

# Cross-Domain Relationship Prediction by Efficient Block Matrix Completion for Social Media Applications

Lizhi Xiao<sup>a</sup>, Zheng Zhang<sup>a</sup>, and Peng Sun<sup>b,\*</sup>

<sup>a</sup>The Information Center Henan Radio Television University, Zhengzhou, 450000, China

<sup>b</sup>School of Computer Science and Engineer, University of Electronic Science and Technology of China, Chengdu, 611731, China

---

## Abstract

The online social media has experienced vigorous evolution. Diversified needs of information acquisition and retrieval on social media platforms have been evoked by massive users. While all sorts of application demands meet with explosive data growth, the development of effective methodologies has become emergent. By taking full advantage of rich context, we propose a heterogeneous object relation matrix completion approach (EBMC) which jointly complements the relationship between the heterogeneous data objects. Specifically, we detect the Place-of-Interest (POI) with mean shift algorithm on the GPS information of the social image collection. Then, a batch matrix completion and learning method is developed by optimizing a unified objective function to learn the POI-specific user-image, image-tag and user-tag relationships. Finally, we decompose the whole learning problem into a set of POI-specific subtasks, which corresponding to the relation data blocks separated by the POI structure. Through experiments on tasks of image annotation and user retrieval based on image similarity of real-world social media datasets, we found that our proposed method achieved good performance.

*Keywords:* matrix factorization; multimedia retrieval; place-of-interest; matrix completion

(Submitted on April 22, 2020; Revised on May 28, 2020; Accepted on June 30, 2020)

© 2020 Totem Publisher, Inc. All rights reserved.

---

## 1. Introduction

On interactive Online Social Media, active users produce massive user-generated content (UGC) with rich attributes and social connections every second. The connection between heterogeneous data objects has become more complicated. In fact, users have raised a variety of information acquisition requirements. While all sorts of application demands meet with explosive data growth, it is impossible to promote the sharing and dissemination of knowledge due to the lack of technology to model the relationship between heterogeneous data objects. Thus, it is necessary to develop an effective method to solve the retrieval problem of heterogeneous data objects.

Matrix completion [1-2] has been hot research issue in a variety of research fields such as data mining, machine learning and information retrieval. We review some of recent works from methodology and application perspectives. The matrix completion problem has been recognized and modeled as a low-rank matrix factorization problem [1-2]. Following this direction, fruitful achievement has been made in the past decade. Moreover, research endeavors have also enhanced the learning ability of matrix completion. For example, inductive matrix completion [3] has also been proposed in recent years as a new way to deal with out-of-sample relation completion under matrix completion framework.

We explore the relationships among three important types of objects from rich social context. Specifically, the user-image relationship reflects the users' attachment to the images, the image-tag relation indicates the social tagging distribution on a different visual content, and the user-tag relationship represents the user behavior and inclination in the tagging activities. Due to the complexity of unpredictable user behavior [4], this object relationship is noisy, and observations are sparse. We jointly complement the intrinsic relationships between objects from multiple observed relationships. By considering visual content as the centric component, we design several pairwise relationship constraints to describe heterogeneous object relationships based on assumptions from different perspectives. Compared with other relationship-specific learning models,

\* Corresponding author.

E-mail address: [psun@uestc.edu.cn](mailto:psun@uestc.edu.cn)

the object relationships can be inferred to be more reliable [5-6] but needs to be learned about more thoroughly.

The rich social environment of heterogeneous objects can be used to deal with the object-relational learning tasks. It is reasonable to constrain the matrix completion task with respect to the Place of Interests (POI) [7-8]. We learn the POI-specific matrix blocks from different observed relationship matrices within a unified objective function, where the information averaging is performed within an appropriate POI-specific range. Our model captures more informative relationship patterns among heterogeneous objects in rich context and the learned relationships are naturally interpretable; thus, it better fits to the information acquisition demands in heterogeneous data object retrieval. On the other hand, by decomposing the tasks into POI-specific relationship block matrix completion sub-tasks, we gain reduced memory cost and improved computational efficiency. The model optimization is well-established in batch learning style. At the same time, the alternate sub-gradient descent algorithm can effectively solve the problem of learning POI specific relation matrix blocks in alternate optimization mode. Therefore, our model can be scaled towards large scale object relationship matrix completion.

By taking full advantage of the rich social context, we propose an efficient block matrix completion model (EBMC) which jointly learns the completed relationships among heterogeneous objects. The relationship matrix completion can be efficiently performed in a block-wise batch learning style. Through experiments on automatic image annotation and image-based user retrieval in object retrieval tasks on real-world social media datasets, we found that our proposed method achieved good performance.

The contributions are summarized as follows.

- We propose a novel joint block matrix completion framework which exploits multiple object relationships and rich social context to perform information averaging within an appropriate POI-specific range.
- Our method is more scalable in relationship learning and completion compared to existing tag refinement, completion and prediction approaches [6] using only the image-tag relationship and recommendation models [8-9] and the user-item relationship.
- Our model may gain an improved computational efficiency since the model optimization can be easily parallelized.

## 2. Approach

### 2.1. Notations and Preliminaries

Given a social image collection with multiple data object types including manually labeled tags, GPS locations and their associated users, the problem we try to solve is how to complement the missing entries in the relation matrix and reduce the amount of noisy entries in three pairwise object relationship matrices: user-image, image-tag and user-tag. We use  $n$  to represent the amount of images provided by  $l$  users, and  $m$  to represent the number of unique tags in the dataset. The three original observed relationship matrices are denoted as:

1) User-Image Matrix:  $\hat{X} \in \{0,1\}^{l \times n}$  denotes the observed binary user-image matrix that encodes how users are associated with images (e.g., ownership, comment, retweet), where  $\hat{X}_{ri}$  is 1 if image  $i$  belongs to user  $r$ ; otherwise  $\hat{X}_{ri}$  is 0. After a simple permutation, the user-image matrix can be rewritten as:

$$\hat{X} = \begin{bmatrix} C_{l-s \times n-s} & 0 \\ 0 & I_s \end{bmatrix} \quad (1)$$

Where  $C \in \{0,1\}^{l-s \times n-s}$  is the observed binary matrix which indicates certain ownership between  $n-s$  images and  $l-s$  users.  $I_s$  is an identity matrix in correspondence with  $s$  images of anonymous users. Incorporating anonymous users in  $\hat{X}$  enables our method to deal with relationship learning on newly uploaded social images.

2) Image-Tag Matrix:  $T \in \{0,1\}^{n \times m}$ , where  $\hat{T}_{ij}$  is 1 if tag  $j$  is associated with image  $i$ ; otherwise  $\hat{T}_{ij}$  is 0. With  $\hat{T}$ , the tags of an image provide some weak labels on the diversified visual content.

3) User-Tag Matrix:  $\hat{U} \in R^{l \times m}$  is the matrix that encode the observed user-tag relation.  $\hat{U}_{rj} = \frac{1}{|r_n|} \sum_i \hat{T}_{ij}$ , if image  $i$  belongs to user  $r$  and  $|r_n|$  denotes the number of user  $r$ 's images. Each row of this matrix can be regarded as a normalized histogram of tags for a single user.

The goal of our study is to obtain  $X \in R^{l \times m}$ ,  $T \in R^{n \times m}$  and  $U \in R^{l \times m}$  by completing the original observed matrices  $\hat{X}, \hat{T}$  and  $\hat{U}$  jointly. The element  $X_{ri}$  in  $X$  reveals the probability of associating image  $i$  to user  $r$ , which represents the preference level of user  $r$  on image  $i$ . We use the feature output of the 5<sup>th</sup> convolution layer the deep CNN [10] trained on ILSVRC 2012 dataset to represent the visual characteristics of the images. We denote  $V \in R^{n \times d}$  as the visual feature matrix where the  $i^{\text{th}}$  row corresponds to the  $d$ -dim feature of image  $i$ .

## 2.2. POI Detection for Matrix Partition

A large number of images make the relationship matrices  $T, U$  and  $X$  very large, which is prohibitive to model them together. We have found that geographically adjacent images show similar visual content and semantic information more likely, which indicates the occurrence of a potential Point of Interests (POI). To detect POI from the dataset, we use mean shift on the GPS information (latitudes and longitudes) of social images. With POI detection, we partition all relationship matrices into a set of sub-matrices according to POIs. Then, we obtain the corresponding data block of each POI.

## 2.3. EMBC

We represent the index of POI by  $k$ . The notations with  $k$  refer to data block with respect to the  $k^{\text{th}}$  POI. We consider the following types of constraint terms for the three relationship matrices.

1) User Related Constraint: For all of the images uploaded by the same user in the same POI, they tend to be assigned the same tags. To formulate the user-wise coherence of users' tagging behavior in single POI, we measure the element-wise difference between  $U_k U_k^T$  and  $X_k T_k T_k^T X_k^T$  with  $F$ -norm. We define a user related loss term as  $\|F_k\|_F^2$ , where:

$$F_k = X_k T_k T_k^T X_k^T - U_k U_k^T \quad (2)$$

As shown in [11], two images sharing more common tags tend to have higher semantic similarity beyond tags. Besides user-wise similarity, the tag-wise similarity which reflects tag co-occurrence should also be consistent, since images attached to the same user share its tagging behavior in single POI. Accordingly, we minimize the constraint term  $\|H_k\|_F^2$  reflecting the tag-wise difference between images  $T_k^T T_k$  and users  $U_k^T U_k$  in  $F$ -norm as:

$$H_k = T_k^T T_k - U_k^T U_k \quad (3)$$

2) Tag Correlation Constraint: Similar as  $T_k^T T_k$ , the tag-wise similarity for the original image-tag sub-matrix is calculated as  $\hat{T}_k^T \hat{T}_k$ . To enforce the tag co-occurrence consistency structure on  $T_k$  before and after the model learning, we define the term to minimize the element-wise difference between  $T_k^T T_k$  and  $\hat{T}_k^T \hat{T}_k$ , denoted as  $\|K_k\|_F^2$ , where

$$K_k = T_k^T T_k - \hat{T}_k^T \hat{T}_k \quad (4)$$

3) Visual Content Related Constraint: The visual features and tags usually have intrinsic consistent relationship. Enforcing such consistency may be beneficial to image-tag matrix completion [6, 12]. To enhance the semantic consistency between visual features and tags, we penalize the difference between similarities in visual feature space and textual semantic label space with an  $F$ -norm  $\|T_k T_k^T - V_k V_k^T\|_F^2$ .

Considering the widely accepted fact that low-level visual features are less capable than tags for semantic labels of a given image, we introduce a feature mapping matrix  $W \in R^{d \times m}$  to reduce the semantic inconsistency between the two

matrices. It can directly map the visual feature into semantic tag space. Such a term can be rewritten as:

$$G_k = T_k T_k^\top - V_k W W^\top V_k^\top \quad (5)$$

4) Regularization: To avoid overly dense solution of the learned relation matrices, it is required that only a small number of  $T_k$  and  $U_k$  entries should be non-zero, *i.e.*, a small number of unique tags can be attached to each image or user. For  $X_k$ , we require a small number of images to be associated with given user. Inspired by existing sparse coding frameworks [13-14], we introduce an  $l_1$ -norm term  $\|T_k\|_1 + \|U_k\|_1 + \|X_k\|_1$  to derive the sparse solution of matrices  $T_k$ ,  $U_k$  and  $X_k$ . For the shared mapping matrix  $W$ , we also add an  $l_1$ -norm to achieve the sparse solution.

5) Overall Lost Function: By jointly considering all of these criteria, we formulate the whole problem as follows:

$$\arg \min_{X_k, T_k, U_k, W} \sum_{k=1}^p \mathcal{L}_k + \tau \|W\|_1 \quad (6)$$

$$\mathcal{L}_k = \alpha \|F_k\|_F^2 + \beta \|G_k\|_F^2 + \gamma \|H_k\|_F^2 + \lambda \|K_k\|_F^2 + \theta (\|T_k\|_1 + \|U_k\|_1 + \|X_k\|_1) \quad (7)$$

Where  $\alpha, \beta, \gamma, \lambda, \tau, \theta > 0$  are parameters and they can be tuned with cross-validation.

### 3. Optimization

We use  $X_k$ ,  $T_k$  and  $U_k$  for POI  $k$  as an example to show the technical details of our proposed algorithm. The optimization problem of Equation (6) is to alternately optimize each data object matrices. It is non-convex due to the non-quadratic terms in Equation (7). We apply sub-gradient descent for nonconvex optimization. During the optimization, it is likely to obtain dense intermediate values of  $X_k^t, T_k^t, U_k^t, k \in \{1, \dots, p\}$  if we directly apply alternating sub-gradient descent on the problem. This may lead to a significantly increase on the computational time cost at each iteration. To avoid this unexpected condition, we propose to decompose Equation (6) into two parts according to the composite function optimization in [6]. Accordingly, we formulate an auxiliary loss function:

$$A_k = \alpha \|F_k\|_F^2 + \beta \|G_k\|_F^2 + \gamma \|H_k\|_F^2 + \lambda \|K_k\|_F^2 \quad (8)$$

The sub-gradients of Equation (8) with respect to each relationship sub-matrix are formulated as:

$$\nabla_{X_k} A_k = 2\alpha F_k X_k T_k T_k^\top \quad (9)$$

$$\nabla_{T_k} A_k = 2\alpha X_k^\top F_k X_k T_k + 2\beta G_k T_k + 2\gamma T_k H_k + 2\lambda T_k K_k \quad (10)$$

$$\nabla_{U_k} A_k = 2\alpha F_k U_k + 2\gamma U_k H_k \quad (11)$$

The sub-gradients with regards to  $W$  are:

$$\nabla_W A_k = -2\beta (\sum_{k=1}^p V_k^\top G_k V_k) W \quad (12)$$

Based on the above setting, we compute intermediate solutions of each matrix from  $t$  to  $t+1$  step with respect to the auxiliary loss function as  $(S \in \{T_k, U_k, X_k, W\})$ :

$$\bar{S}^{t+1} = S^t - \delta_t \nabla_{S^t} A_k \quad (13)$$

$\delta_t$  denotes the step size. Then we formulate the auxiliary optimization problems for each model parameter matrix as follows ( $S \in \{T_k, U_k, X_k\}$  and  $\varphi = \theta$ , or  $S = W$  and  $\varphi = \tau$ ):

$$S^{t+1} = \arg \min_s \frac{1}{2} \|S - \bar{S}^{t+1}\|_F^2 + \varphi \delta_t \|S\| \quad (14)$$

We obtain the new updated solutions for auxiliary problems as:

$$S^{t+1} = \max(\mathbf{0}, \bar{S}^{t+1} - \varphi \delta_t) \quad (15)$$

Each variable in  $\{T_k, U_k, X_k, W\}$  is alternatively updated during the  $t^{\text{th}}$  iteration. Since the images and their related ontologies are divided into different POIs, we conduct the optimization procedure in parallel on each POI-specific data block. For  $T_k, U_k$  and  $X_k$  of POI  $k$ , to achieve correct calculation results and reduce the complexity,  $W$  is shared by all  $V_k$  in the parallel optimization process.

#### 4. Dataset Description and Experiment Settings

We evaluate the performance on image annotation and image-based user retrieval. The experiment selects two Flickr datasets, London and New York, to evaluate the performance on image annotation and image-based user retrieval. London contains 771,099 GPS located images, annotated by more than 100,000 tags uploaded by 16,225 users. New York contains 732,555 images with GPS information uploaded by 15,344 users. Due to the imbalance of the tag distribution between the images, most tags belong to only a few images [15]. The tags are ranked according to the number of annotated images, and the top 1000 are selected as experimental words. We construct both London and New York databases by geographic filtering of the social image data set YFCC100M.

As the geography expands, fix the bandwidth parameter  $h$  in Equation MS to 0.005 to achieve better POI detection [7]. In our experiments, we set the bandwidth parameter corresponding to 500 meters. The feature mapping matrix  $W$  is initialized as identity matrix. Divide 90% of the images in the data set as the training set, and the rest 10% as the test set to obtain  $T$  and  $X$ . We repeat the same experiment 5 times; use the average result as the final index to evaluate the performance of the algorithm. The cross-validation result determines that  $\alpha = 100$ ,  $\beta = 1$ ,  $\gamma = 1$ ,  $\lambda = 1$ ,  $\theta = 1$ ,  $\tau = 1$  and  $\delta_0 = 10^{-7}$ .

We evaluate the performance on image annotation and image-based user retrieval with the following relevant algorithms: TMC [6], LSR [12], TagProp [16], CMC [7], EBMC: our method without tag-wise constraint terms ( $\gamma = 0, \lambda = 0$  in Equation 7), NMF: the baseline method in IBUR [17].

##### 4.1. Image Annotation

Given a query image  $q$  in  $T$ , we rank all the tags in descending order of their probability values attached to image  $q$ . We change the number of initial training tags  $e$  from  $\{2, 4, 6, 8, 10\}$ . Suppose image  $i$  has  $m_i$  manually annotated tags, where  $m_i$  is the non-zero entries of the  $i^{\text{th}}$  row of  $\hat{T}$ . If  $e \geq m_i$ , select  $e$  tags as the initial annotation. If  $e > m_i$ , delete image  $i$  from the training set. We use the Mean Average Precision (MAP) of TOP  $K$  ( $K \in \{5, 20\}$ ) for a different number of completed tags (denoted as MAP@5 and MAP@20) to evaluate the algorithm performance.

Figure 1 shows some annotated results in the setting of  $e = 6$ . EBMC can predict tags with higher accuracy and more accurate meanings, while tags predicted by other approaches are usually wrong or with repeated meaning. As shown in Table 1, the accuracy goes up with the increase of the amount of initially observed tags for all compared methods. In fact, more observed tags for the image dataset provides richer relationship information. We observe that EBMC outperforms other methods because POI-based partition makes geographical local consistency more compact in tag space. The prior knowledge on user tagging behavior in a POI is effective for image annotation.

##### 4.2. Image-based User Retrieval

Image-based user retrieval is an application scenario when an online user with a query image searches possible users. Given the query  $q$ , we rank all the users in descending order according their similarity scores to image  $q$ , corresponding to the  $q^{\text{th}}$  column of  $X$ . For each query in the test set, we expect the position of its uploading user in the ranking list to be the top. Since  $\hat{X}$  contains an identity sub-matrix  $I_s$  for testing images attached to anonymous users, we set the corresponding

diagonal entries in  $X$  to zero. We use Mean Reciprocal Rank (MRR) as  $MRR = \frac{1}{|Q|} \sum_{q=1}^{|Q|} \frac{1}{Rank_q}$ , where  $Q$  is the testing set and  $Rank_q$  is the position of uploading user of query image  $q$  in the  $q^{\text{th}}$  column of  $X$ . A higher MRR score indicates better retrieval result, where the true original user is ranked at top positions. We record the maximum and average MRR score among all POIs.





Photos	Ground Truth	TMC	CMC	TagProp	EBMC
	london england united kingdom great britain river thames	london england uk united kingdom 2012	london england uk westminster united kingdom	london england uk westminster united kingdom	london england bridge River thames united kingdom
	london england uk westminster tower	london england uk united kingdom 2012	london england uk westminster parliament	london england uk westminster parliament	london england uk westminster tower
	london england united kingdom square trafalgar	london england uk united kingdom 2012	london england uk trafalgar square united kingdom	london england uk trafalgar square united kingdom	london england square trafalgar square united kingdom
	london england united kingdom river thames high building	london england uk united kingdom 2012	london england uk united kingdom tower bridge	london england uk tower bridge united kingdom	london england river thames high building United kingdom

Figure 1. Examples of annotation results. Tags in red denotes the wrongly predicted ones, those in green denotes tags with repeated meaning, and those in purple denotes novel but correct tags compared to the ground truth

Table 1. MAP performance of automatic image annotation

New York		MAP@5					
Method	NMF	TMC	LSR	TagProp	CMC	EBMC <sub>T</sub>	EBMC
$e = 2$	63.12	76.16	75.88	75.02	76.43	<b>77.75</b>	77.70
$e = 4$	65.24	76.81	76.99	75.73	77.25	78.33	<b>78.70</b>
$e = 6$	66.17	77.10	77.56	75.75	77.59	78.61	<b>78.67</b>
$e = 8$	66.98	77.34	78.21	76.78	77.71	78.81	<b>78.93</b>
$e = 10$	67.22	77.79	78.87	77.40	78.05	79.10	<b>79.49</b>
London		MAP@5					
Method	NMF	TMC	LSR	TagProp	CMC	EBMC <sub>T</sub>	EBMC
$e = 2$	72.44	81.36	78.57	78.90	84.58	89.06	<b>89.13</b>
$e = 4$	73.16	82.41	79.76	79.46	84.97	89.10	<b>89.24</b>
$e = 6$	74.03	82.44	81.07	79.57	85.21	88.98	<b>89.13</b>
$e = 8$	74.78	83.17	82.63	79.90	85.34	89.02	<b>89.36</b>
$e = 10$	75.29	84.37	84.48	80.23	85.68	89.09	<b>89.39</b>

Table 2 shows the Maximum MRR (MaxMRR) and Average MRR (AveMRR) scores of different methods on London and New York. All of the candidate methods except EBMC and EBMC<sub>T</sub> have the same scores as the baseline NMF, which means they have little improvement over baseline method on this task. At the same time, the Maximum MRR scores on both EBMC and EBMC<sub>T</sub> are eleven times larger than the baseline NMF method. This experimental phenomenon demonstrates that both  $F_k$  and  $G_k$  are indispensable in image-based user retrieval task. Both the visual contents and tags play important roles in sharing high-level semantic information with users.

**Findings and discussions.** In Table 2, we observe a significant difference between MRR scores of New York and London. Similar observation should also be noticed in image annotation experiment (Table 1), where the performance gap between EBMC and TMC in London is much larger than it in New York. It clearly reveals that the #image per user has positive correlation with the MRR scores when comparing two datasets. In the top 10 popular POIs, the uploaded images in New York are mostly contributed by a small number of active users. However, the uploaded images in the top 10 popular POIs in London are more scattered, and the active users tend to upload less images on the site of the top 10 POIs than the active users in New York. Based on this finding, we see that compared with baseline NMF techniques, the model capacity of our proposed method will be fully released to capture the relationships between heterogeneous data objects when the user-image relationship is

denser or more concentrated.

Table 2. MRR performance of image-based user retrieval (in %)

New York		MaxMRR				AveMRR			
Mthod	NMF	CMC	EBMCT	EBMC	NMF	CMC	EBMC <sub>T</sub>	EBMC	
$e = 2$	2.57	31.3	55.31	<b>55.45</b>	1.12	10.31	19.37	<b>19.77</b>	
$e = 4$	3.46	64.6	55.99	<b>56.02</b>	1.48	11.04	20.25	<b>20.36</b>	
$e = 6$	3.85	36.7	56.15	<b>56.20</b>	1.73	11.93	20.64	<b>20.83</b>	
$e = 8$	4.12	38.1	56.58	<b>56.78</b>	1.99	12.65	21.12	<b>21.43</b>	
$e = 10$	4.89	39.4	56.61	<b>56.89</b>	2.17	13.02	21.63	<b>21.97</b>	
London		MaxMRR				AveMRR			
Mthod	NMF	CMC	EBMCT	EBMC	NMF	CMC	EBMC <sub>T</sub>	EBMC	
$e = 2$	0.72	14.53	22.76	<b>22.83</b>	0.39	7.68	10.39	<b>11.64</b>	
$e = 4$	1.56	15.07	22.85	<b>26.04</b>	0.88	7.91	12.86	<b>13.40</b>	
$e = 6$	2.07	15.89	30.64	<b>31.41</b>	1.31	8.04	14.37	<b>15.57</b>	
$e = 8$	2.38	16.34	32.73	<b>35.79</b>	1.69	8.56	15.99	<b>16.61</b>	
$e = 10$	2.95	16.58	34.71	<b>35.79</b>	1.88	8.75	16.05	<b>17.76</b>	

### 4.3. Parameter Sensitivity Analysis

To evaluate the sensitivity of the parameters New York dataset, the experiment sets three significant parameters  $\{\alpha, \beta, \theta\}$  for reflecting the MAP5, MAP20, and the average MRR (AMRR) scores. Fix other parameters as the original value ( $\alpha = 100, \beta = 1, \theta = 1$ ) for each parameter;  $\alpha, \beta$  and  $\theta$  range from  $10^{-3}$  to  $10^3$ . In Figure 2(a), the MAP scores are stable in a wide range of  $\alpha$  except for  $\alpha = 1000$ , and the shape of the AMRR curve shows a trend from rise to decline. It can be seen that  $\alpha$  is sensible in IBUR. Thus, set 100 as the optimal value for  $\alpha$  to balance the performance of image annotation and image-based user retrieval. From Figure 2(b), the trend of average MRR score increases with the increment of  $\beta$ , and all three indicators tend to stabilize in the case of  $\beta \geq 1$ . There is a clear performance trade-off between the average MRR and MAP20 scores observed, which indicates that different values of  $\beta$  balance weights of different relationships among heterogenous objects. Thus, set  $\beta$  with an optimal value of 1. All of the curves in Figure 2(c) keep stable except a small ascent from  $\theta = 100$  to  $\theta = 1000$ . We determine 1 as the optimal value for  $\theta$  through experiments to make our model more flexible.

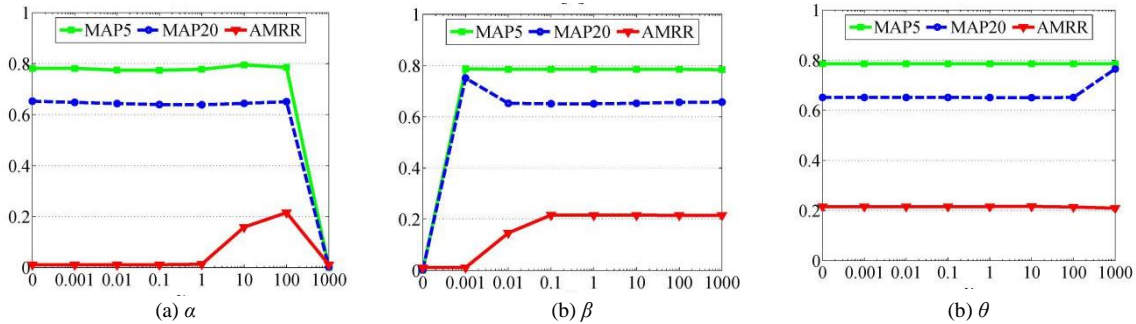


Figure 2. Sensitivity analysis of  $\alpha, \beta, \theta$  in terms of MAP@5, MAP@20 and AveMRR on the New York dataset

**Space Complexity.** The memory scale of TMC, LSR and NMF is  $O(n \times m)$ . TagProp relies on KN, so its memory consumption is  $O(n \times K)$ . The memory consumption of EBMC is  $O(n_{\max} \times m)$ , where  $n_{\max}$  is the number of images in the largest POI. Typically,  $m > 5K$  and  $n > 10n_{\max}$ . Our method achieves the smallest space requirement compared to others.

## 5. Conclusion

This paper proposes an efficient block matrix completion method, EBMC, which jointly factorizes the relationships among different ontologies in a distributed optimization manner. We decompose the whole learning problem into a set of POI-specific subtasks, which corresponding to the relation data blocks separated by the POI structure. Through experiments on automatic image annotation and image-based user retrieval in object retrieval tasks on real-world social media datasets, we found that our proposed method achieved good performance. In future work, we will study the connections between users and locations to form a more flexible relationship modeling framework.

## Acknowledgements

This paper was supported in part by the key scientific and technological project of Henan province (192102210298, 192102210097, 192102210126); and the Open Project Foundation of Information Technology Research Base of Civil Aviation Administration of China (NO. CAAC-ITRB-201607) and the Key scientific research projects of colleges and universities in Henan Province(18A520050).

## References

1. E. J. Candes and B. Recht, "Exact Matrix Completion via Convex Optimization," *Foundations of Computational Mathematics*, Vol. 9, No. 6, pp. 717, 2009
2. R. H. Keshavan, A. Montanari, and S. Oh, "Matrix Completion from a Few Entries," *IEEE Transactions on Information Theory*, Vol. 56, No. 6, pp. 2980-2998, 2010
3. A. Soni, T. Chevalier, and S. Jain, "Noisy Inductive Matrix Completion under Sparse Factor Models," in *Proceedings of 2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 2990-2994, 2017
4. H. Halpin, V. Robu, and H. Shepherd, "The Complex Dynamics of Collaborative Tagging," in *Proceedings of the 16th International Conference on World Wide Web*, pp. 211-220, 2007
5. D. Liu, S. C. Yan, X. S. Hua, and H. J. Zhang, "Image Retagging using Collaborative Tag Propagation," *Multimedia, IEEE Transactions on*, Vol. 13, No. 4, pp. 702-712, 2011
6. L. Wu, R. Jin, and A. K. Jain, "Tag Completion for Image Retrieval," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 35, No. 3, pp. 716-727, 2013
7. D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg, "Mapping the World's Photos," in *Proceedings of the 18th International Conference on World Wide Web*, pp. 761-770, 2009
8. J. Liu, Z. C. Li, J. H. Tang, Y. Jiang, and H. Q. Lu, "Personalized Geo-Specific Tag Recommendation for Photos on Social Websites," *IEEE Transactions on Multimedia*, Vol. 16, No. 3, pp. 588-600, 2014
9. O. Klopp, K. Lounici, and A. B. Tsybakov, "Robust Matrix Completion. Probability Theory and Related Fields," *Probability Theory and Related Fields*, Vol. 169, No. 1-2, pp. 523-564, 2017
10. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012
11. L. Wu, X. S. Hua, N. H. Yu, W. Y. Ma, and S. P. Li, "Flickr Distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 31-40, 2008
12. Z. J. Lin, G. G. Ding, M. Q. Hu, J. M. Wang, and X. J. Ye, "Image Tag Completion via Image-Specific and Tag-Specific Linear Sparse Reconstructions," in *Proceedings/CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1618-1625, 2013
13. J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online Learning for Matrix Factorization and Sparse Coding," *The Journal of Machine Learning Research*, No. 11, pp. 19-60, 2010
14. B. A. Olshausen and D. J. Field, "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by v1," *Vision Research*, Vol. 37, No. 23, pp. 3311-3325, 1997
15. L. Wu, J. Yang, N. H. Yu, and X. S. Hua, "Learning to Tag," in *Proceedings of the 18th International Conference on World Wide Web*, pp. 361-370, 2009
16. M. Guillaumin, T. Mensink, J. Verbeek, and C. Schmid, "Tagprop: Discriminative Metric Learning in Nearest Neighbor Models for Image Auto-Annotation," in *Proceedings of IEEE International Conference on Computer Vision*, pp. 309-316, 2009
17. H. F. Liu, Z. Wu, D. Cai, and T. S. Huang, "Constrained Nonnegative Matrix Factorization for Image Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 7, pp. 1299-1311, 2012