ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

# Propagation Mechanism and user behavior analysis of short video content in mobile social platform

*Wei Zhu*[1,*], *Gui Wang*[2], *Xuan Hao*[1], *Munire Adili*[3] *and Weiqun Cui*[1]

[1] Department of Information Engineering, Shandong Water Conservancy Vocational College, Rizhao 276800, Shandong, China
[2] Office of Registrar, Zibo Polytechnic University, Zibo 255300, Shandong, China
[3] School of Digital Information Engineering, Kashgar Vocational and Technical College, Kashgar 844000, Xinjiang, China

**SUMMARY:** *Aiming at the problem of the separation analysis of short video propagation process and user behavior response in mobile social platforms, a comprehensive analysis framework combining content characteristics, platform mechanism, social structure and behavior feedback was constructed. Based on 52,000 short video samples and corresponding user behavior logs collected from January 2024 to December 2024, multi-modal content representation, propagation network analysis, clustering identification and predictive modeling methods are used to jointly test the propagation trigger mechanism, diffusion evolution mechanism and user behavior transformation path. The results show that topic focus, emotion expression intensity, information density and multi-modal presentation quality form the trigger basis of propagation, and the recommendation strength, hot traffic allocation and social relationship strength significantly affect the speed, scope and persistence of diffusion. Users show obvious heterogeneity in the dimensions of click, stay, comment, forward, follow and repeat visit. Low-active users are more dependent on the accuracy of recommendation, and high-active users are more likely to be stimulated by social interaction. Experimental results show that the Accuracy, F1 and AUC of the constructed model on the user behavior recognition task reach 0.903, 0.895 and 0.941, respectively, and the RMSE and* $R^2$ *on the propagation effect prediction task are 0.109 and 0.872, respectively. This study can provide quantitative basis for short video propagation mechanism explanation, user behavior modeling and platform content governance.*

***KEYWORDS:*** *Short video propagation; User behavior; Recommendation algorithm; Multimodal analysis*

# 1 Introduction

With the continuous development of mobile Internet, intelligent terminals and platform algorithms, short video has quickly become the core content form in mobile social platforms. Compared with images and long videos, short videos have the characteristics of short duration, low entry threshold, strong visual stimulation, fast interactive feedback and high social embedding. Short videos can achieve information transmission, emotional arousal and behavior guidance in a limited time, so they show stronger influence in content dissemination, social interaction, consumption decision-making and public opinion diffusion. Schellewald

*15006335067@163.com
https://doi.org/10.65102/is2026315
Published: 30 April, 2026

(2023) studies the popular supply and user experience logic of TikTok, and points out that the short video platform reconstructs the user participation mode through continuous viewing, lightweight creation and high-frequency interaction [3]. Bhandari and Bimo (2022) proposed that TikTok represents not only a content distribution platform, but also a communication environment with recommendation algorithm as the core, where user expression, viewing and interaction are deeply embedded in the algorithm mechanism [1]. Dekker et al. (2025) studied the impact of algorithmic personalized recommendation on social media engagement, and believed that differentiated distribution significantly changed the content reach efficiency and interaction performance [5]. This shows that short video propagation is not a single information output process, but a dynamic system with content features, platform recommendation, social relationships and user feedback.

Focusing on short video propagation and user behavior, the existing research has been discussed from multiple dimensions. Siles et al. (2024) studied the formation process of TikTok users' algorithmic consciousness and pointed out that users' cognition of the recommendation mechanism would affect its use and acceptance [2]. Obreja (2024) studies the user's narrative construction of the meaning of the algorithm, indicating that the algorithm understanding itself is also involved in the formation of the communication order [4]. Yang et al. (2023) studied the continuous usage behavior of short videos and proposed that flow experience and cognitive lock will jointly promote user retention [8]. Zhang et al. (2024) studied continuous viewing behavior and pointed out that content continuity, social connection and interface convenience will strengthen users' tendency to scroll consumption [19]. Lu et al. (2023) studied the influencing factors of short video user participation and believed that there was a composite effect between content attributes, interactive environment and platform conditions [11]. Cho et al. (2024) further used the multi-modal attention mechanism to predict the popularity of short videos, which provided a new technical path for the recognition of communication effects [20]. In addition, He et al. (2024) studied the driving effect of social learning on short video e-commerce participation [18], and Meng et al. (2024) analyzed the impact of TikTok short video advertisement content characteristics on purchase behavior [16]. These results provide an important foundation for understanding short video propagation and user behavior.

However, there are still some shortcomings in the existing research. On the one hand, many studies pay more attention to the dissemination effect, continuous use or single interaction behavior, and lack a systematic description of the continuous link between content generation, platform distribution, user contact, interaction diffusion and feedback reinforcement. On the other hand, existing studies mostly start from the outcome variables such as overall participation, excessive use or continuous viewing, and do not fully identify the differentiation mechanism of heterogeneous behaviors such as click, stay, like, comment, retweet, follow and revisit. More importantly, the coupling relationship between the dissemination mechanism and user behavior has not been deeply revealed. Current studies often discuss content dissemination and behavior response separately, which makes it difficult to fully explain why short videos diffuse, why users participate in short videos, and how feedback can reverse shape the dissemination.

Based on this, this paper focuses on the dissemination mechanism of short video content and user behavior characteristics on mobile social platform, constructs a comprehensive analysis framework combining content, platform, social and behavior, and investigates its role path on dissemination effect and behavior performance from four dimensions of content characteristics, recommendation mechanism, relationship structure and feedback behavior. The effectiveness of the related mechanism and model is verified by experiments. The innovation of this paper is to integrate the short video propagation mechanism and user behavior into the same analysis system to strengthen the dynamic relationship between them. In the variable design, content, platform, social and feedback factors were introduced to improve the ability to explain the complex communication environment. The mechanism analysis and experimental verification are combined in the research method to enhance the reliability and application value of the conclusion.

# 2 Research design

## 2.1 Data source and sample construction

IThis paper selects short video samples in mobile social platforms as the research object, and constructs a research dataset around the content transmission chain and user behavior feedback process. The samples mainly come from the public display page of the platform, topic aggregation page, user home page and interactive information page, and the collection objects cover typical short video content such as information, knowledge, life, entertainment and commercial promotion. In order to ensure the timeliness and comparability of the samples, the collection interval is set as January 2024 to December 2024, and the observation window is established with weekly as the basic time granularity, and the propagation performance within 24 hours, 72 hours and 7 days after the video is released is tracked and recorded. In the process of data collection, an automatic crawling and parsing process is built based on Python, and multi-source information aggregation is completed by combining page structure recognition, field mapping and timestamp alignment. The original data includes fields such as video title, text description, label information, release time, video duration, account attributes, number of likes, number of comments, number of retweet, number of favorites, number of plays and comment interaction records. The specific sources and field composition are shown in Table 1.

*Table 1: Data sources and core field composition*

| Data Module | Specific Source | Core Fields | Variable Purpose |
| --- | --- | --- | --- |
| Video Content Data | Video detail pages, topic aggregation pages | Title, caption, tags, duration, topic category, posting time | Characterizing content features and publication attributes |
| Dissemination Effect Data | Video statistics pages | View count, exposure volume, completion rate, dissemination duration, growth rate | Measuring dissemination breadth, depth, and speed |
| Interaction Behavior Data | Like/comment/share/favorite pages | Number of likes, comments, shares, favorites, comment reply chains | Capturing user participation intensity and interaction structure |
| Account Attribute Data | User profile pages, verification information pages | Number of followers, following count, account type, historical posting volume | Reflecting creator influence and account basis |
| Platform Environment Data | Trending pages, topic pages, time-window information | Trending tags, topic popularity, posting time slot, recommendation placement information | Representing platform distribution environment and external conditions |

In the sample construction stage, this paper conducts screening based on the principle of "complete content, traceable indicators, comparable time, and abnormal removal". The samples with repeated release, serious field missing, abnormal release time, obvious distortion of interactive data and prominent marketing irrigation traces were removed. The samples with non-short video form content and unable to form effective transmission chain records were not included in the analysis. Then, the text field is segmented, stop words are cleaned, and labels are standardized. The numerical variables are filled with missing values, outliers are censored, and standardized conversion is performed. The cross-table association is completed by combining the account identification and video identification. Among them, the video content information is mainly used to characterize the topic category, duration structure and expression characteristics, the communication effect index is used to characterize the exposure level and diffusion intensity, the interactive behavior index is used to identify the participation performance such as likes, comments, forwarding and favorites, and the platform environment variable is used to reflect the external conditions such as release time window, account influence and topic heat. Thus, it provides data basis for subsequent propagation mechanism analysis and user behavior modeling.

## 2.2 Index system and variable definition

In order to systematically identify the relationship between short video content dissemination mechanism and user behavior in mobile social platforms, this paper constructs four index systems based on data preprocessing, including content feature variables, dissemination mechanism variables, user behavior variables and outcome variables, and completes the definition of variables through field mapping, text coding, behavior measurement and time window alignment. The content feature variables are mainly used to describe the video ontology properties. The topic types are semantically classified into information, knowledge, life, entertainment and marketing categories according to the video title, topic tag and text description. The video duration was recorded in seconds and stratified by short duration, medium duration and medium duration. The emotion expression is obtained by emotion lexicon matching and emotion polarity calculation. The title tag was quantified by combining the explicit words, the number of popular tags and the density of keywords. Multimodal performance is jointly characterized by features such as cover clarity, caption density, background music usage, and screen switching frequency. The propagation mechanism variables mainly reflect the diffusion status of short videos in the platform, including exposure, interaction rate, diffusion depth and diffusion breadth. The exposure is measured by the play volume and recommendation reach volume, the interaction rate is calculated by the ratio of likes, comments, forwarding and collection to the exposure amount, and the diffusion depth is expressed by the forwarding chain level and the comment reply level. Diffusion breadth is measured by the coverage degree of different user groups and topic scenarios. User behavior variables are used to identify user participation paths, including click, stay, like, comment, forward, follow and revisit. The outcome variable is used to comprehensively characterize the communication performance, which is mainly measured by the communication growth rate, interactive conversion rate and continuous activity. At the same time, the platform environment variables were included into the model as control transmission conditions, including recommendation strength, release time and social relationship strength. The recommendation strength was constructed according to the frequency of recommendation bits and the entry of hot list, and the release time was classified by hour window, working days and rest days. The social relationship strength is calculated by combining the connection density of fans, the frequency of interaction and the degree of mutuality of accounts.

## 2.3 Research methods

In this paper, statistical analysis, relational modeling and predictive modeling are combined to identify the short video content propagation mechanism and user behavior characteristics at multiple levels. Descriptive statistics are used to describe the distribution characteristics of the sample in the dimensions of topic type, video duration, emotional expression, interaction scale and communication performance. The basic statistical properties and correlation structure of the variables are tested by the skewness, kurtosis and correlation coefficient matrix. Regression analysis was used to identify the influence direction and effect strength of content feature variables, platform environment variables and user behavior variables on dissemination results. A multiple linear regression model was constructed for continuous outcome variables, and logarithmic transformation was performed on indicators with significant skewed distribution. For binary or stratified outcome variables, Logit or ordered Logit models are introduced to improve estimation fit.

In the part of user behavior recognition, the clustering analysis method is used, and the input variables are the behavior indicators such as click, stay, like, comment, forward, follow and repeat visit. The K-means algorithm is used to complete user clustering, and the silhouette coefficient and the within-class sum of squares are combined to determine the optimal number of clusters, so as to extract the differentiated behavior patterns of high interaction, shallow browsing and continuous participation. The diffusion structure analysis part constructs the "user-content-interaction relationship" diffusion network, and identifies the core propagation nodes, diffusion levels and group aggregation characteristics based on the node degree centrality, proximity centrality, betweenness centrality, network density, average path length and community modularity. In the mechanism testing part, the structural equation model is introduced to estimate the direct effect, indirect effect and mediation path between content characteristics, dissemination mechanism and user behavior, and the stability of the model is verified by the goodness of fit index. In the prediction analysis part, random forest and XGBoost models are used to classify and regression predict the communication performance and user participation level, so as to compare the performance differences of different methods in complex nonlinear relationship recognition.

# 3 Analysis of Short Video Content Dissemination Mechanism

## 3.1 Multimodal Content Representation and Recommendation Triggering Mechanism

The starting point of short video dissemination in mobile social platforms does not depend on whether the content is passively released, but on whether the platform can effectively represent the video and trigger the recommendation distribution in the candidate recall and refinement stage. Short video contains title text, topic tag, key frame sequence, caption layer and audio signal at the same time, so its content representation is a typical heterogeneous data fusion problem. To this end, we introduce natural language processing, computer vision and sequence modeling methods to encode text, visual and audio information in a unified way. Let the original input of the ith short video be Xi={xitext,xivision,xiaudio}, where xitext includes title, copywriting and label text,xivision includes key frames, subtitle area and main target, and xiaudio includes speech rhythm and background audio track. The text mode uses BERT to extract semantic embedding hitext, the visual mode uses ResNet and temporal convolutional network to extract the visual representation hivision, and the audio mode is input into BiGRU after Mel spectrum transformation to obtain the audio representation hiaudio. Three types of features form a unified content vector through the attention fusion mechanism:

$$z_i=\alpha_t h_i^{text}+\alpha_v h_i^{vision}+\alpha_a h_i^{audio}, \quad \alpha_t+\alpha_v+\alpha_a=1 \tag{1}$$

Among them, the modal attention weights $\alpha_t$, $\alpha_v$, and $\alpha_a$ are used to characterize the contribution strength of different modes to propagation discrimination.

Based on the multi-modal representation, the platform further calculates the topic saliency, emotion intensity and information density of the video to improve the recognition ability of the recommendation model for content quality. The topic saliency was output by the text classifier, denoted as Ti. Emotional intensity is jointly calculated by emotional polarity network and emotional fluctuation amplitude, which is denoted as:

$$E_i=\beta_1 p_i^{pos}-\beta_2 p_i^{neg}+\beta_3 \sigma_i^{emo} \tag{2}$$

Here, $p_i^{pos}$ and$p_i^{neg}$ represent positive and negative sentiment probabilities, respectively, and $\sigma_i^{emo}$ represents sentiment fluctuation variance. The information density is composed of the number of effective semantic units per unit time, caption density and shot switching frequency, and is defined as follows.

$$D_i=\gamma_1 \frac{n_i^{sem}}{l_i}+\gamma_2 d_i^{sub}+\gamma_3 f_i^{shot} \tag{3}$$

Here, $n_i^{sem}$ is the number of semantic units, li is the video duration, $d_i^{sub}$ is the caption density, and $f_i^{shot}$ is the shot switching frequency. In this way, short video content is no longer coarse-grained indexed by a single tag or title, but forms a high-dimensional vector representation containing semantic, visual and audio cues with the support of computer technology.

After the multimodal content vector enters the recommendation system, the platform calculates the recommendation trigger score through the candidate recall and refinement model.

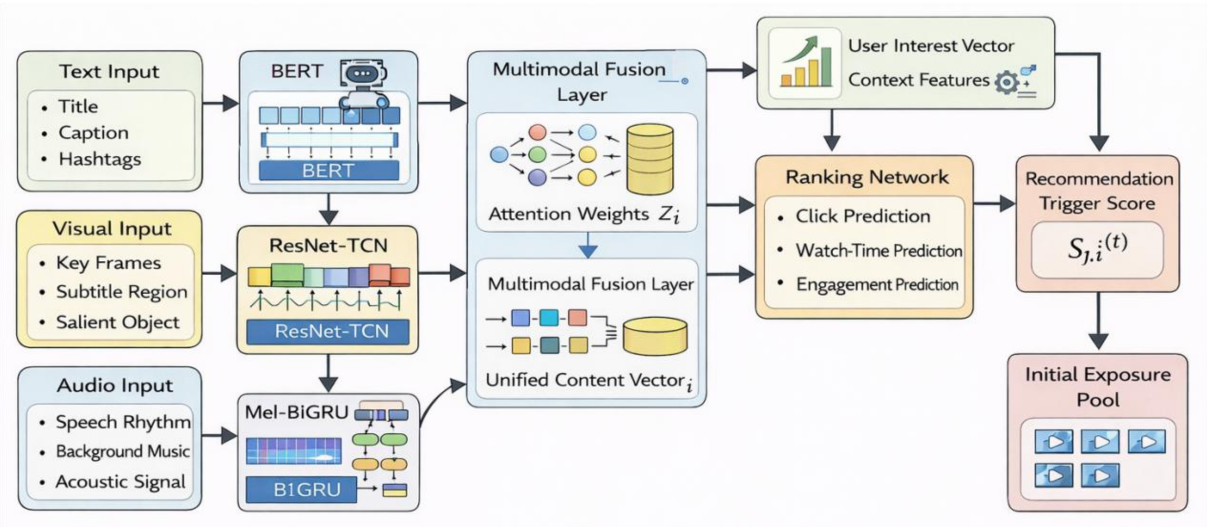


*Figure 1: Framework diagram of short video multimodal content representation and recommendation trigger mechanism computation*

As shown in Figure 1, short video propagation first goes through three Feature Extraction modules: "Text Encoding", "Visual Feature Extraction" and "Audio Rhythm Encoding". In the "Multimodal Fusion Layer", a unified content representation zi is generated, and then input into the ranking network together with the user interest vector uj and the context feature ct to output the recommendation score. For user j, the trigger score of video i at time t can be expressed as follows.

$$S_{j,i}^{(t)}=\lambda_1 \mathrm{sim}(u_j,z_i)+\lambda_2\hat{p}_{j,i}^{click}+\lambda_3\hat{p}_{j,i}^{watch}+\lambda_4\hat{p}_{j,i}^{eng}+\lambda_5 c_t \tag{4}$$

Where sim(uj,zi) represents the interest matching degree, $\hat{p}_{j,i}^{click}$, $\hat{p}_{j,i}^{watch}$, $\hat{p}_{j,i}^{eng}$ represents the prediction probability of click, stay and interaction respectively, and ct represents the context factors such as time window, device environment and heat state. When $S_{j,i}^{(t)}$ exceeds the recommendation threshold τ, the video is assigned to the corresponding traffic pool, and its trigger function can be written as follows.

$$R_{j,i}^{(t)}=\begin{cases}1, & S_{j,i}^{(t)}\geqslant \tau\\ 0, & S_{j,i}^{(t)}<\tau\end{cases} \tag{5}$$

This means that the initial exposure of short video propagation does not occur uniformly, but is determined by multimodal content quality, user interest matching and behavior prediction results.

From the mechanism point of view, the content with high topic focus, moderate emotion expression, reasonable information density and prominent visual subject is easier to obtain higher click prediction value and stay prediction value in the ranking model, thereby improving the recommendation trigger probability. On the contrary, videos with semantic dispersion, redundant subtitles, excessive shot perturbation or unbalanced audio rhythm are difficult to obtain stable exposure in the fine-tuning stage even if they enter the candidate set. It can be seen that the first layer mechanism of short video dissemination is a linkage process of "multimodal content representation - recommendation trigger calculation - initial exposure allocation". The platform does not simply amplify the existing popular content, but relies on computer technology to evaluate the quality of content representation and user matching results in real time, and determine the strength of the dissemination starting point.

## 3.2 Network diffusion evolution and Feedback enhancement Mechanism

After the initial exposure of a short video, its subsequent propagation is not a linear extension, but a dynamic evolution process formed by the joint action of recommendation system, hot spot detection and social network diffusion. The propagation state can be expressed as follows.

$$r_i^{(t)} = [e_i^{(t)}, b_i^{(t)}, d_i^{(t)}, q_i^{(t)}] \tag{6}$$

Here, $e_i^{(t)}$ is the exposure, $b_i^{(t)}$ is the propagation breadth, $d_i^{(t)}$ is the diffusion depth, and $q_i^{(t)}$ is the ranking score. If the video achieves high click rate, completion rate and comment forwarding rate in the early time window, its propagation state will be rapidly enlarged, and the exposure update process can be written as follows.

$$e_i^{(t+1)} = e_i^{(t)} + \eta \cdot \sigma(q_i^{(t)}) \cdot \left(1 + \alpha_1 r_i^{click} + \alpha_2 r_i^{watch} + \alpha_3 r_i^{eng}\right) \tag{7}$$

where σ(· ) is the activation function, riclick, riwatch and rieng represent click feedback, viewing feedback and interactive feedback respectively, and η is the traffic amplification factor. This formula shows that the propagation speed and propagation persistence depend on the results of the platform recalculation of real-time feedback.

At the level of diffusion path, the following, interaction and forwarding relationships between users can form a directed graph G=(V,E). If the relationship strength between user u and user v is wuv, the probability of short video diffusion along the social edge can be expressed as follows.

$$p_{uv}^{(i)} = 1 - \exp(-\beta w_{uv} a_u^{(i)}) \tag{8}$$

where $a_u^{(i)}$ denotes whether user u is activated by the video and β is the diffusion sensitivity coefficient. High centrality nodes and bridging nodes are more likely to extend content from local interest groups to heterogeneous user groups, so social networks determine the depth of diffusion and the ability to spread across circles. At the same time, the platform will also build a hot spot detection model based on the play growth rate, comment growth rate and forwarding growth rate, and push the content with higher growth slope into the larger scale recommendation network. Let the hot spot weight be $H_i^{(t)}$, then it can be defined as follows.

$$H_i^{(t)} = \mu_1 \frac{\Delta P_i^{(t)}}{\Delta t} + \mu_2 \frac{\Delta C_i^{(t)}}{\Delta t} + \mu_3 \frac{\Delta S_i^{(t)}}{\Delta t} \tag{9}$$

where $a_u^{(i)}$ denotes whether user u is activated by the video and β is the diffusion sensitivity coefficient. High centrality nodes and bridging nodes are more likely to extend content from local interest groups to heterogeneous user groups, so social networks determine the depth of diffusion and the ability to spread across circles. At the same time, the platform will also build a hot spot detection model based on the play growth rate, comment growth rate and forwarding growth rate, and push the content with higher growth slope into the larger scale recommendation network. Let the hot spot weight be Hi(t), then it can be defined as follows.

$$G_i^{(t)} = \rho_1 l_i^{(t)} + \rho_2 c_i^{(t)} + \rho_3 s_i^{(t)} + \rho_4 f_i^{(t)} \tag{10}$$

where $l_i^{(t)}$, $c_i^{(t)}$, $s_i^{(t)}$ and $f_i^{(t)}$ are the increments of likes, comments, retweets and followings respectively, then the ranking score is updated as follows.

$$q_i^{(t+1)} = q_i^{(t)} + \gamma G_i^{(t)} \tag{11}$$

Comments and retweets generally have a higher value of information diffusion and thus a stronger gain in feedback learning. Therefore, short video propagation forms a closed-loop mechanism of "content representation - recommendation trigger - network diffusion - interactive feedback - redistribution enhancement", whose internal logic is not automatic diffusion of high-quality content, but a dynamic propagation process driven by multimodal coding, ranking calculation, relationship propagation and feedback reinforcement.

# 4 User behavior analysis

## 4.1 User behavior feature recognition

Short video user behavior on mobile social platforms has obvious hierarchical differences and heterogeneity. Different users have different performance in the behavioral dimensions of viewing, liking, commenting, forwarding, following and re-visiting. This difference not only reflects the difference of user interest preferences and participation depth, but also reflects the joint effect of platform recommendation mechanism, content attraction and social stimulation intensity. In order to realize the quantitative identification of user behavior characteristics, this paper constructs the user behavior vector based on the platform behavior log:

$$b_u = [w_u, l_u, c_u, s_u, f_u, r_u] \tag{12}$$

Among them, $w_u$ represents the effective viewing ratio, $l_u$ represents the frequency of likes, $c_u$ represents the frequency of comments, $s_u$ represents the frequency of forwarding, $f_u$ represents the conversion rate of attention, and $r_u$ represents the probability of re-visiting. Each dimensional indicator is normalized by the user's behavior record within the observation window, for example, the effective viewing ratio can be expressed as follows.

$$w_u = \frac{T_u^{watch}}{T_u^{expo}} \tag{13}$$

where $T_u^{watch}$ is the number of videos that have completed valid viewing, and $T_u^{expo}$ is the number of exposures received by the user. The return visit probability is defined as follows.

$$r_u = \frac{N_u^{return}}{N_u^{session}} \tag{14}$$

where $N_u^{return}$ is the number of returns within the set time window and $N_u^{session}$ is the total number of access sessions.

In the process of behavior recognition, the original behavior log is firstly extracted, missing correction and standardized transformation, and then the standardized user behavior vector is input into the clustering model and the behavior path recognition model to form the calculation process from the original behavior record to the classification of user types, as shown in Figure 2.

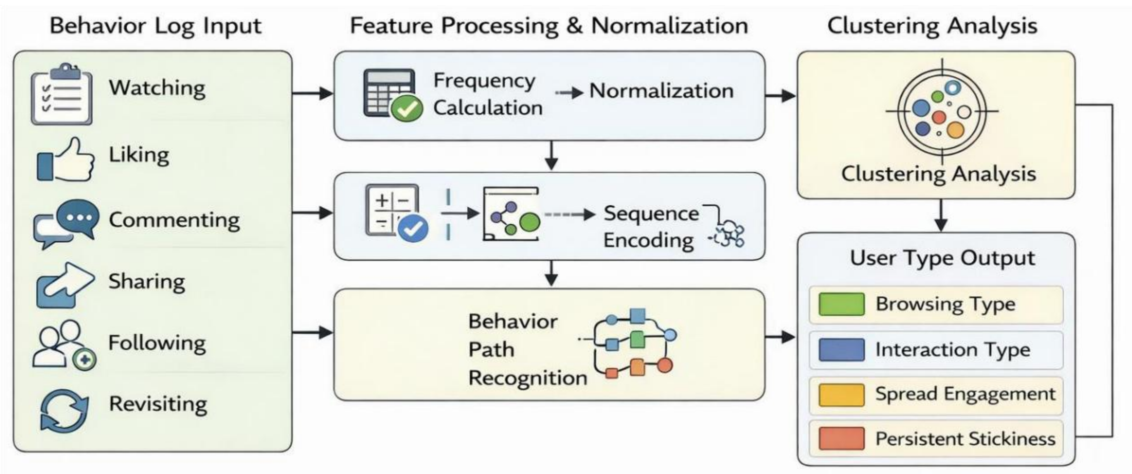


*Figure 2: Computational analysis framework diagram for user behavior feature recognition in short videos*

In Figure 2, the behavior log input layer corresponds to the original records such as viewing, likes, comments, forwarding, following and re-visiting. The feature processing layer completes the behavior frequency calculation, normalization and sequence coding, the cluster analysis layer is used to identify different user groups, and the behavior path identification layer further extracts typical operation sequences. Thus, the comprehensive characterization of user participation characteristics, preference patterns and activity levels can be achieved.

In order to reduce the influence of the difference of different index dimensions on the recognition results, this paper uses the standardized behavior characteristics for K-means clustering analysis, and its objective function is as follows.

$$J = \sum_{k=1}^{K} \sum_{u \in C_k} \| b_u - \mu_k \|^2 \tag{15}$$

Here, $C_k$ denotes the set of KTH class of users and $\mu_k$ denotes the center vector of this class. After determining the optimal number of clusters according to the within-class sum of squares and the contour coefficient, the user groups can be identified as shallow browsing type, interactive response type, high communication participation type and continuous viscosity type. Shallow browsing users usually have higher viewing ratio but lower comments, retweet and follow behaviors, which is manifested as fast consumption and lack of deep participation. Interactive responsive users are more active in the dimensions of likes and comments, and are easily stimulated by emotional expression and controversial content. The users with high communication participation have high forwarding rate and attention conversion rate, and assume the function of content diffusion nodes in social networks. On the other hand, users with persistent stickability maintain a higher level of viewing ratio and repeat visit probability, and have a stronger tendency to retain on the platform.

In order to further quantify the intensity of user participation, this paper constructs a behavior activity index:

$$A_u = \alpha_1 w_u + \alpha_2 l_u + \alpha_3 c_u + \alpha_4 s_u + \alpha_5 f_u + \alpha_6 r_u \tag{16}$$

Here, $\alpha i$ is the weight parameter of each behavioral dimension. Combined with the behavior sequence coding method, typical behavior paths such as "view-like", "view-comment-retweet" and "view-follow-revisit" can also be identified, which indicates that short video user behavior is not independent discrete actions, but a dynamic process with obvious temporal transfer relationship. It can be seen that user behavior feature recognition is not only a statistical summary of single indicators, but also a joint modeling process of behavior intensity, behavior structure and behavior path.

## 4.2 Analysis of influencing factors of user behavior

Users' click, stay and deep participation behaviors on short video platforms are essentially the results of content representation, recommendation matching and social feedback. In order to characterize the influence of different factors on behavioral paths, this paper incorporates content quality, emotional expression, social interaction, recommendation accuracy and information density into a unified analysis framework, and takes click probability, dwell time and depth participation intensity as the core outcome variables. Therein, content quality is characterized by a multimodal quality scoring function, denoted as:

$$Q_i = \alpha_1 q_i^{text} + \alpha_2 q_i^{vision} + \alpha_3 q_i^{audio} \tag{17}$$

Among them, $q_i^{text}$ represents the semantic integrity and theme consistency of the text, $q_i^{vision}$ represents the clarity of the picture, the saliency of the subject and the readability of the caption, and $q_i^{audio}$ represents the clarity of speech and the adaptation of background sound. The emotional expression intensity was denoted as Ei, which was jointly calculated by the polarity probability output by the sentiment classification model and the emotional fluctuation amplitude. The intensity of social interaction is denoted as Si, which is composed of the length of comment reply chain, interaction frequency and relationship connection density. The recommendation accuracy, denoted as Ru,i, is used to measure the matching degree between video i and user u interest vector. The information density, denoted as Di, is comprehensively represented by the number of effective semantic units per unit time, the amount of caption information, and the shot change frequency.

Click behavior can be regarded as the user's immediate response to the recommendation results, and its probability can be expressed by the Logit model as follows.

$$P(\mathrm{click}_{u,i} = 1) = \frac{1}{1 + \exp\left[-(\beta_0 + \beta_1 Q_i + \beta_2 E_i + \beta_3 S_i + \beta_4 R_{u,i} + \beta_5 D_i)\right]} \tag{18}$$

Estimation results show that recommendation accuracy and content quality have the most significant marginal contribution to click behavior. High-matching recommendation can reduce user decision-making cost, and high-quality content can improve the ability of capturing user attention for cover, title and preview information. The stay behavior is more affected by information absorption efficiency and emotional traction strength. Suppose that the stay duration of user to video i is Tu,i, then:

$$T_{u,i} = \gamma_0 + \gamma_1 Q_i + \gamma_2 E_i + \gamma_3 D_i + \gamma_4 R_{u,i} + \varepsilon_{u,i} \tag{19}$$

Among them, information density and emotion expression show obvious interval effect. When the information density is too low, it is difficult for users to form sustained attention. When the information density is too high, the cognitive load increases, but the length of stay decreases. The same is true for emotional expression. Moderate emotional fluctuations are more likely to prolong viewing time, while extreme emotional stimuli may cause short-term clicks, but are not conducive to stable retention.

Deep engagement behaviors include comments, retweet, follow and revisit, which are more dependent on the coupling mechanism between social interaction and content identification. In this paper, deep participation intensity is defined as follows.

$$G_{u,i} = \delta_1 c_{u,i} + \delta_2 s_{u,i} + \delta_3 f_{u,i} + \delta_4 r_{u,i} \tag{20}$$

Here, cu,i, su,i, fu,i, and ru,i represent comment, forwarding, follow, and revisit behaviors, respectively. Further build the path model:

$$G_{u,i} = \theta_0 + \theta_1 Q_i + \theta_2 E_i + \theta_3 S_i + \theta_4 R_{u,i} + \theta_5 D_i + \theta_6 (E_i \times S_i) + \epsilon_{u,i} \tag{21}$$

The interaction term $E_i \times S_i$ reflects the coupling effect of emotional expression and social interaction. The results show that the intensity of social interaction has the most direct impact on deep participation behavior, and the visibility of comments, the number of interactions with others and the activity of reply chain can significantly improve users 'expression and forwarding intention. Content quality and recommendation accuracy indirectly promoted attention and repeat visits by improving cognitive identity and matching efficiency.

### 4.3 Coupling relationship between propagation mechanism and user behavior

The diffusion of short videos on mobile social platforms is not a serial process with independent transmission mechanism and user behavior, but a closed-loop coupling system composed of recommendation distribution, social diffusion and interactive feedback. The platform determines the exposure intensity and reach range of content in different time Windows through the ranking model, hot spot detection model and relationship network propagation model. The user behavior signals formed in the links of click, stay, comment, forward, follow and revisit are written into the recommendation feature space in real time, which is an important basis for subsequent ranking update and traffic redistribution. To characterize this coupling process, let the propagation state of the ith short video at time t be as follows.

$$z_i^{(t)} = [e_i^{(t)}, q_i^{(t)}, b_i^{(t)}, d_i^{(t)}] \tag{22}$$

Here, $e_i^{(t)}$ represents the exposure, $q_i^{(t)}$ represents the ranking score, $b_i^{(t)}$ represents the propagation breadth, and $d_i^{(t)}$ represents the diffusion depth. The behavioral transition probability of user u to video i is defined as follows.

$$P(y_{u,i}^{(t)} = 1) = \sigma\left(\beta_0 + \beta_1 q_i^{(t)} + \beta_2 m_i + \beta_3 s_{u,i}^{(t)} + \beta_4 a_u\right) \tag{23}$$

where $m_i$ is the content representation obtained by the fusion of text semantic coding, visual feature extraction and emotion recognition network, $s_{u,i}^{(t)}$ is the intensity of social stimulus, au is the feature of user activity, and $\sigma(\cdot)$ is the Sigmoid function. This equation shows that the propagation mechanism affects the conversion probability of users from exposure to click and from viewing to deep participation through ranking score, content matching and relationship stimulation.

In the feedback loop stage, the platform re-encodes high-value behaviors such as likes, comments, retweets and follows into supervised signals, and inputs the multi-task learning ranking model to update the distribution status at the next moment. If the integrated feedback gain is defined as follows.

$$G_i^{(t)} = \lambda_1 l_i^{(t)} + \lambda_2 c_i^{(t)} + \lambda_3 sh_i^{(t)} + \lambda_4 f_i^{(t)} \tag{24}$$

where $l_i^{(t)}$, $c_i^{(t)}$, $sh_i^{(t)}$ and $f_i^{(t)}$ represent the increment of likes, comments, retweets and follows respectively, then the ranking score update can be expressed as follows.

$$q_i^{(t+1)} = q_i^{(t)} + \eta G_i^{(t)} \tag{25}$$

Furthermore, the exposure is iterated with the ranking update:

$$e_i^{(t+1)} = e_i^{(t)} + \mu \cdot q_i^{(t+1)} \tag{26}$$

Here, $\eta$ is the feedback weight and $\mu$ is the traffic amplification factor. Comments and retweets are usually given higher weights in the model due to their stronger information diffusion value, so they are more likely to trigger new recommendation gains and cross-community propagation.

The whole coupling structure can be implemented in the time series recommendation model, graph neural network and feedback learning framework, as shown in Figure 3.

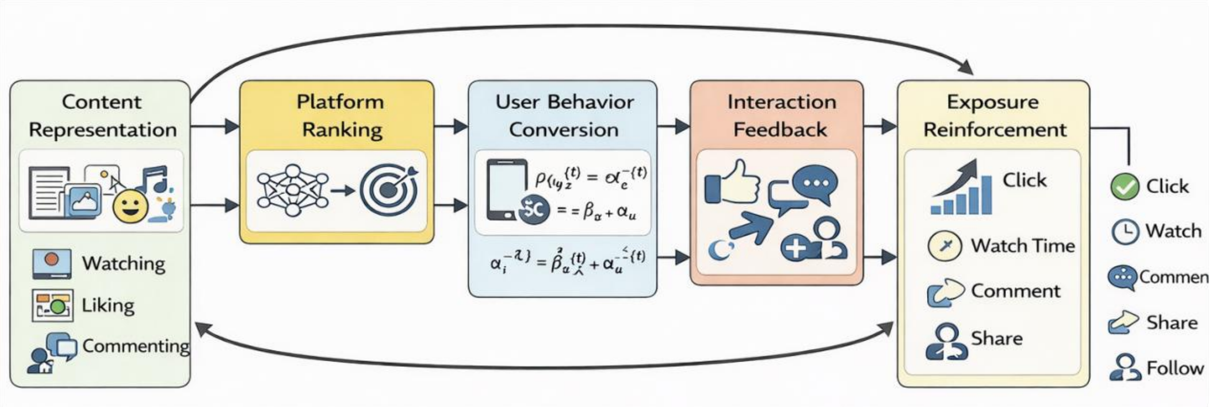


*Figure 3: Computational closed-loop of the coupling relationship between short video propagation mechanism and user behavior*

It can be seen that the propagation in the short video platform is not a single diffusion, but a dynamic evolution process of continuous iteration of computation distribution and behavior response.

# 5 Experimental analysis

## 5.1 Experimental environment and parameter setting

In order to verify the effectiveness of the short video content dissemination mechanism and user behavior analysis framework, experiments are carried out on the basis of the previous sample construction. The experimental data are consistent with Section 2.1, and the samples are from the public pages, topic aggregation pages, user homepages and interactive information pages of mobile social platforms. The observation interval is from January 2024 to December 2024, and the research objects cover short videos of information, knowledge, life, entertainment and commercial promotion. After eliminating duplicate samples, processing missing fields, filtering outliers and cross-table association, a structured sample library containing video content information, communication effect indicators, interactive behavior indicators and platform environment variables was finally formed. In order to avoid information leakage caused by time window crossing, the experiment used the method of dividing the data set according to the release time order, and divided the samples into training set, validation set and test set, and the proportion was set to 70%, 15% and 15%.

The experimental platform uses Python 3.10 for data processing and model training, the statistical analysis and machine learning part is implemented based on Scikit-learn, and the multimodal representation and recommendation calculation module is built based on PyTorch. The text representation module uses BERT-base to semantically encode the title, copy and label, the visual representation module uses ResNet50 to extract key frame features, and the audio representation module uses BiGRU to model the rhythm sequence. AdamW optimizer was used in the model training phase, the initial learning rate was set to $2\times10^{-4}$, the batch size was 64, the training rounds were 50, and the Dropout was set to 0.3. The early stopping mechanism is triggered when there is no significant improvement in the validation set for 5 consecutive rounds, and the cosine annealing strategy is used to update the learning rate to improve the stability of the training process. The experimental environment and key parameters are shown in Table 2.

*Table 2: Experimental environment and key parameter Settings*

| Item | Parameter/Description |
|---|---|
| Data Source | Short-video samples collected from January 2024 to December 2024 |
| Sample Type | Video content, dissemination indicators, interaction behaviors, platform variables |
| Data Split | 70% training set, 15% validation set, 15% test set |
| Experimental Tools | Python 3.10, Scikit-learn, PyTorch |
| Text Representation | BERT-base |
| Visual Representation | ResNet50 |
| Audio Representation | BiGRU |
| Optimizer | AdamW, learning rate ($2\times10^{-4}$) |
| Batch Size | 64 |
| Training Epochs | 50 |
| Dropout | 0.3 |
| Early Stopping | Triggered when there is no improvement on the validation set for 5 consecutive epochs |

Considering that this paper involves three types of tasks, including propagation effect prediction, user behavior recognition and recommendation quality evaluation, the evaluation indicators are set separately according to the task attributes. The classification task uses Accuracy, Precision, Recall and F1-score to measure the recognition effect of user behavior. In the regression task, MAE, RMSE and $R^2$ were used to evaluate the prediction accuracy of playout, stay time and propagation range. The ranking task uses NDCG@10 and HitRate@10 to measure the ranking quality of the recommendation trigger mechanism. The relevant evaluation indicators are shown in Table III. The above Settings are consistent with the data sources and variable system mentioned above, and also provide a unified benchmark for subsequent model effect verification, comparative experiments and robustness analysis.

*Table 3: Evaluation index Settings*

| Task Type | Evaluation Metric | Metric Description |
|---|---|---|
| Classification Task | Accuracy | Overall classification accuracy |
| Classification Task | Precision | Precision of positive-class prediction |
| Classification Task | Recall | Recall of positive-class identification |
| Classification Task | F1-score | Harmonic mean of Precision and Recall |
| Regression Task | MAE | Mean Absolute Error |
| Regression Task | RMSE | Root Mean Squared Error |
| Regression Task | $R^2$ | Goodness of fit |
| Ranking Task | NDCG@10 | Ranking gain quality of the top 10 results |
| Ranking Task | HitRate@10 | Hit rate of the top 10 results |

## 5.2 Model effect verification and result analysis

In order to verify the effectiveness of the proposed analytical framework for short video propagation mechanism, this paper conducts comparative experiments on the propagation effect prediction task and user behavior recognition task. The propagation effect prediction task takes the growth rate of play volume, interaction rate and diffusion persistence as the target variables, and the user behavior recognition task takes the category determination of click, stay and deep participation behavior as the target output. In terms of model Settings, Logistic Regression, Random Forest, XGBoost, BiGRU and the multimodal coupling model proposed in this paper are selected for comparison. In this model, text semantic coding, visual feature extraction, audio rhythm representation, platform ranking variables and social diffusion features are jointly introduced to realize the collaborative modeling of short video propagation process and user behavior path. Experimental results show that the traditional statistical model has certain stability in low-dimensional linear relationship recognition, but its performance is limited when dealing with heterogeneous features, nonlinear coupling relationships and timing dependencies in short video propagation scenarios. The tree model can improve the ability of local feature fitting, but the description of cross-modal information collaboration and dynamic propagation state changes is still insufficient. The sequence model performs better than the static model in the behavior recognition task, but its ability to recognize deeply engaged behaviors is still constrained if the joint input of platform mechanism variables and social communication variables is missing. The comprehensive comparison results are shown in Table 4. The Accuracy, Recall, F1 and AUC of the proposed model on the user behavior recognition task reach 0.903, 0.887, 0.895 and 0.941, respectively. The MAE, RMSE and $R^2$ of the propagation effect prediction task are 0.071, 0.109 and 0.872, respectively, which are better than those of the comparison models.

*Table 4: Performance comparison of different models in propagation effect prediction and user behavior recognition tasks*

| Model | Accuracy | Recall | F1 | AUC | MAE | RMSE | $R^2$ |
|---|---|---|---|---|---|---|---|
| Logistic Regression | 0.781 | 0.754 | 0.766 | 0.823 | 0.126 | 0.184 | 0.691 |
| Random Forest | 0.834 | 0.812 | 0.821 | 0.871 | 0.103 | 0.151 | 0.758 |
| XGBoost | 0.861 | 0.843 | 0.850 | 0.902 | 0.089 | 0.132 | 0.814 |
| BiGRU | 0.879 | 0.861 | 0.868 | 0.918 | 0.081 | 0.121 | 0.843 |
| Proposed Model | 0.903 | 0.887 | 0.895 | 0.941 | 0.071 | 0.109 | 0.872 |

From the perspective of classification indicators, the improvement of Recall and F1 indicators of the proposed model is more obvious, indicating that the constructed framework not only improves the discrimination accuracy of mainstream behavior categories, but also enhances the recognition ability of high-value behaviors such as comments, retweets and follows. This result shows that the joint introduction of multimodal content features and platform distribution variables can effectively improve the quality of modeling user behavior transformation paths. From the perspective of regression indicators, the proposed model achieves the best results in RMSE and $R^2$, indicating that it has lower fitting errors for the propagation growth rate, interactive conversion rate and diffusion persistence, and has stronger explanatory power for the overall fluctuation trend. The model performance comparison is shown in Figure 4, which can intuitively see the advantages of the proposed model in the user behavior recognition task. The convergence of the training process is shown in Figure 5. It can be observed that the proposed model tends to be stable after about the 32nd epoch, and there is no obvious oscillation in the verification loss, indicating that the training process has good convergence and generalization ability.

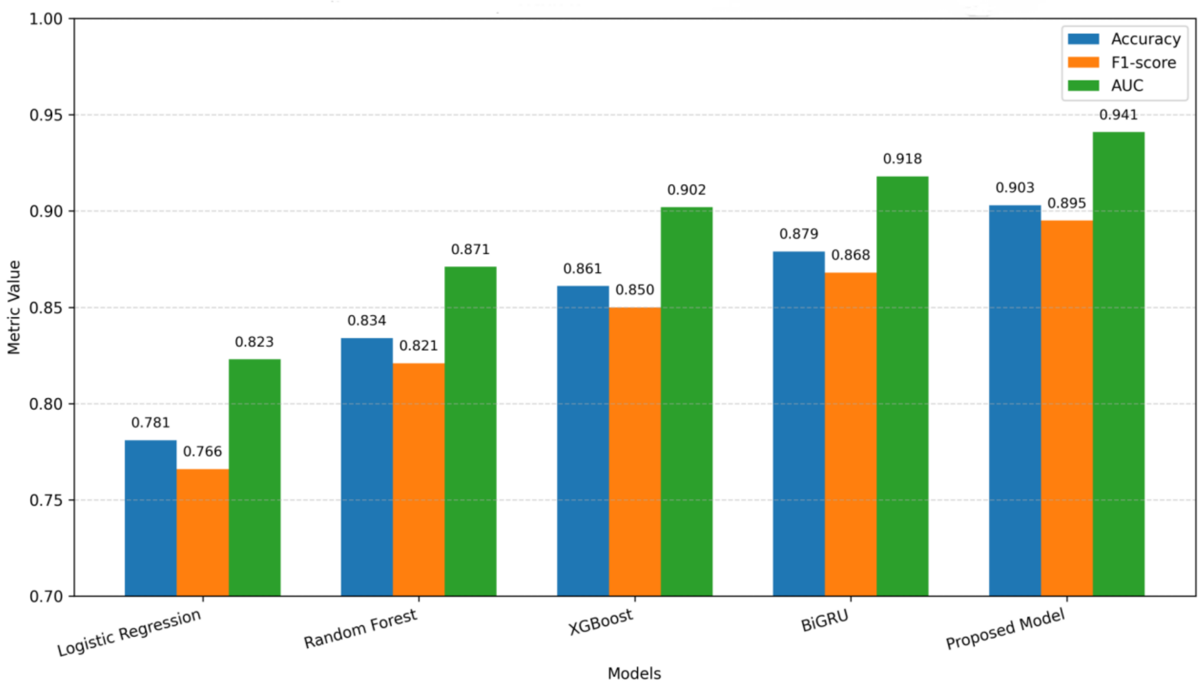


*Figure 4: Performance comparison of different models on the task of user behavior recognition*

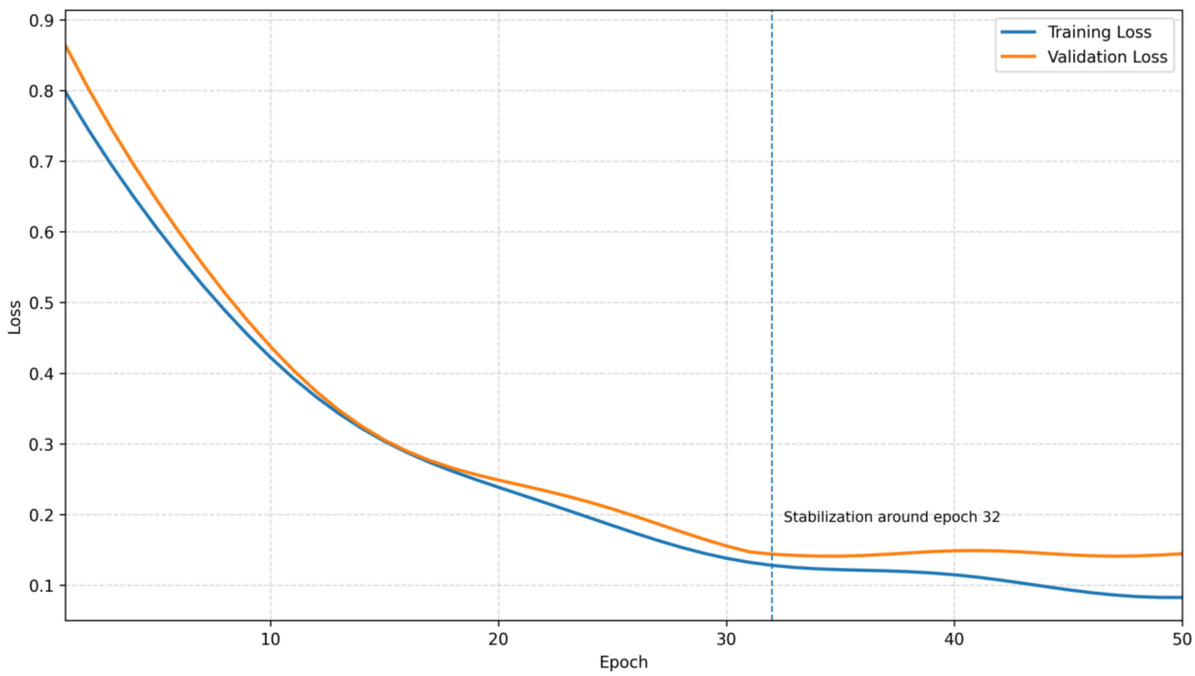


*Figure 5: Convergence of the proposed model training and validation loss*

Overall, the experimental results verify the applicability of the proposed method in two types of tasks: communication effect prediction and user behavior recognition, and show that incorporating content features, platform mechanisms and social communication information into the unified computing framework can more effectively reveal the complex relationship between short video propagation and behavior feedback in mobile social platforms.

## 5.3 Comparative experiments and ablation analysis

On the basis of the model effect verification in Section 5.2, this paper further tested the effectiveness and stability of the constructed method through comparative experiments and ablation experiments. The purpose of the comparison experiment is to identify the performance differences of the multi-factor fusion model compared with traditional statistical methods, single feature models and common machine learning models. The purpose of ablation experiments is to analyze the specific contributions of content features, platform variables and social variables in the prediction of dissemination effects and user behavior recognition. The experiment is still carried out based on the same training set, validation set and test set, and the evaluation indicators are still Accuracy, F1, AUC, RMSE and $R^2$.

In the comparison experiment, this paper sets up five groups of models: Linear Regression/Logistic Regression was used as the traditional statistical baseline, Text-only was used as the single Text feature model, XGBoost was used as the canonical tree model, and BiGRU was used as the time series modeling model. The Proposed Model is the multi-factor fusion model proposed in this paper. The results show that the traditional statistical model can capture a small amount of linear relationships, but lacks the ability to describe the heterogeneous feature coupling, temporal feedback and cross-modal interaction in short video propagation. Single text feature model has certain advantages in topic identification, but it cannot make full use of visual expression, platform recommendation and social diffusion information, so its performance is limited in user deep participation recognition and communication trend fitting. The tree model has strong adaptability to local nonlinear relationships, but the modeling ability of temporal behavior dependence and propagation path evolution is still insufficient. Although BiGRU can depict the characteristics of behavior sequence, its explanatory power for the identification of diffusion persistence and high-value behavior is still lower than that of the multi-factor fusion model without the input of platform variables and social variables. The specific results are shown in Table 5. The proposed model achieves the best results in F1, AUC and $R^2$ indicators, where F1 reaches 0.895, which is 0.089 higher than Text-only and 0.045 higher than XGBoost. AUC reached 0.941, which was 0.023 higher than that of BiGRU. The RMSE of the propagation effect prediction is reduced to 0.109, indicating that the joint modeling of multimodal content representation and platform propagation variables effectively improves the fitting accuracy.

*Table 5: Comparative experimental results of different methods*

| Model | Accuracy | F1 | AUC | RMSE | $R^2$ |
|---|---|---|---|---|---|
| Linear/Logistic Regression | 0.781 | 0.766 | 0.823 | 0.184 | 0.691 |
| Text-only | 0.824 | 0.806 | 0.861 | 0.162 | 0.734 |
| XGBoost | 0.861 | 0.850 | 0.902 | 0.132 | 0.814 |
| BiGRU | 0.879 | 0.868 | 0.918 | 0.121 | 0.843 |
| Proposed Model | 0.903 | 0.895 | 0.941 | 0.109 | 0.872 |

In order to further analyze the marginal contribution of different variable modules, this paper successively removed the content feature module, platform variable module and social variable module based on the full model, and added the "w/o Multimodal Fusion" variant to test the role of multimodal fusion layer in the overall framework. The ablation results are shown in Table 6. After removing the content feature module, the F1 of the model decreases from 0.895 to 0.846, and the AUC decreases from 0.941 to 0.901, indicating that the topic type, emotional expression, tag structure and visual representation are important bases for user click and stay behavior recognition. After removing the platform variable module, RMSE increased from 0.109 to 0.141, and $R^2$ decreased to 0.801, indicating that the variables such as recommendation strength, release time and ranking score had the most significant role in explaining the growth rate and persistence of diffusion. After removing the social variable module, F1 and AUC decreased to 0.857 and 0.914, respectively, indicating that comment response chain, relationship connection density and forwarding diffusion path had obvious contributions to deep participation behavior and cross-group diffusion recognition. After replacing the multi-modal fusion layer with simple stitching, the overall performance also decreases, indicating that the relationship between different modalities is not simple superposition, but there are structural differences that need to be modeled through the attention mechanism.

*Table 6: Results of ablation experiments*

| Model variants | Accuracy | F1 | AUC | RMSE | $R^2$ |
|---|---|---|---|---|---|
| Full Model | 0.903 | 0.895 | 0.941 | 0.109 | 0.872 |
| w/o Content Features | 0.854 | 0.846 | 0.901 | 0.137 | 0.816 |
| w/o Platform Variables | 0.842 | 0.833 | 0.892 | 0.141 | 0.801 |
| w/o Social Variables | 0.866 | 0.857 | 0.914 | 0.128 | 0.837 |
| w/o Multimodal Fusion | 0.873 | 0.864 | 0.919 | 0.124 | 0.844 |

The comprehensive comparison results show that the advantage of the multi-factor fusion method is not only reflected in the improvement of the index value, but also reflected in the ability to explain the coupling relationship between the propagation mechanism and user behavior. The content characteristics determine the quality basis of recommendation trigger, the platform variables determine the range of traffic distribution and exposure enhancement, and the social variables determine the depth of diffusion and the efficiency of interactive conversion. The three factors together constitute the core mechanism chain of short video propagation evolution. If any module is missing, the model's description of short video propagation path and user behavior transformation process will be significantly degraded, which also verifies the rationality of the analysis framework constructed in this paper from the experimental level.

## 5.4 Robustness and heterogeneity tests

In order to verify the reliability of the model effect verification results in Section 5.2 and the comparison experiment and ablation experiment results in Section 5.3, this paper further tests the robustness and heterogeneity of the existing conclusions from four levels: variable substitution, sample re-division, group estimation and scene stratification. The core of the robustness test is to investigate whether the advantage of the proposed model over the baseline model still holds when the explained variable, sample window or estimation method are changed. The core of heterogeneity test is to investigate whether the role strength of content characteristics, platform variables and social variables will systematically change according to different content categories, user activity and communication scenarios. Based on the complete model in Section 5.2, the core outcome variable in the user behavior recognition task is extended from binary click behavior to high-value interactive behavior recognition, and the explained variables in the propagation effect prediction task are replaced by interaction conversion rate and diffusion persistence from play growth rate, and the model estimation is performed again. If the Accuracy, F1, AUC, RMSE and $R^2$ of the model change slightly after replacing the variables, and the ranking relationship with respect to Logistic Regression, XGBoost and BiGRU remains unchanged, the above conclusions can be considered to be robust. The robustness bias rate is defined as follows.

$$\Delta_m = \frac{|Perf_m^{(r)} - Perf_m^{(b)}|}{Perf_m^{(b)}} \tag{27}$$

Here, $Perf_m^{(b)}$ represents the performance index of model m under the benchmark experiment, and $Perf_m^{(r)}$ represents the performance index after replacing variables or redrawing samples. The test results show that the F1 and AUC deviation rate of the proposed model under different Settings remain in a low range, and the increase of RMSE is limited, indicating that the advantages of the model obtained in Section 5.2 are not driven by the definition of a single indicator.

At the sample level, the results of Section 5.2 and 5.3 are further tested by time repartitioning, and the original 70%/15%/15% train-validation-test division is adjusted to 60%/20%/20% and rolling time window division, and Bootstrap resampling is used to repeat the estimation. Assuming that the model parameter under the BTH resampling is $\hat{\beta}^{(b)}$ and the baseline estimate is $\hat{\beta}$, the parameter stability can be expressed as follows.

$$Var(\hat{\beta}) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\beta}^{(b)} - \bar{\beta}\right)^2 \tag{28}$$

where $\bar{\beta}$ is the resampling mean. The results show that the sign directions of the parameters corresponding to content features, platform variables and social variables are consistent with the significance level, which indicates that the conclusion of the ablation experiment in Section 5.3 that "platform variables have the largest contribution to propagation prediction, social variables have prominent contribution to deep participation recognition, and content features have a fundamental role in recommendation trigger and stay judgment" does not depend on a single sample segmentation. However, it has better estimation stability.

The heterogeneity test revolves around the key mechanisms already identified in the previous section. Content category heterogeneity analysis uses information, knowledge, entertainment and business promotion samples to estimate the full model and ablation model respectively, and tests the marginal contribution difference of different modules under different content types. The grouping model can be expressed as follows.

$$Y_{i,g} = \alpha_g + \beta_{1,g}C_i + \beta_{2,g}P_i + \beta_{3,g}S_i + \varepsilon_{i,g} \tag{29}$$

Here, g represents different content categories, and $C_i$, $P_i$, $S_i$ correspond to content, platform, and social variables, respectively. The results show that in the knowledge and information samples, the content feature variables have stronger explanatory power for stay time and comment behavior, indicating that semantic integrity and information density are key supports for such content dissemination. In the entertainment and commercial promotion samples, the marginal contribution of the platform variable is higher, indicating that the recommendation strength and hot spot distribution mechanism are more sensitive to short-term exposure amplification. The heterogeneity test of user group is based on three types of users: low active, medium active and high active. The results show that the recommendation accuracy has a more significant impact on the click conversion of low active users, while the social variables have a stronger impact on the comments, retweet and follow behaviors of high active users, which is consistent with the identification results of user behavior differences in Section 4.1 and 4.2. The heterogeneity test of propagation scenarios divided the samples into common recommendation scenarios, hot traffic scenarios and social forwarding scenarios. The results show that the content characteristics maintain a stable effect in common recommendation scenarios, the coefficient of platform variables increases significantly in hot scenarios, and the explanatory power of social variables on diffusion persistence is significantly enhanced in social forwarding scenarios. This further confirms the closed-loop mechanism of "platform distribution -- user conversion -- feedback recirculation -- redistribution enhancement" proposed in Sections 3.2 and 4.3.

On the whole, the performance improvement results in Section 5.2 and the module contribution conclusions in Section 5.3 are consistent under the conditions of variable replacement, sample re-division and multi-scene grouping, indicating that the multi-factor fusion method proposed in this paper has good robustness in the propagation effect prediction and user behavior recognition tasks. At the same time, there are significant differences in the magnitude of the coefficients under different content categories, user groups and transmission scenarios, indicating that the short video propagation mechanism and user behavior transformation are not homogeneous processes, but heterogeneous evolutionary processes that are regulated by content attributes, platform scheduling and social structure.

# 6 Conclusion and Prospect

Based on 52,000 short video samples and corresponding user behavior logs collected from January 2024 to December 2024, this paper jointly analyzed the short video content dissemination mechanism and user behavior in mobile social platforms. The results show that short video propagation is not determined by single content quality, but the result of content characteristics, platform mechanism and social structure. Among them, topic focus, emotional expression intensity, information density and multi-modal presentation quality constitute the content basis of propagation trigger, recommendation strength, hot traffic allocation and ranking update mechanism determine the speed and scope of propagation, and social relationship strength, comment reply chain and forwarding path significantly affect the depth and persistence of diffusion. In terms of user behavior, viewing, liking, commenting, forwarding, following and re-visiting show obvious hierarchical differences and group heterogeneity. Low-active users are more dependent on recommendation accuracy to complete click conversion, while high-active users are more likely to be stimulated by social interaction to form deep participation. The experimental results show that the Accuracy, F1 and AUC of the constructed model on the user behavior recognition task reach 0.903, 0.895 and 0.941 respectively, and the RMSE and $R^2$ on the propagation effect prediction task are 0.109 and 0.872 respectively. It shows that incorporating content variables, platform variables and social variables into a unified framework can more effectively reveal the coupling relationship between short video propagation and behavior feedback. Related research has theoretically promoted short video communication research from static effect interpretation to dynamic closed-loop analysis of "content representation -- platform distribution -- user conversion -- feedback reinforcement", and also provided multimodal, sequential and structured modeling ideas for user behavior analysis. In practice, it can provide quantitative basis for platform recommendation optimization, hot spot scheduling, content review and fine operation. At the same time, there are still some limitations in this research. The samples are mainly from publicly available data, and the fine-grained private domain interactions and long-term cross-platform migration behaviors have not been included. Emotion recognition, social relationship strength measures, and long-tail propagation modeling still have room for further improvement. In the future, long-term data, multi-platform comparative data, graph neural network, causal inference and reinforcement learning methods can be combined to deepen the research on the evolution mechanism of short video propagation and the dynamic decision-making process of user behavior.

# References


[1] Bhandari A, Bimo S. Why's everyone on TikTok now? The algorithmized self and the future of self-making on social media[J]. Social media+ society, 2022, 8(1): 20563051221086241.

[2] Siles I, Valerio-Alfaro L, Meléndez-Moran A. Learning to like TikTok... and not: Algorithm awareness as process[J]. New Media & Society, 2024, 26(10): 5702-5718.

[3] Schellewald A. Understanding the popularity and affordances of TikTok through user experiences[J]. Media, Culture & Society, 2023, 45(8): 1568-1582.

[4] Obreja D M. When stories turn institutional: How TikTok users legitimate the algorithmic sensemaking[J]. Social Media+ Society, 2024, 10(1): 20563051231224114.

[5] Dekker C A, Baumgartner S E, Sumter S R. For you vs. for everyone: The effectiveness of algorithmic personalization in driving social media engagement[J]. Telematics and Informatics, 2025, 101: 102300.

[6] Diep P P U, Tran H D. Using TikTok as a search Engine: Affordances, perceived Credibility, and evaluative actions[J]. Telematics and Informatics, 2025: 102324.

[7] Zhang N, Hazarika B, Chen K, et al. A cross-national study on the excessive use of short-video applications among college students[J]. Computers in Human Behavior, 2023, 145: 107752.

[8] Yang H, Zhang S, Diao Z, et al. What motivates users to continue using current short video applications? A dual-path examination of flow experience and cognitive lock-in[J]. Telematics and Informatics, 2023, 85: 102050.

[9] Yang Y, Liu R D, Ding Y, et al. Time distortion for short-form video users[J]. Computers in Human Behavior, 2024, 151: 108009.

[10] Pang H, Ruan Y. Disentangling composite influences of social connectivity and system interactivity on continuance intention in mobile short video applications: The pivotal moderation of user-perceived benefits[J]. Journal of Retailing and Consumer Services, 2024, 80: 103923.

[11] Lu S, Yu M, Wang H. What matters for short videos' user engagement: A multiblock model with variable screening[J]. Expert Systems with Applications, 2023, 218: 119542.

[12] Xiao L, Li X, Zhang Y. Exploring the factors influencing consumer engagement behavior regarding short-form video advertising: A big data perspective[J]. Journal of Retailing and Consumer Services, 2023, 70: 103170.

[13] Dong X, Liu H, Xi N, et al. Short video marketing: what, when and how short-branded

videos facilitate consumer engagement[J]. Internet Research, 2024, 34(3): 1104-1128.

[14] Barta S, Belanche D, Fernández A, et al. Influencer marketing on TikTok: The effectiveness of humor and followers' hedonic experience[J]. Journal of Retailing and Consumer Services, 2023, 70: 103149.

[15] Liu Q Q, Yu S K, Yang Y T. The effects of sponsorship disclosure in short-form video: A moderated mediation model of sponsorship literacy and perceived features of sponsored short-form video[J]. Computers in Human Behavior, 2024, 150: 107969.

[16] Meng L M, Kou S, Duan S, et al. The impact of content characteristics of Short-Form video ads on consumer purchase Behavior: Evidence from TikTok[J]. Journal of Business Research, 2024, 183: 114874.

[17] Alcántara-Pilar J M, Rodriguez-López M E, Kalinić Z, et al. From likes to loyalty: Exploring the impact of influencer credibility on purchase intentions in TikTok[J]. Journal of Retailing and Consumer Services, 2024, 78: 103709.

[18] He D, Yao Z, Teo T S H, et al. How social learning drives customer engagement in short video commerce: An attitude transfer perspective[J]. Information & Management, 2024, 61(6): 104018.

[19] Zhang Q, Wang Y, Ariffin S K. Keep scrolling: An investigation of short video users' continuous watching behavior[J]. Information & Management, 2024, 61(6): 104014.

[20] Cho M, Jeong D, Park E. AMPS: Predicting popularity of short-form videos using multi-modal attention mechanisms in social media marketing environments[J]. Journal of Retailing and Consumer Services, 2024, 78: 103778.