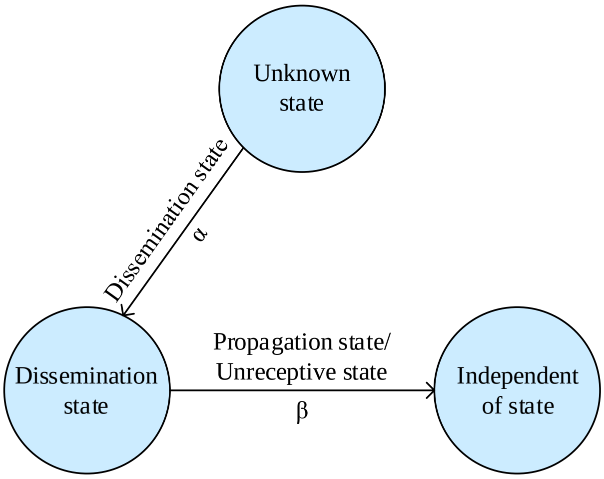


Figure 2. The process of transitioning the state of online public opinion

When an unknown state node meets a propagating state node that has securely exchanged public keys, the unknown state node can accept the information held by the propagating state node. Then the unknown state node will become a propagated state node with a probability of α. When a propagation state node meets another propagation state or non-receiving state node that has already securely exchanged public keys, the node will become a non-receiving state node with a probability of β, because it will be persuaded that all its partners have been informed of the information and therefore no further propagation is necessary. The densities of unknown, propagated, and unaccepted nodes are denoted as $i(t), s(t)$ and $r(t)$, respectively. $i(t), s(t)$ and $r(t)$ are defined as follows:

$$i(t) = N_t(t) / N \tag{5}$$

$$s(t) = N_s(t) / N \tag{6}$$

$$r(t) = N_r(t) / N \tag{7}$$

Where $N_i(t), N_s(t)$ and $N_r(t)$ are the number of nodes in the t-time unknown, propagated, and unafflicted states, respectively. In addition, the state of each node is one of I, S, and R. Therefore, the following equation (8) is the normalization condition of the equation.

$$\mathrm{i}(t) + s(t) + r(t) = 1 \tag{8}$$

The set of mean field equations with respect to the three density variations satisfies the following set of coupled differential equations:

$$\frac{\mathrm{d}i(t)}{dt}=-\alpha k(t)i(t)s(t) \tag{9}$$

$$\frac{ds(t)}{dt}=\alpha k(t)i(t)s(t)-\beta k(t)s(t)[s(t)+r(t)] \tag{10}$$

$$\frac{dr(t)}{dt}=\beta k(t)s(t)[s(t)+r(t)] \tag{11}$$

By replacing $k(t)$ with expressions (1) (2) (3) (4), the differential equations (9) (10) (11) take the following form.

$$\frac{di(t)}{dt}=-\alpha\left\lfloor\left\lfloor\left(2r\cdot vt+\pi r^2\right)\cdot\frac{N}{A}\right\rfloor\times P(t)\right\rfloor i(t)s(t) \tag{12}$$

$$\begin{aligned}\frac{ds(t)}{dt}=&\alpha\left\lfloor\left\lfloor\left(2r\cdot vt+\pi r^2\right)\cdot\frac{N}{A}\right\rfloor\times P(t)\right\rfloor i(t)s(t)\\&-\beta\left\lfloor\left\lfloor\left(2r\cdot vt+\pi r^2\right)\cdot\frac{N}{A}\right\rfloor\times P(t)\right\rfloor s(t)\left[s(t)+r(t)\right]\end{aligned} \tag{13}$$

$$\frac{dr(t)}{dt}=\beta\left\lfloor\left\lfloor\left(2r\cdot vt+\pi r^2\right)\cdot\frac{N}{A}\right\rfloor\times P(t)\right\rfloor s(t)\left[s(t)+r(t)\right] \tag{14}$$

The initial condition for the system of equations is $i(0)=(N-1)/N, s(0)=1/N, r(0)=0$. where α is the propagation rate and β is the rate of decay of the propagating state nodes into unstuck nodes. k is the maximum degree. It can be assumed that $\alpha=1$, meaning that if the unknown state DTN node u and the propagating state node v are in the same communication domain and have securely exchanged their public keys, then node u is sure to accept the information carried by node v.

From the above set of differential equations, it can be seen that $\left\lfloor\left\lfloor\left(2r\cdot vt+\pi r^2\right)\cdot\frac{N}{A}\right\rfloor\times P(t)\right\rfloor$ is in an important position in the set of differential equations. The biggest difference from previous work in the field of rumor research is that $\left\lfloor\left\lfloor\left(2r\cdot vt+\pi r^2\right)\cdot\frac{N}{A}\right\rfloor\times P(t)\right\rfloor$ varies randomly with time. When $\left\lfloor\left\lfloor\left(2r\cdot vt+\pi r^2\right)\cdot\frac{N}{A}\right\rfloor\times P(t)\right\rfloor=0$, the density probability function will remain constant during these time steps because of the lack of message receivers within that communication range.

2.2 Classification of social network opinion users

2.2.1 KNN Classification Method Framework

The kNN, also known as the neighbor algorithm, is one of the most fundamental data mining classification methods. kNN algorithm considers that its k closest neighbors can represent each sample. The algorithm assumes that k closest samples in the feature space belong to the same category, so all samples in this feature space belong to this category. kNN algorithm is based on the limit theorem by defining the category through neighboring node features8 but is only relevant to a very small number of neighboring samples in the decision category process. To determine the correlation between test set data and sample set data, an initial training sample set is defined, and the data is set to corresponding labels. After inputting the test set data, it is compared with the data in the sample set to determine the categories and corresponding features, and finally, the data in the sample set with the most similar features is determined.

Based on the emotional characteristics and evolution rules of online opinion users in the kNN social network environment, the process of building the online opinion user sentiment analysis model is shown in Figure 3:

1) Data pre-processing, word separation based on the subject comment corpus, removal of deactivated and useless words, and extraction of sentiment words in the text.

2) Constructing the sentiment polarity word list, judging the sentiment tendency, including positive, negative and neutral, judging the sentiment polarity based on sentiment analysis for the thesaurus based on the classification word list, and judging the sentiment intensity based on the sentiment intensity word list.

3) Construct a sentiment analysis model to determine the mathematical expression of sentiment polarity and sentiment degree, polarity value and sentimental value.

4) Train and test the text, train the comment set based on the kNN algorithm, train the sentiment polarity, sentiment degree, polarity value and sentiment value of the text in the test set based on the results of the training set, and analyze the evolution of user sentiment of online public opinion in the social network environment.

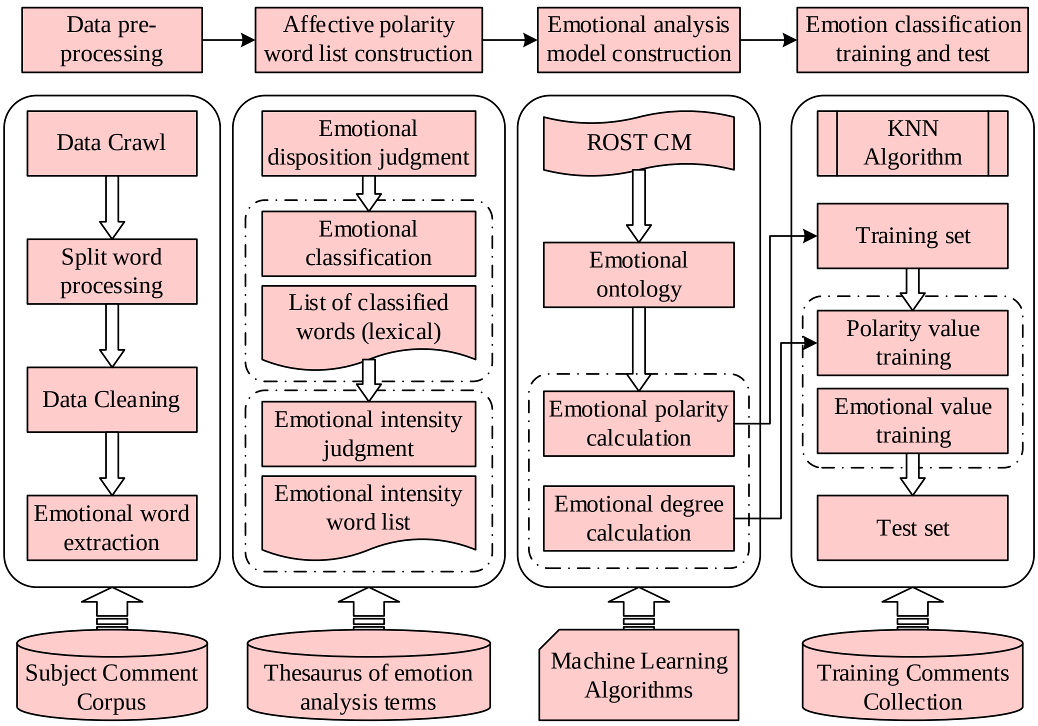


Figure 3. Analysis model of user sentiment of online public opinion

2.2.2 Affective polarity and polarity values

Emotion distribution is obtained based on the comment information of online opinion users in the social network environment, including positive emotion M, neutral emotion K and negative emotion N. The sum of the three is defined as the emotion polarity E, as in Equation (15):

$$E_0 = M_0 * 1 + N_0(-1) + K_0 * 0 \tag{15}$$

Where, M_0 represents the proportion of positive emotions M among all emotion words in the message, N_0 represents negative emotions N among all emotion words in the message, and K_0 represents neutral emotions K among all emotion words in the message.

The training of the polarity value of user comment content is shown in equation (16):

$$\begin{aligned} C_{E_0} &= E_0 * 0.6 + \sum_{i=1}^{i=E_p} W_i * 0.4 \\ &= (M_0 * 1 + N_0 * (-1) + K_0 * 0) * 0.6 + \sum_{i=1}^{i=E_p} W_i * 0.4 \end{aligned} \tag{16}$$

C_{E_0} denotes the polarity value, E_0 denotes the sentiment polarity, and W_i denotes the sentiment intensity of the i th sentiment word. The weight of 0.6 and 0.4 is set for each category of sentiment word proportion and sentiment intensity set, respectively. The training formula (17) for the sentiment value of user comment content is as follows:

$$\begin{aligned} C_{E_f} &= E_f * 0.6 + \sum_{i=1}^{i=E_p} W_i * 0.4 \\ &= \frac{5*A_x + 4*B_x + 3*C_x + 2*D_x + 1*E_x}{100} \%2 * 0.6 + \sum_{i=1}^{i=E_p} W_i * 0.4 \end{aligned} \tag{17}$$

The KNN algorithm-based sentiment analysis model of online opinion users in social network environment includes data preprocessing, sentiment polarity word list construction, sentiment analysis model construction, sentiment classification training and testing.

C_{E_f} denotes sentiment value, E_f denotes sentiment degree, and W_i denotes sentiment intensity of the i th sentiment word. 0.6 and 0.4 are set weightings.

2.2.3 User Text Classification Model Based on Convolutional Neural Network

In this paper, we construct a text topic classification model for social network opinion users based on a convolutional neural network, and the process of social network opinion user sentiment classification is shown in Figure 4.

1) Obtain microblog and WeChat output interface data APIs through web crawler technology and crawl two social media users' comment data.

2) In the data pre-processing stage, the data is cleaned, deactivated words are removed, and lexicality is manually classified and labeled.

3) Use the BosonNLP tool for word separation during the training phase of word vectors, and use Word2Vec to train Chinese word vectors. Lastly, create a text matrix and text corpus.

4) The labeled data is marked as the training set, and the model parameter weights are adjusted by the convolutional layer, pooling layer, and fully connected layer for cross-validation. The correctness, recall, and F1 values are tested in the test set.

5) Determine the optimal model by adjusting the parameters.

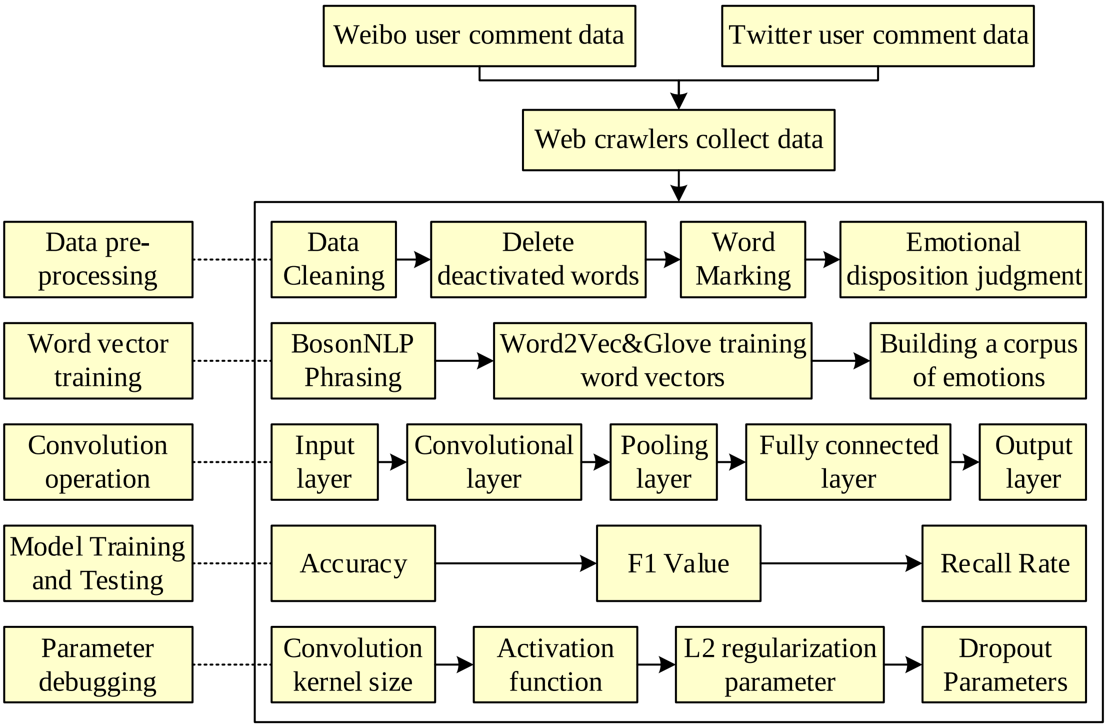


Figure 4. The process of classifying users' emotions in social network opinion

Word2Vec trains word vectors mainly based on a given corpus and expresses a word into vector form by an optimized training model with high speed and efficiency, whose core architecture consists of CBOW and Skip-gram. The HierarcicalSoftmax model for Word2vec optimization has three layers: an input layer, a projection layer, and an output layer. The goal of the training is to maximize the conditional probability of observing the actual output words in a given input context while considering the weights. The vocabulary scores are calculated for each word from the hidden layer to the output layer using the weight matrix W. Then, the posterior distribution of the words is calculated using softmax. This chapter uses AdaGrad's gradient descent algorithm to randomly sample all non-zero elements in matrix X. The learning curve is set to 0.1 and iterated 55 times on vectors smaller than 400 and 110 times on vectors of other sizes until convergence. The operation of vectors determined the semantic similarity between two words. The Weibo and WeChat user comment data were word-sorted using the NLPIR/ICTCLAS2016 tool to compare the results of Word2Vec's training word return volume.

Denote by x_i the word vector corresponding to the i nd word in a user comment message with dimension 400. Since comment messages contain different numbers of words or words, each comment content is expanded to the same length by complementary zeros so that a sentence of length n is represented as:

$$X_{1:n} = x_1 + X_2 + \ldots + X_n \tag{18}$$

The comment information can be transformed into a matrix of sentences of the same size as the model input by a vertical concatenation operation on the word vectors, i.e., summation, i.e., $X_{1:n} \in R^{n*300}$.

The selection and initialization of the filters are considered first when performing the convolution operation on the text sentence matrix. Filters $W \in Q^{kp}$, k are the dimensions of the filter, representing the number of words or words in each convolution operation, and p is the dimension of the word vector. The convolution operation is the feature vector obtained by convolving a filter with a string containing k words. When filter W convolves a particular string $X_{i,i+p-1}$, the resulting feature vector c_i expression is:

$$c_i = f\left(W * X_{i,i+k} + \theta\right) \tag{19}$$

Θ is the bias term and f is the nonlinear activation function. In the process of training word vectors, the input layer vector is denoted as $X = \left(x_0, x_1, \ldots, x_n\right)$, the intermediate layer vector is denoted as $T = \left(t_0, t_1, \ldots, t_n\right)$, the output layer vector is denoted as $Y = \left(y_0, y_1, \ldots, y_n\right)$, the weight from the input layer to the intermediate implied layer is γ_{ij}, the weight from the implied layer to the output layer is δ_{jk}, the bias threshold for the implied layer is σ, and the threshold for the output layer is ρ. Then the formulas from the input layer to the intermediate implied layer and from the implied layer to the output layer are denoted as:

$$t_j = f\left(\sum_{i=0}^{n} \gamma_{ij} x_i + \sigma\right) \tag{20}$$

$$y_j = f\left(\sum_{i=0}^{n} \delta_{ij} t_i + \rho\right) \tag{21}$$

Computational error μ_k and implied layer error μ_k are calculated by comparing output vector m_k with target vector n_k, i.e.:

$$\mu_k = \left(n_k - m_k\right) m_k \left(1 - m_k\right) \tag{22}$$

$$\mu_j = t_j\left(1 - t_j\right) \sum \mu_k \delta_{jk} \tag{23}$$

3 Empirical analysis of online public opinion based on sentiment analysis

3.1 Analysis of the sentiment score of online public opinion

On December 31, 2019, the Wuhan Municipal Health Commission notified that several pneumonia patients had recently appeared in the area and were suspected to be related to the South China Seafood Market. The notification caused immediate concern among the community. The new crown pneumonia outbreak in early 2020 was another major public health emergency since the SARS outbreak 2002. Its long duration and widespread coverage have drawn the attention of netizens. Therefore, studying the evolution of public opinion on Newcastle pneumonia is of great practical importance. In this paper, we first use snowNLP and kNN algorithms to score the sentiment of each microblog comment and calculate the sentiment score of the day based on the weighted average of the sentiment scores of all comments to observe the change in public opinion. The sentiment scores calculated by the snowNLP and kNN algorithms range from 0 to 1, and the closer to 1, the more positive the sentiment is; conversely, the closer to 0, the more negative the sentiment is. To compare the sentiment scores of the now NLP and kNN algorithms in analyzing the change of public opinion, the results of the sentiment scores of the day obtained by weighting the average of all the daily comments are shown in Figure 5.

The results show that during the 80 days from January 20 to April 8, there were 15 days when the sentiment score was less than 0.5, i.e., the sentiment polarity tended to be negative, and 65 days when the sentiment score was greater than 0.5, i.e., the sentiment polarity tended to be positive. The curve's trend shows that although the sentiment fluctuates, the overall sentiment is positive, and the trend line for the sentiment is upward.

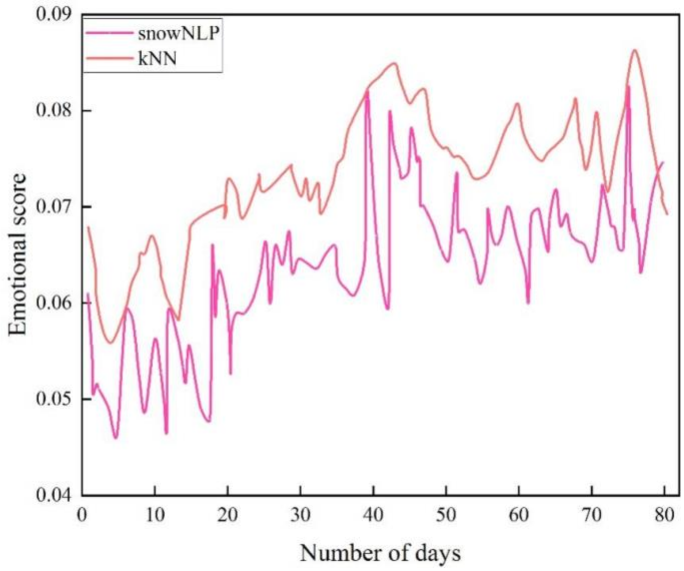


Figure 5. Change in sentiment score of microblog comments

3.2 Analysis of the evolution of public opinion themes

The sentiment analysis results permit dividing public opinion evolution into three stages. The first stage is from January 20 to February 6. The second stage is from February 7, 2020, to February 27, 2020. The content of the microblog text and the comments in these two stages will be modeled and analyzed in conjunction with the actual situation.

1) The first phase (January 20, 2020-February 6, 2020), in which Academician Zhong Nanshan first identified the “human-to-human” transmission of NCCP and new cases began to increase, focuses on the development of NCCP cases and the development of pneumonia in Wuhan, the epicenter of the epidemic. The topics during this period were the development of new cases of Pneumocystis carinii, the development of pneumonia, and policy response in Wuhan, the disease's epicenter. Netizens recognized the epidemic's seriousness, and their emotions changed dramatically, leading to a gradual increase in alertness and worry. From this phase of microblog topic mining, it can be seen that the official microblogs were able to push out the progress related to the epidemic promptly when the epidemic occurred and show the development of cases of the epidemic openly and transparently. However, the negative events

that occurred during the epidemic also caused changes in public opinion, and the negative impact brought by the negative events also required timely control and feedback from the official media. Similarly, the first-stage microblog comment text theme extraction model is constructed, and the microblog comment theme extraction results are shown in Figure 6.

From the extraction results, it can be seen that netizens' concerns are not only about the notification of the epidemic intelligence but also about the supply of materials such as masks during the epidemic, the health of medical personnel fighting on the front line, and so on. During this period, the source of the virus was discussed with a high degree of enthusiasm, and the discussion of "game" and "bats" was 0.0083 and 0.0081, reflecting the public's concern about the source of the new pneumonia outbreak and their tendency to worry and fear in the face of the sudden and unknown disease. In the face of sudden and unknown diseases, generating strong negative emotions such as worry and fear is easy. Unscrupulous individuals influenced people's emotions by rumor-mongering and trouble-making during this period.

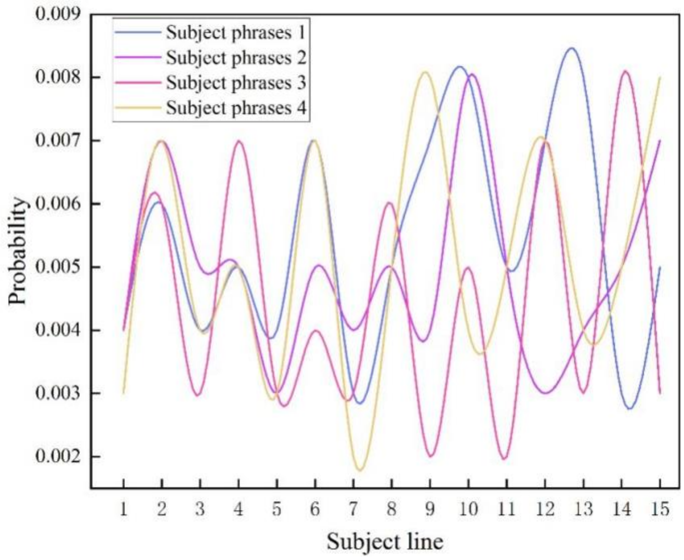


Figure 6. Statistical analysis of the subject words of microblog comments

2) The top liked tweets during this period include: "Support Hubei arrangement", "Dr. Li Wenliang's name is corrected", "The origin of the epidemic is not necessarily from China", "The Red Cross is investigating thoroughly," etc. "Red Cross investigates thoroughly," and so on. In addition to informing the public about the epidemic, the content of the microblogs at this stage responded to the hot issues that netizens were concerned about before, such as the "Red Cross incident", "Dr. Li Wenliang blowing the whistle," and the support for the seriously ill areas in Hubei besides Wuhan. This effectively solved netizens' concerns, controlled the development of public opinion in time, and enhanced the confidence of netizens. Positive microblogs related to people's lives also appeared and were approved by the public. The consistency of modeling themes for the second stage of microblog comments is shown in Table 1. The "further" indicates that public opinion has been improved, and the tension of the people has been relieved compared with the first stage.

Table 1. Consistency statistics of microblog comment topics in the second phase

Number of topics	Consistency
1	0.4825
2	0.4576
3	0.5128
4	0.4604
5	0.5283
6	0.4095
7	0.4472
8	0.5318

Netizens experienced three major stages of anxiety and nervousness: gradual recovery, confidence, and stability in response to the new pneumonia epidemic, a major public health emergency. Positive emotions were generally dominant, with more positive than negative ones. Although more emotional fluctuations occurred at the beginning of the epidemic due to fear of the unknown and the emergence of certain adverse events affected the overall mood, public opinion gradually went on the right track in the middle and later stages. This also confirms the need to pay attention to the preliminary impact when encountering similar emergencies and to intervene in public opinion as soon as possible.

4 Conclusion

In this paper, a convolutional neural network was used to construct a textual topic classification model for social network opinion users, and based on the kNN algorithm, the sentiment scores of different opinion algorithms were compared to analyze opinion changes. The sentiment of public opinion changes on social networks during the New Crown epidemic fluctuates greatly, and the sentiment polarity tends to be positive for 65 days, while the sentiment polarity tends to be negative for 15 days, i.e., the score is less than 0.5. Public opinion presents mostly negative sentiment polarity in the early stages but tends to be positive in the middle and late stages, so timely and rapid intervention in social network public opinion should be achieved. Accordingly, this paper proposes the following paths for timely intervention in online public opinion:

1) Laws and regulations should be used to guarantee and supervise citizens’ freedom of expression in the dissemination of online public opinion to regulate the behavior of online public opinion promptly and prevent the emergence of situations that endanger social stability and the infringement of citizens’ legitimate rights and interests by malicious public opinion. A clear legal restraint mechanism for network public opinion and regulating virtual society management are necessary guarantees for a good environment for disseminating network public opinion.

2) Improving citizens' legal and moral awareness is key to constructing effective public opinion management networks. Citizens take compliance with the normative requirements of network opinion management as the norm for posting comments, cultivate the formation of network cultural literacy, improve resistance to vulgar information and enhance discernment of false information, create healthy and beneficial network opinion, and create a truly harmonious and orderly network environment.

Funding:

2021 Scientific Research Project of Hunan Provincial Department of Education: Research on News Release Optimization Strategy of College Network Public Opinion Management——Taking Hunan Colleges and Universities as an Example. Project number: 21C1364

References

[1] Cheng, Y. (2021). Crowd-sourcing information dissemination based on spatial behavior and social networks. Mobile Information Systems.

[2] Shi, J., Hu, P., Lai, K. K., & Chen, G. (2018). Determinants of users' information dissemination behavior on social networking sites. Internet Research, 28(2), 393-418.

[3] Mcgregor, S. C. (2019). Social media as public opinion: how journalists use social media to represent public opinion. Journalism, 20(8), 1070-1086.

[4] Kim, D., Chung, C. J., & Eom, K. (2022). Measuring online public opinion for decision making: application of deep learning on political context. Sustainability, 14.

[5] Manan, K. A. (2020). Online news and public opinion: how malaysians respond to news on a state by-election. Human Communication Research.

[6] Garritzmann, J. L., Neimanns, E., & Busemeyer, M. R. (2021). Public opinion towards welfare state reform: the role of political trust and government satisfaction. European Journal of Political Research.

[7] He, W., Fang, Y., Malekian, R., & Li, Z. (2019). Time series analysis of online public opinions in colleges and universities and its sustainability. Sustainability, 11(13), 3546.

[8] Xiwei, W., Dan, Z., Mengqing, Y., Lian, D., Meng Meng, X., & Qiuyan, G. (2017). Public opinion dissemination on mobile internet- a case of ebola. Interlending & document supply, 45(2), 87-100.

[9] Zhu, H., Mei, Y., Wei, J., & Shen, C. (2022). Prediction of online topics' popularity patterns: Journal of Information Science, 48(2), 141-151.

[10] Tian, Y. W. L. (2018). Opinion dynamics in social networks with stubborn agents: an issue-based perspective. Automatica, 96.

[11] Chen, Y., Li, Y., Wang, Z., Quintero, A. J., Yang, C., & Ji, W. (2022). Rapid perception of public opinion in emergency events through social media. Natural hazards review, (2), 23.

[12] Zhang, Y., Chen, Z., & Zou, L. (2021). Public opinion communication model under the control of official information. Complexity, 2021.

[13] Wang, G., Wang, Y., Liu, K., & Li, J. (2020). Multidimensional influencing factors of public opinion information dissemination in social media: evidence from weibo dataset. International Journal of Modern Physics B.

[14] Haihong, E., Yingxi, H., Haipeng, P., Wen, Z., Siqi, X., & Peiqing, N. (2019). Theme and sentiment analysis model of public opinion dissemination based on generative adversarial network. Chaos, Solitons & Fractals, 121, 160-167.

[15] A, L. Z., B, C. S., C, Y. J., D, M. G., & E, Z. W. (2018). Cross-network dissemination model of public opinion in coupled networks. Information Sciences, s 451–452, 240-252.

[16] Jia, F., & Chen, C. C. (2020). Emotional characteristics and time series analysis of internet public opinion participants based on emotional feature words: International Journal of Advanced Robotic Systems, 17(1), 105-108.

[17] Li, Q., Du, Y. J., Li, Z. Y., Hu, J. R., & Jia, P. (2021). Hk–seir model of public opinion evolution based on communication factors. Engineering Applications of Artificial Intelligence, 100(2), 104192.

[18] Wang, G., Chi, Y., Liu, Y., & Wang, Y. (2019). Studies on a multidimensional public opinion network model and its topic detection algorithm. Information Processing & Management, 56(3), 584-608.

[19] Wang, Y., Li, H., Zuo, J., & Wang, Z. (2019). Evolution of online public opinions on social impact induced by nimby facility. Environmental Impact Assessment Review, 79(SEP.), 106290.