International Journal of Web Services Research
Volume 20 • Issue 1

User Cold Start Recommendation System Based on Hofstede Cultural Theory

Yunfei Li, Jianghuai College, Anhui University, China*
Shichao Yin, Anhui University, China

ABSTRACT

The main function of recommendation systems is to help users select satisfactory services from many services. Existing recommendation systems usually need to conduct a questionnaire survey of the user or obtain the user's third-party information in the case of cold start users; this operation often infringes on the user's privacy. This article is aimed at providing accurate recommendations for cold start users without infringement on user privacy. Therefore, in response to this problem, this manuscript per the authors proposes a recommendation algorithm based on Hofstede's cultural dimensions theory. The algorithm uses Hofstede's cultural dimensions theory to establish a connection between two cold start users, thus ensuring the stability of QoS prediction accuracy. Then, the prediction results and the dynamic combination of the matrix factorization algorithm are used to obtain a more accurate prediction. The verification results on the real dataset WS-Dream show that the prediction algorithm proposed in this paper effectively alleviates the user cold start problem.

KEYWORDS

Cold Start, Cultural Distance, Matrix Decomposition, Service Recommendation

INTRODUCTION

The number and types of services continue to increase due to the development of technologies like cloud services, internet of things (IoT) services, mobile services, and microservices. Service recommendation technology emerged to identify a service that meet individual users' needs (Qi, 2023; Zhang, Y., Yin, C., et al., 2021). Service recommendation is generally used to model a user's interest by analyzing their historical behavior. Thus, a service can be recommend that meets a user's preference requirements (Zhang, Y., Cui, G., 2021). In a service recommendation system, the quality of service (QoS) should be predicted when identifying a service that meets a user's quality requirements. Therefore, accurate prediction of the QoS value (Zhang, Y., Wang, K., He et al., 2021; Qi et al., 2020; Qi et al., 2022) prior to the user call is an important step in service recommendation (Zhang, Y., Zhang, Yan, 2023; Zhang, Hu, Zhang, 2021; Cui et al., 2020).

Edge computing (Zhang, Pan et al., 2021; Zhang, Cui, Zhao et al., 2016; Zhang, Zhao, Deng, 2018) has introduced new problems to existing recommendation technology. When an emerging technology

DOI: 10.4018/IJWSR.321199
*Corresponding Author

attracts new users, a recommendation system is often used to provide personalized recommendations for new users. The user cold start problem is monumental in recommendation systems (Wang, 2021; Seth & Mehrota, 2021; Amamou et al., 2016). Existing methods to address the cold start of users can be divided into three types. The first, statistical methods, includes the mode method and average method in statistics. These can alleviate the user cold start problem to a certain extent. This type of method is easy to implement; however, it has poor prediction accuracy. The second, demographic methods (Rashid et al., 2008; Liu et al., 2018), uses new users' completed questionnaires to profile users. This type of recommendation method achieves better prediction accuracy than statistical methods; however, it involves user privacy and may lead to a decline in user favorability and user viscosity. The third includes methods based on foreign information sources (Wang et al., 2017; Sahebi & Brusilovsky, 2013). These have good prediction accuracy and recommendation effects. Still, the foreign information sources infringe more on user privacy.

This article focuses on the cold start problem in web services and edge services, including at the protection of user privacy. A recommendation algorithm is proposed based on matrix factorization to solve the cold start problem of users. There is a certain correlation between user preferences and cultural background; therefore, this study introduces Hofstede's cultural dimensions theory (Hofstede et al., 2010). First, it establishes the connection between two users. Second, it ensures QoS for cold start users' stability of prediction accuracy. The experimental results show that the proposed method largely solves the problem of predicting the accuracy of cold starts by users. The main contributions of this article are presented as follows:

1. Using the location information of the user and service, the cultural distance between the user and service is calculated through Hofstede's cultural dimensions theory. The connection among the cold-start user, old user, and service is established without infringement on user privacy.
2. The preliminary prediction of new users is conducted by searching for heuristic services. The preliminary prediction results are dynamically integrated with the matrix decomposition technology. The model parameters are updated by continuously fitting the training set. The final trained model is used to complete the prediction of new users.
3. Several experiments were conducted on the real dataset (WS-Dream). The experimental results show that this method has better prediction accuracy than existing recommendation technology for solving the user cold start.

The remaining content of this article is arranged as follows. The next section introduces the current cold start recommendation methods. Then, it discusses the prediction framework and implementation process proposed in this article. This is followed with a verification of the effectiveness of the algorithm framework via experiments. Finally, the article provides a summary and prospects for follow-up work.

RELATED WORK

Before introducing the proposal, the article introduces solutions to the cold start problem. A direct method to alleviate the user's cold start problem is the mode recommendation method in statistics (or popular recommendation method). The most popular method, as proposed by Park and Chu (Park et al., 2009), is based on this idea. Their method ranks services according to popularity. It provides the same recommendation to all users based on the ranking results. This method is easy to implement and the recommendations are obtained quickly. However, the algorithm disregards the subjectivity of users. In addition, it cannot make personalized recommendations. Moreover, the long tail effect is not conducive to the long-term development of recommendation systems.

Another way to alleviate a user's cold start is by collecting user information. This type of method can be subdivided into two types. The first method obtains the preferences of new users in the form of a questionnaire. Rashid et al. (2008) used the clustering neighborhood method (IGCN) to calculate information gain, apply entropy-based information theory, and use a decision tree to select the services recommended to new users. Zhou et al. (Zhou et al., 2011) introduced a nonfunctional matrix factorization method (FMF). This algorithm creates a problem decision tree to obtain the preferences of new users. Each node of the decision tree represents an item of the questionnaire. This highly customized questionnaire was employed to obtain user preferences, improving the accuracy of recommendations. However, this method requires users to actively cooperate with the questionnaire survey to make accurate predictions. The system overhead caused by the questionnaire process is substantial.

Third-party software authorization or third-party data sources can be used to obtain user preferences. Liu et al. (Liu et al., 2016) proposed a location-aware, personalized recommendation algorithm that uses a network of users and location services to identify users like the new user. However, this method requires the new user to have a record of calling the service. Sehebi and Brusilovsky (2013) proposed an effective evaluation index to measure cross-domain recommendation by sharing user ratings of movies and books. Then, they use collaborative filtering algorithms for recommendation. He et al. (He et al., 2017) proposed a method based on reinforcement learning and neural networks that use group information to make recommendations for cold start users. Sedhain et al. (Sedhain et al.,2014) proposed a domain-based cold start method that uses information in social networks.

The method of collecting user information can alleviate the user's cold start. However, the method has its disadvantages. First, the user information collected through questionnaires must be verified. Inauthentic questionnaire surveys will negatively influence the recommendation results. Additionally, recommendation results obtained via questionnaire surveys are often not diverse. Second, obtaining user information by third-party software or third-party data sources is difficult to operate and incurs significant legal risks. Last, collecting user information is also an infringement on user privacy. Some users feel they have been infringed upon. Accordingly, their favorability toward the recommendation system will be greatly reduced.

Based on these considerations, this article proposes to solve the user's cold start problem by making a trade-off between recommendation accuracy and protection of user privacy. Cultural dimension information is dynamically combined with the matrix decomposition model to provide more accurate predictions while protecting user privacy.

USER COLD START PREDICTION FRAMEWORK AND PROCESS

In this article, the main idea of the method is to combine cultural distance with a matrix factorization model. Cultural distance, one of the central concepts in the study of cultural differences, is widely applied in cross-cultural research and applications. Differences in culture directly affect an individual's beliefs, thoughts, and social behaviors. All calculations of cultural distance are based on the study of cultural dimensions.

Hofstede's cultural dimensions theory provides a quantitative description of abstract cultural concepts. The quantified cultural theory can render culture more intuitive and specific. This cultural dimension theory allows researchers to express culture as data and intuitively compare the differences between cultures and behaviors caused by cultural differences.

Hofstede's theory of cultural dimensions is composed of six cultural dimensions: (1) power distance index; (2) individualism and collectivism; (3) masculinity and femininity; (4) uncertainty avoidance index; (5) long-term orientation and short-term normative orientation; and (6) indulgence and restraint. These represent the independent preferences of each state, which distinguishes countries (rather than individuals) from each other. The country or region score of each dimension is relative; thus, scores can only be meaningfully employed.

Based on Hofstede's cultural dimensions theory, applied a simple mathematical formula to define cultural distance and more simply describe cultural differences (Kogut et al., 1988). The formula is expressed as follows:

$$cd_j = \sum\nolimits_{i-1}^{m} \frac{(I_{ij} - I_{iu})^2}{mV_i} \tag{1}$$

where cd_j is the cultural distance of country j, I_{ij} is the value of the i-th dimension of country j in Hofstede's cultural theory, V_i is the variance of the value of the i-th dimension, and m is the number of cultural dimensions.

The user cold start problem is defined as follows:

Definition 2. Service cold start problem:
Given: (1) Existing user-service rating matrix;
(2) Geographical location information of existing users;
(3) Location information of new users.
Requirement: prediction of the service score of new users.

As the definition shows, the existing user rating information is known. The cold start problem of new users based on existing information is the problem to be solved. In a recommendation system, due to a lack of sufficient historical scores, it is very difficult for the system to make personalized recommendations for cold start users.

Algorithm Framework

User cold start is one of two major problems that coexist with service cold start in service recommendation. However, the difference between user cold start and service cold start is that new services can directly obtain their information and related attributes. For users, user information often involves user privacy. It is also difficult to obtain. In the current environment, the best user information that is easy to obtain is a user's geographic location Thus, it is effective to combine it with the user's geographic location to recommend new users. The cultural dimension can reflect the user's social behavior and preferences to a large extent. In this article, a user's geographic information is used to establish a cold start recommendation model based on Hofstede's cultural dimensions theory.

To solve the cold start problem of new users, this article introduces cultural distance as a similar association between new users and old users. Cultural distance is used as an attribute of new users. With this attribute, this article proposes a matrix decomposition method that combines cultural distance to predict user preferences based on a certain cultural dimension. The experimental model is shown in Figure 1. The method includes three stages: (1) feature learning based on interactive information; (2) cultural distance calculation; and (3) service quality prediction and recommendation after dynamic fusion.

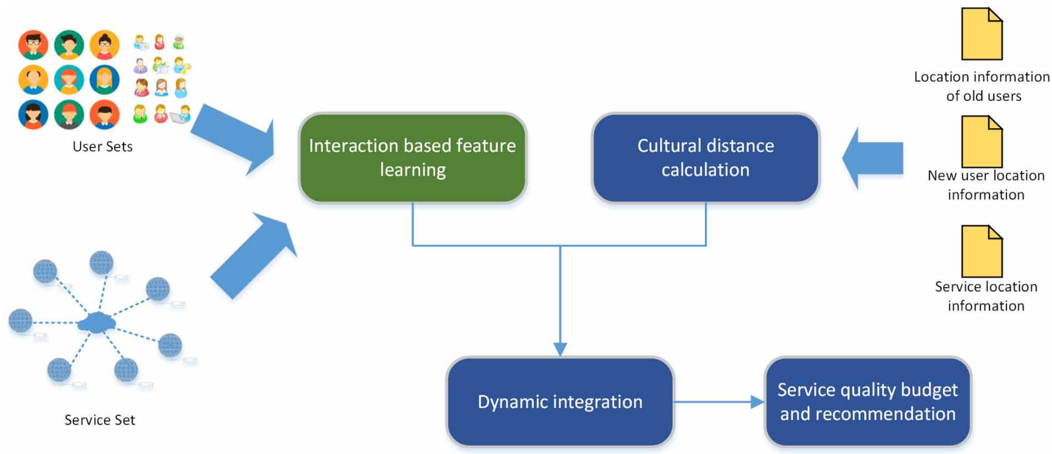
Decomposition Algorithm Based on Cultural Distance

As shown in Figure 1, the algorithm is divided into the following three steps: (1) feature learning based on service quality information; (2) cultural distance calculation based on geographic location information; and (3) service quality prediction after dynamic fusion and recommendation.

Feature Learning Based on Service Quality Information

With the development of service-oriented architecture, an increasing number of services are deployed on the internet. Simultaneously, many services with the same or similar functions are available. For users, choosing the appropriate service should start with functional and nonfunctional requirements like response time and throughput in the service. QoS value is used to describe the nonfunctional attributes of the service.

Figure 1. Decomposition algorithm framework based on cultural distance



Service quality can reflect a user's operating experience. Different operating experiences create different feelings for users. The meaning of a recommendation system is to identify the service with the best experience for users. However, users often cannot make a selection after calling all services. Therefore, the recommendation system must predict the service quality of services that the user has not invoked in the system. According to the prediction results, this article will illustrate how to make service recommendations for users.

Matrix decomposition is a popular and effective method for predicting missing values of scores in service recommendation. The algorithm idea is to map users and services into two recessive feature matrices with dimension D according to the user service matrix. Specifically, the service quality value rij in each system can be expressed as a linear relationship between the called user vector Ui and the service vector Sj. The decomposition model is shown in Figure 2.

In Figure 2, the matrix factorization model can be regarded as a four-layer neural network model. The model is divided into an input layer, embedding layer, inner product layer, and output layer. Each layer performs different functions.

For the input layer, the user number and service number are arranged according to a fixed code. The result is then used as the input of the neural network. The encoding method in this section is one-hot encoding. The idea of one-hot encoding is that each user or service corresponds to a binary vector (only one bit as 1 and the remaining bits as 0).

The function of the embedding layer is to map each code into a dense vector of fixed length. The one-hot code input by the input layer is a sparse, high-dimensional code. If these high-dimensional vectors are directly employed, the system efficiency will be reduced. The vector in the embedding layer is continuously updated when training the neural network. The user feature vector and service feature vector are formed.

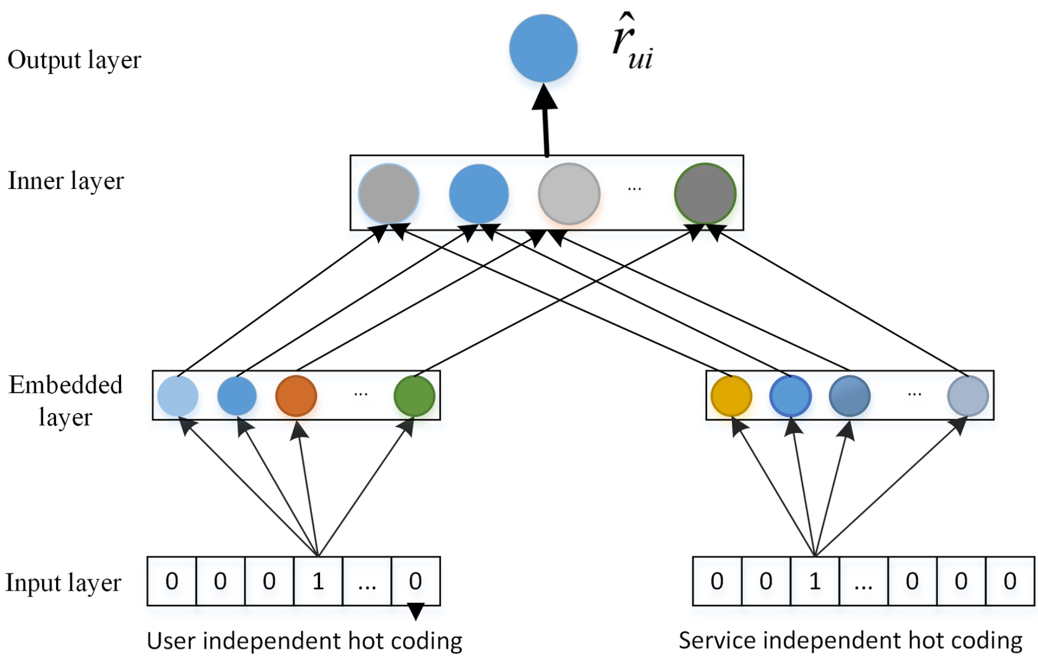
The function of the inner product layer is to multiply the feature vectors obtained by the embedding layer. Then, it adds the result of the multiplication to obtain the result of the interaction between the user and service.

The function of the output layer is to linearly combine the results of the inner product layer with other factors. Then, it obtains the predicted result.

Calculation of Cultural Distance Based on Geographic Location Information

According to Hofstede's cultural dimensions theory, a certain service preference of users can be explained as the inevitable result of users in a certain culture. An increasing number of service

Figure 2. Schematic of matrix decomposition



providers are providing services with different characteristics for users in various regions and cultures. The introduction of cultural distance reflects the potential characteristics of services. In addition, it characterizes users' preferences for services. Therefore, this article use the normalized cultural distance between two users to represent the cultural similarity between two users. A user's preference for service is represented by the cultural distance difference between users and services:

$$C_{uv} = \left(1 - \frac{|cd_u - cd_v|}{\max(cd_u, cd_v)}\right) \quad (2)$$

where Cuv represents the cultural similarity between user u and user v. cdu and cdv refer to the cultural distance between user u and user v.

Like the cultural distance between two users, the cultural distance between two services can also be calculated:

$$C_{ij} = \left(1 - \frac{|cd_i - cd_j|}{\max(cd_i, cd_j)}\right) \quad (3)$$

where Cij represents the cultural distance between service i and service j. cdi and cdj refer to the cultural distance between service i and service j.

User u's preference for service Dus is measured by the absolute value of the difference between the cultural distance between the user and service. The calculation formula is expressed as follows:

$$D_{us} = 1 - \frac{\left|cd_u - cd_s\right|}{\max(cd_u, cd_s)} \tag{4}$$

For a new user, the system lacks historical record. Thus, the user cannot be recommended based on behavior. A new user is likely to call services that have been called by users like themselves and services in the same cultural field. Therefore, these services are chosen as enlightening services. The selection of heuristic services is very important. In traditional collaborative filtering algorithms, the K-nearest neighbor algorithm is used to select similar neighbors. This method selects K users who are most like the user to predict the quality of service of the user.

In the web service recommendation scenario, using the K-nearest neighbor algorithm to select similar neighbors has many disadvantages. Therefore, this article merges the location of users and web services into the selection of enlightening services. The selection method is described as follows.

In Stage 1, the user set N(u) whose cultural similarity with the target user is greater than threshold ε1 is selected. The services jointly invoked by the user set are selected as the heuristic service set SN(u). If there are fewer than K services that have been jointly called, the phase 2 operation is performed. Otherwise, SN(u) is performed and the service quality value of the new users for these heuristic services is predicted. The service quality prediction value of heuristic service is:

$$\hat{r}_{ui} = \bar{r}_i + \frac{\sum_{v \in N_{(u)}} C_{uv}\left(r_{vi} - \bar{r}_i\right)}{\sum_{v \in N_{(u)}} C_{uv}} \tag{5}$$

In Stage 2, the service set S(u) whose user preference is greater than threshold ε2 is obtained. If the number of services in stage 1 and stage 2 is still less than K, the stage 3 operation is performed. Otherwise, S(u)∪SN(u) is performed. This part of the service quality prediction value selects the average value of the user set NS(u) with the same preference greater than threshold ε2 to replace:

$$\hat{r}_{ui} = \frac{\sum_{v \in NS_{(u)}} r_{vi}}{\left|NS_{(u)}\right|} \tag{6}$$

In Stage 3, for each service in S(u), the service set S′(u) whose cultural similarity is greater than threshold ε3 is selected to complete the heuristic service set. If the number of heuristic service sets in the current stage is still less than K, the actual number of services K′ is used to construct the heuristic service set. The service quality prediction method in Stage 3 is presented as follows, where N(i) represents the set of users who have invoked service i:

$$\hat{r}_{ui} = \frac{\sum_{v \in N_{(i)}} r_{vi}}{\left|N_{(i)}\right|} \tag{7}$$

Service Quality Prediction and Recommendation After Dynamic Integration

The decomposition model in step (1) can only predict the missing service quality values of users based on their historical records. Therefore, it is difficult to perform feature learning for new users.

The prediction accuracy of new users' service quality values is poor. The predicted value of the heuristic service selected in step (2) is regarded as the historical record of the new user. It uses the filled matrix to perform the matrix decomposition operation (figure 3).

After determining the heuristic service, we need to incorporate cultural distance into the decomposition model. Before introducing our model, we introduce the SVD++ decomposition model. The prediction formula of the model is expressed as follows:

$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^T \left(p_u + \frac{1}{\sqrt{\left|N_{(u)}\right|}} \sum_{j \in N_{(u)}} y_j \right) \tag{8}$$

Users and services of different cultures will impact the prediction results. As each user grows, their cultural background will develop in different directions. Correspondingly, the services provided by each service provider will also evolve. Therefore, the learning process of the model needs to update the potential factors of users and services, as well as the cultural dimension vectors of users and services. Therefore, this article introduces a culture-based, baseline prediction model:

$$b_{ui}\left(c\right) = \mu + b_u\left(c\right) + b_i\left(c\right) \tag{9}$$

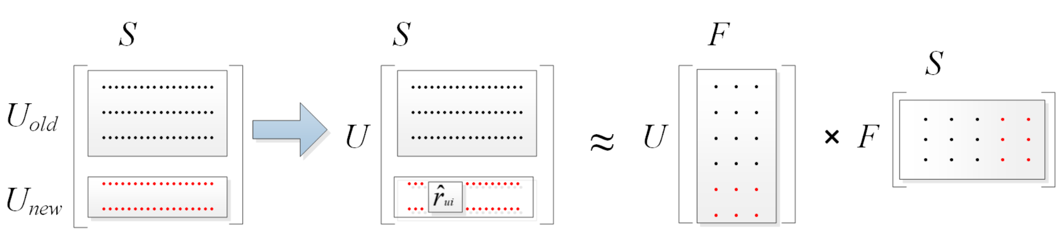
Combining the decomposition forecasting model with the baseline forecasting model, the service quality forecasting formula that integrates cultural dimensions is obtained:

$$\hat{r}_{ui} = \mu + b_u\left(c\right) + b_i\left(c\right) + q_i^T\left(c\right) \left(p_u\left(c\right) + \frac{1}{\left|N_{(u)}\right|} \sum_{j \in N_{(u)}} y_j \right) \tag{10}$$

where:

$$\begin{cases} b_i\left(c\right) = b_i + b_{ic} \\ b_u\left(c\right) = b_u + b_{uc} \\ q_{if}\left(c\right) = q_{if} + q_{ifc} \\ p_{uf}\left(c\right) = p_{uf} + p_{ufc} \end{cases} \tag{11}$$

Figure 3. Decomposition diagram of user cold start



According to the prediction formula, the loss function of the model is expressed as follows:

$$L = \frac{1}{2}\sum_{(u,i\in D)}(r_{ui} - \hat{r}_{ui})^2 + \frac{\lambda}{2}\left(\sum_{i\in S}(\|q_i\|^2 + b_i(c)^2) + \sum_{u\in U}(\|p_u\|^2 + b_u(c)^2)\right) \tag{12}$$

The loss function for each point is expressed as follows:

$$L(u,i) = \frac{1}{2}(r_{ui} - \hat{r}_{ui})^2 + \frac{\lambda}{2}(\|q_i\|^2 + b_i(c)^2 + \|p_u\|^2 + b_u(c)^2) \tag{13}$$

where ||·|| represents the second normal form and D represents the overall score set in the training set. Using the gradient descent operation of the loss function L(u,i) of each point, the prediction model is continuously updated. The parameters of the model are updated by the gradient descent method as follows:

$$\begin{cases} q_i \leftarrow q_i - \alpha\dfrac{\partial L(u,i)}{\partial q_i} \\ p_u \leftarrow p_u - \alpha\dfrac{\partial L(u,i)}{\partial p_u} \\ b_i(c) \leftarrow b_i(c) - \alpha\dfrac{\partial L(u,i)}{\partial b_i(c)} \\ b_u(c) \leftarrow b_u(c) - \alpha\dfrac{\partial L(u,i)}{\partial b_u(c)} \end{cases} \tag{14}$$

The above formula performs derivation and simplification. The final parameter update formula is expressed as follows:

$$\begin{cases} e_{ui} = r_{ui} - \hat{r}_{ui} \\ b_u(c) \leftarrow b_u(c) + \alpha\left(e_{ui} - \lambda b_u(c)\right) \\ b_i(c) \leftarrow b_i(c) + \alpha\left(e_{ui} - \lambda b_i(c)\right) \\ p_u \leftarrow p_u + \alpha\left(e_{ui}\cdot q_i - \lambda p_u\right) \\ q_i \leftarrow q_i + \alpha\left(e_{ui}\cdot\left(p_u + \dfrac{1}{\sqrt{\|N_{(u)}\|}}\sum_{j\in N_{(u)}} y_j\right) - \lambda\cdot q_i\right) \\ y_j \leftarrow y_j + \alpha\left(e_{ui}\cdot\dfrac{1}{\sqrt{\|N_{(u)}\|}}\cdot q_i - \lambda\cdot q_i\right) \end{cases} \tag{15}$$

EXPERIMENTAL RESULTS AND ANALYSIS

This section tests the accuracy of the proposed algorithm. First, it introduces the experimental operation environment. It explains and preprocesses the dataset in the experimental phase. Second, it introduces the evaluation function of the experiment and explains the evaluation function. Third, it introduces the comparison algorithm for comparison with our method. Fourth, it describes the validity of the CAUCS method and influence of each parameter on the model.

Experimental Environment and Dataset

The experimental environment of this part is JetBrains PyCharm 2017.1 x64, python 3.6.5. The experimental code is written in python. The experimental equipment is configured with 16G memory. The processor is Intel core i7-4970. The operating system is win 10.

To verify the effectiveness of the recommendation algorithm, this article conducts experiments on the formal dataset WS-Dream. The dataset records the geographic information and IP information of 339 users, as well as the information of 5,825 service providers and the country to which the service belongs. The dataset contains two scoring matrices—the real response time value and real throughput value of 1,974,675—when 339 users call 5,825 services. This dataset is an ideal record of users calling services. Most users in the real world often only need to call a few services. In most cases, the user-service score matrix is sparse. To better simulate the data sparsity that exists in reality, the article randomly selects 20% of users as cold-start users when dividing the dataset. For the remaining 80% of users, it selects 10% and 15% of their call records (20%, 25% and 30% serve as training sets for training).

Evaluation Function

To evaluate the effectiveness of the proposed method, this article conducted experiments on the WS-Dream dataset according to different dataset division methods. This article chose two commonly employed evaluation indicators—the absolute mean error (MAE) and root mean square error (RMSE)—to test the accuracy of the experiment. Definitions are shown in Formula (16) and Formula (17):

$$MAE = \frac{\sum_{u,i} \left| r_{ui} - \hat{r}_{ui} \right|}{N} \tag{16}$$

$$RMSE = \sqrt{\frac{\sum_{u,i} \left(r_{ui} - \hat{r}_{ui} \right)^2}{N}} \tag{17}$$

where r_{ui} represents the real score value of user u for service i, $\hat{r}_{ui}$ represents the predicted score value of user u for service i, and N represents the number of scores in the test set. The value of MAE is calculated by the average difference between the predicted score and true score for the service in the test set. The smaller its value, the higher the prediction accuracy.

The value of RMSE is the square root of the deviation between the predicted value and true value and the square root of the ratio of the number of observations N. RMSE is sensitive to large errors in the predicted value. It has high requirements for the stability of the experimental method.

Comparison Algorithm

To prove the superiority of the proposed algorithm, this article chose the following four methods to test the performance of the experiment by comparing the two evaluation indicators, MAE and RMSE, with the proposed method:

1. **MODE Recommendation Algorithm Based on Statistics:** The highest quality value of the service user in the system is applied for prediction.
2. **Service-Based Average Prediction (SMEANS):** This algorithm is derived from statistics. It predicts the score of cold-start users based on the average score of all users who call the service. The idea of this method is relatively simple, easy to implement, and has a certain prediction accuracy. Therefore, this method is widely utilized as a comparison method in service recommendation.
3. **Service-Based K-Nearest Neighbor Algorithm (UKNN):** This algorithm is an improved algorithm based on collaborative filtering (Breese et al., 1997). It recommends a service based on the rating information of the first K similar neighbors of each user. It is impossible to identify similar neighbors of cold start users based on historical scores. Thus, this article selects similar users based on cultural distance. According to the weighted average of similar neighbors, the score of cold start users on the new service is predicted.
4. **Matrix Factorization (MF):** MF has always been one of the most famous and effective methods in recommendation algorithms (Zhou et al., 2021). The premise of the algorithm is to decompose the rating matrix into the product of a status user attribute matrix and service attribute matrix. The parameters are updated by continuously fitting the scores in the training set. The final model is utilized to complete the prediction of the missing scores.

Experimental Results and Analysis

This section conducts an experimental evaluation of the proposed method. All experiments verify whether the algorithm obtains an accurate recommendation for user cold start. The best prediction effect of the model is achieved by adjusting parameters in the model. Improvement in the prediction accuracy of the algorithm in this article is determined by comparing it with the recommendation results of other cold-start recommendation algorithms.

First, the article unifies the relevant parameters designed in the experiment. It uses Top-k = 10. In the matrix decomposition process, the dimension of the decomposition is D = 50, the number of iterations is 200, and the matrix update parameters are $\alpha = 0.005$ and $\gamma = 0.02$.

Influence of Training Matrix Density on Prediction Results

To verify the influence of the training matrix density on the prediction results, the article sets the training set matrix density from 0.1 to 0.3, with an increase of 0.05. The experiment was repeated 10 times on the training set of these densities. The average value was obtained. The final experimental results are shown in Table 1.

The experimental results in Table 1 are summarized as follows:

1. The method proposed in this article is always significantly better than the other four methods with the MAE and RMSE indicators. This shows that the forecast accuracy can be improved by using the project attribution rating, which is not a cold start recommendation.
2. As the density increases from 0.1 to 0.3, the MAE and RMSE values of the five methods show an overall downward trend. This indicates that a larger training set can provide more information for prediction and an increasing trend in the training set. As the density increases, the accuracy

Table 1. Data Sparsity Experiment Results

RESPONSE TIME	MAE/Matrix density					RMSE/Matrix density				
	0.1	**0.15**	**0.2**	**0.25**	**0.3**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**
MODE	0.828	0.704	0.629	0.658	0.561	2.135	2.010	1.820	1.837	1.604
SMEANS	0.696	0.671	0.672	0.706	0.664	1.587	1.590	1.609	1.601	1.531
MF	0.867	0.867	0.860	0.864	0.791	1.615	1.577	1.621	1.561	1.418
UKNN	0.732	0.681	0.677	0.676	0.598	1.866	1.733	1.798	1.688	1.347
CAUCS	**0.617**	**0.611**	**0.608**	**0.601**	**0.597**	**1.634**	**1.650**	**1.638**	**1.629**	**1.331**
THROUGHPUT	**MAE/Matrix density**					**RMSE/Matrix density**				
	0.1	**0.15**	**0.2**	**0.25**	**0.3**	**0.1**	**0.15**	**0.2**	**0.25**	**0.3**
MODE	37.06	36.02	35.24	398	343	93.05	88.91	89.05	85.55	83.57
SMEANS	30.38	28.68	29.59	28.03	28.02	73.87	69.05	66.83	67.95	66.83
MF	41.20	36.60	36.59	35.05	28.70	86.72	83.79	92.61	88.44	82.88
UKNN	32.35	29.54	29.29	27.53	27.27	82.38	75.50	71.65	67.40	603
CAUCS	**23.65**	**23.19**	**21.85**	**21.90**	**21.48**	**63.03**	**62.16**	**59.56**	**60.27**	**56.77**

of the experiment worsens. From the theoretical analysis, the accuracy of the prediction should be related not only to the training set matrix density but also to the influence of other elements. This is because the increasing training set density will provide more information. The increasing number of training sets will introduce untrustworthy user information, which will also affect the results of the experiment.

Top-K's Impact on Prediction Accuracy

According to the method proposed in this article, the Top-K parameter is used to control the number of similar items. The number of neighbors is an important factor affecting prediction performance. To explore the influence of Top-K on the experimental results, the article conducted experiments on the training set with a density of 0.1. It simultaneously set $\alpha = 0.005$ and $\gamma = 0.02$. The prediction results for response time and throughput in the dataset are shown in Figure 4.

Experiments were conducted on two datasets to verify the influence of the top-k value on the prediction accuracy. Without changing other parameters, the influence of different Top-K values on the experimental prediction accuracy was tested by adjusting the value of Top-K. Figure 4(a) and (b) show the experimental results of the model in response time. Figure 4(c) and 4(d) show the experimental results of the model in throughput. According to Figure 4, with an increase in Top-K, the change trend of MAE and RMSE decreases and then increases. The prediction effect of the experiment is most accurate when Top-K = 10. The experimental results show that the prediction accuracy will be reduced regardless of whether the value of Top-K is too large or small. If the value of Top-K is too small, then the relationship between two items cannot be fully consider. Too many irrelevant items will be added to the similarity calculation if the Top-K value is too large.

Influence of Parameter α on Prediction Accuracy

The parameter α, the learning rate of our model, determines the convergence speed of this method. To verify the influence of α on the prediction performance, different levels of α are used to verify the prediction accuracy of the model via experiments. Other parameters remain unchanged. The experimental verification results are shown in Figure 5.

Figure 4. Top-K's impact on prediction accuracy

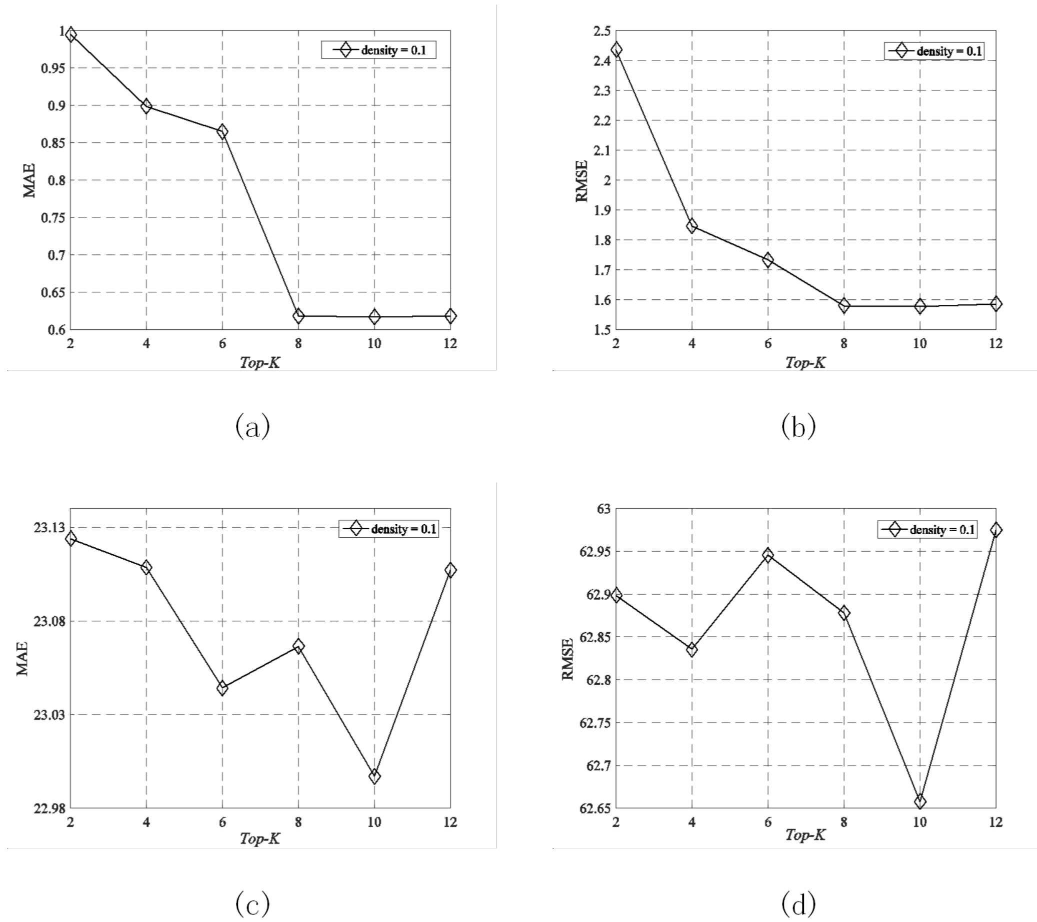


Figure 5 shows the effect of α on the MAE and RMSE in the FASVD model. For both MAE and RMSE, the prediction accuracy decreases and then increases. The best performance was achieved on the order of -3. By theoretical analysis, the experimental prediction results can be explained as follows:

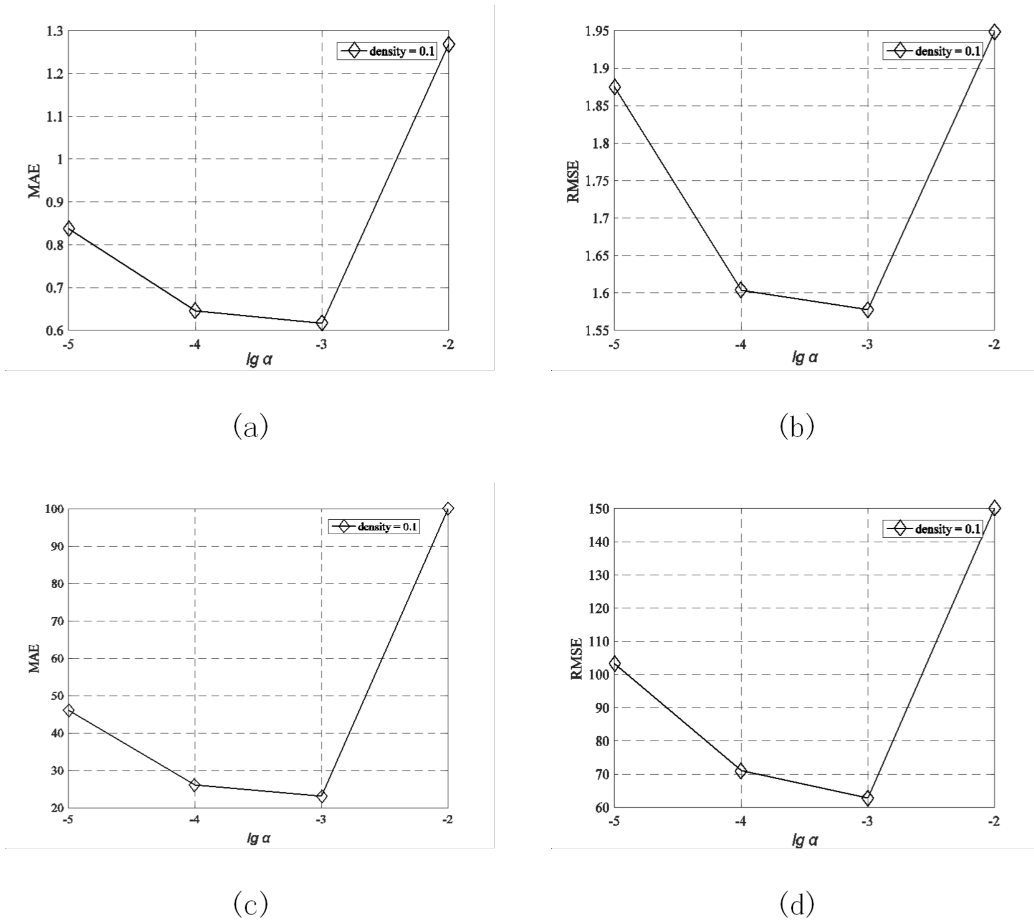
- When the value of α is below -3, the model has not fully converged at the end of the iterations, producing inaccurate prediction results.
- If α is too large, the experimental results may jump out of convergence. They may be unable to achieve the purpose of prediction.

Influence of Parameter γ on Prediction Accuracy

Parameter γ is used to control the degree of regularization of the proposed method and reduce the degree of overfitting of the model. Similarly, according to the beginning of this section, other parameters remain unchanged. The influence of parameter γ on the experimental prediction results is verified by different levels of γ values. The experimental results are shown in Figure 6.

As shown in Figure 6, the experimental prediction results are not accurate if the parameter γ is too large or too small. When the value of γ is too small, overfitting of the model to the training set may occur, yielding an inaccurate prediction. When γ is too large, the penalty for the model is too

Figure 5. Influence of parameter α on prediction accuracy



large. In addition, some features in the training set are lost, resulting in underfitting of the model. Therefore, the best prediction effect cannot be obtained. The model in this article's optimal interval of parameter γ is [0.01, 0.05].

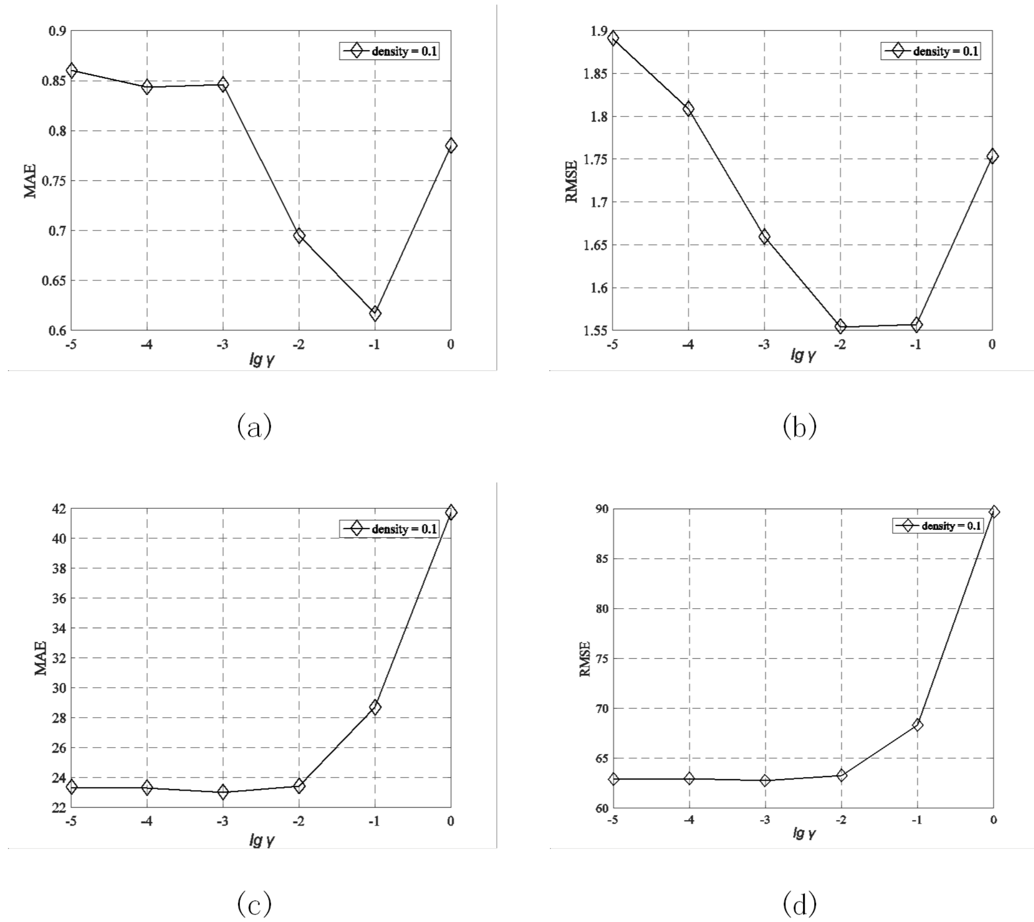
SUMMARY

This article proposes a recommendation algorithm based on cultural distance to solve the cold start problem of new users. The difference from the current method of solving the user cold start problem is that this article only uses the user's location information (without the need for tedious questionnaires). It is key to obtain information about a user's interest without accessing their private information.

This article aimed to provide accurate recommendation of services without infringing on user privacy. Cultural distance was used to identify startup services that are most likely to interest users. These can then dynamically integrate the predicted results with the decomposition model. The model parameters are updated by continuous fitting in the training set. The optimal prediction model is obtained and the prediction of missing values is performed according to the model.

The results on the real dataset, WS-Dream, show that the proposed method can effectively solve the user cold start problem. The algorithms in this article are based on matrix decomposition;

Figure 6. Influence of parameter γ on prediction accuracy



therefore, the computational efficiency of the algorithm may be greatly reduced for large-scale scoring matrices. Subsequent research can improve the efficiency of the algorithm by improving the decomposition model.

ACKNOWLEDGMENT

The authors of this article declare there are no competing interest. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Funding for this research was covered by the authors of the article.

REFERENCES

Amamou, A. A., Kelouwani, S., Boulon, L., & Agbossou, K. (2016). A comprehensive review of solutions and strategies for cold start of automotive proton exchange membrane fuel cells. *IEEE Access : Practical Innovations, Open Solutions*, *4*, 4989–5002. doi:10.1109/ACCESS.2016.2597058

Breese, J. S., Heckerman, D., & Kadie, C. (1997). Empirical analysis of predictive algorithms for collaborative filtering. *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence* (pp. 43–52). Morgan Kaufmann.

Chen, J., Zhang, H., He, X., Nie, L., Liu, W., & Chua, T. (2017). Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention. *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*, (pp. 335-344).ACM. doi:10.1145/3077136.3080797

Cui, G., He, Q., Chen, F., Jin, H., & Yang, Y. (2020). Trading off between user coverage and network robustness for edge server placement. *IEEE Transactions on Cloud Computing*, *10*(3), 2178–2189. doi:10.1109/TCC.2020.3008440

Hofstede, G., Hofstede, J., & Minkov, M. (2010). *Cultures and organizations - Software of the mind: Intercultural cooperation and its importance for survival*. McGraw-Hill.

Kogut, B., & Singh, H. (1988). The effect of national culture on the choice of entry mode. *Journal of International Business Studies*, *19*(3), 411–432. doi:10.1057/palgrave.jibs.8490394

Liu, J., Tang, M., Zheng, Z., Liu, X. F., & Lyu, S. (2016). Location-aware and personalized collaborative filtering for web service recommendation. *IEEE Transactions on Services Computing*, *9*(5), 686–699. doi:10.1109/TSC.2015.2433251

Liu, Y., Han, Y., Iserman, K., & Jin, Z. (2018). Minimizing required user effort for cold-start recommendation by identifying the most important latent factors. *IEEE Access : Practical Innovations, Open Solutions*, *6*, 71846–71856. doi:10.1109/ACCESS.2018.2878407

Park, S. T., & Chu, W. (2009). Pairwise preference regression for cold-start recommendation. *Proceedings of the third ACM conference on Recommender systems*(pp. 21-28). ACM. doi:10.1145/1639714.1639720

Qi, L., Chen, Y., Yuan, Y., Fu, S., Zhang, X., & Xu, X. (2020). A QoS-aware virtual machine scheduling method for energy conservation in cloud-based cyber-physical systems. *World Wide Web Journal*, *23*(2), 1275–1297. doi:10.1007/s11280-019-00684-y

Qi, L., Hu, C., Zhang, X., Khosravi, M. R., Sharma, S., Pang, S., & Wang, T. (2023). Privacy-aware data fusion and prediction with spatial-temporal context for smart city industrial environment. *IEEE Transactions on Industrial Informatics*, *17*(6), 4159–4167. doi:10.1109/TII.2020.3012157

Qi, L., Yang, Y., Zhou, X., Rafique, W., & Ma, J. (2022). Fast anomaly identification based on multi-aspect data streams for intelligent intrusion detection toward secure industry 4.0. *IEEE Transactions on Industrial Informatics*, *18*(9), 6503–6511. doi:10.1109/TII.2021.3139363

Rashid, A. M., Karypis, G., & Riedl, J. (2008). Learning preferences of new users in recommender systems: An information theoretic approach. *SIGKDD Explorations*, *10*(2), 90–100. doi:10.1145/1540276.1540302

Sahebi, S., & Brusilovsky, P. (2013). Cross-domain collaborative recommendation in a cold-start context: The impact of user profile size on the quality of recommendation. *International Conference on User Modeling, Adaptation, and Personalization* (pp. 289–295). Springer. doi:10.1007/978-3-642-38844-6_25

Sedhain, S., Sanner, S., Braziunas, D., Xie, L., & Christensen, J. (2014), Social collaborative filtering for cold-start recommendations, *Proceedings of the 8th ACM Conference on Recommender systems*, (pp. 345-348) ACM.

Sethi, R., & Mehrotra, M. (2021). Cold start in recommender systems-a survey from domain perspective. *Intelligent Data Communication Technologies and Internet of Things,* (pp. 223–232). IEEE.

Wang, H. (2021). ZeroMat: Solving cold-start problem of recommender system with no input data. *Proceedings of the IEEE 4th International Conference on Inf. Syst. Comput. Aided Educ. (ICISCAE)* (pp. 102–105). IEEE. doi:10.1109/ICISCAE52414.2021.9590668

Wang, X., He, X., Nie, L., & Chua, T. (2017). Item Silk Road: Recommending items from information domains to social users. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 185 –194). ACM. doi:10.1145/3077136.3080771

Zhang, C., Zhao, H., & Deng, S. (2018). A density-based offloading strategy for IoT devices in edge computing systems. *IEEE Access : Practical Innovations, Open Solutions*, *6*, 54258–54269. doi:10.1109/ACCESS.2018.2882452

Zhang, Y., Cui, G., Deng, S., Chen, F., Wang, Y., & He, Q. (2021). Efficient query of quality correlation for service composition. *IEEE Transactions on Services Computing*, *14*(3), 695–709. doi:10.1109/TSC.2018.2830773

Zhang, Y., Cui, G., Zhao, S., & Tang, J. (2016). IFOA4WSC: A quick and effective algorithm for QoS-aware service composition. *International Journal of Web and Grid Services*, *12*(1), 81–108. doi:10.1504/IJWGS.2016.074186

Zhang, Y., Hu, B., & Zhang, Y. (2021). Model-driven open ecological cloud enterprise resource planning. *International Journal of Web Services Research*, *18*(3), 82–99. doi:10.4018/IJWSR.2021070105

Zhang, Y., Pan, J., Qi, L., & He, Q. (2021). Privacy-preserving quality prediction for edge-based IoT services. *Future Generation Computer Systems*, *114*, 336–348. doi:10.1016/j.future.2020.08.014

Zhang, Y., Wang, K., He, Q. et al. (2021). Covering-based web service quality prediction via neighborhood-aware matrix factorization. *IEEE Transactions on Services Computing, 14*(5)m 1333–1344.

Zhang, Y., Yin, C., Wu, Q., He, Q., & Zhu, H. (2021). Location-aware deep collaborative filtering for service recommendation. *IEEE Transactions on Systems, Man, and Cybernetics. Systems*, *51*(6), 3796–3807. doi:10.1109/TSMC.2019.2931723

Zhang, Y., Zhang, Y., Yan, D., Deng, S., & Yang, Y. (2023). Revisiting graph-based recommender systems from the perspective of variational auto-encoder. *ACM Transactions on Information Systems*, *41*(3), 1–28. doi:10.1145/3573385

Zhou, C., Shen, J., Guo, X., & Zhou, Z. (2021). Task-oriented battlefield situation information hybrid recommendation model. *Intelligent Automation & Soft Computing*, *27*(1), 127–141. doi:10.32604/iasc.2021.012532

Zhou, K., Yang, S. H., & Zha, H. (2011). Functional matrix factorizations for cold-start recommendation. *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*(pp. 315-324).ACM.