Image Tag Recommendation Algorithm Using Tensor Factorization

Wang Yong-sheng
Lanzhou Jiaotong University, Lanzhou, China
Email: wys_lz@126.com

Abstract—This paper aims to provide high quality tags for digital images according to users’ interest. As there are three main elements in image tag recommendation problem, tensor factorization technology is utilized in this work. In this paper, the parameters of the tensor factorization model are represented as latent variables, and the key functions of the tensor factorization model can be implemented by integrating three matrices (person matrix, image matrix, and tag matrix) into one tensor. The key problem of image tag recommendation is to obtain the top ranked tags which are suitable not only to image visual contents but also to users’ interest. Afterwards, the top ranked tags are obtained by a predictor utilizing the proposed tensor factorization model. Therefore, the image tag recommendation problem can be converted to calculate the ranking scores by maximizing the ranking statistic AUC. Finally, performance evaluation is conducted on the NUS-WIDE dataset using MRR, S@k, P@k, and NDCG metric. Experimental results show that the proposed image tag recommendation algorithm performs better than other methods.

Index Terms—Tag Recommendation; Tensor Factorization; Heaviside Function; Mean Reciprocal Rank; NDCG

I. INTRODUCTION

With the explosive growth of multimedia media applications on the internet, a lot of digital images have been made available in many Web2.0 based multimedia web sites in recent years. Normal users can easily find an image and then share it with their friends or public. Online social networks and social web sites, such as Flickr, Facebook, and PicasaWeb, host billions of digital images, which have been shared and tagged among friends, or published in groups that cover some specific topic of interest. This has provided an open and critical challenge of building an effective digital image search engine to retrieve images over the massive collections of digital images on the internet [1].

On the other hand, it can be seen that there is a rapid development trend of Web 2.0 based multimedia information sharing websites, for example YouTube, Flickr, and so on. These websites can effectively help users to collaboratively upload, browse, and download multimedia data. Furthermore, the above websites permit users to provide some terms to tag the multimedia data which they uploaded. Unfortunately, retrieving useful information on the large-scale multimedia data is fairly hard. Particularly, some photo sharing website (such as Flickr) could provide two ranking methods for image search with tags provided by users [2] [3].

As is illustrated in the above, photo sharing websites allow users to illustrate the multimedia data with terms (also named tags), which can promote the performance of information retrieval. However, manually annotating all social images in a personal album is hard to implement, however and allocating low quality tags to the multimedia data would reduce the tagging precision [4]. Tags are very popular in multimedia information retrieval and can give descriptive information for multimedia data. However, considering there are some noisy words in the user-supplied tags. Hence, it is of great importance to research on tag recommendation systems and algorithms [5] [6]. Particularly, tag recommendation algorithm can effectively lower the data sparsity problem in social tagging systems.

Tag recommendation technology can help users to describe the multimedia data to manage and retrieve personal data. Furthermore, collective tag recommendation refers to let the multimedia data more available to other users by tag recommending process which can facilitate information retrieval [9-12]. However, as tags can not be selected in a fixed vocabulary, users can choose specific terms freely. This process can reduce the usefulness of tags in particular for resources annotated by only a few users [7] [8].

The rest of the paper is organized as the following sections. Section 2 introduces the related works. Section 3 illustrates the proposed scheme for image tag recommendation. In section 4, experiments are conducted to make performance evaluation with comparison to other existing methods. Finally, we conclude the whole paper in section 5.

II. RELATED WORKS

The resources of images are very popular for the Web 2.0 platform, and there are many related works are conducted using digital images. Some typical works are listed as follows.

Zhu et al. proposed a new method to tag personal photos. For a given personal album, an affinity propagation algorithm is used to obtain a set of representative images, and the number of representatives depends on the image’s visual contents. Hence, for extracted representative images, both visual and semantic information are utilized to calculate relevance scores for
each word. Furthermore, random walk algorithm is
exploited re-calculate the above scores. Finally, tags of
the rest images are automatically obtained via a graph-

based semi-supervise algorithm [4].

Si et al. proposed the cross-domain discriminative
locally linear embedding, which can connect the training
and the testing samples by minimizing the quadratic
distance between the distribution of the training samples
and that of the testing samples. In this paper, basically
expect the discriminative information are provided to
separate the concepts in the training set can be shared to
separate the concepts in the testing set as well and thus
we have a chance to address above cross-domain problem
duly [13].

Sun et al. proposed a generic, flexible, and extensible
framework and exploit it for a systematic and
comprehensive empirical evaluation of various methods
for ranking images. Furthermore, the authors identified 5
orthogonal dimensions to quantify the matching score
between an image with descriptive Information and a tag
query [14].

Tag recommendation is very useful for social images
retrieval, and it has been attracted many researchers’
attentions. The tag recommendation related works is
given as follows.

Hsu et al. presented a hybrid method based on
semantic tag-based resource information and user
preference to obtain personalized social tag
recommendation results to users. Experiments show that
the proposed hybrid approach can effectively promote the
precision of social tag recommendation under the metric
of precision and recall [15].

Chen et al. studied on the problem of tag
recommendation for multimedia based Web 2.0 platform.
This problem is as important as the tag recommendation for items, and the reason lies in that the tags can represent
the users’ interests. The authors also presented several
novel features of tags for machine learning which is
named social features as well as textual features.
Afterwards, by the experimental dataset selected from

Flickr, the proposed method can provide accurate tags for

users [5].

Rawashdeh et al. addressed the problem of
recommending suitable tags during folksonomy
development from a graph-based perspective. The
proposed approach adapts the Katz measure, a path-
ensemble based proximity measure, for the use in social
tagging systems. They modeled a folksonomy as a
weighted, undirected tripartite graph, and then applied the
Katz measure to this graph, and exploited it to provide tag
recommendations for individual users [16].

Gedikli et al. et al. proposed new schemes to infer and
exploit context-specific tag preferences in the
recommendation process. An evaluation on two different
datasets reveals that the proposed approach is capable of
providing more accurate recommendations than previous
tag-based recommender algorithms and recent tag-
agnostic matrix factorization techniques [17].

Tensors refer to geometric objects which can represent
relationships between vectors, scalars. Elementary
examples of such relations include the dot product, the
cross product, and linear maps. It can be seen that vectors
and scalars can be regarded as tensors. Therefore, a tensor
can be defined as a multi-dimensional array with
numerical values [18]. In the following section, we will
introduce the related works about applications on tensor
factorization.

Ozerov et al. introduced Coding-based ISS (CISS) and
draw the connection between ISS and source coding.
CISS amounts to encode the sources using not only a
model as in source coding but also the observation of the
mixture. First, it can reach any quality, provided
sufficient bandwidth is available as in source coding.
Second, it makes use of the mixture in order to reduce the
bitrate required to transmit the sources, as in classical ISS
[19]. Other research about applications on tensor
factorization can be found in paper [20-22].

III. THE PROPOSED SCHEME

Tensors refer to generalizations of vectors (first-order
tensors) and matrices (second-order tensors) to arrays of
higher orders (\( N > 2 \) ). Hence, a third-order tensor is an
array with elements \( x_{i,j,k} \).

Particularly, factorizing
tensors have several advantages than 2-way matrix
factorization such as uniqueness of the optimal solution
and component identification even when only a relatively
small fraction of all the data is observed.

Surposing two third-order tensors are represented as
\(A^{i,j,l}\) and \(B^{i,j,l}\). The basic operations of tensors are 1)
scalar multiplication, 2) addition and 3) inner product,
which are defined as follows.

\[
\alpha B = C, \text{ where } c_{i,j,k} = \alpha b_{i,j,k} \tag{1}
\]

\[
A + B = C, \text{ where } c_{i,j,k} = a_{i,j,k} + b_{i,j,k} \tag{2}
\]

\[
< A, B > = \sum_{i,j,k} a_{i,j,k} b_{i,j,k} \tag{3}
\]

We suppose \( X \) can be calculated by the three low rank
matrices and a single tensor based on tensor factorization.
Each element in the proposed image tag recommendation
problem includes “Person”, “Image”, and “Tag”.
Moreover, the parameters of the tensor factorization
model can be represented as latent variables. Therefore,
the core functions of the tensor factorization model are
implemented by integrating three matrices into one key
tensor, and the process is defined as follows.

\[
X = K \times_p P \times_q \hat{I} \times_T T \tag{4}
\]

where the key tensor \( K \) and the three matrices \( P \), \( \hat{I} \),
and \( T \) denote the parameters of the proposed model
which should be estimated. The symbol \( \times \) denotes the
tensor product to multiply a matrix on the specific \( u \) with
a given tensor. The size of the above parameters is given
in the following four conditions.

\[
\text{Condition1: } K \in \mathbb{R}^{k_x \times y_x \times y_T} \tag{5}
\]
Condition 2: \( P \in \mathbb{R}^{[P]} \) 

(6)

Condition 3: \( \hat{I} \in \mathbb{R}^{[P]} \) 

(7)

Condition 4: \( T \in \mathbb{R}^{[P]} \) 

(8)

where \( k_P, k_t, \) and \( k_T \) means the dimensions of the low-rank approximation. Furthermore, the parameter of the tensor factorization model can be represented by the following quadruple process:

\[
\theta = (K, P, \hat{I}, T)
\]

(9)

The image tag recommendation problem is made up three entities: Persons, Images and Tags, therefore, this problem can be solved by the tensor factorization process. The set of all persons, tags and images are defined as \( P = \{p_{i}^t\}_{i=1}^{t}, \ T = \{t_{j}^p\}_{j=1}^{p}, \) and \( I = \{I_{k}\}_{k=1}^{K} \) respectively. The relation between person and tag is defined as \((p, t) \in O\subseteq P \times T\), which denotes the person \( p \) has tagged the given image with tag \( t \). Afterwards, the set of all tags which have been visited by person \( p \) is represented as \( T(p) \). Furthermore, the relationship between different persons are defined as \((p_i, p_j) \in Q\subseteq P \times P\), which means the person \( p_i \) and \( p_j \) are friends for each other. The key problem of image tag recommendation lies in that a personalized image tags according to user interest should be predicted for the specific person.

Supposing there is a predictor \( X \), the set of top ranked tags(belonged to image \( I_{k} \)), which are suitable for the specific user’s(denoted as \( P_{i} \)) interest can be obtained by the following equation.

\[
\text{Top}(P_{j}, I_{i}, N) = \text{Arg} \max_{i,j} y_{j,i}
\]

(10)

where \( N \) means the number of tags which are returned by the proposed tag recommendation system. The image tag recommendation problem can be solved by the tensor factorization model, and the process can be described by Fig. 1.

In Fig. 1, the positive examples are represented by 1 and the negative examples are denoted as 0. Particularly, the recommended tags according to specific person’s interest can be obtained by the tensor factorization process, of which three matrices (“Person”, “Image”, “Tag”) are included.

Based on the above analysis, the parameters of the proposed tensor factorization model computed. There are four main parameters utilized in this paper, which are \( K \), \( P \), \( \hat{I} \), and \( T \). In this paper, the tensor factorization could be calculated by minimizing the following square loss.

\[
\text{Arg} \min_{\theta} \sum_{(p, j, i) \in P \times I \times T} (x_{p,j,i} - \hat{x}_{p,j,i})^2
\]

(11)

Based on the analysis in Eq.11, this optimization problem can be converted to maximize the ranking statistic AUC, which denotes the area under the ROC curve. Hence, the quality measure AUC for a specific person \( p \) with the image \( i \) can be represented as follows.

\[
\text{AUC}(\theta, p, i) = \sum_{l \neq i} \sum_{m \neq i} H_{\theta}(x_{p,l,i} - x_{p,m,i})
\]

(12)

where \( H_{\theta} \) refers to the Heaviside function. Next, the image tag recommendation problem can be converted to compute the ranking scores as follows.

\[
\text{arg} \max_{(p, i) \in S} \text{AUC}(\theta, p, i)
\]

(13)

where \( S \) denotes the observed posts of the digital image dataset. The optimization problem in Eq.13 can be solved by the Eq.14.

\[
\frac{\partial(AUC(\theta, p, i))}{\partial \theta} = \frac{1}{|\hat{I}_{p,i}||I_{p,i}|} \sum_{l \neq i} \sum_{m \neq i} \frac{1}{1 + e^{-H_{\theta}(x_{p,l,i} - x_{p,m,i})}}
\]

(14)

IV. EXPERIMENTS

As is well known that performance evaluation for digital image tag recommendation is quite difficult problem, the reason lies in the lack of a standard dataset
which is suitable to be used as a benchmark. In this section, performance evaluation is conducted on the particularly digital image datasets, and some typical works are compared with the proposed algorithm. In this experiment, the NUS-WIDE digital image dataset is utilized, which is the largest publicly available human-annotated dataset in digital image processing [23].

A. Overview of the Dataset

In the NUS-WIDE digital image dataset, Chua Tat-Seng et al. randomly crawled more than 300,000 digital images together with the user-supplied tags from the Web2.0 based digital image sharing website Flickr using its public API functions. To make this dataset more useful, the images whose sizes are too small or with inappropriate length-width ratios are deleted. Moreover, the authors delete a lot of duplicate images through features matching. The remaining dataset is made up of 269,648 images with a total of 425,059 unique user-supplied tags. For all images together with the user-supplied tags from the annotated dataset in digital image processing [23], utilized, which is the largest publicly available human-annotated dataset in digital image processing [23].

B. Evaluation Metrics

In this experiment, we applied three metrics for performance evaluation, which have been proposed in paper [25]. These three metrics can capture the performance for different aspects, which are: 1) Mean Reciprocal Rank (MRR). MRR measures where in the ranking the first relevant tag is returned by the system, averaged over all the photos. This measure provides insight in the ability of the system to return a relevant tag at the top of the ranking. 2) Success at rank k (S@k). Particularly, in this paper, we choose the success at rank k for two values of k: S@1 and S@5. The success at rank k is defined as the probability of finding a good descriptive tag among the top k recommended tags. 3) Precision at rank k (P@k). The parameter we select for P@k is the precision at rank 5 (P@5). Precision at rank k refers to the proportion of retrieved tags that is relevant, averaged over all photos [25].

Our proposed digital image tag recommendation algorithm can automatically rank all the tags obtained by the proposed algorithm. Hence, the NDCG metric is utilized to make performance evaluation. As is illustrated in paper [26], NDCG is used for element ranking results evaluation. In the case of tag recommendation, for a given digital image, each of its tags is tagged as one of the five degrees: 1) Most Relevant (score 5), 2) Relevant (score 4), 3) Partially Relevant (score 3), 4) Weakly Relevant (score 2), and 5) Irrelevant (score 1).

Given an image with ranked tag set \( \{ t_1, t_2, \ldots, t_n \} \), the NDCG can be represented as follows:

\[
N_n = Z_n \sum_{i=1}^{n} \frac{2^{(r(i))} - 1}{\log(1+i)}
\]

where \( r(i) \) refers to the relevance level of the \( i^{th} \) tag and \( Z_n \) represents a normalization constant which is selected to make sure that the optimal ranking’s NDCG score is equal to one. After calculating the NDCG measures of each digital image’s tag set, we average these scores to achieve an overall performance evaluation of a specific tag ranking algorithm.

C. Performance Evaluation and Analysis

To make performance evaluation more objective, three typical approaches are used for comparison. The
proposed three methods we selected are 1) Tag recommendation by machine learning with textual and social features (TRTS) [5], 2) Digital Image Tag Recommendation by Concept Matching (SITRCM) [24], and 3) Flickr tag recommendation based on collective knowledge (TRCK) [25].

Firstly, we make the performance evaluation on the six categories of the NUS-WIDE social image dataset, which are “Events/Activities”, “Program”, “Scene/Location”, “People”, “Objects”, and “Graphics”.

Table 1 shows that the proposed image tag recommendation algorithm using tensor factorization (ITRTF) performs better than other three methods (TRTS, SITRCM, and TRCK) for all the four metrics (MRR, S@1, S@5, and P@5). Combining all the six categories of NUS-WIDE dataset, we can see that 1) for the MRR metric, the proposed algorithm (ITRTF) promotes the performance 4.86%, 15.3% and 19.2% respectively, 2) for the S@1 metric, the proposed algorithm (ITRTF) promotes the performance 5.39%, 10.5% and 22.1% respectively, 3) for the S@5 metric, the proposed algorithm (ITRTF) promotes the performance 66.2%, 10.1% and 39.2% respectively, and 4) for the S@1 metric, the proposed algorithm (ITRTF) promotes the performance 8.5%, 10.8% and 22.2% respectively.

Secondly, to illustrate the tag recommendation performance more detailedly, the Success@K metric is used to make performance evaluation when the number K, number of user-supplied tags, and social image type changing. The related experimental results are given in Fig. 2 - Fig. 4.

Thirdly, we will test the performance evaluation of tag ranking for the proposed tag recommendation algorithm, because the tags recommended by the proposed algorithm are provided with recommending scores. The performance evaluation on NDCG for different NDCE depth is shown in Fig. 5.

From the experimental results in Fig. 2-Fig. 5, the conclusions can be drawn that for all image categories in NUS-WIDE dataset the proposed algorithm ITRTF can provide the image tags for users with high quality than other three methods.

To give more descriptive results, top ranked images returned by the proposed algorithm for a specific term is illustrated in Fig. 6, and several examples of digital image tag recommendation results by the proposed algorithm is shown in Table 2.

Experimental results in Fig. 6 demonstrate that there are many pictures returned by the image search engine which contain the object dog. It proves that the salient objects (such dog) can be detected by the proposed tag recommendation algorithm effectively.

As is shown in Table 2, we can see that the tag recommendation results obtained by the proposed algorithm can effectively supplement the semantic in user-supplied tags even when there are no user-supplied tags in a given image (please see the sixth image in Table 2).
TABLE II. EXAMPLES OF DIGITAL IMAGE TAG RECOMMENDATION RESULTS BY THE PROPOSED ALGORITHM

<table>
<thead>
<tr>
<th>Digital image</th>
<th>User-supplied tags</th>
<th>Tag recommendation results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greece, Travel, Boat</td>
<td>Sea, Sailboat, Yacht, Sky, Beach</td>
<td>Football, Grass, Children, Running</td>
</tr>
<tr>
<td>Soccer</td>
<td>Girl, Smiling, Finger, Face, Cell Phone</td>
<td></td>
</tr>
<tr>
<td>Digital image</td>
<td>Helgufoss, Island, Iceland</td>
<td>Carr, brae, viewpoint, dorne, inverinate, west, coast, Scotland, hills, glenshiel, mountains, View, allanamciver, snow</td>
</tr>
<tr>
<td>Tree, Lake, Water, Waterfall, Stone</td>
<td>Tree, Cloud, Tree, River, Building, Grass, Snow mountain</td>
<td>Human, Tree, Bicycle, Grass, Lane</td>
</tr>
</tbody>
</table>

Figure 6. Top results for the query term “dog” utilizing the proposed tag recommendation algorithm

From the above experimental results, it can be seen that the proposed scheme is superior to other schemes. The main reasons lie in the following aspects:

1. **TRTS** is a method to effectively find representative tags for users that are related to the users’ favorite topics in large-scale online communities. However, the information of user’s personalized interest are not effectively represented and extracted in this paper. Therefore, the performance of this method is not satisfied.

2. **SITRCM** is knowledge-based approach for image tag recommendation that exploits tag concepts, which are derived based on the collective knowledge embedded in tag cooccurrence pairs. However, the tag recommendation results of this approach are influenced by the initial user-supplied tags. If the quality of the initial tags is not very high, the tag recommendation results are low as well. Furthermore, our proposed algorithm can effectively solve this problem.

3. **TRCK** is a method for recommending tags, and this approach deploys the collective knowledge that resides in Flickr without introducing tag class specific heuristics. Particularly, this approach extracted tag co-occurrence statistics, which can combine with the two tag aggregation strategies. However, this approach can not effectively analyze the relationship between person, image and tags.

On the other hand, the proposed tag recommendation algorithm can effectively solve the above problems in the former methods.

V. CONCLUSIONS

In this paper, we present a novel image tag recommendation algorithm based on tensor factorization. The core functions of the tensor factorization model is implemented by integrating “person matrix”, “image matrix”, and “tag matrix” into one tensor. The recommended tags are obtained through tag ranking process by a predictor utilizing the tensor factorization model. The main innovations of this paper lie in that we convert the image tag recommendation process to calculate the ranking scores by maximizing the ranking statistic AUC.

REFERENCES

[1] Jin-Woo Jeong, Hyun-Ki Hong, Jee-Uk Heu, Qasim I., Dong-Ho Lee, Visual Summarization of the Social Image Collection Using Image Attractiveness Learned from


[9] Zhang Xiaoming, Zhao Xiaojian, Li Zhouchun, Social image tagging using graph-based reinforcement on multi-type interrelated objects, Signal Processing, 2013, 93(8) pp. 2178-2189


[17] Gedikli Fatih, Jannach Dietmar, Improving Recommendation Accuracy Based on Item-Specific Tag Preferences, ACM Transactions on Intelligent Systems and Technology 2013, 4(1), Article No: UNSP 11

[18] Li Li, Yan Hua, Cost aggregation strategy with bilateral filter based on multi-scale nonlinear structure tensor, Journal of Networks, 2011, 6(7) pp. 958-965


[22] Hayashi Kohei, Takenouchi Takashi, Shibata Tomohiro, Exponential family tensor factorization: an online extension and applications, Knowledge and Information Systems 2012, 33(1) pp. 57-88


