An Improved Collaborative Filtering Recommendation Method Using User Learning Tree Postprint

Authors: Ma Li

Date: 2017-10-11T00:00:00+00:00

Abstract

Objective: Utilizing knowledge point attributes and learning access sequences in the learning tree, predictive ratings are generated for knowledge points, enabling user similarity clustering for collaborative filtering recommendations, thereby improving traditional online learning recommendation methods and enhancing recommendation quality. Method: Knowledge point attributes, knowledge point learning access sequences, learning frequency, and learning time from users are standardized to construct a learning tree. Based on this learning tree, predictive ratings are generated for knowledge points within the tree. User similarity is then calculated using Pearson similarity and cosine similarity based on predictive ratings, knowledge point attributes, and knowledge point learning sequences. Similar users are clustered using the K-means clustering method, and collaborative filtering recommendation methods are subsequently employed for online learning recommendations. Results: Experimental evaluation using the F-measure metric demonstrates that compared with traditional online learning collaborative filtering recommendation methods, the proposed method' s F-measure exceeds singular value decomposition collaborative filtering by 8.22% and average score prediction collaborative filtering by 3.75%. Limitations: Modeling and testing are conducted solely based on 52,456 student learning records and logs from a specific online learning platform, without further validation on other datasets. Conclusion: This approach addresses the limitations of collaborative filtering recommendations that rely on user ratings, while also considering the impact of user interest drift on recommendation accuracy, offering valuable guidance for solving cold start and scalability issues in online learning.

Full Text

Abstract

[**Objective**] This paper aims to improve traditional online learning recommendation methods and enhance recommendation quality by utilizing the attributes and access sequences of knowledge points in learning trees to predict ratings for knowledge points, thereby performing user similarity clustering for collaborative filtering recommendation.

[**Methods**] We standardized the attributes of knowledge points learned by users, knowledge point learning access sequences, learning frequency, and learning time to construct learning trees. Based on the learning trees, we predicted ratings for knowledge points within the trees. Using the predicted ratings and knowledge point attributes and sequences, we calculated user similarity through Pearson similarity and cosine similarity respectively, employed K-means clustering to group similar users, and subsequently implemented online learning collaborative filtering recommendations.

[**Results**] Experimental evaluation using the F-measure metric demonstrated that compared with traditional online learning collaborative filtering recommendation methods, the proposed method' s F-measure exceeded singular value decomposition collaborative filtering by 8.22% and average score prediction collaborative filtering by 3.75%.

[**Limitations**] The modeling and testing were conducted solely based on 52,456 student learning records and logs from a single online learning platform, without further validation on other datasets.

[**Conclusions**] The proposed method addresses the shortcomings of collaborative filtering recommendations that rely on user ratings, while considering the impact of user interest migration on recommendation accuracy, offering valuable guidance for solving the cold start and scalability issues in online learning.

Keywords: Online learning recommendation; Collaborative filtering; Learning tree; Learning access sequence
Classification codes: TP301.6; G35

Introduction

In recent years, online learning has become widespread, and locating the most needed learning materials among massive resources represents the greatest challenge in online learning. The emergence of online learning recommendation systems enables better positioning of learning resources that best match learners' interests. Over the past decade, the TEL community has conducted in-depth research on online learning recommendations[1], with recommendation systems being one of the technologies designed to facilitate online learning by creating personalized learning environments for users.

Currently, mainstream recommendation algorithms primarily include content-based online learning recommendation, collaborative filtering (CF) online learning recommendation, and hybrid recommendation algorithms. Content-based recommendation establishes a learning interest model for users and matches this model with resource attributes to recommend resources with the highest matching degree. Collaborative filtering recommendation clusters users with similar interests based on their preferences and performs cross-recommendations among similar users, representing the most widely applied and mature online recommendation system to date. Both methods have distinct advantages and disadvantages: content-based recommendation only considers users' existing interests in learned resources without discovering potential future learning interests, while collaborative filtering suffers from cold start problems for newly registered learners and sparsity issues when users are unwilling to leave learning records or comments. Hybrid recommendation methods aim to combine the strengths of both approaches, but effectively integrating these two algorithms remains the greatest challenge for hybrid recommendation.

This paper builds upon collaborative filtering recommendation by standardizing resource attributes, access sequences, learning frequency, and learning time to construct a learning tree for each user. User similarity calculation and clustering are then performed based on these learning trees to implement online learning collaborative filtering recommendations. Compared with traditional collaborative filtering algorithms that rely on user ratings, the proposed method demonstrates better recommendation accuracy while considering the impact of user learning interest migration on recommendation quality, and effectively addresses the sparsity and cold start problems inherent in collaborative filtering recommendation systems.

Literature Review

Currently, online learning recommendation methods proposed by scholars domestically and internationally primarily include content-based online learning recommendation, collaborative filtering online learning recommendation, and hybrid recommendation algorithms. The recommendation principles, advantages, and disadvantages of each algorithm are analyzed as follows.

Content-based online learning recommendation has been explored by researchers such as Khribi et al.[2], who matched users' recent navigation history with learning resource content to automatically generate online learning suggestions. Sharif et al.[3] designed a recommendation framework that matches learning resource keywords with users' learning interest tags to implement recommendations, assigning weights to learning resources and further ranking them by importance level. However, the content-based recommendation algorithm only considers matching between learning resources and user interest features without accounting for similarity between users, resulting in recommendations limited to resources users have already learned and shown interest in[4], while failing to recommend unencountered learning resources. To avoid these drawbacks of content-based learning recommendation, researchers have proposed new personalization schemes, such as collaborative filtering recommendation technology[5].

Collaborative filtering recommendation is employed by the vast majority of online learning recommendation systems and is currently the most widely applied recommendation method, which can be divided into three categories. Neighborhood-based collaborative filtering discovers similarity between users or learning resources through user rating data on learning content and predicts recommendations for unrated resources. Model-based collaborative filtering utilizes user rating matrices to build models that predict user ratings. Demographic-based collaborative filtering employs demographic features for user similarity calculation and recommends learning resources among similar users[6]. In online learning environments, learning resources exhibit diverse media formats, including text, hypertext, images, videos, audio, and slideshows, making it difficult to measure resource similarity. Typically, recommendations in online learning environments are implemented based on user preferences for learning resources. Collaborative filtering is a popular recommendation technology, but it has two obvious defects[7]. The primary defect is the sparsity problem: many users are unwilling to rate learning resources, leading to missing foundational data for user similarity calculation and consequently affecting recommendation accuracy. Many researchers have supplemented sparse data by mining implicit valuable information through data mining techniques. The second defect is the cold start problem: newly launched learning resources with few ratings are difficult to recommend even if valuable. Aher et al.[7] alleviated sparsity and cold start problems by classifying learning resources according to relevant computational rules. Based on attributes of learning resources in multi-dimensional space, Salehi et al.[8] proposed a tree model for user interest modeling and employed a novel similarity calculation method in the scholar tree model to generate recommendations. Experimental results demonstrated that their proposed method effectively alleviated cold start and sparsity problems.

To overcome the limitations of content-based and collaborative filtering recommendations, most scholars have attempted to combine these two approaches. Ge et al.[9] proposed a recommendation method integrating content-based and collaborative filtering recommendations. Some scholars have attempted to feed content-based recommendation results into collaborative filtering systems for secondary screening using collaborative filtering techniques, while others have attempted to filter collaborative filtering results through content-based recommendation[10]. Since content-based and collaborative filtering recommendations employ different recommendation philosophies, effectively combining them constitutes the core challenge of hybrid recommendation. Although numerous hybrid methods have been proposed, recommendation accuracy and efficiency warrant further improvement.

To better represent resource features and user learning interests, models are typically established for both resources and users. Wang et al.[11] proposed representing resources through attributes of their classification categories. Kim et al.[12] further proposed building resource models through content feature attributes of resources and ranking them by feature importance. Other scholars have suggested modeling resources based on keyword frequency, where higher keyword frequency indicates greater weight for that keyword in representing the resource. Regarding user modeling, the earliest proposed method involved building user models based on users' personal social attribute information, which might involve privacy concerns. Jalali et al.[13] built user models using features of resources accessed by users, which could better reflect user learning preferences. However, this method can only discover historically expressed preferences and cannot mine potential future learning preferences. Albadvi et al.[14] proposed modeling current users based on the common preferences of user clusters with similar preferences discovered through clustering methods, achieving favorable results.

We model users based on basic information such as access records, access time, duration, and frequency of resources accessed by users. The user model consists of predicted ratings of resources by users, resource access sequences, and user preference transitions. We model resources through basic resource information, with the resource model comprising resource attribute information and attribute weights. Subsequently, we construct user learning trees through user models and resource models, perform user similarity clustering through user learning trees, calculate resource recommendation degrees, and implement recommendations.

The research framework of this paper is illustrated in Figure 1 [Figure 1: see original paper].

Methodology

Resource Modeling

Learning resources can be classified according to resource types, such as mathematics, physics, computer science, etc., with each type further subdivided (e.g., computer science can be divided into software, computer networks, etc.). The types to which learning resources belong can serve as resource attributes. A learning resource may possess multiple attributes; for instance, a resource may belong to computer science, contain mathematical knowledge attributes, and also have author, learning type, and other attributes. The attributes of learning resources should be multi-dimensional. Users may learn because they are interested in certain attributes of resources. By counting the number of resource accesses due to each attribute, we can obtain the weight contribution of that attribute to the learning resource. A larger weight indicates that the attribute is more important relative to the resource and more attractive to users. Based on this, resources can be modeled as follows:

$$M = [(Ak_1, Aw_1), (Ak_2, Aw_2), ..., (Ak_m, Aw_m)]$$

where Ak_m represents the name of the m-th attribute and Aw_m represents the weight contribution of the m-th attribute to learning resource M. In this paper, we set $\sum(Aw_m) = 1$. An example of a learning resource model is: $M = [(\text{Linear Algebra}, 0.35), (\text{Probability Theory}, 0.3), (\text{Master's Thesis}, 0.2), (\text{Some Author}, 0.15)]$.

User Modeling

User models reflect users' preference levels for learning resources, typically represented through user ratings of resources. However, according to Nielsen' s[15] 90-9-1 theory, 90% of users only search, read, and browse online without participating in interactions (such as rating resources); 9% of users may occasionally participate in online interactions, but spend most of their time browsing; and only 1% of users are willing to participate in online interactions while browsing. Consequently, obtaining complete user ratings for modeling purposes to reflect user interests is difficult. Nevertheless, user interest preferences for learning resources can be derived by extracting and processing user access records. This paper primarily processes two indicators—user access time and access frequency to learning resources—to obtain user learning preferences.

We establish a set $Ls_i = \{M_1, M_2, ..., M_n\}$ for the sequence of learning resources accessed by user i, where the set is sorted by the most recent access time, meaning M_1 is the most recently accessed resource and M_n is the least recently accessed. Generally, the more time users spend on a resource, the more important that resource is to them. However, sometimes users spend more time on a resource simply because it contains more information, while less time might be due to lower information content. Considering these factors comprehensively, we process user access time to resources using Equation (1) to obtain users' temporal interest degree.

$$\text{Time}(L_i, M_j) = \frac{\text{TotalTime}(L_i, M_j)/\text{size}(M_j)}{\max_{q \neq s}(\text{TotalTime}(L_i, M_q)/\text{size}(M_q))}$$

where $\text{Time}(L_i, M_j)$ represents temporal interest degree, $\text{TotalTime}(L_i, M_j)$ represents the learning time spent by user i on resource j, and $\text{size}(M_j)$ represents the information volume of resource j, typically the storage capacity of the resource.

The more frequently users access a resource, the more attractive that resource is to them. We define users' frequency interest degree for resources using Equation (2):

$$\text{Frequency}(L_i, M_j) = \frac{\text{Number\_of\_visits}(L_i, M_j)}{\max_{q \neq s}(\text{Number\_of\_visits}(L_i, M_q))}$$

where $\text{Frequency}(L_i, M_j)$ represents frequency interest degree, $\text{Number\_of\_visits}(L_i, M_j)$ represents the number of times user i accessed resource j, and $\max_{q \neq s}(\text{Number\_of\_visits}(L_i, M_q))$

represents the access count of resource M_q that user i visited most frequently among resources other than j.

Comprehensively considering both temporal and frequency interest degrees for a learning resource, we standardize them using Equation (3) to obtain the predicted rating of the resource by the user.

$$\mathrm{MR}(L_i, M_j) = 5 \times \mathrm{Nor}(\mathrm{Frequency}(L_i, M_j) + \mathrm{Time}(L_i, M_j))$$

$\mathrm{MR}(L_i, M_j)$ is the predicted rating of resource j by user i based on temporal and frequency interest degrees. Nor() is a normalization function that processes temporal and frequency interest degrees into values between 0 and 1. The predicted rating here ranges from 1 to 5 points and updates dynamically as users' resource access patterns change.

It should be noted that the most recent access time to learning resources can reflect the dynamic shift of user learning interests. In E-Learning environments, user interest in learning resources changes dynamically, with recently accessed resources better reflecting future learning preferences. Previous user learning models treated all learning resources equally, ignoring the impact of resource access temporal order on user preferences. The forgetting function curve proposed by German psychologist Ebbinghaus[16] reflects human forgetting patterns for new information. Based on this forgetting function, we design an exponential function to reflect the dynamic shift of user preferences for learning resources, as shown in Equation (4):

$$h(x(M_j)) = \frac{\lambda^{(x(M_j)-1)} - 1}{\lambda - 1}, \quad 0 \leq h(x) \leq 1$$

$x(M_j)$ represents the order of resource j in user i' s resource access sequence set Ls_i. As evident, the later the order of $x(M_j)$ in Ls_i (the larger the value), the weaker user i' s preference for resource j, and the smaller $h(x)$ becomes. λ is a tuning parameter reflecting the rate of change in user preference for resources; a larger λ results in more pronounced changes in $h(x)$, indicating more obvious forgetting. When λ is set to 0.95, the variation of $h(x)$ is shown in Figure 2 [Figure 2: see original paper].

Based on users' resource access sequences, predicted ratings of resources, and $h(x)$ reflecting preference shifts, we establish a learning tree model for users as illustrated in Figure 3 [Figure 3: see original paper].

The learning tree is an $m + 1$ level tree structure, where m represents the number of resource attributes accessed by the user. The bottom level consists of leaf nodes representing learning resources accessed by the user, expressed as a quadruple $\{MID, OR, NH, MR\}$, where MID represents resource ID, OR represents the access order of the resource by the user, NH represents the normalized $h(x)$ value of the accessed resource, and MR represents the predicted rating based on temporal and frequency interest degrees. Non-leaf nodes in the tree can be defined as a triple $\{KA, NH, MR\}$, where KA represents an attribute keyword of the resource at that level. The NH value of a node at level i can be represented as the sum of NH values of its successor nodes at level $i+1$, and the MR value of a node at level i can be represented as the average MR of all leaf nodes in its subtree. Based on the learning tree shown in Figure 3, the user' s NH value for information technology is 0.24 and for mathematics is 0.76, reflecting a shift in the author' s interests.

Whenever a user accesses a learning resource, the learning tree is updated based on access duration, frequency, and order. If the learning resource node does not exist in the learning tree, the node is added and the tree is updated. The dynamic update process of the learning tree is as follows:

(1) When user accesses an existing resource node M_1 (leaf node) in the learning tree, the user' s learning time and frequency are recorded, and the user' s resource access sequence is modified. Equation (1) is used to calculate the user' s temporal interest degree in the resource, Equation (2) calculates the frequency interest degree, and Equation (3) yields the predicted rating to update the MR in leaf node $\{MID, OR, NH, MR\}$. The changing resource access sequence is used to recalculate dynamic interest degree using Equation (4) to update NH, and the changed sequence updates OR. Since NH and MR values in upper-level non-leaf nodes $\{KA, NH, MR\}$ are averages of their lower-level nodes, after leaf node updates are completed, upper-level non-leaf nodes in the learning tree are updated accordingly.

(2) If the user accesses a resource node (leaf node) that does not exist in the learning tree, the leaf node $\{MID, OR, NH, MR\}$ values are calculated as above, and a new leaf node is added under the corresponding non-leaf node in the learning tree based on resource attributes, with dynamic updates to its upper-level non-leaf nodes.

Similar Learning User Clustering

Currently, most collaborative filtering recommendation systems calculate user similarity and implement recommendations based on user rating matrices for learning resources. This approach has two drawbacks: (1) Some users are unwilling to leave ratings for resources, or some newly launched resources have not yet been evaluated by users, both leading to sparsity problems. (2) This method overly relies on user ratings of learning resources while ignoring resource attributes and user learning context, also resulting in reduced recommendation accuracy.

The user learning tree model contains user learning resource attributes, predicted ratings of learning resources by users, learning order of resources, and shifts in user learning preferences. This paper proposes a method for clustering similar users based on user learning trees, which can effectively avoid the drawbacks of traditional collaborative filtering while maintaining clustering effectiveness.

The user similarity calculation based on learning trees proposed in this paper follows three principles: (1) The more similar the attributes of learning resources in learning trees, the more similar the users' learning interests. (2) The more similar the learning order of resources in learning trees, the more similar the users' learning interests. (3) The more similar the predicted ratings of learning resources in learning trees, the more similar the users' learning interests.

User similarity calculation in this paper is divided into two parts: similarity calculation based on learning tree resource attributes and similarity calculation based on learning tree user ratings.

(1) Similarity calculation based on learning tree resource attributes, $\mathrm{sim}_A(L_a, L_b)$, is as follows:

$$\mathrm{sim}_A(L_a, L_b) = \frac{\sum_i [\mathrm{AV}(L_a, L_b) \cap \mathrm{MW}_i \times \mathrm{NH}_{ai}]}{\sum_i [\mathrm{MW}_i \times \mathrm{NH}_{ai}]}$$

where $\mathrm{AV}(L_a, L_b)$ represents the intersection set of identical attributes in learning trees of user a and user b, MW_i represents the weight of attributes at level i in the learning tree. In this paper, we define $i\mathrm{MW} = 2^{-i}$. NH_{ai} represents the NH value of user a' s learning tree node at level i. The deeper the node level where $i\mathrm{MW}$ resides, the larger the value.

(2) Similarity calculation based on learning tree resource predicted ratings, $\mathrm{sim}_R(L_a, L_b)$, is as follows:

$$\mathrm{sim}_R(L_a, L_b) = \frac{|\sum_i (\mathrm{MR}_{ai} - \overline{\mathrm{MR}}_a)(\mathrm{MR}_{bi} - \overline{\mathrm{MR}}_b)|}{\sqrt{[\sum_i (\mathrm{MR}_{ai} - \overline{\mathrm{MR}}_a)^2 \times \sum_i (\mathrm{MR}_{bi} - \overline{\mathrm{MR}}_b)^2]}}$$

where L represents the leaf node set, MR_{ai} and MR_{bi} represent the predicted ratings of the i-th leaf node by user a and user b respectively, and $\overline{\mathrm{MR}}_a$ and $\overline{\mathrm{MR}}_b$ represent the average predicted ratings of user a and user b.

The above two similarity calculation formulas consider only learning resource attributes and predicted ratings, effectively eliminating cold start and sparsity problems. The final similarity between user a and user b is:

$$\mathrm{LearnerSim}(L_a, L_b) = \alpha \times \mathrm{sim}_R(L_a, L_b) + (1 - \alpha) \times \mathrm{sim}_A(L_a, L_b)$$

where α is the weight of $\mathrm{sim}_R(L_a, L_b)$ and $\mathrm{sim}_A(L_a, L_b)$. Through testing with different α values on the dataset, we found that $\alpha = 0.7$ yields the best recommendation performance.

Collaborative Filtering Recommendation

The recommendation process is as follows: (1) As users' learning progresses, learning trees are generated and dynamically updated for users; see Section 4.2 for the learning tree generation and update process. (2) User similarity is calculated based on resource attributes and predicted ratings in user learning trees; see Section 4.3 for the user similarity calculation process. (3) To determine whether resource M_j that user L_i has not encountered is worth recommending to that user, we propose a recommendation degree indicator $\mathrm{RD}(L_i, M_j)$, calculated as follows:

$$\mathrm{RD}(L_i, M_j) = \sum_{q \in \mathrm{LM}j} [\mathrm{LearnerSim}(L_i, L_q) \times (\mathrm{MR}_{qj} - \overline{\mathrm{MR}}_q)]$$

LM_j is the set of all users who accessed resource M_j, q is a learning user in this set, $\mathrm{LearnerSim}(L_i, L_q)$ is the similarity between user i and user q obtained based on Equation (7), MR_{qj} is the predicted rating of resource M_j by user q obtained based on Equation (3), and $\overline{\mathrm{MR}}_q$ is the average predicted rating of all resources by user q.

Based on the recommendation degree indicator, we can recommend the Top-n unlearned resources with the highest recommendation degrees to learning users.

Experimental Data

The experimental data in this paper consists of access data from a foreign online learning resource platform, containing complete user access records and basic resource information. We extracted access data from September 2009 to February 2011. This dataset includes 52,456 learning records from 2,354 users and 3,254 learning resources, with complete basic information. Learning resources contain basic attributes including resource ID, resource address, upload time, resource size, appropriate learning level, resource classification, difficulty level, etc., among which resource ID and classification can be used for resource modeling. User access logs contain information such as user ID, access path, and timestamp, where resource ID, resource size, and timestamp can be used to calculate temporal interest degree (Equation (1)) and user interest shift (Equation (4)), while user ID and resource ID can be used to calculate frequency interest degree (Equation (2)). Rating logs contain basic information including user ID, resource ID, and rating.

Comprehensive information including resource modeling, temporal interest degree, frequency interest degree, and user ratings can be utilized for user modeling (user learning tree). Based on user learning trees, user similarity calculation can be performed (Equations (5) and (6)), and collaborative filtering recommendations can be implemented.

Experimental Results

Recommendation Precision, Recall and F-measure

Recommendation quality is typically measured using two metrics: recommendation precision and recall. Precision is the ratio of recommended relevant items to total recommended items[17]. Recall is the ratio of recommended relevant items to total relevant items (that should be retrieved). The calculation formulas for precision and recall are as follows[18]:

$$\text{Precision} = \frac{|\{\text{relevant\_items}\} \cap \{\text{recommended\_items}\}|}{|\{\text{recommended\_items}\}|}$$

$$\text{Recall} = \frac{|\{\text{relevant\_items}\} \cap \{\text{recommended\_items}\}|}{|\{\text{relevant\_items}\}|}$$

Since precision and recall are contradictory metrics, this paper adopts the F-measure metric for experimental evaluation, which combines both precision and recall.

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Impact of Different α Values on F-measure

Experiments show that when α is set to 0.7, a higher F-measure value is obtained, while smaller values yield poorer recommendation performance. User predicted ratings fully reflect the similarity of users' access duration and frequency to learning resources, thus better indicating the proximity of user learning preferences compared to user attribute similarity (similarity in accessed resources). As shown in Figure 4 [Figure 4: see original paper], when α is small, the weight of user attribute similarity is larger and the F-measure value is lower. As α increases, the weight of user predicted ratings grows and the F-measure value rises accordingly. However, when α exceeds 0.7, the F-measure value decreases again. This is because user predicted rating similarity only considers users' interest in leaf nodes (learning resources) without considering users' interest in resource categories from the perspective of the entire learning tree. Therefore, an excessively large α value negatively impacts recommendation quality.

The experimental results objectively demonstrate that user similarity based on predicted ratings has a greater influence on user clustering effectiveness than user similarity based on resource attributes.

Comparison with Other Recommendation Methods

Currently, two effective collaborative filtering recommendation methods for eliminating sparsity and cold start problems are the classic singular value decomposition collaborative filtering proposed by Sarwar et al.[19] and the average score prediction collaborative filtering[17].

The singular value decomposition collaborative filtering recommendation algorithm extracts the most essential features from the original user rating matrix to provide a simplified approximation matrix. This method eliminates weakly correlated data, thereby reducing the dimensionality of data to be processed. Since the recommendation system only processes the simplified matrix and considers only low-dimensional data after dimensionality reduction, computational complexity is reduced to some extent, making it one of the more classic collaborative filtering recommendation algorithms.

The average score prediction collaborative filtering recommendation method, proposed by Devi et al.[17], pre-clusters users with similar ratings and predicts unrated data for users based on similarity within clusters. Its essence lies in predicting ratings for unrated products based on similarity of rating values from similar users on rated products. This prediction method has also achieved good results, offering some improvement in recommendation accuracy over singular value decomposition collaborative filtering, and is currently a mainstream collaborative filtering recommendation method with high recommendation accuracy.

We compare our proposed method with singular value decomposition collaborative filtering and average score prediction collaborative filtering, with F-measure values shown in Figure 5 [Figure 5: see original paper].

The experimental results are analyzed as follows: (1) The F-measure of our method exceeds singular value decomposition collaborative filtering by 8.22%. (2) The F-measure of our method exceeds average score prediction collaborative filtering by 3.75%.

The experimental results demonstrate that the proposed recommendation method achieves better recommendation quality compared to the other two classic recommendation methods, particularly showing much better performance than the singular value decomposition collaborative filtering algorithm. The singular value decomposition recommendation algorithm eliminates weakly correlated data and extracts only the most essential features from the original user rating matrix, but its decomposition effect significantly impacts recommendation quality, resulting in large variations and unstable recommendation performance. The average score prediction collaborative filtering performs slightly better, but it can only cluster users and implement recommendations based on learning resources that users have already rated. If there are many unrated data (sparsity problem), its recommendation performance also suffers. By removing some rated data to increase sparsity and conducting experiments, we found that our method clearly demonstrates advantages. Additionally, it can recommend "hot-click" resources to newly registered users, alleviating the cold start problem to some extent. The experimental results are shown in Figure 6 [Figure 6: see original paper].

Conclusion

The user learning tree proposed in this paper fully considers user learning resource attributes, learning order of resources, predicted ratings of learning resources by users, and shifts in user learning interests. Based on these factors, user similarity clustering is performed, and learning resource recommendation degrees are calculated through learned user similarities. This method can effectively avoid cold start and sparsity problems in collaborative filtering recommendation algorithms. Experimental evaluation results demonstrate that the proposed recommendation method achieves high recommendation quality in online learning.

References

[1] Verbert K, Manouselis N, Ochoa X, et al. Context-Aware Recommender Systems for Learning: A Survey and Future Challenges[J]. IEEE Transactions on Learning Technologies, 2012, 5(4): 318-335.

[2] Khribi M K, Jemni M, Nasraoui O. Automatic Recommendations for E-learning Personalization Based on Web Usage Mining Techniques and Information Retrieval [J]. Educational Technology and Society, 2009, 12(4): 30-42.

[3] Sharif N, Afzal M T, Helic D. A Framework for Resource Recommendations for Learners Using Social Bookmarking [C]. In: Proceedings of the 8th International Conference on Computing and Networking Technology. IEEE, 2012: 71-76.

[4] Salehi M, Kamalabadi I N, Ghoushchi M B G. Personalized Recommendation of Learning Material Using Sequential Pattern Mining and Attribute Based Collaborative Filtering [J]. Education and Information Technologies, 2014, 19(4): 725-742.

[5] Salehi M, Kamalabadi I N. Hybrid Recommendation Approach for Learning Material Based on Sequential Pattern of the Accessed Material and the Learner' s Preference Tree [J]. Knowledge-Based Systems, 2013, 48: 57-69.

[6] Chen W, Niu Z, Zhao X. A Hybrid Recommendation Algorithm Adapted in E-learning Environments [J]. 2014, 17(2): 271-284.

[7] Aher S B, Lobo L. Applicability of Data Mining Algorithms for Recommendation System in E-learning [C]. In: Proceedings of the International Conference on Advances in Computing, Communications and Informatics. 2012: 1034-1040.

[8] Salehi M, Kamalabadi I N, Attribute-based Recommender System for Learning Resource by Learner Preference Tree[C]. In: Proceedings of the 2nd International e-Conference on Computer and Knowledge Engineering. IEEE, 2012: 133-138.

[9] Ge L, Kong W, Luo J. Courseware Recommendation in E-learning System [C]. In: Proceedings of the 5th International Conference on Advances in Web

Based Learning. 2006: 10-24.

[10] Wan L, Zhao C. A Hybrid Learning Object Recommendation Algorithm in E-learning Context [J]. International Journal of Digital Content Technology and Its Applications, 2012, 6(18): 371-379.

[11] Wang S, Xie Y, Fang M. A Collaborative Filtering Recommendation Algorithm Based on Item and Cloud Model [J]. Wuhan University Journal of Natural Sciences, 2011, 16(1): 16-20.

[12] Kim K, Ahn H. A Recommender System Using GA K-means Clustering in an Online Shopping Market [J]. Expert Systems with Applications, 2008, 34(2): 1200-1209.

[13] Jalali M, Mustapha N, Sulaiman M N B, et al. OPWUMP: An Architecture for Online Predicting in WUM-Based Personalization System [A]. // Advances in Computer Science and Engineering [M]. Springer Berlin Heidelberg, 2009.

[14] Albadvi A, Shahbazi M. Integrating Rating-based Collaborative Filtering with Customer Lifetime Value: New Product Recommendation Technique [J]. Intelligent Data Analysis, 2010, 14(1): 143-155.

[15] Nielsen J. The 90-9-1 Rule for Participation Inequality in Social Media and Online Communities [EB/OL]. [2015-09-01]. http://www.useit.com/alertbox/participation_{inequality}.html.

[16] Ebbinghaus H. Memory: A Contribution to Experimental Psychology [M]. New York: Dover, 1885.

[17] Devi M K K, Venkatesh P. Kernel Based Collaborative Recommender System for E-Purchasing [J]. Academy of Sciences, 2010, 35(5): 513-524.

[18] Klašnja-Milićević A, Vesin B, Ivanovic M, et al. E-learning Personalization Based on Hybrid Recommendation Strategy and Learning Style Identification [J]. Computers & Education, 2011, 56(3): 885-899.

[19] Sarwar B M, Karypis G, Konstan J A, et al. Application of Dimensionality Reduction in Recommender Systems [EB/OL]. [2015-10-25]. http://robotics.stanford.edu/users/ronnyk/WEBKDD2000/papers/sarwar.pdf.

Conflict of Interest Statement

The authors declare no conflict of interest.

Supporting Data

The supporting data is self-archived by the author, E-mail: mali8321@tjfsu.edu.cn.

Dates

Received: November 9, 2015
Revised: January 20, 2016