CrossMark

An adjustable re-ranking approach for improving the individual and aggregate diversities of product recommendations

Qian Wang[1] · Jijun Yu[1] · Weiwei Deng[2] 📷

Abstract

The effectiveness of product recommendations is previously assessed based on recommendation accuracy. Recently, individual diversity and aggregate diversity of product recommendations have been recognized as important dimensions in evaluating the recommendation effectiveness. However, the gain of either diversity is usually at the cost of accuracy and the increase of one diversity does not guarantee a significant improvement in the other. A few attempts have been made to achieve reasonable trade-offs either between recommendation accuracy and individual diversity or between recommendation accuracy and aggregate diversity. Little attention has been paid to obtain a balance among the three important aspects of product recommendations. To address this problem, we propose an adjustable re-ranking approach that incorporates two new ranking criteria for improving both diversities. Three ranking lists are generated to guarantee recommendation accuracy, individual diversity, and aggregate diversity, respectively. The three ranking lists are finally merged with tunable parameters to generate a recommendation list. To evaluate the proposed method, experiments are conducted on a data set obtained from Alibaba. The results show that the proposed method achieves much higher improvements in both diversities than the baseline methods when sacrificing the same amount of recommendation accuracy.

Qian Wang, Jijun Yu and Weiwei Deng have contributed equally to this work.

✉ Weiwei Deng
wwdeng3-c@my.cityu.edu.hk

[1] Department of Information System and Engineering, Business School of Sun Yat-sen University, Guang Zhou, Guang Dong, People's Republic of China

[2] Department of Information Systems, College of Business, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Hong Kong, People's Republic of China

🙋 Springer

1 Introduction

Recommender systems are widely used in electronic commerce to help customers find products. A recommender system learns customers' preferences and recommends products that customers are likely to purchase. Studies have consistently shown that product recommendations can increase the sales and profits of electronic commerce [1, 2]. The effectiveness of a recommender system is usually assessed by the accuracy of recommendation results, i.e. how well a customer's preference on a product is predicted by the system. Accuracy is no doubt of great importance in evaluating the recommendation effectiveness, which can be exemplified by a large pool of studies dedicated to improve the accuracy of recommendation methods [3–7]. Recently, recommendation diversity has been recognized as another important aspect for increasing customers' satisfaction and the sales of electronic commerce [8–11].

There are two types of recommendation diversity, namely individual diversity (ID) and aggregate diversity (AD). ID refers to the diversity or dissimilarity of products in a recommendation list for a customer. Recommender systems are expected to recommend diverse products to meet the needs of a customer. A diverse set of products also leads to more serendipity in recommendations. Therefore, ID is important regarding customers' satisfaction [12, 13]. AD is the diversity or dissimilarity of products across all recommendation lists for all customers. Low AD indicates that only a small portion of products is recommended to customers, which will exacerbate the long tail phenomenon in online marketplaces. Increasing AD, therefore having a higher possibility to recommend more products in the long tail, has great potential for gaining higher profits since products in the long tail are extremely abundant. From this point of view, AD is significant to ecommerce business [14, 15].

Although recommendation accuracy and diversities are important, product recommendations are facing the dilemma between high accuracy and high diversities. The increase of either diversity comes with a decrease in recommendation accuracy [16]. In addition, increasing ID does not necessarily lead to a significant improvement in AD, and vice versa [8, 15, 17]. For example, high ID can be obtained by recommending the same diverse set of products to all customers, but AD still remains low. A few methods have been proposed to achieve reasonable trade-offs between recommendation accuracy and diversities, such as ranking-based techniques [18, 19] and optimization-based approaches [20, 21]. However, previous methods focus on maintaining the balance either between accuracy and ID or between accuracy and AD. Little effort has been made to maintain the balance among the three aspects.

In this paper, we propose an adjustable re-ranking approach to address the above problem. Specifically, we first obtain a candidate list using accuracy oriented recommendation methods. Two copies of the candidate list are re-ranked based on product pairs' distance in the concept hierarchy of products and products' degree centrality in the co-purchased network of products, respectively. The distance of product pairs is used to guarantee high ID while the degree centrality is used to guarantee high AD. To obtain reasonable levels of recommendation accuracy, ID, and AD, the three ranking

lists are merged with tunable parameters (i.e. the weight assigned to each list). Top *N* products from the merged list are then recommended to customers. Experiments are conducted on a real-world data set obtained from Alibaba, the largest ecommerce company in China. The experimental results demonstrate that the proposed method can produce different trade-offs between recommendation accuracy and both diversities by changing the length and the weight of the three lists. The results also show that the proposed method achieves much higher improvements in both diversities than several baseline methods when sacrificing the same amount of recommendation accuracy. To summarize, this paper proposes two novel criteria to guarantee ID and AD respectively, i.e. the distance between products in the concept hierarchy and the degree centrality of products in the co-purchased network. A re-ranking approach is designed to obtain adjustable balances among recommendation accuracy, ID, and AD. Experiments are also conducted to demonstrate the effectiveness of the proposed method.

The rest of this paper is organized as follows. The second section reviews accuracy oriented recommendation methods, accuracy and diversity measures, and studies on maintaining the trade-offs between recommendation accuracy and diversities. The third section introduces the proposed re-ranking approach. The fourth section provides details of the experiments and the experimental results. The last section concludes this study with major contributions and possible future research.

2 Related work

2.1 Accuracy oriented recommendation methods

Most existing recommendation methods are designed with the goal of high recommendation accuracy. These methods can be classified into three categories, namely collaborative filtering (CF), content-based filtering (CB), and hybrid method [22, 23]. For a given customer, CF recommends products that his or her like-minded customers have liked before, and CB recommends products with features liked by the customer. Hybrid method can combine both types of methods in different ways, such as switching mechanism, cascade mechanism, and feature augmentation [24]. A recommender system generally takes two stages for recommending products to customers. The first one is to estimate customers' preferences on products using certain recommendation algorithms. Based on the estimated preferences, the second stage is to select and recommend products that maximize customers' satisfaction. The proposed adjustable re-ranking approach is used at the second stage and it can be combined with any accuracy oriented recommendation method. To illustrate and evaluate the proposed re-ranking approach, we combine it with the most popular accuracy oriented method, i.e. the user-based CF [25]. Given a set of customers *U*, a set of products *I*, and a rating matrix *R* representing customers' preferences on products, the user-based CF estimates the unknown preference of customer *u* on product *i* as follows:

$$sim(u, u') = \frac{\sum_{i \in I(u, u')} R(u, i) \cdot R(u', i)}{\sqrt{\sum_{i \in I(u, u')} R(u, i)^2} \sqrt{\sum_{i \in I(u, u')} R(u', i)^2}} \tag{1}$$

$$R^*(u, i) = \frac{\sum_{u' \in N(u)} sim(u, u') \cdot R(u', i)}{\sum_{u' \in N(u)} |sim(u, u')|} \tag{2}$$

where $I(u, u')$ is the set of products rated by both customer u and u', $sim(u, u')$ is the similarity between customer u and u', and $N(u)$ is the set of neighbors nearest to customer u.

2.2 Accuracy and diversity measures

Different measures have been proposed to evaluate recommendation accuracy and diversities. To select suitable measures for evaluating the proposed method, we review the most commonly used accuracy and diversity measures in this subsection. There are two popular types of accuracy metrics, namely predictive accuracy metrics (such as *Mean Absolute Error* and *Root Mean Square Error*) and classification accuracy metrics (such as *Precision*, *Recall*, and *F* measure) [26, 27]. Predictive accuracy metrics measure how close the predicted preferences are to the true customer preferences. They are used when preferences are scalar value. Classification accuracy metrics measure the percentage of correct predictions about whether or not a customer likes a product. They are used when preferences are binary value. In the context of electronic commerce, customer preferences are usually expressed by binary value, such as viewing or not viewing a web page, clicking or not clicking a link, and buying or not buying a product.

ID is measured by the average dissimilarity or distance of all product pairs in one recommendation list [13, 27]. Previously, the dissimilarity or distance between products is calculated based on the ratings given by customers [28]. However, the obtained diversity may not be explicitly sensed by customers because products with similar features could have very different rating patterns. For example, two smart phones with similar design and functions may get totally different ratings from the same set of customers due to the difference of brands. Therefore, we propose to measure ID by the average distance between products in products' concept hierarchy which explicitly shows the diversification of products. The proposed criterion is introduced in Sect. 3.

AD is measured either by the number of distinct products recommended to all customers (e.g. *Diversity_in_top_N*) [19, 20] or the distribution of recommended products among all recommendation lists (e.g. *Gini_diversity*) [14, 15]. *Diversity_in_top_N* favors recommending more products, but it does not consider the distribution of recommended products. Consequently, recommended products could have high *Diversity_in_top_N* while bearing an unbalanced distribution. For example, recommending every possible product to a different customer and filling the remaining recommendation lists with the same set of products could produce high *Diversity_in_top_N*. But the product distribution among all recommendation lists is unbalanced. To compensate for this drawback, *Gini_diversity* is proposed to measure how evenly the products are recommended to all customers. This metric is also

important to product recommendations because it ensures a more even amount of exposure for each of the recommended products.

2.3 Accuracy-diversity trade-off

According to the stage where diversities are taken into account, previous methods balancing recommendation accuracy and diversities can be categorized into two types. The first type of methods considers the diversities at the stage of preference estimation. For example, Niemann and Wolpers [29] improved the rating predictions for niche items by considering their jointly consumed items when measuring item similarities. Gan and Jiang [16] identified the adverse effect of item popularity and proposed to adjust user similarity through a power function for more correct rating predictions, and thus for higher recommendation accuracy and AD. Gan [30] balanced recommendation accuracy and AD by calculating discriminant scores for items using a random walk with restart model on a user similarity network. These methods try to identify problems that lead to the lack of recommendation diversities and therefore propose corresponding solutions to enhance diversities. Although such methods can increase diversities at the stage of preference estimation, they provide no control over the extents of recommendation accuracy and diversities.

Another type of methods guarantees recommendation diversities at the product selection stage. Such methods predict customers' preferences on products using previous accuracy oriented methods and then select products based on the predicted preferences and diversity oriented criteria. For instance, several studies [14, 21, 31] formulated the accuracy-diversity trade-off problem as a multi-objective optimization problem with certain constraints and proposed corresponding solutions to solve the optimization problem. It is possible for optimization methods to adjust the trade-offs between recommendation accuracy and diversities by setting certain parameters, but it is very difficult to find the optimal solution [32, 33]. Some studies proposed new ranking mechanisms to generate balanced recommendation lists such as parameterized item-popularity-based ranking approach [19] and graph-based re-ranking approach [15]. Compared to other methods, ranking methods are more efficient and can easily adjust the balance between recommendation accuracy and diversities through tunable parameters. However, previous methods aim at achieving a fair trade-off either between recommendation accuracy and ID or between recommendation accuracy and AD. Little attention has been paid to balance the three important aspects at the same time. This paper addresses this problem by proposing an adjustable re-ranking approach which is employed at the stage of product selection.

3 The proposed re-ranking approach

In this section, we introduce the proposed re-ranking approach for improving the individual and aggregate diversities of product recommendations.

Figure 1 shows the framework of the proposed re-ranking approach and illustrates how to improve ID and AD while keeping acceptable recommendation accuracy.
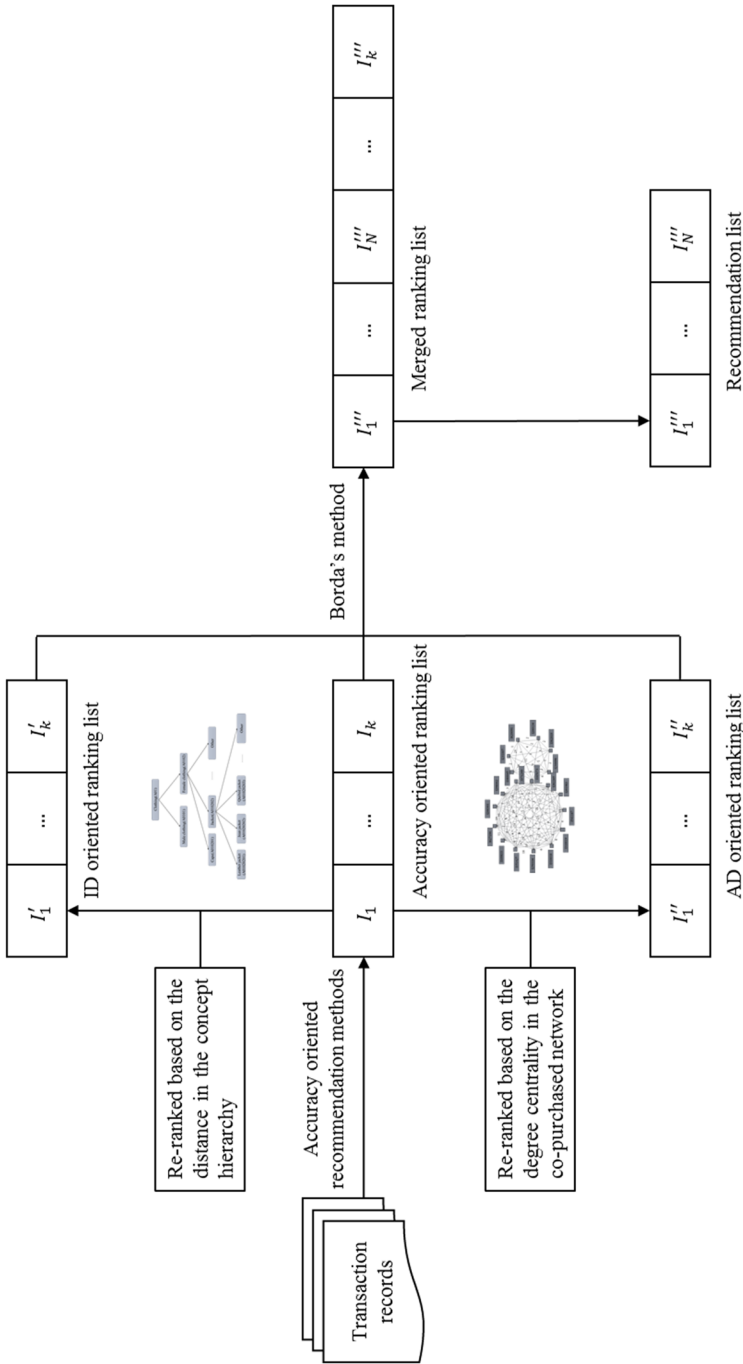
Fig. 1 The framework of the proposed re-ranking approach

The proposed method has three steps. The first one is to predict customers' preferences and generate an accuracy oriented ranking list by accuracy oriented recommendation methods. The second step is to make two copies of the accuracy oriented ranking list and re-rank them to produce ID and AD oriented ranking lists respectively. The final step is to merge the three ranking lists and obtain the top N recommendation list. Details are introduced in the following subsections.

3.1 Accuracy oriented ranking list

The generation of an accuracy oriented ranking list is the same as that of previous recommender systems. Customers' preferences on products are first estimated by accuracy oriented recommendation methods. For each customer, the list is then filled with products that are of the highest estimated preferences and have not been consumed by the customer. For the sake of re-ranking efficiency, only top k products with the highest estimated preferences are included in the accuracy oriented ranking list. To leave space for improving ID and AD, k should be larger than the intended number of recommended products. Specifically, given a customer and his or her estimated preferences p, the accuracy oriented ranking list for the customer is as follows:

$$List_{accuracy} = \{I_1, I_2, \dots, I_k\} \tag{3}$$

where $p_{I_1} \geq p_{I_2} \geq \cdots \geq p_{I_k}$ and k is larger than the intended number of recommended products.

3.2 Individual diversity oriented ranking list

To show ID in a way that can be explicitly sensed by customers, we propose to re-rank products based on the distance of product pairs in products' concept hierarchy. A concept hierarchy is a tree-like structure that organizes concepts from groups of low-level and specific concepts to several high-level and general concepts, and finally to one universal concept [34, 35]. Due to the explicit representation of relationships between concepts, concept hierarchies have been widely used in applications that require measuring relationships between objects, such as segmenting customers [35], recommending documents [36], and finding similar patients [37]. In the context of electronic commerce, products are related to certain product concepts that are well organized as a concept hierarchy. Figure 2 shows part of the concept hierarchy on clothing products.

Given such a concept hierarchy, the distance between two products can be calculated by a commonly used measure [38, 39] defined below:

$$dis(I_i, I_j) = 1 - \frac{2 \times N_3}{N_1 + N_2 + 2 \times N_3} \tag{4}$$

where N_1 and N_2 are the numbers of edges from the lowest-level concepts related to I_i and I_j to their least common subsumer, i.e. the lowest-level ancestor that includes
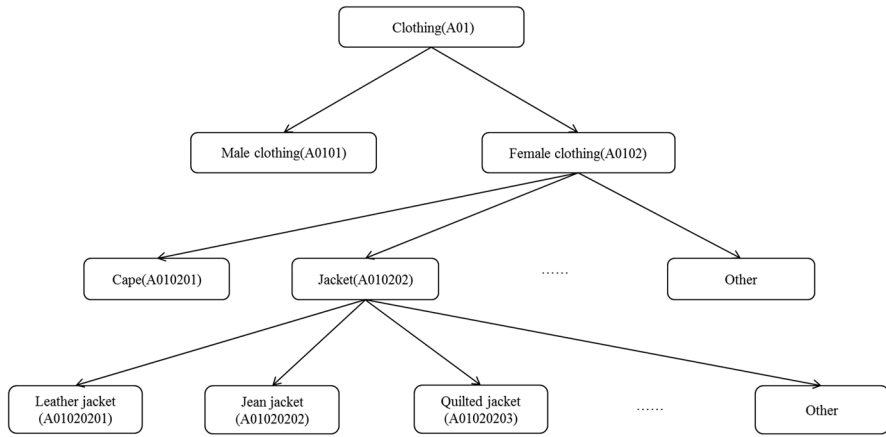
Fig. 2 Part of the concept hierarchy on clothing products

both concepts. N_3 is the number of edges from the least common subsumer to the root of the concept hierarchy. For example, if I_1 and I_2 are products that belong to category *A01020201* and *A010201* respectively in the concept hierarchy shown in Fig. 2, then their least common subsumer is category *A0102*. Therefore, $N_1 = 2$, $N_2 = 1$, $N_3 = 1$, and $dis(I_1, I_2) = 1 - \frac{2 \times 1}{2 + 1 + 2 \times 1} = \frac{3}{5}$. There are other types of products, such as digital products and food. Different types of products belong to different sub-hierarchies, and all sub-hierarchies constitute the concept hierarchy that contains all types of products. Therefore, the distance between products of any categories can be calculated within the same concept hierarchy.

Since ID is measured by the average distance between products in a recommendation list, it can be improved by positioning the product pair with the largest distance in the first and second places of the list, and continuously putting in the next place the product that has the largest average distance to the existing products in the list. Therefore, to guarantee certain level of ID, we copy the accuracy oriented ranking list and re-rank the copied list according to the proposed mechanism. Specifically, the ID oriented ranking list can be represented as follows:

$$List_{ID} = \{I'_1, I'_2, \ldots, I'_k\} \tag{5}$$

where $dis(I'_1, I'_2) \geq dis(I'_i, I'_j)$ for $\forall i, j \leq k$ and $\frac{1}{s-1} \sum_{I'_t \in \{I'_1, I'_2, \ldots, I'_{s-1}\}} dis(I'_s, I'_t) \geq \frac{1}{s-1} \sum_{I'_t \in \{I'_1, I'_2, \ldots, I'_{s-1}\}} dis(I'_{s'}, I'_t)$ for $\forall I'_{s'} \in \{I'_{s+1}, I'_{s+2}, \ldots, I'_k\}$. The positions of I'_1 and I'_2 are determined by their estimated preferences, i.e. the product with larger estimated preference is ranked first.

3.3 Aggregate diversity oriented ranking list

Previous studies have shown that recommendation agents can lower AD of product recommendations since they are usually made based on products' co-purchased patterns [8, 40]. It is thus reasonable to conjecture that re-ranking products against their
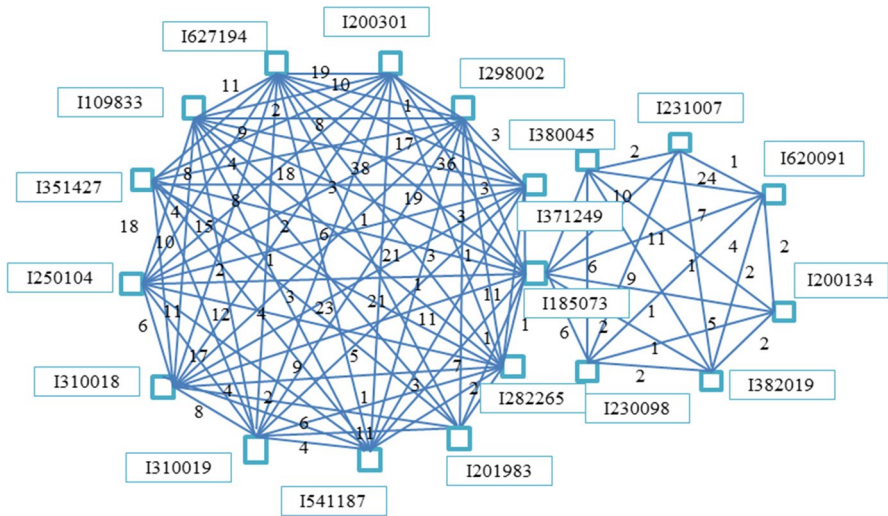
Fig. 3 An example of the co-purchased network

co-purchased relationships can improve AD. Co-purchased relationships among products can be constructed as a co-purchased network in which products are represented by nodes and co-purchasing relationships between products are represented by links between corresponding nodes. In graph and network analysis domain, degree centrality represents the number of links incident upon a node [41] and indicates the co-purchased frequency of products in the context of the co-purchased network [42]. We conjecture that ranking against the degree centrality can increase AD. An example of the co-purchased network is presented in Fig. 3. In the example, numbers beside the links refer to the co-purchased frequency between two products.

Formally, the degree centrality of products in the co-purchased network is defined as follows:

$$degree\_centrality(i) = \sum_{j \in I} f_{i,j} \tag{6}$$

where $f_{i,j}$ refers to the co-purchased frequency of product i and j. To examine our conjecture, we empirically investigated the correlation between the degree centrality and AD by increasing the average degree centrality of products in recommendation lists and observing the resulted AD. The empirical results shown in Fig. 4 verify our conjecture, i.e. *Gini_diversity* and *Diversity_in_top_N* are negatively related to the average degree centrality of products in all recommendation lists. Therefore, to improve AD, we copy the accuracy oriented ranking list and re-rank the products in non-descending order of the degree centrality in the co-purchased network. The AD oriented ranking list is then represented as follows:

$$List_{AD} = \left\{ I_1'', I_2'', \dots, I_k'' \right\} \tag{7}$$

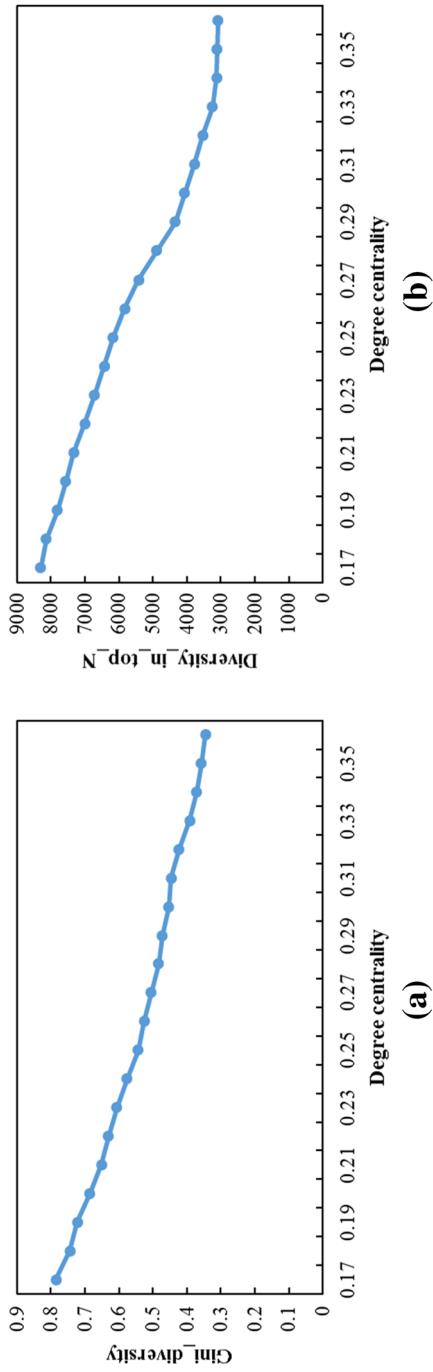where $degree\_centrality\left(I_i''\right) \leq degree\_centrality\left(I_{i+1}''\right)$.

Fig. 4 The correlation between the degree centrality and AD

3.4 Merging of ranking lists

After obtaining the three ranking lists, the balance among recommendation accuracy, ID, and AD can be achieved by merging the three lists. Rank aggregation is a well-studied area and many aggregate functions have been proposed to generate an aggregated ranking list from multiple ranking lists. Among existing aggregate functions, Borda's method has gained wide popularity because of its effectiveness and simplicity [43]. Therefore, Borda's method is employed to merge the three ranking lists in this paper. Borda's method works by assigning a score to each position of a ranking list and sorting products according to their total scores calculated based on their positions in different ranking lists. In this study, the score assigned to a position is set to the ranking of the position. Consequently, the higher a product scores, the lower it ranks. Traditionally, all ranking lists are treated equally in Borda's method. To make the balance among recommendation accuracy, ID, and AD adjustable, we adapt Borda's method by assigning different weight to the three ranking lists. Given an item and its rankings in the three lists, the integrated ranking score obtained by the adapted Borada's method is as follows:

$$Score_i = \alpha \times Ranking_i^{List_{accuracy}} + \beta \times Ranking_i^{List_{ID}} + \gamma \times Ranking_i^{List_{AD}} \qquad (8)$$

where $Ranking_i^{List}$ is the ranking of item i in the corresponding ranking list, α, β, and γ are weight assigned to corresponding ranking lists, $0 \leq \alpha, \beta, \gamma \leq 1$ and $\alpha + \beta + \gamma = 1$. After merging the three ranking lists, the final recommendation list can be represented as follows:

$$Recommendation\_list = \left\{ I_1''', I_2''', \ldots, I_N''' \right\} \qquad (9)$$

where $Score\left(I_i'''\right) \leq Score\left(I_{i+1}'''\right)$.

4 Experimental evaluation

4.1 Data set

To evaluate the proposed re-ranking approach, we conduct experiments on a real-world data set obtained from Alibaba, the largest ecommerce company in China. The original data set[1] contains more than one hundred million consumer-product interaction records. Without compromising the evaluation, we randomly select 50,000 transaction records from the original data set. Each transaction record contains one customer and at least one product. Each product belongs to at least one product category. A product category can range from the most general level to the finest level in the concept hierarchy of products. The statistical description of the selected data set is provided in Table 1.

---

[1] https://tianchi.aliyun.com/datalab/dataSet.html?spm=5176.100073.0.0.6f3635eezHHOtK&dataId=649.

Table 1 The statistical description of the selected data set

Data field	Value
#Transaction	50,000
#Customer–product purchase record	146,998
#Customer	16,260
#Product	95,416
#Product category	3658
#Co-purchased product	82,540
#Co-purchased product pair	272,164

4.2 Experimental design

The selected data set is randomly divided into training and testing sets with an 80:20 split. Specifically, 40,000 transaction records are used to tune the parameter of and then used as the input of the accuracy-oriented recommendation method (i.e., the user-based CF). The remaining 10,000 transaction records are used to test the recommendation results re-ranked by the proposed method as well as the baseline methods. In our experiments, customers' preferences are binary value (i.e. purchasing or not purchasing a product). Therefore, we use the following classification accuracy metric to measure the accuracy of recommendation results:

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{10}$$

where $Precision = \frac{\#|RS \cap TS|}{\#|RS|}$, $Recall = \frac{\#|RS \cap TS|}{\#|TS|}$, RS is the set of products recommended to customers, and TS represents test set, the set of products that customers actually like. *Precision* measures the percentage of products in the recommendation set that are actually liked by customers, *Recall* measures that of products in the test set that are recommended to customers, and F is the harmonic mean of both *Precision* and *Recall* [23, 26].

To measure ID, *Individual_Diversity* metric defined below is used because it explicitly shows the diversification of products.

$$Individual\_Diversity = \frac{1}{N(N-1)} \sum_{i \in RS} \sum_{j \in RS, j \neq i} dis(i, j) \tag{11}$$

where RS is a recommendation list, N is the number of products in the recommendation list, and $dis(i, j)$ is the dissimilarity or distance between product i and j as defined by Eq. (4).

In terms of measuring AD, both *Diversity_in_top_N* and *Gini_diversity* are used in our experiments. The two metrics emphasize different but important aspects of AD as discussed in Sect. 2.2. They are defined as follows:

$$Diversity\_in\_top\_N = \left| \bigcup_{u \in U} RS_N(u) \right| \tag{12}$$

$$Gini\_diversity = 2 \sum_{i=1}^{|I|} \left[\left(\frac{|I| + 1 - i}{|I| + 1} \right) \times \left(\frac{rec_i}{total} \right) \right] \tag{13}$$

where N is the number of products recommended to one customer, $RS_N(u)$ is the set of products recommended to customer u, i refers to the index of a product in the product set I, rec_i is the number of customers to whom product i is recommended, products in I are arranged in non-descending order of rec_i, and $total$ indicates the total number of recommendations across all customers.

For a target customer, the user-based CF is applied to estimate his or her preferences on products. Ranking approaches are then used to rank and thus recommend suitable products to the customer. Three ranking approaches are used as our baselines, namely the standard ranking approach [25], the enhancing ID approach [44], and the enhancing AD approach [19]. The standard ranking approach arranges products in descending order of the estimated preferences and then recommends the top N ranked products to the target customer.

The enhancing ID approach first selects a candidate list S which contains k candidate products with the highest estimated preferences. The candidate products are then re-ranked according to the re-estimated preferences defined below:

$$Score_i = \theta \times R^*(u, i) - (1 - \theta) \times \max_{j \in S} sim(i, j) \tag{14}$$

$$sim(i, j) = \frac{\sum_{a \in A} w_{u(a)} \times sim_a(i, j)}{\max(w_u) \times |A|} \tag{15}$$

where θ is ranging from 0 to 1 and controls the balance between recommendation accuracy and ID, $R^*(u, i)$ is the estimated preference of user u on product i, $sim(i, j)$ is the similarity between product i and j, $sim_a(i, j)$ is the similarity between product i and j on attribute a, A is the collection of product attributes which are defined as product categories in this paper, and $w_{u(a)}$ is the importance of attribute a to user u and is decided based on the number of products previously purchased by the user and his or her propensity towards diversities on attribute a. The latter one is measured through Shannon's entropy. Further calculation details can be found in [44].

Similarly, the enhancing AD approach selects candidate products whose estimated preferences are higher than a predefined threshold T_H. The selected candidate products are then re-ranked based on the following mechanism:

$$rank(i, T_R) = \begin{cases} rank_{random}(i), & if \ R^*(u, i) \in [T_R, T_{max}] \\ \delta_u + rank_{standard}(i), & if \ R^*(u, i) \in [T_H, T_R] \end{cases} \tag{16}$$

where $\delta_u = \max_{i \in I_u(T_R)} rank_{random}(i)$ and $I_u(T_R) = \{i \in I | R^*(u, i) \geq T_R\}$. Simply put, candidate products with estimated preferences higher than the re-ranking threshold T_R are re-ranked randomly, while products that are lower than T_R are ranked using the

standard ranking approach. Besides, all candidate products which are above T_R get ranked ahead of all candidate products which are below T_R. This is ensured by δ_u.

4.3 Experimental results and analysis

To show the ability of the proposed method in balancing recommendation accuracy with ID and AD, we conduct a series of recommendations by varying the values of α, β, and γ. Figure 5 shows the resulting recommendation accuracy, ID, and AD when α, β, and γ are ranging from 0 to 1. In the experiment, the length of the recommendation list (i.e. N) is fixed at 5, and the length of the accuracy oriented ranking list (i.e. k) is fixed at 10, 20, 30, 40, and 50 respectively. And α is changing from 1 to 0 with step decrement of 0.1 while $\beta = \gamma = (1 - \alpha)/2$.

There are many other possible combinations of α, β, and γ, but we are not interested in enumerating all of them. The combinations in Fig. 5 are used to illustrate that the proposed method can be adjusted to obtain different trade-offs between recommendation accuracy and diversities. The results also demonstrate that varying k results in different trade-offs. For a specific combination of α, β, and γ, higher k leads to higher diversities and lower accuracy. Thus, the proposed method allows electronic commerce platforms to achieve expected trade-offs between recommendation accuracy and diversities by setting suitable values for α, β, γ, and k.

We also calculate the changes of F, *Iidividual_diversity*, *Gini_diversity*, and *Diversity_in_top_N* respectively when decreasing α from 1 to 0 and setting $\beta = \gamma = (1 - \alpha)/2$. Table 2 presents the results which show that the proposed method can significantly improve ID and AD at the cost of only a small decrease in recommendation accuracy. For example, the proposed method achieves 13.00%, 12.56%, and 16.43% increases in *Individual_diversity*, *Gini_diveisity*, and *Diversity_in_top_N* respectively with only a 4.08% decrease in F by changing α from 1 to 0.8 when $k=10$.

To show the superiority of the proposed method further, we compare the proposed method with the enhancing ID approach and the enhancing AD approach in terms of the changes in F, *Individual_diversity*, *Gini_diveisity*, and *Diversity_in_top_N*. In this experiment, the length of the recommendation list (i.e. N) is fixed at 5, the length of the accuracy oriented ranking list (i.e. k) is fixed at 20, and the recommendation performance of the standard ranking approach is used as a comparison standard. The parameters of the enhancing ID approach, the enhancing AD approach, and the proposed method are tuned to improve recommendation diversities at the cost of certain decreases in recommendation accuracy. The extents of the changes compared to the standard then can be calculated. The results are shown in Fig. 6. They demonstrate that the proposed method achieves higher improvements in *Individual_diversity*, *Gini_diveisity*, and *Diversity_in_top_N* than the enhancing ID approach and the enhancing AD approach when sacrificing the same amount of F.
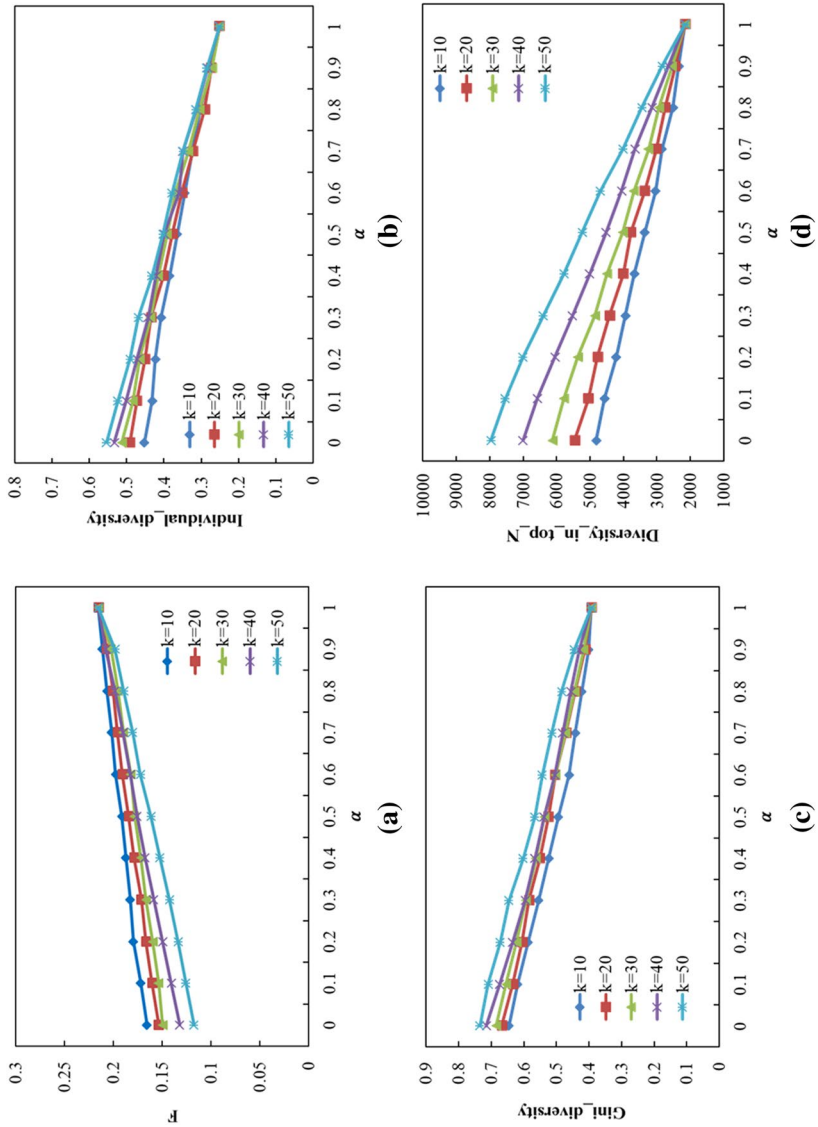
Fig. 5 The recommendation performance of the proposed method

Table 2 The changes of the performance of the proposed method

F/%	Individual_diversity/%	Gini_diversity/%	Diversity_in_top_N/%	F/%	Individual_diversity/%	Gini_diversity/%	Diversity_in_top_N/%
k=10				k=20			
− 01.90	04.93	07.27	08.56	− 03.99	08.91	04.75	13.65
− 04.08	13.00	12.56	16.43	− 07.85	11.93	10.02	27.54
− 06.36	25.05	20.04	31.94	− 10.96	20.28	17.29	39.81
− 08.31	37.07	27.49	40.74	− 13.84	28.24	25.59	55.78
− 11.14	45.14	30.93	55.78	− 16.21	42.24	30.83	75.23
− 13.09	53.22	35.79	70.37	− 17.09	52.02	35.18	85.64
− 15.18	65.31	43.75	82.63	− 20.39	65.23	43.55	104.39
− 16.58	68.45	49.47	95.13	− 22.66	75.57	51.48	121.29
− 20.15	71.79	57.51	111.34	− 25.54	84.12	58.78	133.79
− 23.03	79.99	64.25	123.14	− 28.65	95.74	67.36	153.00
k=30				k=40			
− 05.99	07.83	06.00	18.05	− 06.78	12.41	09.28	22.91
− 09.33	17.10	13.02	35.87	− 11.56	25.01	15.79	45.83
− 11.89	28.99	20.24	50.46	− 13.09	35.04	19.89	69.21
− 15.42	41.05	27.85	70.60	− 16.25	40.05	28.31	87.96
− 17.18	52.02	34.13	85.87	− 18.43	52.82	36.88	109.25
− 20.34	65.19	40.47	108.10	− 22.24	63.72	44.03	131.94
− 22.71	73.26	48.63	124.53	− 26.38	71.95	51.64	155.78
− 25.91	85.08	55.66	148.37	− 30.65	87.98	58.89	179.86
− 28.61	92.32	61.35	167.59	− 34.64	99.40	66.49	204.62
− 30.65	103.93	69.19	183.56	− 38.59	111.97	74.33	224.30
k=50							
− 07.94	17.58	10.70	31.71				
− 12.03	29.15	18.11	59.72				
− 16.34	42.64	28.28	86.11				
− 19.97	54.41	35.79	117.12				
− 25.03	60.42	44.08	142.36				
− 29.26	71.87	53.21	167.59				
− 33.76	86.35	64.28	196.06				
− 37.90	95.38	70.99	224.53				
− 41.66	108.55	80.46	248.84				
− 45.42	120.48	87.07	268.75				

5 Conclusions

In this research, we propose an adjustable re-ranking approach for improving the individual and aggregate diversities of product recommendations. This approach balances recommendation accuracy with both diversities at the product selection stage. To improve ID and AD while keeping acceptable accuracy, three ranking lists (i.e. the
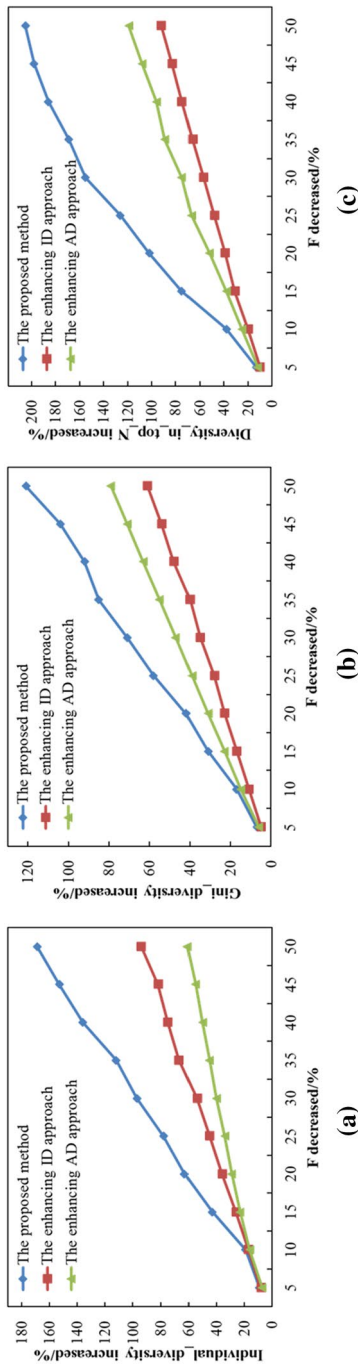
Fig. 6 The trade-offs between recommendation accuracy and diversities

accuracy, ID, and AD oriented ranking lists, respectively) are produced. The accuracy oriented ranking list consists of top k products with the highest ratings estimated by accuracy oriented recommendation methods (e.g., the user-based CF). The ID oriented ranking list copies the accuracy oriented ranking list and re-ranks the copied list by maximizing ID of the sub list with top N products. Similarly, the AD oriented ranking list copies the accuracy oriented ranking list and arranges the products in non-descending order of the degree centrality in the co-purchased network of products. The three lists are merged by a rank aggregation method, i.e. Borda's method. Experiments are conducted on a real-world data set. The experimental results demonstrate that the proposed method can achieve different trade-offs between recommendation accuracy and diversities by changing the length and the weight of the three lists. Compared to the baseline methods, the proposed method can generate much higher improvements in both diversities when sacrificing the same amount of recommendation accuracy.

The contribution of this research is twofold. First, two novel criteria are proposed to ensure ID and AD of product recommendations respectively, i.e. the distance between products in the concept hierarchy and the degree centrality of products in the co-purchased network. Second, a re-ranking approach is designed to balance recommendation accuracy with both diversities for product recommendations. Experimental evaluation demonstrates the advantages of the proposed approach. With the proposed method, ecommerce platforms can better balance recommendation accuracy with ID and AD. The proposed method may be adapted for recommendations in other contexts, such as expert recommendation and literature recommendation. In terms of expert recommendation, a researcher looking for cross-discipline collaborations may expect recommended experts with different discipline backgrounds, and a platform may want to balance the collaboration workload of experts. In this situation, a discipline classification tree can be used to ensure ID and a collaboration network can be used to guarantee AD. In the context of literature recommendation, an author writing a research paper may expect recommended citations covering different topics of the paper, and a platform may want to disseminate as many publications as possible. In this case, a research domain classification tree can be used to ensure ID and a citation network can be used to guarantee AD. Future research will investigate the generalizability of the proposed method in these recommendation contexts.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Dias, M. B., Locher, D., Li, M., El-Deredy, W., & Lisboa, P. J. G. (2008). The value of personalised recommender systems to e-business: A case study. In *Proceedings of the 2008 ACM conference on recommender systems* (pp. 291–294). New York, NY: ACM. https://doi.org/10.1145/1454008.1454054.

2. Pathak, B., Garfinkel, R., Gopal, R. D., Venkatesan, R., & Yin, F. (2010). Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems, 27*(2), 159–188. https://doi.org/10.2753/MIS0742-1222270205.

3. Ghoshal, A., Menon, S., & Sarkar, S. (2015). Recommendations using information from multiple association rules: A probabilistic approach. *Information Systems Research, 26*(3), 532–551. https://doi.org/10.1287/isre.2015.0583.

4. Zhang, Q., Wu, D., Lu, J., Liu, F., & Zhang, G. (2017). A cross-domain recommender system with consistent information transfer. *Decision Support Systems, 104,* 49–63. https://doi.org/10.1016/j.dss.2017.10.002.

5. Ma, Y., Chen, G., & Wei, Q. (2017). Finding users preferences from large-scale online reviews for personalized recommendation. *Electronic Commerce Research, 17*(1), 3–29. https://doi.org/10.1007/s10660-016-9240-9.

6. Hafshejani, Z. Y., Kaedi, M., & Fatemi, A. (2018). Improving sparsity and new user problems in collaborative filtering by clustering the personality factors. *Electronic Commerce Research, 18,* 1–24. https://doi.org/10.1007/s10660-018-9287-x.

7. Jing, N., Jiang, T., Du, J., & Sugumaran, V. (2018). Personalized recommendation based on customer preference mining and sentiment assessment from a Chinese e-commerce website. *Electronic Commerce Research, 18*(1), 159–179. https://doi.org/10.1007/s10660-017-9275-6.

8. Fleder, D., & Hosanagar, K. (2009). Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science, 55*(5), 697–712. https://doi.org/10.1287/mnsc.1080.0974.

9. Brynjolfsson, E., Hu, Y., & Simester, D. (2011). Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science, 57*(8), 1373–1386. https://doi.org/10.1287/mnsc.1110.1371.

10. Kunaver, M., & Požrl, T. (2017). Diversity in recommender systems—A survey. *Knowledge-Based Systems, 123,* 154–162. https://doi.org/10.1016/j.knosys.2017.02.009.

11. Oestreicher-Singer, G., & Sundararajan, A. (2012). Recommendation networks and the long tail of electronic commerce. *MIS Quarterly, 36*(1), 65–83.

12. Chen, L., Zhang, G., & Zhou, H. (2017). *Improving the diversity of top-N recommendation* via *determinantal point process*. arXiv:1709.05135 [cs]. Retrieved from http://arxiv.org/abs/1709.05135.

13. Zhang, M., & Hurley, N. (2008). Avoiding monotony: Improving the diversity of recommendation lists. In *Proceedings of the 2008 ACM conference on recommender systems* (pp. 123–130). New York, NY: ACM. https://doi.org/10.1145/1454008.1454030.

14. Muter, I., & Aytekin, T. (2017). Incorporating aggregate diversity in recommender systems using scalable optimization approaches. *INFORMS Journal on Computing, 29*(3), 405–421. https://doi.org/10.1287/ijoc.2016.0741.

15. Karakaya, M. Ö., & Aytekin, T. (2017). Effective methods for increasing aggregate diversity in recommender systems. *Knowledge and Information Systems, 56,* 1–18. https://doi.org/10.1007/s10115-017-1135-0.

16. Gan, M., & Jiang, R. (2013). Improving accuracy and diversity of personalized recommendation through power law adjustments of user similarities. *Decision Support Systems, 55*(3), 811–821. https://doi.org/10.1016/j.dss.2013.03.006.

17. Gogna, A., & Majumdar, A. (2017). Balancing accuracy and diversity in recommendations using matrix completion framework. *Knowledge-Based Systems, 125,* 83–95. https://doi.org/10.1016/j.knosys.2017.03.023.

18. Adomavicius, G., & Kwon, Y. O. (2009). Toward more diverse recommendations: Item re-ranking methods for recommender systems. In *19th workshop on information technologies and systems, WITS 2009* (pp. 79–84). Presented at the 19th workshop on information technologies and systems, WITS 2009, social science research network. Retrieved from https://experts.umn.edu/en/publications/toward-more-diverse-recommendations-item-re-ranking-methods-for-r.

19. Adomavicius, G., & Kwon, Y. (2012). Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering, 24*(5), 896–911. https://doi.org/10.1109/TKDE.2011.15.

20. Adomavicius, G., & Kwon, Y. (2014). Optimization-based approaches for maximizing aggregate recommendation diversity. *INFORMS Journal on Computing, 26*(2), 351–369. https://doi.org/10.1287/ijoc.2013.0570.

21. Ding, S., Xia, C., Wang, C., Wu, D., & Zhang, Y. (2017). Multi-objective optimization based ranking prediction for cloud service recommendation. *Decision Support Systems, 101,* 106–114. https://doi.org/10.1016/j.dss.2017.06.005.

22. Adomavicius, G., & Tuzhilin, A. (2005). 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, 17*(6), 734–749. https://doi.org/10.1109/TKDE.2005.99.

23. Wang, Q., Ma, J., Liao, X., & Du, W. (2017). A context-aware researcher recommendation system for university-industry collaboration on R&D projects. *Decision Support Systems, 103,* 46–57. https://doi.org/10.1016/j.dss.2017.09.001.

24. Burke, R. (2007). Hybrid web recommender systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The adaptive web* (pp. 377–408). Berlin: Springer. https://doi.org/10.1007/978-3-540-72079-9_12.

25. Breese, J. S., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the fourteenth conference on uncertainty in artificial intelligence* (pp. 43–52). San Francisco, CA: Morgan Kaufmann Publishers Inc. Retrieved from http://dl.acm.org/citation.cfm?id=2074094.2074100.

26. Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS), 22*(1), 5–53. https://doi.org/10.1145/963770.963772.

27. Shani, G., & Gunawardana, A. (2011). Evaluating recommendation systems. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), *Recommender systems handbook* (pp. 257–297). Boston, MA: Springer. https://doi.org/10.1007/978-0-387-85820-3_8.

28. Ziegler, C.-N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web* (pp. 22–32). New York, NY: ACM. https://doi.org/10.1145/1060745.1060754.

29. Niemann, K., & Wolpers, M. (2013). A new collaborative filtering approach for increasing the aggregate diversity of recommender systems. In *Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 955–963). New York, NY: ACM. https://doi.org/10.1145/2487575.2487656.

30. Gan, M. (2014). Walking on a user similarity network towards personalized recommendations. *PLoS ONE, 9*(12), e114662. https://doi.org/10.1371/journal.pone.0114662.

31. Wang, J., Liu, Y., Sun, J., Jiang, Y., & Sun, C. (2016). Diversified recommendation incorporating item information based on MOEA/D. In *2016 49th Hawaii international conference on system sciences (HICSS)* (pp. 688–696). Presented at the 2016 49th Hawaii international conference on system sciences (HICSS). https://doi.org/10.1109/hicss.2016.91.

32. Deb, K., & Deb, K. (2014). Multi-objective Optimization. In E. K. Burke & G. Kendall (Eds.), *Search methodologies* (pp. 403–449). Boston, MA: Springer. https://doi.org/10.1007/978-1-4614-6940-7_15.

33. Konak, A., Coit, D. W., & Smith, A. E. (2006). Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering and System Safety, 91*(9), 992–1007. https://doi.org/10.1016/j.ress.2005.11.018.

34. Chien, B.-C., Hu, C.-H., & Ju, M.-Y. (2009). Learning fuzzy concept hierarchy and measurement with node labeling. *Information Systems Frontiers, 11*(5), 551. https://doi.org/10.1007/s10796-008-9126-z.

35. Hsu, F.-M., Lu, L.-P., & Lin, C.-M. (2012). Segmenting customers by transaction data with concept hierarchy. *Expert Systems with Applications, 39*(6), 6221–6228. https://doi.org/10.1016/j.eswa.2011.12.005.

36. Weng, S.-S., & Chang, H.-L. (2008). Using ontology network analysis for research document recommendation. *Expert Systems with Applications, 34*(3), 1857–1869. https://doi.org/10.1016/j.eswa.2007.02.023.

37. Girardi, D., Wartner, S., Halmerbauer, G., Ehrenmüller, M., Kosorus, H., & Dreiseitl, S. (2016). Using concept hierarchies to improve calculation of patient similarity. *Journal of Biomedical Informatics, 63,* 66–73. https://doi.org/10.1016/j.jbi.2016.07.021.
38. Budanitsky, A., & Hirst, G. (2006). Evaluating WordNet-based measures of lexical semantic relatedness. *Computational Linguistics, 32*(1), 13–47. https://doi.org/10.1162/coli.2006.32.1.13.
39. Lika, B., Kolomvatsos, K., & Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert Systems with Applications, 41*(4, Part 2), 2065–2073. https://doi.org/10.1016/j.eswa.2013.09.005.
40. Tucker, C., & Zhang, J. (2007). Long tail or steep tail? A field investigation into how online popularity information affects the distribution of customer choices. MIT Sloan Working Paper 4655–07, Cambridge. http://hdl.handle.net/1721.1/39811.
41. Borgatti, S. P., & Everett, M. G. (2006). A Graph-theoretic perspective on centrality. *Social Networks, 28*(4), 466–484. https://doi.org/10.1016/j.socnet.2005.11.005.
42. Kim, H. K., Kim, J. K., & Chen, Q. Y. (2012). A product network analysis for extending the market basket analysis. *Expert Systems with Applications, 39*(8), 7403–7410. https://doi.org/10.1016/j.eswa.2012.01.066.
43. Dwork, C., Kumar, R., Naor, M., & Sivakumar, D. (2001). Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web* (pp. 613–622). New York, NY: ACM. https://doi.org/10.1145/371920.372165.
44. Di Noia, T., Ostuni, V. C., Rosati, J., Tomeo, P., & Di Sciascio, E. (2014). An analysis of users' propensity toward diversity in recommendations. In *Proceedings of the 8th ACM conference on recommender systems* (pp. 285–288). New York, NY: ACM. https://doi.org/10.1145/2645710.2645774.