Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 4, pp. 2358–2370, DOI 10.12694/scpe.v25i4.2812

BOOK CLASSIFICATION AND RECOMMENDATION FOR UNIVERSITY LIBRARIES USING GREY CORRELATION AND BAYESIAN PROBABILITY

FANG CUI* AND LIANG GAO†

Abstract. In today's information era, collaborative filtering algorithms are widely used and their distinct knowledge discovery techniques can effectively address numerous issues. However, conventional collaborative filtering algorithms encounter cold-start and data sparsity issues, which restrict their performance and accuracy. The study selected the multi-feature method to improve the traditional collaborative filtering algorithm, and introduced gray correlation calculation and Bayesian probability for user preference analysis. A learning resource recommendation model based on collaborative filtering was developed by comparing the target user's characteristics with those of other users, calculating their similarity, selecting users with high similarity to the target user and forming a neighbor set. Using Bayesian probability and grey correlation to analyze user preferences in library systems can be well applied in book classification and recommendation problems in university libraries. The computing layer, which includes the collaborative filtering calculation stage and the group recommendation calculation stage, is the model's main functional component. The smaller the value of mean absolute error, the higher the prediction accuracy of the model. The mean absolute error value of the multi-feature collaborative filtering algorithm was inferior to the traditional collaborative filtering algorithm, indicating that the classification accuracy of the former is higher than that of the latter. When the training set to test set ratio steadily became bigger, the mean absolute error value reached the lowest and smoothest point at 80%. In dataset A, the minimum mean absolute error values of multi-feature collaborative filtering and collaborative filtering were 0.765 and 0.809. Compared with traditional filtering algorithms, the mean absolute error value has decreased by 0.044. In dataset B, the mean absolute error values of multi-feature collaborative filtering and collaborative filtering were 0.796 and 0.836. Compared with traditional filtering algorithms, the mean absolute error value has decreased by 0.040. In dataset C, the minimum mean absolute error values of multi-feature collaborative filtering and collaborative filtering were 0.815 and 0.848. Compared with traditional filtering algorithms, the mean absolute error value has decreased by 0.033. When the accuracy was the highest; the mean absolute error value was the smallest at the grey correlation, which means that the technique improves the reliability of the recommendations compared with other methods. This means that the method has a positive impact on the accuracy of the recommendations compared to other methods. Grey correlation degree can comprehensively consider the interrelationships between multiple factors, handle uncertain and incomplete information, and explore potential user needs and behavior patterns. The implementation of the grey correlation degree has transformed the collaborative filtering algorithm into a group filtering algorithm, thereby enhancing its precision. The research on book classification and recommendation in university libraries, which enhances the group filtering algorithm, can address a range of issues such as improving classification accuracy, augmenting recommendation diversity, enhancing library management efficiency among others. This, in turn, enables more precise book recommendations to users.

Key words: Collaborative filtering system; Grey correlation; Multi-feature algorithm; Book recommendation

1. Introduction. For libraries, museums and other databases with large volumes of books, accurate and precise classification and recommendation systems are necessary [1]. Collaborative filtering (CF) algorithms are widely used in many recommendation systems [2]. The unique knowledge discovery technology of resource recommendation can identify users with similar information by examining their past history, forming a group of similar neighbors. And based on the past data in the collection, customized suggestions can be provided to target users to solve problems such as information overload [3]. However, data sparsity is a challenge for the conventional CF method. As time passes and the number of users grows, the system accumulates more data, the computational workload increases in size and complexity, and the users' ratings of items decrease. Consequently, the system experiences greater sparsity, which significantly impacts its recommendation accuracy and performance. It is difficult to give accurate recommendations to new users with no or sparse rating data, which is a problem that is difficult to ignore [4, 5, 6]. To address the problem that data sparsity can lead to lower accuracy and diversity of collaborative filtering recommendation algorithms, Yan H et al. used

*Library, Inner Mongolia University, Hohhot 010021, China (ndcuifang@imu.edu.cn)
†Library, Suqian University, Suqian 223800, China

a granularity computation model to achieve nearest neighbor clustering and proposed a coverage roughness granularity computation model for optimization of collaborative filtering recommendation algorithms to solve the problem [7]. Yang, Y et al. proposed the design and application of a handicraft recommendation system based on an improved hybrid algorithm. Based on the theory of e-commerce system, a personalized e-commerce system with hybrid algorithm is designed and analyzed by traditional user collaborative filtering algorithm. Further proposed a personalized recommendation system for e-commerce based on hybrid algorithm [8]. To enhance the constraints in the system, the paper adopted a multi-feature strategy to enhance the standard CF algorithm. It incorporated grey correlation into the recommendation system for computation and utilized Bayesian probability to analyze user preferences. The system obtained a set of neighbors with high similarity to the user to supplement the recommendation process, alongside constructing a CF-based learning resource recommendation model. Multi-feature collaborative filtering (MCF) calculation and group recommendation computation were applied to learning recommendation resources. Also for the problem of the accuracy of feature selection caused by data sparsity, the introduction of gray correlation computation and Bayesian probability to improve the user preference-based recommend system chosen in this paper has the advantage that: the gray correlation computation can excavate the regularity and characteristics of the data, and the Bayesian probability can provide a more accurate recommend result through the a priori knowledge and the user's feedback; both of these two methods have a certain degree of robustness, which can reduce the impact of external interference on the recommend result.

The innovation of the research on book classification and recommendation in university libraries based on group filtering algorithms lies in its ability to break through traditional methods, improve recommendation accuracy, promote interdisciplinary integration, and achieve dynamic recommendations. It can better meet the personalized needs of readers, improve the service quality and borrowing rate of libraries.

2. Related Works. To address the problems in using multiple types of representations in a knowledge graph recommendation scenario, Wu et al. suggested a new method, which combines the benefits of path-based and propagation-based approaches [9]. A method for classifying and retrieving items in hierarchy was designed by Li et al. A merchandise matching recommendation algorithm was proposed to analyze the research status of image retrieval method. Correspondence between various item visual characteristics was applied for item recommendation [10]. Li et al. proposed an algorithm for recommending explanation documents with questions and answers to alleviate the problem of CQA websites. The Q&A documents were modeled using a two-terms topic model and the Q&A documents were clustered using a growing neural gas algorithm [12]. Li et al. proposed a balanced resource recommendation engine to address the issues of resource imbalance and low trust in traditional sports online education resource recommendations. The sports network's instructional materials were first classified using the Support Vector Machine (SVM) method, and then any incorrect information was eliminated [12]. Liang et al. suggested an image recommendation algorithm in social networks. Deep neural networks showed a wide range of successful applications in various fields [13].

Deep networks have recently achieved good performance in classification tasks. When the training set is very small, the behaviour of this model diminishes significantly. Classifiers based on linear representations have been widely used in many fields. Building on these observations, Du et al. suggested a new competitive and categorization via cooperative representation that uses properties of training data with L2 parametric regularization to create a competitive environment that allows the correct class to contribute more to the encoding [14]. Chi et al. proposed a new CRC-based classifier utilizing class-mean weighted discriminative co-representation [15]. Currently, classification techniques have become increasingly important. Classifiers based on collaborative representation have been applied to many practical cognitive domains due to their advantages in terms of efficiency and effectiveness. A new neighbourhood prior constrained collaborative representation model was suggested by Gou et al. The guidance of the neighbourhood prior in the encoding process was emphasized [16]. Hyper-spectral pictures' extensive spectral data enabled several applications with tremendous advantages. The RPnet model overcame the redundant bands with a few training sample numbers [17]. Wu et al. proposed a weighted multi-view cooperative fuzzy C-mean clustering algorithm (DPSO-WCoFCM). The weighted clustering method's clustering centers were optimized using the common particle swarm optimization technique in conjunction with the WCoFCM algorithm [18].

In summary, many scholars are interested in recommendation and classification systems and combine various

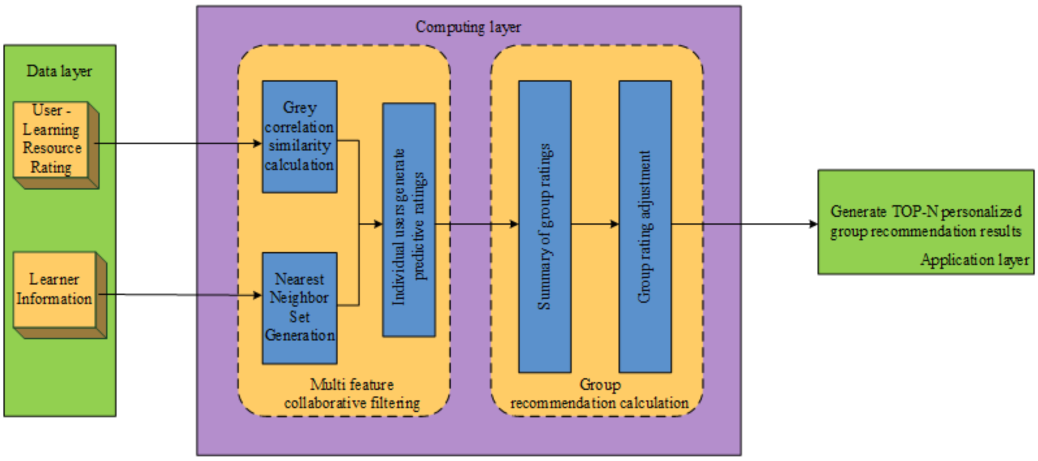


Fig. 3.1: The Group Recommendation Model for Learning Resources

algorithms with them to optimize the efficiency and accuracy of system recommendation and classification. There is relatively little research on combining collaborative algorithms with classification recommendation systems. Therefore, improving collaborative algorithms and combining them to recommend and classify library books is a worthwhile research direction. Based on the analysis of user attributes, the MCF algorithm has been improved through the optimization of feature selection and extraction to overcome the limitations of conventional group filtering algorithms. The benefit of the enhanced CF algorithm suggested in this study lies in its ability to contemplate the interaction between various features concurrently. This approach produces more comprehensive recommendation outcomes and caters to the assorted requirements of users. Nevertheless, a setback is that the dataset size employed in the experiment might contrast with the size of a real-world dataset, impeding full and precise simulation.

3. Collaborative population filtering algorithm for book recommendation model and classification in university libraries applications.

3.1. A CF-based book recommendation model for university libraries. The CF algorithm offers several benefits in recommendation models. However, with growing computational volume and increasing data complexity, the algorithm confronts limitations, particularly regarding cold starts and data sparsity [19]. The cold start problem pertains to the challenge of traditional recommendation algorithms in producing precise personalized recommendations for new users or projects in recommendation systems because of the lack of data and inadequate information. For new users, the system lacks sufficient personal preference data. The problem of data sparsity in recommendation systems is characterized by a large number of users and items, yet a relatively small amount of interaction data between them. This poses a challenge for accurately assessing similarity between users and items, thereby impacting recommendation accuracy. In the face of the problems of group recommendation and learning resource recommendation, CF algorithm is used to build a group recommendation model for learning resources based on the idea of hierarchical software engineering systems, as shown in Figure 3.1.

In Figure 3.1, the computational layer of the learning resource group recommendation model plays a key role. This layer comprises a CF computation phase and a group recommendation computation phase. The MCF calculation phase shown in Figure 3.1 is based on this algorithm and is calculated in three main steps: grey correlation, nearest neighbour generation, and prediction ratings generated by individual users. Within this model, there is a MCF of diverse recommendation lists with individuals as the focus, and the recommendation lists produced by the learning resource group recommendation model comprise solely of the predicted ratings of individual users that can be utilized as a group. The group suggestion framework generates

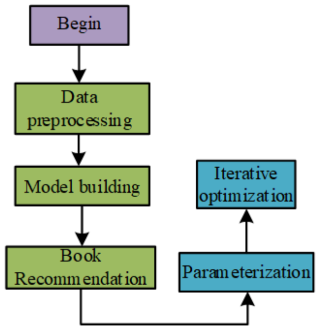


Fig. 3.2: Flow chart of MCF algorithm

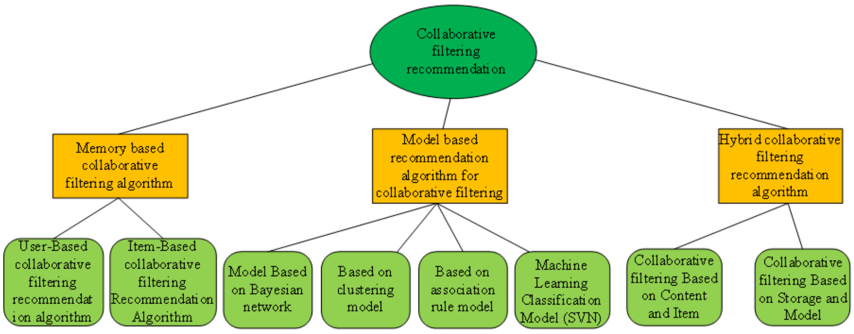


Fig. 3.3: Classification of CF algorithms

a recommendation list solely based on the predicted ratings of individual users. This list can serve as input data for calculating group recommendations. The MCF algorithm is an enhancement of the CF algorithm, which possesses a succinct operational approach capable of remedying intricate and varied recommendation issues. The algorithm extensively utilizes multiple features and requires processing vast amounts of data and precise parameter tuning. In practical applications, it is essential to choose suitable multi-feature algorithms depending on specific tasks and data characteristics. Such algorithms have found widespread application in various internet products [20]. The flowchart of the MCF algorithm is shown in Figure 3.2.

In Figure 3.2, there are five steps: data preprocessing, model construction, book recommendation, parameter adjustment, and iterative optimization. Figure 3.3 shows the memory-based, model-based, and hybrid CF algorithms.

The memory-based recommendation algorithm offers several advantages over earlier similar systems. It is more widely used, efficient, and easier to implement in its application. The algorithm differs from earlier similar systems in that existing historical data on user ratings are collected together as a training set and a training model is built for predictive evaluation. The initial step in the data representation stage involves formalizing the recommendation problem by dichotomizing the recommendation system through a dichotomous network, as illustrated in Figure 3.4.

In the dichotomous network, the users of the personalized recommendation system are regarded as the set of user nodes U and the whole item as the set of item nodes O . The relationship between the U and O is a set of edges with weights. All the values taken by the item's user evaluation are represented by the set of weights of the edges between the item node set and the user node set by , as shown in equation 3.1.

$$U \times O \to \{R_{u_a}(o_k)\}_{m\times n} \tag{3.1}$$

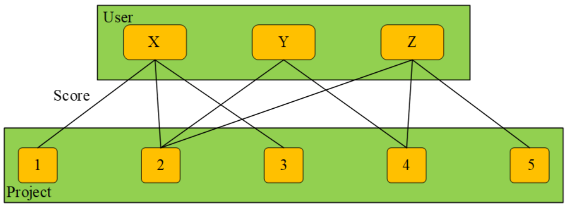


Fig. 3.4: Binary network of recommendation systems

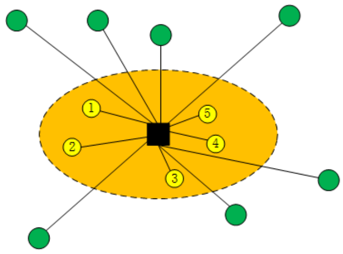


Fig. 3.5: An illustration of the closest neighbour set

In equation 3.1, $\{R_{u_a}(o_k)\}$ is the rating of the item by the user o_k and u_a. The nearest neighbor query constructs a set of nearest neighbors for the user and utilizes this information in the subsequent stage to generate recommendations. The diagram for constructing the set of users' neighbors is shown in Figure 3.5.

In Figure 3.5, the rectangle in the middle of the graph is the target user. The spheres at the end points of the line segment represent users with similarity to the target user. The similar users within the dashed box ellipse constitute the set of nearest neighbors U . Weighted similarity calculations are the typical calculation techniques, Pearson similarity calculation and cosine exponential similarity calculation. The formula for the weighted similarity calculation is shown in equation 3.2.

$$sim_{(u_\alpha,u_\beta)} = \frac{\sum_{o_k \in o(u_\alpha,u_\beta)} R_{u_a}(o_k) R_{u_\beta(o_k)}}{|o_k|_{o_k \in o(u_\alpha,u_\beta)}} \tag{3.2}$$

As shown in equation 3.2, $|o_k|_{o_k \in o(u_\alpha,u_\beta)}$ denotes the quantity of items for which the user u_α, u_β has a common rating on the item o_k. $o_k \in o(u_\alpha, u_\beta)$ denotes the set of items for which the user u_α, u_β has a ordinary rating. The technique for calculating Pearson similarity focuses on the correlation between vectors and is shown in equation 3.3.

$$sim_{(u_\alpha,u_\beta)} = \frac{\sum_{o_k \in o(u_\alpha,u_\beta)} (R_{u_\alpha}(o_k) - \overline{R_{u_\alpha}})(R_{u_\beta}(o_k) - \overline{R_{u_\beta}})}{\sqrt{\sum_{o_k \in o(u_\alpha,u_\beta)} (R_{u_\alpha}(o_k) - \overline{R_{u_\alpha}})^2} \sqrt{\sum_{o_k \in o(u_\alpha,u_\beta)} (R_{u_\beta}(o_k) - \overline{R_{u_\theta}})^2}} \tag{3.3}$$

As shown in equation 3.3, $\overline{R_{u_\alpha}}$ denotes the mean value of user u_α's rating and $\overline{R_{u_\beta}}$ is the mean value of user u_β's rating. The cosine index similarity calculation method uses the user vector as the unit of calculation by measuring the tutor cosine value between the user vectors, which is calculated as shown in equation 3.4.

$$sim_{(u_\alpha,u_\beta)} = \frac{\overline{R_{u_\alpha}} \cdot \overline{R_{u_\alpha}}}{\|R_{u_\alpha}\| \times \|R_{u_\alpha}\|} \tag{3.4}$$

In equation 3.4, R_{u_α} represents the rating vector of user u_α and R_{u_β} represents the rating vector of user u_β. The list of recommendation results is obtained in two ways, namely the weighted sum approach and the

Table 3.1: User Learning Resource Rating

Field name	Data type	Field Description
U-id	Int	User ID
I-id	Int	Learning Resource Number
U-Rate	Int	User Learning Resource Rating
G-Rate	Double	Group Learning Resource Rating
R-time	Datatimme	Scoring time

weighted average method. The formula for the weighted summation method of predicted ratings is shown in equation (5).

$$P_{u_\alpha}^{(o_k)} = R_{u_\alpha} \frac{\sum_{u_\beta \in \alpha} sim(u_\alpha, u_\beta) \times R_{u_\beta}(o_k) - \overline{R_{u_\beta}}}{\sum_{u_\beta \in \alpha} sim(u_\alpha, u_\beta)} \tag{3.5}$$

In equation 3.5, $sim(u_\alpha, u_\beta)$ denotes the similarity between user u_α to be recommended and user u_β in its neighbor set A. $\overline{R_{u_\alpha}}$ and $\overline{R_{u_\beta}}$ represent the average score of the project. The formula for calculating the weighted average of predicted scores is shown in equation (6).

$$P_{u_\alpha}^{(o_k)} = \frac{\sum_{u_\beta \in A} sim(u_\alpha, u_\beta) \times R_{u_\beta}(o_k)}{\sum_{u_\beta \in \alpha} sim(u_\alpha, u_\beta)} \tag{3.6}$$

The initial stage of incorporating the individual predicted scores of call signs from the multi-feature synergy phase into the group recommendation calculation is to amalgamate the group's scores. This calculation can be found in equation 3.7.

$$R_{G_i}(O_k) = \frac{1}{|G_i|} \sum_{j=1}^{|G_i|} R_{u_j}(O_k) \tag{3.7}$$

In equation 3.7, G_i represents the group of users in the learning resource recommendation. $R_{G_i}(O_k)$ represents the rating of the learning resource O_k by the group G_i. $|G_i|$ is the total number of users in the group for learning resource recommendations. $R_{G_i}(O_k)$ represents the rating of the item O_k by the individual user u_j. The second step in the recommendation calculation is the group rating adjustment, which is calculated as shown in equation (8).

$$d(G_i(O_k)) = \frac{\sum \sqrt{(R_{u_i}(O_k) - R_{u_j}(O_k))^2}}{|G_i|(|G_i| - 1)/2} \tag{3.8}$$

In equation 3.8, $d(G_i(O_k))$ represents the difference in the ratings of the learning resource recommendation group G_i against the learning resource O_k . The aggregated group ratings of the learning resources are then adjusted to give the final adjusted group ratings of against the learning resource, as shown in equation (9).

$$r_{G_i}(O_k) = w_1 \cdot R_{G_i}(O_k) + w_2 \cdot (1 \rightarrow d(G_i(O_k))) \tag{3.9}$$

In equation 3.9, w_1 is the percentage of aggregated ratings in the group ratings. w_2 is the percentage of variance accounted for in the group ratings, and $r_{G_i}(O_k)$ represents the adjusted ratings of G_i for the learning resources O_k in the group recommendations of learning resources. The computation is modular and divided into three computational layers. The input data in the data layer contains information on user-learning-resource ratings and learner characteristic attributes. Table 3.1 demonstrates the user-learning-resource table for the data layer.

The MCF learning resource recommendations initially score the computational layer. Subsequently, the group learning resource recommendations aggregate and adjust the scores of the user-learning resources. Finally, the resulting scores are incorporated into the learning resource scores in the data layer. The primary responsibility of the application layer is to present user and group data, and to arrange recommendations in a decreasing order based on adjusted group learning resources.

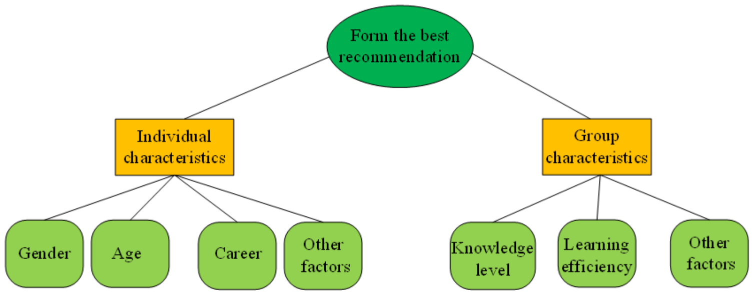


Fig. 4.1: Schematic diagram of user feature selection

4. Improved MCF algorithm design. CF algorithm has the limitations of difficulties with data sparsity and cold start. The problem with cold starts occurs in the system when recommending accurately becomes challenging with new users who lack rating data or have sparse data. This problem cannot be ignored during operation, making it a difficult issue to address [21] . The issue of data sparsity is predominantly a result of the increasing number of users and data stored over time. As a result, calculating and rating the items becomes more challenging, and the number of users contributing to these ratings decreases. This leads to greater sparsity, which directly impacts the accuracy and overall performance of the system [22] To address the common problems in CF algorithms, a fresh approach to calculating similarity was chosen to enhance the algorithm by incorporating user attribute analysis to reduce the impact of slow start and data shortage issues. The formula for calculating the grey correlation degree is shown in equation 4.1.

$$\gamma_{u_\alpha,u_\beta}^{o_k} = \frac{min_{o_k \in o(u_\alpha,u_\beta)}|R_{u_\alpha}(o_k) - R_{u_\beta}(o_k)| + \xi max_{o_k \in o(u_\alpha,u_\beta)}|R_{u_\alpha}(o_k) - R_{u_\beta}(o_k)|}{|R_{u_\alpha}(o_k) - R_{u_\beta}(o_k)| + \xi max_{o_k \in o(u_\alpha,u_\beta)}|R_{u_\alpha}(o_k) - R_{u_\beta}(o_k)|} \tag{4.1}$$

In equation 4.1, $\xi \in (0,1)$, and $o(u_\alpha, u_\beta)$ indicates the group of objects that the recommended users and peculiar users are rated by together. Finally, the grey correlation of users is calculated and its formula is shown in equation 4.2.

$$\gamma_{(u_\alpha,u_\beta)} = \frac{\sum_{o_k \in O(u_\alpha,u_\beta)} r_{(u_\alpha,u_\beta)}^{o_k}}{|o_k|_{o_k \in o(u_\alpha,u_\beta)}} \tag{4.2}$$

In equation 4.2, $r_{(u_\alpha,u_\beta)}$ is the number of grey associations of users u_α, u_β, about items o_k , and $|o_k|_{o_k \in o(u_\alpha,u_\beta)}$ is the number of items with common ratings of users u_α and u_β. The multi-feature closest neighbour set modifies the ranking based on similarity by include user multi-feature qualities in the similarity computation, and its calculation steps include: attribute extraction, pre-processing, and adjusting the nearest neighbour set selection. The user multi-feature attribute extraction includes many evaluation criteria, and the individual features and group features are extracted by hierarchical analysis, and the features are selected as shown in Figure 3.6.

User multi-trait attributes cannot be calculated directly, they should be converted into orders of magnitude before they can be calculated. Equation 4.3 illustrates the formula for determining a user's interest in a particular item.

$$p(u_\alpha, u_\beta) = \frac{\sum_{u_\beta \in U_{o_k}} \gamma(u_\alpha, u_\beta)}{|U_{o_k}|} \tag{4.3}$$

As shown in equation (12), $\gamma(u_\alpha, u_\beta)$ is the user's grey correlation similarity. $|U_{o_k}|$ is the set of users who have scored item o_k. $|U_{o_k}|$ is the number of users who have scored item o_k. From the Bayesian probability, the probability of a user liking an item when the user characteristic $c_m \in u_\alpha$ is shown in equation 4.4.

$$P_{o_k}^{u_\alpha}(M|c_m) = \frac{P_{o_k}(c_m|M)P_{o_k}(M)}{P_{o_k}(c_m|M)P_{o_k(M)} + P_{o_k}(c_m|N)P_{o_k(N)}} \tag{4.4}$$

In equation 4.4, $P_{o_k}(c_m|M)$ is the likelihood that an element will appear in the occurrence M and $P_{o_k}(M)$ is the likelihood that a consumer will like a product. The likelihood that a user would adore a product when the user features is shown in equation (14).

$$P_{o_k}^{u_\alpha}(M|C_a) = \frac{P_{o_k}^{u_\alpha}(M|C_1)P_{o_k}^{u_\alpha}(M|C_2)\dots P_{o_k}^{u_\alpha}(M|C_m)}{\prod_{i=1}^{m} P_{o_k}^{u_\alpha}(M|C_1) + \prod_{i=1}^{m}(1 - P_{o_k}^{u_\alpha}(M|C_i))} \tag{4.5}$$

The first task in adjusting the nearest neighbour set selection is to adjust the similarity of the users, based on the theory shown in equation 4.6.

$$nsim(u_\alpha, u_\beta) = \begin{cases} P_{o_k}^{u_\alpha} * \gamma(u_\alpha, u_\beta), u_\alpha \in U_{like} \\ (1 - P_{o_k}^{u_\alpha}(M|c_a)) * \gamma(u_\alpha, u_\beta), u_\alpha \in U_{unlike} \end{cases} \tag{4.6}$$

As shown in equation 4.6, $P_{o_k}^{u_\alpha}(M|c_a))$ is the likelihood that a user would adore a product when the user features and $\gamma(u_\alpha, u_\beta)$ is the grey correlation of the user. The list of recommendations is generated for the user and arranged in descending order of predicted value. The prediction score, calculated in accordance with equation 4.7, is based on grey correlation.

$$P_{o_k}^{u_\alpha} = \overline{R_{u_\alpha}} + \frac{\sum_{U_\beta \in A} nsim(u_\alpha, u_\beta) \times R_{u_\beta}(o_k - \overline{R_{u_\beta}})}{\sum_{U_\beta \in A} nsim(u_\alpha, u_\beta)} \tag{4.7}$$

In equation 4.7, $nsim(u_\alpha, u_\beta)$ is the adjusted similarity between the recommended users and the neighbouring users in the neighbourhood set. $\overline{R_{u_\alpha}}$ and $\overline{R_{u_\beta}}$ represent the mean item ratings of the two users in the user-item rating matrix, and $R_{u_\beta(o_k)}$ is the user ratings of the items in the user-item rating matrix.

5. Results of testing the recommendation capability of the multi-feature group filtering algorithm.

5.1. Effect of the ratio of dataset algorithm to training set on recommendation accuracy. Faced with cold start issues and data scarcity issues in CF algorithm, the study proposes a grey correlation similarity algorithm for improvement, conducts simulation experiments on the improved MCF algorithm, and analyzes the results. Common evaluation criteria used in recommendation engines to measure the precision of the system are selected for the experiments. To examine the impact of various datasets on the recommendation algorithm's accuracy, comparison experiments are conducted for each dataset. Therefore, data of the same size as the data set A was selected from the three data sets A, B and C. 940 data sets were selected from the user pool, while 1660 data sets were chosen from the item group to serve as the test data for this experiment. One-fifth of the test set was randomly chosen as the test set, while the remaining four-fifths of the data comprised the training set. The validity of the recommendation algorithm was tested on three different datasets. According to the aforementioned experimental criteria and different datasets, the experimental outcomes were segregated into three categories: groups A, B, and C. The validity of the algorithm was compared with the traditional CF algorithm and MCF algorithm, and the experimental results obtained are shown in Figure 5.1.

As depicted in Figure 5.1, the graph of experimental results with the data set as a variable demonstrates that the MCF algorithm has considerably higher accuracy compared to the traditional CF algorithm. It is verified that the recommendation accuracy of the MCF algorithm has been significantly improved compared with that of the traditional CF algorithm. To investigate the effect of different training set to test set ratios, i.e. data sparsity on performance, a comparative study of the recommendation accuracy obtained with different test set training set ratios was conducted. The experimental group's training set to test set ratio was increased from 0.1 to 0.9 to compare algorithmic recommendation accuracy with various data sparsity ratios. A dataset with 50 predetermined neighbors was utilized for experimentation purposes. The objective was to examine the impact of varying ratios between the training and testing sets on the performance of the recommendation algorithm. This allowed for the assessment of the algorithm's effectiveness with different levels of data sparsity. To test the validity of different sparsity of the data set on the recommendation accuracy, three test sets A, B and C were selected. The ratio of the training set to the test set was set to 0.1 as the initial ratio, and the

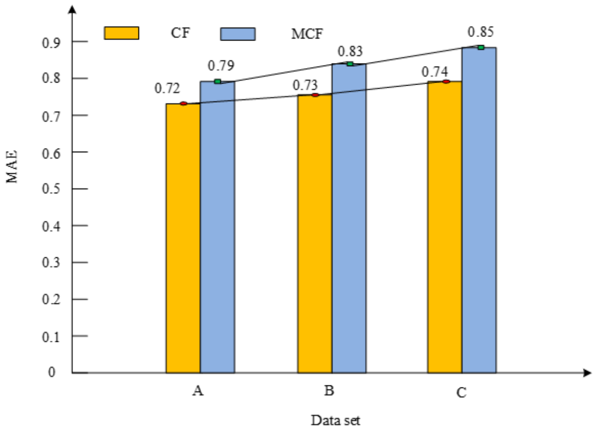
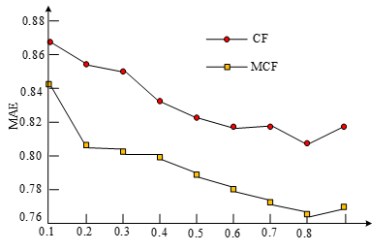
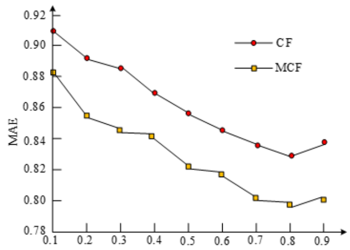


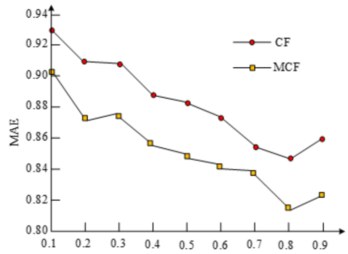
Fig. 5.1: Comparison of Algorithm Accuracy under Different Datasets



(a) Performance testing under different sparsities(Date A)



(b) Performance testing under different sparsities(Date B)



(c) Performance testing under different sparsities(Date C)

Fig. 5.2: Performance testing under different sparsities (Data A and B)

interval was set to 0.1, which was gradually increased until the data sparsity ended at 0.9. The accuracy of the recommendations and advice decreased as the average mistake value increased. The results of the Mean Absolute Error (MAE) values of the MCF and the traditional CF algorithms with different proportions of training sets in the A, B and C data sets are shown in Figure 5.2.

In Figure 5.2, the average error value of CF algorithm gradually leveled off as the ratio of the training set to the test set increased steadily for two different sized data sets A, B and C. The average error value of CF algorithm tended to be optimal when the training set was 0.8. The average error value is minimized when the training set is 0.8. In dataset A, the minimum MAE values of MCF and CF were 0.765 and 0.809. Compared with traditional filtering algorithms, the MAE value has decreased by 0.044. In dataset B, the MAE value of MCF and CF was 0.796 and 0.836. Compared with traditional filtering algorithms, the MAE value has decreased by 0.040. In dataset C, the minimum MAE values of MCF and CF were 0.815 and 0.848. Compared with traditional filtering algorithms, the MAE value has decreased by 0.033. The recommendation accuracy of the algorithm is at its highest point, and the performance of the MCF algorithm in terms of recommendation precision has considerably increased compared with the traditional CF algorithm. The number of neighbors is also a factor in the MAE, and the experimental comparison of the MAE data with different sets of neighbors is shown in Figure 5.3.

In Figure 5.3, the MAE value of the MCF algorithm is consistently lower than that of the traditional CF algorithm as the number of neighbourhood sets increases. Thus, the MCF algorithm can effectively enhance

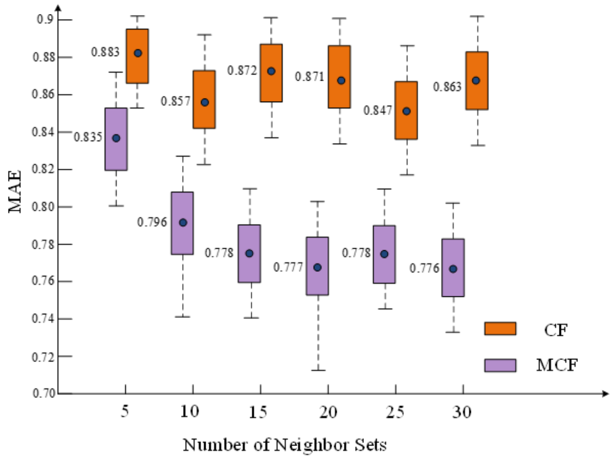


Fig. 5.3: MAE data values for different neighbor sets

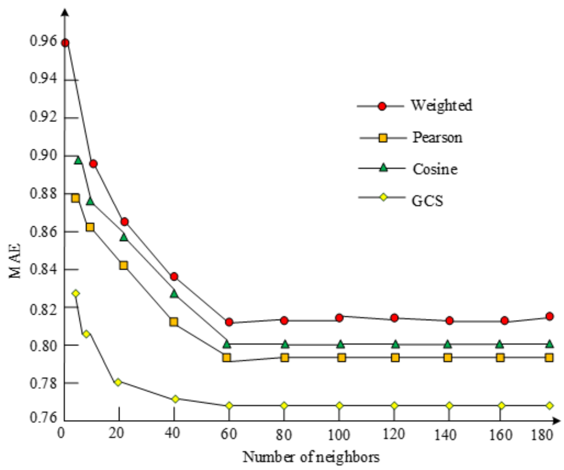


Fig. 5.4: The impact of the number of neighbors on MAE

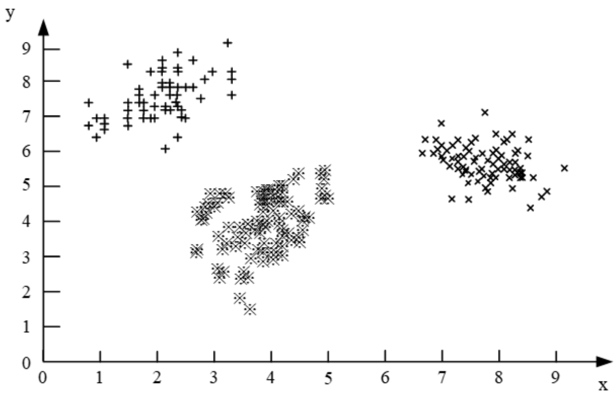
the recommendation quality of the recommendation system.

5.2. Impact of similarity measures on recommendation accuracy and optimization results of improved algorithms for book classification. To explore the impact of similarity measurement on the accuracy of recommendation systems, three commonly used similarity measures were selected: weighted similarity, Pearson, and cosine index calculation. The three indicators were compared with the grey correlation similarity selected in the MCF used in this study. 0.8 in the dataset is used as the training set and 0.2 as the test set to verify the effectiveness of grey correlation. To assess the effectiveness of grey correlation as a similarity calculation, this paper conducted experiments comparing it with three commonly used similarity calculations: weighted, Pearson, and cosine exponential similarity calculations. The data set used was B, and the training set was set at 80%. The results are displayed in Figure 10.

In Figure 5.4, under MCF, the minimum MAE value obtained by using grey correlation similarity as similarity indicates that using this method in recommendation algorithms can improve recommendation quality to a certain extent. It can be found that the recommendation accuracy of the MCF is higher when the choice of

Data set	Number of samples	Number of categories	Dimension
C dataset	153	3	2
Library dataset	2240	10	4

(a) Dataset Parameters



(b) Cluster Results of Library Datasets

Fig. 5.5: Dataset parameters and cluster results of library datasets

Table 5.1: Comparison of classification results of improved algorithms

Algorithm	F1-measure	Iterations
Traditional CF algorithm (Data A)	0.9752	12
MCF group algorithm (Data A)	0.9830	10
Traditional CF algorithm (Data B)	0.9734	12
MCF group algorithm (Data B)	0.9851	10
Traditional CF algorithm (Data C)	0.9721	12
MCF group algorithm (Data C)	0.9863	10

neighbors of the user is relatively large. However, when the number of neighbors is chosen to be relatively large, it will have an impact on the efficiency of the algorithm's execution. The MCF works better when the number of neighbors is in the interval of 50 to 70, when the number of neighbors is 60, the MAE value obtained by GCS calculation is the smallest, which is about 0.767; when the number of neighbors is 60, the MAE value obtained by all four algorithms is the smallest, which is 0.814 by weighted similarity calculation, 0.806 by Pearson's similarity calculation, and 0.795 by cosine similarity calculation method. The clustering effect of the improved system was explored, and the improved population filtering algorithm was used to cluster the library dataset and a dataset C. The clustering effect obtained experimentally on the two datasets is shown in Figure 5.5.

In Figure 5.5, the clustering results of the improved population filtering algorithm are excellent and the correct number of clusters can be obtained, verifying its feasibility. For the judgement of the classification effect, the common judgement criterion is F1-measure. The collection of a library was chosen as the database for the experiment, and the two algorithms were repeatedly run 10 times on the datasets A, B and C respectively to take the average value. Table 5.1 displays a comparison of the categorization outcomes produced by the algorithms.

In Table 5.1, with the improvement of the algorithm, the F1-measure in dataset A grew from 0.9752 to

0.9830, the F1-measure in dataset B grew from 0.9734 to 0.9851, and the F1-measure in dataset C grew from 0.9721 to 0.9863. The number of iterations increased from 12 to 10 on all three datasets reduced to 10 iterations. The experiments show that the improved population MCF algorithm classifies library books better than the traditional algorithm, with increased classification accuracy and better stability.

The research results show that introducing Bayesian probability and grey correlation to improve the collaborative filtering algorithm has better accuracy compared to traditional collaborative filtering algorithms. This is consistent with the research results obtained in Wang N's research on the ideological and political education recommendation system of the improved collaborative optimization algorithm. The personalized recommendation proposed in his study is feasible and has important accuracy and convergence [23].And it also matches with the findings of Sri S L R et al. Predicting user's preference for a specific item, collaborative filtering based on user clustering is used for venue recommendation, where clusters are formed by bio-inspired grey wolf optimization algorithm. The use of clustering eliminates the drawbacks of collaborative filtering in terms of scalability, sparsity and accuracy. In the study Sri S L R et al simulated and validated new mobile based recommendation application framework for developing urban venue recommendation in smart cities. The experimental and evaluation results demonstrate the utility of the newly generated recommendations and demonstrate user satisfaction with the suggested recommendation techniques [24, 25, 26, 27].

6. Conclusion. The classification and personalized recommendation of learning resources such as library books are necessary. The CF group algorithm has a good application environment for book classification and recommendation. However, traditional CF algorithms have issues with cold start and data sparsity, which can constrain the application of the algorithm and affect its recommendation accuracy. This study selected multiple feature methods to improve the traditional CF algorithm. By introducing grey correlation degree into the recommendation system for calculation, a neighbor set with high similarity to users was obtained, and the system's classification and recommendation were completed. In addition, a learning resource recommendation model based on the CF algorithm was constructed, which applies MCF calculation and group recommendation calculation. After several experiments the factors influencing the accuracy of the recommendation sessions were explored. It was verified that the improved MCF group algorithm classified library books better than the traditional algorithms. The MAE value was used as the evaluation criterion to measure the recommendation effectiveness of the system. The average error value of the MCF algorithm was inferior to the traditional CF algorithm, indicating that the classification accuracy of the former is higher than that of the latter. The sparsity of the data set was also an important influencing factor, with the ratio of the training set to the test set gradually increasing, the MAE value reached the lowest and smoothest point at 80%. Because of the improvement of the algorithm, the F1-measure value increased from 0.9752 to 0.9830 in dataset A, from 0.9734 to 0.9851 in dataset B and from 0.9721 to 0.9863 in dataset C. The number of iterations decreased from 12 to 10, suggesting that the enhanced population MCF algorithm performs better in classifying library books than the traditional algorithm. Moreover, the classification accuracy has increased and stability has improved. However, the paper still has some limitations, including the discrepancy in dataset size between the selected dataset in the experiment and the actual dataset, which cannot be simulated accurately. Moreover, although the model has been designed and the main system functions have been implemented, the effectiveness of the model's recommendations has not been practically verified. To further advance research efforts, the model should be applied to specific practical scenarios for verification.

REFERENCES

[1] Pinar-Mendez, A., Fernandez, S. & And, B. and improved identification of drinking water bacteria using the Drinking Water Library. *A Dedicated Water Research: A Journal Of The International Water Association.* **203**, 31-40 (2021)

[2] Xiao, S., Shao, Y., Yawen L. I. Hongzhi Y. Yanyan, S. & Bin, C. Recommendation via learnable edge collaborative filtering. *Chinese Science.* **65**, 112-126 (2022)

[3] Díez, J., Pérez-Núez, P., Luaces, O., Remeseiro, B. & Bahamonde, A. Towards explainable personalized recommendations by learning from users' photos - ScienceDirect. *Information Sciences.* **520**, 416-430 (2020)

[4] Li, L., Zhang, Z. & Zhang, S. Hybrid Algorithm Based on Content and Collaborative Filtering in Recommendation System Optimization and Simulation. *Scientific Programming.* **21**, 101-111 (2021)

[5] Chen, L. restaurant recommendation approach with the contextual information. *Journal Of Intelligent & Fuzzy Systems: Applications In Engineering And Technology.* **40**, 81-89 (2021)

[6] Ryu, T., Jeong, J. & Sang, W. Ethylene trapping of palladium-impregnated zeolites for cold-start emission control. chemical engineering journal. (2022)

[7] Yan, H., Wang, Z., Niu, J. & Xue, T. Application of covering rough granular computing model in collaborative filtering recommendation algorithm optimization. *Advanced Engineering Informatics.* **51**, 1485-10149 (2022)

[8] Yang, Y. & And, K. and Application of Handicraft Recommendation System Based on Improved Hybrid Algorithm. *International Journal Of Pattern Recognition And Artificial Intelligence.* **36**, 50008-22500 (2022)

[9] Wu, C., Liu, S., Zeng, Z., Chen, M., Tang, X. & Peng, X. Knowledge graph-based multi-context-aware recommendation algorithm. *Information Sciences.* **595**, 179-194 (2022)

[10] Li, B., Li, J. & Ou, X. Hybrid recommendation algorithm of cross-border e-commerce items based on artificial intelligence and multiview collaborative fusion. neural computing & applications. (2022)

[11] Li, M., Li, Y., Xu, Y. & Wang, L. Explanatory Q&A recommendation algorithm in community question answering. *Data Technologies And Applications.* **54**, 437-459 (2020)

[12] Li, S. Heuristic Video Recommendation Algorithm based on Similarity Computation for Multiple Features Analysis. recent advances in computer science and communications. (2022)

[13] Liang, X. & Yin, J. Recommendation Algorithm for Equilibrium of Teaching Resources in Physical Education Network Based on Trust Relationship. of Internet Technology. (2022)

[14] Du, S., Chen, Z., Wu, H., Tang, Y. & Li, Y. Image Recommendation Algorithm Combined with Deep Neural Network Designed for Social Networks. *Complexity.* **20** pp. 191-199 (2021)

[15] Chi, H., Xia, H., Zhang, L., Zhang, C. & And, T. and collaborative representation for classification. *Pattern Recognition Letters.* **132**, 46-55 (2020)

[16] Gou, J., Song, J., Du, L., Zeng, S. & Yi, Z. Class mean eighted discriminative collaborative representation for classification. *International Journal Of Intelligent Systems.* **36**, 44-73 (2021)

[17] Yabin, H., Ren, G., Yi, M., Yang, J. & Wang, J. Coastal wetland hyperspectral classification under the collaborative of subspace partition and infinite probabilistic latent graph ranking. *Chinese Science: Technical Science.* **65**, 759-777 (2022)

[18] Wu, Y., Shi, B., Wang, J., Wang, Q., Li, H., Lei, Z., Zhang, N. & Cao, Q. An improved multi-view collaborative fuzzy C-means clustering algorithm and its application in an improved multi-view collaborative fuzzy C-means clustering algorithm and its application in overseas oil and gas exploration. *Journal Of Petroleum Science & Engineering.* **197**, 31-38 (2021)

[19] Masood, F., Masood, J., Zahir, H. & Others Novel approach to evaluate classification algorithms and feature selection filter algorithms using medical data. *Journal Of Computational And Cognitive Engineering.* **2**, 57-67 (2023)

[20] Srinivas, R., Verma, N., Larson, K. & Eric, C. Deep Learning-Based Ligand Design Using Shared Latent Implicit Fingerprints from Collaborative Filtering. journal of Chemical Information and Modeling. (2021)

[21] Wei, L., Liao, Z. & Dafalla A. M. Jiang, F. Effects of Endplate Assembly on Cold Start Performance of Proton Exchange Membrane Fuel Cell Stacks. *Energy Technology.* **9**, 21-32 (2021)

[22] Chui, C., Lin, S., Zhang, B. & Zhou, D. Realization of Spatial Sparseness by Deep ReLU Nets with Massive Data. *IEEE Transactions On Neural Networks And Learning Systems.* **33**, 229-243 (2020)

[23] Ideological, W. & Education, P. Recommendation System Based on AHP and Improved Collaborative Filtering Algorithm. *Scientific Programming.* **2021**, 48352-26483 (2021)

[24] Sivaramakrishnan, S., V. , R., Vijayakumar, V. & Gao, X. Sri S L R. *N Effective User Clustering-based Collaborative Filtering Recommender System With Grey Wolf Optimisation. International Journal Of Bio-Inspired Computation.* **16**, 44-55 (2020)

[25] Diwaker, C., Tomar, P., Solanki, A., Nayyar, A., Jhanjhi, N., Abdullah, A. & Supramaniam, M. A new model for predicting component-based software reliability using soft computing. *IEEE Access.* **7** pp. 147191-147203 (2019)

[26] SANDEEP, V., KAUR, S. & DANDA, B. Intelligent Framework using IoT-based WSNs for Wildfire Detection.

[27] Ghosh, G., Verma, S., Jhanjhi, N., Talib, M. & Others Secure surveillance system using chaotic image encryption technique. *IOP Conference Series: Materials Science And Engineering.* **993**, 012062 (2020)

Edited by: Mudasir Mohd
Special issue on: Scalable Computing in Online and Blended Learning Environments: Challenges and Solutions
Received: Nov 8, 2023
Accepted: Jan 10, 2024