Research Article

A Triangular Personalized Recommendation Algorithm for Improving Diversity

Biao Cai [iD],[1] Xiaowang Yang [iD],[2] Yusheng Huang,[2] Hongjun Li,[1] and Qiang Sang [iD][1]

[1]*Department of Digital Media Technology, College of Information Science & Technology,*
 Chengdu University of Technology, Chengdu, Sichuan, China
[2]*College of Information Science & Technology, Chengdu University of Technology, Chengdu, Sichuan, China*

Correspondence should be addressed to Xiaowang Yang; 1096404495@qq.com

Recommendation systems are used when searching online databases. As such they are very important tools because they provide users with predictions of the outcomes of different potential choices and help users to avoid information overload. They can be used on e-commerce websites and have attracted considerable attention in the scientific community. To date, many personalized recommendation algorithms have aimed to improve recommendation accuracy from the perspective of vertex similarities, such as collaborative filtering and mass diffusion. However, diversity is also an important evaluation index in the recommendation algorithm. In order to study both the accuracy and diversity of a recommendation algorithm at the same time, this study introduced a "third dimension" to the commonly used user/product two-dimensional recommendation, and a recommendation algorithm is proposed that is based on a triangular area (TR algorithm). The proposed algorithm combines the Markov chain and collaborative filtering method to make recommendations for users by building a triangle model, making use of the triangulated area. Additionally, recommendation algorithms based on a triangulated area are parameter-free and are more suitable for applications in real environments. Furthermore, the experimental results showed that the TR algorithm had better performance on diversity and novelty for real datasets of MovieLens-100K and MovieLens-1M than did the other benchmark methods.

1. Introduction

With the rapid development of the Internet and e-commerce, which have tremendous impacts on our lifestyles, the way that information is accessed has been changed. On the one hand, hundreds of millions of products are available online, making life much more convenient [1]. On the other hand, there is a problem of information overload. The large amount of data generated every day makes it difficult for items we want to be chosen as easily as previously [2]. Personalized recommendation is considered the most effective way to solve the problem of information overload [3, 4] and thus far, recommendation systems have been used in many fields [5–7]. However, it is still difficult to meet the growing demand for product information services [8]. Investigation results show that the underlying reasons for this problem can vary but the most prominent reason is the lack of adaptability of the recommendation algorithm, that is, the constraints exerted

by the two performance evaluation standards of accuracy and diversity for the conventional recommendation algorithm [8].

A variety of personalized recommendation algorithms have been proposed previously [9–13], and the most representative recommendation algorithm is collaborative filtering, which includes user-based collaborative filtering (UCF) and item-based collaborative filtering (ICF) [14]. UCF and ICF are based, respectively, on the weighted combination of similar users' opinions and the similarity between items [15]. There is also a series of methods for recommendations based on content filtering [16–18], and there is a "community structure" in complex networks [19, 20]. The "time role method" [21] utilizes a time-aware modification of an existing recommendation method and exploits that by combining the temporal and structural aspects. In recent years, many physical dynamics-based methods have been used in recommendation algorithms such as mass diffusion (MD) [22],

heat conduction (HC) [23], and a hybrid recommendation method that combines MD and HC [24–26].

Essentially, all of these methods make recommendations for users by studying vertex similarity and focus primarily on recommendation accuracy, and therefore recommendation diversity is relatively poor. However, the latest study found that although the most popular methods do not achieve very high accuracy [27], its diversity performance is outstanding. Research results have shown that excessive attention to accuracy can have a detrimental effect on the quality of the recommendation results, especially in respect of their relevance to user interests [28, 29].

New developments in the e-commerce have taken place at the two big Chinese companies (Alibaba and Jingdong). For example, these companies are now developing vigorously their offline entities to operate online. The traditional two-dimensional recommendations (user and product model) cannot fully accommodate the roles of location and high diversity in that recommendation model. The present paper proposes a data model that introduces a "third dimension" into the traditional two-dimensional user/product recommendation relationship. A recommendation algorithm based on triangulated area is designed by using the third-dimensional data relationship. In this algorithm, the relationship weighs each of the two relationships in the three-dimensional model and uses the relationship weights between each of the two relationships as the side lengths of a triangle and then the triangular area is calculated using Heron's formula. As each triangle corresponds to a product, recommendations can be made for the target user according to the sorted results of triangular areas, from large to small.

The triangle recommendation algorithm that is proposed in this paper has been tested on two real datasets. The results show that TR was more effective than other benchmark methods in terms of recommendation diversity. In addition, the parameter-free character of TR is an important feature as it is, therefore, easier to apply to reality than some parameter-based benchmark methods. The primary focus of the present work was not limited to overfocusing on recommendation accuracy, and therefore recommendation diversity increased dramatically.

The rest of this paper is organized as follows. In the second part, a detailed description is given of how to build a triangle model and how to apply the model to recommendation systems, and the diversity analysis of the triangle model is explained. In the third section, the experimental data are presented and the performance evaluation standards of the recommendation algorithm are introduced. The fourth section shows the performance of the proposed method on a public dataset and compares its performance with other benchmark methods. Finally, the work is summarized and an indication of the direction for future work is provided.

2. Triangle Recommendation Model

A user-object network without weights and directions $G(U, O, E)$ is used in the study of recommendation system. Its function is to describe the relationship between the user and the product, where $U = \{u_1, u_2, \ldots, u_n\}$, $O = \{o_1, o_2, \ldots, o_m\}$,

and $E = \{e_1, e_2, \ldots, e_z\}$ represent the user set, the object set, and the edge set, respectively. To distinguish between user and object, Latin and Greek indices are used to represent them, respectively. Meanwhile, $G(U, O, E)$ can be represented by an adjacency matrix A, and $a_{i\alpha}$ is defined as an element in the adjacency matrix A. If o_α is collected by the user u_i, then $a_{i\alpha} = 1$; otherwise, $a_{i\alpha} = 0$. The ultimate purpose of any recommendation algorithm is to provide the target user with a list of rankings for objects that are collected. For user i, the recommendation list is defined with a length of L as o_i^L. That is, for user i, o_i^L is a set of the first L uncollected objects that has the highest recommendation points.

2.1. Construction of the Triangle Model. Recommendation accuracy and diversity are studied when evaluating the performance of recommendation algorithms. In order to achieve recommendations with high diversity, a "third dimension ($T1$)" was introduced. In the present study, the category attributes (in the MovieLens-100K dataset there are 19 product categories) of the products were tested as the third dimension. Furthermore, the third dimension in this study could be changed according to the actual application scenario (i.e., business and distance). Then the relationship between any two of the three dimensions is studied. Weighting is used to measure the strength of the relationship and, finally, the three weights are treated as the three sides of a triangle. Each triangle corresponds to an object and the final recommendations are determined based on the triangle area, as shown in Figure 1.

The composition of the three dimensions is U (user), O (object), and $T1$ (third dimension).

In this investigation, the relationship between two dimensions was not studied directly from the perspective of their similarity. In the user/object bipartite graph, the relationship weightings are calculated by combining the K-mean with Markov chain. The user's historical behaviors are applied as the basis of the study at the moment, and the user's future behaviors then are predicted. The obtained prediction values are used as the relationship weight in the user/object relationship, which is denoted by w_{uo}. In the other two relationships, the relationship weightings are calculated using the collaborative filtering method, which are denoted by w_{oT_1}, w_{uT_1}. When u and t are determined, they can be used to construct a three-dimensional spatial structure with multiple objects ($o1$, $o2$ in Figure 2). The recommendation list L is composed of multiple objects. Then u, t, and o can jointly form a three-dimensional structure, as shown in the figure V ($u, t, \{o1, o2\}$). Similarly, a three-dimensional spatial structure V ($u, \{t1, t2, \ldots\}, \{o1, o2, \ldots\}$) can be constructed as shown in Figure 2. While multiple t can correspond to a target u, t can correspond to multiple o values, o can correspond to multiple t values, and so on. Furthermore, the final recommendation list corresponds to the spatial structure shown in Figure 2.

2.2. Solution of the Triangle Model. Firstly, the calculation of weightings w_{uo} between each user and object, using the "K-means Markov chain," was introduced. Here, the K-means
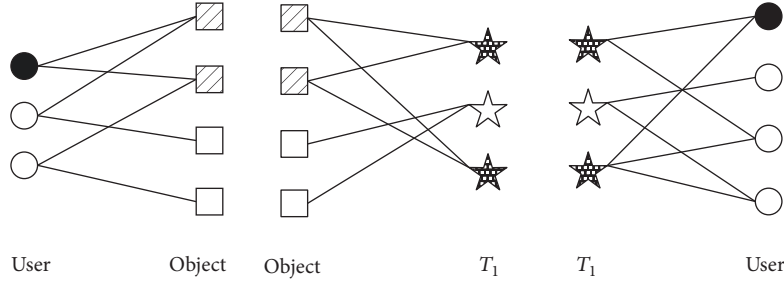
FIGURE 1: These schematics illustrate the relationship between any of the two dimensions in the three dimensions of the proposed recommendation algorithm.
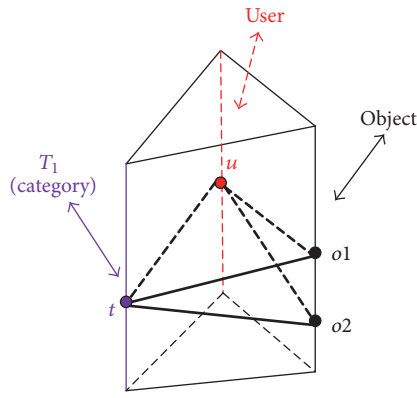

FIGURE 2: The construction of three-dimensional space. The relationships between any of the two dimensions in the three-dimensional model are treated as the three sides of a triangle, and T_1 represents the third dimension in experiments of this paper, which is the category of product.

methodology was used first to conduct user clustering for U_j ($j = 1, 2, \ldots, n$), taking user U_1 as an example.

Step 1. Because the sameness between U_1 and U_j ($j = 2, 3, \ldots, n$) (in order to ensure the accuracy of the recommendation) must be considered in the clustering process and the differences between U_1 and U_j ($j = 2, 3, \ldots, n$) (in order to ensure the recommendation diversity) must also be taken into account, the k-means two-dimensional clustering method was used in the present study, where the number of common objects (CountSame_{1j}) between U_1 and U_j ($j = 2, 3, \ldots, n$) was defined as the first dimension, and $D_{u_j} - \text{CountSame}_{1j}$ was defined as the second dimension. Therefore, the data structure of other user data during the U_1 clustering is

$$\left\{ U_j \left(\text{CountSame}_{1j}, D_{u_j} - \text{CountSame}_{ij} \right) \right\}, \quad (1)$$

where D_{u_j} represents the degree of U_j, $j = 2, 3, \ldots, n$.

The data of U_1 used for clustering is

$$U_1 \left(\frac{1}{n-1} \sum \text{CountSame}_{1j}, D_{u_1} \right.$$
$$\left. - \frac{1}{n-1} \sum \text{CountSame}_{1j} \right). \quad (2)$$

The number of classes is the count of different sets: that is,

$$K = \left| \left\{ D_{u_j} - \text{CountSame}_{1j} \right\} \right.$$
$$\left. \cup \left\{ D_{u_1} - \frac{1}{n-1} \sum_{i=2}^{n} \text{CountSame}_{1i} \right\} \right| \quad (3)$$
$$(j = 2, 3, \ldots, n).$$

The cluster about the user is formed by the clustering according to the Euclidean distance of two-dimensional vector.

Step 2. The cluster of U_1 is found, which is represented by C. The number of different items $N_c = |\text{ActionC}|$, including the behavior records of all the users in C, was counted, where ActionC represents all the users and the product purchase behavior records in C.

The records of each user in ActionC were sorted by the ascending order of timestamps and a sequence then was obtained. Here, the timestamp is a factor in the construction of the Markov Chain.

No specific study was available of a time-based dynamic model as a reference [21]. Due to the uniqueness of object ID, the object ID was taken as the object status and the state transfer matrix $T\text{Matrix}_{N_c * N_c} = 0$ then was constructed.

Step 3. $T\text{Matrix}_{ij}$ is assigned. $T\text{Matrix}_{ij}$ equals the total occurrence number of (i, j) in ActionC divided by the total occurrence number of i in ActionC.

$$T\text{Matrix}_{ij} = \frac{\text{Count}\left(\text{ActionC}\left(i, j\right) \right)}{\text{Count}\left(\text{ActionC}\left(i\right) \right)}. \quad (4)$$

Step 4. The initial behavior vector f^{u_1} of item for U_1 is obtained by sorting the behavioral data of U_1 according to

the order of TMatrix, where $f_j^{u_1}$ is equal to the given score of the target user U_1 on the item j.

Thus, the final result of $W_{u_1 o}$ is

$$W_{u_1 o} = f^{u_1} * T\text{Matrix}^{\text{trans}}. \tag{5}$$

In order to reduce the iteration numbers, the state transfer number was set as $\text{trans} = 3$ after comparing experiments and the collaborative filtering. Thus, the final weight was calculated as

$$W_{u_1 o} = f^{u_1} * T\text{Matrix}^3. \tag{6}$$

Step 5. Next, in the other two relationships, which are the relationship weights of the object and the category w_{oT_1} and the relationship of the user and the category w_{uT_1}, w_{oT_1} was calculated using the collaborative filtering method, which is denoted as follows:

$$w_{oT_1} = \text{CF}\left(\text{Object ID}, \text{Category ID}, w_{0t_1^j}\right), \tag{7}$$

where $w_{0t_1^j}$ is the average weight of product and category, descripted as

$$w_{0t_1^j} = \frac{1}{c} * \left(\frac{1}{n'}\sum_{i=1}^{n} r_{io}\right) \tag{8}$$

in which n' is the number of users that have actions on the given object, r_{io} is the ranking of the ith user on the given object, n is the full user number, and c is the number of categories of the given object.

In addition, the weight of the user and category w_{uT_1} is calculated with collaborative filtering method too and is denoted as follows:

$$w_{uT_1} = \text{CF}\left(\text{User ID}, \text{Category ID}, w_{ut_1^j}\right), \tag{9}$$

where $w_{ut_1^j}$ is described as

$$w_{ut_1^j} = \begin{cases} \dfrac{1}{c} * r_{io}, & n'' = 1 \\ \dfrac{1}{n''}\sum_{j=1}^{n''} w_{ut_1^j}, & n'' > 1, \end{cases} \tag{10}$$

where $w_{ut_1^j}$ is the average weight of user and category and n'' is number of each category about the given user's actions.

Step 6. Each triangle contains an object. As all the three side lengths are known, the triangle area can be obtained according to Heron's formula as follows:

$$R\left(U, O, T_1\right) = \sqrt{p\left(p - w_{uT_1}\right)\left(p - w_{oT_1}\right)\left(p - w_{uo}\right)}, \tag{11}$$

where p represents half of the perimeter:

$$p = \frac{w_{uT_1} + w_{oT_1} + w_{uo}}{2}. \tag{12}$$

Step 7. To make sure the three weights can construct a triangle, w_{uT_1} must be increased or decreased so that a triangle can be constructed with the other two weights. Because it will have an impact on the area whether it increases or decreases w_{uT_1}, the area must be corrected, as illustrated in Section 2.3.

2.3. The Changes of w_{uT_1}

2.3.1. Increase w_{uT_1}. When increasing w_{uT_1}, the area must be decreased accordingly. The increased w_{uT_1} is denoted as

$$w'_{uT_1} = x w_{uT_1}. \tag{13}$$

The unknown x represents the multiple of weight increase. No matter how much it increases, it must be within a small range. If the x value is taken that meets the requirements, the obtained area is

$$\begin{aligned} R'\left(U, O, T_1\right) \\ = \sqrt{p'\left(p' - w'_{uT_1}\right)\left(p' - w_{oT_1}\right)\left(p' - w_{uo}\right)}, \end{aligned} \tag{14}$$

where p' is

$$p' = \frac{w'_{uT_1} + w_{oT_1} + w_{uo}}{2}. \tag{15}$$

Here, an important assumption must be made that the original three weighting values can construct a triangle. When calculating the triangular area, the absolute value must be taken for each product factor in the root of (11); then we can get the median, R_l, where

$$R_l = \sqrt{p\left|p - w_{uT_1}\right| * \left|p - w_{oT_1}\right| * \left|p - w_{uo}\right|}. \tag{16}$$

Then, the final equation for calculating the triangle area is

$$R = \left(1 - P_1\right) * R', \tag{17}$$

where P_1 represents the area correction factor:

$$P_1 = \frac{\left|R_l - R'\right|}{\max\left\{R_l, R'\right\}}. \tag{18}$$

It is important to note that the main purpose of the proposed recommendation algorithm based on the triangle model is to improve the diversity of the recommendation. Thus, when correcting the area after increasing w_{uT_1}, the maximum value of the denominator in (18) is taken.

2.3.2. Decreasing w_{uT_1}. Processing of the area correction for the case of decreasing w_{uT_1} is similar to that for increasing. Only (17) and (18) need to be changed, and the following processing method is obtained:

$$R = \left(1 + P_1\right) * R' \tag{19}$$

$$P_1 = \frac{\left|R_l - R'\right|}{\min\left\{R_l, R'\right\}}. \tag{20}$$

It should be noted that to ensure recommendation accuracy, the minimum value of the denominator in (20) is taken. Although the research focus of this project is on the recommendation diversity, that does not mean giving up accuracy.

2.4. Analysis of Parameters and Diversity

2.4.1. The Analysis of $w_{u T_1}$ Changes. In the construction of a triangle model, there will be cases when the three relationship weightings cannot meet the requirements for constructing a triangle. When w_{uo}, w_{oT_1}, and w_{uT_1} cannot meet the requirements for constructing a triangle, the processing method is as follows: keep the two weighting values unchanged and change the other weighting value. In practical applications, the "third dimension" is a variable factor: for example, it can be replaced by a seller. This investigation does not consider changes of w_{oT1} and w_{uo} because both the relationship ⟨users, objects⟩ and relationship ⟨objects, $T1$⟩ can be obtained from known information (for example, the training dataset). Below is an analysis showing that changes in w_{oT1} and w_{uo} will not affect

user behavior but a change in $w_{u T_1}$ will have an impact on the recommendation result.

Analysis 1. Considering w_{oT1}/w_{uo} will have a direct impact on w_{uo}/w_{oT1}. Because there is no direct information representing the relationship ⟨users, $T1$⟩ in the known information, the change of w_{oT1}/w_{uo} does not directly affect $w_{u T_1}$. Similarly, the discussion of w_{oT1} or w_{uo} is the same, so only w_{oT1} is given here as an example.

Let the function $f(w_{oT1}, w_{uo}, w_{uT1}) = b_1$, the change amount of w_{oT1} is Δw_{oT1}, the change amount of w_{uo} is Δw_{uo}, and the change amount of w_{uT1} is $\Delta w_{uT1} = 0$; then we have $f(w_{oT1} + \Delta w_{oT1}, w_{uo} + \Delta w_{uo}, w_{uT1} + \Delta w_{ou}) = b_2$.

Then,

$$\lim_{\substack{\Delta w_{oT1} \to 0 \\ \Delta w_{uo} \to 0}} \frac{\left| f\left(w_{oT1} + \Delta w_{oT1}, w_{uo} + \Delta w_{uo}, w_{uT1} + \Delta w_{ou}\right) - f\left(w_{oT1}, w_{uo}, w_{uT1}\right)\right|}{\sqrt{\Delta w_{oT1}^2 + \Delta w_{uo}^2}} = \lim_{\substack{\Delta w_{oT1} \to 0 \\ \Delta w_{uo} \to 0}} \frac{|b_2 - b_1|}{\sqrt{\Delta w_{oT1}^2 + \Delta w_{uo}^2}} = 0. \quad (21)$$

It can be obtained from the formula above that the user's behavior will not be affected when w_{oT1} is changed.

Analysis 2. Because there is no direct information representing the relationship ⟨users, $T1$⟩ in the known information, the choice of a different $T1$ will lead to different user behavior. This is the value of the existence of $T1$. In the same way as for Analysis 1, the concept of derivatives can still be used to explain why changes in $w_{u T_1}$ have an impact on the user

behavior. Because $T1$ is related to the product and the product is related to the user, the changes of w_{uT1} will lead to changes in w_{oT1} and w_{uo} at the same time.

If $\Delta w_{uT1} = x_0 \neq 0$, the changes of w_{uT1} will lead to the changes of w_{oT1} and w_{uo} at the same time.

Then, $f(w_{oT1} + \Delta w_{oT1}, w_{uo} + \Delta w_{uo}, w_{uT1} + \Delta w_{uT1}) = b_3$. Hence,

$$\lim_{\substack{\Delta w_{oT1} \to 0 \\ \Delta w_{uo} \to 0 \\ \Delta w_{uT1} \to x_0}} \frac{\left| f\left(w_{oT1} + \Delta w_{oT1}, w_{uo} + \Delta w_{uo}, w_{uT1} + x\right) - f\left(w_{oT1}, w_{uo}, w_{uT1}\right)\right|}{\sqrt{\Delta w_{oT1}^2 + \Delta w_{uo}^2 + \Delta w_{uT1}^2}} = \lim_{\substack{\Delta w_{oT1} \to 0 \\ \Delta w_{uo} \to 0 \\ \Delta w_{uT1} \to x_0}} \frac{|b_3 - b_1|}{\sqrt{\Delta w_{oT1}^2 + \Delta w_{uo}^2 + \Delta w_{uT1}^2}}$$

$$= \frac{|b_3 - b_1|}{|x_0|} \neq 0. \quad (22)$$

It can be concluded from the above formula that changes in w_{uT1} will have impacts on the user behavior factor. Thus, the changed parameter in this study was $w_{u T_1}$.

2.4.2. Diversity Analysis of the Triangle Model. This algorithm introduces the "third dimension" to the traditional two-dimensional recommendation algorithm and proposes a method based on the triangular area to improve the diversity of the recommendation algorithm. In that case, the question arises: Is the diversity performance getting better with the increase of dimensions? To this end, this section of the manuscript presents an analysis using the "squeeze theorem" that is used to determine whether limits exist and demonstrates the method of equation-solving in linear algebra. The deterministic problem of diversity in two dimensions, four dimensions, and higher dimensions is analyzed.

Theorem 1. *The triangular recommendation algorithm has better diversity performance than the two-dimensional recommendation algorithm.*

Proof. The existing recommendation algorithms have introduced various methods from various fields and then the bipartite graph is used to describe how to make recommendations for the target users. These algorithms mainly focus on recommendation accuracy and have poor performance in diversity and novelty. When so implemented, the most popular recommendation algorithm does not have higher accuracy but it has better performance in the diversity and novelty [27]. With the following assumptions, the two dimensions can be mapped onto a linear function:

$$y = a * \max\{x_1, x_2, \ldots, x_n\} \pm b \quad (a \geq 0, \ b \geq 0). \quad (23)$$
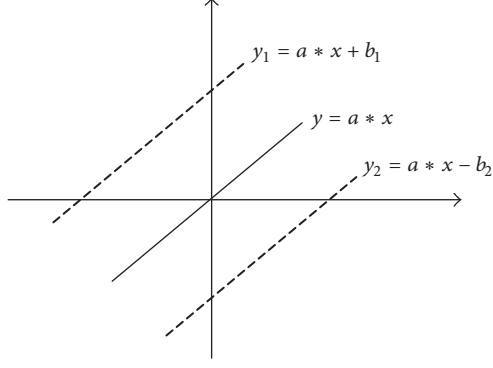
□

FIGURE 3: The schematic represents the contents in (23). In the traditional two-dimensional relationship, the recommendation accuracy dominates and the objects between $y_1 = a * x + b_1$ and $y_2 = a * x - b_2$ are recommended to the user.

FIGURE 4: w_{ij} represents the relationship weights between two points; the larger the value, the stronger the relationship, where S_k $(k = 1, 2)$ represents the areas of the top and bottom triangles.

The y and x_i values represent the user and the object, respectively. When the slope a (indicating the strength of the relationship between the user and the object) is larger, indicating that the user is more interested in this object, the object x_i will be recommended to the user. Thus, the recommendation accuracy is improved. As the L objects with large y values cannot be distributed regularly on a line, they need to be translated for b_1 and b_2 units, as shown in Figure 3.

Thus, at least L objects have the shortest distance in the space with distance of D, and then the L objects are recommended to target users, where D is expressed as follows:

$$D = \frac{b_1 + b_2}{\sqrt{1 + a^2}}. \tag{24}$$

If the "third dimension" is introduced, it is equivalent to adding an unknown parameter in (23) as

$$y = a * \max\{x_1, x_2, \ldots, x_n\} + c * z \pm b$$
$$(a \geq 0, \ b \geq 0). \tag{25}$$

In a two-dimensional relationship, the parameters (y, x) are bound, and z is a free variable. This is similar to the solving of multivariate linear equations system in linear algebra, so a suitable z value can be found in the three-dimensional space to satisfy (25), and there are many values that satisfy the condition. This is the reason for better diversity performance after introducing the "third dimension."

Theorem 2. *Four-dimensional or even higher dimensional recommendation algorithms can be represented by the triangle recommendation algorithm.*

Proof. Assume a diversity recommendation algorithm is constructed with user, object, T_1, and fourth dimension T_2. It is clear that no matter how many dimensions are added, the new added dimension must be closely linked with the object. Then the situations illustrated in Figure 4 should be considered to illustrate all possible distributions of the four dimensions. Of
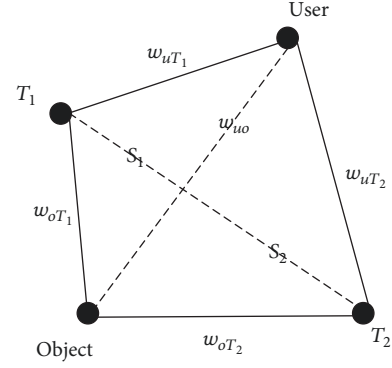
course, the area of the constructed graph is still used under each possible state in order to make recommendations. The following work is illustrated with Figure 4. □

As the quadrilateral formed by four points varies, that is, it may not necessarily be a regular figure, so the area of the quadrilateral is denoted by S, and it can be represented by

$$S = S_1\left(w_{uT_1}, w_{oT_1}, w_{uo}\right) + S_2\left(w_{uo}, w_{oT_2}, w_{uT_2}\right). \tag{26}$$

It is not difficult to see that w_{uo} is the most important weighting in S_1, S_2, and S. It is also the research focus of the existing two-dimensional recommendation algorithm. At this point, two cases of S_1, S_2 should be considered:

(1) When $S_1 = S_2$, for the target user, only one extra dimension is needed. Either T_1 or T_2 can be chosen and only three dimensions are needed.

(2) When $S_1 < S_2$ or $S_1 > S_2$, S is the sum of the two dimensions (in the three-dimensional relationship). The recommendation list of target users is the only obviously affected one of the new dimension. Therefore, another new dimension is redundant and only three dimensions are needed.

To sum up, in the two cases discussed above, the "squeeze theorem" can be used to illustrate the importance of adding a "third dimension" into the two-dimensional relationships. In other words, the purpose is to improve the diversity of the recommendation, but the accuracy of the recommendation cannot be abandoned completely. In the two-dimensional relationship, it is difficult to achieve good recommendation diversity; in the four-dimensional relationship, as illustrated by the discussion above, it is not necessary to have the fourth dimension. Hence, all cases for higher dimensions can be transformed into the research of several three-dimensional relationships and there is no need for higher multidimensional models. In the three-dimensional relationship, it not only retains the recommendation accuracy in the two-dimensional relationship but also improves the recommendation diversity. Thus, the method proposed in this paper, which is adding the third dimension to the two dimensions, is effective in improving diversity.

TABLE 1: The basic statistics for two real online rating datasets.

Data	Users	Objects	Links	Sparsity
MovieLens-100K	943	1574	82520	$5.56 * 10^{-2}$
MovieLens-1M	6039	3628	836478	$3.82 * 10^{-2}$

3. Data and Evaluation Indices

3.1. Data Description. In the research of traditional recommendation algorithms, there has never been a research method working with three dimensions. Therefore, this paper has difficulties in the data collection. Finally, two commonly used real datasets were selected for the purposes of this paper, which were the MovieLens-100K dataset and the MovieLens-1M dataset. These often are used as the database for different recommendation algorithms. The MovieLens datasets were provided by the Group Lens project at the University of Minnesota. The dataset uses a 5-point rating, and the higher the score, the better the data. In constructing the bipartite model, only data with a rating greater than or equal to 3 were considered. After coarse-graining, the small dataset contains 82520 links and the large one contains 836478 links. It should be noted that the bipartite model has no weighting in the following analysis. In other words, the rating is ignored. The basic statistics of the dataset are shown in Table 1.

3.2. Evaluation Indices. In order to evaluate the performance of recommendation algorithms in practical applications, cross-validation usually is used to assess how the results are extended to independent datasets [30]. A cross-validation process involves partitioning the datasets into complementary subsets, the training set, and the testing set. The training set is used in the recommendation algorithm, and the recommendation results are obtained. The testing set then is used to verify the result. In the following experiment, a 10-fold cross-validation strategy was used to evaluate the performance of the proposed algorithm in each independent realization. In particular, the user's evaluation (i.e., at the edges of the bipartite graph) is divided randomly into 10 equal length subsamples and the partitioning process is independent of the user and the object. Then, one of the 10 subsamples is retained to test the performance of the recommendation algorithm. The remaining 9 subsamples are combined and used as a training set for the recommendations. In other words, 90% of the entire dataset is used for recommendation algorithms, and the remaining 10% is used to evaluate the recommendation results. Specifically, the 10-fold cross-validation process is repeated 10 times; that is, each of the 10 subsamples will be used as a testing set. Finally, the 10 test results are averaged to obtain a single result, which is used as the basis for evaluating the performance of the algorithm.

In previous literature, extensive research has been done on how to evaluate the performance of recommendation algorithms [31, 32]. In this study, seven widely used indices were used to evaluate the performance of the recommendation algorithm, including four accuracy indices (AUC, MAP, Precision, and Recall), two diversity indices (Intrasimilarity and Intersimilarity), and a novelty index (Popularity). The following is a brief overview of the seven indices.

Accuracy is one of the important indices for evaluating the quality of recommendation algorithms. First, the AUC (area under the ROC curve) [33] is introduced. The AUC value can be interpreted as the probability that a randomly chosen collected object is ranked higher than a randomly chosen uncollected object. In the calculation of AUC, a side randomly selected from the testing set is compared with a nonexistent randomly selected side. If the weight of side in the testing set is larger than that of the nonexistent side, then 1 point is added. If the two values are equal, then 0.5 points are added. After N times independent comparisons, if the side value in the testing set is higher than that of the nonexistent side for N' times, and the two values are equal for N'' times, then the AUC can be defined as

$$\text{AUC} = \frac{1}{m}\sum_{i=1}^{m}\frac{\left(N' + 0.5N''\right)}{N}. \tag{27}$$

The greater the of AUC value, the higher the accuracy of the algorithm.

Then, three L-dependent accuracy indexes are introduced, namely, the MAP (Mean Average Precision) index [34], the Precision index, and the Recall index [35]. The MAP is a standard rank-aware measure of the overall ranking accuracy in the field of information retrieval, which is similar to the average ranking score [31, 32]. The average precision for user i is defined as

$$\overline{P}_i(L) = \frac{1}{D(i)}\sum_{s=1}^{d_i(L)}\frac{s}{r_s}. \tag{28}$$

Here, $D(i)$ represents the number of objects in the testing set; $d_i(L)$ represents the number of common objects of the testing set and the recommendation list with length of L. r_s represents the ranking of the sth object in the recommendation list, and $r_s \in [1, L]$. Then, the MAP index can be obtained by averaging $\overline{P}_i(L)$ of all users:

$$\text{MAP} = \frac{1}{m}\sum_{i=1}^{m}\overline{P}_i(L), \tag{29}$$

where m represents the number of users. A larger MAP index corresponds to a better overall accuracy ranking.

The Precision index is defined as the ratio of the number of common objects that appear in the recommended list and testing set to the length of the recommendation list. For all the users, the average Precision is defined as

$$P(L) = \frac{1}{m}\sum_{i=1}^{m}\frac{d_i(L)}{L}. \tag{30}$$

Recall is also defined as a ratio

$$R(L) = \frac{1}{m}\sum_{i=1}^{m}\frac{d_i(L)}{D(i)}. \tag{31}$$

The higher the values of Precision and Recall, the higher the recommendation accuracy.

In the personalized recommendation algorithm, diversity is an important index to evaluate the diversity of the objects recommended by the algorithm. As it is difficult to obtain external sources of object similarity information, the diversity measure usually is based on the evaluation matrix. The Intersimilarity is one of the widely used diversity indices and it can be quantified by Hamming distance [36]. The average Hamming distance value of all users can be defined as

$$H(L) = \frac{1}{m(m-1)} \sum_{i=1}^{m} \sum_{j=1}^{m} \left(1 - \frac{C(i,j)}{L} \right), \qquad (32)$$

where $C(i,j) = |o_i^L \cap o_j^L|$ represents the number of common objects in the recommendation list of user i and user j. The greater the Hamming distance value, the higher the diversity. Intrasimilarity is another similarity index [37]; in other words, it is measured by the cosine similarity between objects in the recommendation list of target user. Mathematically, the average Intrasimilarity [38] of all users is defined as

$$I(L) = \frac{1}{m(m-1)} \sum_{i=1}^{m} \sum_{o_\alpha, o_\beta \in o_i^L, \alpha \neq \beta} S_{\alpha\beta}^{Cos}, \qquad (33)$$

where $S_{\alpha\beta}^{Cos}$ is the cosine similarity between the object α and object β in the recommendation list o_i^L of user i with length of L. The smaller the Intrasimilarity value, the higher the recommendation diversity.

Novelty is an important index, which aims to quantify the ability of an algorithm to produce novel (i.e., unpopular or unwelcome) and unexpected results. Here, novelty is quantified using the average popularity of the recommended object. It can be defined as

$$N(L) = \frac{1}{mL} \sum_{i=1}^{m} \sum_{o_\alpha \in o_i^L} k_\alpha, \qquad (34)$$

where k_α is the degree of the object α in the recommendation list o_i^L of user i. Lower popularity indicates higher novelty and, potentially, better user experience.

4. Experiments and Results

The recommendation algorithm based on triangle area was applied to two real online rating datasets. In order to facilitate the comparison, some existing recommendation algorithms also were considered, including global ranking (GR) [31], user-based collaboration filtering (UCF) [14], item-based collaboration filtering (ICF) [14], mass diffusion (MD) [22], heat conduction (HC) [23], and CosRA-based filtering [38].

In GR, from the perspective of item degree, all items are sorted in descending order according to their degrees and then items with high degrees are recommended to the target user [31] at one time. In UCF, the recommendation list of the target user is obtained by analyzing the historical behavior of other users with similar interests and preferences [39]. Similarly, in ICF, by analyzing the past preferences of the target user, the similarity of the items is obtained and

the recommendation list then can be obtained [40]. MD can be viewed as the resource allocation process in the user-object network [41, 42]. The CosRA-based index combines the cosine index with the RA index and a new similarity index, CosRA, is generated. Through the two steps of resource allocation processes, the uncollected items are sorted in descending order according to the size of the resource and recommended to the user from large to small [38]. The realization details of the previous five recommendation algorithms can be found in the survey report [9].

The results of the seven evaluation metrics are shown in Table 2. As the Intrasimilarity value has good performance in both of the datasets, 4 decimal places were retained for the seven evaluation standards to better observe the change of the Intrasimilarity value in the two datasets. This study focused mainly on improvement of recommendation diversity and the proposed recommendation algorithm based on triangular area (TR) has better performance than other algorithms. This is discussed in two parts as shown in Tables 2 and 3. First, comparing TR with the 5 recommendation algorithms other than HC, it can be observed that TR has the highest Hamming distance value, the lowest Intrasimilarity value, and the highest novelty value. Although TR does not perform as well as other recommendation algorithms in the four accuracy metrics, it has the best performance in the three metrics that measure diversity and novelty.

By comparing TR and HC with other recommendation algorithms, it is evident that these two algorithms have better performance than other algorithms in terms of diversity and novelty. However, the performances of these two algorithms are inferior to other algorithms in terms of accuracy. Of the seven evaluation criteria, for the dataset MovieLens-100K, HC is only better than TR in terms of AUC value and the novelty value. TR is better than HC from the perspective of accuracy. For the dataset MovieLens-1M, the performance difference of HC and TR in diversity and novelty is not significant. However, the performance difference in accuracy is very evident. This is because HC studies the similarity between items from the perspective of degree and decides how to allocate resources accordingly. This is helpful in improving accuracy. There is no effective way to achieve the best performance in all the three aspects of accuracy, diversity, and novelty at the same time [27]. The latest study found that although the most popular method did not achieve very high accuracy [27], its diversity performance is outstanding.

From Tables 2-3, we can see intuitively that the performance of TR is much better than that of other recommendation algorithms in the two aspects of diversity and novelty.

5. Summary and Future Work

In this study, research of the recommendation algorithm differs from the traditional two-dimensional recommendation algorithm. This was the research focus of the investigation. Due to the rapid development of the Internet, the volume of available information has increased dramatically. Thus, information overload has become an unavoidable problem, which needs to be solved urgently. If too much attention is paid to the accuracy of the recommendation, it will bring

TABLE 2: The values of the seven evaluation metrics after applying the different recommendation algorithms on the two datasets.

	AUC	MAP	P	R	H	I	N
MovieLens-100K							
GR	0.863	0.208	0.058	0.358	0.395	0.408	255
UCF	0.887	0.315	0.070	0.476	0.550	0.394	242
ICF	0.888	**0.385**	0.073	0.494	0.674	0.413	211
MD	0.898	0.325	0.075	0.527	0.618	0.355	230
CosRA	**0.908**	0.380	**0.082**	**0.575**	0.724	0.335	204
TR	0.6105	0.0482	0.0446	0.3196	**0.8862**	**0.0024**	**138**
MovieLens-1M							
GR	0.856	0.144	0.053	0.222	0.403	0.415	1660
UCF	0.872	0.176	0.061	0.263	0.458	0.415	1640
ICF	0.885	**0.289**	0.072	0.314	0.629	0.404	1445
MD	0.885	0.188	0.066	0.297	0.504	0.403	1618
CosRA	**0.895**	0.223	**0.074**	**0.350**	0.598	0.387	1541
TR	0.5717	0.0447	0.0334	0.2940	**0.8043**	**0.0017**	**531**

Notes. The recommendation length is defined as $L = 50$. All the seven values are the average values of 10 independent realizations. The bold values indicate the best performance of the recommendation algorithm on a metric.

TABLE 3: The performance comparison between TR and HC in seven evaluation metrics.

	AUC	MAP	P	R	H	I	N
MovieLens-100K							
HC	**0.842**	0.037	0.021	0.123	0.858	0.056	23
TR	0.6105	**0.0482**	**0.0446**	**0.3196**	**0.8862**	**0.0024**	138
MovieLens-1M							
HC	**0.881**	0.052	0.034	0.162	**0.861**	0.045	**198**
TR	0.5717	0.0447	0.0334	**0.2940**	0.8043	**0.0017**	531

trouble for users, and the users may feel that a recommendation is unnecessary. Therefore, the focus of the present study was recommendation diversity. A "third dimension" was introduced alongside the two normal dimensions, and the relationship between any of the two dimensions in the three-dimensional model is studied and quantified. The three dimensions then were treated as three sides of a triangle, and each triangle corresponded to an object. The area of the triangle was calculated using Heron's formula. Finally, the recommendation list for target users was obtained according to the triangle area in descending order. In the second part of the investigation, the effectiveness of the three-dimensional recommendation was illustrated by the method of solving multivariate equations and application of the "squeeze theorem" that is used to decide whether a limit exists. The experimental results showed also that the proposed algorithm performs well in terms of recommendation diversity.

In the realization of the proposed recommendation algorithm, user behaviors are clustered and processed by combining the K-mean and Markov chain in a user-object relationship. This makes the algorithm relatively complex. In future work, one of the focuses will be to reduce the time complexity while keeping the effectiveness of the technique and the recommendation performance of the model will be evaluated from the perspective of time as reference [21]. Due to the limited number of commonly used public datasets that are available to construct the triangle model, the experiments in this paper only used the MovieLens dataset. It is planned that more test data will be collected to validate the model. Furthermore, an investigation of the balance of accuracy and diversity will be helpful and is the anticipated focus of future work.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] K. C. Laudon, C. G. Traver, and A. V. R. Elizondo, *E-Commerce*, Pearson - Addison Wesley, 2007.

[2] Q.-M. Zhang, A. Zeng, and M.-S. Shang, "Extracting the information backbone in online system," *PLoS ONE*, vol. 8, no. 5, Article ID e62624, 2013.

[3] J. B. Schafer, J. Konstan, and J. Riedl, "Recommender systems in e-commerce," in *Proceedings of the 1st ACM Conference on Electronic Commerce*, pp. 158–166, USA, November 1999.

[4] C. Porcel and E. Herrera-Viedma, "Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries," *Knowledge-Based Systems*, vol. 23, no. 1, pp. 32–39, 2010.

[5] C. He, D. Parra, and K. Verbert, "Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities," *Expert Systems with Applications*, vol. 56, pp. 9–27, 2016.

[6] G.-D. Xu, Z. Wu, Y. Zhang, and J. Cao, "Social networking meets recommender systems: survey," *International Journal of Social Network Mining*, vol. 2, no. 1, pp. 64–100, 2015.

[7] F. Ricci, L. Rokach, and B. Shapira, "Recommender systems: introduction and challenges," in *Recommender Systems Handbook*, pp. 1–34, Springer, 2015.

[8] G. Chen, T. Gao, X. Zhu, H. Tian, and Z. Yang, "Personalized recommendation based on preferential bidirectional mass diffusion," *Physica A: Statistical Mechanics and its Applications*, vol. 469, pp. 397–404, 2017.

[9] L. Lü, M. Medo, C. H. Yeung, Y. Zhang, Z. Zhang, and T. Zhou, "Recommender systems," *Physics Reports*, vol. 519, no. 1, pp. 1–49, 2012.

[10] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Systems*, vol. 46, pp. 109–132, 2013.

[11] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, 2005.

[12] A. Zeng, A. Vidmer, M. Medo, and Y.-C. Zhang, "Information filtering by similarity-preferential diffusion processes," *EPL (Europhysics Letters)*, vol. 105, no. 5, Article ID e58002, 2014.

[13] M. Gan, "Walking on a user similarity network towards personalized recommendations," *PLoS ONE*, vol. 9, no. 12, Article ID e114662, 2014.

[14] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th International Conference on World Wide Web (WWW '01)*, pp. 285–295, New York, NY, USA, 2001.

[15] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, vol. 35, no. 12, pp. 61–70, 1992.

[16] M. J. Pazzani and D. Billsus, *Content-Based Recommendation Systems*, Adaptive Web Springer-Verlag, 2007.

[17] R. V. Meteren, Using Content-Based Filtering for Recommendation, 2007.

[18] P. Lops, M. D. Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in *Recommender Systems Handbook*, pp. 73–105, 2011.

[19] M. E. Newman, "The structure and function of complex networks," *SIAM Review*, pp. 167–256, 2006.

[20] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 99, no. 12, pp. 7821–7826, 2002.

[21] A. Vidmer and M. Medo, "The essential role of time in network-based recommendation," *EPL (Europhysics Letters)*, vol. 116, no. 3, Article ID 30007, 2016.

[22] Y.-C. Zhang, M. Medo, J. Ren, T. Zhou, T. Li, and F. Yang, "Recommendation model based on opinion diffusion," *EPL (Europhysics Letters)*, vol. 80, no. 6, Article ID 68003, 2007.

[23] Y. C. Zhang, M. Blattner, and Y. K. Yu, "Heat conduction process on community networks as a recommendation model," *Physical Review Letters*, vol. 99, no. 15, Article ID 154301, 2007.

[24] R. Burke, "Hybrid recommender systems: survey and experiments," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, 2002.

[25] T. Tran and R. Cohen, *Hybrid recommender systems for electronic commerce*, 2000.

[26] A. Lelu, *Hybrid Recommender Systems*, 2009.

[27] F. Yu, A. Zeng, S. Gillard, and M. Medo, "Network-based recommendation algorithms: a review," *Physica A: Statistical Mechanics and its Applications*, vol. 452, pp. 192–208, 2016.

[28] S. M. McNee, J. Riedl, and J. A. Konstan, "Being accurate is not enough: how accuracy metrics have hurt recommender systems," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI EA '06)*, pp. 1097–1101, Montreal, Canada, April 2006.

[29] M. Zhang and N. Hurley, "Avoiding monotony: Improving the diversity of recommendation lists," in *Proceedings of the 2008 2nd ACM International Conference on Recommender Systems (RecSys '08)*, pp. 123–130, ACM, Switzerland, October 2008.

[30] G. McLachlan, K. A. Do, and C. Ambroise, *Analyzing Microarray Gene Expression Data*, John Wiley and Sons, 2005.

[31] T. Zhou, J. Ren, M. Medo, and Y. Zhang, "Bipartite network projection and personal recommendation," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 76, no. 4, Article ID 046115, 2007.

[32] J.-G. Liu, K. Shi, and Q. Guo, "Solving the accuracy-diversity dilemma via directed random walks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 85, no. 1, Article ID 016118, 2012.

[33] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[34] T. Y. Liu, J. Xu, T. Qin, W. Xiong, and H. Li, "Letor: Benchmark dataset for research on learning to rank for information retrieval," in *Proceedings of SIGIR2007 Workshop on Learning to Rank for Information Retrieval (LR4IR '07)*, pp. 3–10, ACM, Amsterdam, The Netherlands, 2007.

[35] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information and System Security*, vol. 22, no. 1, pp. 5–53, 2004.

[36] T. Zhou, L. L. Jiang, R. Q. Su, and Y. C. Zhang, "Effect of initial configuration on network-based recommendation," *EPL (Europhysics Letters)*, vol. 81, no. 5, Article ID 58004, 2008.

[37] K. Choi and Y. Suh, "A new similarity function for selecting neighbors for each target item in collaborative filtering," *Knowledge-Based Systems*, vol. 37, no. 1, pp. 146–153, 2013.

[38] L. J. Chen, Z. K. Zhang, J. H. Liu, J. Gao, and T. Zhou, "A vertex similarity index for better personalized recommendation," *Physical A Statistical Mechanics and Its Applications*, vol. 466, pp. 607–615, 2017.

[39] R.-R. Liu, C.-X. Jia, T. Zhou, D. Sun, and B.-H. Wang, "Personal recommendation via modified collaborative filtering," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 4, pp. 462–468, 2009.

[40] T. Zhou, R. Q. Su, R. R. Liu, L. L. Jiang, B. H. Wang, and Y. Zhang, "Accurate and diverse recommendations via eliminating redundant correlations," *New Journal of Physics*, vol. 11, Article ID 123008, 2009.

[41] T. Zhou, Z. Kuscsik, J. Liu, M. Medo, J. R. Wakeling, and Y. Zhang, "Solving the apparent diversity-accuracy dilemma of recommender systems," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 107, no. 10, pp. 4511–4515, 2010.

[42] Q. Ou, Y.-D. Jin, T. Zhou, B.-H. Wang, and B.-Q. Yin, "Power-law strength-degree correlation from resource-allocation dynamics on weighted networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 75, no. 2, Article ID 021102, 2007.