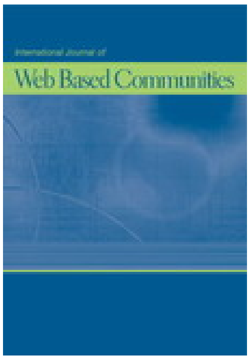


International Journal of Web Based Communities

ISSN online: 1741-8216 - ISSN print: 1477-8394
https://www.inderscience.com/ijwbc

Research on personalised short video push on social media platforms based on affinity propagation clustering

Miao Wang

DOI: 10.1504/IJWBC.2024.10061793

Article History:

Received:	22 May 2023
Last revised:	10 July 2023
Accepted:	10 October 2023
Published online:	04 November 2024

Research on personalised short video push on social media platforms based on affinity propagation clustering

Miao Wang

School of Culture, Tourism and International Education,
Henan Polytechnic Institute,
Nanyang, 473000, China
Email: gongyuanwm@163.com

Abstract: Personalised short video push on social media platforms can help enterprises improve user experience, competitiveness, and marketing effectiveness. This article proposes a personalised short video push method on social media platforms based on affinity propagation clustering. By determining the attractiveness, belonging, and reference between data points, the optimal clustering centre of the affinity propagation clustering algorithm is selected to achieve user behaviour data collection. Based on the data collection results, Markov matrix is used to extract user sentiment labels, combined with sentiment labels and XGBoost model to predict user personalised preferences. The i Expand algorithm is used to determine user interest vectors and generate recommendation lists, achieving personalised short video push on social media platforms. The experimental results show that the maximum push accuracy of this method is 97%, the maximum time consumption is 97.4 ms, and the maximum satisfaction with push results is 98.6.

Keywords: affinity propagation clustering; social media platforms; short video; personalised; push; sentiment labels; XGBoost model; i Expand algorithm.

Reference to this paper should be made as follows: Wang, M. (2024) 'Research on personalised short video push on social media platforms based on affinity propagation clustering', *Int. J. Web Based Communities*, Vol. 20, Nos. 3/4, pp.263–277.

Biographical notes: Miao Wang received her Master's degree in Design from Henan University. She is currently a Lecturer at the School of Culture, Tourism and International Education of Henan Polytechnic Institute. She was awarded as the Provincial Backbone Teacher of Henan Vocational Colleges in 2021 and the Excellent Instructor of the National College Digital Art Design Competition in 2021. She has conducted in-depth research in the fields of art design, digital media technology, and has high attainments.

1 Introduction

With the rapid development of the internet, more and more people are starting to use social media platforms to watch short videos. However, due to the excessive content of short videos, users often need to spend a lot of time searching for their favourite content. Therefore, social media platforms are exploring a more efficient way to provide users with short video content that is more in line with their interests and hobbies through personalised recommendation algorithms (Almeida et al., 2022). Personalised short video push on social media platforms refers to the ability to recommend short video content that best suits users' preferences by analysing their browsing history, interests, and other data. This personalised push model also provides more opportunities for content creators to attract more user attention and fans through high-quality content, thereby gaining more exposure and revenue (Cai et al., 2021). However, there are also some issues with personalised push, such as the possibility of users becoming addicted to certain types of content and lacking diversity. Therefore, it is necessary to maintain rationality and objectivity while enjoying personalised push, and enrich one's perspective (Su et al., 2021). In short, the research significance of personalised short video push on social media platforms lies in improving user browsing efficiency and satisfaction, bringing more exposure opportunities to content creators, and providing users with short video content that suits their taste through analysis and processing of user data, thereby improving user experience and satisfaction, and promoting the application of artificial intelligence technology in the field of the internet. Therefore, researching a new personalised short video push method on social media platforms has important research significance.

Zhu et al. (2021) proposed a personalised push method for short videos on social media platforms based on bullet screen emotion analysis and topic model. Firstly, conduct in-depth mining and clustering of short video barrage comments on social media platforms, and use knowledge graph methods to determine the user sentiment vector of short videos, thereby determining the emotional similarity of different short videos. Secondly, the topic model is constructed based on short video tags to determine the emotional similarity between different short videos. The emotional similarity and emotional similarity analysis results are deeply fused to determine the comprehensive similarity coefficient between different short videos. Finally, based on the historical browsing data of users, personal preferences of users are determined, and their comprehensive recognition of short videos is determined. Combined with relevant analysis results, a personalised push list is generated to achieve personalised short video push on social media platforms. However, in testing, it was found that due to the difficulty in using relevant data to determine emotional similarity, the problem of low push accuracy has arisen, making it difficult to widely apply in practice. Gu et al. (2021) proposed a personalised short video push method for social media platforms based on autoencoder and multimodal data fusion. This method utilises data mining technology to obtain historical browsing data of social media platform users, and integrates multimodal data for processing. Using word bags and TF-IDF methods to describe comment and comment data, text data features are obtained. Deep convolutional neural networks are used to multimodally describe video documents, and self encoders are used to determine user personalised features. Relevant push lists are generated, and appropriate social media platform short videos are pushed to users based on user personalised features. However, the execution steps of this method are too complex, resulting in a significant increase in push time consumption, and the actual application effect is not good. Gao et al. (2021) proposed a personalised short video push method based on graph neural networks for social media platforms. Firstly, a personalised push model based on graph neural networks was constructed. This model mainly models the user's video viewing sequence behaviour as a graph structure, using nodes to represent the user and the video, and edges to represent the behaviour between them. At the same time, two types of vector propagation methods were introduced to model users' long-term and short-term interests, respectively. In this model, the long-term interests of users are depicted through bidirectional propagation between users and short videos. By one-way propagation of video node switching relationships, users' short-term interests can be characterised. In addition, capturing high-order adjacency information on the graph through multi-layer vector propagation can determine user personalised preferences and recommend suitable short videos on social media platforms for users. However, in practical applications, it has been found that this method has the problem of low satisfaction with short video push results on social media platforms, and there is still a certain gap between it and the expected goals.

Due to the lack of rich user behaviour data collected during the design process of the current method, it is difficult to predict user interest, resulting in low push accuracy, long recommendation time consumption, and low satisfaction. With the goal of solving the thank you problem, a new personalised short video push method on social media platforms based on affinity propagation clustering has been proposed, and the effectiveness of this method was tested through experiments.

2 Personalised short video push method on social media platforms

2.1 User behaviour data collection based on affinity propagation clustering algorithm

Affinity propagation clustering is a graph-based clustering algorithm. It infers the cluster to which each data point belongs by constructing a similarity graph between data points and utilising the similarity between neighbours.

The objectives of using neighbour propagation clustering algorithm for user behaviour data collection mainly include the following aspects:

1 Discovering user interests and user groups: through neighbour propagation clustering algorithms, relationships and behaviours between users can be mined and analysed to discover their interests and preferences. By clustering user groups, it is possible to better understand user behaviour patterns and group characteristics.

2 Personalised recommendation and personalised marketing: by collecting user behaviour data and conducting cluster analysis, users' historical behaviour and preference patterns can be obtained. Based on these data, personalised recommendation services and customised marketing strategies can be provided for each user to improve their engagement and satisfaction.

3 Target advertising and precise marketing: by collecting user behaviour data and performing cluster analysis, potential customer groups can be identified, and target advertising and targeted marketing activities can be carried out at precise time points according to the needs and interests of different groups, so as to improve advertising effect and user responsivity.

4 Mining social influence: collect user relationship network and behaviour data, combined with neighbour propagation clustering algorithm, can identify users with social influence, help enterprises or platforms find key opinion leader and social influencers, and use their influence to promote products, services or brands.

The advantages of affinity propagation clustering algorithm in user behaviour data collection (Gao, 2021) mainly include:

1 Strong adaptability: affinity propagation clustering algorithm can adaptively adjust the number and size of clusters, suitable for various datasets of different sizes and densities.

2 Strong robustness: affinity propagation clustering algorithm has good robustness to noise and outliers, and can effectively handle abnormal situations in data.

3 High computational efficiency: affinity propagation clustering algorithms have low computational complexity and can handle large-scale datasets.

4 Good stability of results: the results of the neighbouring propagation clustering algorithm are relatively stable, and there will be no significant differences in results due to different initialisation.

Assuming that the user behaviour dataset is represented by $X = \{x_1, x_2, \ldots, x_N\}$, there are closely related clusters in a certain feature space, $C = \{C_1, C_2, \ldots, C_K\}$, $k \in N$. Each data is only relative to one cluster. Assuming that there is a representative cluster $x_{C(i)}$ and x_i in this cluster, the error function in the clustering process is expressed by the following formula:

$$J(C) = \sum_{i=1}^{N} d^2\left(x_i, x_{C(i)}\right) \tag{1}$$

In the above formula, d represents clustering similarity.

The objective of the affinity propagation clustering algorithm is to determine the optimal set of representative points of the class and minimise the error function, then the objective function is as follows:

$$C^* = \arg\min[J(C)] \tag{2}$$

Assuming $x_{C(i)}$ exists at another data point x_j, determine the distance between x_i and x_j to obtain the similarity between these two data points. The specific calculation formula is as follows:

$$d = -\sqrt{\left(x_i - x_j\right)^2} \qquad i \neq j \tag{3}$$

To select the optimal clustering centre as the goal, it is necessary to update the attractiveness $r(i, j)$ and belonging $a(i, j)$ between data points. The information dissemination process during the update process is shown in Figure 1.

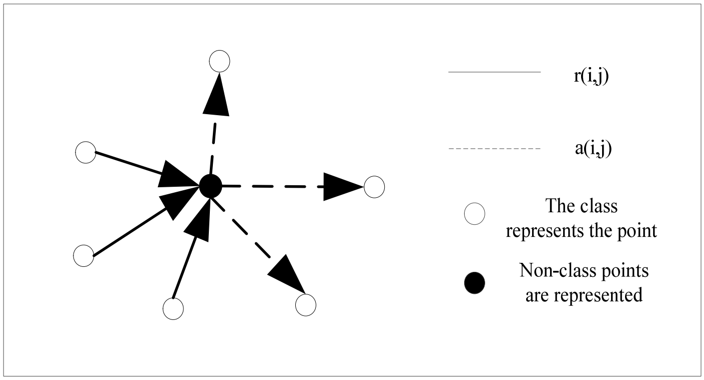
The specific update formula is as follows:

$$r(i, j) = s(i, j) - \max_{k \neq j}\{a(i, k) + s(i, k)\} \tag{4}$$

$$a(i, j) = \min\left\{0, r(i, j) + \sum_{k \neq i, j} \max\{0, r(k, j)\}\right\} \qquad i \neq j \tag{5}$$

In the above formula, $s(i, j)$ represents the reference degree for selecting the j^{th} point as the class representative point for the i^{th} point, $a(i, k)$ and $s(i, k)$ represent the initial and optimal clustering centres, respectively, and $r(k, j)$ represents the current clustering centre.

Figure 1 Information dissemination process



In the process of updating the affinity propagation clustering algorithm, in order to avoid oscillations, a damping factor of λ is introduced to continuously update the attractiveness and affiliation. The specific update formula is as follows:

$$r_{new}(i, k) = \lambda^* r_{old}(i, k) + (1-\lambda)^* r(i, k) \tag{6}$$

$$a_{new}(i, k) = \lambda^* a_{old}(i, k) + (1-\lambda)^* a(i, k) \tag{7}$$

During the continuous iteration process, determine whether the number of representative points in the obtained cluster meets the requirements. If not, change the value, repeat the program until the number of representative points in the cluster meets the requirements, and output the final clustering result. This result is the user behaviour data collection result, specifically represented by the following formula.

$$E = R + r_{new}(i, k) - \max_{k' \neq k} a_{new}(i, k) \tag{8}$$

In the formula, R represents the iterative update matrix.

2.2 User sentiment label extraction

Based on the data collection results, Markov matrix is used to extract user emotional labels, laying a solid foundation for subsequent personalised preference prediction of users. Assuming that there are D documents in the user behaviour data collection result C, with $C = \{d_1, d_2, \ldots, d_D\}$, w representing one of the words in document d, l representing the emotional label of document d (Yang and Lin, 2022), and z representing the theme of the document, the joint probability of topic/emotional label allocation and words in the user behaviour data can be expressed by the following formula:

$$P(w, z, l) = P(w|z, l)P(z|l, d)P(l|d) \tag{9}$$

In the above formula, $P(w|z, l)$ represents the probability of topic label allocation, $P(z|l, d)$ represents the probability of emotional label allocation, and $P(l|d)$ represents the probability of emotional word allocation (Li, 2022).

The calculation formula for $P(w|z, l)$ is as follows:

$$P(w|z,l)=\left(\frac{\Gamma(V\beta)}{\Gamma(\beta)^{V}}\right)^{T*S}\prod_{j}\prod_{k}\frac{\prod_{i}\Gamma\left(N_{i,j,k}+\beta\right)}{\Gamma\left(N_{j,k}+V\beta\right)} \tag{10}$$

In the above formula, V represents the emotion dictionary, T represents the number of topic words, S represents the number of emotion labels, $N_{i,j,k}$ represents the number of times word i appears under topic j and emotion label k, $N_{j,k}$ represents the number of times word is assigned to topic j and emotion label k, and Γ represents the gamma function.

The calculation formula for $P(z|l, d)$ is as follows:

$$P(z|l,d)=\left(\frac{\Gamma(T\alpha)}{\Gamma(\alpha)^{T}}\right)^{S*D}\prod_{k}\prod_{d}\frac{\prod_{j}\Gamma\left(N_{j,k,d}+\alpha\right)}{\Gamma\left(N_{k,d}+T\alpha\right)} \tag{11}$$

In the above formula, $N_{j,k,d}$ represents the number of times a word appears in document d with topic j and label k, and $N_{k,d}$ represents the number of times emotional label k is assigned to some words in document d (He et al., 2021).

The calculation formula for $P(l|d)$ is as follows:

$$P(l|d)=\left(\frac{\Gamma(S\gamma)}{\Gamma(\gamma)^{S}}\right)^{D}\prod_{d}\frac{\prod_{k}\Gamma\left(N_{k,d}+\gamma\right)}{\Gamma\left(N_{d}+S\gamma\right)} \tag{12}$$

In the above formula, N_d represents the total number of words in the document set.

Combining Markov chains to determine the distribution of words in topic and emotional labels (Dias et al., 2021), the specific calculation formula is as follows:

$$\varphi_{i,j,k}=\frac{N_{i,j,k}+\beta}{N_{j,k}+V\beta} \tag{13}$$

In the above formula, β represents the transition probability of the Markov matrix, and the results of user sentiment label extraction are as follows:

$$\pi_{k,d}=\frac{\varphi_{i,j,k}\left(N_{k,d}+\gamma\right)}{N_{d}+S\gamma} \tag{14}$$

In the above formula, γ represents the feature parameter of emotional labels (Hong, 2021).

2.3 User personalised preference prediction

Combining emotional labels and XGBoost model to predict user personalised preferences can provide important directions for personalised short video push. XtremeGradient boosting (XGBoost) is a boosting ensemble learning algorithm that integrates multiple weak classifiers. It has the advantages of high accuracy, strong robustness, processing high-dimensional features, strong interpretability, and good scalability, and can accurately predict user personalised preferences.

Assuming that the user behaviour data collection result can be represented by $C = \{(x_i, y_i)\}$, where $x_i \in R^m$, $y_i \in r$, $i = 1, 2, \ldots, n$. x_i represents the feature dataset, m represents the number of features (Konapure and Lobo, 2021), and y_i represents the label value as the target variable. Assuming that there are a total of K trees in the XGBoost model, the XGBoost model can be represented by the following formula:

$$\hat{y}_i = F_{K-1}(x_i) + f_K(x_i) \tag{15}$$

In the above formula, $f_k(x_i)$ represents the prediction result of the K^{th} tree of XGBoost model, and $\hat{y}_i$ represents the sum of the prediction results. The calculation formula of the objective loss function of XGBoost is as follows:

$$Obj = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{16}$$

In the above formula, $L(y_i, \hat{y}_i)$ represents the loss function and $\Omega(f_k)$ represents the regularisation parameter. The specific calculation formula is as follows:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|\omega\|^2 \tag{17}$$

In the above formula, γ, λ represents different hyperparameter, and T, ω represents different leaf node weight values. The objective function described in formula (16) is difficult to optimise in Euclidean space, and the XGBoost model uses approximate rewriting to solve this problem. The rewritten objective function is as follows:

$$Obj^{(s)} = \sum_{i=1}^{n} L\left(y_i, \hat{y}_i^{(s-1)} + f_s(x_i)\right) + \sum_{k=1}^{K} \Omega(f_s) \tag{18}$$

In the above formula, $\hat{y}_i^{(s-1)}$ represents the personalised preference prediction results for users in $S - 1$ rounds (Ikram and Farooq, 2022), and $f_s(x_i)$ represents the single tree prediction results that exist in the S^{th} round.

Assuming $f(x)$ has a n^{th} order continuous derivative in $[a, b]$ and a $n + 1^{\text{th}}$ order continuous derivative in (a, b). If $x_0 \in [a, b]$ is a certain point, then for any $x \in [a, b]$, there exists:

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \ldots \\ &+ \frac{f^n(x_0)}{n!}(x - x_0)^n + r_n(x) \end{aligned} \tag{19}$$

If the value of ξ is between x and x_0, the remaining term can be represented by the following formula:

$$r_n(x) = \frac{f^{n+1}(\xi)}{(n+1)!}(x - x_0)^{n+1} \tag{20}$$

Consider $\hat{y}_i^{(s-1)} + f_s(x_i)$ as x and $\hat{y}_i^{(s-1)}$ as x_0 in the objective function to obtain the Taylor expansion:

$$Obj^{(s)} \cong \sum_{i=1}^{n} L\left(y_i, \hat{y}_i^{(s-1)}\right) + g_i f_s(x_i) + \frac{1}{2} h_i f_s^2(x_i) + \Omega(f_s) \tag{21}$$

In the above formula, g_i and h_i represent the sum of the first and second steps of the loss function respectively, and $L(y_i, \hat{y}_i^{(s-1)})$ represent the constant term of the loss function. When $\Omega(f_s)$ specific formulas are introduced into formula (21), we can get:

$$Obj' = \sum_{i=1}^{n} \left[g_i \omega_{q(x_i)} + \frac{1}{2} h_i \omega^2{}_{q(x_i)} \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=i}^{T} \omega_j^2 \tag{22}$$

In the above formula, $\omega_{q(x_i)}$ represents the weight values of T leaf nodes in the S^{th} round of training a single tree model, traversing all leaf nodes to obtain personalised preference prediction results for users, as follows:

$$\omega_j^* = -\frac{Obj' \sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{23}$$

2.4 Personalised short video push

Utilise the i Expand algorithm to determine user interest vectors and generate recommendation lists, achieving personalised short video push on social media platforms. The i Expand algorithm is a push algorithm based on user behaviour sequences. Its core idea is to improve the accuracy and coverage of push by utilising user behaviour sequences and similar user behaviour information, and has good push performance and scalability. The i Expand algorithm provides a novel way to personalised push short videos on social media platforms, introducing some unique recommendation ideas, applying novel mining methods, or improving the core process of existing algorithms. Moreover, the algorithm utilises a new data processing method and incorporates sorting techniques to ensure the scientific and reliable recommendation results.

Determine the probability $\overrightarrow{vl}$ of interest T_i for each social media platform user (Chen, 2021), and the results are as follows:

$$\overrightarrow{vl} = P(T_i) = \frac{C_{ij}^{MK} + \alpha}{\sum_{k=1}^{K} \sum_{m=1}^{M} C_{mk}^{MK} + K\alpha} \tag{24}$$

In the above formula, T_i represents the user interest of social media platforms, C_{ij}^{MK} and C_{mk}^{MK} represent the user interest and sentiment matrix, respectively, K represents the number of user interests, and α represents the user interest feature parameters.

In order to facilitate the analysis and calculation of the transfer probability between interests, and to construct a user interest correlation graph, let φ_j represent the probability distribution of the project on user interests on social media platforms (Huang, 2021). The specific calculation formula is as follows:

$$\varphi_j = P(T_j, T_i) = \frac{P(T_j, I_i)}{P(I_i)} = \frac{\vartheta_{ij} \overrightarrow{vl}}{\sum_{k=1}^{K} \overrightarrow{vk} \theta_{ik}} \tag{25}$$

In the above formula, T_i represents the i^{th} interest of the user, I_i represents the i^{th} short video on social media platforms, θ_{ik} represents the probability of interest transfer, and ϑ_{ij}

represents the edge weight of interest T_i to item I_i. The specific calculation formula is as follows:

$$\vartheta_{ij} = P\left(T_j \left| T_i\right.\right) = \sum\nolimits_{n=1}^{N} P\left(T_j \left| I_n\right.\right) P\left(I_n \left| T_i\right.\right) = \sum\nolimits_{n=1}^{N} \vartheta_{nj} \vartheta_{ni} \tag{26}$$

In the above formula, ϑ_{nj} and ϑ_{ni} represent interest feature vectors and interest distribution vectors, respectively.

Assuming $\vec{\theta}_l^{(0)}$ represents the user's initial interest vector, when it swims to the j^{th} node T_j, it will continue to randomly select among all possible neighbours of T_j, and then continue to walk. The decision analysis of each step of the vector will finally return to the initial state, where the return probability value is 7. When performing random walk with U_i, the user interest vector in step $s + 1$ is expressed by the following formula:

$$\vec{\theta}_l^{(s+1)} = (1-c)\vec{\theta}_1^{(s)} \vartheta + c\vec{\theta}_l^{(0)} \tag{27}$$

In the above formula, $\vec{\theta}_1^{(s)}$ represents the neighbourhood user interest feature vector, and ϑ represents the decision parameter.

The i Expand algorithm sorts short videos based on the preference level between users and social media platforms. This method analyses the random distribution of users' interests:

$$P\left(I_j \left| U_i\right.\right) = \sum\nolimits_{k=1}^{K} P\left(I_j \left| t = k\right.\right) P_s\left(t = k \left| U_i\right.\right) = \sum\nolimits_{k=1}^{K} \vartheta_{jk} \vec{\theta}_l^{(s+1)} \tag{28}$$

In the above formula, ϑ_{jk} represents the feature vector of short videos on social media platforms. Using the Pearson correlation calculation method to process and expand the interest similarity between users, the interest similarity can be represented as $sim(U_i, U_h)$. Through this method, it is possible to obtain the nearest neighbour set $Neighbour(U_i)$ of all individual users U_i. The rating of each user on short videos on social media platforms can be expressed as:

$$\hat{r}_{ij} = \hat{r}_i + \frac{\sum\nolimits_{uh \in Neighbour(U_i)} sim\left(U_i, U_h\right) * \left(r_{h,j} - \hat{r}_h\right)}{\sum\nolimits_{uh \in Neighbour(U_i)} \left| sim\left(U_i, U_h\right) \right|} \tag{29}$$

In the above formula, $\hat{r}_i$ and $\hat{r}_h$ represent the average score of U_i and U_h, $r_{h,j}$ represents the predicted rating of user U_h for short videos I_j on social media platforms.

Using the Gibbs sampling method to sample and update their interests, during the algorithm iteration calculation process, user U_u interest in short videos on social media platforms is assigned a value, and the results are as follows:

$$\beta = \frac{C_{I_i^u}^{NK} + \vec{C}_{j-u}^{u} + \beta}{\sum\nolimits_{n=1}^{N} C_{nj}^{NK} + \vec{C}_j^u + NB - 1 \sum\nolimits_{k=1}^{K} C_k^u + K\alpha - 1} \tag{30}$$

In the above formula, t_i^u represents assigning interest T_j to I_i^u, and $\vec{C}_j^u$ represents how many selected items from $\vec{C}^u$ have been assigned interest T_j to U_u. After all interests have been reassigned, the probability calculation formula for each interest is as follows:

$$\theta_{uj} = P\left(T_j \middle| U_u\right) = \frac{\vec{C}_j^u + \alpha\beta}{\sum_{k=1}^{K} \vec{C}_k^u + K\alpha} \tag{31}$$

Optimise and adjust the user interest feature parameter α and interest assignment parameter β. The calculation formula for the optimised and adjusted parameters is as follows:

$$\alpha^* \leftarrow \frac{\alpha \sum_{m=1}^{M} \sum_{k=1}^{K} \left[\delta\left(C_{mk}^{MK} + \alpha\right) - \delta(\alpha)\right]}{K \sum_{m=1}^{M} \left[\delta\left(\sum_{k=1}^{K} C_{mk}^{MK} + K\alpha\right) - \delta(K\alpha)\right]} \tag{32}$$

$$\beta^* \leftarrow \frac{\beta \sum_{k=1}^{K} \sum_{n=1}^{N} \left[\delta\left(C_{nk}^{MK} + \beta\right) - \delta(\beta)\right]}{N \sum_{k=1}^{M} \left[\delta\left(\sum_{n=1}^{N} C_{nk}^{MK} + N\beta\right) - \delta(N\beta)\right]} \tag{33}$$

Generate a recommendation list based on user interest characteristics, and push the top n short videos from social media platforms to relevant users. The specific recommendation results are as follows:

$$Rec(r_i) = \alpha \cdot Q'(r_j) + \beta \cdot Pop(r_j) \tag{34}$$

In the above formula, $Pop(r_j)$ represents the popularity of short videos on social media platforms, and $Q'(r_j)$ represents the popularity of short videos on social media platforms.

3 Experimental design

The research objective was to verify the effectiveness of the personalised short video push method on social media platforms based on affinity propagation clustering proposed in this article on social media platforms. Relevant experimental tests were conducted, and the specific experimental plan is as follows:

3.1 Experimental data

Utilise data mining techniques to obtain relevant data on social media platforms, including user behaviour data, video feature data, and real-time feedback data. Divide the obtained experimental data into test sets and experimental sets in a 4:6 ratio. Input the data from the test set into MATLAB 7.2 simulation software, locate the data import function, select the data file in the test set to be imported, and set it according to software requirements. After importing data, it is necessary to verify it to ensure its correctness and completeness. If there are issues with the data, it needs to be modified or re imported. After completing data import and multiple verifications, the optimal operating parameters are obtained and used as the initial simulation parameters for MATLAB 7.2 simulation software. The data from the test set is then imported into the simulation software to complete the relevant simulation experiments. The experimental data types are shown in Table 1.

Table 1 Experimental data types

Primary classification of experimental data	*Secondary classification of experimental data*
User behaviour data	Social media platform users' viewing history, likes, comments, sharing and other behavioural data, used to understand user interests, preferences, consumption habits and other information
Video feature data	Social media platforms use feature data such as short video titles, covers, duration, tags, classifications, etc. to analyse video content and attributes
Real time feedback data	Real time feedback data from users, such as dwell time, sliding speed, etc.

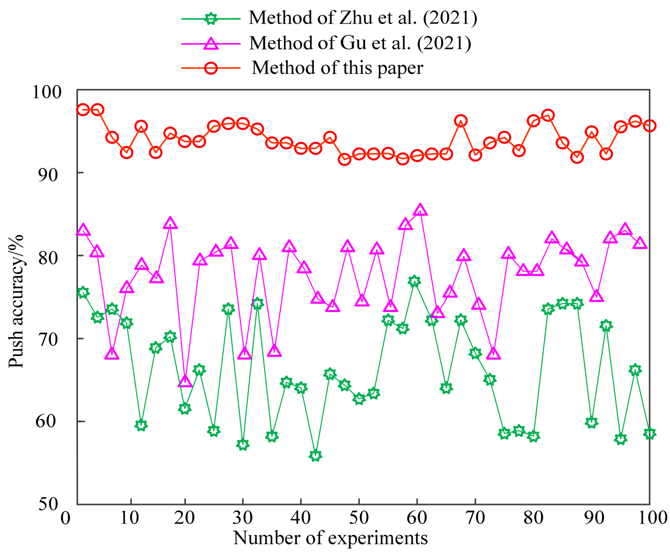
3.2 Evaluation indicators

Using method of Zhu et al. (2021), method of Gu et al. (2021), and method of this paper as experimental comparison methods, the application effectiveness of different methods was tested by comparing the accuracy, time consumption, and satisfaction with short video personalised push on social media platforms using different methods. Push accuracy refers to the degree of matching between short videos recommended by different methods on social media platforms and videos of actual interest to users. The higher this value, the higher the recommendation accuracy; The time consumption of short video push on social media platforms refers to the time spent by recommendation algorithms from receiving user requests to returning recommendation results. For personalised short video push on social media platforms, the speed of push time consumption directly affects the user experience and platform service quality; the goal of personalised short video push on social media platforms is to provide users with interesting and valuable video content, thereby increasing user retention and platform revenue. Therefore, satisfaction with the push results is an important indicator for measuring recommendation algorithms and strategies. In this experiment, this article mainly conducted personalised short video push experiments on social media platforms on 10,000 users. The satisfaction ratings of users with the short video push results on social media platforms were obtained through a survey questionnaire. The users were divided into an average of ten groups, and the satisfaction scores of each group were mainly taken as the average.

The comparison results of the accuracy of personalised short video push on social media platforms using three methods are shown in Figure 2.

The comparison results in Figure 2 show that the maximum accuracy of personalised short video push on the social media platform of method of Zhu et al. (2021) is 76%, and the maximum accuracy of personalised short video push on the social media platform of method of Gu et al. (2021) is 85%. The maximum accuracy of personalised short video push on the social media platform of this method is 97%, which is 21% higher than method of Zhu et al. (2021) and method of Gu et al. (2021), respectively 12%; the minimum accuracy of personalised short video push on the social media platform of method of Zhu et al. (2021) is 56%, and the minimum accuracy of personalised short video push on the social media platform of method of Gu et al. (2021) is 64%. The minimum accuracy of personalised short video push on the social media platform of this article is 91%, which is 35% and 27% higher than method of Zhu et al. (2021) and method of Gu et al. (2021), respectively. No matter from which aspects, the push accuracy of the method in this article is the highest, indicating that this method can provide short videos on social media that users are truly interested in. The reason is that this method combines emotional labels and XGBoost model to predict user personalised preferences, uses the i Expand algorithm to determine user interest vectors and generate recommendation lists, achieving personalised short video push on social media platforms, thus possessing high recommendation accuracy.

Figure 2 Comparison results of push accuracy of three methods



The comparison results of the time consumption for personalised short video push on social media platforms using three methods are shown in Table 2.

Table 2 Comparison results of push time consumption (unit: ms)

Number of experiments	*Method of Zhu et al. (2021)*	*Method of Gu et al. (2021)*	*Method of this paper*
10	156.2	266.1	89.6
20	147.2	247.5	97.4
30	146.3	211.6	85.4
40	125.8	247.6	75.6
50	169.3	135.2	62.3
60	155.1	314.1	58.9
70	169.4	227.7	68.4
80	145.3	253.6	82.3
90	138.2	253.4	84.7
100	176.3	274.1	82.1
Mean value	152.9	243.1	78.7

Analysing the data in Table 2, it can be seen that the maximum, minimum, and average consumption of personalised short video push time on the social media platform of method of Zhu et al. (2021) are 176.3 ms, 125.8 ms, and 152.98 ms, respectively. The maximum, minimum, and average consumption of personalised short video push time on the social media platform of method of Gu et al. (2021) are 314.1 ms, 135.2 ms, and 243.1 ms, respectively. The maximum, minimum, and average time consumption for personalised short video push on social media platforms using this method are 97.4 ms, 58.9 ms, and 78.7 ms, respectively. Overall, the personalised push time consumption of short videos on social media platforms using this method is lower, indicating that this method can push short videos on social media platforms to relevant users. The reason is that this method utilises the i Expand algorithm to determine user interest vectors and generate recommendation lists, thereby ensuring recommendation accuracy and efficiency. Therefore, the push time consumption of this method is relatively low.

The satisfaction comparison results of personalised short video push results on social media platforms using three methods are shown in Table 3.

Table 3 Comparison of satisfaction with push results

Number of experiments	*Method of Zhu et al. (2021)*	*Method of Gu et al. (2021)*	*Method of this paper*
Group 1	84.7	65.2	91.4
Group 2	86.3	63.5	96.3
Group 3	84.1	54.7	94.7
Group 4	75.2	71.2	95.2
Group 5	71.3	70.6	92.3
Group 6	75.6	71.1	94.5
Group 7	71.9	75.4	96.3
Group 8	75.6	70.3	94.4
Group 9	80.1	71.2	92.5
Group 10	78.3	69.8	98.6
Mean value	78.3	68.3	94.6

By analysing the results in Table 3, it can be seen that the maximum satisfaction value of the social media platform short video personalised push results proposed in this article is 98.6, which is 12.3% higher than method of Zhu et al. (2021) and 27.4% higher than method of Gu et al. (2021). The minimum satisfaction value for personalised short video push results on social media platforms using this method is 91.4, which is 19.9% higher than method of Zhu et al. (2021) and 36.7% higher than method of Gu et al. (2021). The average satisfaction rate with personalised short video push results on social media platforms using this method is 94.6, which is 16.3 higher than method of Zhu et al. (2021) and 26.7 higher than method of Gu et al. (2021). From these data, it can be seen that the satisfaction of the push results in this article is higher, indicating that the application effect of this method is better. The reason is that this method uses Markov matrix to extract user sentiment labels, combines sentiment labels with XGBoost model to predict user personalised preferences, in order to ensure that recommendation results can meet user needs to the greatest extent, and thus maintain a high level of user satisfaction.

4 Conclusions

The development of personalised short video push on social media platforms has become a trend in the current internet field. Through personalised recommendation algorithms, social media platforms can provide users with short video content that is more in line with their interests and hobbies, improve user experience and satisfaction, and also bring more exposure opportunities for high-quality content creators. However, there are also some issues with personalised short video push on social media platforms, such as low push accuracy, high time consumption, and low satisfaction. Therefore, this article proposes a new personalised short video push method on social media platforms based on affinity propagation clustering. Through experimental testing, it has been proven that the maximum accuracy of personalised short video push on social media platforms using this method is 97%, the maximum push time consumption is 97.4 ms, and the maximum satisfaction with push results is 98.6. It has the characteristics of high push accuracy, low time consumption, and high satisfaction. In future research, while ensuring user experience and exposure opportunities for creators, attention should be paid to the diversity and objectivity of information to promote the healthy development of personalised short video push on social media platforms.

References

Almeida, A., Villiers, J., Freitas, A.D. et al. (2022) 'The complementarity of a diverse range of deep learning features extracted from video content for video recommendation', *Expert Systems with Applications*, Vol. 192, No. 1, pp.116335–116348.

Cai, D., Qian, S., Fang, Q. et al. (2021) 'Heterogeneous hierarchical feature aggregation network for personalized micro-video recommendation', *IEEE Transactions on Multimedia*, Vol. 24, No. 1, pp.805–818.

Chen, J. (2021) 'Intelligent recommendation system of dance art video resources based on the wireless network', *Security and Communication Networks*, Vol. 17, No. 1, pp.59–62.

Dias, L.L., Barrére, E. and Souza, J. (2021) 'The impact of semantic annotation techniques on content-based video lecture recommendation', *Journal of Information Science*, Vol. 47, No. 6, pp.740–752.

Gao, C., Li, Y. and Jin, D.P. (2021) 'Video recommender system with graph neural networks', *ZTE Technology Journal*, Vol. 27, No. 1, pp.27–32.

Gao, T. (2021) 'Research on short video recommendation strategy based on big data analysis', *Journal of Physics: Conference Series*, Vol. 1941, No. 1, pp.012071–012078.

Gu, Q.Y., Ju, C.H. and Wu, G.X. (2021) 'Fusion of auto encoders and multi-modal data based video recommendation method', *Telecommunications Science*, Vol. 37, No. 2, pp.82–98.

He, P., Ma, S. and Li, W. (2021) 'Efficient barrage video recommendation algorithm based on convolutional and recursive neural network', *Journal of Internet Technology*, Vol. 22, No. 6, pp.1241–1251.

Hong, K. (2021) 'Automatic recommendation algorithm for video background music based on deep learning', *Complexity*, Vol. 21, No. 4, pp.1–11.

Huang, F. (2021) 'Personalized marketing recommendation system of new media short video based on deep neural network data fusion', Vol. 9, No. 1, pp.1–10, Hindawi Limited.

Ikram, F. and Farooq, H. (2022) 'Multimedia recommendation system for video game based on high-level visual semantic features', *Scientific Programming*, Vol. 25, No. 1, pp.1–12.

Konapure, R.C. and Lobo, L. (2021) 'Video content-based advertisement recommendation system using classification technique of machine learning', *Journal of Physics: Conference Series*, Vol. 15, No. 6, pp.182–196.

Li, S. (2022) 'A heuristic video recommendation algorithm based on similarity computation for multiple features analysis', *Recent Advances in Computer Science and Communications*, Vol. 15, No. 8, pp.1017–1025.

Su, C., Zhou, H., Wang, C. et al. (2021) 'Individualized video recommendation modulates functional connectivity between large scale networks', *Human Brain Mapping*, Vol. 42, No. 16, pp.5288–5299.

Yang, Z. and Lin, Z. (2022) 'Interpretable video tag recommendation with multimedia deep learning framework', *Internet Research: Electronic Networking Applications and Policy*, Vol. 32, No. 2, pp.518–535.

Zhu, S.M., Wei, S.W., Wei, S.H. et al. (2021) 'Video recommendation algorithm based on danmaku sentiment analysis and topic model', *Journal of Computer Applications*, Vol. 41, No. 10, pp.2813–2819.