Tag Recommendation Based on High Order Singular Value Decomposition

Yuliang Shi [a)], Shuo Liu [b)]

Department of Informatics, Beijing University of Technology, Beijing, 100124, China.

[a)] shiyl@bjut.edu.cn
[b)] Corresponding author: 450550193@qq.com

Abstract. Social tag recommendation system provides the information sharing platform for users, allowing users to "tag" for the browsing of the items marked information. Tags both describe the semantics of the item and reflect the user's preferences. However, existing tag recommendation systems face the problem that different users may have different tags for the same item due to different interests, or the same tag may have different semantics for different users. To solve this problem, a three-dimensional tensor model is introduced. Three dimensions of the three-dimensional tensor are used to describe three types of entities in the tag recommendation system: users, objects and tags respectively. The high order singular value decomposition (HOSVD) is used to reduce the tensor model and realize the potential semantic association analysis among the three types of entities, so as to improve the accuracy of the tag recommendation system. Experimental results show that this method has significantly improved in accuracy and recall performance.

Key words: tag recommendation, tensor decomposition, HOSVD.

INTRODUCTION

While the explosive growth of Internet information, the recommendation system has become an effective tool to solve the problem of information overload. The emergence of massive information, on the one hand, makes it difficult for users to find the information they are interested in [1]. On the other hand, it also makes a lot of useful information become "dark information" in the network, which cannot be obtained by users.

Compared with other explicit or implicit rating information user information, tag information is lightweight, it is easy to understand, and versatility [2]. Unlike other types of content information, tags directly provided by the user, users can use keywords to describe any type of content can be music, video, pictures and other multimedia content, it also can be web pages, documents and other information. Because of their simplicity and multiple functions, tag widely used in various applications, such as, Delicious, Flickr, Amazon and Last.fm and other sites.

In this paper, we analyze the tag metadata three-dimensionally and improves the accuracy of the tag recommendation system according to the potential semantic relations between the three types (users, items, tags) [3]. Under normal circumstances, such a ternary relationship only part of the data records, most of the relationship is no data record, but we want to get some missing data records, which requires the use of tensor decomposition will be the original space The relationships are mapped to low-dimensional spaces, and the latent tensors are used to derive the approximate tensor of the original tensor, and the approximate tensor can recover the missing data, giving an explanation of some potential relationships in the original tensor. The recommendation system uses these approximate data as a reference to decide which tags to recommend to the user.

RELATED WORK

Present situation of the concept of social bookmarking is still no uniform definition, but scholars generally agree that social bookmarking is generated by the user, the previous system of Web content is organized and shared bottom-up organization classification system compared , scientific classification of resources are mostly relying on a small number of experts in the field, although this method is more scientific reason and authority, but not well express your personal thoughts and ideas, and this is a social bookmarking been widely welcomed.

Social tag system is not only a tool for classification of information resources, but also a reflection of user's personal preferences [4]. Social tag system provides a new source of data for recommending systems to gain user preference information, while presenting new challenges to traditional two-dimensional data-based recommendation techniques. The subjects of recommended research in social tag systems include users, items (products) and tags.

General recommendation system includes a user tag, these three elements of the article and the tag, the user can annotate items plurality of tags, tagging methods are mainly two: one is a manual annotation, to choose the right from its full vocabulary by the user Words to tag resources; the other is semi-automatic annotation, that social tag system for the user recommended tag list, the user selects the appropriate tag for tagging[5].

TENSOR DECOMPOSITION TECHNIQUES

Traditional information expressions such as matrices and vectors can no longer handle such complex multidimensional information. For this reason, tensors should come out [6]. Tensors can be seen as an extension of vectors and matrices, the vector are considered as one-dimensional tensors, matrices are considered as two-dimensional tensors, and multi-dimensional relations of three dimensions and above are represented by noun tensors. But here the tensor is different from the tensors in physics and engineering, such as the stress tensor, but specifically refers to a concept in mathematics [7]. Each dimension of tensor is called the Mode or Order of the tensor. From the perspective of structure, tensor can not only store the information of each dimension, but also the relationship between dimension and dimension. The field of dealing with tensors in mathematics is called multiple linear algebra. A third-order tensor can be expressed as Fig 1.
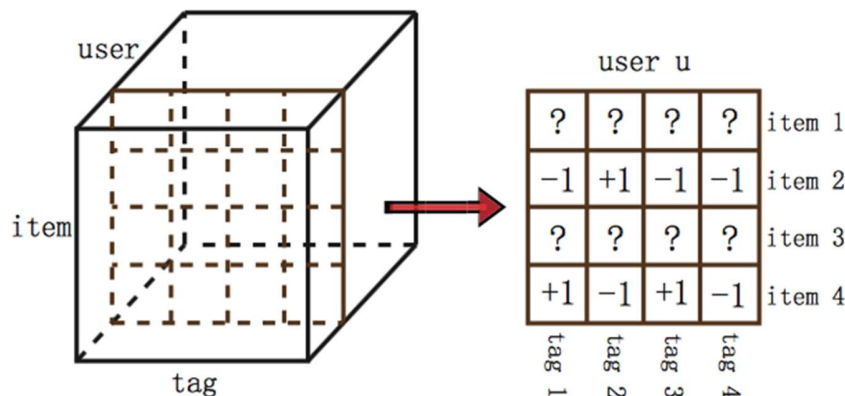

FIGURE 1. Tensor construction

Singular Value Decomposition

To decompose the tensor A by a high-order singular value, we first need to expand the tensor by a matrix, that is, rearrange the tensors into new matrices according to different dimensions (n-mode). The tensor in the paper is a three-dimensional tensor, so the three dimensions of the tensor A are respectively expanded by definition to form a tensor 1-mode, a 2-mode and a 3-mode expansion A1, A2 and A3 respectively as follows:

$$A_1 \in R^{I_u \times I_t I_i}, A_2 \in R^{I_t \times I_u I_i}, A_3 \in R^{I_i \times I_u I_t} \tag{1}$$

Then the singular value decomposition of the expanded matrix is obtained respectively, and the singular value decomposition is used to obtain the diagonal matrix S(n) which is composed of singular values of the matrix An. The decomposition is as follows:

$$A_n = U^{(n)}.S^{(n)} \cdot \left(V^{(n)}\right)^T, 1 \leq n \leq 3 \tag{2}$$

In the process of tensor decomposition, the most important one is the calculation of the low rank approximation of the matrix, that is, the singular values of the matrix an are truncated (c large singular values are retained before c<min{I1,I2}, where c can be determined experimentally by preserving the percentage of the original information in the diagonal matrix $s^i(1 \leq i \leq 3)$. Low-rank approximation can well filter out the noise caused by small singular values, so as to achieve the purpose of noise reduction.

High Order Singular Value Decomposition

Higher-order singular value decomposition is the generalization of singular value decomposition in tensor. The higher-order singular value decomposition of tensor is the product form of decomposing the tensor into the same size of the core tensor and multiple matrices. In this paper, the high-order singular value decomposition of three-dimensional tensor A is expressed as:

$$A = S \times_1^{\square} U^{(1)} \times_2^{\square} U^{(2)} \times_3^{\square} U^{(3)} \tag{3}$$

In the formula, the core tensor $S \in R$ is an orthogonal tensor with the same dimension as the tensor A, which determines the interaction between the entity user, item and tag. The mathematical expression of S is:

$$\hat{S} = A \times_1^{\square} U_{c_1}^{(1)^T} \times_2^{\square} U_{c_2}^{(2)^T} \times_3^{\square} U_{c_3}^{(3)^T}. \tag{4}$$

Finally, due to the large amount of noise in the tensor data, the tensor A does not have low rank, and it is necessary to construct the approximate tensor A of the tensor A through the HOSVD. The mathematical expression for reconstructing the tensor A is as follows:

$$\hat{A} \approx \hat{S} \times_1^{\square} U_{c_1}^{(1)} \times_2^{\square} U_{c_2}^{(2)} \times_1^{\square} U_{c_3}^{(3)}. \tag{5}$$

EXPERIMENTAL RESULTS

Data Set

MovieLens was founded by the Group Lens project team at the University of Minnesota in the United States. The oldest recommended system at present is a non-commercial, research-oriented movie recommendation experimental site. In 2006 Movie Lens introduced the characteristics of social tags and began testing for tag recommendations. The data set triple {u, i, t} represents the user, the movie, and the public tag, respectively.

Experimental Setup

The accuracy of the traditional classic evaluation index describes how many percent of the recommended recommendations in the final recommendation table are sent, and the recall describes how many of the recommended recommendations are included in the final recommendation table. Based on this, this paper makes some adjustments to the traditional evaluation indexes and divides the tags of each test user into two groups, in which the test set sizes of the first group and the second group default to 50%.

Group 1: Tested user previously tagged (defined as $T_1(u)$);
Group 2: A tag (defined as $T_2(u)$) that the predictive test user will hit;

In this paper, the accuracy and the recall rate are used to evaluate the accuracy and validity of the proposed algorithm. The two evaluation indexes are defined as follows:

$$precision = \frac{R(u) \cap T_2(u)}{N} \tag{6}$$

$$recall = \frac{R(u) \cap T(u)}{T_2(u) \cap T(u)} \tag{7}$$

Result Analysis

This section compares the tensor decomposition algorithm with the Folk Rank and PR algorithms using the accuracy and recall as the evaluation criteria. The robustness of each algorithm is obtained by obtaining the higher recall rate at the cost of reducing the accuracy as much as possible.

There are three curves in the Fig.2, each node has 5 nodes, each node represents a tag, and the position of the node from left to right represents the rank of this tag in the top-N list, which is 1 to 5 in turn. As can be seen from the figure, the accuracy of each algorithm decreases with the rise of N; on the contrary, as the value of N increases, the recall rates of the three algorithms all increase. According to the tag ranking in the recommendation list, the accuracy and recall rate of the Zhang decomposition algorithm are higher than those of the algorithms Folk Rank and PR. The experimental results show that the tensor decomposition algorithm is robust.
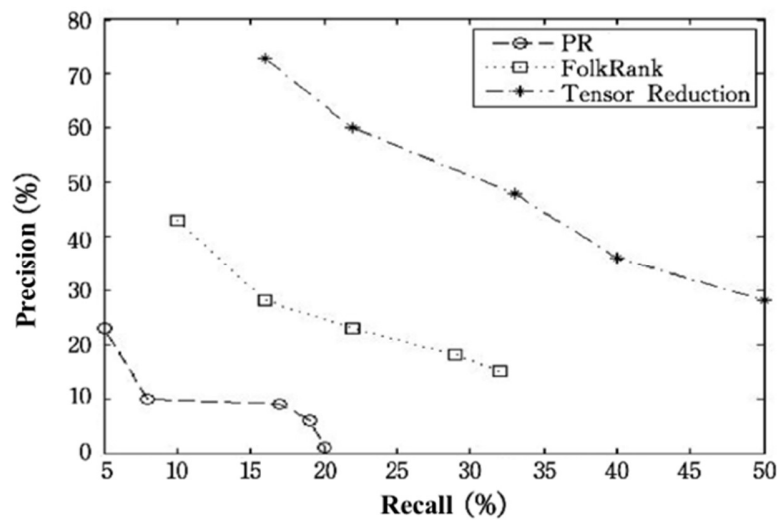

FIGURE 2. Comparison of experimental data based on Movie Lens dataset

Experimental results show that the tensor decomposition algorithm in two different data sets has higher accuracy and robustness than the tags provided by Folk Rank and PR algorithms.

CONCLUSION

In this paper, a three-dimensional tensor model is introduced to describe the three types of physical users, objects and tags existing in social tagging recommendation system. The three-dimensional tensor is used to store the metadata of the three types of entities in the system, and the HOSVD algorithm is used to reduce the dimensionality of the three-dimensional tensor to mine potential semantic associations among the three types of entities. The tensor decomposition is compared with two classical algorithms Folk Rank and PR. The results show that the tag recommendation algorithm based on tensor decomposition is accurate and the recall rate has greatly improved.

REFERENCES

1. Yuen M C, King I, Leung K S. An online-updating algorithm on probabilistic matrix factorization with active learning for task recommendation in crowdsourcing systems[J]. Big Data Analytics, 2016, 1(1):14.
2. Zou B Y, Cui-Ping L I, Tan L W, et al. Social Recommendations Based on User Trust and Tensor Factorization[J]. Journal of Software, 2014.
3. Qin X. A Study of the Personalized Micro-Blog Recommendation Algorithm Based on Tensor Factorization[J]. Software Engineering, 2017.
4. Du L, Li X, Shen Y D. User Graph Regularized Pairwise Matrix Factorization for Item Recommendation[M]// Advanced Data Mining and Applications. Springer Berlin Heidelberg, 2011:372-385.
5. Li G, Zhang X, Zheng S, et al. Semi-supervised Convex Nonnegative Matrix Factorizations with Graph Regularized for Image Representation[J]. Neurocomputing, 2016, 237(C):1-11.
6. Kolda T G, Bader B W. Tensor Decompositions and Applications[J]. Siam Review, 2009, 51(3):455-500.
7. Marinho L B, Nanopoulos A, Schmidtthieme L, et al. Social Tagging Recommender Systems[J]. 2014:20:57–20:75.