International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 6443-6456

https://doi.org/10.70517/ijhsa463554

A Research on the Analysis Model of User Behavior and Emotional Tendency of Social Platforms Based on Support Vector Machines

Ming Lu[1,*], Rongfa Chen[1], Xiuzhe Meng[1] and Kai Yang[2]
[1] College of Management and Economy, Tianjin University, Tianjin, 300072, China
[2] Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
Corresponding authors: (e-mail: jarratt_cn@163.com).

Abstract With the rapid development of social media, massive user behavior data and sentiment expression have become important research resources. Existing sentiment analysis methods still have limitations in dealing with complex text and user behavior features. In this study, we constructed a support vector machine-based model for analyzing user behavior and sentiment tendency of social platforms, processed unstructured text data through vector space model, and built a high-dimensional mixed-feature sentiment classifier by combining mutual information value feature extraction and latent semantic analysis. In terms of methodology, firstly, text data is preprocessed by vector space model, and data annotation is performed by using lexicon and weakly labeled information; secondly, feature extraction is performed by using mutual information value computation, and SVM algorithm is used to construct sentiment classifiers; finally, empirical analysis is performed to analyze the user behaviors and emotional tendencies of social platforms. The results of the study show that: short video playing, liking, commenting and sharing behaviors conform to the power law distribution with long-tail effect; the actual conversion rate of short video is only 2.83%, which shows that the user participation is low; in terms of sentiment analysis, the sentiment density of user J in the 38-day observation period reaches 0.8957227, which is significantly higher than that of other users; and the characteristic value of the sentiment transmissibility of user X is 2.83, which is significantly higher than that of other users. The conclusion of the study shows that the constructed high-dimensional mixed-feature SVM model can effectively reflect users' behavioral characteristics and emotional tendencies, providing a technical method for social platform user behavior prediction, emotion monitoring and crisis warning.

Index Terms Social platform, User behavior analysis, Emotional tendency, Support Vector Machine (SVM), Feature extraction, Latent semantic analysis

I. Introduction

In recent years, the explosive development of information science and technology makes the interaction between people from the reality of the physical world gradually penetrate into the network world, people are not satisfied with passive access to information from the network, but tend to take the initiative and randomly to the network world to publish information, disseminate information, and to realize the social interaction between people [1]-[3]. This is due to this warm and frequent interaction between people and networks, networks and people, and people and people in the computer makes the Internet Web 2.0 services came into being. The birth of Web 2.0 undoubtedly adds fun to the Internet world, and at the same time gave birth to many Internet social platforms, such as Facebook, YouTube, blogs, microblogs, Tik Tok, Xiaohongshu, etc. [4], [5]. The whole social network is developing towards the goal of transferring people's life, information flow, and reducing its cost after transferring the information flow from offline to online, thus realizing the intersection and integration of real socialization and online socialization [6].

Social platform has penetrated into people's daily life, it not only affects the way of information acquisition and transmission, but also has an inestimable impact on people's friendship, life and thinking way. Therefore, the analysis of user behavior and emotional tendency of social network platform information has become an important research tool to grasp hot topics. Through the analysis of user behavior and emotional tendency of social platforms, on the one hand, we can instantly obtain the direction of public opinion in a certain region, a certain type of group, or a certain event during a certain period of time, especially in recent years, the frequent occurrence of cyber violence, the release of false information, and the discussion of emergencies after the occurrence of an emergency have caused panic and interference in the society, which influences the management of the society [7]-[10]. On the other hand, user profiles can be constructed to predict the demand for products and services, develop accurate marketing plans, achieve personalized services and product optimization, and enhance the competitiveness of products and enterprise markets [11]-[13].

With the development of intelligent algorithms, machine learning techniques are widely used, among which Support Vector Machine (SVM) is a common machine learning algorithm which can be used for classification and regression problems. It performs well in dealing with high-dimensional data and small-sample data, so it is widely used in the fields of image recognition, text categorization, bioinformatics, etc., and also has advantages in sentiment analysis and user behavior analysis [14], [15]. Literature [16] used the chi-square method to deeply analyze customer profiles and behavioral data collected on social platforms, revealing the influence relationship between customer profiles and products, and customers and products. Literature [17] proposed a mathematical model of opinion dynamics for obtaining the individual behavior of the public's posting information and information dissemination in social networks, and analyzed this in conjunction with the Porter-Hamilton system theory to effectively analyze the behavior of public opinion in the context of cooperative competition. Literature [18] analyzed the activity behavior of users on social networking sites through the developed nine-factor analysis method, genetic weighting-K-mean clustering, and negative selection algorithm, explored the presence of criminal behavior, and recommended people whose difficulties belong to criminals, which provided convenience for public security management. Literature [19] designed a sentiment analysis model with neural networks for the determination of textual sentiment-related data shared by users in social platforms, constructed an emotional tendency model to analyze user textual and behavioral data based on these data, and upgraded the model using association rule mining to achieve user behavior and emotional tendency analysis.

In terms of sentiment tendency analysis, literature [20] developed an OCC sentiment rule system for online opinion sentiment tendency analysis, which was realized by Word2Vec and convolutional neural network, and the accuracy of the sentiment tendency analysis of this system was improved by 3%-8% compared with manual annotation. Literature [21] explored the influence factors of public emotional tendency and its sentiment changes in early tweets of the epidemic using two-way long and short-term memory model and logistic regression analysis. Literature [22] analyzed the sentiments of different text granularities in social platforms and classified these sentiments using stochastic subspace integration learning with binary particle swarm optimization as a way to analyze the sentiment tendencies of users on the platforms. Literature [23] applied deep learning techniques incorporating ChatGPT4 to reveal the sentiment tendencies of Thai people towards the Thailand-China high-speed railroad project on YouTube platform based on time series data.

As an important carrier of information dissemination and emotional communication, social platforms have become an indispensable part of people's daily life. With the rapid growth of user scale, social platforms have generated a huge amount of user behavioral data and emotional expression content, which not only reflect the user's emotional state and behavioral preferences, but also contain rich social psychology information. However, the text data in social platforms are mostly unstructured and integrate the multidimensional behavioral characteristics of users, which brings great challenges to sentiment analysis. Most of the current sentiment analysis methods are limited to single text content or simple sentiment lexicon statistics, which are difficult to effectively deal with the complex context and diverse user behaviors in social platforms. In the face of this problem, this study proposes a support vector machine-based model for analyzing user behavior and sentiment tendency in social platforms, and constructs a more complete framework for user sentiment analysis by integrating text feature extraction, latent semantic analysis, and user behavioral feature analysis. Firstly, this study adopts vector space model to preprocess the text data of social platforms, so that the unstructured data is transformed into a structured form that can be processed by computers. On this basis, the text is sentiment labeled by combining lexicon and weakly labeled information, and the modifying effect of negative words and degree adverbs on sentiment expression is considered. Secondly, this study carries out feature extraction through the mutual information value calculation method, and selects lexicality, degree adverbs, negation words, positive emotion words, and negative emotion words as the key feature dimensions, which effectively reduces the dimension of the feature space and improves the classification efficiency. At the same time, to address the problem of “multiple meanings of words”, the potential semantic analysis technique is introduced to further explore the deep semantic information of the text. Based on the extracted high-dimensional hybrid features, this study adopts the support vector machine algorithm to construct a sentiment classifier, which has a significant advantage in dealing with high-dimensional sparse vector matrices. In the empirical analysis part, this study analyzes the behavioral characteristics of users on social platforms from the dimensions of time series and user preferences, revealing the regularity and paroxysm of users' behaviors; it explores the users' emotional tendency from the dimensions of emotional density, implicit emotional features, and explicit emotional features, and constructs a multi-dimensional emotional portrait of users. The study not only enriches the research framework of user behavior and emotion analysis of social platforms theoretically, but also provides technical support and methodological guidance for the optimization of social platform operation, monitoring of public opinion, and improvement of user experience in practice. By digging deeper into the intrinsic connection between user behavior and emotion, this study contributes to a more comprehensive understanding of the psychology and behavioral patterns of social platform users, and provides new ideas for research and application in related fields.

II. Social platform user behavior and sentiment tendency analysis model

II. A. Social Platform User Emotion Extraction

II. A. 1) Preprocessing of Social Platform Text Data

Text is usually expressed as a string, which expresses rich information, but cannot be directly used for sentiment analysis. Data preprocessing is a necessary stage of text sentiment tendency analysis, the main purpose is to process the massive unstructured data that cannot be recognized by the computer, so that it can meet the requirements of computer processing.

In this paper, the preprocessing of data can be done in the way of vector space model (VSM). Vector space is mostly utilized for natural language query, based on this, the query result can be processed as a small piece of information, then a certain item of information within the vector space can be represented as:

$$D_j = (w_{j1}, w_{j2}, \cdots, w_{jn}) \tag{1}$$

In the formula, n represents all indexed items, and w_{jn} denotes the weights of indexed items within information items.

Set D_j as the text item, k_i as the index item, the frequency of occurrence of k_i in D_j is tf_{ij}, and the inverse document rate is idf_i, the more the number of text items is, the smaller the rate of inverse document is, and the better the distinguishing ability of the word w, where for the computation of the weight of the index item is carried out by utilizing the TD-IDF method, and its computation formula is as follows:

$$w_{ij} = tf_{ij} * idf_i \tag{2}$$

where the expression for idf_i is as follows:

$$idf_i = \ln \frac{N}{n_i} \tag{3}$$

In the formula, N represents all text items and n_i is the number of information items indexing k_i.

In order to make the structured characteristics of data more obvious, the formula for calculating the weights of indexed items is improved, and the improved expression is as follows:

$$W(t,d) = \frac{tf(t,d)^* \ln\left(N / n_t + 0.01\right)}{\sqrt{\sum_{t\in d}\left[tf(t,d)^* \ln\left(N / n_i + 0.01\right)\right]^2}} \tag{4}$$

In summary, for the user model that needs to be queried, if through the VSM model can be represented as:

$$Q = (w_i, w_j, \cdots, w_n) \tag{5}$$

Therefore the vector similarity can be calculated using the matching degree model of the user model Q and the text D_j, the degree of similarity of these two vectors depends on the angle of the vector pinch, which is then calculated as:

$$\cos\theta = \frac{v_1 - v_2}{\|v_1\| - \|v_2\|} \tag{6}$$

Assuming that all index items k_i are independent of each other, data preprocessing can be completed through the calculation of the above similarity measure, so that all text information has structured characteristics, laying the foundation for characteristic text feature extraction.

II. A. 2) Data annotation based on dictionary and weakly labeled information

In this paper, we use a combination of lexicon and weakly labeled information to analyze the sentiment of social platform user comment data.

For a single comment, this paper first processes it with participle and lexical labeling, and then finds the sentiment words in the comment according to the domain sentiment dictionary constructed in the previous section. If a sentiment word is found, the position of the sentiment word is marked, and then the negative words and degree adverbs modifying the sentiment word are looked up forward, so that each sentiment word and its related negative words and degree adverbs are called a sentiment word class. Its sentiment analysis algorithm is described as follows:

1) Comment preprocessing, including word division and lexical labeling

2) While emotion words do, find the negative words and degree adverbs that modify the emotion words, and calculate the emotion word class sentiment value.

3) Calculate the comment sentiment value, the sentiment value is greater than or equal to 0, the comment sentiment tendency is marked as 1, otherwise it is marked as -1. The formula for calculating the sentiment value of each sentiment word class is:

$$s(w) = n(w) \times d(w) \times p(w) \times l(w) \tag{7}$$

where, $s(w)$ denotes the sentiment polarity of the sentiment word class w; $n(w)$ denotes the sentiment weight of the negation. A negative word indicates a reversal of sentiment, but a double negative sentiment is unchanged, and $n(w)$ is -1 when the number of negatives is odd, and 1 when it is even, which is calculated as in equation (8):

$$n(w) = \begin{cases} 1, & count(neg)\%2 = 0 \\ -1, & count(neg)\%2 \neq 0 \end{cases} \tag{8}$$

where $count(neg)$ denotes the number of negatives. $d(w)$ denotes the cumulative sum of the weights of multiple degree adverbs modifying the sentiment word, which is calculated as in equation (9):

$$d(w) = \sum_{i=1}^{m} dg_i \tag{9}$$

where, m is the number of degree adverbs modifying the sentiment word class, and dg_i is the degree adverb weight.

$p(w)$ denotes the polarity of the emotion word, positive emotion word is 1, negative emotion word is -1; $l(w)$ denotes the relative position of the negative word and the degree adverb, the position between them is different, the emotion is different, for example, commenting on “this movie is very bad” and “this movie is not good”, expresses a completely different emotion; when the negative word is in front of the degree adverb, it expresses a completely different emotion; when the negative word is in front of the degree adverb, it expresses a completely different emotion. For example, commenting “this movie is very bad” and “this movie is not very good” express completely different emotions; when the negative word is in front of the adverb of degree, $l(w)$ is set to 0.5, and vice versa, $l(w)$ is set to -1, which is calculated as in equation (10):

$$l(w) = \begin{cases} 0.5, & loc(neg) < loc(dg) \\ -1, & loc(neg) > loc(dg) \end{cases} \tag{10}$$

where, $loc(neg)$ denotes the position of negation in the sentiment word class, $loc(dg)$ denotes the position of adverbs of degree in the sentiment word class, and “<” denotes that the relative position is anterior and “>” denotes that the relative position after.

A single comment consists of multiple sentiment word classes, so the sentiment polarity of a single comment is calculated as in equation (11):

$$sen(r) = \sum_{w \in r} s(w) \tag{11}$$

where, r denotes all the sentiment lexical categories in a single comment, and $sen(r)$ denotes the sentiment polarity of a single comment. Using equation (11), this paper can calculate the lexicon-based sentiment value of each comment, and $sen(r) \geq 0$ denotes the positive sentiment of the comment and vice versa for the negative sentiment.

II. B. Mutual information value based feature extraction and dimensionality reduction

II. B. 1) Emotional Resource Building

This paper analyzes some of the negative words, conditional word transitions, etc. in the literary texts to establish emotional resources as shown in Tables 1 to 3.

Table 1: Denser

Never	Not necessarily	Difficulty	No	No
Moo	Nothing	Not	No	No
Don't	No	No	No	Not enough

Table 2: Conditional list

Like it	Given that	Hypothesize	If	If
For example	Once	As long as	Provided that	If

Table 3: Turning point

But	But	Well	However
Or	However	But	However

The degree adverbs were collected and different intensity values were calculated and the degree adverbs and intensities are shown in Table 4.

Table 4: Degree and strength table

Degree adverb	Strength value
light, any	0.27
A little, A little bit	0.55
More, More and more	1.3
Absolute, Full	2.2

II. B. 2) Feature extraction rules

In this paper, a total of five feature extraction rules are formulated as follows:

(1) Keyword features in the text, determine the number of nouns, verbs, adjectives and adverbs.

(2) Negative word characteristics, the total number of times to the remainder of 2, if the value is 1, this time the polarity of the emotion word is reversed; if the value is equal to 0, the polarity of the emotion word does not change.

(3) Degree adverbial property, if there is a degree adverb in front of the emotion word, the intensity of the emotion of the text changes accordingly.

(4) Conditional syntactic properties, if there is a conditional word in front of the emotion word, the value of the conditional syntactic properties is equal to 1, and vice versa is 0. After analyzing the random corpus, the conditional word has a weakening effect on the intensity of emotional expression.

(5) Transitive syntactic properties, when there is only one kind of transitive word in front of the emotion word, the polarity of the emotion word is judged to be inverse; if there are two kinds of words, it means that no change occurs.

II. B. 3) Text Feature Extraction

Feature extraction is expressed in mathematical form as selecting a true subset $T' = \{t_1, t_2, \cdots, t_n\}$ from the initial set of features $T = \{t_1, t_2, \cdots, t_{n'}\}$, where $n' \le n$, the number of extracted features is less than the initial number of features [24]. In this paper, text feature extraction is performed by calculating the mutual information value (MI).

The mutual information value essentially describes the correlation existing between two sets of events, for feature extraction, the mutual information shows the correlation existing between the lexical item t' and the category c_i, and the expression is as follows:

$$MI(t', c_i) = \ln \frac{p(t' \mid c_i)}{p(t)} \tag{12}$$

The global MI value for lexicon t' with all categories is calculated as:

$$MI(t') = \sum_{i=1}^{m} \ln \frac{p(t' \mid c_i)}{p(t)} \tag{13}$$

In the formula, i denotes the number of categories, $p(t)$ represents the number of times t appears in the training set, and $p(t' \mid c_i)$ denotes the probability that t' and c_i appear together. For all the calculated MI values, they are sorted from high to low, and the words with higher thresholds are treated as feature words.

II. B. 4) Feature selection

The core of machine learning based sentiment analysis is feature selection, which is related to the accuracy of sentiment classification. At present, the common feature selection are: one-word features, binary word features, three-word features, word frequency, word properties, sentiment words and so on. Among them, the feature dimensions of monadic word features, binary word features, and ternary word features are related to the amount of corpus, and when the corpus is large, the feature dimensions will reach the level of thousands of dimensions, which is difficult to deal with; the word frequency can reflect the importance of a word, but not all the words are related to the text emotion, and the introduction of the word frequency will lead to the noise of the data.

In this paper, we choose the five feature dimensions of lexicality, degree adverbs, negative words, positive emotion words and negative emotion words, in which a text is composed of multiple words and their lexicality, and lexicality plays a big role in it; emotion words are the key core of a text's emotion classification, and negative words usually make a text's emotion polarity reversed; at the same time, the degree adverbs can change the intensity of the emotion words, when a text At the same time, adverbs of degree can change the intensity of emotion words, when both positive and negative emotion words appear in a text, it is difficult to judge the emotional tendency of a text if only relying on the polarity of emotion words, and adverbs of degree can help to make a choice.

II. C.Latent Semantic Analysis

After the above feature extraction, we can obtain the more significant features of the text, but we have not considered the problem of "multiple meanings of the word", therefore, we need to further study the potential semantics.

Firstly, the following probability variables are determined: $P'(D_j)$ represents the probability of selecting text D_j in the text set, $P'(w_j \mid z_k)$ represents the conditional probability of a word w_j under the constraints of latent variable z_k, and $P'(z_k \mid D_j)$ is the probability distribution of text D_j in the latent variable.

Based on the above definition, a generative model can be composed after the following steps:

Step 1: Combine $P'(D_j)$ to randomly select a text D_j.

Step 2: Based on the text D_j, select the potential variable z_k expressed by the text by means of $P'(z_k \mid D_j)$.

Step 3: Obtain a pair (d_i, w_j) of observed variables without latent variables, and change the generation process into the form of joint probability distribution:

$$P'(D_j, w_j) = P'(D_j) P(w_j \mid D_j) \tag{14}$$

$$P'(w_j \mid D_j) = \sum_{j=1}^{\infty} P'(w_j \mid D_j) \tag{15}$$

After constructing the generative model, the parameters are determined by the maximum similarity expression for potential semantic mining. The expression is as follows:

$$L = \sum_{i=1}^{N} \sum_{j=1}^{M} n(D_j, w_j) \log P'(D_j, w_j) \tag{16}$$

II. D.SVM High-Dimensional Mixed-Feature Sentiment Classifier

II. D. 1) Theoretical foundations

Sentiment short text obtained after feature extraction is a high-dimensional sparse vector matrix, which is directly used as the training and testing data of the classifier, and the SVM algorithm suitable for dealing with large-scale text categorization is chosen to construct the sentiment classifier. Given a set of sample sets $\{x_i, y_i\}$, $i = 1, 2, \cdots, l$, $x_i \in R^n$, $y_i \in \{-1, +1\}$, the SVM needs to solve the following unconstrained optimization problem:

$$\min_{w} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi(w; x_i, y_i) \tag{17}$$

where $\xi(w; x_i, y_i)$ is the loss function; C is the penalty coefficient; and l is the total number of samples.

The standard C-SVM (L1-SVM) is usually used as an effective classification algorithm in classification problems.The loss function of the L1-SVM is a first-order paradigm, whereas the loss function of the second-order L2-SVM adds a dual method of a Hessian matrix inverted by the diagonal matrix of the penalty factors. This improves the stability of the solution process.The loss function formulas for L1-SVM and L2-SVM are given below respectively:

$$\max(1 - y_i w^T x_i, 0) \tag{18}$$

$$\max(1 - y_i w^T x_i, 0)^2 \tag{19}$$

It is common to add a bias term b to the classification problem of SVMs, and the bias term b is treated in the text as shown below:

$$\begin{aligned} x^T &\leftarrow [x_i^T, B] \\ w^T &\leftarrow [w^T, b] \end{aligned} \tag{20}$$

where B is a constant.

Eq. (17) is called the original form of the SVM, which is transformed into the dyadic form in the solution:

$$\begin{cases} \min\limits_{\alpha} f(\alpha) = \frac{1}{2} \alpha^T \bar{Q} \alpha - e^T \alpha \\ s.t.\ 0 \le \alpha_i \le U, \forall i \end{cases} \tag{21}$$

where e is the all-1 matrix; $\bar{Q} = Q + D$, D is the diagonal matrix, and $Q_{ij} = y_i y_j x_i^T x_j$.

II. D. 2) Framework implementation

The representation of sentiment text features is a key step in sentiment classification, including three parts: preprocessing, Chinese word segmentation, and feature extraction.

The experimental idea: firstly, crawling comments, microblogs and other data from the target website for labeling; then using the k-fold crossover method for training and testing; finally, after the sentiment classifier outputs the sentiment polarity (positive, negative, neutral), and statistics of the experimental results.

III. Analysis of social platform user behavior and emotional tendencies

III. A. Statistical Analysis of User Behavioral Characteristics

III. A. 1) Time series analysis

To collect data, this collection period is 7 days in total, and by counting the final value of the last day, it is found that there are orders of magnitude differences between the number of playbacks, likes, comments and shares, excluding a total of 8 data that close or restrict the comment function (the lowest number of comments is 1). Therefore, the user's behavioral preference can be obtained: play>like>comment>share, as shown in Fig. 1~Fig. 4, in which the black curve is the change trend. Meanwhile, according to the mean value statistics, the conversion rate of short videos = (number of likes + number of comments + number of shares)/playback, the actual conversion rate is only 2.83%, indicating that the majority of users still stay at the browsing level, and the participation in the video's behavior of liking, commenting and sharing is relatively low.

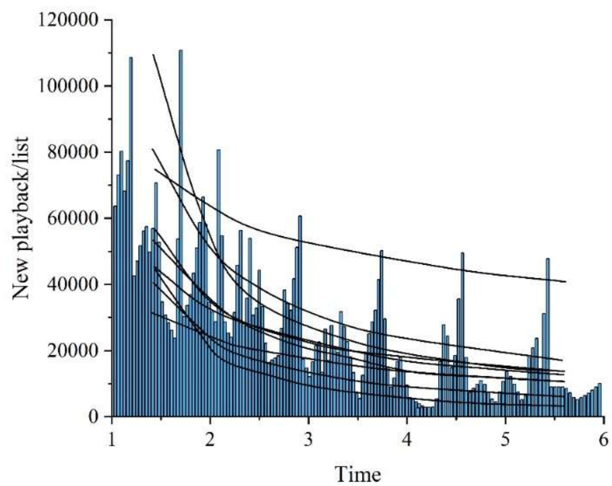

Figure 1: Time sequence distribution of new playback quantity

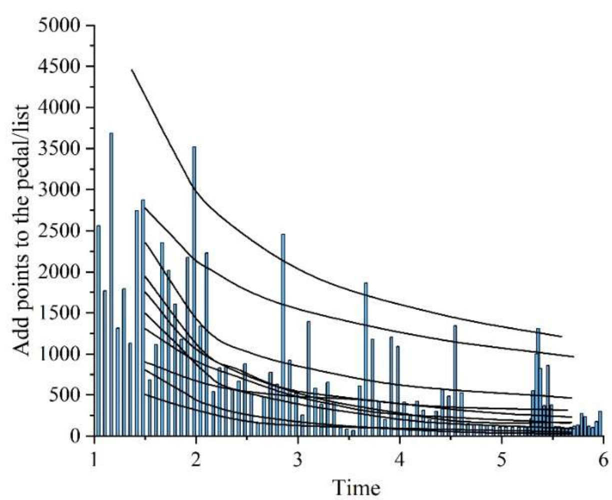

Figure 2: The time sequence distribution of new zazes

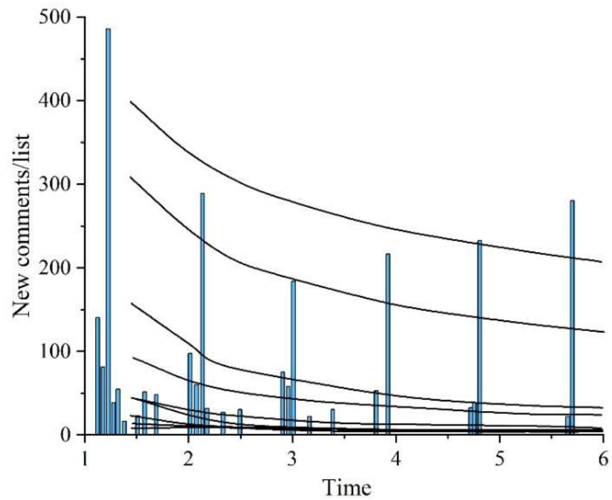

Figure 3: The time sequence distribution of new comments

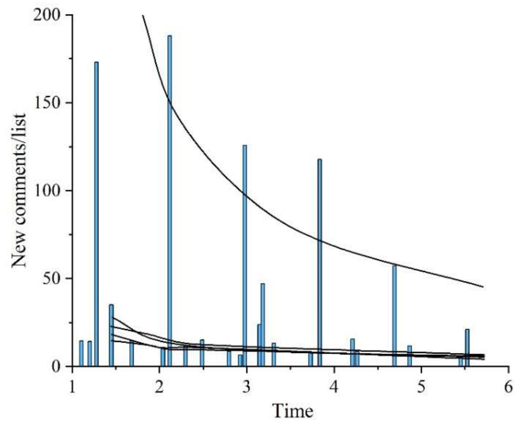

Figure 4: The time sequence distribution of the added share

The time series distribution of short videos posted by users is shown in Figure 5. Through the statistics of video posting time, it can be found that users' posting behavior is concentrated in the time period from 8:00 to 23:00, and reaches its peak from 18:00 to 20:00, and the rest of the time reaches a small peak from 10:00 to 14:00 and 20:00 to 22:00 respectively, which is exactly in line with people's daily work and rest patterns. Every day from 0:00 to 8:00 is the sleep time, so in addition to working hours most users will post videos during lunch breaks or night breaks, this behavior is more in line with people's habits of entertainment and leisure, and this type of video receives a relatively higher degree of attention.

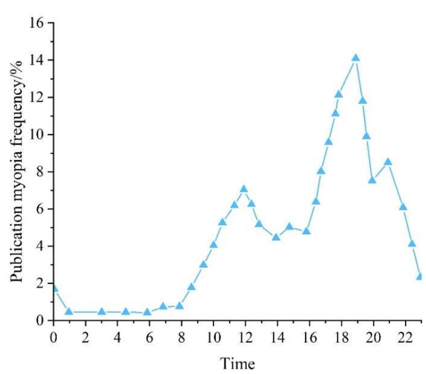

Figure 5: The user publishes the time sequence distribution of the short-time frequency

Due to the limitation of the number of statistics, this paper chooses to take the total list data as an example, in which the statistics of the number of comments removes a total of 16 data that closes or restricts the comment function. From the time series distribution of new playback, new likes, new comments and new shares, it can be seen that the trend line fitting is in line with the law of power law distribution, so that the playback, likes, comments and sharing behaviors over time obey the power law distribution, and from the perspective of the trend of change, they all have the long-tail effect.

Since the ranking data can only reflect the overall trend of change, and it is difficult to reflect the user behavior characteristics of individual short videos, this paper selects the individual short video data of Li Ziqi, a popular user of "Meipai" short video self-media with a high degree of attention, to conduct a specific analysis and explore the behavioral characteristics of the obtained data. Specific analysis, to explore the pattern of change in behavior as shown in Figure 6 to Figure 9.

As can be seen from the figure, the change trend of browsing, liking, commenting and sharing behaviors is relatively consistent, which indicates that users' behaviors have certain regularity and bouts. Within one hour of the release of the short video, the number of plays, likes, comments and shares is the highest, indicating that the attention of the short video reaches its maximum value during this period. The respective numbers then declined due to the rest period from 0:00 to 8:00, but showed a clear growth trend again during the day. In summary, it can be seen that the short videos receive the most attention within one hour of release, and then change according to the user's rest and rest patterns, and finally show a clear long-tail effect.

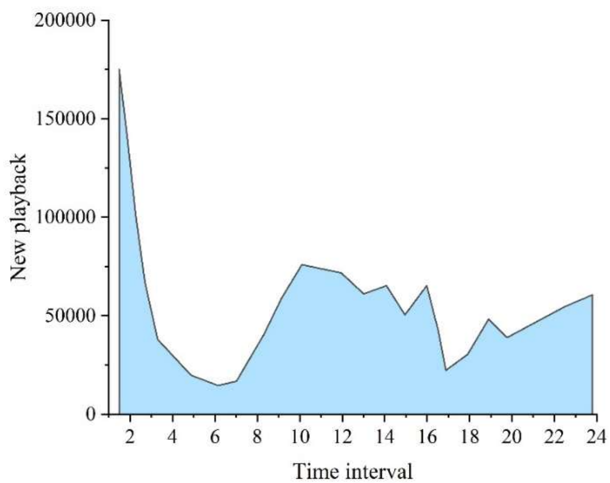

Figure 6: The time sequence distribution of the new playback amount of a single user

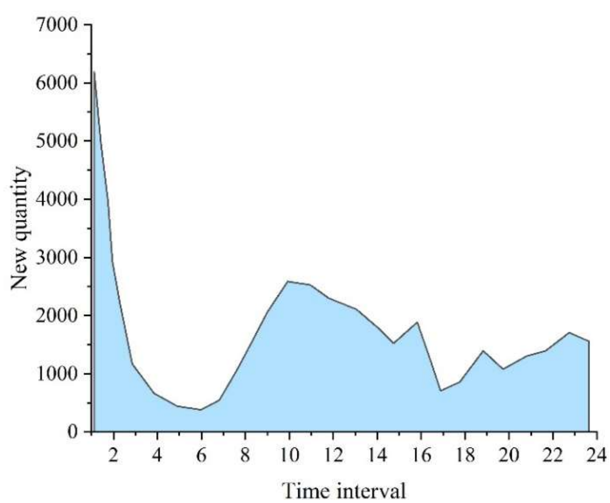

Figure 7: The time sequence distribution of the new number of new likes for a single user

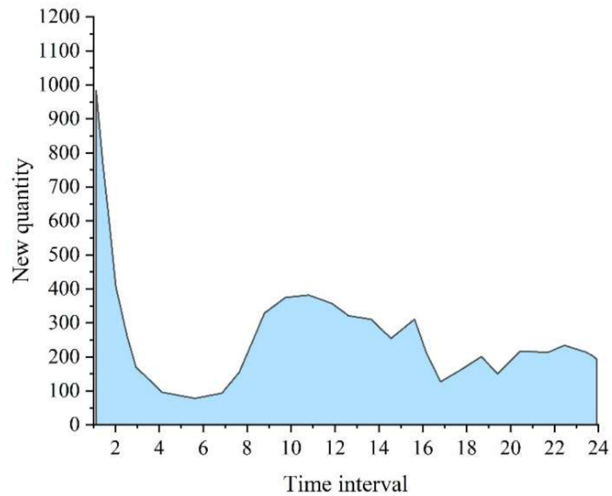

Figure 8: The time sequence distribution of individual users' new comments

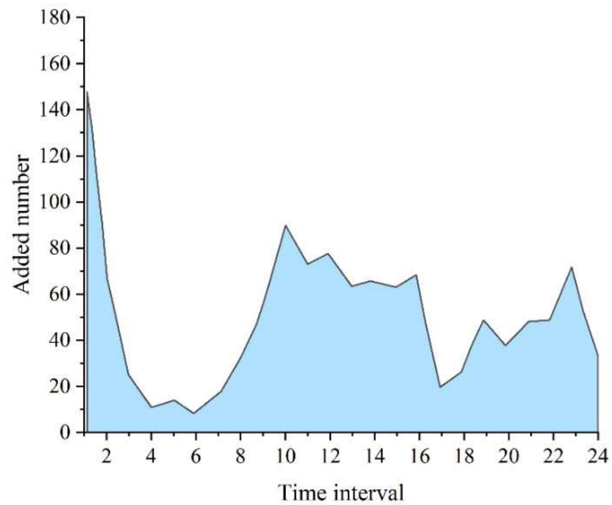

Figure 9: The time sequence distribution of individual users with new sharing

III. A. 2) User behavioral preference analysis

The short video data collected includes the number of likes, comments and shares of the short video, and in order to have a clearer understanding of the user's behavioral preferences, this study introduces the concept of “the number of attention to the short video”. It is expressed by the formula:

Attention number of short videos = number of likes + number of comments + number of shares

The attention number of short videos represents the hotness of the video. The more users' likes, comments and shares, the more attention they pay to the short video, and the higher the heat of the short video. The scatter distribution of the attention number of short videos based on double logarithmic coordinates is shown in Figure 10. From the figure, it can be seen that the number of attention of short videos obeys the power law distribution, and the number of attention of most short videos is relatively small, concentrated in the interval of 110 to 5200. Only a small number of short videos are heavily liked, commented and shared, with the highest number of times reaching more than 81000. In practice, ordinary users tend to pay attention to those short videos that have received a high number of likes, comments or analysis, which results in the polarization of short video information in the dissemination process of the “Matthew effect”. Such behavioral characteristics are generally called “cascade characteristics”, i.e., most users will give priority to popular microblogging in the process of liking, commenting and sharing, which greatly reduces the time interval between events, which also better explains the cause of the power law distribution.

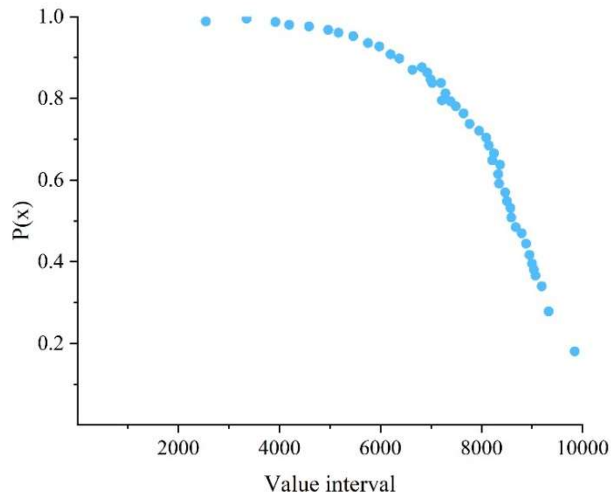

Figure 10: The short-frequency attention of the two-logarithmic coordinate is distributed

According to the definition of the number of attention of short videos above, the short videos in each category are categorized and analyzed, then the average number of attention of short videos in each category is expressed as:

Average number of short videos receiving attention in this category = (total number of likes + total number of comments + total number of shares) / total number of short videos in this category

The average number of attention of short videos by category is shown in Figure 11. From the perspective of content type, “Meipai” short videos cover a wide range of information such as entertainment, news, sports, life and so on. In the era of “pan-entertainment”, the browsing interests of user groups are different, and the number of attention to the corresponding short-video self-media information content is also very different. For example, on the “Meipai” platform, funny, food, eating show and handicrafts receive high attention, while cute pets, travel and sports receive less attention. This is related to the platform's own content layout and the main user's interest orientation, while the short video self-media has adopted a highly accurate personalized recommendation system, and the promotion of content of interest to the user is stronger, so the information flow of the content with a higher degree of concern revolves quickly and efficiently, while the information flow of the content with a low degree of concern is slow and inefficient, which has formed the “Cocoon effect”.

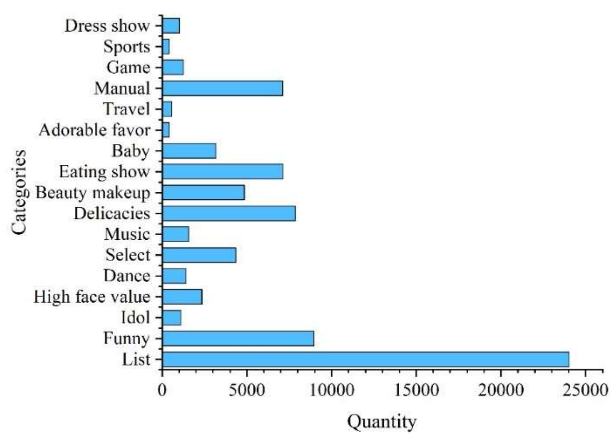

Figure 11: The average attention count chart of the category classification

III. B. User Sentiment Analysis

In order to evaluate the effectiveness of the model in this paper, this section will conduct experiments and analyze the results according to the user sentiment space model.

Firstly, web crawler is used to crawl the user data of Sina Weibo, in order to ensure the timeliness of user data, this paper only extracts the user data within half a year. Here, X, D, and J are used as user nickname codes, respectively.

III. B. 1) Emotional density analysis

This section counts the sentiment frequency of the three users. Sentiment frequency refers to the total number of tweets sent by a user in a single day, which reflects how often the user expresses his/her emotions. Figure 12 shows the statistical comparison of the trend of emotion frequency of users of the three social platforms from 2021-12-1 to 2023-1-7 for a total of 38 days. From the figure, it can be seen that the emotional frequency peak of user D is the largest, and the emotional frequency fluctuation is also larger, and the frequency of microblog messages sent by this user from 2021-12-1 to 2023-1-7 is larger.

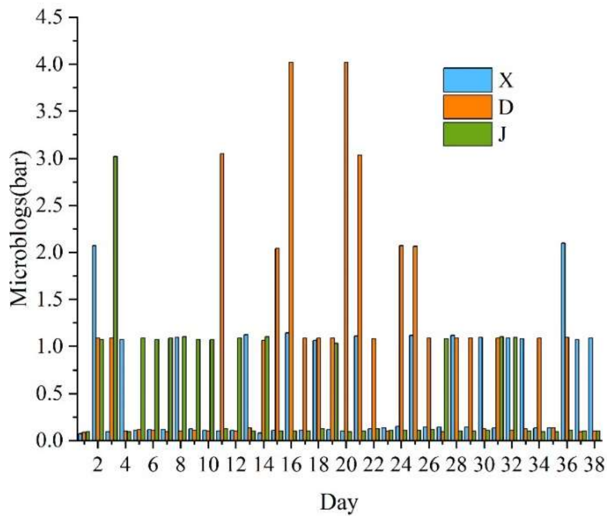

Figure 12: User emotional frequency performance statistics

Emotion density is the average daily emotion frequency value of users over a period of time, and the statistics of emotion density of users corresponding to Figure 12 are shown in Table 5. From the table, it can be seen that the emotion density of user J in the above time period is significantly higher than that of users X and D, which indicates that user J is more inclined to use microblogging to express emotions in that time period.

Table 5: Statistics on the emotional density of each user

	X	J	D
Affective density ω	0.4729583	0.8957227	0.4175441

III. B. 2) Implicit Emotional Characterization

In this section, the results of the experiments based on the values of each feature in the implicit emotion space are discussed, and the statistics of the implicit emotion feature values of each user are shown in Table 6.

From the table, it can be seen that the ei value of user X is much larger than the values of users D and J. According to the correspondence between user characteristics and implicit emotional features, it indicates that the user has a very strong emotional transmissibility. By comparing its personal data, it is found that the number of fans of this user is 8921231, and the number of concerned users is 578, which does have strong public opinion appeal and influence on the microblogging platform, and should be the key monitoring object of public opinion transmission guidance and supervision. Its da value is relatively large, indicating that it has a strong controllability and the probability of needing external intervention is small.

User D's cu and en values are both relatively large, indicating that the emotional fluctuation of this user D is large, the stability of emotion is low, and it may be necessary to use external intervention in its emotional state when necessary. Its ei value is small, indicating that its emotional transmissibility is relatively small, and through personal data verification, it is found that its number of followers is only 136, and its influence on the microblogging platform is very small, which is in line with the quantitative judgment rule of transmissibility.

User J has a relatively large ei value, and through comparing personal data, its number of followers is 7831277 and the number of concerned users is 1187, which also has a strong transmissibility on the microblogging platform.

Table 6: Different user implicit emotional spatial eigenvalue statistics

	X	D	J
ei	2.83	0.0245	0.1781
cu	0.0933	0.1352	0.0996
da	0.271	0.157	0.163
en	0.00077	0.000141	0.00109

III. B. 3) Explicit Sentiment Profiling

Table 7 shows the mean statistics for each emotion categorization on the explicit emotion space for the three users. It can be seen from the table that the average values of "happiness" and "shock" of user X are much higher than the average values of other negative emotions, while the average values of "evil" emotions of users D and J are larger. The results of the analysis of the sentiment averages broadly fit the user sentiment profile.

Table 7: Different users' emotional classification mean statistics

	X	D	J
Joyfulness	0.0287	0.019	0.0113
Excitation	0.0379	0.047	0.054
Anger	0.0005	0.0011	0.0009
Mourning	0.0047	0.0114	0.0151
Fear	0.0033	0.0051	0.0032
Revulsion	0.0145	0.0363	0.0195
Surprise	0.038	0.0	0.0

From the above analysis, it is shown that the high-dimensional hybrid feature classification model can reflect the actual emotional characteristics and entity characteristics of users to a certain extent.

IV. Conclusion

The study constructed an analysis model of user behavior and emotional tendency of social platform based on support vector machine, and the following conclusions were drawn through empirical research: firstly, the user behavior of social platform presents obvious regularity and paroxysm, and the order of user behavioral preference of short videos is playing, liking, commenting, and sharing, in which the time series distributions of the amount of playback, the number of liking, the number of commenting, and the number of sharing are all in line with the law of power law distribution, and there is an obvious long-tail effect. Short video release behavior is concentrated in the time period from 8:00 to 23:00, and reaches its peak from 18:00 to 20:00, which is highly consistent with people's daily routine. Secondly, users' emotional expressions show individual differences. In terms of emotional density, user J reaches 0.8957227, which is significantly higher than user X's 0.4729583 and user D's 0.4175441; in terms of implicit emotional characteristics, user X's emotional transmissibility (ei value) is 2.83, which is much higher than the other users, which matches with its social influence of 8.91 million followers; In terms of explicit affective features, different users showed their own prominent affective categories, with user X's “joy” and “surprise” affective values of 0.0287 and 0.0380, respectively, while users D and J were prominent in “evil” affective aspects. The high-dimensional mixed-feature SVM classification model developed in this study can effectively capture the emotional characteristics of users, and combined with the analysis of user behavior, it provides a technical method for the prediction of user behavior, emotion monitoring and crisis warning on social platforms, and at the same time, it provides a theoretical basis for the personalized recommendation of social platforms, precision marketing and user experience optimization.

References

[1] Roblek, V., Meško, M., Bach, M. P., Thorpe, O., & Šprajc, P. (2020). The interaction between internet, sustainable development, and emergence of society 5.0. Data, 5(3), 80.

[2] Shukhratovna, U. M., & Bayazovna, G. N. (2025). FOREIGN EXPERIENCE IN THE DEVELOPMENT OF MOBILE INTERNET. INNOVATION IN THE MODERN EDUCATION SYSTEM, 6(49), 250-256.

[3] Just, N., & Latzer, M. (2017). Governance by algorithms: reality construction by algorithmic selection on the Internet. Media, culture & society, 39(2), 238-258.

[4] zhou Hua, X., Hasan, N. A. M., & De Costa, F. (2023). The Performance of Corporate Social Responsibility Communication in the Web2. 0 Era: A Bibliometric Analysis of CSR Communication in Social Media Field. Studies in Media and Communication, 11(7), 233-250.

[5] Fang, Y. (2020). In the Web2. 0 Era, Social Media Marketing Brings Business Opportunities to Enterprises. Open Journal of Business and Management, 8(02), 923.

[6] Marler, W. (2022). "You can connect with like, the world!": Social platforms, survival support, and digital inequalities for people experiencing homelessness. Journal of Computer-Mediated Communication, 27(1), zmab020.

[7] Wang, S., Zhu, X., Ding, W., & Yengejeh, A. A. (2022). Cyberbullying and cyberviolence detection: A triangular user-activity-content view. IEEE/CAA Journal of Automatica Sinica, 9(8), 1384-1405.

[8] Indu, V., & Thampi, S. M. (2024). Misinformation detection in social networks using emotion analysis and user behavior analysis. Pattern Recognition Letters, 182, 60-66.

[9] Liu, C., Tian, Q., & Zhu, X. (2023). Social media emergency information assessment of major emergencies: a case study of additional emotional characteristics. Library Hi Tech, 41(3), 939-968.

[10] Hu, N. (2022). Sentiment analysis of texts on public health emergencies based on social media data mining. Computational and mathematical methods in medicine, 2022(1), 3964473.

[11] Molina, A. U., & Subias, M. H. (2024). Personalization of Content in Video-on-Demand Services: Insights from Satisfaction over Social Media Algorithms. Revista ComHumanitas, 15(2), 175-187.

[12] Luo, B., & Luo, R. (2024). Application and Empirical Analysis of Fuzzy Neural Networks in Mining Social Media Users' Behavioral Characteristics and Formulating Accurate Online Marketing Strategies. International Journal of Computational Intelligence Systems, 17(1), 273.

[13] Al-Qurishi, M., Alhuzami, S., AlRubaian, M., Hossain, M. S., Alamri, A., & Rahman, M. A. (2018). User profiling for big social media data using standing ovation model. Multimedia Tools and Applications, 77, 11179-11201.

[14] Han, K. X., Chien, W., Chiu, C. C., & Cheng, Y. T. (2020). Application of support vector machine (SVM) in the sentiment analysis of twitter dataset. Applied Sciences, 10(3), 1125.

[15] Pramudito, D. K., Tubagus, M., Ismail, A., Anggraini, F. D., & Paramita, C. C. P. (2023). Application of Support Vector Machine and Naive Bayes Method to Analyze Mobile Banking User Behavior Through social media Twitter. Jurnal Informasi Dan Teknologi, 325-330.

[16] Sundararaj, V., & Rejeesh, M. R. (2021). A detailed behavioral analysis on consumer and customer changing behavior with respect to social networking sites. Journal of Retailing and Consumer Services, 58, 102190.

[17] Xue, D., Hirche, S., & Cao, M. (2019). Opinion behavior analysis in social networks under the influence of coopetitive media. IEEE Transactions on Network Science and Engineering, 7(3), 961-974.

[18] Soundarya, V., Kanimozhi, U., & Manjula, D. (2017). Recommendation System for Criminal Behavioral Analysis on Social Network using Genetic Weighted K-Means Clustering. J. Comput., 12(3), 212-220.

[19] Lu, M., Chen, R., Meng, X., & Yang, K. (2025). Computer model-based analysis of user behaviour and emotional disposition on social platforms. J. COMBIN. MATH. COMBIN. COMPUT, 127, 8613-8625.

[20] Li, J., Wang, Y., & Wang, J. (2021). An analysis of emotional tendency under the network public opinion: Deep learning. Informatica, 45(1).

[21] Shen, L., Yao, R., Zhang, W., Evans, R., Cao, G., & Zhang, Z. (2021). Emotional attitudes of Chinese citizens on social distancing during the COVID-19 outbreak: analysis of social media data. JMIR medical informatics, 9(3), e27079.

[22] Sun, H., Wang, G., & Xia, S. (2021). Text tendency analysis based on multi-granularity emotional chunks and integrated learning. Neural Computing and Applications, 33, 8119-8129.

[23] Nokkaew, M., Nongpong, K., Yeophantong, T., Ploykitikoon, P., Arjharn, W., Phonak, D., ... & Surawanitkun, C. (2024). Hidden emotional trends on social media regarding the Thailand–China high-speed railway project: A deep learning approach with ChatGPT integration. Social Network Analysis and Mining, 14(1), 175.

[24] Poonam Narang,Ajay Vikram Singh & Himanshu Monga. (2024). Enhanced detection of fabricated news through sentiment analysis and text feature extraction. International Journal of Information Technology,16(6),3891-3900.