International Journal of Cognitive Computing in Engineering 7 (2026) 1–11

Contents lists available at ScienceDirect

International Journal of Cognitive Computing in Engineering

journal homepage: www.keaipublishing.com/en/journals/international-journal-of-cognitive-computing-in-engineering/

Analysis of emotional tendencies and discourse patterns in VKontakte social comments based on Nvivo12 encoding☆

Jiaxing Han

Journalism and Communication College, Xinjiang University, Urumqi 830046, PR China

A R T I C L E I N F O

Keywords:
Nvivo12
Natural language processing
VKontakte
Discourse patterns
Emotional tendencies

A B S T R A C T

To study the emotional changes of the public during the COVID-19 epidemic, the experiment conducted an analysis of emotional tendencies and discourse patterns for comments on the VKontakte platform. The study first used Nvivo12 to classify comments on social platforms into topics, emotions, and relationship nodes. Then, a bidirectional long short-term memory network was introduced to comprehensively understand the context and classify positive and negative emotions. In addition, natural language processing toolkits were used to analyze the discourse structure of social comments, and support vector machines were used to discriminate the emotional tendencies of comments. According to the experimental analysis, during the period of rapid incidence rate increased, 27.6 % of the public exhibited positive emotional tendencies, while <39.3 % exhibited negative emotional tendencies. In the following three stages, the proportion of negative emotions in the public was greater than that of positive emotions. In the fourth stage of the epidemic, comments mainly concerned the supply of medical drugs, masks, and the construction and opening of hospitals. The existing problems indicate that the epidemic has had a significant impact on public emotions, and effective measures need to be taken to alleviate the negative emotions of the public. The research results are helpful in revealing the dynamic changes of public emotions and their discourse patterns during the COVID-19 epidemic, and provide a new perspective for understanding public emotions.

1. Introduction

During the COVID-19 epidemic, VKontakte users expressed their views on the epidemic through comments and shared their life experiences and information (Mirzaei et al., 2021). These comments not only reflect the emotional tendencies of users, but also reflect the distribution of discourse power and the characteristics of information dissemination (Gheisari et al., 2023). However, traditional methods in comment analysis are often based on manual annotation or expert evaluation. These methods are highly subjective. Therefore, different people may interpret the same text differently, resulting in inconsistent analysis (Che & Kim, 2024). Nvivo12 encoding provides multiple encoding methods, including automatic, associative, and core encoding, and suitable encoding methods can be selected according to actual needs (Sremanakova et al., 2024; Arbieu et al., 2021). In comment analysis, Nvivo12 encoding provides a powerful search and query tool that can help users quickly locate and filter the content that needs to be encoded. The Natural Language Processing (NLP) toolkit provides sentiment analysis functionality that can automatically determine the emotional tone of text. This helps people quickly understand the tone of a large number of comments. Meanwhile, the NLP toolkit can extract keywords and topics from comments, which helps to understand the core content and topic distribution of comments. Therefore, this study uses Nvivo12 coding and NLP toolkit to analyze the emotional tendencies and discourse patterns of VKontakte social comments. To improve the accuracy of emotion classification and topic classification, the Bidirectional Long Short-Term Memory (BiLSTM) and Support Vector Machine (SVM) algorithms are introduced to improve the original methods. First, relevant social comment data is captured from social networking sites and preprocessed. Using NVivo's sentiment analysis function, positive and negative emotions in text are automatically identified and encoded. Then, the classified samples are input to train the BiLSTM model, which can then be used for sentiment classification. Finally, an NLP toolkit is used to analyze the discourse structure of social comments. SVM is then utilized to discriminate the emotional tendencies of comments. Combining qualitative and quantitative analysis is a methodological approach that can provide researchers with a more comprehensive perspective. Therefore, it can facilitates a deeper understanding of the

☆ Peer review under the responsibility of KeAi Communications Co., Ltd
E-mail address: jiaxing_han@xju.edu.cn.

https://doi.org/10.1016/j.ijcce.2025.07.001
Received 28 April 2024; Received in revised form 10 July 2025; Accepted 12 July 2025
Available online 13 July 2025

emotional dynamics and discourse characteristics of social comments. When applied to the analysis of user data on social platforms, it can promote the visualization and interpretation of social comment data. It can also support decision-making processes, improve user experience and market strategies, and provide important support and insight for the development of online platforms and social research. The innovation of this study lies in the comprehensive application of various methods and technologies such as Nvivo12 coding, NLP toolkit, BiLSTM, and SVM to achieve a comprehensive and in-depth analysis of VKontakte social comments. The contributions of this study are as follows: (1) The study can promote academic research in the field of emotion analysis in public health emergencies and enrich the theory and methodology of emotion analysis. (2) The study can assist government departments and related organizations to better understand the public's emotional state and needs during an epidemic. Therefore, it can improve social governance capabilities and public health emergency management levels.

Firstly, a review of relevant research is conducted, introducing the current application status and existing problems of Nvivo12 encoding and NLP toolkit. Then, the research methods and implementation steps of Nvivo12 encoding and NLP toolkit are introduced in detail. Next, a case study is conducted to verify the effectiveness and feasibility of the research method through experiments. Finally, the research results are summarized and future prospects are proposed.

2. Literature review

The sentiment analysis of social comments aims to determine their emotional tendencies by analyzing the content of comments on social media. Kuznetsova used tools that employed sentiment predicate analysis to automatically extract emotional situations from text for online sentiment analysis. The results indicated that the emotional attitudes in the online environment were generally balanced, with no significant negative or positive emotions (Kuznetsova, 2022). K ü chler et al. developed a supervised machine learning model that used pre-trained word embeddings and word frequency features to analyze emotional tendencies in the comment sections of news articles. They used 303342 user comments as samples to provide insights into the dynamics of online discussions (Küchler et al., 2023). Zhang et al. proposed a new method that uses an efficient learning encoder to accurately classify token replacements and a hybrid neural network to more accurately capture emotional features in text. The results indicated that the proposed model had effectively improved accuracy (Zhang et al., 2022). Using tools such as computer vision, NLP, and sentiment analysis, Cheng et al. examined 101,292 TikTok news videos and found that those with higher audience engagement often conveyed more negative emotions (Cheng & Li, 2024).

The COVID-19 has a profound impact on the global economy and society and on the discourse patterns and emotional tendencies on social media platforms. Prasad et al. proposed a COVID Vision system based on convolutional neural network for face recognition to ensure that people maintained social distance and reduced the spread of COVID-19 (Prasad et al., 2022). To measure the impact of COVID-19 on passenger travel emotions, C.-H et al. collected and examined flight reviews on TripAdvisor from January 2016 to August 2020. Data driven and visual analysis were conducted. The proposed method was superior to baseline comparison and helpful for theoretical and management literature (C.-H et al., 2022). To explore the role of social media in the analysis of tourists' emotions, Floresruiz used social media data to analyze tourists' emotions and behaviors during the COVID-19 pandemic. Since the outbreak of the epidemic, tourists placed greater emphasis on safety and preferred to travel alone to nearby less crowded destinations (Floresruiz & Elizondosalto, 2021). Luo et al. extracted topic, emotion, machine learning and other means. Further analysis was conducted on online public opinion's properties during crisis events. Information's richness et al. might affect emotions and dissemination (Luo et al., 2023). Chaney et al. conducted 53 interviews on people's attitudes towards COVID-19 vaccination plans using a consumer perspective and qualitative methods. These results confirmed that hesitation in COVID-19 vaccine administration was mainly due to two factors: stability factors and background factors (Chaney & Lee, 2021). Since the COVID-19 pandemic, anti-Asian sentiment in the United States rose significantly. Willnat et al. conducted an online survey of 913 white Americans in 2021. They studied exposure to news related to the pandemic and analyzed possible connections between stigmatization of Asians and a sense of entitlement towards Asian immigrants (Willnat et al., 2023).

Emotion analysis is a branch of NLP that refers to the process of automatically interpreting and classifying emotions based on textual data. Shaik A et al. proposed an emotion analysis model based on Recurrent Neural networks and Salp Swarm (RNN-SS) algorithm for user sentiment analysis. The results indicated that the proposed model could effectively analyze users' emoticons on social media platforms and improve the accuracy of sentiment analysis (Shaik et al., 2024). Damayanti et al. proposed an SVM algorithm with Word2Vec feature extraction to analyze emotional tendencies reflected in online conversations, which helped to quickly and comprehensively understand public opinions and preferences. The results showed that the proposed algorithm was superior to traditional SVM methods, with an accuracy of 88.94 % and a recall rate of 93.08 % (Damayanti & Lhaksmana, 2024). Ramasamy M et al. proposed an sentiment analysis model that combined an improved LSTM and a Hybrid Chameleon Rat Swarm Optimization (HCRSO) algorithm for sentiment analysis of social media content. The results showed that the accuracy and specificity of the proposed model were 94 % and 93 %, respectively, demonstrating good performance in sentiment analysis (Ramasamy & Elangovan, 2024).

In summary, NVivo12 software and NLP can be effectively applied to text and comment analysis. However, at present, there are few technologies that use NVivo12 software and NLP in the comment analysis of social software in the context of COVID-19. Therefore, this study uses these two technologies to analyze changes in public sentiment during the pandemic, providing a useful reference for policy formulation, public opinion guidance, and social development.

3. Social comments combining Nvivo12 encoding with natural language processing toolkits

The COVID-19 epidemic is a global challenge that requires international cooperation and exchanges. Analyzing social comments can help people better understand the needs and concerns of the country's and region's residents. It can also promote cooperation and understanding between the people and the government. Therefore, this study is based on the Russian VKontakte social platform to investigate emotional tendencies and discourse patterns in comments.

3.1. Statistical analysis of emotional tendencies and discourse patterns based on Nvivo

Since the beginning of 2020, a serious COVID-19 epidemic breaks out worldwide. In Russia, the spread of the COVID-19 epidemic brings enormous challenges to the social economy, public health and people's lives (Syropoulos et al., 2024; Xie et al., 2022). In this context, social media becomes an important platform for people to obtain information, exchange opinions, and express emotions (Yalva & Gaynor, 2021). As one of the largest social media platforms in Russia, VKontakte's user generated content is very important for understanding the emotions and discourse patterns of Russian people in the COVID-19 epidemic (Han et al., 2021). Nvivo12 provides a complete set of tools to not only quickly identify and tag emotional expressions in text using automatic coding wizards, but also to support multiple coding methods. Using Nvivo12 to classify social media comments into themes, emotions, and relationships can quickly identify key themes and emotional trends in comments, providing a foundation for further in-depth analysis. However, for emotional trend analysis during COVID-19, Nvivo12 may have

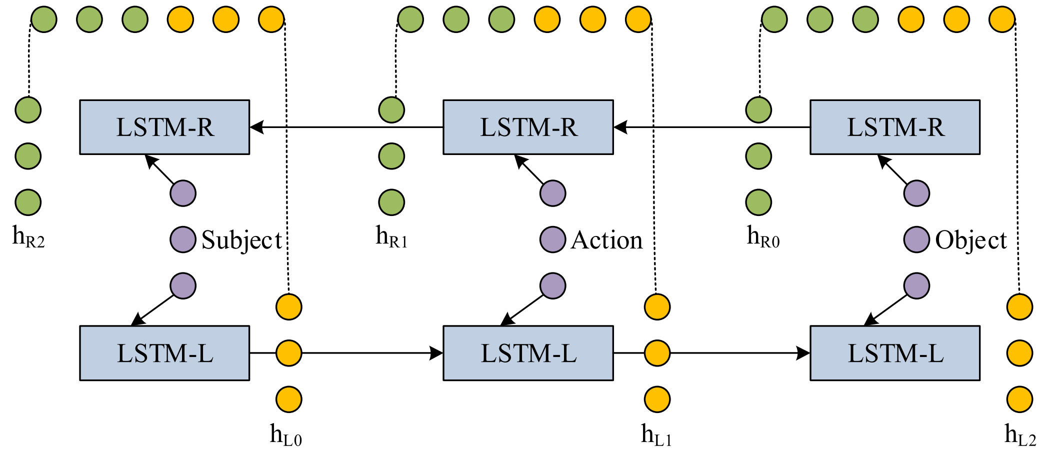

Fig. 1. Structure diagram of BiLSTM.

some format incompatibilities and format conversion is required.

Therefore, the study aims to use Nvivo12 encoding to classify comments on VKontakte social platform into topic, emotion, and relationship nodes. The analysis steps are as follows. First, relevant social comment data are captured from VKontakte. Then, the data are cleaned up and irrelevant information, duplicate content, and obvious errors or outliers are removed. Text preprocessing is necessary for subsequent sentiment and discourse pattern analysis. This includes text cleanup, segmentation, deletion of deactivated words, stemming, and spelling correction. Specifically, text cleaning aims to remove irrelevant characters, special symbols, HTML tags, etc. to ensure text quality. Segmentation refers to the process of breaking down text into words or subunits. Deletion of deactivated words refers to the removal of commonly used words that are not meaningful for sentiment analysis. Stemming is rule-based and may result in truncation of certain words. Spelling correction is the process of correcting spelling errors in text to improve its accuracy. Next, a coding scheme is established and the use of automatic emotion coding is explored, using NVivo's emotion analysis feature to automatically identify and encode positive and negative emotions in text. The emotion score is assigned between negative and positive, and the score is adjusted based on contextual information. The framework and process of automatic emotion encoding mainly includes file selection, parameter setting, encoding initiation, and result display. Specifically, the first step is to select the social comment text to be automatically encoded in NVivo. Then, the sentiment vocabulary is set, such as which words or phrases express positive emotions and which express negative emotions. After clicking the "start coding" button, NVivo automatically analyzes the text data and categorizes the relevant text segments into the appropriate sentiment categories. Finally, the coding results and related statistical information are reviewed.

Nvivo12 has performance and storage issues with very large social media datasets. BiLSTM simultaneously captures information in both directions, before and after a sequence, to capture long-term dependencies. Therefore, it has significant advantages in emotion classification tasks that require rich contextual information for decision-making. However, traditional machine learning models, such as logistic regression, K-nearest neighbors, and convolutional neural networks, have difficulty capturing the complex contextual dependencies found in text data. This limits their effectiveness in emotion recognition tasks (Sasangohar et al., 2020; Bahr, 2021). Therefore, the study introduces BiLSTM on the basis of Nvivo12. BiLSTM combines forward and backward LSTM. Compared to traditional LSTM, BiLSTM can process data from both forward and backward time series perspectives, allowing for a more complete understanding of context. This enables BiLSTM to perform better in tasks such as text classification, sentiment analysis, machine translation, and time series prediction in NLP. The structure of BiLSTM is shown in Fig. 1.

In Fig. 1, there are three gate structures in BiLSTM. The expression for the forget gate is represented by Eq. (1).

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

In Eq. (1), σ is an activation function. h_{t-1} represents the output information of the previous cell. x_t is the input information for the current cell. b_f is a bias term. W_f is the forget gate's weight matrix. The input gate of BiLSTM updates the information using Eq. (2).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

In Eq. (2), W_i is a matrix weight of the input gate. b_i is a bias term of the input gate. The candidate values for storage units in the input gate

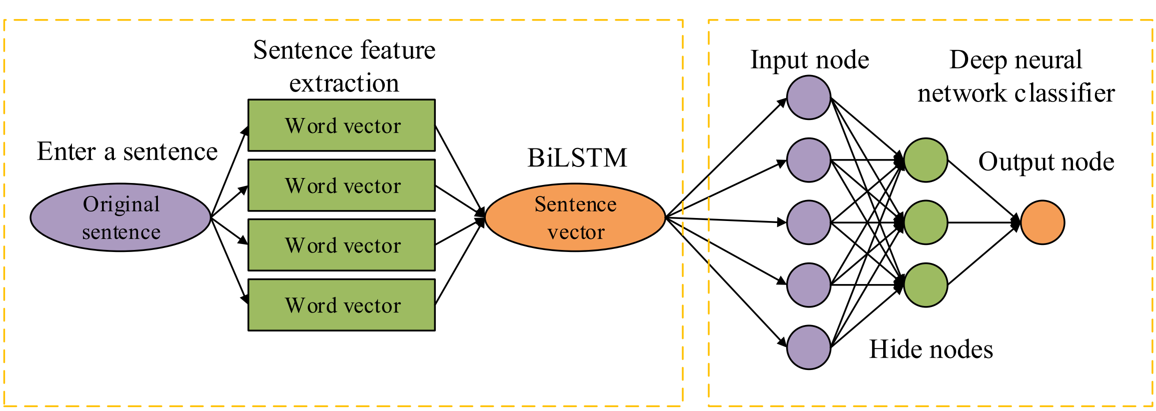

Fig. 2. Positive and negative emotion classification model based on BiLSTM.

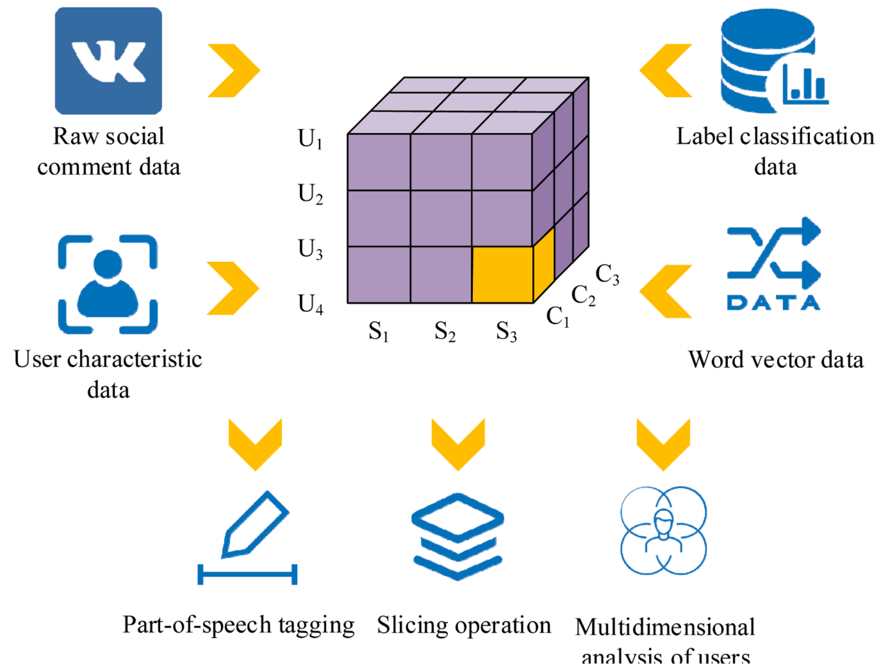

Fig. 3. Schematic diagram of discourse pattern analysis based on Nvivo12 and BiLSTM.

are represented by Eq. (3).

$$\overline{C_t} = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

In Eq. (3), W_c is the candidate content's matrix weight. b_c is the alternative content's bias term. The update of the storage unit in the input gate is represented by Eq. (4).

$$C_t = it * \overline{C_t} + f_t * C_{t-1} \tag{4}$$

Eq. (5) is a bias term for the corresponding output gate.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

In Eq. (5), W_o is the output gate's matrix weight. b_i is the output gate's bias term. The input information of the cell can be obtained from the update equation of the storage unit, represented by Eq. (6).

$$x_t = \sigma_t * \tanh(C_t) \tag{6}$$

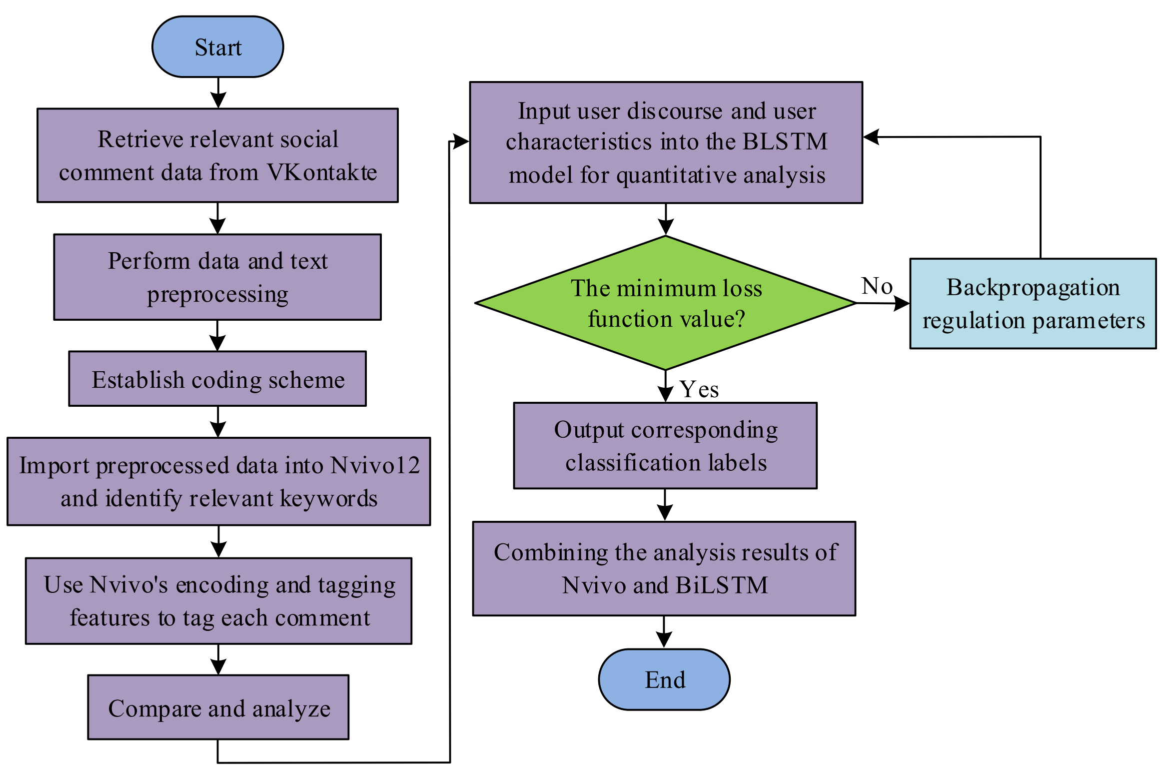

Fig. 4. The flowchart of the emotion classification model based on Nvivo and BiLSTM.

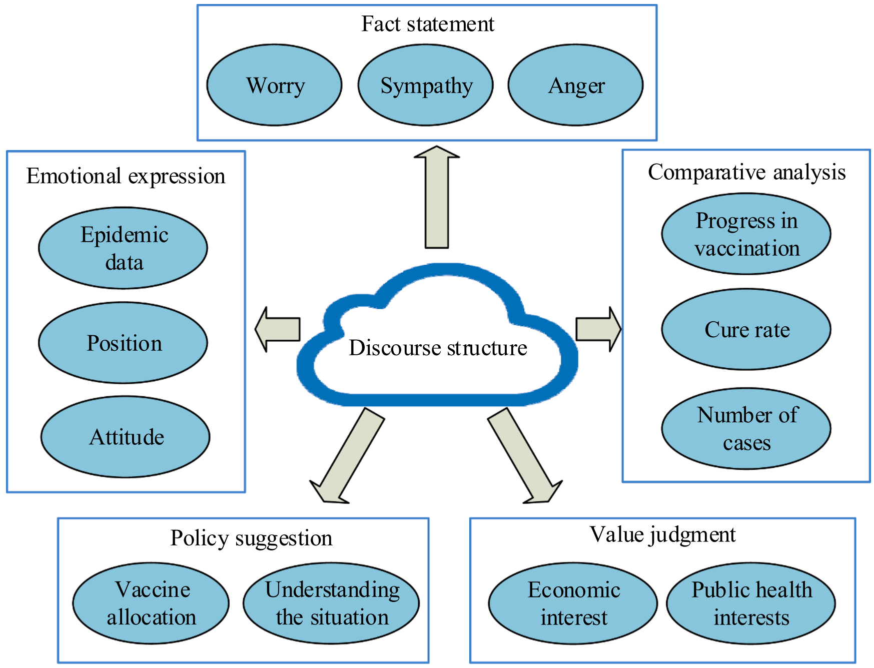

Fig. 5. Comment structure of netizens under daily epidemic data.

Fig. 2 is a positive and negative sentiment classification model based on BiLSTM.

In the positive and negative sentiment classification model, the pre-trained word vectors are first used to transform the text into fixed-dimensional feature vectors. Then, an emotion classification model is constructed using BiLSTM. This model typically includes one or more BiLSTM layers to capture contextual and semantic information in the text. Next, this model is trained using a labeled sentiment dataset. During training, this model's parameters are optimized using back-propagation algorithm to minimize classification error. The cross-entropy loss function is used in sentiment classification, represented by Eq. (7).

$$loss = -\sum_{i=1}^{n} y_i \log y_i + (1 + \bar{y}_i)\log(1 - \bar{y}_i) \tag{7}$$

In Eq. (7), y_i is a value of the true label. $\bar{y}_i$ is a value of the predicted label. The corresponding variation of the cross-entropy loss function can be obtained, represented by Eq. (8).

$$\frac{\partial loss}{\partial y} = -\sum_{i=1}^{n} \frac{\bar{y}_i}{y_i} - \frac{1 - \bar{y}_i}{1 - y_i} \tag{8}$$

After the BiLSTM layer, an Attention layer is usually added. The outputs of BiLSTM and Attention layers are converted into the final sentiment classification result. By using the Softmax function, the output can be converted into a probability distribution for classification purposes. α is the output result of the Match module, represented by Eq. (9).

$$\alpha = h^T W z \tag{9}$$

In Eq. (9), h refers to the hidden state of the previous time step. W is the weight matrix. α is normalized in Softmax, as shown in Eq. (10).

$$c^0 = \sum_{i=1}^{n} \alpha_0^i h^i \tag{10}$$

In Eq. (10), α_0^i is a normalized weight. The schematic diagram of discourse pattern analysis based on Nvivo12 and BiLSTM is shown in Fig. 3.

Fig. 3 illustrates the discourse pattern research process. First, the original social comment data from VKontakte is collected. Then, the preprocessed data is imported into Nvivo12 to identify related subject words, such as Covid-19. Then, Nvivo's encoding and labeling functions are used to label each comment to identify different discourse patterns or trends. Subsequently, a comparison and comparative analysis are conducted to gain a deeper understanding of the discourse patterns and characteristics of different user groups. Next, quantitative analysis is conducted using BiLSTM, with user discourse and user features as inputs and corresponding classification labels as outputs. Finally, the analysis results of Nvivo and BiLSTM are combined to comprehensively understand the discourse patterns of different feature user groups. Nvivo provides in-depth understanding of themes and trends, while BiLSTM provides quantitative, data-driven insights. In summary, the flowchart of the emotion classification model based on Nvivo and BiLSTM is shown in Fig. 4.

In Fig. 4, the sentiment classification model based on Nvivo and BLSTM uses Nvivo's coding and labeling capabilities to label each comment to identify different discourse patterns or trends. Input samples containing positive and negative emotions are used to train BiLSTM, and the trained BiLSTM can be used to predict emotional tendencies in social comments.

3.2. Social comment analysis based on natural language processing toolkits

In this study, the aforementioned emotion classification model based on Nvivo and BLSTM preliminarily identifies key emotional nodes and discourse structures in the comments. However, due to the vast amount of textual data on VKontakte, it is difficult to capture deep semantic associations and dynamic changes using only qualitative analysis. To address this issue, the study introduces NLP toolkits and SVM algorithms to complement Nvivo's limitations through quantitative modeling. Therefore, it can achieve automation of emotion classification and topic recognition. Fig. 5 shows the comment structure of netizens under daily epidemic data.

Under daily epidemic data, the discourse structure of comments by netizens may vary due to factors such as personal views, geography, and group. By using the NLP toolkit, the following common discourse structures can be derived. First, there is emotional expression. Many comments directly express the emotions of netizens, such as concerns

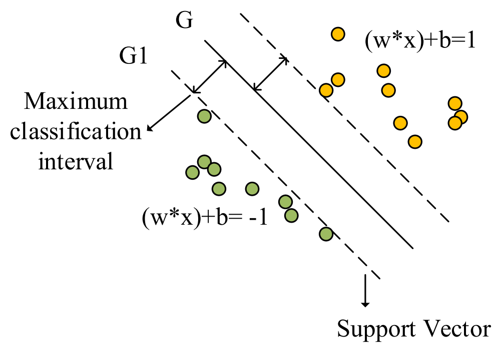

Fig. 6. Optimal classification hyperplane for SVM.

about the epidemic, sympathy for victims, anger towards the government or relevant institutions, etc. Then comes the factual statement. Some comments can provide specific facts or data related to the epidemic, such as the number of cases, cure rates, progress of vaccination, etc. These factual statements are often used to support the commentator's viewpoint or refute the views of others. Next is comparative analysis. Some comments may express concerns or appreciation for other regions by comparing epidemic data between different regions and countries. This comparative analysis can highlight the commentator's position and attitude. Then there are policy recommendations. With the development of the epidemic, many comments will propose specific policy suggestions, such as strengthening prevention and control measures, optimizing vaccine distribution plans, etc. These suggestions are usually based on the commentator's understanding and expectations of the current situation. Finally, there is value judgment. Discussions on morality, ethics, and values often appear in comments. These value judgments are typically associated with decisions and behaviors related to the epidemic. Examples include whether vaccination should be mandatory and how to balance economic and public health interests.

The NLP toolkit can extract keywords and themes from comments. Moreover, its sentiment analysis function automatically determines the emotional tone of text. This is useful when processing large amounts of text, particularly when analyzing comments on VKontakte (Shafi et al., 2023; Liu et al., 2023). However, applying NLP to VKontakte comments may also face some challenges, especially considering that these comments may differ in language, informal expression, or grammar. For example, users may frequently make grammatical and spelling mistakes in comments, which can affect the performance of NLP toolkits. Therefore, this study uses the Valence Aware Dictionary and Emotion Reasoner (VADER) sentiment analyzer to calculate the sentiment score for each comment. VADER is a powerful NLP tool based on dictionaries and rules. It can evaluate the emotional orientation of text and return scores for positive, negative, and neutral emotions. It also provides an overall score representing the text's emotional orientation. The range of positive, neutral, and negative scores is 0 to 1, representing the proportion of positive, neutral, and negative text. The overall score ranges from −1 (extremely negative) to 1 (extremely positive), indicating the overall emotional valence of the text. In this way, researchers can quantify the emotional tendencies of comments and then analyze the emotional changes of the public during the COVID-19 epidemic. However, the sentiment analysis methods of NLP toolkits may be limited by vocabulary, context, and culture. Sometimes, the meaning of emotional words can differ in different contexts, and emotional expressions can vary across different cultures. SVM is a class of generalized linear classifiers that perform binary classification on data using supervised learning methods (Saulsberry et al., 2023; Luo et al., 2024). In SVM, input samples are mapped to a high-dimensional space, and different sample categories are segmented by finding an optimal hyperplane. SVM only cares about the support vectors, not the entire data set. Therefore, it has good robustness to noise and outliers in the data. In addition, since the optimization problem of SVM is a convex optimization problem, the solution is the globally optimal solution and will not fall into local minima. SVM can also perform well in small sample situations, making it suitable for problems with small sample sizes. Moreover, the model has strong interpretability. Therefore, SVM is used for social comment analysis in the study. Fig. 6 shows the optimal classification hyperplane for SVM.

Fig. 6 shows the optimal classification hyperplane of SVM, which is determined by maximizing the interval and aims to reduce classification errors and improve generalization ability. Therefore, the study introduces SVM on the basis of NLP tools to improve the classification performance and generalization ability in comments. The calculation of the optimal classification plane and lines G, G1, and G2 in SVM is represented by Eq. (11).

$$f(x) = w^T x + b \tag{11}$$

In Eq. (11), w is the normal vector. b is the bias amount. (x_1,y_1), (x_2, y_2), … (x_i,y_i) in all classification samples need to meet the equation conditions and be represented by Eq. (12).

$$y_i(wx_i + b) \geq 1 \tag{12}$$

Different w and b can determine different position planes. The optimal classification plane is calculated using Eq. (13).

$$\begin{cases} \max \dfrac{1}{\| w \|} = \min \dfrac{1}{2} \| w\|^2 \\ s.t, y_i(wx_i + b) \geq 1, i = 1, 2, 3 \ldots n \end{cases} \tag{13}$$

In Eq. (13), $\min \frac{1}{2}\| w \|^2$ refers to the minimum confidence range. Eq. (11) can be converted to Eq. (14).

$$\begin{cases} \min \dfrac{1}{2} \| w\|^2 \\ s.t, y_i - wx_i - b \leq \varepsilon \\ wx_i - y_i + b \leq \varepsilon, i = 1, 2, 3 \ldots, n \end{cases} \tag{14}$$

In Eq. (14), $\min \frac{1}{2}\| w \|^2$ can be regarded as a prerequisite for the optimization problem. When the current conditions cannot be met, it is necessary to introduce relaxation variables ξ_i and ξ_i^* to relax the range of prerequisite conditions. At this point, the objective of the function is the minimum confidence range in Eq. (15).

$$\begin{cases} \min \dfrac{1}{2} \| w \| + C \sum_{i=1}^{N} (\xi_i + \xi_i{}^*) \\ s.t, y_i - wx_i - b \leq \varepsilon + \xi_i \\ wx_i - y_i + b \leq \varepsilon + \xi_i{}^*, i = 1, 2, 3 \ldots n \\ \xi_i \geq 0 \\ \xi_i{}^* \geq 0 \end{cases} \tag{15}$$

In Eq. (15), C is a penalty factor that refers to the degree of punishment for data beyond the range of ε. In summary, the social comment analysis method based on NLP toolkit first uses the NLP toolkit to analyze the discourse structure of VKontakte social comments. Then it uses SVM multi class machine learning algorithm to distinguish the emotional tendency of comments. The study aims to analyze the emotional tendencies and discourse patterns of VKontakte social comments. This can help us better understand people's emotional reactions and speech characteristics during the pandemic, providing valuable information for the government, medical institutions, and other relevant organizations.

4. Analysis of emotional tendencies and discourse patterns based on Nvivo and natural language processing

Based on the above hybrid method framework, this study conducted empirical analysis on the comment data of the VKontakte platform from

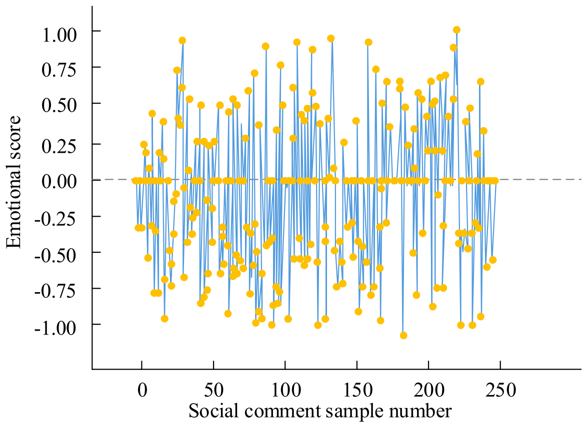

Fig. 7. Emotional propensity score in the rapid rise period of incidence rate.

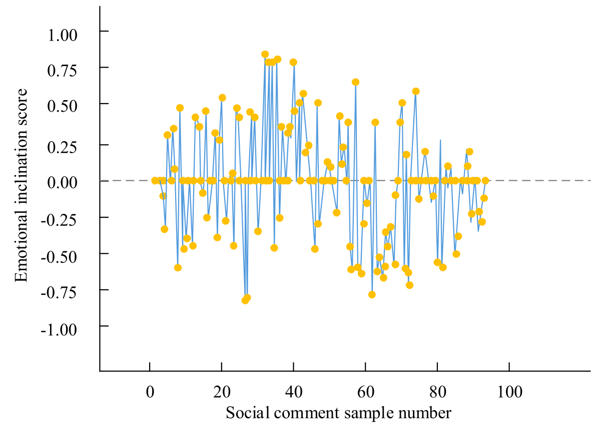

Fig. 8. Emotional propensity score in the rapid decline period of incidence rate.

March 2020 to March 2023. In the following section, the research results were presented and their social significance was explored from two dimensions: the temporal changes of emotional tendencies and the structural characteristics of discourse patterns.

4.1. Analysis of emotional tendencies in VKontakte social comments

VKontakte users were mainly concentrated in Russian-speaking countries, which limited its user base geographically. However, VKontakte had a huge user base and high user activity, which to some extent could reflect the emotions and public opinion trends of the Russian domestic public. Therefore, although VKontakte had certain limitations in terms of user groups and regions, it could form a more comprehensive social media ecosystem together with other social media platforms, providing richer data sources for studying public emotions during the epidemic. To ensure the robustness and representativeness of discovering public sentiment trends, the study collected a large amount of relevant social commentary data from VKontakte to ensure the data's richness and diversity. Moreover, the collected data was cleaned to remove irrelevant information, duplicate content, and obvious errors or outliers. The chi-square test was used to cross-analyze VKontakte's social commentary on different topics during different epidemic periods. It also analyzed whether there were significant differences in the public's attention to different topics and emotional tendencies in different periods by calculating the adjusted residuals and other indicators. The study optimized parameters through grid search and cross validation, setting dropout to 0.5, batch size to 16, epoch to 5, and iteration count to 200. The emotional tendency analysis results of the method designed for the research in different stages of the epidemic were first divided into four development stages based on the changes in the epidemic situation in the study. They included the rapid rise period of the incidence rate (March to April 2020), the rapid decline period of the incidence rate (May to July 2020), the rapid decline and fluctuation period of the epidemic (August to December 2020), and the continuous decline period of the epidemic (January 2021 to March 2023). The emotional propensity score in the rapid rise period of incidence rate is shown in Fig. 7.

In Fig. 7, during the period of rapid increase in the incidence rate, 27.6 % of the Russian population had positive emotional tendencies, <39.3 % of the negative emotional tendencies. The analysis revealed that, during the rapid formation of the first wave of the pandemic, public debate focused on protecting the population from the epidemic's impact and the government's anti-epidemic policies. The relatively high prevalence of negative emotions was mainly because people did not believe that the increase and decrease of incidence rate could be effectively monitored. Moreover, the COVID-19 epidemic posed a serious threat to human life safety, leading to widespread fear and concern. During the peak period of the epidemic, this emotion might become more severe, leading to a higher number of people with a tendency towards negative

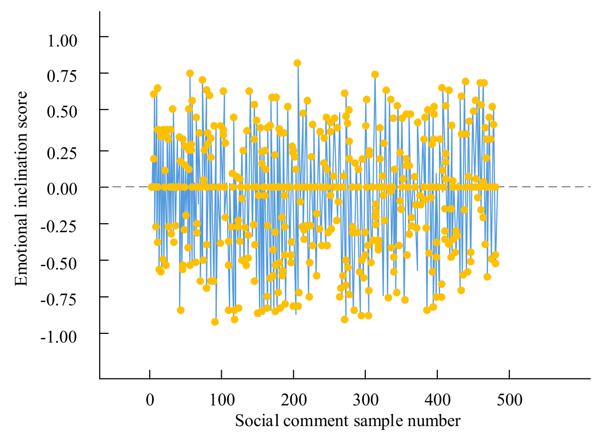

Fig. 9. Emotional propensity score during the rapid decline and fluctuation period of the epidemic.

emotions. Then the emotional propensity score in the rapid decline period of incidence rate is analyzed in Fig. 8.

Fig. 8 shows the analysis results of emotional tendencies of VKontakte social comments during the period of rapid decline of COVID-19 incidence rate (May to July 2020). During the period of rapid decline, the proportion of negative emotions was 33.4 %, which was 11.3 % higher than the proportion of positive emotions. Compared with the rapid period of COVID-19 incidence rate, the negative emotions in the comments decreased by 5.9 %. The main reasons were as follows: as the control of the pandemic gradually improved, the number of new cases per day decreased and the mortality rate declined. People gradually returned to normal production and life. This reduced people's panic and concern about the epidemic, and the proportion of negative emotions also decreased. Subsequently, the emotional propensity score during the rapid decline and fluctuation period of the epidemic (August to December 2020) is analyzed in Fig. 9.

In Fig. 9, the proportion of negative emotions in the comment section accounted for 39.6 %, which increased by 6.2 % compared to the period of rapid decline in the epidemic. During this period, the proportion of positive emotions reached 21.3 %, while negative emotions still accounted for a higher proportion in the comment group. The reason for this was that during this period, the prevention and control policies related to the inconvenience in people's lives, such as the lack of protective equipment, social isolation, and other issues. These led to a decline in the attitude of the Russian people towards the government and society. Although the epidemic was partially under control, strict social isolation measures were implemented to prevent a resurgence. These measures significantly restricted people's lifestyles and work

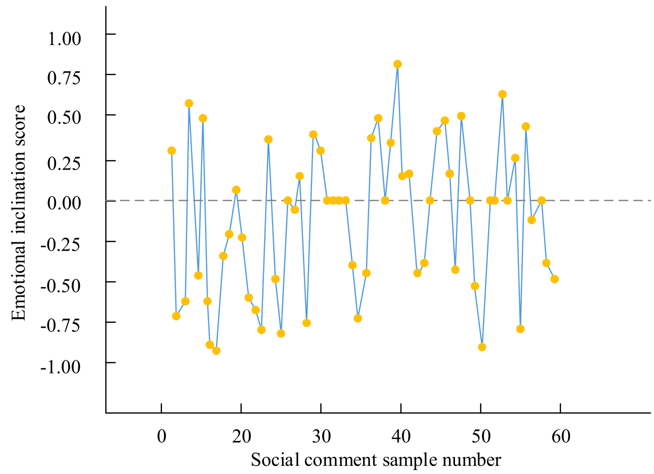

Fig. 10. Emotional inclination score during the continuous decline period of the epidemic.

styles, increased economic pressure, and caused negative emotions. Finally, an analysis was conducted on the changes in social comment sentiment tendencies of VKontakte during the continuous decline period of the epidemic (January 2021 to March 2023) in Fig. 10.

In Fig. 10, during the continuous decline of the epidemic, negative emotions in VKontakte's social comments reached 45.3 %. Compared to the first three periods, this stage had the highest ratio of negative emotions. The proportion of positive emotions during this period was 21.3 %. The trend of negative emotions in the comments was opposite to the continuous decline in the number of cases during the epidemic, showing an increasing trend. The main reason for this was that, although the daily increase in cases was in a steady decline according to the statistical data, the impact of the prevention and pandemic control policies on the population still existed. Moreover, the negative emotions of the population had not weakened. The above research analyzed the emotional tendencies of VKontakte's social comments during different epidemic periods. Therefore, emotional changes of the public could be timely understood, policy formulation and adjustment could be optimized, and social unity and cooperation can be promoted. In the three subsequent stages of the epidemic, negative emotions outweighed positive ones. This could be due to the government's prevention and control measures, which restricted people's lifestyles and work styles for a long time. It caused great inconvenience to the public, and increased economic pressure, thereby triggering negative emotions. In addition, the media's extensive coverage and dissemination of negative information about the epidemic could also increase the public's perception of risk, triggering panic and fear. To further analyze these changes, the study could use NLP technology and sentiment analysis tools to analyze a large amount of textual data to better understand the emotional state of the public at different stages.

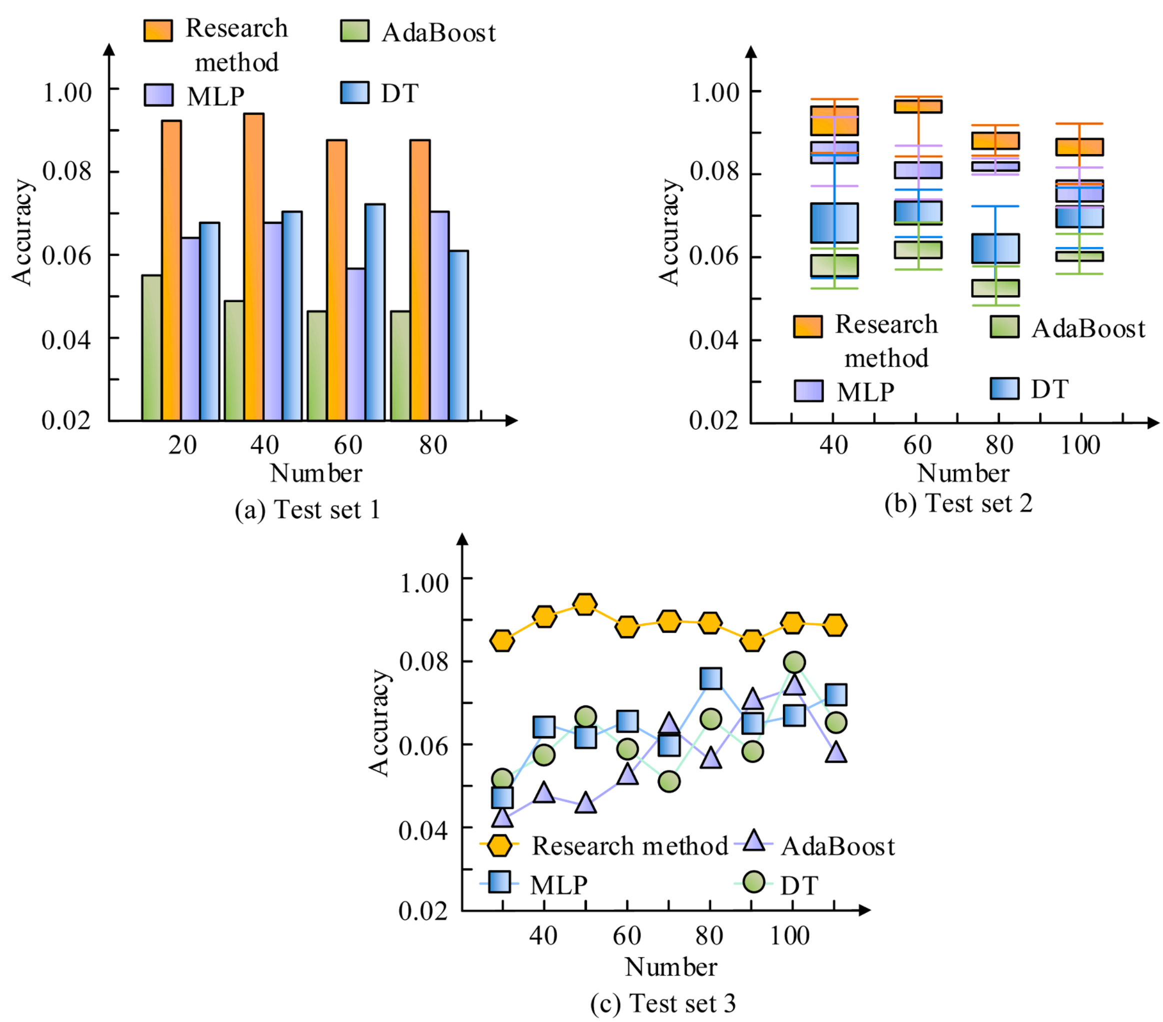

Fig. 11. Comparison of test set accuracy for different methods.

Table 1
Statistical results of average accuracy.

Method	Test set		
	Test set 1	Test set 2	Test set 3
DT	67.9 %	68.2 %	69.2 %
MLP	63.8 %	80.1 %	69.6 %
AdaBoost	53.8 %	54.4 %	55.3 %
Research method	89.2 %$^{\#*@}$	92.3 %$^{\#*@}$	92.4 %$^{\#*@}$

Note: $^{\#}$ represents a significant difference compared to DT, $p < 0.05$. * represents a significant difference compared to MLP, $p < 0.05$. $^{@}$ represents a significant difference compared to AdaBoost, $p < 0.05$.

4.2. Analysis of discourse patterns in VKontakte social comments

Next, an analysis was conducted on the discourse patterns of VKontakte social comments during the epidemic period. First, a comparison of the accuracy of multiple model tests was conducted. The study set the threshold to 0.5, Dropout to 0.5, bath3 to 16, epoch to 5, and iteration count to 200. Due to the fact that SVM used in this research method was a multi-class machine learning algorithm, this study compared three multi-class machine learning algorithms with their own characteristics and advantages in classification tasks: Decision Tree Classifier (DT), Multilayer Perceptron Network (MLP), and Adaptive Boosting (AdaBoost), to provide different perspectives and performance benchmarks. DT was an easy-to-interpret, non-parametric model that could quickly process and classify data, making it suitable for handling complex textual data such as social media comments. MLP was a feed-forward artificial neural network consisting of multiple layers of nodes that could capture complex nonlinear relationships and perform well in NLP. The AdaBoost algorithm was an ensemble learning method that could automatically identify important features and was suitable for handling data such as social media comments that contain a lot of noise. The comparison of test set accuracy for different methods is shown in Fig. 11.

Fig. 11(a) to Fig. 11(c) show the accuracy comparison of the four methods in test sets 1, 2, and 3, respectively. The average accuracy of the research method in test set 1 was 89.2 %, which was 21.3 %, 25.4 %, and 35.4 % higher than the average accuracy of DT, MLP, and AdaBoost, respectively. In test sets 2 and 3, the accuracy of the research methods was also the highest compared to other methods, with 92.3 % and 92.4 %, respectively. The average accuracy statistics of the above four methods on three test sets are shown in Table 1.

According to Table 1, the average accuracy of the research method on the three test sets was significantly higher than that of DT, MLP and AdaBoost ($p < 0.05$). This indicated that the method proposed by the research had high accuracy in analyzing discourse patterns. Chi-square test was a statistical method used to test whether there was a significant relationship between two categorical variables. This study used a chi-squared test to analyze whether significant differences existed in public attention and emotional tendencies toward different issues during different time periods. The adjusted balance was the balance value obtained by adjusting the original data after performing a chi-square test. This value was typically used to measure the difference between observed and expected frequencies in order to evaluate the strength of associations between different categories. In sentiment analysis and discourse pattern analysis, this value helped to understand whether there were significant changes in public attention and emotional tendencies toward a particular topic over different time periods. In chi-square tests, the adjusted residual was an indicator used to measure the difference between the observed and expected frequencies. It standardized residual values, making them more intuitive representations of the difference between observed and expected frequencies. In discourse pattern analysis, the adjusted residual could help people understand whether the emotional tendencies of a particular topic changed significantly over different time periods. Then, this research method was used

Table 2
Cross-tabulation of the chi-square test.

Time		Thematic tree						
		Policy measures to prevent a pandemic	Reliability of epidemiological data	Protecting civilians from a pandemic	Pandemic trends	Medical prevention of the spread of the virus	Pandemic and people's livelihoods	Spiritual support and encouragement
2020/03–2020/04	Number of comments	67	33	69	36	7	5	15
	Adjusted balances	2.741	−2.629	9.182	0.16	−4.002	−4.178	−2.526
2020/05–2020/07	Number of comments	36	78	11	44	12	4	5
	Adjusted balances	−1.264	7.978	−3.109	3.422	−1.811	−3.702	−4.113
2020/08–2020/12	Number of comments	85	75	37	58	56	74	82
	Adjusted balances	−3.03	−3.165	−4.158	−2.342	2.225	7.305	6.396
2021/01–2023/03	Number of comments	24	6	1	6	18	2	2
	Adjusted balances	3.487	−1.99	−2.584	−1.109	5.519	−1.548	−1.924
(total) amount	Number of comments	212	192	118	144	93	85	104
	Percentage of time intervals	22,36 %	20,25 %	12,45 %	15,19 %	9,81 %	8,97 %	10,97 %

Note: χ^2=284.484, $p = 0.000$.

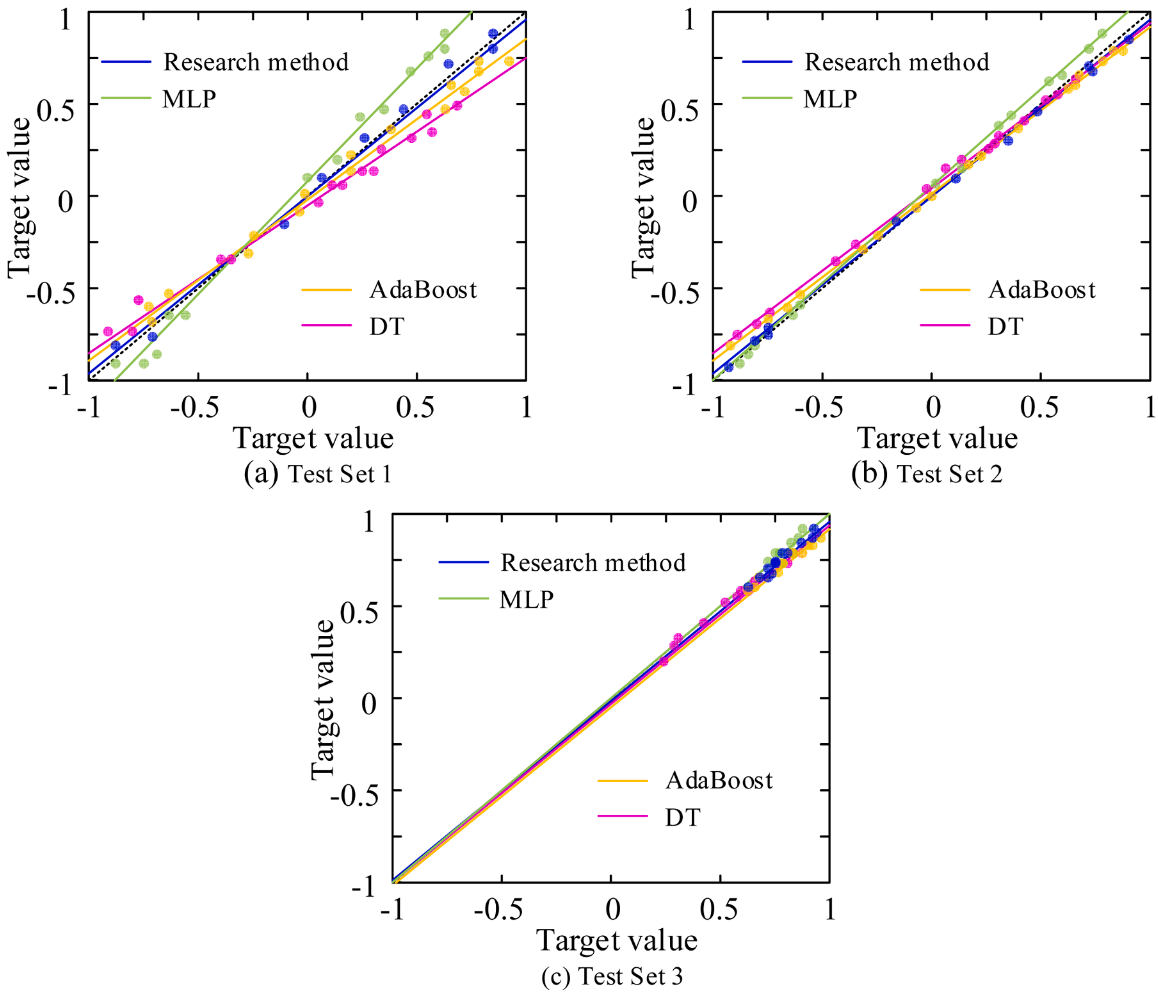

Fig. 12. Accuracy comparison of discourse classification among four methods.

to conduct cross analysis of VKontakte social comments on different topics during different epidemic periods. Table 2 shows the cross-tabulation of the chi-square test.

According to Table 2, the chi-square statistic of the chi-square test was 284.484, with a *p*-value of 0.00. This indicated that there were significant differences in public attention and emotional tendencies toward different topics at different times. In the first stage, the main topic in the comments was to prevent virus transmission, and the adjusted residual value was 2.741. It indicated that this topic had received significant attention during this time period. In the second stage, the review focused on the data related to the decline of incidence rate, and the corresponding adjusted residual value was 7.978. In the third stage, the comments mainly focused on discussing life support under epidemics, with a residual of 7.305. In the fourth stage, the comments mainly discussed the supply of medical drugs, masks, hospital construction, and opening up. In summary, the main commentator Hong focused on policy measures to prevent the spread of diseases, medical measures to prevent virus transmission, reliability of epidemiological data, and trends in pandemics. Finally, the accuracy comparison of discourse classification among four methods is analyzed in Fig. 12.

Fig. 12(a) to Fig. 12(c) show the comparison of topic classification accuracy for four methods in test sets 1, 2, and 3, respectively. In test set 1, the topic classification accuracy of the research method was 98.2 %, which was 12.4 %, 15.2 %, and 17.4 % higher than the topic classification accuracy of DT, MLP, and AdaBoost, respectively. In test set 2, the topic classification accuracy of the research method was 98.2 %, which was 10.3 %, 11.4 %, and 13.1 % higher than the topic classification accuracy of DT, MLP, and AdaBoost, respectively. In test set 3, the topic classification accuracy of the research method was also the highest, at 99.6 %. In summary, the use of Nvivo12 encoding and NLP toolkit could provide accurate sentiment analysis and topic classification for comments. It provided accurate public sentiment and topic attention for government service optimization.

5. Conclusion

As a global health crisis, the COVID-19 epidemic has attracted people's attention. VKontakte, as a major social media platform in Russia, has a large number of users and its comments are highly representative. Therefore, to understand the public's emotional state and attitude changes during the new epidemic, this study used the Nvivo 12 encoding tool and the NLP toolkit to analyze the emotional tendencies and discourse patterns of VKontakte comments. These analysis results confirmed that during the four stages of the epidemic, there was a fluctuating increase in negative emotions among the public, with a proportion of 45.3 in the fourth stage. In discourse pattern analysis, the average accuracy of research methods in different test sets was above 89 %, which was better than other methods. In the first stage, the main topic in the comments was to prevent virus transmission, and the adjusted residual value was 2.741. According to the analysis results, it can be concluded that the public's emotional reactions showed different characteristics at different stages of the epidemic, such as a shift from negative emotions to positive emotions. This emotional change may be related to the progression of the epidemic, the implementation of prevention and control measures, and how the public perceives and acts toward the epidemic. In addition, the high accuracy of discourse pattern analysis indicated that this research method could effectively identify and classify public discourse patterns during the epidemic period. This was significant for understanding the public's response to, concerns about, and needs regarding the epidemic. It helped the government and relevant departments formulate more targeted policies and measures. In summary, Nvivo12′s encoding and the NLP toolkit can effectively analyze emotional changes and the focus of public discussions during an epidemic. This is crucial for managing public opinion during emergencies. Applying it to the public opinion analysis process of emergencies can help real-time grasp public emotions and opinions, provide basis for decision-making, and take measures in advance to prevent the situation from worsening. Therefore, it can enhance public understanding and trust. This study combines the qualitative analysis of Nvivo12, the contextual understanding of BiLSTM, and the emotion classification method of SVM to create a multidimensional analysis system. This system overcomes the limitations of single-method analysis and can more accurately extract emotional information from social comments. The proposed method has important academic value and social significance in the field of public sentiment analysis during sudden public health emergencies. It not only fills the gap in the integration and application of technology and the study of emotional transmission mechanisms, but also helps improve social governance capabilities and the level of public health emergency management. However, the study only analyzed the comment data on the VKontakte social platform, which may lead to certain regional and cultural limitations in the conclusions. Moreover, the online data collection did not consider the issue of verifying users' personal information, and the results of sentiment analysis may be affected by the problem of water army flooding the screen. In addition, this study mainly analyzes the immediate reactions of public emotions during the epidemic period, and lacks exploration of the long-term impact of public emotions after the epidemic. Based on the results of this study, future research can be conducted in various areas, such as cross-platform comparisons, emotion diffusion, multimodal fusion, long-term tracking, associations between user behavior and emotions, and cross-cultural research. This research can further enrich our understanding of social media users' emotions. Twitter, Facebook and other platforms have their own characteristics in emotional trends due to differences in user demographics and platform usage patterns. Understanding these differences is critical to accurately gauging public sentiment and formulating targeted strategies.

Funding

The research is supported by the Xinjiang Uygur Autonomous Region's 'Tianchi Talent' Young Doctoral Project (5105240150f); Progressive results of the University Internal Cultivation programme in Philosophy and Social Sciences 2024 (24BPY005).

CRediT authorship contribution statement

Jiaxing Han: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Arbieu, U., Helsper, K., Dadvar, M., Mueller, T., & Niamir, A. (2021). Natural language processing as a tool to evaluate emotions in conservation conflicts. *Biological Conservation, 8*(1), 564–575.

Bahr, T. (2021). COVID-19 containment measures analysis. *European Journal of Public Health, 8*(2), 65–78, 31.

C.-H, K., Y.-C, C., & Nguyen, D. L (2022). Predicting aspect-based sentiment using deep learning and information visualization: The impact of COVID-19 on the airline industry. *Information & Management, 59*(2), 12–15.

Chaney, D., & Lee, M. S (2021). COVID-19 vaccines and anti-consumption: Understanding anti-vaxxers hesitancy. *Psychology & Marketing, 7*(9), 231–245.

Che, S. P., & Kim, J. H (2024). Sentiment impact of Public Health Agency communication strategies on TikTok under COVID-19 normalization: Deep learning exploration. *Journal of Public Health-Heidelberg, 32*(8), 1559–1570.

Cheng, Z., & Li, Y. (2024). Like, comment, and share on TikTok: Exploring the effect of sentiment and second-person view on the user engagement with TikTok news videos. *Social Science Computer Review, 42*(1), 201–223.

Damayanti, L., & Lhaksmana, K. M (2024). Sentiment analysis of the 2024 Indonesia presidential election on twitter. *Sinkron: Jurnal Dan Penelitian Teknik Informatika, 8*(2), 938–946.

Floresruiz, D., & Elizondosalto, A. (2021). Using social Media in tourist sentiment analysis: A case study of Andalusia during the Covid-19 pandemic. *Sustainability, 13*(2), 273–287.

Gheisari, M., Hamidpour, H., Liu, Y., Saedi, P., Raza, A., Jalili, A., Rokhsati, H., & Amin, R. (2023). Data mining techniques for web mining: A survey. *Artificial Intelligence and Applications, 1*(1), 3–10.

Han, X., Zhang, Z., & Liu, Z. (2021). Knowledgeable machine learning for natural language processing. *Communications of The ACM, 64*(11), 50–51.

Küchler, C., Stoll, A., Ziegele, M., & Naab, T. K (2023). Gender-related differences in online comment sections: Findings from a large-scale content analysis of commenting behavior. *Social Science Computer Review, 41*(3), 728–747.

Kuznetsova, Y. M (2022). Emotional attitudes towards the components of the digital environment (based on the text analysis of network comments). *RUDN Journal of Psychology and Pedagogics, 19*(2), 253–281.

Liu, S., Wen, A., Wang, L., He, H., Fu, S., Miller, R., et al. (2023). An open natural language processing (NLP) framework for EHR-based clinical research: A case demonstration using the National COVID Cohort Collaborative (N3C). *Journal of The American Medical Informatics Association, 30*(12), 2036–2040.

Luo, H., Meng, X., Zhao, Y., & Cai, M. (2023). Exploring the impact of sentiment on multi-dimensional information dissemination using COVID-19 data in China. *Computers in Human Behavior, 8*(9), 546–558.

Luo, J., Zhang, Y., Gao, Y., & Zhang, J. (2024). A novel method based on knowledge adoption model and non-kernel SVM for predicting the helpfulness of online reviews. *Journal of The Operational Research Society, 75*(6), 1205–1222.

Mirzaei, R., Attar, A., Papizadeh, S., Jeda, A., Hosseini-Fard, S., Jamasbi, E., et al. (2021). The emerging role of probiotics as a mitigation strategy against coronavirus disease 2019 (COVID-19). *Archives of Virology, 166*(7), 1819–1840.

Prasad, J., Jain, A., Velho, D., & KS, S. K (2022). COVID vision: An integrated face mask detector and social distancing tracker. *International Journal of Cognitive Computing in Engineering, 3*, 106–113.

Ramasamy, M., & Elangovan, M. (2024). Optimized neural attention mechanism for aspect-based sentiment analysis framework with optimal polarity-based weighted features. *Knowledge and Information Systems, 66*(4), 2501–2535.

Sasangohar, F., Dhala, A., Zheng, F., Zheng, F., Ahmadi, N., & Masud, F. (2020). Use of telecritical care for family visitation to ICU during the COVID-19 pandemic: An interview study and sentiment analysis. *BMJ Quality & Safety, 30*(9), 715–725.

Saulsberry, L., Bhargava, A., Zeng, S., Gibbons, J. B., Brannan, C., Lauderdale, D. S., & Gibbons, R. D (2023). The social vulnerability metric (SVM) as a new tool for public health. *Health Services Research, 58*(4), 873–881.

Shafi, J., Iqbal, H. R., Nawab, R. M. A., & Rayson, P. (2023). UNLT: Urdu Natural Language Toolkit. *Natural Language Engineering, 29*(4), 942–977.

Shaik, A., Devi, B. A., Baskaran, R., Bojjawar, S., Vidyullatha, P., & Balaji, P. (2024). Recurrent neural network with emperor penguin-based Salp swarm (RNN-EPS2) algorithm for emoji based sentiment analysis. *Multimedia Tools and Applications, 83*(12), 35097–35116.

Sremanakova, J., Sowerbutts, A. M., Todd, C., Cooke, R., Pearce, L., Leiberman, D., et al. (2024). Healthy eating and active lifestyle after Bowel cancer (HEAL ABC)—Feasibility randomised controlled trial. *European Journal of Clinical Nutrition, 78*(12), 1095–1104.

Syropoulos, C., Felbermayr, G., Kirilakha, A., Yalcin, E., & Yotov, Y. V (2024). The global sanctions data base–Release 3: COVID-19, Russia, and multilateral sanctions. *Review of International Economics, 32*(1), 12–48.

Willnat, L., Shi, J., & Coninck, D. D. (2023). Covid-19 and xenophobia in America: Media exposure, anti-Asian stigmatization, and deservingness of Asian immigrants. *Asian Journal of Communication, 33*(1), 87–104.

Xie, M., Liu, Z., & Guo, C. (2022). Effect of the congruity of emotional contexts at encoding on source memory: Evidence from ERPs. *International Journal of Psychophysiology, 173*, 45–57.

Yalva, E. B. K., & Gaynor, K. (2021). Emotional dysregulation in adults: The influence of rumination and negative secondary appraisals of emotion. *Journal of Affective Disorders, 282*(7), 656–661.

Zhang, S., Yu, H., & Zhu, G. (2022). An emotional classification method of Chinese short comment text based on ELECTRA. *Connection Science, 34*(1), 254–273.