JEREMIE: Joint Semantic Feature Learning via Multi-relational Matrix Completion

Jiaming Zhang[1], Shuhui Wang[1(✉)], Qiang Qu[2], and Qingming Huang[1]

[1] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China
jiaming.zhang@vipl.ict.ac.cn, wangshuhui@ict.ac.cn, qmhuang@ucas.ac.cn

[2] The Global Center for Big Mobile Intelligence, Frontier Science and Technology Research Centre, Shenzhen Institutes of Advanced Technology, CAS, Shenzhen 518055, China
qiang@siat.ac.cn

Abstract. The relations among heterogeneous data objects (e.g., image, tag, user and geographical point-of-interest (POI)) on interactive Online Social Media (OSM) play an important information source in describing complicated connections among Web entities (users and POIs) and items (images). Jointly predicting multiple relations instead of single relation completion in separate tasks facilitates sufficient knowledge sharing among heterogeneous relations and mitigate the information imbalance among different tasks. In this paper, we propose JEREMIE, a **J**oint S**E**mantic Featu**R**e L**E**arning model via **M**ulti-relational Matr**I**x Compl**E**tion, which jointly complements the semantic features of different entities from heterogeneous domains. Specifically, to perform appropriate information averaging, we first divide the social image collection into data blocks according to the affiliated user and POI information, where POIs are detected by mean shift from the GPS information. Then we develop a block-wise batch learning method which jointly learns the semantic features (e.g., *image-tag*, *POI-tag* and *user-tag* relations) by optimizing a transductive matrix completion framework with structure preservation and appropriate information averaging functionality. Experimental results on automatic image annotation, image-based user retrieval and image-based POI retrieval demonstrate that our approach achieves promising performance in various relation prediction tasks on six city-scale OSM datasets.

Keywords: Multi-relational learning · Semantic feature
Matrix completion · Information retrieval

1 Introduction

Multiple relations among heterogeneous data objects (e.g., image, tag, user and point-of-interest) on interactive Online Social Media (OSM) play an important role in describing complicated connections among entities (user and point-of-interest) in real world and items (image) on the Internet. In this context, diversified information acquisition needs have been evoked by the users for enjoying

C. Doulkeridis et al. (Eds.): MATES 2017, LNCS 10731, pp. 87–108, 2018.
https://doi.org/10.1007/978-3-319-73521-4_6

better daily life. For example, users tend to find potential friends with similar preference on the image content and semantics. They may be pleased if the recommended points-of-interest (POIs) contain the locations actually meet their purposes. Therefore, it has become more and more emergent to develop effective and efficient methodologies for heterogeneous data object retrieval.

In this paper, we prescribe the heterogeneous data object retrieval problem into several relation prediction tasks, e.g., automatic image annotation (*image-tag*), image-based user retrieval (*image-user*), image-based POI retrieval (*image-POI*), etc. Traditionally, the statistical dependency among certain types of data objects are modeled by specific learning tasks. For example, to obtain a sufficient number of high-quality tags for social images based on existing manually labeled tags, tag refinement [12,14,30], tag completion [13,25,29] and multi-label learning [8] are widely studied to deal with the absence and noise in the collected social tags [6]. Recommendation models are employed to discover the relation between users and content [16,19,24]. Then for POI recommendation, current works [3,27,28] mainly focus on leveraging the geographical and social properties to improve recommendation effectiveness. The correlation among heterogeneous objects on OSM is uniformly modeled with multi-layered graphs [15]. All of these works either focus on a specific single relational learning task or use a unified graph structure to represent heterogeneous objects and their correlation.

In fact, entities like users and POIs or web items like images are all associated with tags with rich semantics. This common semantic attribute makes it possible to facilitate the knowledge sharing and propagation by jointly modeling the relation and interaction between heterogeneous data objects (image, user and POI) and semantic tags. Specifically, we regard the tag-oriented relation *X-tag* (X can be user, POI or image) as the semantic feature of each data object, which reflects the users' interest in the tagging activities for *user-tag* and POI properties for *POI-tag*, visual content for *image-tag* respectively. The semantic feature of a given entity can be regarded as a representation in the shared semantic space. In addition, each image has a visual feature to represent the visual content, and the content similarity is regarded as the core structure to determine how labels are propagated in standard image tagging models.

Besides the semantic feature, there exist rich attributes among OSM images. For example, each image is uploaded by a single user, while a user may have uploaded multiple images. Each image belongs to a detected POI, while a POI may have multiple images within its geographical range. The social context existing in the attributes of heterogeneous objects can be adopted to deploy the multi-relational learning tasks. For example, the image location may be strongly related to the geographical tags, and may also have direct implication on the scene and object in the image [5,16,18]. From the users' perspective, users with similar social attributes tend to co-occur in the same places, and they are more likely to be interested in the same type of images [25]. Since the rich attributes describe natural cluster structure of the OSM images, we divide the OSM image dataset into a set of relation data blocks according to the user structure and POI structure. Then the learning process of the *X-tag* relation can be performed by cluster-specific block matrix completion and factorization, which are actually information averaging mechanisms by latent factor modeling among the given relation sub-matrices within an appropriate POI-specific or user-specific range. Furthermore, by constructing block-based relation learning mechanism, the computational efficiency can be enhanced, especially the reduced storage consumption. Consequently, the multi-relational matrix completion problem on large scale OSM data can be solved with parallel optimization.

In our research context, the semantic feature of each entity is noisy and sparsely observed due to the complicated user behaviors [7], which brings great challenges in quantifying the true relations among heterogeneous data objects. Therefore, considering the intrinsic sharing nature among these relations, we jointly learn the semantic features from their corresponding noisy and sparse observation in a multi-relational matrix completion framework. Our goal is to learn representative semantic features for user, POI and image, regularized by visual similarity among images and the relation data block structure. The *image-user* and *image-POI* relation can be reconstructed based on the learned model by calculating the inner product on the corresponding learned semantic features, which is similar in spirit as the reconstruction procedure in classical matrix factorization approaches. Based on the reconstructed relations, the *image-user* and *image-POI* retrieval can be performed by relation score ranking.

Accordingly, our framework consists of two steps, one is the preprocessing step dealing with the data partition by users or POIs, and the second step is the semantic feature learning step where the semantic features are learned by the proposed multi-relational matrix completion model. In the preprocessing step, we individually separate images into data blocks with respect to users and POIs, respectively. Then the whole learning procedure can be formulated into a set of sub-tasks. In the learning step, the unified objective loss function contains two main modules, i.e., a structure preservation module using grouped affiliation relations and an information averaging module using visual feature of images. The structure preservation module penalizes the difference defined by the entity similarity calculated by the semantic feature between user and image with the assistance of the cluster structure. The information averaging module adopts the same formulation with the structure preservation module. However, the information averaging module employs both the visual feature and the semantic feature of image. Consequently, the relations among heterogeneous objects can be more reliably learned than other relational learning models [14,21,25]. Moreover, our method better reflects the intrinsic nature of heterogeneous object correlations than unified graph-based model [15], thus the relations can be learned more accurately.

The contribution of our framework is listed as below:

– We propose the JEREMIE approach, a novel **J**oint s**E**mantic featu**R**e l**E**arning fra**M**ework via **M**ulti-relational Matr**I**x compl**E**tion which involves both mapping relationship for structure preservation and intrinsic feature for information averaging.

– The data partition step builds a localized neighborhood environment for more compact semantic feature learning, which is much closer to the intrinsic data distribution for information averaging within local neighbors compared with other irrelevant samples.
– The better completion results of semantic features of images promote that the learned semantic features of users and POIs to have better representation for their properties in return, which indicates better performance of the reconstruction of the *image-user* and *image-POI* relation.
– The relation structure in our proposed framework makes better use of rich social context among multi-relational domains than unified graph-based models and collective learning algorithms.

In summary, by combining nonnegative matrix factorization and transductive matrix completion approaches, we propose JEREMIE to jointly complements the semantic features of different entities from heterogeneous domains. An efficient alternating sub-gradient descent algorithm can be adopted to simultaneously learn the POI-specific and user-specific relation matrix blocks in an alternating optimization style. Experiments on real-world OSM datasets demonstrate the promising power of our approach. The framework is illustrated in Fig. 1.

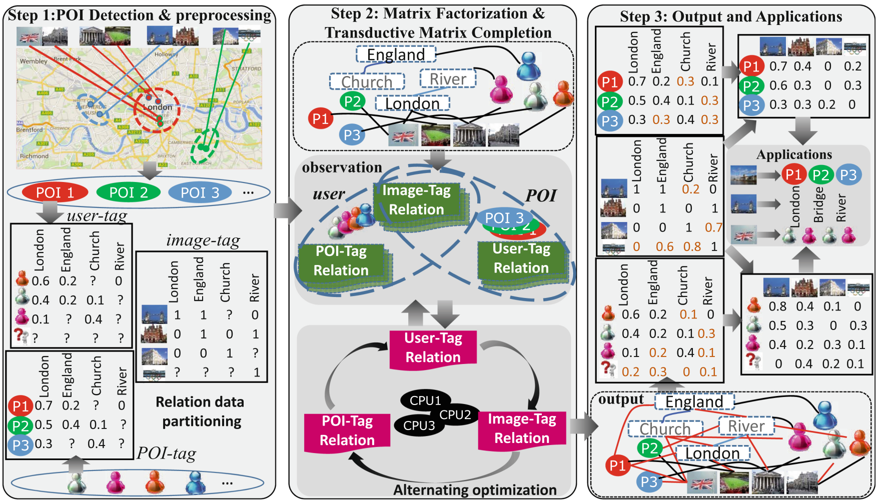


Fig. 1. Framework of our proposed method JEREMIE (best viewed in color). JEREMIE consists of two steps. The first step is the data partition preprocessing step dealing with GPS locations (by Mean Shift procedure) and users of images. The second step is the proposed block-wise multi-relational matrix completion model for joint semantic feature learning. (Color figure online)

2 Related Work

In this section, we first briefly survey the existing approaches from the perspective of both the application (automatic image annotation, attribute-based social media analysis) and the multi-relational learning theory, and motivate our study for designing multi-relational matrix completion method.

2.1 Automatic Image Annotation

Besides general content-based automatic image annotation techniques [20], there are also many recent works exploiting multi-label learning techniques [8] to deal with image annotation problem.

Several researchers solve the image annotation problem by image retagging, tag recommendation, tag propagation, etc. Li et al. [12] propose a neighbor voting method for social tagging. Guillaumin et al. [6] propose a tag propagation (TagProp) method to transfer tags through a weighted nearest neighbor graph.

All the approaches mentioned above usually need high-quality annotated tags and good class assignments for confident model training, contrary to the reality that manually annotated tags contain many incorrect and noisy ones. Furthermore, matrix completion techniques are also introduced to address poor initial image annotation problem [13,25], getting convincible experiment results. More recently, Wu et al. [25] build a concise tag matrix completion computational framework based on both semantic-content and tag co-occurrence consistency. Motivated by this work, we adopt matrix completion in our proposed joint semantic feature learning framework.

2.2 Attribute-Based Social Media Analysis

In social media analysis, many algorithms for social user recommendation have been proposed in recent years. Combining users and visual content, social recommendation models are studied to discover the connection between users and content [19,24]. From the perspective of correlation among heterogeneous objects in OSM, graph-based approaches are utilized to uniformly model the recommendation problems [15]. These correlation based recommendation algorithms usually build on global relationship, which will inevitably encounter computational complexity problems as the scalability goes up.

Besides numerous research works on social media user analysis, there are many research works focusing on combining geographical attributes and visual content. To analyze large scale online image collections with both geographical and visual content information, Crandall et al. [5] formularize the image location estimation as a classification problem by classifying images into POI categories. Moxley et al. [18] adopt geographical based search strategy to provide candidate tags and images which are similar in visual content. Liu et al. [16] propose a unified framework using subspace learning in personalized and geo-specific tag recommendation for images on OSM.

2.3 Multi-relational Learning

User-item clustering or matrix factorization techniques in collaborative filtering have been successfully utilized to represent non-trivial similarities between the relation patterns of entities in single relational data. Easily extended to multi-relational learning, several clustering based approaches have been proposed. Sutskever et al. [22] use a non-parametric Bayesian clustering of entities embedding in a collective matrix factorization formulation. Zhu et al. [31] refine the clustering setting to allow entities to have a mixed clusters membership. To share parameters between relations, [22,31] build models that cluster not only entities but relations as well.

Similarly, as pointed out by [9], a lot of methods for multi-relational data have been designed within the framework of relational learning from latent attributes, which learn latent embeddings of the entities by operating on relations through factorization (matrix or tensor) framework. The fact that relations can be similar or highly related suggests that a superposition of independently learned models for each relation would be highly inefficient especially since the relationships observed for each relation are extremely sparse. The basic idea of collectively learning several relations simultaneously is to share the common embedding via collective matrix factorization (CMF) [21]. The insight we model pairwise relations is similar to CMF [21]. However we adopt transductive matrix completion instead of CMF in a similarity propagation manner, which is quite different from CMF.

3 Notations and Preliminaries

We denote n as the number of images uploaded by u users, m as the number of unique tags in the dataset and p as the number of clustered POIs. Then the two types of relation matrices are denoted as:

Observed Image Semantic Feature: $\widehat{T} \in \{0,1\}^{n\times m}$ is a binary matrix, where $\widehat{T}_{st}$ is set to 1 if tag t is assigned to image s and otherwise 0. $\widehat{T}$ represents the tagging behavior on each image, while the tags of an image provide some weak annotations on the diversified visual content.

Observed User Semantic Feature: $\widehat{U} \in \mathbb{R}^{u\times m}$ is the observed user-tag matrix, where $\widehat{U}_{kt} = \frac{1}{|k_n|}\sum_s \widehat{T}_{st}$ and $|k_n|$ denotes the number of user k's images. It can be considered as a normalized histogram of tags for users.

Observed POI Semantic Feature: $\widehat{P} \in \mathbb{R}^{p\times m}$ is the observed POI-tag matrix, where $\widehat{P}_{lt} = \frac{1}{|l_n|}\sum_s \widehat{T}_{st}$ and $|l_n|$ denotes the number of images in the l-th POI. Similarly, it can be considered as a normalized histogram of tags for POIs.

Image-User Relation Matrix: $R \in \{0,1\}^{u\times n}$ is the binary user-image matrix that represents the affiliation between users and images, where R_{ks} is 1 if image s belongs to user k and otherwise 0.

Image-POI Relation Matrix: $S \in \{0,1\}^{p \times n}$ is the binary POI-image matrix that represents the aggregation distribution between POIs and images, where S_{ls} is 1 if image s belongs to POI l and otherwise 0.

The goal of our proposed framework is to obtain $\widetilde{R}$, $\widetilde{S}$, T, U and P by completing the three corresponding original observed matrices $\widehat{T}, \widehat{U}$ and $\widehat{P}$, where $\widetilde{R} = UT^\top$ and $\widetilde{S} = PT^\top$. Each entry T_{st} in completed image-tag matrix T indicates the probability of assigning tag t to image s, with U_{kt} in U the probability of attaching tag t to user k and P_{lt} in P the probability of attaching tag t to POI l respectively.

For the representation of the visual content of images, we utilize the feature (the output of the fifth layer) of the deep convolutional neural networks (CNN) [10] trained on ILSVRC 2012 dataset. We denote the visual feature matrix as $V \in \mathbb{R}^{n \times d}$ where the i-th row corresponds to a d-dimension visual feature of image i.

4 Approach

First we introduce the data partition including the POI detection procedure as the preprocessing step of our algorithm. Then we provide details about our multi-relational matrix completion framework.

4.1 POI Clustering via Mean Shift

The points-of-interest (POIs) that we refer to in this paper is different from the definition based on users' check-in actions [28]. As discussed in previous section, geographically adjacent images may have similar visual content or semantic information with higher probability, which indicates potential POIs. Thus we apply a density-based clustering method Mean Shift [4] to detect POIs from the GPS information (latitudes and longitudes) of social images. With the POI detection results, each image obtains a new aggregated attribute POI. Then we can partition T into a set of sub-matrices according to POIs. After that we can build index for images based on the *image-POI* relation.

4.2 Data Partition with Users and POIs

Without loss of generality, we refer to the index of user as i and POI as j respectively for the interpretation of the data partition procedure. Within the two affiliation relations, *image-User* and clustered *image-POI*, we individually separate images into user blocks and POI blocks respectively. That means, the images which belong to user i (POI j) will be assigned to data block i (j). Then the relation matrices T, U and R will be partitioned into corresponding sub-matrices T_i, U_i and $R_i (i \in \{1, 2, \cdots, u\})$ within user blocks, with T_j, P_j and $S_j (j \in \{1, 2, \cdots, p\})$ for T, P and S within POI blocks respectively.

Algorithm 1. Optimization procedure of **JEREMIE**

Input: Observed Relation Matrices: $R_i, \widehat{T}_i, \widehat{U}_i, S_j, \widehat{T}_j, \widehat{P}_j$
Visual Feature Matrices: V_i, V_j, $i \in \{1, \ldots, u\}, j \in \{1, \ldots, p\}$
Output: Completed Relation Matrices: $\widetilde{R}, \widetilde{S}, T^t, U^t, P^t$;
Initial: $W_U^1 = W_P^1 = I_{d,m}, T_i^1 = \widehat{T}_i^1, U_i^1 = \widehat{U}_i^1, T_j^1 = \widehat{T}_j^1, P_j^1 = \widehat{P}_j^1, t = 1$;
while $\|\mathcal{L}^{t+1} - \mathcal{L}^t\| \geq \epsilon \mathcal{L}^t$ **do**
  Step size $\delta_t = \delta_0 / t$;//**The loops below are both executed in block-wise manner**
  for $i = 1$ to u **do**
    Calculate $\overline{T}_i^{t+1}, \overline{U}_i^{t+1}, \overline{W_U}^{t+1}$: Eq. 14
    Update and Concatenate T_i^{t+1}, U_i^{t+1}: Eq. 16
  end for
  for $j = 1$ to p **do**
    Calculate $\overline{T}_j^{t+1}, \overline{P}_j^{t+1}, \overline{W_P}^{t+1}$: Eq. 14
    Update and Concatenate T_j^{t+1}, P_j^{t+1}: Eq. 16
  end for
  Update W_U^{t+1} and W_P^{t+1}: Eq. 16
  t = t + 1;
end while
$\widetilde{R} = U^t T^{t\top}, \widetilde{S} = P^t T^{t\top}, T^t = \min\left(T^{Pt}, T^{Ut}\right)$
return $\widetilde{R}, \widetilde{S}, T^t, U^t, P^t$

4.3 JEREMIE Algorithm

Without loss of generality, we take user as the example to describe the details of our framework, and directly give the symmetric formulation for POI respectively.

Basic NMF Framework. Firstly proposed in [11], NMF is a commonly used matrix factorization technique which require the elements both in original matrix and factor matrices must be nonnegative. In our framework, we decompose the relation matrix R_i into two nonnegative latent factor matrices U_i and T_i aiming to find the semantics supervised reconstruction of R_i from tag space, represented as $R_i \approx U_i V_i^\top$. Usually, the approximation is quantified by a cost function constructed by the Frobenius norm denoted as $\|E_i\|_F^2$, where:

$$E_i = R_i - U_i T_i^\top. \tag{1}$$

Similarly, we obtain the formulation for POI denoted as $\|E_j\|_F^2$, where:

$$E_j = S_j - P_j T_j^\top. \tag{2}$$

In our framework, we force the $k = m$ to ensure the latent factor matrices are isomorphic with the tag matrices.

Modeling Tagging Preference via Structure Preservation. For all of the images uploaded by the same user in the same POI, they tend to assign the same tagging results from themselves and other users. The k-th row of U_i depicts the actual tag distribution of user k. Similarly, the k-th row in the product R_iT_i reveals the weighted sum of image tag distribution for the images related to user k. So R_iT_i can be regarded as a refined estimation for U_i using the group of images for each user. To formulate the tagging preference via user interest, we calculate the difference between the user-wise similarity $U_iU_i^\top$ and its estimation $R_iT_iT_i^\top R_i^\top$ with Frobenius norm. Then we define the structure preservation constraint term as $\|F_i\|_F^2$, where:

$$F_i = R_iT_iT_i^\top R_i^\top - U_iU_i^\top. \tag{3}$$

Similarly, we have the structure preservation constraint term for POI denoted as $\|F_j\|_F^2$, where:

$$F_j = S_jT_jT_j^\top S_j^\top - P_jP_j^\top. \tag{4}$$

Modeling Tagging Preference via Information Averaging. The visual content and annotated tags usually have intrinsic semantic relation. It makes great improvement to utilize visual information in image-tag matrix completion [13,25]. To enhance the semantic coherence between visual content and tags, we penalize the difference of similarities in visual feature space and textual semantic space with a Frobenius norm denoted as $\|T_iT_i^\top - V_iV_i^\top\|_F^2$.

However, low level visual features are less capable than tags for semantic representation of a given image. To reduce the semantic gap, we introduce a feature mapping matrix $W_U \in \mathbb{R}^{d\times m}$, which can directly map the visual feature into textual semantic space. Then the visual constraint term can be rewritten as $\|G\|_F^2$, where:

$$G_i = T_iT_i^\top - V_iW_UW_U^\top V_i^\top. \tag{5}$$

Similarly, we obtain the visual constraint term for POI denoted as $\|G_j\|_F^2$, where:

$$G_j = T_jT_j^\top - V_jW_PW_P^\top V_j^\top. \tag{6}$$

Regularization Terms. To avoid dense solution of semantic feature matrices, we require that only a small number of entries of T_i, T_j, U_i and P_j are nonzero, *i.e.*, a small number of unique tags are attached to each object (image, user or POI). As in many sparse coding literatures [17], we introduce two ℓ_1-norm regularization terms $\|T_i\|_1 + \|U_i\|_1$ to pursue a sparse solution of T_i and U_i, with $\|T_j\|_1 + \|P_j\|_1$ for POI respectively. For the shared mapping matrices W_U and W_P, we also add ℓ_1-norm regularization terms to enforce the sparsity.

Unified Loss Function. Finally, with respect to all of these criteria, we formulate our optimization problem as follows:

$$\operatorname*{arg\,min}_{T_i,U_i,W_U,T_j,P_j,W_P} \sum\nolimits_{i=1}^{u} \mathcal{L}_i + \sum\nolimits_{j=1}^{p} \mathcal{L}_j + \tau\left(\|W_U\|_1 + \|W_P\|_1\right) \tag{7}$$

$$\mathcal{L}_i = \|E_i\|_F^2 + \alpha\|F_i\|_F^2 + \beta\|G_i\|_F^2 + \theta(\|T_i\|_1 + \|U_i\|_1) \tag{8}$$

$$\mathcal{L}_j = \|E_j\|_F^2 + \alpha\|F_j\|_F^2 + \beta\|G_j\|_F^2 + \theta(\|T_j\|_1 + \|P_j\|_1) \tag{9}$$

where $\alpha, \beta, \theta, \tau > 0$ are parameters whose values can be easily tuned in the cross-validation procedure.

4.4 Optimization

Without loss of generality, we take T_i and U_i within user i and W_U as an example to describe the details of our algorithm. According to symmetry, the same optimization strategy are applied to T_j and P_j within POI j and W_P respectively. As we can see, the objective function is non-convex because of several non-quadric regularization terms. Accordingly, we design an alternating optimization procedure using subgradient descent method to solve the problem on large scale OSM datasets.

On the other hand, we may get dense intermediate solutions $T_i^t, U_i^t, i \in \{1, \ldots, u\}$ if we directly use the alternating subgradient descent approach to solve the original problem. It will significantly increase the computational time cost in each iteration. To avoid this ill-posed condition, we decompose the objective function into two parts according to the composite function optimization method [1]. In particular, we construct an auxiliary function as below:

$$A_i = \|E_i\|_F^2 + \alpha\|F_i\|_F^2 + \beta\|G_i\|_F^2. \tag{10}$$

The subgradients of the auxiliary function with respect to each relation submatrix are:

$$\nabla_{T_i} A_i = 2E_i^\top U_i + 2\alpha R_i^\top F_i R_i T_i + 2\beta G_i T_i \tag{11}$$

$$\nabla_{U_i} A_i = 2E_i T_i + 2\alpha F_i U_i. \tag{12}$$

And the subgradients with respect to W are:

$$\nabla_{W_U} A_i = -2\beta(\sum\nolimits_{i=1}^{u} V_i^\top G_i V_i) W_U. \tag{13}$$

Then we can compute the intermediate optimal solutions of each objective matrix from iteration t to $t+1$ referring to the auxiliary function by ($Q \in \{T_i, U_i, W_U\}$):

$$\overline{Q}^{t+1} = Q^t - \delta_t \nabla_{Q^t} A_i, \tag{14}$$

where δ_t is the step size. Based on these intermediate solutions and the rest of nonconvex regularization terms, we formulate the auxiliary optimization problems as follows ($Q \in \{T_i, U_i\}$ and $\varphi = \theta$, or $Q = W_U$ and $\varphi = \tau$):

$$Q^{t+1} = \arg\min_{Q} \frac{1}{2}\|Q - \overline{Q}^{t+1}\|_F^2 + \varphi\delta_t\|Q\|_1 \tag{15}$$

Combined with the intermediate solutions, we obtain the new solutions for auxiliary problems as follows:

$$Q^{t+1} = \max(\mathbf{0}, \overline{Q}^{t+1} - \varphi\delta_t). \tag{16}$$

Each variable in $\{T_i, U_i, W_U\}$ is alternatively updated during the t-th iteration.

Block-Wise Batch Processing. Here we discuss the whole optimization procedure in all of the user blocks, with the same strategy applying to POIs respectively. For T_i and U_i of user i, its calculation progress is independent from the corresponding matrices of other users. In order to achieve correct calculation results and reduce the complexity of our model, we make W_U shared by all of the sub-matrices V_i in parallel computation, with W_P for V_j respectively. Algorithm 1 illustrates the main steps in our solution for the optimization problem.

5 Experimental Evaluation

We evaluate the performance of our model **JEREMIE** on three object retrieval tasks: automatic image annotation (AIA), image-based user retrieval (IBUR) and image-base POI retrieval (IBPR).

5.1 Dataset and Experiment Settings

We use two Flickr datasets which are named after several international metropolises attracting crowds of tourists to conduct the experiments for AIA, IBUR and IBPR. All of these datasets are built by giving a geographical filtration to a large scale social image dataset YFCC100M[1] [23] published by Yahoo Web Lab[2].

As studied in [26], the tag distribution among images is extremely imbalanced and the majority of tags belong to a few images. Then we rank the tags according to its number of annotated images and select top 1000 to serve as the vocabulary in experiment. Table 1 shows some statistics of these datasets.

Users' tagging behavior in specific POI may become uncertain as the geographical scope goes larger. So we fix the bandwidth parameter in Mean Shift as 0.005 empirically to achieve '*landmark-scale*' POI detection, as defined in [5]. This bandwidth parameter is in correspondence with 500 meters as the maximum geographical radius of the POIs detected in our experiments.

Table 1. Statistics of the datasets used in the experiments

Dataset	Image		User		Tag				
	#Total	#GPS	#User	#Image per user	#Tag	#Tag per image		#Image per tag	
						Mean	Max	Mean	Max
London	1,338,388	771,099	16,225	47.52	1,000	5.2	63	4,016.4	481,957
New York	1,210,094	732,555	15,344	47.74	1,000	5.5	71	4,031.5	299,867
Paris	532,562	425,282	10,510	40.46	1,000	4.7	45	2,012.5	342,987
Beijing	146,541	79,140	2,016	39.26	1,000	4.7	38	370.6	57,232
Shanghai	123,436	68,964	1,701	40.54	1,000	4.7	51	322.6	49,147

[1] https://multimediacommons.wordpress.com/.
[2] http://webscope.sandbox.yahoo.com/.

We utilize identity matrix to initialize the feature mapping matrices W_U and W_P. Then we randomly divide each dataset into training set with 90% images and testing set with the rest 10%, respectively. As we obtain T, $\widetilde{R}$ and $\widetilde{S}$ simultaneously, we share the same training set and testing set for all of the three tasks. We repeat the experiment five times and report the average performance, while each run adopts a new separation of the datasets. After the cross-validation procedure, we determine that $\alpha = 1000, \beta = 10$, $\theta = 1$ and $\tau = 1$. For TMC and other methods, we use the optimal parameter settings reported in related literature. The initial step size δ_0 is set as 10^{-7} empirically.

5.2 Compared Methods

Our proposed framework is evaluated on the three tasks with the following relevant algorithms:

- **FastTag** (**Fast**): Baseline tagging method [2].
- **TagProp** (**Prop**): Label propagation on weighted nearest neighbor graph [6].
- **TagRel** (**Rel**): Tagging based on neighbor voting [12].
- **TMC**: Tag completion using content-semantic consistency and tag co-occurrence consistency [25].
- **PTC**: Tagging among POIs with users' tagging behavior prior [29].
- **CMF**: Multi-relational Learning method using collective matrix factorization [21].
- **JEREMIE**: our method.
- **JEREMIE$_\mathbf{U}$** (**J-U**): our method **only with user-related** constraint terms (with $\mathcal{L}_j$ removed in Eq. 7).
- **JEREMIE$_\mathbf{P}$** (**J-P**): our method **only with POI-related** constraint terms (with $\mathcal{L}_i$ removed in Eq. 7).
- **JEREMIE$_\mathbf{F}$** (**J-F**): our method **without visual content related** constraint terms (corresponding to $\beta = 0$ in Eqs. 8 and 9).
- **JEREMIE$_\mathbf{G}$** (**J-G**): our method **without structure related** constraint terms (corresponding to $\alpha = 0$ in Eqs. 8 and 9).

5.3 Automatic Image Annotation

The task of automatic image annotation is to automatically associate unlabelled images with semantically related tags. Given a query image q, we simply rank all the tags in descending order of their probability scores attached to image q, corresponding to the q-th row in T. Since we obtain T^U from user and T^P from POI simultaneously, we adopt the minimum pooling strategy

$$T_{st} = \min\left(T^P_{st}, T^U_{st}\right), \tag{17}$$

for the final tagging results to avoid very small positive completed entries (e.g. 1×10^{-8}). In particular, to test the robustness of our proposed method to the number of initial tags allocated to each image, we vary the number of initial training tags (denoted as e) for each image from $\{1, 2, 3, 4, 5\}$. Without loss of generality, we suppose image i has m_i manually annotated tags, corresponding to m_i non-zero entries in the i-th row of $\widehat{T}$ in training set. If $e \leq m_i$, we randomly select e tags as the initial annotation for image i. Otherwise if $e > m_i$, we remove image i from the training set. We use the Precision at TOP 20 (denoted as Precision@20) and Recall at TOP 20 (denoted as Recall@20) to measure the performance of different algorithms.

As shown in Tables 2 and 3, the annotation accuracy goes up with the increase of the number of initially observed tags in general for all methods. It can be explained by the fact that more observed tags for each image provides richer semantic information for knowledge propagation.

We also observe that all of our proposed methods outperform other compared methods. It provides evidence for that the matrix partition strategy makes both geographical locality and user interest consistency more compact in tag semantic space. Compared with state-of-the-art tagging methods (FastTag, TMC and TagProp), our algorithm utilizes auxiliary information such as the user-image and POI-image relations. The experiment results are dominated by the local consistency of tag distribution after matrix partition instead of the coherence between visual content and semantics.

5.4 Image-Based User and POI Retrieval

Closely related to friend recommendation, IBUR is a typical application scenario when a user of OSM wants to search possible users who has taken the query image, with IBPR searching for potential POIs respectively. Given a query image q, we simply rank all the known users in descending order of their relationship intensity scores to image q, corresponding to the q-th column in R and S for POI respectively. For each query image in the testing set, we expect the position of its original user or POI in the ranking list to be as high as possible. Basic settings such as initial training tags are the same with AIA, as our framework simultaneously outputs three *X-tag* relation matrices. In particular, we take Mean Reciprocal Rank (MRR) as performance measurement for different methods, which is defined as:

$$MRR = \frac{1}{|Q|} \sum\nolimits_{q=1}^{|Q|} \frac{1}{Rank_q}, \tag{18}$$

where Q is the testing set and $Rank_q$ is the position of original user or POI of testing image q in the ranked q-th column of R or S.

Table 4 shows the MRR scores of different methods on IBUR and IBPR. Since the two tasks are relatively independent of each other within the unified framework **JEREMIE**, we select the experiment results from corresponding branch, *i.e.* **J-U** for IBUR and **J-P** for IBPR respectively. We can see that the MRR scores of both **J-U** and **J-G** are even times larger than the baseline Matrix Factorization methods. While compared with general multi-relational learning method CMF, our algorithm adopts similarity consistency instead of

Table 2. Precision@20 of image tag recommendation

London	Precision@20						
Method	FastTag	TagRel	TagProp	CMF	TMC	PTC	JEREMIE
e = 1	16.67	16.26	16.92	19.43	26.16	35.19	**40.40**
e = 2	17.52	16.54	17.05	19.97	27.28	37.12	**42.26**
e = 3	17.37	17.07	17.21	19.29	27.63	37.39	**43.12**
e = 4	17.79	17.46	18.32	19.83	28.34	38.24	**43.50**
e = 5	18.61	17.73	18.62	20.19	28.05	38.38	**43.78**
New York	Precision@20						
Method	FastTag	TagRel	TagProp	CMF	TMC	PTC	JEREMIE
e = 1	13.13	12.67	15.51	16.77	17.93	23.79	**37.23**
e = 2	13.43	12.98	15.53	16.27	18.36	25.32	**38.14**
e = 3	13.88	13.15	15.90	16.83	18.38	26.14	**39.41**
e = 4	14.26	13.23	15.77	16.61	18.37	26.94	**39.87**
e = 5	14.68	13.62	15.96	17.02	18.40	27.06	**40.43**
Paris	Precision@20						
Method	FastTag	TagRel	TagProp	CMF	TMC	PTC	JEREMIE
e = 1	13.26	13.81	14.74	16.79	18.07	24.63	**33.68**
e = 2	13.63	12.90	14.61	16.12	18.72	25.56	**34.95**
e = 3	13.81	13.19	14.82	16.33	18.72	25.90	**35.51**
e = 4	14.36	13.57	15.05	16.64	18.89	25.90	**35.67**
e = 5	14.48	14.23	15.07	16.98	18.85	26.23	**35.91**
Beijing	Precision@20						
Method	FastTag	TagRel	TagProp	CMF	TMC	PTC	JEREMIE
e = 1	11.67	12.61	12.16	15.43	15.79	27.23	**33.34**
e = 2	12.15	13.21	13.74	15.92	15.70	29.23	**34.15**
e = 3	13.52	13.25	14.38	15.29	15.80	29.97	**34.96**
e = 4	13.40	13.56	14.42	15.77	15.67	30.01	**35.48**
e = 5	13.62	13.75	14.56	15.15	15.68	30.04	**36.72**
Shanghai	Precision@20						
Method	FastTag	TagRel	TagProp	CMF	TMC	PTC	JEREMIE
e = 1	11.84	12.01	11.70	15.29	15.94	23.07	**32.18**
e = 2	11.93	11.30	13.32	15.06	15.76	24.23	**34.13**
e = 3	12.66	12.46	13.59	15.51	15.83	24.90	**34.21**
e = 4	13.26	13.20	14.05	15.73	15.79	25.20	**35.78**
e = 5	14.20	12.86	14.20	15.33	15.78	25.32	**37.57**

Table 3. Recall@20 of Image Tag Recommendation

London	Recall@20						
Method	FastTag	TagRel	TagProp	CMF	TMC	PTC	JEREMIE
e = 1	25.16	23.27	27.50	33.76	36.28	63.69	**74.21**
e = 2	26.67	25.11	27.68	33.25	38.10	66.90	**78.26**
e = 3	26.88	26.37	28.38	33.98	38.63	67.98	**80.23**
e = 4	27.58	27.16	29.04	34.70	39.12	68.95	**81.08**
e = 5	28.43	27.86	29.15	34.33	39.97	69.28	**81.81**
New York	Recall@20						
Method	FastTag	TagRel	TagProp	CMF	TMC	PTC	JEREMIE
e = 1	28.66	26.32	33.64	37.91	42.15	54.17	**75.16**
e = 2	29.25	27.54	33.99	37.87	43.07	57.71	**79.29**
e = 3	29.34	28.41	34.57	38.02	43.18	59.72	**81.37**
e = 4	29.96	28.80	34.36	37.45	43.20	61.37	**82.56**
e = 5	30.84	29.66	34.91	37.99	43.17	61.67	**83.33**
Beijing	Recall@20						
Method	FastTag	TagRel	TagProp	CMF	TMC	PTC	JEREMIE
e = 1	32.97	33.77	33.62	43.26	43.35	71.91	**85.85**
e = 2	33.98	34.03	37.90	43.90	42.91	76.65	**89.80**
e = 3	36.88	34.36	39.54	43.53	43.18	78.24	**90.98**
e = 4	36.90	33.99	39.64	43.83	42.81	78.55	**91.26**
e = 5	39.97	33.86	40.03	43.01	42.82	78.51	**92.96**
Paris	Recall@20						
Method	FastTag	TagRel	TagProp	CMF	TMC	PTC	JEREMIE
e = 1	38.91	39.81	42.23	46.24	53.81	67.89	**84.74**
e = 2	39.56	38.16	42.14	45.95	54.91	70.15	**87.96**
e = 3	40.15	38.19	42.36	45.44	47.19	70.71	**89.24**
e = 4	41.47	39.57	43.84	45.77	48.29	71.05	**89.75**
e = 5	42.18	40.23	43.97	45.38	48.32	71.52	**90.24**
Shanghai	Recall@20						
Method	FastTag	TagRel	TagProp	CMF	TMC	PTC	JEREMIE
e = 1	32.09	31.50	32.64	41.64	43.98	64.36	**81.84**
e = 2	33.38	30.56	36.49	41.17	43.55	66.46	**84.87**
e = 3	35.80	32.66	37.06	42.07	43.76	68.68	**85.82**
e = 4	37.45	32.77	38.30	41.55	43.62	69.00	**86.57**
e = 5	38.39	33.54	38.76	41.93	43.62	69.11	**87.82**

Table 4. MRR performance of IBUR and IBPR

London	MRR @ IBUR				MRR @ IBPR			
Method	NMF	CMF	J-G	J-U	NMF	CMF	J-G	J-P
e = 1		12.41	19.70	**33.46**		6.67	13.16	**20.87**
e = 2		12.15	20.88	**34.70**		6.38	15.40	**22.74**
e = 3	7.96	12.69	21.37	**35.63**	3.67	6.89	15.88	**23.34**
e = 4		13.01	21.96	**36.93**		7.27	16.66	**23.86**
e = 5		13.27	23.04	**38.68**		7.34	17.83	**24.95**
New York	MRR @ IBUR				MRR @ IBPR			
Method	NMF	CMF	J-G	J-U	NMF	CMF	J-G	J-P
e = 1		16.44	26.13	**41.83**		10.62	17.17	**23.76**
e = 2		16.62	28.36	**43.10**		10.34	18.45	**25.55**
e = 3	11.07	16.86	28.78	**43.72**	6.41	10.85	18.90	**26.31**
e = 4		17.11	29.66	**44.26**		11.20	19.33	**27.15**
e = 5		17.48	31.05	**46.65**		11.36	20.62	**28.88**
Paris	MRR @ IBUR				MRR @ IBPR			
Method	NMF	CMF	J-G	J-U	NMF	CMF	J-G	J-P
e = 1		24.11	33.52	**51.65**		11.43	17.62	**28.61**
e = 2		23.98	34.49	**52.59**		11.05	18.27	**28.91**
e = 3	15.34	24.45	34.52	**54.01**	5.56	11.77	19.03	**30.89**
e = 4		24.76	35.10	**56.30**		12.06	20.50	**33.08**
e = 5		25.22	36.34	**57.61**		12.31	21.41	**35.24**
Beijing	MRR @ IBUR				MRR @ IBPR			
Method	NMF	CMF	J-G	J-U	NMF	CMF	J-G	J-P
e = 1		11.12	36.52	**46.32**		27.32	38.24	**56.34**
e = 2		10.99	36.15	**47.21**		26.64	39.95	**56.32**
e = 3	7.58	11.46	37.49	**48.37**	18.25	27.39	41.01	**58.81**
e = 4		11.37	38.37	**48.70**		26.91	41.56	**59.63**
e = 5		12.03	38.23	**48.85**		27.88	41.77	**59.92**
Shanghai	MRR @ IBUR				MRR @ IBPR			
Method	NMF	CMF	J-G	J-U	NMF	CMF	J-G	J-P
e = 1		14.05	18.02	**38.28**		18.37	36.89	**40.48**
e = 2		13.65	19.43	**41.04**		18.62	37.45	**42.83**
e = 3	10.01	14.21	19.69	**41.67**	12.13	17.80	38.76	**44.86**
e = 4		14.16	20.80	**43.87**		18.45	40.87	**47.74**
e = 5		14.64	20.31	**43.81**		18.92	40.12	**47.27**

shared latent embedding. The performance gap between **J-G** and **J-U** in the IBUR task gives a strong support for the effectiveness of structure preservation in the procedure of jointly learning semantic features. Similar conclusion can be derive with **J-G** and **J-P** in the IBPR task. This experimental phenomenon demonstrates that the semantic features T_i and U_i are indispensable for the reconstruction of R_i in IBUR task, and we can derive similar conclusion in IBPR. Moreover, both the visual contents and structure information make contributions to the high-level semantic information sharing with users and POIs.

5.5 Evaluation of Model Effectiveness

To evaluate the effectiveness of the learning model of **JEREMIE**, we propose to consider different combinations of components by setting the corresponding parameters α and β. Since our learning model is symmetric, we first generate two individual sub-models $\mathbf{JEREMIE_U}$ and $\mathbf{JEREMIE_P}$ from the original model by wiping out corresponding components. Then we consider to test the effectiveness of structure preservation in $\mathbf{JEREMIE_G}$ by setting the parameter α to 0. Similarly, we obtain $\mathbf{JEREMIE_F}$ to test the effectiveness of information averaging via visual content by setting the parameter β to 0. All of the four types of weakening variant of our method have been introduced in Sect. 5.2.

Table 5 shows the scores of all tasks for this experiment in *Beijing* and *Shanghai*. In AIA task, we can see that the performance of fusion results with our **JEREMIE** model are better than separate results (**J-U** and **J-P**). It gives strong support for the effectiveness of minimum pooling strategy. When comparing the performance in AIA between **J-F** and **J-G** we can see that the information averaging via visual content is more important than the structure preserva-

Table 5. Experiment results of different settings

Beijing	Precision@20					MRR@IBUR		
Methods	J-P	J-U	J-F	J-G	JEREMIE	J-F	J-G	J-U
e = 1	22.54	26.25	23.62	29.34	**33.34**	33.46	36.52	**46.32**
e = 2	23.76	28.67	24.45	30.66	**34.15**	33.37	36.15	**47.21**
e = 3	23.63	28.68	24.87	31.14	**34.96**	34.21	37.49	**48.37**
e = 4	24.83	29.73	25.36	31.59	**35.48**	35.43	38.37	**48.70**
e = 5	25.48	30.32	26.77	31.71	**36.72**	35.77	38.23	**48.85**
Beijing	Recall@20					MRR@IBPR		
Methods	J-P	J-U	J-F	J-G	JEREMIE	J-F	J-G	J-P
e = 1	58.81	69.56	59.41	76.03	**85.85**	31.64	38.24	**56.34**
e = 2	60.83	74.29	63.17	78.08	**89.80**	33.15	39.95	**56.32**
e = 3	61.85	73.94	62.84	79.10	**90.98**	33.58	41.01	**58.81**
e = 4	62.35	76.02	64.31	80.05	**91.26**	34.65	41.56	**59.63**
e = 5	62.99	77.48	66.22	80.96	**92.96**	34.23	41.77	**59.92**

tion in our semantic feature learning framework. However, the performance in IBUR and IBPR between **J-F** and **J-G** is the opposite of the one in AIA demonstrates that the structure preservation makes greater contribution in both the IBUR and IBPR tasks. This result demonstrates that our proposed model is effective even when the two tasks are not bound together. Combined with the results in all of the three tasks, we can see that our proposed relation structure simultaneously improve the performance of all the three tasks within a unified learning process.

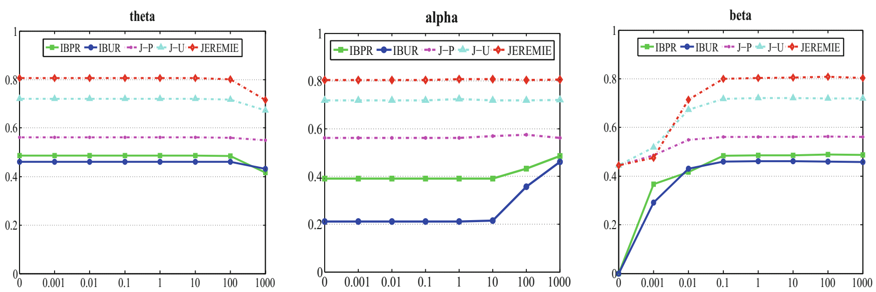
5.6 Parameter Sensitivity Analysis

We evaluate the sensitivity of the parameters in JEREMIE with *Shanghai* dataset in the condition of $e = 5$. We select three significant parameters $\{\alpha, \beta, \theta\}$ to illustrate the MAP@5 and MRR scores. For each parameter to be analysed, we fix other parameters as the original value ($\alpha = 1000, \beta = 10, \theta = 1$). The parameters α, β and θ range from 10^{-3} to 10^{3} (also includes 0) for sensitivity test.

All of the curves in Fig. 2(a) keep stable when $\theta \leq 100$, and they all tend to descent from $\theta = 100$ to $\theta = 1000$. While we choose 1 instead of 100 as the optimal value for θ to make our model more flexible in parameter tuning.

In Fig. 2(b) we can see that the MAP scores of AIA are almost stable in a wide range of α. However, the shape of the MRR curves (IBUR and IBPR) show an ascend trend, which means the user interest constraints F_i and F_j is sensible in IBUR and IBPR respectively. To balance the performance, we choose 1000 as the optimal value for α.

Illustrated in Fig. 2(c), the trend of all scores goes up with the increment of β and all of the metrics level off in the condition of $\beta \geq 1$. To obtain a relatively better performance, we choose 10 as the optimal value for β respectively.



(a) θ (b) α (c) β

Fig. 2. Sensitivity analysis of θ, α, β in terms of MAP@5 and MRR on the Shanghai dataset

5.7 Space Complexity Analysis

The space requirement for TMC and NMF is $O(n \times m)$. Since TagProp utilizes the KNN technique to reduce the size of search space, its space requirement is $O(n \times K)$. By reducing the problem scale with POI-specific relation matrix partition, the space requirement of our method is $O(n_{max} \times m)$, where n_{max} is the number of images in the largest POI. Typically, $m > 5K$ and $n > 10n_{max}$. Therefore, our method JEREMIE achieves the smallest space requirement than other state-of-the-art models.

5.8 Time Complexity Analysis

For time complexity analysis, we first calculate time complexity for each term in our loss function, then we aggregate them into a final result. The time complexity of each term and corresponding subgradients is listed as:

$F_i \ : O(u * m * n_i + u^2 * m)$
$G_i \ : O(d * m * n_i + n_i^2 * m)$
$E_i \ : O(u * m * n_i)$
$\nabla_{T_i} A_i \ : O(u * m * n_i + u * n_i^2 + n_i * u^2 + n_i^2 * m)$
$\nabla_{U_i} A_i \ : O(u * m * n_i + m * u^2)$
$\nabla_{W_U} A_i \ : O(d * n_i^2 * p + d^2 * n_i * p + d^2 * m * p)$

In our experiment we fix m, d as constants. We approximation u by $u \approx n_k/\omega$, where ω is a constant. The time complexity with user blocks is $O(n_i^3)$. We can obtain similar results by analyzing the symmetric terms of POI blocks. The time complexity of our framework is $O((\max(n_i, n_j))^3)$, where n_i is the number of images in user block i and n_j is the number of images in POI block i.

5.9 Efficiency Analysis

We evaluate the computational efficiency of proposed JEREMIE method in different parallel mode on a 100 K image subset of London. Our algorithm is implemented on MATLAB R2014a, and run on a Intel i7-4770K CPU and 32 GB RAM configured PC. Figure 3 illustrates the running time per iteration of our proposed JEREMIE method in serial and 2-parallel modes. The shape of the curves demonstrates that the parallelism of our proposed JEREMIE method effectively reduces the computational time as the increment of scalability. It is no surprising that our parallel computational framework conducted by matrix partition strategy is the key point of efficiency improvement.

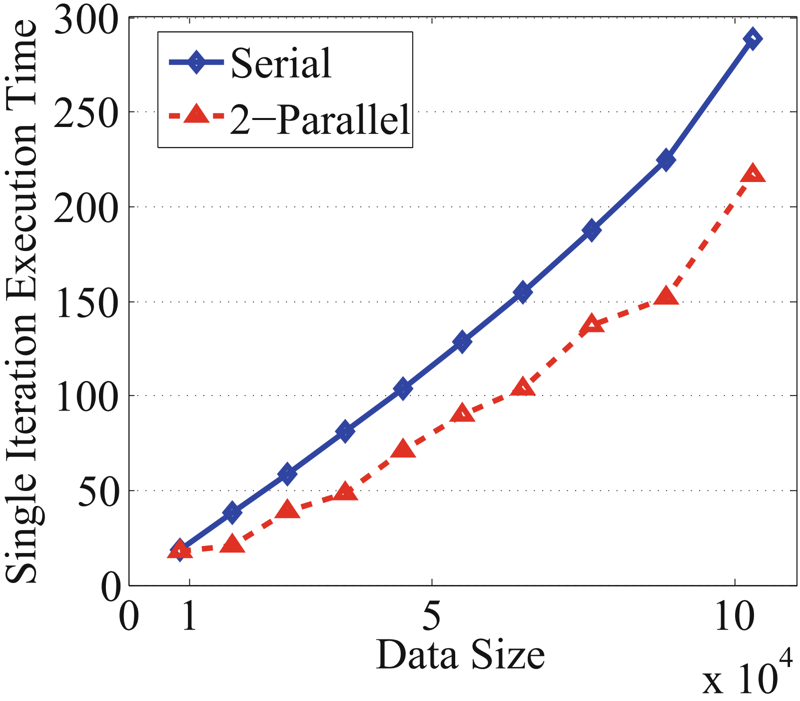


Fig. 3. Serial and parallel execution time of single iteration

6 Conclusion

We propose the JEREMIE method, a **J**oint S**E**mantic Featu**R**e L**E**arning model via **M**ulti-relational Matr**I**x Compl**E**tion, which jointly complements the semantic features of different entities from heterogeneous domains. We learn the blockwise *image-tag*, *POI-tag* and *user-tag* relations by optimizing a unified objective function which consists of matrix factorization and transductive matrix completion. The overall learning problem can be decomposed into a set of POI-specific and user-specific sub-tasks. Experimental results on automatic image annotation and image-based user (POI) retrieval demonstrate that our approach achieves promising performance on real world social media datasets. In future work, we will fully investigate the connection among users, content, concepts and locations towards a more flexible relation modeling framework.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant 61672497, Grant 61332016, Grant 61620106009, Grant 61650202 and Grant U1636214, in part by the National Basic Research Program of China (973 Program) under Grant 2015CB351802, and in part by the Key Research Program of Frontier Sciences of CAS under Grant QYZDJ-SSW-SYS013. This work was also partially supported by CAS Pioneer Hundred Talents Program by Dr. Qiang Qu.

References

1. Cartis, C., Gould, N.I., Toint, P.L.: On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming. SIAM J. Optim. **21**(4), 1721–1739 (2011)
2. Chen, M., Zheng, A., Weinberger, K.: Fast image tagging. In: ICML, pp. 1274–1282 (2013)
3. Cho, E., Myers, S.A., Leskovec, J.: Friendship and mobility: user movement in location-based social networks. In: KDD, pp. 1082–1090. ACM (2011)
4. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. TPAMI **24**(5), 603–619 (2002)
5. Crandall, D.J., Backstrom, L., Huttenlocher, D., Kleinberg, J.: Mapping the world's photos. In: WWW 2009, pp. 761–770. ACM (2009)
6. Guillaumin, M., Mensink, T., Verbeek, J., Schmid, C.: TagProp: discriminative metric learning in nearest neighbor models for image auto-annotation. In: ICCV, pp. 309–316. IEEE (2009)
7. Halpin, H., Robu, V., Shepherd, H.: The complex dynamics of collaborative tagging. In: WWW, pp. 211–220. ACM (2007)
8. Hariharan, B., Zelnik-Manor, L., Varma, M., Vishwanathan, S.: Large scale max-margin multi-label classification with priors. In: ICML, pp. 423–430 (2010)
9. Jenatton, R., Roux, N.L., Bordes, A., Obozinski, G.R.: A latent factor model for highly multi-relational data. In: NIPS, pp. 3167–3175 (2012)
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105 (2012)
11. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature **401**(6755), 788–791 (1999)
12. Li, X., Snoek, C.G., Worring, M.: Learning social tag relevance by neighbor voting. TMM **11**(7), 1310–1322 (2009)
13. Lin, Z., Ding, G., Hu, M., Wang, J., Ye, X.: Image tag completion via image-specific and tag-specific linear sparse reconstructions. In: CVPR, pp. 1618–1625. IEEE (2013)
14. Liu, D., Yan, S., Hua, X.S., Zhang, H.J.: Image retagging using collaborative tag propagation. TMM **13**(4), 702–712 (2011)
15. Liu, D., Ye, G., Chen, C.T., Yan, S., Chang, S.F.: Hybrid social media network. In: ACMMM, pp. 659–668. ACM (2012)
16. Liu, J., Li, Z., Tang, J., Jiang, Y., Lu, H.: Personalized geo-specific tag recommendation for photos on social websites. TMM **16**(3), 588–600 (2014)
17. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. JMLR **11**, 19–60 (2010)
18. Moxley, E., Kleban, J., Manjunath, B.: Spirittagger: a geo-aware tag suggestion tool mined from flickr. In: ACM MIR, pp. 24–30. ACM (2008)
19. Qian, X., Feng, H., Zhao, G., Mei, T.: Personalized recommendation combining user interest and social circle. TKDE **26**(7), 1763–1777 (2014)
20. Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: LabelMe: a database and web-based tool for image annotation. IJCV **77**(1–3), 157–173 (2008)
21. Singh, A.P., Gordon, G.J.: Relational learning via collective matrix factorization. In: KDD, pp. 650–658. ACM (2008)
22. Sutskever, I., Tenenbaum, J.B., Salakhutdinov, R.R.: Modelling relational data using bayesian clustered tensor factorization. In: NIPS, pp. 1821–1828 (2009)

23. Thomee, B., Shamma, D.A., Friedland, G., Elizalde, B., Ni, K., Poland, D., Borth, D., Li, L.J.: YFCC100M: the new data in multimedia research. Commun. ACM **59**(2), 64–73 (2016)
24. Wang, Z., Zhu, W., Cui, P., Sun, L., Yang, S.: Social media recommendation. In: Ramzan, N., van Zwol, R., Lee, J.S., Clüver, K., Hua, X.S. (eds.) Social Media Retrieval. CCN, pp. 23–42. Springer, London (2013). https://doi.org/10.1007/978-1-4471-4555-4_2
25. Wu, L., Jin, R., Jain, A.K.: Tag completion for image retrieval. TPAMI **35**(3), 716–727 (2013)
26. Wu, L., Yang, L., Yu, N., Hua, X.S.: Learning to tag. In: WWW 2009, pp. 361–370. ACM (2009)
27. Ye, M., Yin, P., Lee, W.C., Lee, D.L.: Exploiting geographical influence for collaborative point-of-interest recommendation. In: SIGIR, pp. 325–334. ACM (2011)
28. Yin, H., Sun, Y., Cui, B., Hu, Z., Chen, L.: LCARS: a location-content-aware recommender system. In: KDD, pp. 221–229. ACM (2013)
29. Zhang, J., Wang, S., Huang, Q.: Location-based parallel tag completion for geo-tagged social image retrieval. In: ICMR, pp. 355–362. ACM (2015)
30. Zhu, G., Yan, S., Ma, Y.: Image tag refinement towards low-rank, content-tag prior and error sparsity. In: ACMMM, pp. 461–470. ACM (2010)
31. Zhu, J.: Max-margin nonparametric latent feature models for link prediction. In: ICML, pp. 719–726 (2012)