A Recommender System for Virtual Cultural Heritage Tourism: Matrix Factorization and Collaborative Filtering Approach

Xi Zhao*, Guangyun Yu
School of Art, Zhejiang Shuren University, Hangzhou, 310000, China
E-mail: ceecee123456@163.com
*Corresponding author

Keywords: matrix factorization, coordinated filtering, virtual reality, recommend, user behavior, classification

Received: February 8, 2025

In the digital age, digital collection and recording technology can handle various types of tangible and intangible cultural heritage. Virtual tourism technology for cultural heritage has great potential in providing users with personalized experiences, but it also faces the problem of ignoring the personalized needs of different users. To this end, a user behavior classification model for cultural heritage virtual tourism technology and a cultural heritage virtual tourism recommendation model based on matrix factorization and coordinated filtering were developed. In the classification task, this study used Virtual Reality scene action data collected from HTC VIVO devices. In the recommendation task, MovieLens, Amazon-charts, ciao, and Epinions datasets were used. The findings denoted that the accuracy of the raised user behavior classification model was 85.47%, 94.62%, and 80.17% in the controller, head mounted display, and button data, respectively. In the mixed data source, the classification accuracy of the proposed model was 98.42%, and the F1 value was 97.74%. The Recall@20 of virtual tour recommendation model in MovieLens and Amazon-charts Dataset were 72.36% and 72.84%, respectively, with diversity values ranging from 0.7 to 0.9. On the Ciao dataset and Epinions dataset, the Root Mean Squared Error and Mean Absolute Error of the proposed model were 0.937 and 0.701, 1.033 and 0.796, respectively. The experimental results demonstrated that the proposed model improved classification and recommendation performance by innovatively combining additive attention mechanism, contextual multi-arm slot machine algorithm, and deep analysis of user behavior, surpassing standard matrix factorization and collaborative filtering methods. The research results help improve the display and service quality of cultural heritage virtual exhibition halls, effectively protect and inherit intangible cultural heritage, and promote the digital development of cultural resources.

Povzetek: Nova metoda združi LSTM+pozornost za razvrščanje VR vedenja in MF+CF s kontekstnim banditom za osebna priporočila virtualnega kulturnega turizma, z večjo raznolikostjo, obvladovanjem hladnega zagona in sprotnim posodabljanjem uporabnikov.

1 Introduction

As the technology rapidly develops, the usage of artificial intelligence technology in cultural dissemination has greatly enriched the forms of expression and dissemination effects of cultural products. It not only enriches the forms of expression of cultural products, but also takes a critical part in the protection and inheritance of traditional culture, breaking through the constraints of time and space and expanding the breadth of cultural dissemination [1]. Cultural Heritage Virtual Tourism (CHVT) technology refers to the usage of Virtual Reality (VR) technology or Augmented Reality (AR) technology to simulate the environment and atmosphere of real tourist attractions or overlay virtual information in tourist attractions, providing users with immersive and interactive tourism experiences [2-3]. Melo et al. investigated the impact of multi sensory VR settings on tourist satisfaction to further explore the potential application of VR technology in tourism marketing. The results indicated that VR technology could effectively improve tourists' satisfaction and positive emotions, thereby promoting their consumption behavior towards tourism products and services. Wu et al. conducted a survey and analysis of 320 samples to explore how 360-degree VR technology can stimulate tourists' willingness to truly hike on the mountain, which can provide inspiration for the development of Virtual Tourism (VT) technology [4]. The results indicated that 360-degree VR technology had a positive impact on tourists' emotional participation, presence, and enjoyment, which in turn motivates users to engage in mountain climbing tourism [5]. Yang et al. proposed to identify the motivation of virtual tourists based on the method purpose chain theory in response to the current immature research on the motivation of VT, and conducted step-by-step interviews with the respondents. The results indicated that the motivation for VT largely depends on whether tourists were attracted by the technological features, safety, and experiential conditions of VT [6]. Zeng et al. indicated that VR is currently being employed extensively within the tourism industry. To investigate further how VR may encourage tourists to engage in cultural dissemination activities, a moderated mediation model was developed. The results indicated that VR experienced could stimulate tourists' cultural dissemination behavior [7]. Bretos et al. conducted a critical review of literature on VR and AR technologies in the tourism industry to explore their role in the tourism industry. The results indicated that the usage of VR and AR in the tourism industry received widespread attention from academia and could help improve tourists' travel experience [8]. Cham et al. addressed the issue of low adoption rate of VR facilities in tourist attractions among elderly tourists. Based on a cross-sectional method, they collected data from elderly tourists through a survey questionnaire and continued to work on data cleaning and statistical analysis. The results indicated that technological and psychological barriers were the main reasons affecting the adoption rate of VR facilities in tourist attractions by elderly tourists [9].

In CHVT, Collaborative Filtering (CF) technology can recommend relevant Cultural Heritage (CH) content based on user preferences and behavior patterns, which not only enhances user participation and learning interest, but also effectively spreads and popularizes CH knowledge [10]. The Matrix Factorization (MF) model is a model-based CF that models the features of users and items by multiplying the user item rating MF into two small matrices, thereby enabling recommendation [11]. Papadakis et al. reviewed the research on CF recommendation systems for the prediction of user preferences in Internet recommendation systems, classified each method, and compared different CF recommendation systems [12]. Fkih tested the performance of different similarity measures through experimental research to address the sensitivity of CF technology in quantifying the strength of dependence between two users. The results showed that in the user-based CF recommendation system, the Mean Absolute Error (MAE) of Improved Triangle Similarity (ITR) on the MovieLens100k standard dataset was 0.786 [13]. Widayanti et al. proposed a more personalized recommendation paradigm that integrates CF and content-based filtering techniques. The outcomes indicated that the raised method could generate recommendations with enhanced diversity and accuracy, effectively solving the “cold start” problem of a single CF method and the problem of poor recommendation diversity in content-based filtering techniques [14]. Anwar et al. proposed memory-based CF method and model-based CF method to generate similarity matrix and prediction matrix to solve the issues in CF recommendation system. The results indicated that the proposed method could effectively solve problems of CF recommendation systems, and provide more recommended items [15]. D'Amico et al. found that different random initialization could result in the same MF technique generating different recommendation lists. Therefore, they proposed a nearest neighbor MF method, which learns the embeddings of each user and item as weighted linear combinations of their respective nearest neighbor representations. The outcomes denoted that the raisedmethod improved the stability and accuracy of recommendations [16]. Sankari et al. pointed out that MF was a key technology in recommendation systems, explored three MF techniques, and conducted experimental evaluations on real-world datasets. The experiment findings validated the recommendation performance of MF technology [17]. The summary table of the above related work is shown in Table 1.

Table 1: Summary table of related work

Study	Data set	Index	Key findings	Limitation
Papadakis et al. [12]	No specific dataset available	Completeness of classification system	Propose a multidimensional classification framework for CF recommendation system	Unverified emerging deep learning methods
Fkih [13]	MovieLens 100k	MAE	ITR similarity performs the best in user-based CF, with an MAE of 0.786, which is 9.2% higher than cosine similarity	Only test a single dataset
Widayanti et al. [14]	MovieLens 1M	Accuracy, diversity, and coverage	Mixing CF and content filtering increases the coverage of cold start scenarios by 37%, with a diversity index of 0.68	Feature engineering relies on domain knowledge
Anwar et al. [15]	Amazon Electronics	Root mean squared error (RMSE), F1 value	CF method integrating kNN improves F1 value by 22% among cold start users	High computational complexity
D'Amico et al. [16]	Netflix Prize	Stability Index and RMSE	The nearest neighbor MF will improve the recommended stability by 43%	Sacrificing some accuracy
Sankari et al. [17]	Yelp, BookCrossing	MAE, normalized discounted cumulative gain	Singular value decomposition performs the best in BookCrossing (MAE=0.72)	Unresolved long tail distribution problem

In summary, although the value of VT technology in the dissemination of CH has been recognized and affirmed, the cultural and tourism industry needs to constantly adapt and adopt new technologies to further improve user experience. To this end, a user behavior classification model for CHVT technology and a CHVT recommendation model based on MF-CF are proposed. This study aims to accurately analyze user behavior and dynamic preferences by combining MF and CF technologies, thereby improving the diversity of VT recommendations, enhancing user experience, and ensuring the dissemination effect of CH.

The innovation of this study includes: (1) Combining Long Short-Term Memory Networks (LSTM) and Additive Attention (AM) mechanisms for user behavior classification, which can effectively integrate operational data from different VR devices. (2) Combining MF and CF technologies to build a VT recommendation model for CH, which can provide more accurate and diverse recommendation results.

The main contributions of this study include: (1) Providing personalized CH virtual tour recommendation services through the combination of user behavior classification models and recommendation models. (2) By introducing user behavior analysis and AM mechanism, the problems of cold start and data sparsity in recommendation systems have been effectively alleviated, enhancing the reliability and stability of recommendation results. (3) Promoted the popularization and development of VT technology for CH, providing new methods and tools for the digitization and dissemination of cultural resources.

2 Methods and materials

To enhance the user experience and effectiveness of CHVT, a user behavior classification model for CHVT technology is developed. Based on the analysis of user behavior and the dynamic preferences of different users, a CHVT recommendation model based on MF-CF is built.

2.1 User behavior classification model for virtual tourism technology of CH

Intangible CH constitutes an essential element of China's exemplary traditional culture, serving as a tangible testament to the uninterrupted transmission of Chinese civilization. CH encompasses the collective legacy bequeathed to humanity by history, which can be categorized into two principal forms: tangible CH and intangible CH. The protection and inheritance of CH are of paramount importance for the maintenance of human cultural diversity and historical continuity. The development and application of VR technology can provide new solutions for the protection and dissemination of CH [18-19]. VT technology for CH has been widely applied in recent years, which can digitally model, reproduce, and display CH, allowing the public to experience the charm of historical culture through VR devices. Famous attractions in multiple countries, such as the Imperial Palace in Beijing, Stonehenge in the UK, the Twin Towers in Malaysia, and Notre Dame Cathedral in France, have provided virtual tours and historical and cultural experiences to visitors through VR technology. This not only strengthens the visitor experience but also makes the protection and inheritance of CH more effective. However, the current VT technology for CH mostly remains at the display stage in practice, neglecting the analysis of user behavior, resulting in the inability to provide personalized navigation and interactive experiences [20-22]. By analyzing users' behavior in VR environments, their needs and preferences can be better understood, thereby optimizing the design and functionality of VR systems. The schematic diagram of user behavior classification in VR environment is shown in Figure 1.

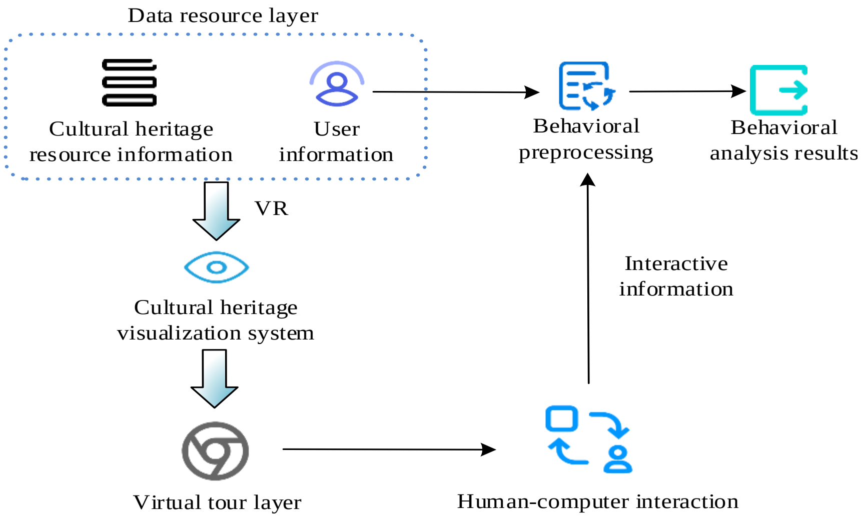


Figure 1: Schematic diagram of user behavior classification in VR environment

In Figure 1, the data resource layer contains CH resource information and user information. VR technology is used to visualize the CH resource information, presenting an immersive virtual travel environment for users. Users conduct virtual tours in the CH visualization system and interact with the virtual environment through VR devices. The user data and interaction data are preprocessed and analyzed to get the results of user behavior classification. For this purpose, the study proposes a user behavior classification model for CHVT technology. In the VR environment, user operation data is a type of time series data, so the research adopts LSTM, which has good application effects in processing time series data, to build a user behavior classification model. LSTM is a specific type of Recurrent Neural Network (RNN) that incorporates specialized storage units and gating mechanisms to more effectively capture and process long-term dependencies in sequential data [23-25]. This is done with the aim of addressing the problems of gradient vanishing and finding that are commonly encountered by conventional RNNs when processing long sequence data. An LSTM comprises three gating mechanisms, namely an Input Gate (IG), a Forget Gate (FG) and an Output Gate (OG). The FG is responsible for determining whether information should be retained or discarded. Its state is represented by the value given in formula (1).

$$f_t = \sigma\left(W_f h_{t-1} + V_f x_t + b_f\right) \qquad (1)$$

In equation (1), σ indicates the sigmoid activation function, W_f indicates the weight matrix of the FG, h_{t-1} means the input at the previous time, V denotes the weight matrix of h_{t-1}, x_t indicates the inputting value at the current time, and b represents the bias term. The IG is utilized to control whether the input information is updated into the cell unit, as shown in formula (2).

$$\begin{cases} i_t = \sigma(W_i h_{t-1} + V_i x_i + b_i) \\ \bar{c}_t = \tanh(W_c h_{t-1} + V_c x_t + b_c) \\ c_t = f_t \odot c_{t-1} + i_t \odot \bar{c}_t \end{cases} \qquad (2)$$

In equation (2), i_t refers to the state of the IG, W_i means the weight matrix of the IG, tanh denotes the tanh activation function, c_t represents the candidate cell, W_c represents the weight matrix of the vector cell, and $\bar{c}_t$ denotes the latest state of the memory cell node. The OG is used to control the output of memory unit state values, as shown in formula (3).

$$\begin{cases} o_t = \sigma(W_o h_{t-1} + V_o x_t + b_o) \\ h_t = o_t \odot \tanh(c_t) \end{cases} \qquad (3)$$

In equation (3), o_t denotes the state of the OG, W_o refers to the weight matrix of the OG, and h_t means the output at the current time. VR devices have multiple input sources, and this study uses multiple independent LSTM models to process operational data from different VR devices, such as controllers, head mounted displays, and buttons. AM has a strong ability to enhance feature fusion. By calculating attention weights through an additional nonlinear function, it can provide more flexible attention allocation methods in sequence processing tasks and adaptively adjust the calculation of attention weights. However, traditional self attention mechanisms may face high computational complexity when dealing with long sequences, and transformers are more suitable for processing large-scale sequence data. Therefore, to integrate feature vectors from different input sources, the study introduces an AM mechanism to aggregate LSTM output information from different VR devices. The concept of attention originates from the study of human vision and is concerned with the simulation of the selective attention abilities of humans when processing information. This approach allows models to focus on the most relevant aspects of the current task, thereby enhancing processing efficiency and accuracy. The calculation of the AM is indicated in Figure 2.

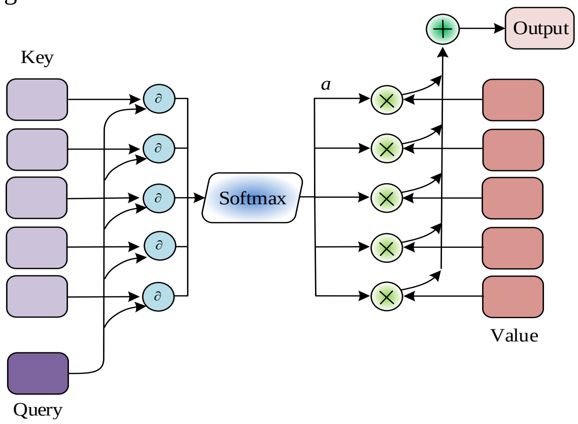


Figure 2: The calculation process of AM

In Figure 2, the AM first receives two inputs: a query and a key. Secondly, based on the query and key information, a score matrix is obtained to represent the similarity or matching degree between the query and each key. Then, it normalizes the score matrix using the Softmax function to obtain the attention weights corresponding to each key. Finally, the value is multiplied with the corresponding attention weight, and then all values are weighted and summed to obtain the final output. AM is utilized in deep learning to strengthen the model's ability to capture key information. The core idea is to calculate attention weights by combining query and key information through an additive function, and then the values are weighted and summed based on these weights. It assumes that there is a query $q \in R^q$ and the key value pairs are $(k_1, v_1), \ldots (k_m, v_m)$, where m means the amount of key value pairs, the expression of the attention function f is denoted in formula (4).

$$f\left[q, (k_1, v_1), \ldots, (k_m, v_m)\right] = \sum_{i=1}^{m} a(q, k_i) v_i \qquad (4)$$

In equation (4), a represents the attention weight, indicating the degree to which the model values different parts when processing sequence data. a is usually calculated using the Softmax function, as shown in formula (5).

$$a(q, k_i) = \text{softmax}(\partial(q, k_i)) = \frac{\exp(\partial(q, k_i))}{\sum_{j=1}^{n} \exp(\partial(q, k_j))} \qquad (5)$$

In equation (5), ∂ represents the attention rating function. The AM calculates attention weights through an additional nonlinear function, providing a more flexible way of attention allocation in sequence processing tasks. In the AM, the expression of the attention rating function ∂ is denoted in formula (6).

$$\partial(q,k) = W_v^T \tanh(W_q q + W_k k) \qquad (6)$$

In equation (6), W_v, W_q, and W_k represent learnable parameters, and T represents transposition operation. Adding the results of q and the key yields a vector of length h, which is then multiplied by the tanh activation function and weight matrix to obtain a. For the case where q and key length are consistent, the study adopts scaled dot product attention. Scaling dot product attention is mainly used for modeling sequential data in deep learning. When the batch size is n, the calculation of scaled dot product attention is shown in formula (7).

$$\partial\left[q_1,\cdots q_n,(k_1,v_1),\cdots,(k_m,v_m)\right] = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \qquad (7)$$

In equation (7), Q refers to the matrix corresponding to n queries, K means key, V denotes value, and d denotes the length of q and K. In summary, the structure of the user behavior classification model for CHVT technology proposed by the research is shown in Figure 3.

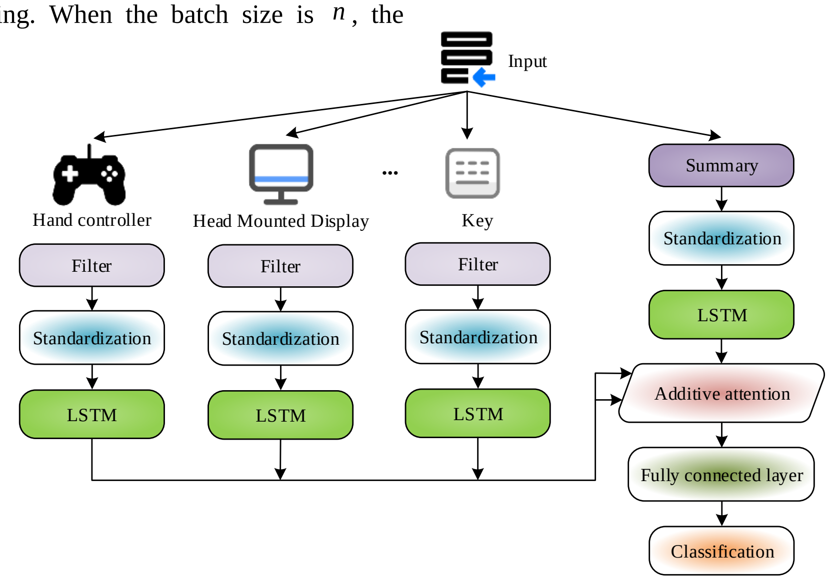
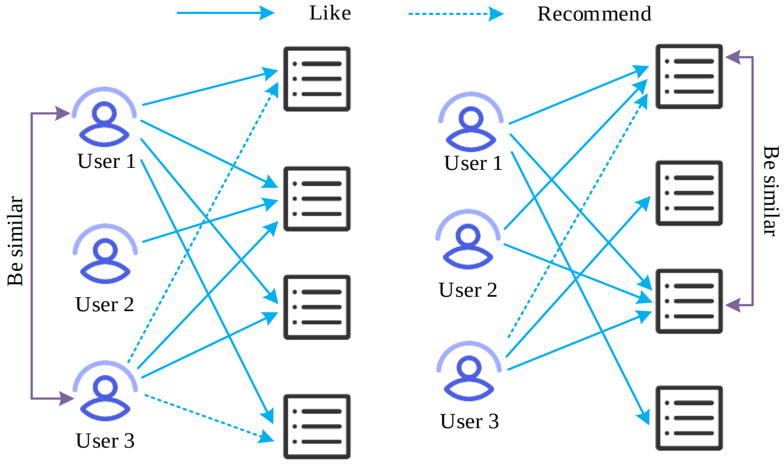


Figure 3: Structure diagram of user behavior classification model

In Figure 3, the user behavior classification model proposed by the research first takes the data generated by users operating VR devices as input. Then, the data is filtered and standardized, and independent LSTM models are used to process the operation data from different VR device sources. At the same time, new sequences are obtained by summarizing data from different sources and integrating them with operational data from different VR devices through AM. Finally, the final user behavior classification is achieved through a fully connected layer.

2.2 A virtual tourism recommendation model for CH based on MF-CF

The user behavior classification model developed through research can classify the behavior of users when participating in CH virtual tours, thereby helping virtual tour platforms better understand users' needs and preferences. However, it can only capture users' explicit behavior and cannot directly provide personalized content recommendations based on users' behavior and preferences. Therefore, based on analyzing user behavior, in order to obtain a more personalized and autonomous VT experience of CH, accurate classification services can be further provided to users. The CF recommendation model is currently one of the mainstream recommendation models, which can be broken into user-based CF, project-based CF, and model-based CF. The schematic diagram of user-based CF and project-based CF is denoted in Figure 4.



(a) User-based collaborative filtering (b) Project-based collaborative filtering

Figure 4: Schematic diagram of CF

From Figure 4, project-based CF can capture the similarity between items by analyzing users' ratings or behaviors towards the project and identifying other items that are similar to the project. The method is based on user collaboration filtering, whereby the target user's preferences are compared with those of other users with similar profiles. Items that these users have previously indicated a preference for are then recommended to the target user. This approach allows for the capture of dynamic preference changes over time. Therefore, to capture users' dynamic preferences and better understand their needs and expectations during the virtual tour of CH, the research chooses to build a virtual tour recommendation model based on user CF. It is of great significance to perform multi-information representation before clustering similar users. In the stage of multi-information representation, firstly, to obtain the feature vector matrices of users and items, MF is utilized to decompose the user item rating matrix, and mean square error is used for iteration until convergence. The loss function is denoted in formula (8).

$$L=\sum_{(u,i)}(r_{ui}-y_{ui})^2+\lambda(\|I\|_F^2+\|Z\|_F^2) \qquad (8)$$

In equation (8), r_{iz} refers to the true rating of user u on project i, y_{iz} means the predicted rating of user z on project i, λ indicates the hyperparameter, $\|I\|_F^2$ represents the norm square of the user matrix, and $\|Z\|_F^2$ represents the norm square of the project matrix. Secondly, to strengthen the representation ability, the original features of the input user and project are used to obtain the embedded features of the user and project, which are then connected to the user and project feature vector matrix obtained by MF to obtain the embedded vectors of the user and project. Finally, the obtained embedding vector is expanded into several dimensions, and the importance of each weight is determined using the Softmax function, which is then added and averaged to obtain the final project features. After obtaining the user's feature vector, the Pearson correlation coefficient is utilized to calculate the user similarity, as denoted in formula (9).

$$sim(u,v)=\frac{\sum_{i\in I_{uv}}\left((r_{ui}-\bar{r}_u)(r_{vi}-\bar{r}_v)\right)}{\sqrt{\sum_{i\in I_{uv}}(r_{ui}-\bar{r}_u)^2\sum_{i\in I_{uv}}(r_{vi}-\bar{r}_v)^2}} \qquad (9)$$

In equation (9), $\bar{r}_u$ means the average rating of user u, r_{vi} expresses the rating of user v on item i, $\bar{r}_v$ expresses the average rating of user v, and I_{uv} denotes the set of items rated jointly by users u and v. Users with high similarity are divided into clusters for personalized recommendations, and users are re-divided based on feedback after each recommendation to capture their constantly changing preferences in a timely manner. To comprehensively update the embedded information of users and projects, the gradient descent algorithm is applied in the study. The calculation of the comprehensive feature vector for user updates is shown in formula (10).

$$P'=\bar{X}\frac{Z}{\theta}+P \qquad (10)$$

In equation (10), $\bar{X}$ means the average feature vector of the recommended item, Z represents the total reward value of the recommended item, θ indicates the hyperparameter, and P means the user's comprehensive feature vector. To better classify projects, a filtering mechanism is established, and the schematic diagram of the filtering mechanism is denoted in Figure 5.

In Figure 5, the filtering mechanism proposed by the research first analyzes the scoring records in the initial set of projects to obtain a set of recommended high scoring projects. Secondly, other low scoring items in the initial project set are shuffled to obtain a candidate project set. Finally, based on the user's interests, high rated project categories are used as connection points to find projects in the same category. In addition, to better capture users' dynamically changing preferences, the study combines CF with reinforcement learning, models the recommendation process as a Markov decision process, and uses the Contextual Multi-Armed Bandit (CMAB) algorithm to select the largest arm for output.

The core concepts in reinforcement learning include agency, state, action, and reward. In reinforcement learning modeling, the main objective is to guide agents to make choices that maximize their benefits [26]. The CMAB algorithm is an algorithm that introduces contextual information based on the problem of multi arm slot machines. In the CMAB algorithm, the reward of each arm depends not only on its own state, but also on an additional contextual information, which can help the algorithm better understand the behavior of each arm and improve the accuracy of reward prediction [27-28]. The schematic diagram of the CMAB algorithm is shown in Figure 6.

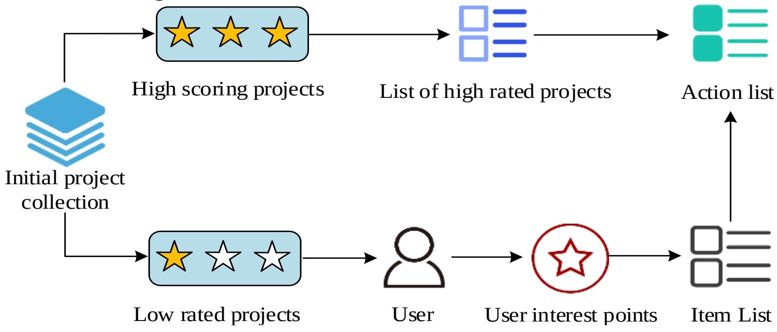


Figure 5: Schematic diagram of filtering mechanism

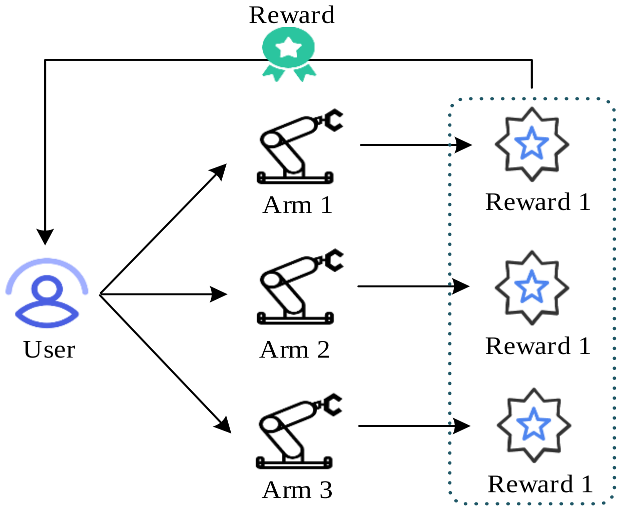


Figure 6: Schematic diagram of multi-arm slot machine algorithm

In Figure 6, in the CMAB algorithm, the user's goal is to maximize the long-term accumulated rewards through a series of choices. Drawing on this idea, the study models the project as an arm in a slot machine. In the recommendation, the project is divided into multiple groups using the filtering mechanism shown in Figure 5, with each group being an arm and containing CH projects of the same category. When recommending, CMAB calculates the expected reward value for each arm based on the current user status and selects the arm with the highest reward value for recommendation. Research uses the agents to traverse each item and action, obtains recommended values for all arms, and selects the arm with the highest recommended value as the recommended item list. The calculation of recommended values is shown in formula (11).

$$i_j = \arg\max\left[\sigma^2(x_j) + \beta\sqrt{x_j^{'T} W_n^{-1} x_j^{'} \log(1+t)}\right] \quad (11)$$

In equation (11), i_j expresses the recommended value of the j th project group, β represents hyperparameters, $\sigma^2(x_j)$ indicates the variance of the j th project group, $x_j^{'}$ indicates the average feature vector of the j th project group, t represents reinforcement learning epochs, and W_n^{-1} represents the inverse matrix of the feature matrix of the n th user group corresponding to the target user. The state is a description of the agent's current situation in the environment. At any given moment, the state of a specific user's u is $S_u = (a_1, a_2, a_3, \ldots)$, and the state is a set of actions. The agent selects the action with the highest profit and updates the user's feature vector based on the reward after each recommendation. An action is an operation performed by an agent in the environment, which determines how the agent interacts with the environment in its state. The candidate action for user u is $a_u = (c_1, c_2, \ldots, c_{t,a})$, where $c_{t,a}$ represents the feature vector obtained at time t after the user takes action a_u and pulls their arm. When making decisions, agents will facilitate the feature vectors of each item to obtain recommended values for recommendation output. Reward is the feedback provided by the environment to the agent. In the study, the total reward for each project is recorded as Z, and user interests are obtained in real-time based on the reward when updating user characteristics [29]. The loss function L_2 of reinforcement learning is shown in formula (12).

$$L_2 = \max(i - b\log L) \qquad (12)$$

In equation (12), i represents the output of reinforcement learning, b denotes the loss weight, and L denotes the output of the MF loss function. In summary, the framework of the MF-CF-based VT recommendation model for CH proposed by the research is shown in Figure 7.

In Figure 7, the MF-CF-based VT recommendation model for CH proposed in this study first uses the MF algorithm to obtain feature vectors of users and projects. Secondly, clustering similar users based on feature vectors allows users to communicate within the community. Then, the projects are divided into different groups for agents to choose from through filtering mechanisms, with each project group serving as an arm. Finally, the agent recommends based on the user's preferences, selects the arm with the highest expectation, and sets a reward mechanism according to the user's feedback to update the user's feature vector, to achieve real-time personalized recommendation.

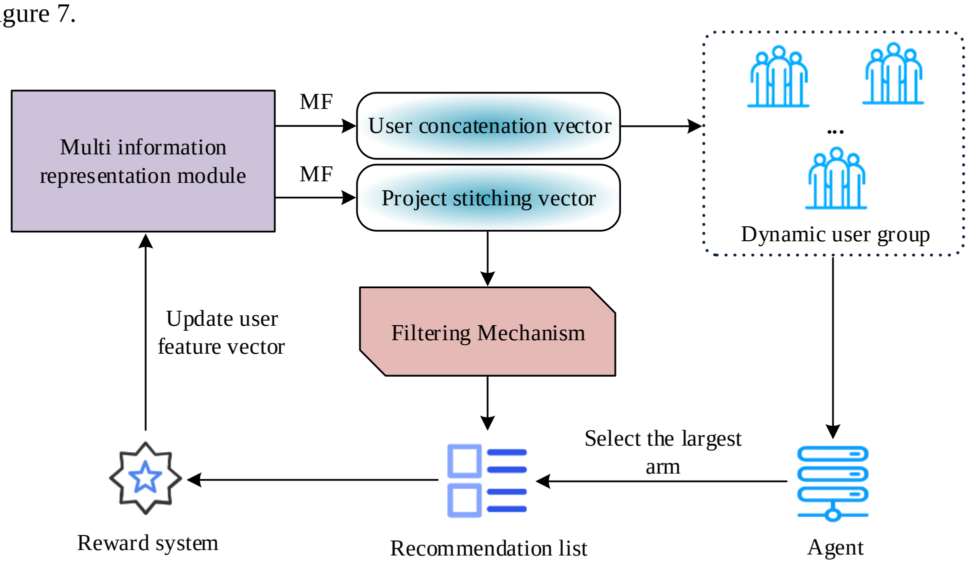


Figure 7: Framework diagram of VT recommendation model for CH based on MF-CF

3 Results

The research developed a user behavior classification model for CHVT technology and a CHVT recommendation model based on MF-CF, but their performance still needs further validation. To this end, the study first analyzed the performance of user behavior classification models, and then explored the feasibility of a CH virtual tour recommendation model grounded on MF-CF.

3.1 Performance analysis of user behavior classification model

To assess the effectiveness of the user behavior classification model for CHVT technology, the Pytorch framework was used to build the model, and HTC VIVO was used for operation. User data on controllers, head mounted displays, and button operations were collected in three different VR scenes, with 200 action data for each scene. The collected data were preprocessed, duplicate action records were checked and removed, data collected by different VR devices were standardized to the same timestamped format, and missing values were filled in with averages. Subsequently, the sparsity of the user item interaction matrix was calculated. For users with few interactions, they were sampled in the training phase and their proportion in the training data was appropriately reduced to increase data density. The study used 10-fold cross validation to divide the dataset into training and testing sets in a 9:1 ratio. This process was repeated 10 times, with different subsets selected as the testing set each time. Finally, the performance indicators of the 10 test results were averaged as the final evaluation result of the model. The iteration count was set to 250, the LSTM hidden layer dimension was set to 128, the key/query dimension was set to 64, and Dropout was set to 0.2, using Adam optimizer and cross entropy loss function. To verify the impact of different hyperparameters on model performance, the learning rates were set to 0.01, 0.001, and 0.0001, respectively, and BatchSize was set to 4, 8, 16, and 32. The results of hyperparameter sensitivity analysis are shown in Table 2. From Table 2, excessively high learning rates can lead to oscillatory convergence, while excessively low learning rates can result in slow convergence. When the learning rate was 0.001 and BatchSize was 8, the accuracy and F1 value of the model were the highest, at 92.43% and

91.71%, respectively.

Table 2: Hyperparameter sensitivity analysis results

Learning rate	BatchSize	Accuracy/%	F1/%
0.01	8	86.14	84.86
0.001	8	92.43	91.71
0.0001	8	88.25	87.06
0.001	4	91.82	91.09
0.001	16	90.35	89.41
0.001	32	87.77	87.63

Using a single data source, the Classification Accuracy (CA) of the proposed model was compared with traditional RNN, LSTM, and GRU, and the outcomes are denoted in Figure 8. From Figure 8(a), in the handle dataset, the CA of the raised model was higher than that of traditional RNN, LSTM, and GRU models, reaching 85.47%. Next was the GRU model, with an accuracy of 62.48%. From Figure 8(b), the CA of the proposed model was still the highest in the head mounted display dataset, at 94.62%. From Figure 8(c), the CA of the proposed model in the key data set was 80.17%. The results indicated that the user behavior classification model proposed by the research for CHVT technology showed good accuracy in user behavior classification in different VR device data sources, and had certain feasibility and effectiveness.

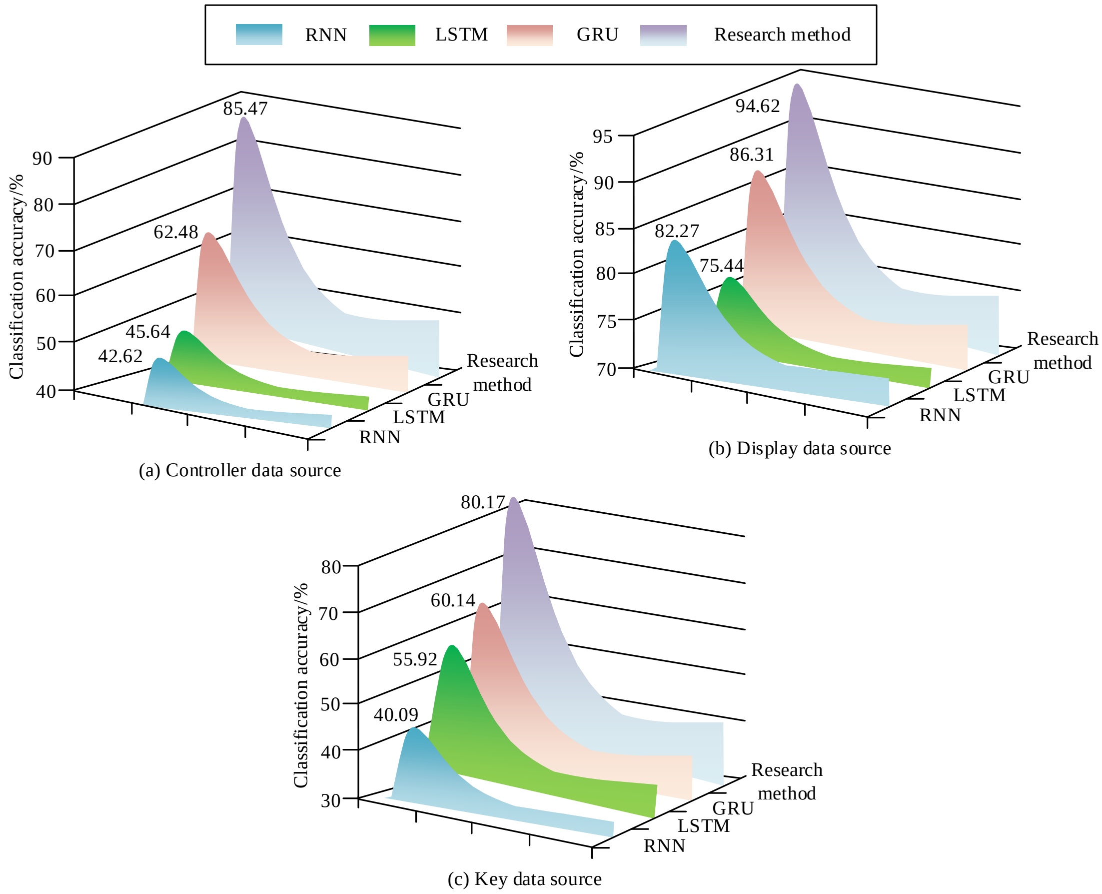


Figure 8: Comparison of CA of four methods under single source data

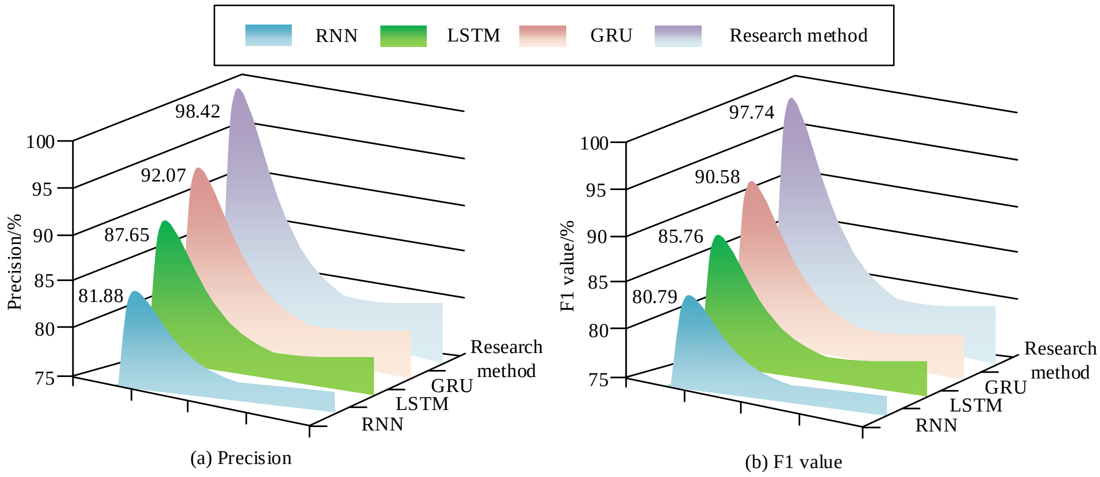


Figure 9: Comparison of CA and F1 value of four models

Using datasets from all data sources, the CA and F1 value of the above four models were compared, and the outcomes are denoted in Figure 9. From Figure 9(a), the CA of the proposed model was the highest, at 98.42%. Next was the GRU algorithm, with an accuracy of 92.097%. The recommendation accuracy of the RNN model was the lowest, at 81.88%. From Figure 9(b), the F1 value of the proposed model was still the highest, at 97.74%. In addition, comparing Figure 8 and Figure 9, using mixed data sources could effectively improve the user behavior classification performance of different models. Higher accuracy and precision mean the ability to effectively classify user behavior, thereby better understanding user needs and preferences, and providing reliable basis for recommendation systems.

To assess the user behavior classification performance of the proposed model in different VR scenarios, the CA of the proposed model was compared with traditional Convolutional Neural Networks (CNN), Convolutional 3D (C3D), and TimeSformer models. The findings are denoted in Figure 10. From Figure 10(a), in scenario 1, compared to the other three models, the CA of the proposed model was the highest, at 97.72%. The second was the TimeSformer model, and the CNN model had the lowest CA. From Figure 10(b), in scenario 2, the CA of the proposed model was the highest, at 94.86%. From Figure 10(c), in scenario 3, the CA of the proposed model was 92.47%. The outcomes indicate that the proposed model has good user behavior classification performance in different VR scenarios and has certain practical application value.

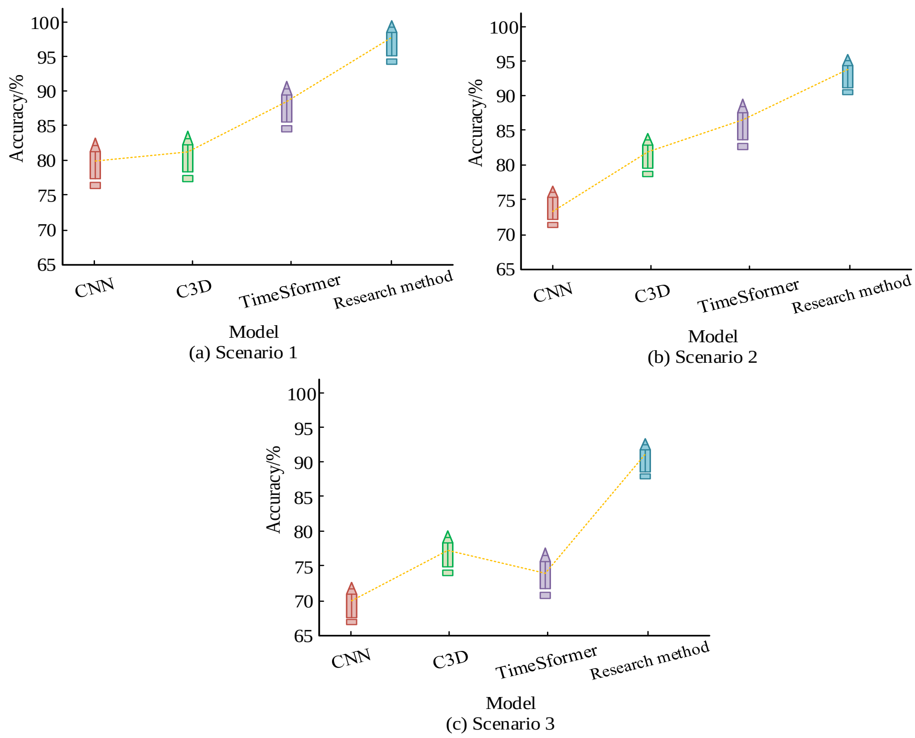


Figure 10: Comparison of CA of different models in three VR scenarios

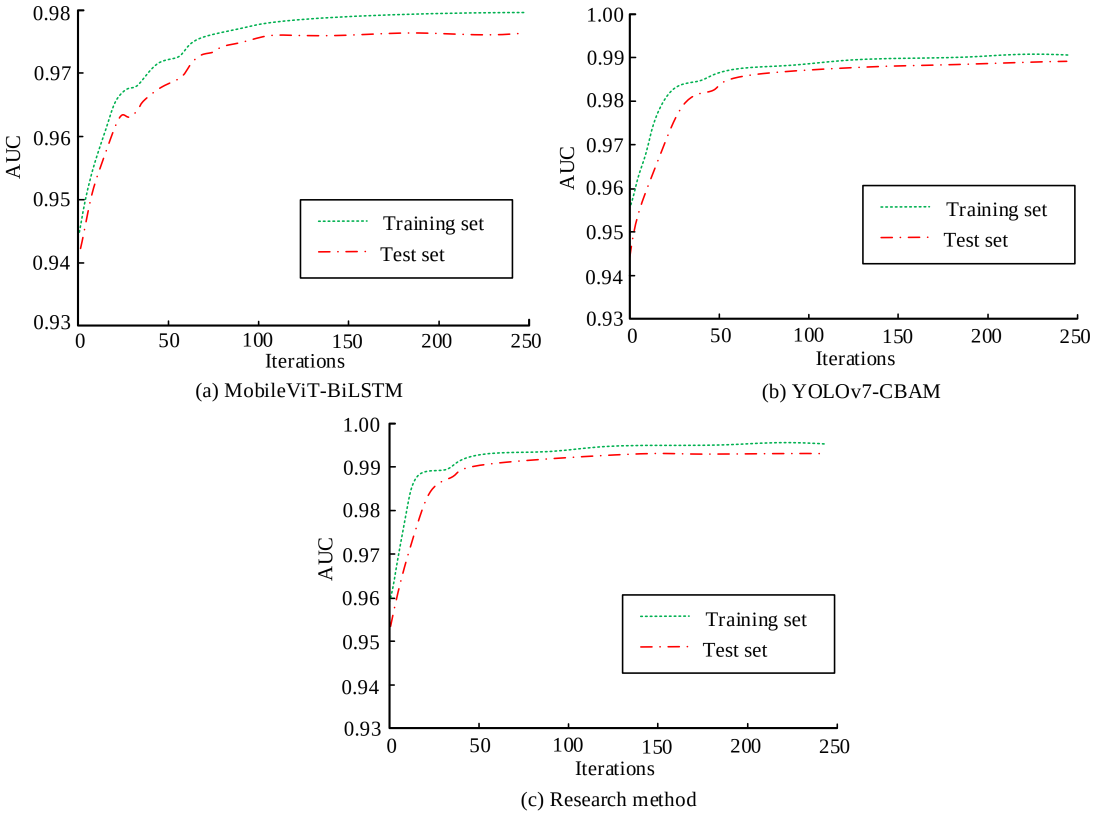


Figure 11: Comparison of classification performance of four models

To further validate the superiority and generalization of the proposed model, the Area Under the Curve (AUC) was compared with the currently advanced classification models that integrate MobileViT and bidirectional LSTM (MobileViT-BiLSTM) and the YOLOv7-CBAM model. The outcomes are denoted in Figure 11. Comparing Figures 11 (a), (b), and (c), compared to the MobileViT-BiLSTM and YOLOv7-CBAM models, the proposed model performed better in AUC metrics and converges faster, demonstrating certain superiority.

3.2 Feasibility analysis of virtual tour recommendation model

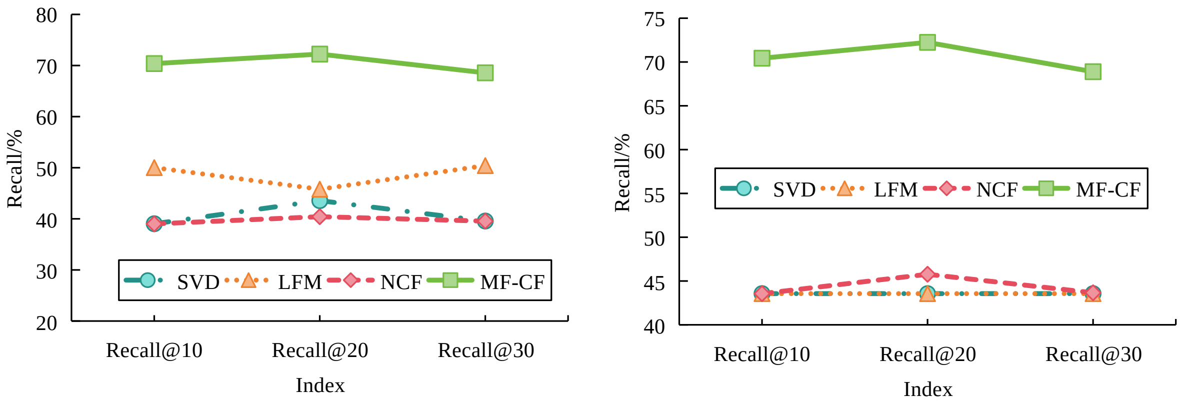
To prove the feasibility of the proposed MF-CF-based VT recommendation model for CH, the MovieLens and Amazon-charts datasets were used for testing and divided into training, testing, and validation sets in an 8:1:1 ratio. The MovieLens dataset contains rich user behavior data and a clear rating system, which can effectively simulate users' rating behavior towards CHVT projects. In addition, the rating behavior of users towards movies is similar to that towards virtual tours of CH, both involving users' interests and subjective evaluations of the content. The Amazon-charts dataset contains rating data for goods and services in related categories such as culture, art, and tourism. This study mainly used data from tourism products and cultural experience products. BatchSize was set to 128, top-k was set to 10-30, learning rate was set to 0.001, Dropout was set to 0.2, Embedding dimension was set to 64, regularization coefficient was set to 0.01, and SGD optimizer was used. The project-based CF (Item CF), multi-layer perceptron (MLP), neural collaborative filtering (NCF), and graph neural network-based (GNN) recommendation models were compared with the proposed MF-CF model as baseline models. The comparison results of the recommendation performance of the 5 models are shown in Table 3. From Table 3, the accuracy and recall@20 of the proposed model were superior to the other four baseline models, accounting for 94.77% and 72.36%, respectively. But the training time of the proposed model was relatively high, at 35.26 minutes, only lower than the recommendation model based on GNN. The results indicated that the proposed MF-CF model not only improved recommendation performance, but also increased the complexity of the model, resulting in an increase in training time.

Table 3: Comparison of recommendation performance among 5 models

Models	Accuracy/%	Recall@20/%	Training time/min
Item-CF	66.84	68.19	12.49
MLP	72.25	70.12	25.31
NCF	74.09	71.27	30.78
GNN	75.88	71.86	40.67
MF-CF	94.77	72.36	35.26

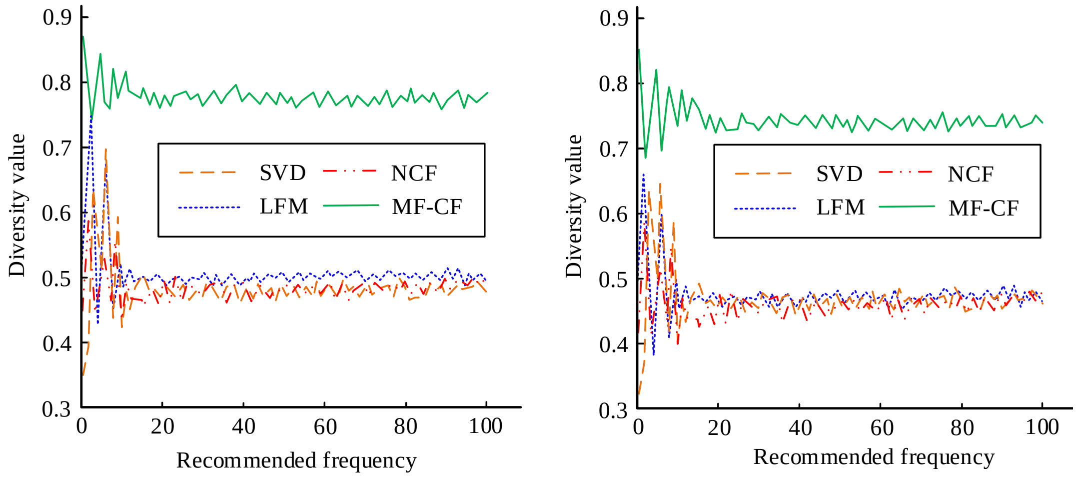
The recall rate measures the proportion of CH virtual tour projects that users are truly interested in, and a higher recall rate means that the model can display more CH projects that users may be interested in to them.

Therefore, the study adopted recall rate as the evaluation index. The recall rate of the proposed model was compared with traditional Singular Value Decomposition (SVD) algorithm, Latent Factor Model (LFM), and NCF, and the findings are denoted in Figure 12. From Figure 12(a), in the MovieLens dataset, compared with the other three models, the proposed model performed the best in terms of recall rate, its Recall@20 was 72.36%. From Figure 12(b), in the Amazon-charts dataset, the proposed model still performed the best in terms of recall rate, and its Recall@20 was 72.84%. Recall@20 measures the proportion of the top 20 recommended items that the model successfully recommends that the user is truly interested in. In the VT scene of CH, higher Recall@20 value indicates that the recommendation system can effectively explore users' points of interest and display more CH projects that users may be interested in, thereby improving the user experience. The results indicate that the MF-CF-based VT recommendation model for CH has good recommendation performance, and has certain feasibility and effectiveness.



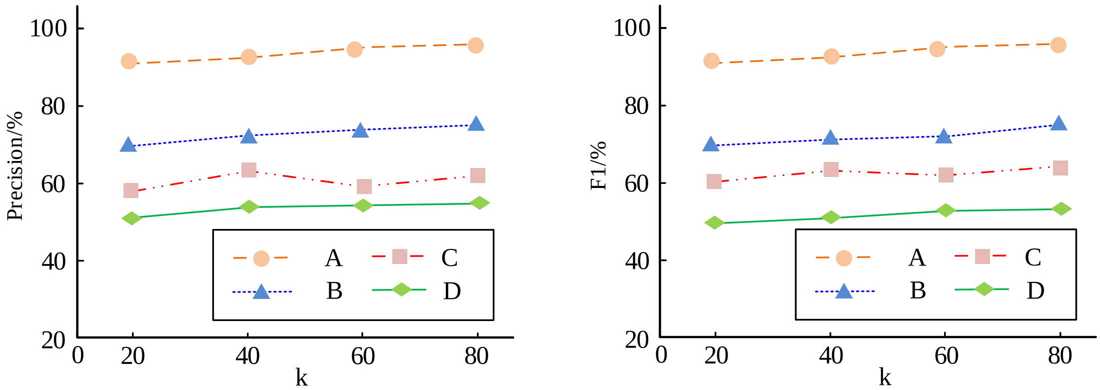
(a) MovieLens dataset (b) Amazon-charts dataset

Figure 12: Comparison of recall rates among four models



(a) MovieLens dataset (b) Amazon-charts dataset

Figure 13: Comparison of recommendation diversity among 4 models



(a) Precision (b) F1 value

Figure 14: Results of ablation experiment

Diversity assessment evaluates the degree of diversity of items in the recommended list. In VT of CH, diversity indicators can ensure that the recommendation system does not overly focus on a few types of projects, thereby bringing users a richer experience. Comparing the recommended diversity of the four models mentioned above, the results are denoted in Figure 13. From Figures 13 (a) and (b), the recommendation diversity of the proposed model was superior to SVD algorithm, LFM algorithm, and NCF algorithm on both datasets, with diversity values ranging from 0.7 to 0.9. The outcomes indicate that the MF-CF-based VT recommendation model for CH raised by the research has good diversity in recommendations.

To verify the contribution of each strategy module (filtering mechanism, user feature update, and similar user segmentation) in the proposed CH virtual tour recommendation model to the overall recommendation performance of the model, and to demonstrate the effectiveness and necessity of these strategies, this study conducted ablation experiments. Using recommendation precision and F1 value as evaluation metrics, the complete model (A) that includes all strategies, the model (B) that removes filtering mechanisms based on the complete model, the model (C) that removes user feature updates, and the model (D) that removes similar user partitions were compared. The results of the ablation experiment are shown in Figure 14. From Figure 14 (a), the recommendation accuracy of the complete model was the highest, at 94.57%. The second was the model without filtering mechanism, and the model without similar user segmentation had the lowest recommendation accuracy. From Figure 14 (b), the F1 value of the complete model was the highest, at 92.66%. The results indicate that the proposed strategy can effectively raise the recommendation effectiveness of the model, with the contribution of the similar user segmentation module being the greatest.

Due to the fact that the MovieLens and Amazon-charts datasets mainly focus on movie and music recommendations, and have relatively few users and projects, the sparsity of the datasets is relatively low. Therefore, to comprehensively evaluate the performance and generalization ability of the proposed model, the Ciao and Epinions datasets with larger data scales and higher sparsity were used for testing. The Ciao dataset is a dataset of the entire DVD category for UK websites in December 2013, extracted from DVD. ciao.co.uk. It includes user ratings of the items they have purchased and social connections between users. The Epinions dataset is a website where users can comment on products, including user ratings of products and social information between users. Root Mean Square Error (RMSE) and MAE measure the degree of difference between the model's predicted rating and the user's actual rating. Lower RMSE and MAE values mean that the model can more accurately predict the user's rating, thereby better meeting the user's personalized needs and ensuring the effectiveness and credibility of the recommendation system. Therefore, this study used RMSE and MAE as evaluation indicators. The proposed model was compared with the recommendation models based on Attention-based RNN (ARNN), Van Dat [30], Wu [31], and Chen [10]. The comparison outcomes of the recommendation efficacy of the five models are denoted in Table 4. From Table 4, compared to the other four recommendation models, the MF-CF model proposed by the study had lower RMSE and MAE indicators on both datasets. On the Ciao dataset, both values of the raised model were 0.937 and 0.701. On the Epinions dataset, both values of the raised model were 1.033 and 0.796. RMSE was used to evaluate the difference between predicted ratings and actual user ratings, while MAE reflected the mean absolute error between predicted ratings and actual ratings. In the CH virtual tour recommendation model, lower RMSE and MAE values mean that the model's prediction of user ratings is more accurate, which helps to better meet users' personalized needs. The results indicate that the MF-CF-based VT recommendation model for CH proposed by the research shows low recommendation errors in different datasets, demonstrating certain superiority and generalization.

Table 4: Comparison of recommendation performance among five models

Models	Ciao		Epinions	
	RMSE	MAE	RMSE	MAE
ARNN	1.217	0.946	1.335	1.008
Van Dat et al	1.065	0.841	1.144	0.896
Wu et al	0.977	0.740	1.064	0.820
Chen et al	0.974	0.732	1.057	0.805
MF-CF	0.937	0.701	1.033	0.796

In summary, this study used smaller datasets (MovieLens and Amazon-charts) and larger datasets (Ciao and Epinions) for testing to evaluate the impact of dataset size on model performance. A larger dataset provides more user and project information, enabling the model to capture more complex user behavior patterns and project features. However, smaller datasets may not provide enough information to capture the complex relationships between users and items, leading to limitations in the model's recommendation diversity. However, the proposed MF-CF recommendation model still exhibits good performance, indicating that it can work effectively on smaller datasets.

4 Discussion

To further improve the user experience and cultural dissemination effect of CHVT technology, a user behavior classification model for CHVT technology and a CHVT recommendation model based on MF-CF were studied and constructed. The results showed that in the controller dataset, the CA of the proposed user behavior classification model was higher than that of traditional RNN, LSTM, and GRU models, reaching 85.47%. Next was the GRU model, with an accuracy of 62.48%. In the dataset of head mounted displays, the CA of the proposed model was still the highest, at 94.62%. In the key data set, the CA of the proposed model was 80.17%. In the mixed data source, the CA of the raised model was the highest, at 98.42%, and the F1 value was still the highest, at 97.74%. It also has good user behavior classification performance in different VR scenarios. The Recall@20 of virtual tour recommendation model in MovieLens and Amazon-charts Dataset were 72.36% and 72.84%, respectively. The recommendation diversity of the proposed model was superior to SVD, LFM, and NCF algorithms on both datasets, with diversity values ranging from 0.7 to 0.9. The recommendation accuracy of the complete model was the highest, at 94.57%, and the F1 value was the highest, at 92.66%. The second was the model without filtering mechanism, and the model without similar user segmentation had the lowest recommendation accuracy. The MF-CF model had lower RMSE and MAE indicators on both datasets. On the Ciao dataset, the RMSE and MAE values of the proposed model were 0.937 and 0.701. On the Epinions dataset, both values of the proposed model were 1.033 and 0.796.

Compared with the improved triangle similarity method proposed by Fkih et al. [13], this research model dynamically adjusted user clustering through reinforcement learning, solving the problem of traditional CF being sensitive to similarity measurement. The user behavior classification model proposed in this study dynamically weighted the LSTM outputs of different VR devices through an AM mechanism, solving the problem of classification bias from a single data source, and achieving an F1 value of 97.74% under mixed data, which was 5.6% higher than GRU. Compared to the graph CF proposed by Chen et al. [10], which only focuses on popular nodes, this model extracted real-time operation sequences through LSTM, supplementing the shortcomings of static social graphs. On the Epinions dataset, Recall@20 improved by 7.2%. In practical deployment, the proposed model supports mainstream VR devices and can reduce users' on-site travel costs through CH virtual tour technology. It can also provide personalized recommendations based on user behavior and preferences, and has certain practical application value and prospects. However, although the proposed model performs well on experimental datasets, differences in cultural backgrounds, user group characteristics, and device differences may limit its generalization performance in practical applications. The interpretability of the model is also a challenge, stemming from the black box nature of deep feature interaction and the lack of transparent presentation of recommendation logic. In addition, there is a trade-off between computational cost and accuracy, and the proposed improvement strategy significantly increases the complexity of the model while improving recommendation accuracy, which may hinder deployment in resource constrained scenarios. To further improve recommendation performance, it is necessary to apply a distillation pruning compression model to balance efficiency and accuracy.

5 Conclusion

In summary, the model proposed by the research can effectively analyze user behavior in virtual scenes and make rich and accurate recommendations. However, the recommendation model proposed in the study not only improves recommendation performance, but also increases the complexity of the model to a certain extent, leading to an increase in the demand for computing resources. This may limit the application of the model in resource constrained environments. Therefore, in future research, more lightweight models should be further explored, such as pruning to remove unimportant connections or neurons in the model, or training a smaller student model through model distillation to mimic the output of larger teacher models, to preserve the main model performance, improve the model's operational efficiency and deployment flexibility.

References

[1] N. Raimo, I. De Turi, A. Ricciardelli and F. Vitolla, “Digitalization in the cultural industry: evidence from Italian museums,” in Int. J. Entrep. Behav. Res., vol. 28, no. 8, pp. 1962-1974, November 2022, DOI: 10.1108/IJEBR-01-2021-0082.

[2] D. DeWitt, S. F. Chan and R. Loban, “Virtual reality for developing intercultural communication competence in Mandarin as a Foreign language,” in Educ. Technol. Res. Dev., vol. 70, no. 2, pp. 615-638, April 2022, DOI: 10.1007/s11423-021-10074-9.

[3] J. Li, “Impact of Metaverse cultural communication on the mental health of international students in China: highlighting effects of healthcare anxiety and cyberchondria,” in Am. J. Health Behav., vol. 46, no. 6, pp. 809-820, December 2022, DOI: 10.5993/AJHB.46.6.21.

[4] M. Melo, H. Coelho, G. Gonçalves, N. Losada, F. Jorge, M. S. Teixeira and M. Bessa, “Immersive multisensory virtual reality technologies for virtual tourism: A study of the user’s sense of presence, satisfaction, emotions, and attitudes,” in Multimed. Syst., vol. 28, no. 3, pp. 1027-1037, June 2022, DOI: 10.1007/s00530-022-00898-7.

[5] X. Wu and I. K. W. Lai, “The use of 360-degree virtual tours to promote mountain walking tourism: Stimulus–organism–response model,” in Inf. Technol. Tour., vol. 24, no. 1, pp. 85-107, March 2022, DOI: 10.1007/s40558-021-00218-1.

[6] T. Yang, W. Ruan, Y. Li and S. Zhang, “Virtual tourist motivation: the differences between virtual tourism and on-site tourism,” in Tour. Rev., vol. 78, no. 5, pp. 1280-1297, September 2023, DOI: 10.1108/TR-07-2022-0331.

[7] Y. Zeng, L. Liu and R. Xu, “The effects of a virtual reality tourism experience on tourist’s cultural dissemination behavior,” in Tour. Hosp., vol. 3, no. 1, pp. 314-329, March 2022, DOI: 10.3390/tourhosp3010021.

[8] M. A. Bretos, S. Ibáñez-Sánchez and C. Orús, “Applying virtual reality and augmented reality to

the tourism experience: a comparative literature review," in Span. J. Mark. -ESIC, vol. 28, no. 3, pp. 287-309, July 2024, DOI: 10.1108/SJME-03-2023-0052.

[9] T. H. Cham, G. Wei-Han Tan, E. C. X. Aw, K. B. Ooi, T. W. Jee and C. K. Pek, "Virtual reality in tourism: adoption scepticism and resistance," in Tour. Rev., vol. 79, no. 2, pp. 337-354, February 2024, DOI: 10.1108/TR-10-2022-0479.

[10] J. Chen, X. Xin, X. Liang, X. He and J. Liu, "GDSRec: Graph-Based Decentralized Collaborative Filtering for Social Recommendation," in IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 5, pp. 4813-4824, May 2023, DOI: 10.1109/TKDE.2022.3153284.

[11] V. R. Yannam, J. Kumar, K. S. Babu and B. Sahoo, "Improving group recommendation using deep collaborative filtering approach," in Int. J. Inf. Technol., vol. 15, no. 3, pp. 1489-1497, March 2023, DOI: 10.1007/s41870-023-01205-x.

[12] H. Papadakis, A. Papagrigoriou, C. Panagiotakis, E. Kosmas and P. Fragopoulou, "Collaborative filtering recommender systems taxonomy," in Knowl. Inf. Syst., vol. 64, no. 1, pp. 35-74, January 2022, DOI: 10.1007/s10115-021-01628-7.

[13] F. Fkih, "Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison," in J. King Saud Univ.-Comput. Inf. Sci., vol. 34, no. 9, pp. 7645-7669, October 2022, DOI: 10.1016/j.jksuci.2021.09.014.

[14] R. Widayanti, M. H. R. Chakim, C. Lukita, U. Rahardja and N. Lutfiani, "Improving recommender systems using hybrid techniques of collaborative filtering and content-based filtering," in J. Appl. Data Sci., vol. 4, no. 3, pp. 289-302, 2023, DOI: 10.47738/jads.v4i3.115.

[15] T. Anwar, V. Uma, M. I. Hussain and M. Pantula, "Collaborative filtering and kNN based recommendation to overcome cold start and sparsity issues: A comparative analysis," in Multimed. Tools Appl., vol. 81, no. 25, pp. 35693-35711, October 2022, DOI: 10.1007/s11042-021-11883-z.

[16] E. D'Amico, G. Gabbolini, C. Bernardis and P. Cremonesi, "Analyzing and improving stability of matrix factorization for recommender systems," in J. Intell. Inf. Syst., vol. 58, no. 2, pp. 255-285, April 2022, DOI: 10.1007/s10844-021-00686-1.

[17] A. Sankari, S. Masih and M. Ingle, "Exploring Matrix Decomposition Methods for Recommender Systems," in J. Sci. Res., vol. 16, no. 3, pp. 705-712, September 2024, DOI: 10.3329/jsr. v16i3.70831.

[18] X. Xu, "Multimedia VR image improvement and simulation analysis based on visual VR restructuring algorithm," Informatica, vol. 48, no. 1, pp. 107-118, 2024, DOI: 10.31449/inf. v48i1.5368.

[19] T. Yu, C. Lin, S. Zhang, C. Wang, X. Ding, H. An, X. Liu, T. Qu, L. Wan, S. You, J. Wu and J. Zhang, "Artificial intelligence for Dunhuang cultural heritage protection: the project and the dataset," in Int. J. Comput. Vis., vol. 130, no. 11, pp. 2646-2673, January 2022, DOI: 10.1007/s11263-022-01665-x.

[20] S. N. Amin, P. Shivakumara, T. X. Jun, Chong Yang, Zan D. Leong Lon and R. Rahavendra, "An Augmented Reality-Based Approach for Designing Interactive Food Menu of Restaurant Using Android," in Artif. Intell. Appl., vol. 1, no. 1, pp. 26-34, November 2023, DOI: 10.47852/bonviewAIA2202354.

[21] Y. Zhang, Y. Gao, Z. Zhao, "Research on Operation and Anomaly Detection of Smart Power Grid Based on Information Technology Using CNN+ Bidirectional LSTM," Informatica, vol, 49, no. 7, pp. 157-164, 2025, DOI: 10.31449/inf. v49i7.7037.

[22] A. S. Gaafar, J. M. Dahr, A, K. Hamoud, "Comparative analysis of performance of deep learning classification approach based on LSTM-RNN for textual and image datasets," Informatica, vol, 46, no. 5, pp. 21-28, 2022, DOI: 10.31449/inf. v46i5.3872.

[23] D. Soydaner, "Attention mechanism in neural networks: where it comes and where it goes," in Neural Comput. Appl., vol. 34, no. 16, pp. 13371-13385, August 2022, DOI: 10.1007/s00521-022-07366-3.

[24] Z. Chen, L. Zhang, J. Sun, R. Meng, S. Yin and Q. Zhao, "DCAMCP: A deep learning model based on capsule network and attention mechanism for molecular carcinogenicity prediction," in J. Cell. Mol. Med., vol. 27, no. 20, pp. 3117-3126, July 2023, DOI: 10.1111/jcmm.17889.

[25] C. Su, Z.S. Zhu. Research on non-legacy creative design based on shadow play elements. Advanced Management Science, 2023, 12(1).

[26] L. Brunke, M. Greeff, A. W. Hall, Z. Yuan, S. Zhou, J. Panerati and A. P. Schoellig, "Safe learning in robotics: From learning-based control to safe reinforcement learning," in Annu. Rev. Control, Robot. Auton. Syst., vol. 5, no. 1, pp. 411-444, May 2022, DOI: 10.1146/annurev-control-042920-020211.

[27] C. Wu, Y. Zhu, R. Zhang, Y. Chen, F. Wang and S. Cui, "FedAB: Truthful Federated Learning with Auction-Based Combinatorial Multi-Armed Bandit," in IEEE Internet Things J., vol. 10, no. 17, pp. 15159-15170, 1 September 2023, DOI: 10.1109/JIOT.2023.3264677.

[28] T. Yang, S. Gao, J. Li, M. Qin, X. Sun, R. Zhang, M. Wang and X. Li, "Multi-Armed Bandits Learning for Task Offloading in Maritime Edge Intelligence Networks," in IEEE Trans. Veh. Technol., vol. 71, no. 4, pp. 4212-4224, April 2022, DOI: 10.1109/TVT.2022.3141740.

[29] F. O. Isinkaye, "Matrix factorization in recommender systems: algorithms, applications, and peculiar challenges," in IETE J. Res., vol. 69, no. 9, pp. 6087-6100, 2023.

[30] N. Van Dat, P. Van Toan and T. M. Thanh, "Solving distribution problems in content-based recommendation system with gaussian mixture model," in Appl. Intell., vol. 52, no. 2, pp.

1602-1614, January 2022, DOI: 10.1007/s10489-021-02429-9.

[31] L. Wu, J. Li, P. Sun, R. Hong, Y. Ge and M. Wang, "DiffNet++: A Neural Influence and Interest Diffusion Network for Social Recommendation," in IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 10, pp. 4753-4766, 1 October. 2022, DOI: 10.1109/TKDE.2020.3048414.