# Research on Personalized Recommendation Algorithm Based on Mobile Social Network Data

Xiang Yan[1] and Wen-hua Deng[2](✉)

[1] Information Center of Pu'er University, Puer 665000, China
[2] Wuhan Railway Vocational College of Technology, Wuhan 430205, China

**Abstract.** Due to the lack of mining of hidden data in traditional personalized recommendation algorithms, the algorithm is interfered by the mobile social network environment, and it is difficult to accurately recommend targeted data for users. Therefore, research on personalized recommendation algorithms based on mobile social network data. By dividing mobile social network user categories, user information is obtained; based on mobile social network data, user demand characteristics are extracted; potential association rules between users and service needs are mined to build personalized recommendation algorithms. The experimental results show that compared with the traditional recommendation algorithm, the research algorithm has stronger perception and recognition ability, and it can recommend more matching information for users according to different user needs when facing different network environments.

**Keywords:** Mobile social network data · Personalization · Recommendation algorithm · Association rules

## 1 Introduction

At present, with the rapid development of Internet and computer technology, the fields of industry, commerce, agriculture, aerospace, environmental measurement, education and teaching are inseparable from the use of the network, so a large number of network users bring great work to the network Pressure, in order to improve the efficiency of the network, research a new personalized recommendation algorithm, for users of different ages, different work fields and different living environments, recommend network information that meets their respective needs [1].

Literature [2] proposed a hybrid algorithm for personalized recommendation for interactive experience devices. The algorithm sequentially uses collaborative filtering and content-based recommendation methods for recommendation. In the initial recommendation, the latent Dirichlet allocation (LDA) topic model is used to reduce the dimensionality of high-dimensional user behavior data, and the user writing topic matrix is established to reduce the recommendation inaccuracy caused by the high sparse data in the collaborative filtering algorithm. The user interest list is obtained by calculating the similarity between users. Then, on the basis of the preliminary recommendation results,


S. Liu and X. Ma (Eds.): ADHIP 2021, LNICST 416, pp. 191–203, 2022.
https://doi.org/10.1007/978-3-030-94551-0_16

the VGG16 model is used to extract the feature vector of the calligraphy image, and the similarity between the user's calligraphy words and the primary recommendation calligraphy words is calculated, so as to obtain the final recommendation result. The recommendation accuracy of this algorithm is high, but the efficiency is poor. Literature [3] proposes a personalized recommendation system for academic libraries based on a hybrid recommendation algorithm. First, the paper studies the application of collaborative filtering and content-based recommendation algorithms in college book recommendation, involving reader classification, user-item scoring matrix establishment, vector space model construction, and similarity calculation. user. And considering the characteristics of college books and readers, the user-item rating matrix is improved, and clustering is used to alleviate the problem of data sparseness. This method can quickly realize personalized recommendation, but the recommendation matching degree is low.

This research combines the problems of a variety of traditional algorithms and introduces mobile social network data. This type of data has the characteristics of many types, large amounts, and wide coverage. These data provide the most basic raw data for the recommendation work [4]. This research combines the characteristics of mobile social network data to establish a new personalized recommendation algorithm. Obtain user information by classifying mobile social network users; extract user demand characteristics based on mobile social network data; mine potential association rules between users and service demands, and build personalized recommendation algorithms. Compared with the traditional algorithm, the research algorithm has stronger perception and recognition ability, and can recommend more matching information for users according to different user needs.

## 2 Personalized Recommendation Algorithm Based on Mobile Social Network Data

### 2.1 Classification of Mobile Social Network Users

According to the browsing trajectory of the mobile social network user when browsing the web page information, the user attribute preference is calculated, and the score is used as the basis for user classification. Set user tags, including: finance, technology, digital, social, transportation, weather, news, law, brand, food, insurance, etc. Using mathematical algorithms, calculate the logic and similar preferences that exist in the user's browsing trajectory to form a classification definition. Assuming that there is $m$ user preference $A$, the calculation result of its mathematical algorithm is:

$$Y_i \sim Z\left(\alpha_i, \beta^2\right), i = 1, 2, \ldots, m \tag{1}$$

In the formula: $Y_i$ represents the sample data set; $\alpha_i$ represents the label offset threshold affected by the change of preference $A$; $\beta$ represents the preference difference. The problem of saliency is transformed into a question of whether preference $A$ is in the $Z$ space, and whether it affects web browsing selection behavior, that is, whether $G_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_m$ is valid or not. The following equations are given, and the various parameters are the indicators required for verification.

$$\begin{cases} \bar{y} = \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij} & S_1^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}\right)^2 \\ S_2^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2 & S_3^2 = \sum_{i=1}^{w}\sum_{j=1}^{n_i}(y_i - \bar{y})^2 \end{cases} \tag{2}$$

In the above formula, $n$ represents the total number of results; $\bar{y}$ represents the total mean; $S_1^2$ represents the sum of squares of the total variance; $S_2^2$ represents the sum of squares within the group; $S_3^2$ represents the sum of squares between the groups [5]. According to the above indicators, $G_0$ rejection domains are obtained:

$$W = \left\{ \frac{(n-m)S_3^2}{(m-1)S_2^2} > G(m-1, n-m) \right\} \tag{3}$$

The test results obtained can be divided into four cases, namely highly significant impact, significant impact, certain impact and no significant impact. According to the result, the influence degree of the mobile social network data selection under the change of user preference $A$ is obtained, and a user classification data table is established, as shown in Table 1.

**Table 1.** Classification data table of mobile social users

| High-end business crowd | White-collar user population | Campus user population | Rural user population | Other user groups |
|---|---|---|---|---|
| Financial management | Shopping | Home | Medical treatment | Communication |
| Country | Film and television | Social | Real estate | The internet |
| Finance | Weibo | Digital | Cell phone | Keep in good health |
| Fashion | Game | Real estate | Technology | Music |
| Read | Diet | Apparel | Education | News |

According to the above decomposition results, a detailed user classification strategy is formulated, and the improved svm design classification model is used to classify mobile social network users. In the case where the user's non-linearity is separable, assuming that the selection vectors of the two users are $a$ and $b$, the non-linear function $f$ of the improved svm is used to map the user selection vector into the feature space $T$. Then the Euclidean of the two vectors The distance is:

$$d^T(a, b) = \sqrt{H(*) - 2H(a, b) + H(a, b)} \tag{4}$$

In the formula: $H(*)$ represents the kernel function. Then the center vector $C$ of the feature space is:

$$C_f = \frac{1}{n}\sum_{i=1}^{n} f(a_i) \tag{5}$$

Calculate the class center according to the above formula, and then calculate the distance between the two class centers, the formula is:

$$D = \left|C^+ - C^-\right| \tag{6}$$

In the formula: $C^+$ represents the center of the positive class; $C^-$ represents the center of the negative class [6]. Calculate the distance between the two types of samples and other user sample information. When the distance is less than the calculation result of formula (6), the sample is taken as a valid candidate support vector, that is, there is:

$$D' = |a_i - C| \tag{7}$$

Figure 1 below is a schematic diagram of the sample as a valid candidate support vector when the reservation satisfies $D' < D$.

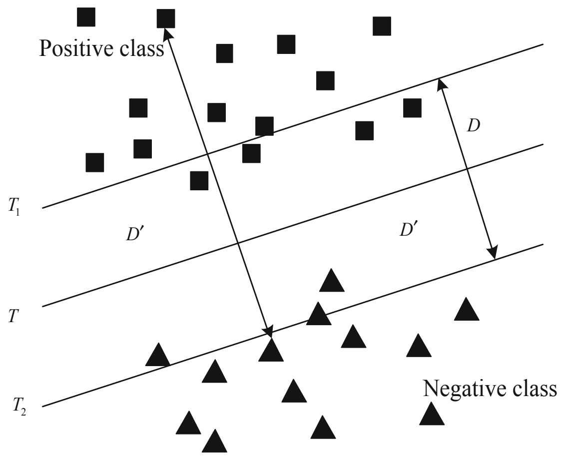



**Fig. 1.** Pre-select valid candidate support vectors

According to the above schematic diagram, the mobile social network users in the feature space are divided according to $D'$ and $D$, so that users with the same selection preference are grouped into a data set, and the classification results shown in Table 1 above are obtained.

The needs of mobile social network users are diverse, so the need elements between different users are classified fuzzy in order to find out the correlation characteristics between needs. Collect raw data for the types of users who use personalized recommendation services. The source of the data is mobile social network data, and the data should be collected according to factors such as different age groups, different genders, and different occupations to ensure the universality and breadth of the data. Assuming that user needs are represented by $q$, there is user demand set $Q = \{q_1, q_2, \ldots, q_n\}$, where $q_n$ represents $n$ different needs of the user; at the same time, there are $w$ characteristic indicators under the influence of different users for each demand, that is, there is $q_c = \{q_{c1}, q_{c2}, \ldots, q_{cn}\}$, so the user is obtained The original data matrix of the demand [7]. According to the original mobile social network data actually collected, the translation standard deviation transformation process is implemented, and the calculated result is:

$$q'_{ci} = \left(q_{ci} - \overline{q}_i\right)/D'B_i \tag{8}$$

In the formula: $q'_{ci}$ represents the standard deviation value of the demand data after translational change; $q_{ci}$ represents the original value of the $i$ original data under the influence of the characteristic index $c$; $\overline{q}_i$ represents the average result of the $i$ original data; $B_i$ represents the transformation matrix; of which $c = \{1, 2, \ldots, n\}$, $i = \{1, 2, \ldots, m\}$. After the data transformation of the above formula, the mean value of each demand variable is 0, and the standard deviation is 1, eliminating the influence of dimensions. It is known that there is a demand set $Q$, from which two demand quantities are randomly selected, denoted as $q_a$ and $q_b$ respectively, and the degree of similarity between the two is calculated as $s_{ab} = f(q_a, q_b)$, the calculation expression is:

$$s_{ab} = \sum_{i=1}^{w} q_{ai} \cdot q_{bi} / \sqrt{\sum_{i=1}^{w} q_{ai}^2} \sqrt{\sum_{i=1}^{w} q_{bi}^2} \tag{9}$$

Use the above formula to find the degree of similarity between the demand data, and realize the fuzzy clustering of user demand through the optimal threshold method:

$$q_i^{(\mu)} = \frac{1}{s_{ab}} \sum_{k=1}^{n_\mu} q_{ki}^{(\mu)} \tag{10}$$

In the formula: $q_i^{(\mu)}$ represents the clustering result of $i$ sample data under the influence of the $\mu$ types of demand; $q_{ki}^{(\mu)}$ represents the dynamic change constant of the demand data under the action of the change factor $k$ [8]. According to the above calculation steps, fuzzy clustering of requirements is realized. The requirements after clustering have different characteristic quantities, so on the basis of demand correlation, by quantitatively analyzing the correlation between demand data, judging the hidden correlation between demands, so as to determine the degree of dispersion of demand, so as to extract different demands Similar characteristics of type data. Assuming that each cluster subset is represented by a feature vector $\lambda$, the category attributes of different data are calculated for $n$ different user needs. When real number type demand data exists, $\lambda_u = [\lambda_{u1}, \lambda_{u2}, \ldots, \lambda_{un}]$, $\lambda_v = [\lambda_{v1}, \lambda_{v2}, \ldots, \lambda_{vn}]$. Then the distance components between different vector subsets are:

$$l_{uv} = \frac{|\lambda_{ui} - \lambda_{vi}|}{\max|\lambda_{ui} - \lambda_{vi}|} \tag{11}$$

In the formula: $l_{uv}$ represents the distance between the two subsets whose subset vector is $u$, $v$; $\lambda_{ui}$, $\lambda_{vi}$. represents the value under the influence of attribute 4. At this time, in order to meet the universality of the data, it is necessary to extract the sample features uniformly. When extracting data features, it is necessary to ensure that the data features are evenly distributed in different spaces, and for both strong feature data and weak feature data, extraction must not be ignored. Then the extraction result of the demand feature quantity is:

$$q' = \sum_{k=1}^{n} \lambda_k / l_{uv} \tag{12}$$

In the formula: $q'$ represents the user demand feature; $\lambda_k$ represents the vector feature value under the control of feature intensity $k$. According to the above formula, the user needs feature extraction is completed.

### 2.2 Mining the Association Rules Between Users and Service Needs

Combining the above classification results and demand feature extraction results, mining association rules between users and service needs. First, describe the entire sample document abstractly, and represent the text in a form that can be quickly processed by the computer. Therefore, all user information is represented by a set $Z$, and there is $Z=\{Z_1, Z_2, \ldots, Z_n\}$, of which $Z_i \in Z_n$ indicates that the $i$ user information is included in the transaction set with $n$ user information. The document is abstracted into a collection of things $X = \{X_1, X_2, \ldots, X_n\}$, where $X_n$ indicates that the collection consists of $n$ documents. The keywords are abstracted into item sets $e$, and $e = \{e_1, e_2, \ldots, e_n\}$ and $e_n$ represent $n$ item sets in the document, which are the keywords of user demand information. The service demand problem is regarded as an abstract collection, and this collection is scanned multiple times and data is mined [9]. Assume that the structure list $I$ is a set of service demand keywords, and $I = \{I_1, I_2, \ldots, I_n\}$ is used to represent each demand information, that is, a sample document. When mining the association rules between requirements and basic information, all feature keywords are collected. Assuming that there are $i$ sample documents in a certain category, and each document has $j$ feature words, then all keywords are collected, totaling $i \times j$ A. But there will be some repetitive keywords, so before mining association rules, remove these repetitive data to get a set of all keywords that are not repetitive [10–12]. The process is shown in Fig. 2 below.

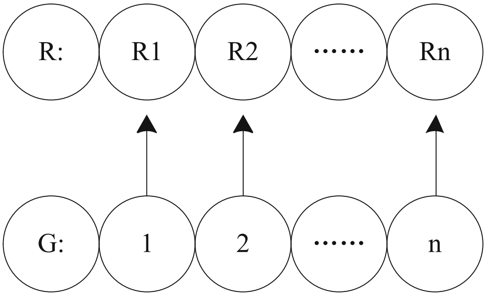



**Fig. 2.** Schematic diagram of deduplication

The 1 in Fig. 2 represents the marking of all keywords. In actual mining, category tags should be added, so the category occupancy should be set to 0 occupancy. The mining restrictions are: Rule: keyword 1, keyword 2,…, keyword n → category name. According to the association rules between users and service needs obtained by the above mining, a personalized recommendation algorithm is designed.

### 2.3 Building a Personalized Recommendation Algorithm

According to the above association rules, a personalized recommendation algorithm based on the user's temporal characteristics is constructed. The algorithm is mainly composed of two parts, which are perceptual data compression and perceptual feature recognition and recommendation. The construction plan of the algorithm is: the algorithm needs to meet the preprocessing network flow, sort it according to the timestamp of the data packet, and construct a time series set. Then the algorithm performs perceptual compression on the data to obtain a significantly robust private data set. The new data set is identified and recommended, and finally a traceability summary is generated according to the hash principle. According to the above construction scheme, the perception data is compressed first. Think of the network flow as a set of data packets with the same 5-tuple, and then divide the network flow with fixed-length time slots, so that the network flow can start from a random point in time to form a sequence set of several time slots, which means The formula is $T = \{t_1, t_2, \ldots, t_n\}$, where $T$ represents the data packet contained in a certain time slot. In order to effectively distinguish each user demand network flow, by extracting burst period data, the unique perceptual characteristics of the network flow are retained [13, 14]. In the time series set divided by time slot, the bursty traffic period represents the time slot corresponding to the peak flow rate, so it is regarded as the new time series data set $T'$. The result is as follows:

$$T' = Dcomp(T) \tag{13}$$

When extracting the time slot according to the above calculation formula and calculating the flow rate of the time slot at the same time, the default time slot $T$ contains $m$ data packets, and the flow rate rate corresponding to the time slot is:

$$\varphi_i = \frac{L_{i,1} + L_{i,2} + \cdots + L_{i,j} + \cdots + L_{i,m}}{D} \tag{14}$$

In the formula: $D$ represents the length of time slot $i$; $L_{i,j}$ represents the length of the $j$ data packet in time slot $i$. The time slot corresponding to $\varphi_i$ is selected from the set $T$, and a brand new time slot set $T' = \{t'_1, t'_2, \ldots, t'_n\}$ is obtained. In order to ensure the robustness of the data set, information entropy is used to further process the set $T'$:

$$T'' = Hcomep(T') \tag{15}$$

Through the above calculation process, the algorithm realizes the perceptual data compression of the mobile social network data. Then use the traceability function to perceive feature extraction and recommendation. According to the research results, it can be known that the data rate reflects the burst characteristics of network traffic, so $\varphi_i$ feature is extracted, and its recommended encoding is set as a perceptible hash sequence to obtain a set containing $\gamma$ time slots. Divide each time slot into smaller time slices to reflect the detailed characteristics of the network flow, and the new time slice matrix after the original time slot set is mapped is:

$$Matrix_\theta = \begin{bmatrix} \theta_{1,1} & \cdots & \theta_{1,\gamma} \\ \vdots & \ddots & \vdots \\ \theta_{\gamma,1} & \cdots & \theta_{\gamma,\gamma} \end{bmatrix} \tag{16}$$

In the formula: $\theta$ is the length of the time slice, which means that there are several data packets in each time slice. According to the calculation result of formula (16), the characteristic matrix is obtained:

$$Matrix_\varphi = \begin{bmatrix} \varphi_{1,1} & \cdots & \varphi_{1,\gamma} \\ \vdots & \ddots & \vdots \\ \theta_{\gamma,1} & \cdots & \theta_{\gamma,\gamma} \end{bmatrix} \tag{17}$$

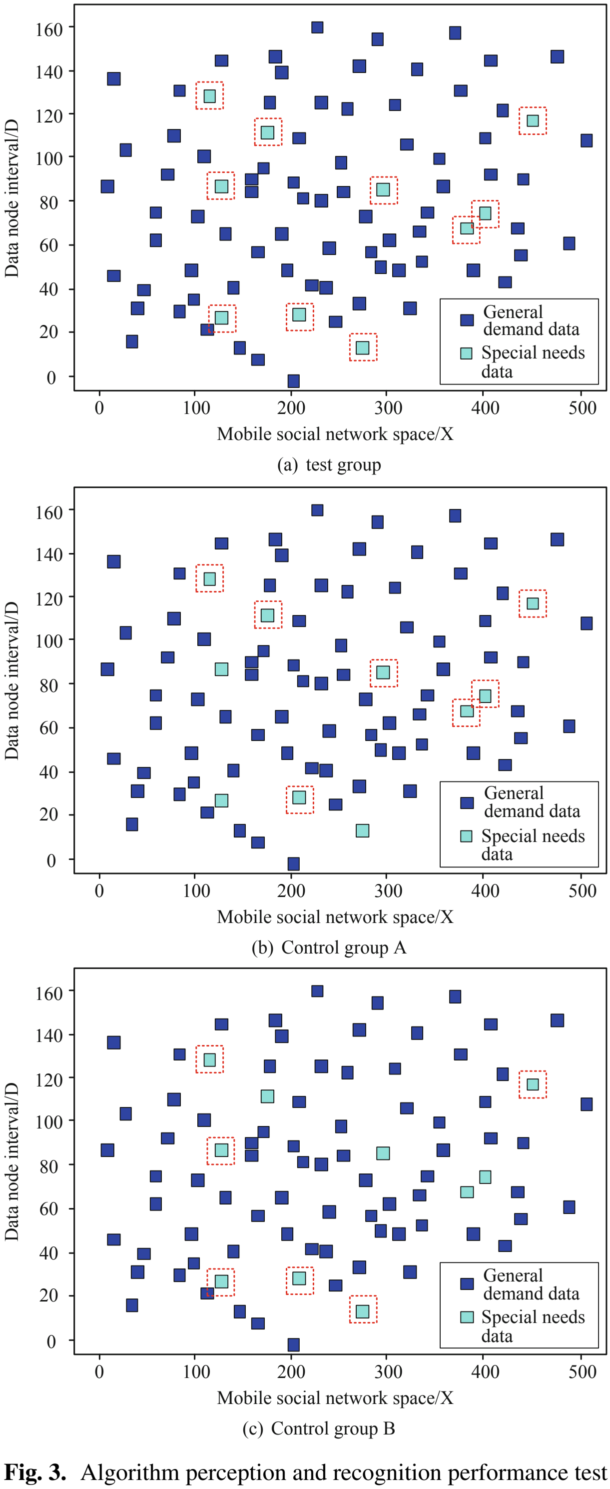
It is known that $\theta$ and $\varphi$ are corresponding. Calculate the average data rate, represented by $\overline{\varphi}$ [15]. Perform an average hash on each column of the data, then compare the value of each data with the average of the set, and encode:

$$K(i) = \begin{cases} 0 \; if \; \varphi_{i,j} \le \overline{\varphi}_i \\ 1 \; if \; \varphi_{i,j} > \overline{\varphi}_i \end{cases} \tag{18}$$

When the above calculation result is less than the average value, the code is 0, otherwise it is 1. So far, the construction of a personalized recommendation algorithm based on mobile social network data has been completed.

## 3 Experimental Research

Apply the designed algorithm and the traditional algorithm to the same system separately, set up a simulated experiment environment, use efficient blockchain users as the test object, map the network flow into summary information, and send it to the test system. From the perspectives of perception recognition and demand recommendation, the performance of different recommendation algorithms is analyzed. In order to facilitate the description of the experimental test results, the algorithm of this study is used as the experimental group, and the traditional algorithm is used as the control group, and a two-stage experimental test task is carried out.




(a) test group

(b) Control group A

(c) Control group B

**Fig. 3.** Algorithm perception and recognition performance test

### 3.1 Perceptual Recognition Performance Test

Design two sets of experiments to test whether the two algorithms can identify the different needs of a large number of users in mobile social networks. The experiment set up a large number of conventional demand data and a type of special demand data. The following Fig. 3 shows the test results of the perception and recognition performance of different recommendation algorithms.

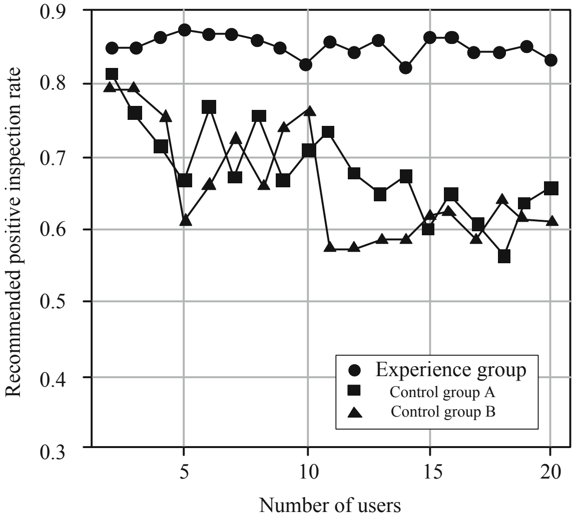
According to the test results in the above figure, it can be seen that under the same test conditions, the experimental group algorithm found all the special demand nodes among the massive network nodes; while the two control group algorithms only found 10 special demand nodes respectively. 7 and 6 of the nodes. The perceptual efficiency of calculating traditional algorithms is 70% and 60%, which are far lower than the perceptual recognition effect of the algorithm in the article. It can be seen that the recommendation algorithm of this research has a better perception and recognition effect when facing special demand nodes hidden in massive data. The method in this paper performs perceptual data compression on mobile social network data. Then use the traceability function to perceive feature extraction and recommendation. The data rate reflects the burst characteristics of network traffic, thereby extracting the characteristics, and realizing the personalized recommendation corresponding to the characteristics.

### 3.2 Recommended Effect Test

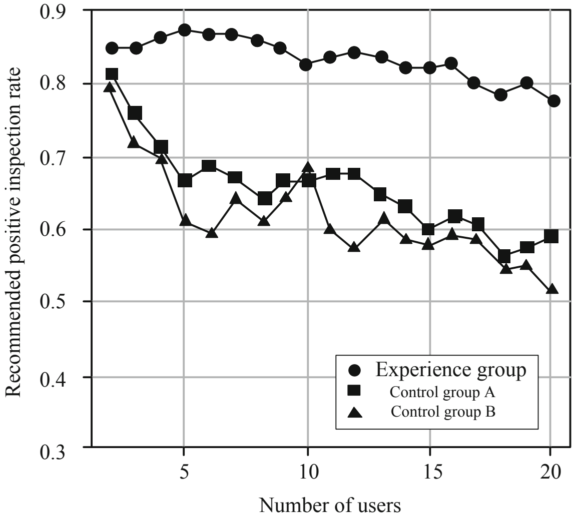
Set up a multi-stream mixed model, and perform multi-stream recommendation tests on the experimental group and the control group. The model contains multiple network flows, and the demand nodes among them increase with the increase of projects. The experiment sets up four different addition schemes, establishes four different test environments, compares the three groups of different algorithms when facing different user needs, and their ability to recommend information according to the needs. The test results are shown in Fig. 4 below.

According to the test results shown in Fig. 4, it can be seen that in the face of jitter environment, packet loss environment, packet injection environment, and multiple interference environment, the experimental group algorithm has better recommendation ability. The algorithm recommendation ability of the other two control groups is significantly weaker than the experimental group due to the influence of different test environments. It can be seen that the algorithm constructed this time can be used to solve network user service work.

Combined with the above experimental results, it can be seen that the personalized recommendation algorithm based on mobile social network data proposed in this paper has more accurate data perception and recognition effects, and can achieve more accurate personalized recommendations.

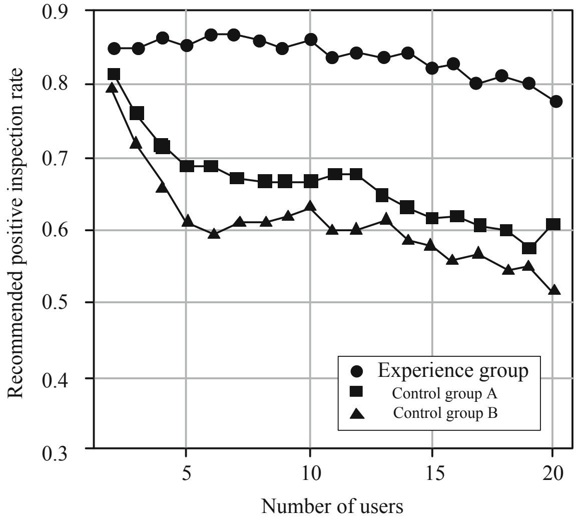



(a) Recommended test in jitter environment

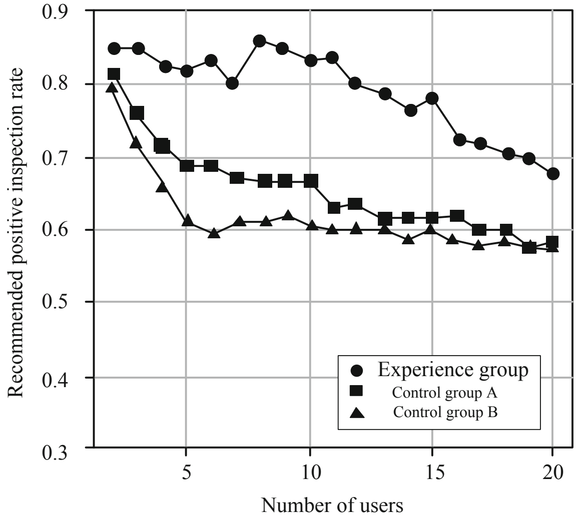



(b) Recommended test in packet loss environment

**Fig. 4.** Multi-stream hybrid traceability performance test of the algorithm




(c) Recommended testing in package injection environment




(d) Recommended test under multiple interference

**Fig. 4.** continued

## 4 Concluding Remarks

This research optimizes the personalized recommendation algorithm based on traditional algorithms and combined with mobile social network data, and has achieved relatively satisfactory research results. However, considering the design process of the entire algorithm, there are a large number of calculation formulas in this design, so calculation errors are prone to occur during calculations. At the same time, the calculation efficiency of this algorithm may be weaker than other algorithms. Therefore, future research should focus on calculation errors and efficiency. To provide users with more reliable network technology.

## References


1. Shi, Y., Zeng, Z.: Temporal segment networks based on feature propagation for action recognition. J. Comput.-Aided Des. Comput. Graph. **32**(04), 582–589 (2020)
2. Chen, S., Huang, L., Lei, Z., et al.: Research on personalized recommendation hybrid algorithm for interactive experience equipment. Comput. Intell. **36**(3), 1348–1373 (2020)
3. Yt, A., Bing, Z.A., Yw, A., et al.: College library personalized recommendation system based on hybrid recommendation algorithm. Procedia CIRP **83**, 490–494 (2019)
4. Li, J., Hong, T., Wang, X., et al.: A data dissemination mechanism based on group structure in opportunistic mobile social networks. J. Comput. Res. Dev. **56**(11), 2494–2505 (2019)
5. Cao, H., Ma, Z., Yang, X., et al.: Pilot allocation based on users' location and user classification in massive MIMO system. Telecommun. Sci. **35**(10), 92–99 (2019)
6. Xie, G.-R., Zheng, H., Lin, W., et al.: Power outage sensitive customers classification based on improved random forest algorithm. Comput. Syst. Appl. **28**(03), 104–110 (2019)
7. Liu, C., Huang, J., Zhou, T.: Hierarchical characteristics and driving force of user demand for paid digital reading based on grounded theory. Modern Inf. **39**(11), 80–89 (2019)
8. Sun, Y., Yang, G., He, L.: Prediction of advertising conversion rate based on store characteristics and user needs. J. Qingdao Univ. (Eng. Technol. Edition) **34**(03), 16–20 (2019)
9. Wang, X., Zhu, J., Meng, X., et al.: A model of safety monitoring and early warning for coal mine based on ontology and association rules. Mining Saf. Environ. Prot. **46**(03), 27–31 (2019)
10. Li, P., Wang, Y.: Simulation study on the periodic association of book circulation data association rules. Comput. Simul. **37**(05), 354–357+452 (2020)
11. Zhao, C., Zhang, K., Liang, J.: Asymmetric recommendation algorithm in heterogeneous information network. J. Front. Comput. Sci. Technol. **14**(06), 939–946 (2020)
12. Zhang, Q., Wang, B., Cui, N., et al.: Attention-based regularized matrix factorization for recommendation. J. Softw. **31**(03), 778–793 (2020)
13. Liu, S., Liu, G., Zhou, H.: A robust parallel object tracking method for illumination variations. Mob. Networks Appl. **24**(1), 5–17 (2018). https://doi.org/10.1007/s11036-018-1134-8
14. Liu, S., Bai, W., Liu, G., et al.: Parallel fractal compression method for big video data. Complexity **2018**, 2016976 (2018)
15. Liu, S., He, T., Dai, J.: A survey of CRF algorithm based knowledge extraction of elementary mathematics in Chinese. Mob. Networks Appl. (2021). https://doi.org/10.1007/s11036-020-01725-x