Research on Personalized Recommendation Algorithm Based on Improved Tensor Decomposition Model (Postprint)

Authors: Chen Meimei, Xue Kangjie

Date: 2017-11-08T00:00:00+00:00

Abstract

Purpose: To address the degradation of recommendation accuracy in tensor decomposition-based personalized recommendation systems arising from UGC tag redundancy and the influence of popular tags and resources on users' personalized interests.

Method: This paper proposes an improved personalized recommendation algorithm based on tensor decomposition, which integrates a tag comprehensive co-occurrence approach with spectral clustering. Inspired by the IDF concept in TF-IDF, a popularity penalty mechanism based on co-occurring tags and resources is introduced to redefine the initial tensor constructed upon the triple relationship of <user, tag cluster, resource>.

Results: Simulation experiments conducted on the Last.fm dataset demonstrate that the proposed algorithm achieves favorable performance across precision, recall, and F1-score metrics. The incorporation of comprehensive co-occurrence spectral clustering yields an average F1-score improvement of 5.91%, while the IDF-based enhancement of the initial tensor contributes an average F1-score improvement of 1.29%.

Limitations: The algorithm has not been validated on datasets from other domains, such as Weibo or Delicious.

Conclusion: The personalized recommendation algorithm based on the improved tensor decomposition model significantly enhances accuracy and facilitates the provision of more satisfactory resources to users in social network environments.

Full Text

Preamble

Personalized Recommendation Algorithm Based on Modified Tensor Decomposition Model

Chen Meimei, Xue Kangjie
(Glorious Sun School of Business & Management, Donghua University, Shanghai 200051, China)

Abstract

[**Objective**] This study addresses the degradation of recommendation accuracy in tensor decomposition-based personalized recommendation caused by redundant UGC tags, popular tags, and resource bias affecting users' personalized interests. [**Methods**] We propose an improved personalized recommendation algorithm based on tensor decomposition that introduces a spectral clustering method incorporating comprehensive tag co-occurrence. Drawing on the IDF concept from TF-IDF, we develop a penalty mechanism for popular tags and resources based on co-occurrence patterns, which redefines the initial tensor structure around the triple relationship of <user, tag cluster, resource>. [**Results**] Simulation experiments on the Last.fm dataset demonstrate strong performance across precision, recall, and F1 metrics. The introduction of combined co-occurrence spectral clustering improves the F1 measure by an average of 5.91%, while the IDF-based modification to the initial tensor yields an average F1 improvement of 1.29%. [**Limitations**] The algorithm has not been validated on datasets from other domains such as Weibo or Delicious. [**Conclusions**] The improved tensor decomposition model significantly enhances recommendation accuracy and facilitates more satisfactory resource provision in social network environments.

Keywords: Personalized Recommendation; UGC; Tag; Tag Co-occurrence; Spectral Clustering; Tensor Decomposition

1. Introduction

Social networks such as Facebook and Weibo have become primary channels for maintaining social connections and obtaining information, making resource recommendation in social networks a prominent research area. Rich feature information enables recommendation systems to provide more personalized suggestions. Tags, as a form of User Generated Content (UGC), represent an effective approach to information classification, organization, and management emerging from collective intelligence in internet-based social environments. Users spontaneously annotate online resources with tags to describe them, making UGC tags a crucial link between users and resources and an important data source for reflecting user interests and resource characteristics.

To effectively integrate tag data, personalized recommendation algorithms must fully preserve the characteristics of the <user, tag, resource> triple relationship. In recent years, tensor decomposition models have provided important theoretical support for tag-based recommendation systems due to their superior adaptability to high-dimensional data. Tensor decomposition is a higher-order generalization of matrix decomposition that maps triple relationships into three-dimensional matrix space. By extracting principal tensor eigenvalues, it obtains a compressed approximation of the original tensor, eliminating noisy data while effectively revealing implicit relationships between variables. This makes it particularly suitable for addressing recommendation accuracy issues caused by substantial noise in UGC tags and has become mainstream in tag-based recommendation algorithms. Common tensor decomposition methods include CP decomposition and Tucker decomposition, which have matured since their development in 1927. Current research primarily focuses on domain-specific improvements.

In tag-based recommendation systems, Symeonidis et al. designed a general framework for tensor decomposition in recommendation and found that Higher-Order Singular Value Decomposition (HOSVD) significantly outperformed the FolkRank algorithm in accuracy. Liao Zhifang et al. proposed an incremental model for new user tag recommendation based on Tucker and CP decomposition, substantially reducing computational time. While these studies successfully leveraged tensor decomposition to highlight relationships between tags and users/resources, tag semantic ambiguity and redundancy issues hinder further accuracy improvements.

Some research addresses this problem at the tensor model construction stage. For instance, Rendle et al. performed positive and negative filling of missing values within tensors and optimized AUC values to obtain optimal decomposition results. Wu Huijuan et al. compared the superiority between tags under the same user-resource pair, extending the triple relationship. Utilizing clustering algorithms to clean tag data represents an effective approach to solving tag redundancy and semantic ambiguity at its core.

Additionally, popular resources and tags affect recommendation results, particularly in social network-based recommendation algorithms. Since popular resources often receive larger weights, recommendation results tend to be biased toward these resources while neglecting numerous long-tail resources, thereby reducing recommendation accuracy. In two-dimensional spaces, TF or IDF concepts from TF-IDF are typically used to set penalty terms that reduce the influence of popular tags or resources. Term Frequency-Inverse Document Frequency (TF-IDF) measures a term' s ability to express document characteristics based on the principle that if a word or phrase appears frequently in one document but rarely in others (high TF and high IDF), it possesses strong category discrimination capability. For popular resource penalization, Fleder et al. applied TF concepts to calculate product similarity in product recommendation research, finding that penalizing popular resources helped increase sales. Wang Cheng et al. penalized popular resources using IDF concepts when calculating user similarity, improving the accuracy and recall of user-based collaborative filtering algorithms. For popular tag penalization, Cantador et al. found that popular tags provide no additional information for distinguishing user preferences and resource characteristics, instead reducing recommendation accuracy, though they ignored popular resource effects. Xiang Liang treated tags as features connecting users and resources, penalizing both popular tags and resources simultaneously, achieving good results in tag-based recommendation accuracy.

However, the situation differs in three-dimensional space when tags are applied to tensor decomposition-based recommendation algorithms. Rafailidis et al. split the triple relationship into two binary relationships when defining the initial tensor after tag clustering, setting separate penalty terms for each. While this helps highlight the modified triple relationship, the negative impact caused by the high sparsity of the tensor itself may be more severe than the positive impact of highlighting the triple relationship, potentially decreasing recommendation accuracy.

Therefore, this paper proposes a tensor decomposition model-based recommendation algorithm that integrates comprehensive tag co-occurrence spectral clustering and an improved popularity penalty mechanism. First, we introduce a spectral clustering method based on comprehensive tag co-occurrence in tag data preprocessing to address tag semantic ambiguity and redundancy while preserving triple relationships. Second, to tackle the problem of popular tags and resources affecting recommendation accuracy, we introduce an improved penalty term in the initial tensor redefinition while fully preserving the <user, tag cluster, resource> triple relationship, thereby further enhancing the accuracy of personalized recommendation algorithms based on tensor decomposition.

2. Tag Data Preprocessing Based on Combined Co-occurrence Spectral Clustering

UGC tags derived from folksonomy suffer from semantic ambiguity, synonymy, and polysemy issues that significantly degrade recommendation algorithm accuracy. Before tensor decomposition, it is necessary to cluster tag data to reduce the impact of tag redundancy and semantic ambiguity, eliminate noisy data, and highlight semantic relationships to improve recommendation quality. Tag clustering divides tag data into multiple clusters where tags within each cluster are as similar as possible while being dissimilar to tags in other clusters. This allows infrequently used tags to be replaced by a tag group and semantically similar tags to be grouped into the same cluster, thereby highlighting user preferences and resource themes to improve recommendation accuracy.

Clustering algorithms are typically used to address sparsity problems by selecting a smaller number of clusters. However, for tag data characteristics, the ability of clustering algorithms to correctly identify semantically ambiguous and redundant tags significantly affects the rationality of clustering results. Leginus et al. compared several clustering methods under tensor models and found spectral clustering outperformed others. Due to its efficiency and ability to discover irregular clusters, spectral clustering has been increasingly applied to personalized recommendation based on tag clustering. For spectral clustering algorithms, the tag similarity matrix is a crucial input. How to define tag similarity to comprehensively capture relationships between tags, users, and resources significantly impacts final clustering effectiveness. However, Leginus et al. did not consider differences in user annotation preferences or annotation variations across different resources in their initial tensor definition, thus affecting model accuracy. Symeonidis improved upon this by considering relationships between users and tag clusters and between resources and tag clusters, using vector space models to calculate cosine similarity between tags to form similarity matrices for spectral clustering.

Typically, tag similarity calculation employs two methods: vector space and tag co-occurrence. The vector space model defines each tag as a vector whose elements generally represent associations between users or resources and the tag. However, such two-dimensional vector forms struggle to represent three-dimensional spatial relationships. Consequently, Symeonidis' s vector space-based tag similarity calculation cannot integrate user, tag cluster, and resource relationships holistically, causing separation of the triple relationship and theoretically weakening tags' important role in connecting semantic relationships between users and resources. Furthermore, homogenizing all users and resources in the tag vector space model leads to exponential growth in vector dimensions as users and resources increase, causing severe sparsity issues that affect clustering effectiveness.

In contrast, tag co-occurrence methods based on graph theory facilitate direct representation of multi-relational structures. Li et al. proposed an improved tag co-occurrence combined with spectral clustering approach, dividing tag similarity into individual co-occurrence similarity and group co-occurrence similarity. Individual co-occurrence similarity captures the most fundamental connection between two tags, while group similarity enhances semantic relationships between tags and serves as a supplement to individual similarity. This core concept aligns with Li Ruimin et al.' s view that if a resource and a user share more common tags, the association between that user and resource is stronger. By combining individual and group co-occurrence similarities into comprehensive co-occurrence similarity, tag similarity relationships can be better expressed. This approach neither splits the triple relationship into binary forms nor homogenizes users and resources, enabling differentiation between users and resources while completely preserving semantic relationships among users, tags, and resources, thereby helping clustering algorithms better identify semantically ambiguous and redundant tags.

Since spectral clustering is used before tensor decomposition to address inherent semantic ambiguity and redundancy in tags and improve decomposition quality, the choice of clustering method is critical. Partition-based clustering methods such as K-means cluster based on distance to cluster centers, and modifying points far from cluster centers would cause significant errors in complex tag networks. In contrast, this paper employs graph theory-based spectral clustering for tag clustering, which cuts graphs to form tag clusters with the objective of minimizing graph weights. Without cluster centers, this approach facilitates aggregating scattered tags, ensures high similarity among tags within the same cluster regardless of distance to a center, and is more advantageous than K-means for discovering irregular clusters, thereby minimizing semantic loss caused by clustering.

3. Initial Tensor Improvement

Since tag clustering is performed before tensor decomposition, transforming the <user, tag, resource> triple relationship into <user, tag cluster, resource> form and changing the dimension definitions, the initial tensor must be adaptively modified to reflect the correlations among these three elements while introducing a popularity penalty mechanism to further weaken the influence of popular tags and resources on recommendation results.

In folksonomy, we define a quadruple $F = (U, T, R, \Omega)$. Here $U = \{u_1, u_2, \dots, u_l\}$ represents the set of l user IDs, $T = \{t_1, t_2, \dots, t_m\}$ represents the set of m tag IDs, $R = \{r_1, r_2, \dots, r_n\}$ represents the set of n resource IDs, and $\Omega \subseteq U \times T \times R$ represents the set of possibilities that user u_i annotated resource r_k with tag t_j. If an annotation record exists, $\omega_{i,j,k} = 1$; otherwise it is 0. This quadruple can be converted into tensor form: define tensor $\mathcal{X} \in \mathbb{R}^{|I_u| \times |I_C| \times |I_r|}$, where $|I_u|$, $|I_C|$, and $|I_r|$ represent the numbers of users, tag clusters, and resources in the dataset, respectively. The elements in the tensor are $x_{u_i,C_j,r_k} = \omega_{i,j,k}$. Through tensor decomposition algorithms, the initial tensor is dimensionally reduced, removing noisy eigenvalues to obtain an approximate tensor $\hat{\mathcal{X}}$. The elements $\hat{x}_{u_i,C_j,r_k}$ represent values after iterative convergence.

Typically, tensor elements represent the degree of association among users, tags, and resources. When tags are clustered into tag clusters, elements in the initial tensor based on <user, tag cluster, resource> relationships can be transformed as:

$$x_{u_i,C_j,r_k} = \sum_{t_j \in C_j} \omega_{i,j,k} \quad (1)$$

Formula (1) sums the number of times user u_i annotated resource r_k with tags in cluster C_j as the weight of user u_i' s association with resource r_k under cluster C_j.

According to formula (1), if many users annotate resource r_1 using tags from cluster C_1, making x_{u_i,C_1,r_1} large, the system will inevitably bias toward resource r_1 under cluster C_1 when making recommendations. Even if users have selected other tag clusters, the algorithm cannot objectively reflect users' personalized interests and struggles to discover other resource features.

To address this issue, Xiang Liang proposed combining IDF with logarithmic functions to penalize resources and tags separately in tag-based recommendation algorithms. For popular resources, the penalty term is:

$$\log\left(1+\frac{N}{kn(r_k)}\right)$$

where $kn(r_k)$ represents the number of times resource r_k appears across different users. A high $kn(r_k)$ indicates that many users have annotation records for this resource, demonstrating its universality and enabling identification of popular resources. Adding 1 before taking the logarithm avoids division by zero, and since logarithmic functions grow slower than linear functions, this prevents the denominator from becoming too large and the entire formula from approaching zero, thereby avoiding information loss and effectively solving the popularity penalty problem. This paper introduces this method into three-dimensional space for initial tensor definition.

However, Gemmell et al. observed that the most popular tags are often semantically ambiguous, causing these tags to be popular only in certain meanings. Penalizing tags separately might incorrectly punish popular tags in non-popular meanings. Therefore, this paper abandons the practice of penalizing resources and tags separately, instead considering penalizing co-occurring popular resources and popular tags. This creates a correspondence between tags and resources to determine the actual meaning of tags within those resources, avoiding incorrect penalties caused by tag ambiguity.

Thus, the initial tensor definition incorporating the popularity penalty mechanism based on co-occurring tags and resources is:

$$x_{u_i,C_j,r_k}=\sum_{t_j\in C_j}\omega_{i,j,k}\cdot\log\left(1+\frac{N}{|U_{j,k}|}\right)\qquad(2)$$

where $|U_{j,k}|$ represents the number of different users who used both tag t_j and resource r_k. This indicates that if a tag is consistently used by different users to annotate a particular resource, the <tag, resource> pair is considered popular and receives appropriate penalization. $|U_{j,k}|$ essentially represents the intersection frequency of tag and resource usage, which can identify and penalize truly popular tags and resources while avoiding over-penalization from separate tag and resource penalization.

4. Personalized Recommendation Algorithm Based on Improved Tensor Decomposition Model

In the recommendation phase, to maximally utilize tag clusters and discover latent relationships between users and resources, tensor decomposition is required.

This paper employs the HOSVD-HOOI algorithm to decompose the initial tensor, choosing to retain 70% of the original information. First, the Higher-Order Singular Value Decomposition (HOSVD) algorithm removes useless eigenvalues to reduce noise in the tensor and obtain a good initial solution. Then, the Higher-Order Orthogonal Iteration (HOOI) algorithm iteratively optimizes this initial solution to obtain the optimal approximate tensor, which contains more accurate semantic relationships among the triplets and helps the system discover users' latent interests for better recommendations. Research has shown that this combined algorithm yields more precise approximate tensors compared to other tensor decomposition methods.

The entire recommendation process proceeds from a practical usage scenario: when user u_i selects a tag t_j, the system queries the tag cluster C_j to which t_j belongs, then identifies the top N resources with the highest (u_i, C_j, r_k) values in that user' s tensor, and recommends these resources to the user.

5. Experiments

5.1 Dataset Selection

This paper utilizes the Last.fm dataset, which has been widely applied in related research since its release at the 5th International Conference on Recommender Systems in 2011. The dataset includes annotation and listening records from 1,892 users on 17,632 artists between 2005-2011, generating 11,946 tags and 186,479 annotation behaviors. To improve computational efficiency, the original dataset was filtered. First, to avoid cold-start problems, users and artists with more than 70 annotations were selected. Second, to prevent malicious bot ratings from affecting dataset quality, users with fewer than 3,000 annotations were retained. Finally, to avoid excessive sparsity affecting clustering effectiveness, tags used more than 20 times were selected. The resulting core subset includes 444 users, 275 tags, and 372 artists, comprising 37,749 valid records (20.24% of total annotations). From this, 80% of data were randomly selected as the training set and the remaining 20% as the test set.

All users in the training set were randomly assigned a tag they had previously used. The algorithm generated a Top-N list for each user, and recommendation performance metrics were calculated by comparing these with resources annotated by corresponding users in the test set.

5.2 Performance Metrics

Recommendation accuracy is a crucial metric for evaluating recommendation algorithms. Current mainstream evaluation metrics for Top-N recommendation results include Precision, Recall, and F1 measure. Since the first two metrics influence each other, this paper uses Precision-Recall curves to qualitatively reflect algorithmic accuracy trends, and employs the F1 measure (the harmonic mean of precision and recall) to quantitatively reflect differences between algorithms. Simulation experiments were repeated 10 times, with the mean value of each metric across the 10 runs taken as the final result.

5.3 Performance Comparison

The simulation experiment mimics a scenario where users obtain a Top-N list by selecting a previously used tag, comparing performance metrics of different algorithms as recommendation length increases from Top-10 to Top-50 in increments of 5. The optimal number of clusters was determined to be 5 using the Modularity Metric.

To evaluate the performance of the proposed improved tensor decomposition model-based recommendation algorithm (CoSCluIDF), the simulation selected three comparison algorithms: (1) K-means clustering based on tag-user-resource matrices combined with IDF initial tensor improvement (KmeansIDF), to test the improvement effect of introducing tag co-occurrence spectral clustering on tensor decomposition model accuracy. The number of clusters was set to 5 for consistency. (2) Combined co-occurrence spectral clustering with traditional tensor decomposition without initial tensor modification (CoSClu), to test the impact of initial tensor improvements on recommendation performance. (3) Traditional tensor decomposition-based recommendation without any modifications (TD), to test the combined effect of both the comprehensive co-occurrence spectral clustering preprocessing and the improved popularity penalty mechanism in the initial tensor definition.

5.4 Accuracy Analysis

[Figure 1: see original paper] shows the Precision-Recall curve results, where each curve represents the precision and recall changes of an algorithm across different N values. When N is small, precision is high while recall is low; as N increases, precision decreases and recall increases. Curves closer to the upper-right corner indicate better recommendation performance.

Analysis of [Figure 1: see original paper]:
(1) The proposed algorithm generally outperforms the other three algorithms for recommendation lengths of 10-50. Compared with the second-best CoSClu, it achieves an average precision improvement of 2.69% and average recall improvement of 2.71%, demonstrating that using comprehensive co-occurrence spectral clustering combined with IDF and co-occurrence-based popularity penalty mechanisms improves recommendation accuracy.
(2) CoSCluIDF shows more significant improvement over KmeansIDF than over CoSClu, with average precision improvement of 5.00% and average recall improvement of 5.08%. This indicates that appropriate tag clustering contributes more to recommendation accuracy improvement than penalizing popular tags and resources.
(3) KmeansIDF shows slight disadvantage compared to CoSClu, with average precision difference of 2.21% and average recall difference of 2.26%, further

confirming that comprehensive co-occurrence spectral clustering improves recommendation accuracy more noticeably than penalizing popular tags and resources.

[Figure 2: see original paper] compares the F1 measures of the four algorithms.

Analysis of [Figure 2: see original paper]:
(1) Compared with KmeansIDF, CoSCluIDF achieves an average F1 improvement of 5.04% and maximum improvement of 5.91% (at N=15), indicating that comprehensive co-occurrence spectral clustering is more beneficial for improving recommendation accuracy than traditional K-means methods.
(2) Compared with CoSClu, CoSCluIDF achieves an average F1 improvement of 1.29% and maximum improvement of 1.90% (at N=35), demonstrating that the improved initial tensor definition incorporating IDF concepts and co-occurrence-based popularity penalty effectively eliminates the negative impact of popular tags and resources on algorithm accuracy.

Comparing [Figure 1: see original paper] and [Figure 2: see original paper] reveals that CoSClu, KmeansIDF, and CoSCluIDF all show substantial performance improvements over TD in both Precision-Recall curves and F1 measures. This indicates that sparsity most significantly affects recommendation accuracy, and clustering algorithms effectively address sparsity by setting a smaller number of clusters. However, when cluster numbers are identical, different clustering algorithms produce similar sparsity reduction effects, though comprehensive co-occurrence spectral clustering offers greater advantages in better identifying tag redundancy and semantic ambiguity to improve recommendation accuracy.

In practical applications, especially when users have already annotated some resources, systems rarely recommend only 10 or fewer resources. Instead, all resources above a certain threshold are sorted by weight and recommended. Therefore, larger recommendation lengths have more practical significance. The simulation results in [Figure 2: see original paper] show that CoSCluIDF achieves maximum F1 at N=20, which is practically meaningful. Thus, the ideal recommendation length is suggested to be N=20.

To avoid simulation results being affected by tag filtering, this paper specifically compares simulation results on a core subset where tag occurrence frequency is greater than 8 (denoted as Tag8) while keeping other filtering conditions unchanged. The mean performance of each algorithm across three accuracy metrics on different datasets is shown in .

**** shows that sparsity affects all algorithms' accuracy. However, even under high sparsity, the proposed CoSCluIDF performs best across all metrics, though its performance advantage slightly diminishes compared to the lower-sparsity Tag20 dataset. This indicates that noisy tags in highly sparse tag data affect the ability of comprehensive co-occurrence spectral clustering to discover reasonable tag clusters. Therefore, in practical applications, tag data should be cleaned as much as possible beforehand to reduce sparsity effects on the algorithm.

6. Conclusion

To optimize the accuracy of personalized recommendation results based on UGC tags, this paper introduces a comprehensive co-occurrence spectral clustering method in the tensor decomposition model. By preserving the semantic integrity of the <user, tag, resource> triple relationship, this approach effectively identifies similar tags and mitigates the impact of tag redundancy and semantic ambiguity on recommendation accuracy. Furthermore, to address the influence of popular tags and resources on recommendation accuracy in three-dimensional space, we propose a popularity penalty mechanism for co-occurring tags and resources based on IDF concepts from TF-IDF, redefining the initial tensor to both preserve triple semantic relationships and highlight users' personalized interests. Simulation experiments demonstrate that this method can fully utilize tag data information and effectively improve recommendation algorithm performance.

With the popularization of online social networking sites, personalized recommendation based on tags and trust relationships will receive broader attention. Future work will focus on further optimizing tensor decomposition model-based recommendation algorithms by making full use of tag cluster information and social network trust relationships.

References

[1] Moens M F, Li J, Chua T S. Mining User Generated Content[M]. CRC Press, 2014: 7-9.
[2] Marinho L B, Nanopoulos A, Schmidt-Thieme L, et al. Social Tagging Recommender Systems[M]. USA: Springer US, 2011: 615-644.
[3] Hitchcock F L. The Expression of a Tensor or a Polyadic as a Sum of Products[J]. Journal of Mathematics & Physics, 1927, 6(1): 164-189.
[4] Symeonidis P, Nanopoulos A, Manolopoulos Y. Tag Recommendations Based on Tensor Dimensionality Reduction[C]//Proceedings of the 2008 ACM Conference on Recommender Systems, Lausanne, Switzerland. ACM, 2008: 43-50.
[5] Liao Zhifang, Wang Chaoqun, Li Xiaoqing, et al. Tag Recommendation and New User Tag Recommendation Algorithms Based on Tensor Decomposition[J]. Journal of Chinese Computer Systems, 2013, 34(11): 2472-2476.
[6] Rendle S, BalbyMarinho L, Nanopoulos A, et al. Learning Optimal Ranking with Tensor Factorization for Tag Recommendation[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2009: 727-736.
[7] Wu Huijuan, Xu Baoxiang, Wang Yanyan. Optimization Research of Personalized Tag Recommendation Method Based on Tensor Decomposition[J]. Information Science, 2014, 32(6): 134-137.
[8] Celma S, Cano P. From Hits to Niches? or How Popular Artists Can Bias Music Recommendation and Discovery[C]//Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition, Las Vegas, Nevada. ACM, 2008: 1-8.

[9] Salton G, Buckley C. Term-weighting Approaches in Automatic Text Retrieval[J]. Information Processing & Management, 1988, 24(5): 513-523.
[10] Fleder D, Hosanagar K. Blockbuster Culture' s Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity[J]. Management Science, 2007, 55(5): 697-712.
[11] Wang Cheng, Zhu Zhigang, Zhang Yuxia, et al. Improvement in Recommendation Efficiency and Personalized of User-based Collaborative Filtering Algorithm[J]. Journal of Chinese Computer Systems, 2016, 37(3): 428-432.
[12] Cantador I, Bellogín A, Vallet D. Content-based Recommendation in Social Tagging Systems[C]//Proceedings of the 4th ACM Conference on Recommender Systems, Barcelona, Spain. ACM, 2010: 237-240.
[13] Xiang Liang. Practice of Recommendation System[M]. Posts & Telecom Press, 2012: 107-108.
[14] Rafailidis D, Daras P. The TFC Model: Tensor Factorization and Tag Clustering for Item Recommendation in Social Tagging Systems[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2013, 43(3): 673-688.
[15] Gemmell J, Ramezani M, Schimoler T, et al. The Impact of Ambiguity and Redundancy on Tag Recommendation in Folksonomies[C]//Proceedings of the 3rd ACM Conference on Recommender Systems, New York. ACM, 2009: 45-52.
[16] Leginus M, Dolog P, Žemaitis V. Improving Tensor Based Recommenders with Clustering[C]//Proceedings of the 20th International Conference on User Modeling, Adaptation, and Personalization, Montreal, Canada. Springer-Verlag, 2012: 301-307.
[17] Symeonidis P. ClustHOSVD: Item Recommendation by Combining Semantically Enhanced Tag Clustering with Tensor HOSVD[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2015, 46(9): 1-12.
[18] Shepitsen A, Gemmell J, Mobasher B, et al. Personalized Recommendation in Social Tagging Systems Using Hierarchical Clustering[C]//Proceedings of the 2008 ACM Conference on Recommender Systems, Lausanne, Switzerland. ACM, 2008: 259-266.
[19] Li H, Hu X, Lin Y, et al. A Social Tag Clustering Method Based on Common Co-occurrence Group Similarity[J]. Frontiers of Information Technology & Electronic Engineering, 2016, 17(2): 122-134.
[20] Li Ruimin, Lin Hongfei, Yan Jun. Mining Latent Semantic on User-Tag-Item for Personalized Music Recommendation[J]. Journal of Computer Research and Development, 2014, 51(10): 2270-2276.
[21] Symeonidis P, Nanopoulos A, Manolopoulos Y. A Unified Framework for Providing Recommendations in Social Tagging Systems Based on Ternary Semantic Analysis[J]. IEEE Transactions on Knowledge & Data Engineering, 2010, 22(2): 179-192.
[22] Lathauwer L D, Moor B D, Vandewalle J. On the Best Rank-1 and Rank-(R1, R2,···, RN) Approximation of Higher-Order Tensors[J]. SIAM Journal on Matrix Analysis & Applications, 2000, 21(4): 1324-1342.
[23] Kolda T G, Bader B W. Tensor Decompositions and Applications[J]. SIAM

Review, 2009, 51(3): 455-500.
[24] Pazzani M, Billsus D. Learning and Revising User Profiles: The Identification of Interesting Web Sites[J]. Machine Learning, 1997, 27(3): 313-331.
[25] White S, Smyth P. A Spectral Clustering Approach to Finding Communities in Graph[C]//Proceedings of the 2005 SIAM International Conference on Data Mining, Newport Beach, CA, USA. SIAM, 2005: 274-285.

Author Contributions

Chen Meimei: Conceived research ideas and specific research plan, wrote and revised the manuscript.
Xue Kangjie: Designed methods and implemented simulations, wrote and revised the manuscript.

Conflict of Interest Statement

All authors declare no conflict of interest.

Supporting Data

Supporting data is self-archived by the authors, E-mail: cmm@dhu.edu.cn.

Educators and institutions wishing to create or modify OER content can continue using the VitalSource Content Studio platform and VitalSource' s proprietary digital authoring tools, which provide intuitive operations for creating standards-based, responsive, interactive, and accessible content. Content created in VitalSource Content Studio can be distributed to students through the Bookshelf platform.

"This partnership will enable individuals and institutions to conveniently, scalably, and sustainably use OER," said Gerry Hanley, CSU Associate Vice Chancellor. "In the future, we will have a digital marketplace that provides educators and learners with the most affordable educational content, along with convenient and reliable distribution services."

(Compiled from: http://press.vitalsource.com/oer-adoption-made-easy-through-vitalsource-and-california-state-university)
(Journal correspondence)

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.