Research Article

A Novel Recommendation Algorithm Based on Diffusion of Innovation Theory

Liang Zhang[1], Xuesheng Qian[1], Ping Lv[1*] and Xue Zhou[2]

[1]*School of economics and management, Guizhou Normal University, Guiyang 550001, China*
[2]*School of management, University of Chicago, Chicago, IL 60637, United States*

_____

Abstract

Determining whether an item is "novel" for target users is a crucial problem and understanding their preference and awareness degree of the item has been widely investigated in novelty recommendations. A recommendation algorithm based on "diffusion of innovation" theory was proposed in this study to verify the novelty of recommendation results on the precondition of ensuring their accuracy. Items were clustered using the K-means clustering algorithm. Users' positivity of innovation adoption for each item was calculated on the basis of the items adopted by the users, and then the novelty of the items for target users was calculated by combining their popularity change, user preference, and user difference. Results of traditional recommendation algorithms were integrated for recommendation on the basis of fusion strategy results. The effectiveness of the proposed algorithm was verified through an offline experiment. Results indicate that the novelty of the recommendation list of the proposed algorithm is remarkably higher than that of traditional algorithms. The novelty is high when the quantity of alternative sets reaches 400 or more, where the average popularity of the recommendation list declines by 40%, and the coverage is elevated by 100%, thereby improving the ability of the proposed system to extract all kinds of items. This study serves as reference for the improvement of user satisfaction with recommendation systems.

Keywords: innovation diffusion, novelty, recommendation algorithm, clustering

_____

1. Introduction

A recommendation system serves as a compass for us to travel in complex conceptual spaces [1] and plays a significant role in remitting information overload, shortening user search time, and improving individualized experience by recommending articles of interest to users in accordance to their historical behavior. Commonly used recommendation algorithms mainly include collaborative filtering (CF) and article feature-based recommendation. Most of these algorithms only focus on the similarities of purchase history or behavior of users in the recommendation process, thereby resulting in the redundancy of recommendation results and serious homogeneity problem and making it difficult to bring users into contact with fresh and diversified contents, lowering users' satisfaction [2, 3]. High accuracy of recommendation systems will result in diversity and novelty degradations, thereby inducing problems, such as homogeneity, polarization, echo chamber, and wrong information [4, 5]. Novelty metrics have attracted much attention and have been widely investigated because of their high correlation with user satisfaction. However, most novelty recommendation algorithms have resulted in accuracy degradation [6]. Novelty recommendation requires that the recommended items should appear as "new" items for target users. Thus, ensuring the novelty and accuracy of recommendation systems remains difficult.

Most scholars have focused on user preferences with regard to user modeling, and few researchers have modeled users' adoption attitudes and behavior for new items. "Diffusion of innovation" (DI) theory divides innovation adopters into five classes on the basis of time dimension, where the users at early time have high positivity in adopting innovations, and those at later time are conservative. Therefore, the term "new" embodied in novelty is a kind of user perception, and different users have different positivity of innovation adoption. Thus, exclusively ensuring the novelty of recommendation systems by excluding popular products will certainly sacrifice their accuracy.

This study presented a users' positivity of innovation adoption (PIA) model by calculating the product life cycle where their adopted items are located. The variation tendency of item popularity was combined to provide recommendations for target users in accordance to their PIA to ensure the novelty without influencing the accuracy of recommendation list.

2. State of the art

Novelty recommendations have gradually attracted increasing attention in terms of their accuracy in recent years. Yu Hong found that the newly recommended items with extremely low popularity are not known and novel by the majority of users and measured the novelty of recommendation results on the basis of the proportion of new items occupied in the recommendation list [7]. Chou S. Y. et al. defined the percentage occupied by artists, who are known by users, as the novel music recommendation list [8]. The abovementioned studies have only evaluated the overall

novelty of recommendation results and ignored the novelty of items. In terms of scoring time, Chen Lingjiao et al. introduced the concepts of "renovator" and "potential follower" and believed that the items rated by renovators are novel for target users [9]. Kapoor K. et al. deemed that the novel items for a system are unrated and newly added items that are unpopular or forgotten by users [10]. Considering that item popularity is easy to calculate, commodities with low popularity have high novelty [11, 12]. However, global values cannot reflect user interests or represent the novelty of the same item for different users. Wu Hao et al. used parameters to combine random walk with heat conduction theory to improve accuracy and novelty on the basis of a user–item bipartite graph [13]. Wang Bin et al. proposed a discovery-based user–item relation recommendation model and compared it with traditional k-nearest neighbor and min-min roughness-based classical algorithms in terms of improvement to effectively improve the novelty and diversity of recommendation lists [14]. Yu Qian et al. presented the association between users and communities in a neighborhood and a user–community distance measurement method to calculate the novelty of candidate communities and provide novel community recommendations, thereby improving the accuracy of recommendation results [15]. Pietro Gravino et al. introduced the concept of "adjacent possible" to redesign a recommendation system to meet user demands [16]. For users, novelty is a "new" thing and differs from known things, making it a kind of user perception [17]. K. G. Saranya et al. used maximum cosine and average cosine distances to measure the distance of new documents to those already known by users for verifying their novelty [18]. Wael Alkhatib et al. used a deep semantic similarity model to implicitly measure the semantic similarity between user interest and recommended available resources for the recommendation of novel learning resources [19]. Jorge Diez et al. optimized the novelty and diversity of recommendation list through factorization [20]. M. Kminkas and D. Bridge analyzed the differences and associations of four metrics, namely, diversity, surprise, novelty, and coverage, through several experiments on the basis of their definitions and a summary of optimal technologies. They proposed multiple optimal strategies to improve the four metrics by analyzing their influences on accuracy [21]. A. Pathak et al. introduced the characteristic space-based diversification technology by investigating the associations of diversity and novelty principles of information retrieved using recommendation systems [22]. FM. M. et al. presented theme association attribute random forest and diversity and novelty improvement algorithms for implicit tag recommendation by integrating three aspects, namely relevance of tag suggestions, diversity, and novelty of explicit themes [23].

Few of the above studies have explored the acceptance level and process of new things. DI theory defines "innovation" as a new concept, new approach or new thing [24]. The term "new" refers to the perception of receivers. Although the process of consumers adopting new products cannot be fully perceived, their attitudes and time in adopting new products can be analyzed through historical data, indicating that novel items can be recommended to users at an appropriate time. The K-means algorithm was used in this study to cluster item sets and model innovation adoption attitudes and behavior of users through their historical data. The variation trend of popularity of items to be recommended was combined to calculate novelty. Items

with maximum novelty were chosen from favorite alternative item sets of users for recommendation to ensure the novelty and accuracy of recommendation systems.

The remainder of this study is organized as follows. Section 3 introduces the users' PIA model and novelty recommendation algorithm based on DI theory. Section 4 verifies the change degrees of metrics, such as accuracy and novelty, of the proposed algorithm through an offline experiment. Section 5 summarizes the conclusions.

3. Methodology

3.1 DI and product life cycle
Rogers indicated that consumer decision making in innovation adoption mainly undergoes five stages, namely, awareness, interest, evaluation, trial use, and adoption (Fig. 1). A consumer becomes aware of a new product but lacks related information. The consumer will seek for related detailed information when he/she is interested in this new product and then evaluates whether to use it on the basis of the obtained information. The trial effect is the determining factor whether the consumer will purchase the new product or not [24].
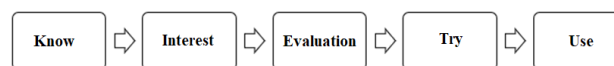

Fig. 1. Consumer decision-making process in innovation adoption

Consumers obviously have different attitudes toward new products that they are about to use. In specific product fields, some people tend to become "consumer leaders," whereas some consumers start using new products at a later time. Rogers classifies innovation adopters into five types, which present a normal distribution, as shown in Fig. 2. Fig. 3 displays the histogram of average evaluation time of users in the MovieLens dataset, which presents a clock-shaped distribution.

The theory of product life cycle is a significant decision-making basis and method for enterprise production and sales management. As shown in Table 1, the products in different life cycle phases have different market statuses, where marketing objectives and strategies must be adopted. Product diffusion in the market varies from life cycle phases that they are located because of the marketing activities of enterprises and word-of-mouth spreading between consumers.
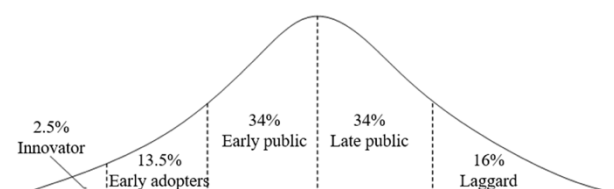

Fig. 2. Classification of innovation adopters

From the angle of recommendation system, innovators should be recommended with items in the introduction phase and early adopters with items in the growth phase. The possibility for early adopters to adopt novel items and the accuracy of the recommendation system are low when these items are recommended in the introduction phase. The early adopter is likely to adopt a mature item when it is recommended to ensure accuracy, thereby degrading the novelty of the recommendation system and making the recommendation invalid. Therefore, the corresponding

recommendation based on the classification of innovation adopters among users cannot only ensure relative novelty for target users but also increases the willingness of users to accept for ensure the novelty and accuracy of the recommendation system.
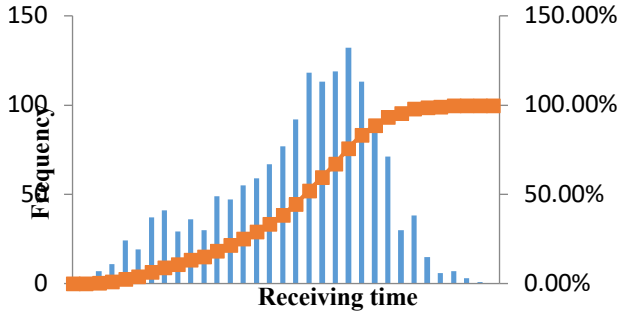

Fig. 3. Time distribution of user innovation adoption in the MovieLens dataset

Table 1. Characteristics, objectives, and marketing strategies in different life cycle phases of products

Item	Introduction phase	Growth phase	Mature phase	Decline phase
Sale	Low sales volume	Intense growth of sales volume	Sales peak	Sales decline
Cost	High cost per customer	General cost per customer	Low cost per customer	Low cost per customer
Profit	Loss	Profit growth	High profit	Profit decline
Customer	Innovator	Early adopters	The public	Laggards
Competitors	Very few	Increasing	Stable with a slight decline	Reducing
Marketing objectives				
	Establish product awareness and elevate product utilization rate	Maximize market shares	Protect market shares and strive for maximum profit	Reduce the expenditure and squeeze brand value

3.2 PIA

Rogers classifies innovation adopters into five categories on the basis of time dimension, that is, the users adopting earlier have higher PIA and those adopting later are more conservative. Therefore, users' positivity in adopting a new product can be measured by calculating the time difference when users adopt the new product and when the product enters the market, as expressed in Formula 1, where *PIA(i,u)* is the PIA of user *u* for item *i*, and T_{iu} is the time when user *u* adopts item *i* and T_i is the time when item *i* enters the market.

$$PIA(i,u) = \log(2 + T_{iu} - T_i) \tag{1}$$

In terms of a recommendation system, whether the to-be-recommended item accords with the user's *PIA* should be predicted. Thus, the user's *PIA* for a to-be-recommended item must be known. However, the recommendation system is not aware whether the user will adopt the to-be-recommended item or not and cannot directly calculate *PIA (i,u)*. A simple method uses the mean value of user's *PIA* for

already rated items to measure user's *PIA* (Formula 2), where   refers to a set of items already adopted by user u.

$$PIA(u) = \frac{1}{|I_u|} \sum_{i \in I_u} \log(2 + T_{iu} - T_i) \tag{2}$$

Users' PIA is influenced by their internal factors and is a consumptive attitude. Conservative users usually adopt new products slowly, whereas radical users are more willing to try. Using Formula 2 to measure users' *PIA* has certain reasonability. Users' *PIA* for items is closely related to item characteristics. The cognitive degree and *PIA* will be high when a user has high interest in an item. Item classification should be considered in users' *PIA*. $PIA(C_j,u)$ of user u for item class C_j is defined as the mean value of *PIA* of the user for class K items, as shown in Formula 3.

$$PIA(C_j,u) = \frac{1}{|I_u \cap C_j|} \sum_{i \in I_u \cap C_j} \log(2 + T_{iu} - T_i) \tag{3}$$

3.3 K-means clustering-based item classification

3.3.1 K-means clustering
The item sets must be classified to accurately calculate users' PIA. Item classification methods include universal classification, eigenvector-based classification, implicit category model, and clustering. K-means clustering can realize item classification in accordance to user behavior and can reflect users' different consuming behavior and attitudes towards different items. The concrete steps of this method are described as follows:

(1) Initial centroids are first selected, where K is a designated cluster number, and the common approach used is random selection.

(2) Each data point is assigned to the nearest cluster formed by centroids.

(3) The centroids are recalculated on the basis of K formed clusters. The sum of squared error (SSE) is taken as the objective function in centroid selection (Formula 4). For centroid selection, the SSE of the clusters is minimum. The centroid that contributes to minimum SSE of the clusters is the mean value of all data points in the clusters at different dimensionalities (Formula 5), where c is the centroid of cluster C_j, *dis(c,i)* represents the distance from data points *i* to *c*, and c_k and i_k are the centroid and k (th) dimension of data points, respectively. The equations are expressed as follows:

$$SSE = \sum_{j=1}^{K} \sum_{i \in C_j} dis(c,i)^2 \tag{4}$$

$$c_k = \frac{1}{m} \sum_{i \in C_j} i_k \tag{5}$$

(4) Steps (2) and (3) are repeated until the centroid becomes constant.

3.3.2 Determination of cluster number via silhouette coefficient (SC)
As a metric used to evaluate clustering effect, SC integrates two factors, namely, cohesion and separation degrees. For any item i in a data set, the average distance a_i from *i* to all

items in the cluster where it is located is calculated, similar to the average distance from i to all items in other clusters. The minimum b_i value is obtained, and then the SC of item i is defined in Formula 6, which can be expressed as:

$$silouette(i) = \frac{b_i - a_i}{\max(a_i, b_i)} \tag{6}$$

The value range of *silhouette(i)* is (−1,1). Item i is immensely different from the items in other clusters when the value of *silhouette(i)* is one. The classification of item i is not obvious when its value is zero, whereas item i is allocated to a wrong cluster when its value is −1. The SC of each item can be calculated using Formula 6. The overall clustering effect can be measured through the mean value of SCs of all items, as shown in Formula 7. The greater the SC value is, the better the clustering effect will be.

$$SC = \frac{1}{|I|}\sum_{i \in I} silhouette(i) \tag{7}$$

K-means clustering first needs to determine cluster number K. The corresponding SC value should be calculated for the clustering result generated for each K value. Thus, the optimal cluster number can be determined on the basis of the SC value. Clustering should be repeated multiple times for each K value because of the instability of K-means algorithm, and the effect of cluster number K is measured by calculating the average SC value.

All items can be classified into K categories through K-means clustering. However, the items already rated by a user may not include all categories. At the same time, the *PIA* can be measured using the mean *PIA* of all items rated by the user, and $PIA(C_j, u)$ is redefined in Formula 8. Hence, the set of user's *PIAs* for all item categories can be used to describe the *PIA* of this user, which can be expressed as:

$$PIA(C_j, u) = \begin{cases} \frac{1}{|I_u \cap C_j|} \sum_{i \in I_u \cap C_j} \log(2 + T_{iu} - T_i) & I_u \cap C_j \neq \varnothing \\ \frac{1}{|I_u|} \sum_{i \in I_u} \log(2 + T_{iu} - T_i) & I_u \cap C_j = \varnothing \end{cases} \tag{8}$$

3.4 DI-based novelty recommendation algorithm

People will seek for information in different innovative decision-making phases to reduce the uncertainty of expected consequences of an innovation. Different people have different willingness to take risk in accepting new products or new concepts. A minority of people are willing to accept risks of new products or new concepts, whereas most people are conservative and tend to wait for others to make initial trial. As shown in Tab. 1, innovators are recommended with items in the introduction phase, early users with items in the growth phase, the public with items in the mature phase, and laggards with items in the decline phase.

DI theory is actually an ideal state for classification and distribution of innovation adopters. Deviation may occur under the influence of many factors, and product life cycle phases are difficult to demarcate through quantitative analysis. Therefore, the corresponding relationship between previously mentioned innovation adopters and product life cycle phases is simplified, as shown in Fig. 4. Product 1 is included in the market at time T_1, and T_{11}, T_{12}, and T_{13} represent the times when three users adopt this type of

product. The novelty of product 1 is low for user 1, whereas its novelty for user 3 is extremely high when the recommendation system recommends it at time T. However, the possibility for this user to adopt this innovation is extremely low. The novelty and accuracy of product 1 can be guaranteed for user 2. Therefore, user PIA should be considered in measuring the degree in which the user adopts the items. Assume that a user adopts item i at time T, the closer the *PIA (i, u)* for this item to the user *PIA* for the category this item belongs to, the higher the probability that the user will adopt this item.

Measuring the possibility for the user to adopt this item only through prediction of the item popularity will remain the same for all users without considering the user difference in PIA. The DI-based novelty recommendation algorithm (hereinafter abbreviated as DI algorithm) defines the possibility for item i to be adopted by user u, which can be expressed as Formula 9, where i_{n+1} is the popularity prediction for item i, $PIA(i,u)$ is the *PIA* when the user adopts item i at recommendation time, $PIA(C,u)$ is the user *PIA* for items in the category that item i belongs to, and $AP(i,u)$ is the possibility for user u to adopt item i. As shown in Formula 9, the closer the $PIA(i,u)$ value to $PIA(C,u)$ is, the greater the $AP(i,u)$ value will be. In other words, item i recommended at this time accords with the *PIA* of user u for the category this item belongs to, and the possibility for the user to adopt this item is high.

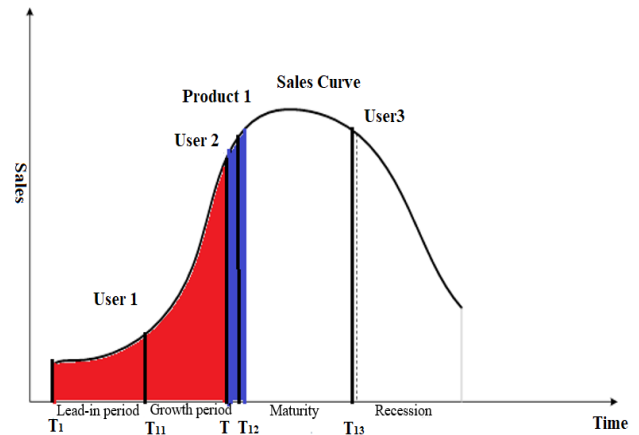$$AP(i,u) = \frac{i_{n+1}}{1 + PIA(i,u) - PIA(C,u)} \quad i \in C \tag{9}$$

Fig. 4. Schematic of the proposed novelty algorithm based on DI theory

The DI algorithm is mainly divided into the following steps. Step 1: K-means clustering method is used to classify all items into K categories, where category number K is determined on the basis of SC metrics. Step 2: The user PIA is modeled, and the user PIA for each item category is calculated. Step 3: The degree in which the user likes each item is predicted, and N items that may be liked are selected as an alternative set. Step 4: The novelty of each item is calculated using Formula 10, followed by recommendation. CF and case-based reasoning (CBR) algorithms, which are extensively applied, are selected and combined with the results of fusion strategy to generate two algorithms, namely, DI-CF and DI-CBR.

$$Novelty(i,u) = p(i \mid like, u) \times \frac{AP(i,u)}{\log(\sum_{k=1}^{n}(n-k+1)i_k)} \times \min_{j \in I_u}(1 - \cos ine(i,j)) \tag{10}$$

4 Result Analysis and Discussion

4.1 Experiment and design of evaluation metrics

In Top-N recommendation, the traditional experimental method hides some item ratings in the existing datasets and then measures the algorithm performance in terms of recall rate and accuracy. Here, the assumption that the hidden items cater to user preferences is reasonable, whereas the hidden items are already known by the user on the basis of the training dataset. The novelty of each item is relative to specific users at a specified time. A user behaviour dataset with a timestamp can be used and divided into two subsets by setting a time point. The items with high ratings at this time point are novel for the user. Following the above idea, a detailed offline experimental scheme for novelty recommendation is designed, as shown in Fig. 5 [25]. The experimental results are comprehensively evaluated using three metrics in the recommendation list, namely novelty (Formula 11), average popularity (Formula 12), and coverage (Formula 13).
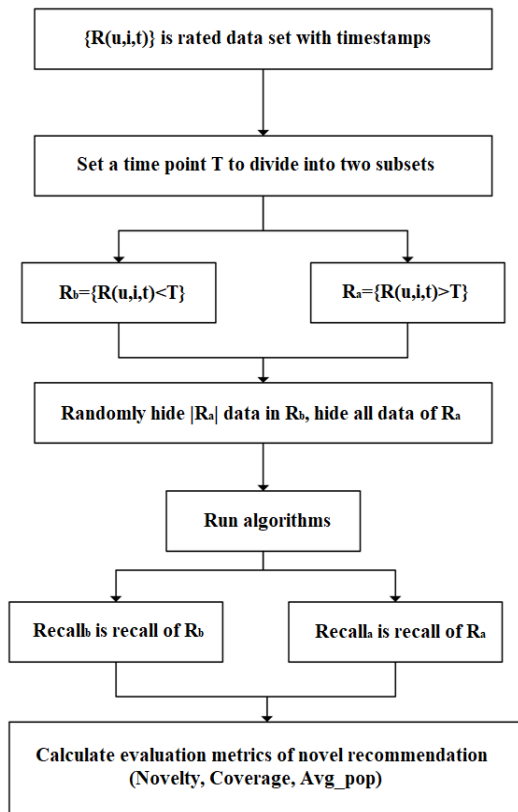

Fig. 5. Offline experimental scheme of novel recommendation

$$Novelty = \frac{1}{|U|} \sum_{u \in U} (2 \operatorname{Re} call_{au} - \operatorname{Re} call_{bu})$$

$$= \frac{1}{|U|} \sum_{u \in U} (2 \frac{|\operatorname{Re} c_u \cap Hide_{au}|}{|Hide_{au}|} - \frac{|\operatorname{Re} c_u \cap Hide_{bu}|}{|Hide_{bu}|}) \qquad (11)$$

$$Avg\_pop = \frac{1}{|U|} \sum_{u \in U} \frac{\sum_{i \in \operatorname{Re} c_o} RN_i}{|\operatorname{Re} c_u|} \qquad (12)$$

$$Coverage = \frac{|\cup_{u \in U} \operatorname{Re} c_u|}{|I|} \qquad (13)$$

where $Hide_{au}$ and $Hide_{bu}$ are the sets of items hidden by user u in R_a and R_b, respectively, $\operatorname{Re} c_u$ is the set of recommended items for user u, RN_i represents the number of evaluation times of item i, and U and I represent the user and item sets, respectively. As shown in the formula, $\operatorname{Re} call_b$ is the recall rate in the traditional experiment, and $\operatorname{Re} call_a$ is the accuracy measurement for the prediction of future user demand by the recommendation system. The quantity of recommendation lists is extremely limited. Thus, each recommendation list is expected to contain items meeting future user demands and reduce the items already known by the user, thereby accurately defining the meaning of novelty metrics. Average popularity and coverage refer to the measurement of the ability of the recommendation algorithm to extract long-tail products and also reflect the algorithm's ability in novelty recommendation.

The two datasets, namely, MovieLens and LastFM, are adopted in the offline experiment. The time interval is set to 14 days. Five experimental points are uniformly selected from the datasets of the previous year, previous rating data of each experimental point is set as R_b, the data within 14 days after these experimental data are used as the test data, namely R_a, and random rating data concealed by each user in R_b with the same quantity as that in R_a are taken as the training set. Each user is recommended with 20 items on the basis of this training set. The data obtained through the five repeated experiments are averaged to compare the differences of the two commonly used algorithms, namely, CF and CBR. The reasons are analyzed on the basis of various evaluation metrics. Considering that the differences of metrics are mainly considered, the change amplitude of experimental data is calculated using Formula 14, where M_n and M_s represent the new experimental and experimental data for reference, respectively, which can be expressed as:

$$M = \frac{M_n - M_s}{|M_s|} \qquad (14)$$

4.2 Item classification using the K-means clustering algorithm

The K-means algorithm first needs to formulate cluster number K, which starts from 10 and progressively increases by 10. Each cluster number K is operated for ten times and their average SC value is then calculated to measure the clustering effect of cluster number K because the K-means algorithm is extremely sensitive to the initial centroid. The results are shown in Figs. 6. The SC value of the LastFM dataset tends to be stable when the cluster number reaches 50, whereas that of the MovieLens dataset is the maximum when the cluster number reaches 70. Table 2 lists the classified items from the MovieLens dataset using the K-means clustering algorithm. Ten items in three categories are randomly selected, which are described as: Category I: feature films, love movies, and comedy movies, Category II: thriller films and horror movies, and the feature of Category II is not obvious. Thus, the majority of classification made using the K-means clustering algorithm is reasonable and represents the preferences of users of the same category, thereby realizing the item clustering function via user behavior.

Table 2. Example of item classification in the MovieLens dataset using the K-means clustering algorithm

Category	Name	Years	Types
1	When Night Is Falling	1995	Plot/love/same sex
	Once Upon a Time... When We Were Colored	1995	Plot/love
	The Journey of August King	1995	Plot
	Beyond Rangoon	1995	Plot
	The Stars Fell on Henrietta	1995	Plot
	Unstrung Heroes	1995	Plot/Comedy
	Nina Takes a Lover	1994	Plot/love
	Tui shou	1992	Plot/family
	Roommates	1995	Plot/Comedy
2	The Jerky Boys	1995	Comedy/crime
	Harem suar	1999	Story/History
	L'oeil de Vichy	1993	Documentary/War
	The Rich Man's Wife	1996	Thriller
	A Damsel in Distress	1937	Comedy/love/song dance
	Amityville 3D	1983	Plot/horror
	Amityville: A New Generation	1993	Terror
	The Amityville Curse	1990	Terror
	Nosferatu a Venezia	1988	Terror
	Turbulence	1997	Action/Thriller
3	Open Season	2006	Comedy/animation/adventure
	Above the Rim	1994	Plot/crime/sports
	JLG/JLG - autoportrait de décembre	1994	Plot/documentary
	The Allnighter	1987	Comedy/love
	Belly	1998	Plot/crime
	Foolish Wives	1922	Plot
	Cobra	1986	Action/horror/crime
	The Omega Code	1999	Action/thriller/adventure
	Urban Legends: Final Cut	2000	Suspense/horror
	Carman: The Champion	2001	Plot/action

Fig. 7 shows the statistical data of PIA of one user for items under various categories in the MovieLens dataset. The overall mean *PIA* of the user for various items fluctuates, indicating that the user presents different PIAs for different items. The variance is lower than the overall variance, indicating that the user's *PIA* for different items fluctuates within a small range.

4.3 Novelty and accuracy of the DI algorithm
The recommendation idea of the DI algorithm is to first take N items with maximum $p(i|like,u)$ as an alternative set and then use Formula 10 to calculate the novelty of items in the alternative set for recommendation. $p(i|like,u)$ should be calculated using the traditional recommendation algorithm. CF and CBR algorithms, which are extensively applied, are selected, and the results of fusion strategy are used to generate two new algorithms, namely, DI-CF and DI-CBR. The size N of alternative set is an important parameter influencing the algorithm's performance. Five experimental points are selected for experiments, and evaluation metrics

are selected as $\mathrm{Re}call_b$, $\mathrm{Re}call_a$, *novelty, avg_pop*, and *coverage*, and the mean values of evaluation metrics obtained through five repeated experiments are calculated. Figs. 8 and 9 display the influences of N value on the evaluation metrics of four algorithms. The data in the figures are the relative change values of various metrics compared with the evaluation data of traditional CF and CBR algorithms.
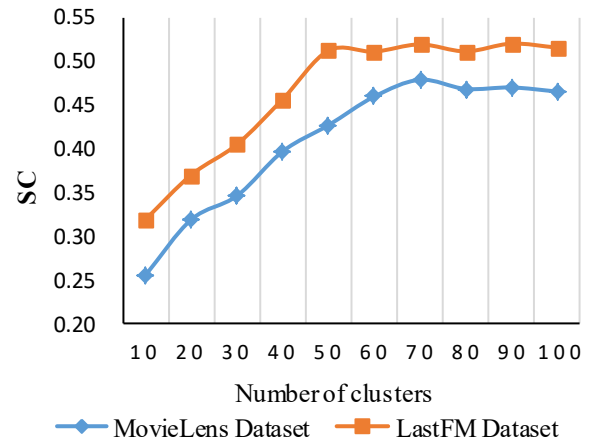
Fig. 6. SC change curve of cluster number K using the K-means clustering algorithm
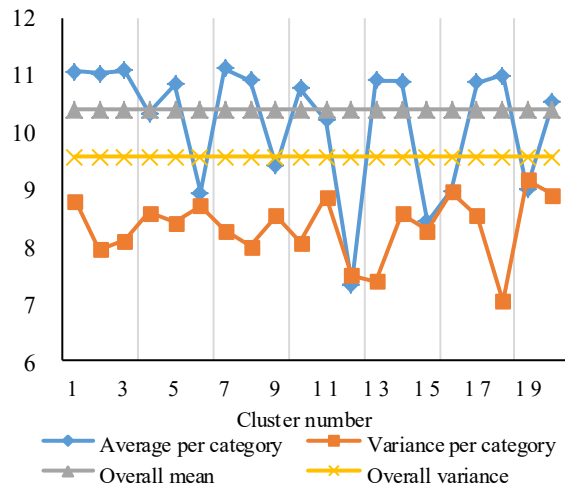
Fig. 7. Statistical data of PIC of one user for different items in the MovieLens dataset

The novelty metrics show that the novelty of the DI algorithm is remarkably higher than that of traditional algorithms. The novelty of the DI algorithm increases with the increase in the N value. The maximum novelty value is reached and then decreases when the N value is within 300–400, and the novelty of the DI algorithm is constantly higher than that of traditional algorithms regardless of the change in the N value. Therefore, the inclusion of user PIA metrics in the item novelty calculation can improve the novelty of the recommendation system. The effects of PIA on the novelty of the recommendation system can be observed through the analysis of two metrics, namely, $\mathrm{Re}call_a$ and $\mathrm{Re}call_b$.
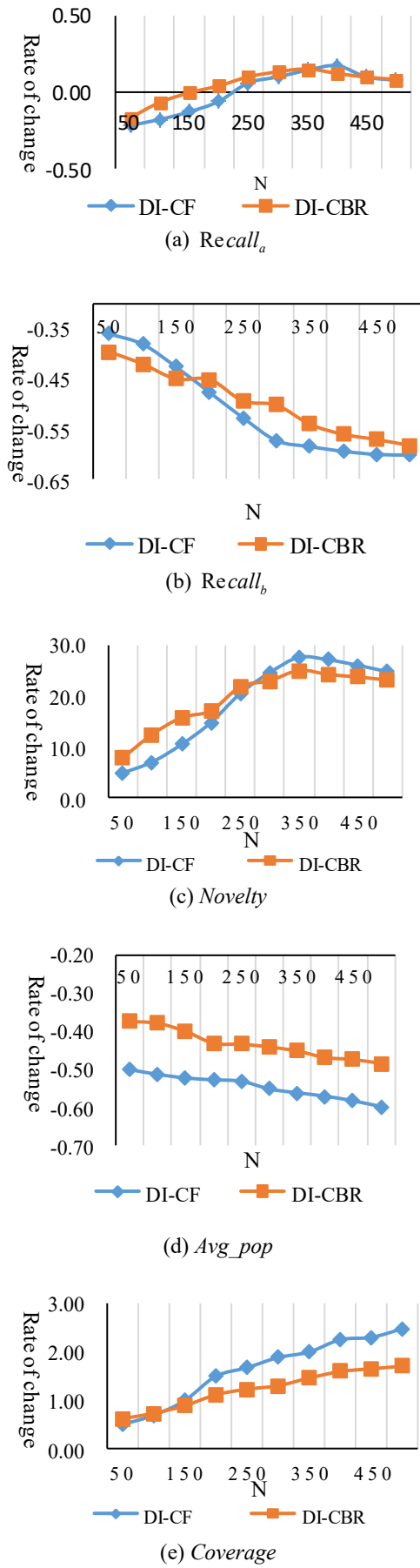
(a) Re*call*$_a$

(b) Re*call*$_b$

(c) *Novelty*

(d) *Avg_pop*

(e) *Coverage*

Fig. 8. Experimental results of the MovieLens dataset

(a) Re*call*$_a$

(b) Re*call*$_b$

(c) *Novelty*

(d) *Avg_pop*

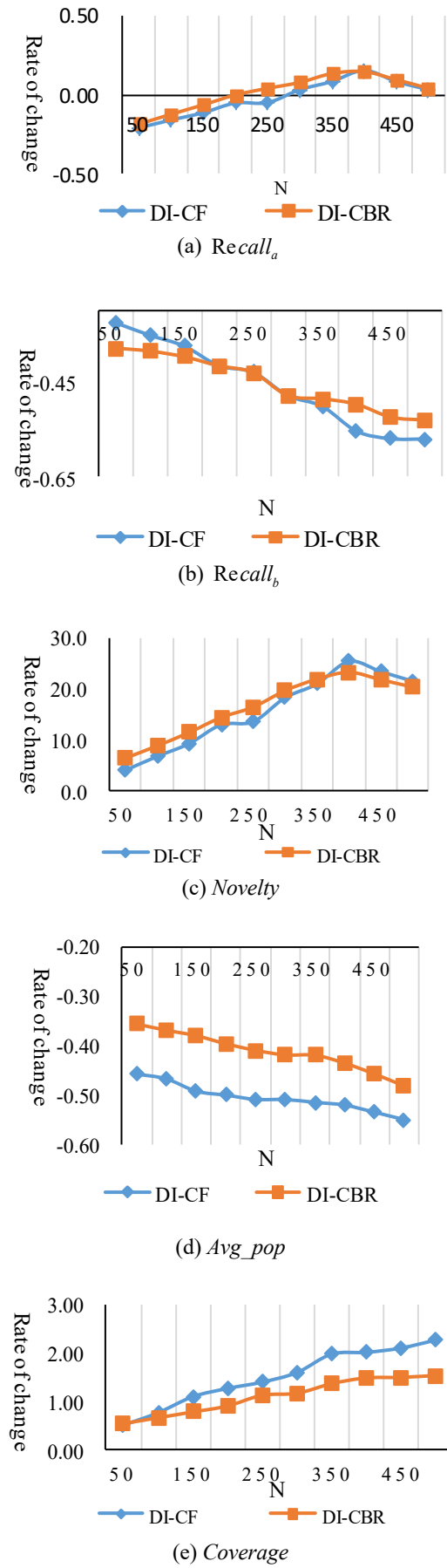

(e) *Coverage*

Fig. 9. Experimental results of the MovieLens dataset

The metric $\mathrm{Re}call_a$ of the DI algorithm is notably higher than those of commonly used algorithms because it focuses on recommending items in accordance with user preferences and is inclined to items with high popularity. The DI algorithm can recommend some popular and old items with low popularity to new users to improve the recommendation accuracy. The metric $\mathrm{Re}call_b$ of the DI algorithm is lower than those of traditional algorithms, and its decline amplitude gradually stabilizes with the increase in the N value. Therefore, the substantial novelty improvement of the DI algorithm is mainly ascribed to the increase in metric $\mathrm{Re}call_a$ and sharp decrease in metric $\mathrm{Re}call_b$. By calculating PIA, The DI algorithm can contribute to the "novelty" of two item categories, namely unpopular and popular items that can be perceived by different users by calculating the PIA. Long-tail items are recommended to users who know about this item category, whereas popular items are recommended to user who does not know about this item category. Therefore, the metric *Avg_pop* of the DI algorithm is superior to those of traditional commonly used algorithms. The average popularity change of the recommendation list is insensitive to the size N of the alternative set. The coverage of the recommendation list of the DI algorithm rapidly increases with the increase in N increases compared with traditional algorithms, and the alternative set is calculated using *p(i|like,u)*. Thus, the greater the N value is, the more the items will be covered in the alternative set, and the DI algorithm can improve the ability of the recommendation system to extract all types of items without lowering its accuracy.

5 Conclusion

The K-means clustering algorithm was used to classify items for providing "novelty" recommendations in accordance to user's acceptance degree. The PIA of each item category was calculated for modeling users' PIA, and the PIA metrics were integrated in the probability calculation for the user to adopt the item. Items with different popularity degrees were recommended to different users. The conclusions are summarized as follows:

(1) The DI algorithm improves the novelty without influencing the accuracy of the recommendation list;

(2) The DI algorithm lowers the average popularity of the recommendation list;

(3) The DI algorithm increases the coverage of the recommendation list.

The modeling calculation of users' PIA and item popularity prediction were combined in this study. Different users were recommended with items that conformed to their preferences and might become popular in the future for enabling the recommendation system to effectively recommend items of different categories, which has certain significance in improving the user satisfaction with the recommendation system and improving user experience. The DI algorithm provides improvements with regard to user preference and unawareness without considering the differences. Traditional differences are mainly measured by calculating the minimum distance or average distance from the target item to the items already rated by the user, although the two methods are simple. The clustering method will be applied to novelty recommendation in future studies to clearly describe user preferences and enhance the differences in recommendation lists between target users.

References

1. Bakshy, E., Messing, S., Adamic, L.A., "Exposure to ideologically diverse news and opinion on facebook". *Science*, 348 (6239), 2015, pp.1130-1132.
2. Han, J., Yamana, H., "A survey on recommendation methods beyond accuracy". *IEICE Trans on Information & Systems*, E100 (12), 2017, pp.2931-2944.
3. Barbera, P., Jost, J. T., Nagler, J.A., Tucker, R., "Tweeting from left to right: is online political communication more than an echo chamber?". *Psychol. Science*, 26(10), 2015, pp.1531–1542.
4. Vicario, M. D., Bessi, A., Zollo, F., Petroni, A., Scala, G., Caldarelli, H.E., Stanley, W., Quattrociocchi, F., "The spreading of misinformation online". *Proceedings of the National Academy of Sciences*, 113(3), 2016, pp.554–559.
5. Gravino, P., Caminiti, S., Sîrbu, A., Tria, F., Servedio, V.D.P., Loreto, V., "Unveiling political opinion structures with a web-experiment". In: *1st International Conference on Complex Information Systems*, Rome, Italy: SciTe Press, 2016, pp.22–24.
6. Zhou, T., Kuscsik, Z., et al., "Solving the apparent diversity accuracy dilemma of recommender systems". *Proceedings of the National Academy of Sciences of the United States of America,* 107(10), 2010, pp.4511-4515.
7. Hong, Y., Junhua. L., "A recommendation algorithm solving cold boot problem of new items". *Journal of Software*, 26(6), 2015, pp.1395-1408.

8. Chou S. Y., Yang Y. H., Jang J. S. R., et al. "Addressing cold start for next-song recommendation". In: *Proceeding of the 10th ACM Conference on Recommender Systems*, New York, USA: ACM Press, 2016, pp.115-118.
9. Chen L. J., Gao J., "A trust-based recommendation method using network diffusion processes". *Physica A:Statistic Mechanic and Its Application*, 506,2018, pp.679-691.
10. Kapoor K., Kumar V., Terveen L., et al., ""I like to explore sometimes": Adapting to dynamic user novelty preferences". In: *Proceeding of ACM Conference on Recommender Systems*, New York, USA: ACM Press, 2015, pp.19-26.
11. Ma W. P., Feng X., Wang S.F., et al. "Personalized recommendation based on heat bidirectional transfer". *Physica A: Statistical Mechanics & Its Applications*, 444, 2016, pp.713-721.
12. Hu J. M., Lin X., "Design of fusion recommendation algorithm for socialized users'minority based on the energy diffusion theory of heat spreading". *Information Studies: Theory & Application*, 39 (4), 2016, pp.123-127.
13. Wu H., Cui X. H., He J., et al., "On improving aggregate recommendation diversity and novelty in folksonomy-based social systems". *Personal & Ubiquitous Computing*, 18(8), 2014, pp.1855-1869.
14. Wang B., Cao H., "Study of a tourism recommendation model based on novelty and diversity". *Computer Engineering and Applications*, 52 (6), 2016, pp.219-222.

15. Yu Q., Peng Z. Y., Hong L., Wan Y. L., "A novelty Web community recommendation method based on user neighborhood and them". *Journal of Software*, 27 (5), 2016, pp.1266-1284.
16. Pietro G., Bernardo M., Vittorio L., "Towards novelty-driven recommender systems". *Comptes Rendus Physique*, 20(4), 2019, pp.371–379.
17. Marcelo M., Nicolas T., "Evaluating content novelty in recommender system". *Journal of Intelligent Information Systems*, 2019, doi: 10.1007/s10844-019-00548-x.
18. Saranya, K., G., Sadasivam, G., "Personalized news article recommendation with novelty using collaborative filtering based rough set theory". *Mobile networks & applications*, 22(4), 2017, pp.719–729.
19. Wael A., Eid A., Christoph R., Steffen S., "Ensuring Novelty and Transparency in Learning Resource-Recommendation Based on Deep Learning Techniques". In: *13th European Conference on Technology Enhanced Learning,* Leeds, UK: Springer, 2018, pp.609-612.

20. Jorge D., David M. R., Amparo A. B., Oscar L., Antonio B., "Optimizing novelty and diversity in recommendations". *Progress in Artificial Intelligence*, 8, 2019, pp.101–109.
21. Kaminkas, M., Bridge. D., "Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems". *Acm Transactions on Interactive Intelligent Systems*, 7(1), 2016, Article no.: 2.
22. Pathak, A., Patra, B. K., "A knowledge reuse framework for improving novelty and diversity in recommendations". In: *Proceedings of the Second ACM IKDD Conference on Data Sciences,* Bangalore, India: ACM Press, 2015, pp.11-19.
23. FM. M, Batista, C. S., Santos, R. L. T., et al. "Beyond Relevance: Explicitly Promoting Novelty and Diversity in Tag Recommendation". *ACM Transactions on Intelligent Systems & Technology*, 7(3), 2016, Article no.: 26.
24. Rogers E M., "*Diffusion of innovations*". New York: Simon and Schuster, USA, 2010, pp.14-15.
25. Zhang, L., Peng, L. F., Phelan, C.A., "Novel recommendation of user-based collaborative filtering". *Journal of Digital Information Management*, 12(3), 2014, pp.165-175.