TGTM: Temporal-Geographical Topic Model for Point-of-Interest Recommendation

Cong Zheng(✉), Haihong E, Meina Song, and Junde Song

School of Computer Science, Beijing University of Posts and Telecommunications, No. 10, Xitucheng Road, Haidian District, Beijing, China
zhengcongbupt@163.com

Abstract. The wide spread use of location based social networks (LBSNs) and Micro-blogging services generated large volume of users' check-in data, which consists of user ids, textual contents, posting timestamps, geographic information and so on. Point-of-interest (POI) recommendation is a task to provide personalized recommendations of interesting places to enhance the user experience in LBSNs. In this paper, we propose 2 novel time-location-content aware POI recommendation models which jointly integrate auxiliary temporal, textual and spatial information to improve the performance of POI recommendation. Specifically, we utilize temporal information to partition the original user-POI check-in frequency matrix into sub-matrices so that behavior in similar temporal scenario can be grouped. Then, we take advantage of Latent Dirichlet Allocation (LDA) model and spatial coordinates to infer the POIs. Comprehensive experiments conducted using real-world datasets demonstrate the superiority of our approach.

Keywords: Recommendation · Point of interest · Location-based social networks · Human mobility

1 Introduction

Given the phenomenal growth rate of user population and the vast amount of mobile devices with positioning function, location-based social networking service has become immensely popular with hoards of online web sites, such as Facebook Places, Google latitude, Twitter and Foursquare, etc. LBSNs now allow users share not only their physical location coordinates and time stamps in the form of "check-in", but also write textual opinions on the POIs they have visited. Mining and modeling the check-in behavior using the location log data is of great value to both users and POIs, because POI recommendation can help people discover attractive places and may foster more potential business for the owners of POIs.

To this end, POI recommendation has become a popular research topic in the past few years. One of the most crucial challenges in POI recommendation is how to cope with the extreme sparsity of user-POI check-in frequency matrix.

S.B. Navathe et al. (Eds.): DASFAA 2016, Part I, LNCS 9642, pp. 348–363, 2016.
DOI: 10.1007/978-3-319-32025-0_22

In addition, unlike traditional recommendations which only take the user-item rating into account, POI recommender systems is much more complex, since the mobile behavior in this scenario could be a mixture of many aspects. Particularly, when there are rich text such as comments or microblogs, the POI recommendation should be personalized, location-content aware and context depended. In light of this, diverse types of information pose another big challenge which is how to incorporate them in a unified way systematically.

Intuitively, users tend to visit restaurants within a short distance near office for lunch at noon on weekday. But on weekends, users may pay a long visit to bars at night for fun. We believe that mobile patterns can be easier found in similar temporal context. In light of this, we partition the original user-POI information into subgroups based on temporal stamp at first. Then, we employ LDA algorithm [2] to infer users' interests and the topic distributions of POIs. In [4], the authors found users' displacements (distance between 2 consecutive check-in locations) trend can be approximated by a power-law distribution. Taking into account the textual and geographical influence, on the basis of matrix factorization model, we can infer to what extent does a user prefer a POI.

In summary, the contributions of this paper are threefold:

(1) We study the relationship between users' implicit feedback check-in behavior and auxiliary information on LBSNs in terms of textual content information, geographical coordinates and temporal influence.
(2) We propose 2 novel probabilistic matrix factorization models for POI recommendation, and each of them incorporates the above three types of auxiliary information.
(3) We evaluate the presented methods by comprehensive experiments on real world LBSN datasets extracted from Twitter's API. The results demonstrate the effectiveness of our methods.

We have the usual organization: Survey, problem definition, proposed method, experiments and conclusions.

2 Related Work

2.1 Matrix Factorization

Collaborative Filtering (CF) in recommender systems can predict personalized preferences to unconsumed items [1,19]. There are two major categories of Collaborative Filtering methods. Neighborhood-based solutions predict users' potential interests by finding like-minded users [5,18], which compute similar users or items using similarity functions such as Cosine Distance or Pearson Correlation. Model-based [9,17] solutions utilize the observed ratings or tags to model the user-item interaction. A variety of successful realizations of model-based models are based on matrix factorization which decomposes the user-item rating matrix. In this paper, we resort to approaches based on matrix factorization method, so we briefly introduce it here.

Basically, matrix factorization (MF) methods factorize a rating matrix $\boldsymbol{R}$ into one user-specific matrix $\boldsymbol{U}$ and one item-specific matrix $\boldsymbol{C}$ [10]. Then the original rating matrix can be approximated by multiplying the two factorized matrices, while avoiding over-fitting using regularization terms:

$$\begin{aligned} \arg min_{\boldsymbol{U},\boldsymbol{C}} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij} \left(R_{ij} - \boldsymbol{U_i}^{\boldsymbol{T}} \boldsymbol{C_j} \right)^2 \\ + \lambda \left(\|\boldsymbol{U}\|_F{}^2 + \|\boldsymbol{C}\|_F{}^2 \right) \end{aligned} \tag{1}$$

where $\|\cdot\|_F^2$ denotes the Frobenius norm, and I_{ij} is the indicator function that is equal to 1 if user i rated movie j and equal to 0 otherwise. The constant λ controls the extent of regularization in order to alleviate the over-fitting problem. The probabilistic matrix factorization (PMF) [16] gives a probabilistic explanation for the regularization.

2.2 POI Recommendation

POI recommendation was firstly studied on GPS trajectory data [23]. With the growing popularity of LBSNs, POI recommendation has drawn a lot of attention. There is a vast literature on POI recommendation exploiting geographical influence. A mutual observation has been discovered by existing studies is that people tend to visit nearby locations. User-based CF and item-based CF are studied in [11,20]. Ye et al. [20] explored social and geographical factors together under an user-based CF framework to make POI recommendation. There has been recent interest in leveraging the social and geographical properties to improve the effectiveness of POI recommendation. In [3], they proposed a matrix factorization scheme incorporating the geographical and social information by a Gaussian mixture model (GMM). In [6], authors discovered that preference derived from similar users was more important for in-town users while friendship became more important for out-of-town users.

Liu et al. [13] proposed a graph-based method which exploits temporal and geographical information in an integrated way. Yuan et al. [21] incorporated temporal influence in an user-based CF manner, and the final preference score for a candidate POI are linearly combined by the scores computed based on temporal influence and geographical influence respectively.

When it comes to the textual content information, Gao et al. [7] studied the content information on LBSNs w.r.t. POI properties, user interests, and sentiment indications under an unified POI recommendation framework. Liu and Xiong [12] studied the effect of POI-associated tags with an aggregated LDA model. Hu and Ester [8] investigated the user-interest from Twitter and Yelp according to topic modeling method. Yuan et al. [22] jointly model individual user's mobility behavior from spatial, temporal and content aspects. They captured them in a probabilistic generative model.

3 Proposed Approach

3.1 Problem Definition

The problem of POI recommendation is to recommend POIs potentially attractive to users. For ease of exposition, let $\boldsymbol{u} = \{u_1, u_2, ..., u_m\}$ be the set of users and $\boldsymbol{c} = \{c_1, c_2, ..., c_n\}$, be the set of POIs, where m and n denote the number of users and POIs, respectively. Each user has observable properties x_i (e.g., user's comments and check-in history). Also, each POI keeps observable properties x_j (e.g., POI's related textual description and spatial coordinates in terms of longitude and latitude). $\boldsymbol{R} \in \mathbb{R}^{m \times n}$ is a check-in frequency matrix with each entry R_{ij} representing the observed check-in frequency made by u_i at c_j. Applying LDA model to all the textual descriptions related to all the users and POIs can help us obtain the topic distribution $\boldsymbol{\theta_i}$ for each user u_i or $\boldsymbol{\pi_j}$ for POI c_j.

3.2 Time Aware User-POI Subgrouping

Subgrouping the original user-item-rating matrix based on contextual information has proven to be a promising way of improving recommender systems [15], because the generated sub-groups contain similar ratings which have higher correlations. Specifically, we use temporal information (i.e., day-of-week and hour-of-day) to partition the original user-POI check-in frequency matrix into 4 subgroups. Intuitively, people's activity may be more regular and predictable on weekday while more various and similar on weekend. So we first partition the original matrix according to day-of-week (i.e., weekday versus weekend). Next, a day (24 h) can be divided into working segment (from 08:00 am to 17:59 pm) and leisure hours (from 18:00 pm to 7:59 am of the next day). Then, the generated 2 submatrices according to day-of-week can be further derived into 4 based on hour-of-day. At last, we get 4 user-POI check-in frequency sub-matrices, in which we can make predictions by leveraging geographical and textual information.

3.3 Exploiting Textual Content and Spatial Information

LDA algorithm is known to have poor performance on short text documents such as short tweets. In this paper, we aggregate all the textual comments associated with same POI into a POI document. We also combine all the textual comments of the POIs that each user has checked in into a user document. As a consequence, we get a large document collection, and each document corresponds to one POI or user. Then LDA [2] is utilized to analyze the topic distribution of every document. Like [12], we define the matching score between user u_i and POI c_j as the similarity with regards to user's topic distribution $\boldsymbol{\theta_i}$ and the topic distribution of POI $\boldsymbol{\pi_j}$. We resort to Jensen-Shannon divergence to measure the similarity between above 2 multinomial topic distributions. The symmetric Jensen-Shannon divergence between them is:

$$D_{JS}(u_i, c_j) = \frac{1}{2} D(\boldsymbol{\theta_i} \parallel \boldsymbol{M}) + \frac{1}{2} D(\boldsymbol{\pi_j} \parallel \boldsymbol{M}) \tag{2}$$

where $\boldsymbol{M} = \frac{1}{2}(\boldsymbol{\theta_i} + \boldsymbol{\pi_j})$ and $D(\cdot \parallel \cdot)$ is the Kullback-Leibler distance. The matching score is defined as:

$$S(u_i, c_j) = 1 - D_{JS}(u_i, c_j) \tag{3}$$

Next, we exploit the spatial information. Previous work [4] investigated the distance-based displacements of consecutive check-in made by users. They found the trend of check-in frequency with regard to displacement can be approximated by a power-law distribution. To be consistent with this observed property, supposing an user is at POI c_j now, we model the probability that a user may visit another POI c_k in terms of distance as [13]:

$${p_{jk}}^d = [Dist(j, k)]^{-1} \tag{4}$$

where ${p_{jk}}^d$ is the probability that a user will check-in at POI c_k from POI c_j, and $Dist(j, k)$ is the distance between the 2 positions.

In practice, users may not share their position information in time, so Eq. (4) can not be simply applied in this scenario without precise spatial coordinates. Intuitively, if a user frequently check-in at a certain POI, this POI can be very influencial to his daily life because he may live or work there. While user's personal behavior is less prone to rely on the POIs which the user rarely occur. According to this phenomenon, we postulate that the attraction of POI c_j to user u_i is the weighted average of probability calculated by Eq. (4) between candidate POI c_j and the POIs in user's (user u_i) check-in log list. The weights are proportional to the user's check-in frequency at each POI in his own history. For example, suppose there are K POIs in the historical list $\boldsymbol{f_i}$ of user u_i. The geographical attractiveness of POI c_j to u_i is calculated as:

$${p_{ji}}^d = \frac{\sum_{k=1}^{K} f_{ik} {p_{jk}}^d}{\sum_{k=1}^{K} f_{ik}} \tag{5}$$

where f_{ik} is the probability with regard to check-in frequency of user u_i at POI c_k and ${p_{jk}}^d$ is computed by using Eq. (4). The displacement is between POI c_k and c_j.

We assume that the probability of checking in at a POI should reflect both the content and geographical influence. So we fuse these 2 factors and derive the integrated attractiveness of POI c_j to user u_i as:

$$W_{ij} = [S(u_i, c_j)]^a \times [{p_{ji}}^d]^b \tag{6}$$

where ${p_{ji}}^d$ is calculated by Eq. (5) and $S(u_i, c_j)$ is calculated by Eq. (3). a and b are 2 decaying parameters.

3.4 Temporal-Geographical Topic Model for Personalized POI Recommendation

To jointly leverage temporal, spatial and content information, we integrate them into a probabilistic matrix factorization model (TGTM-1). Figure 1(a) shows the graphical model of TGTM-1. We define the conditional distribution of observed check-in frequencies as:

$$p(\boldsymbol{R} \mid \boldsymbol{U}, \boldsymbol{C}, \boldsymbol{W}, \sigma^2) = \prod_{i=1}^{m} \prod_{j=1}^{n} [\mathcal{N}(R_{ij} \mid f(\boldsymbol{U_i}, \boldsymbol{C_j}, W_{ij}), \sigma^2)]^{I_{ij}} \tag{7}$$

where $\mathcal{N}(x \mid \mu, \sigma^2)$ is the probabilistic density function of Gaussian distribution with mean μ and variance σ^2, and I_{ij} is the indicator function that is equal to 1 if user u_i visited POI c_j and equal to 0 otherwise. We use function $f(\boldsymbol{U_i}, \boldsymbol{C_j}, W_{ij})$ to approximate the check-in frequency of user u_i at POI c_j.

Taking spatial and content influence into consideration, we define:

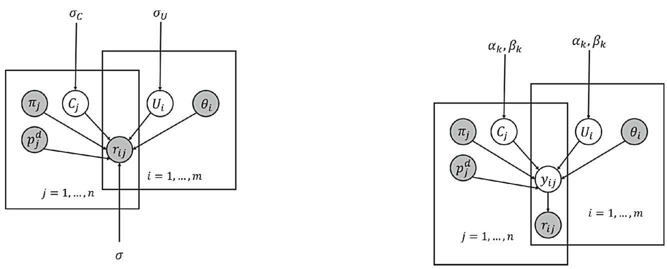
$$f(\boldsymbol{U_i}, \boldsymbol{C_j}, W_{ij}) = W_{ij} \cdot \boldsymbol{U_i^T} \boldsymbol{C_j} \tag{8}$$

where W_{ij} is computed by using Eq. (6). The weighted product of user-specific and POI-specific latent feature vectors allow our model to take full advantage of all the information and make personalized POI recommendation.

We place zero-mean spherical Gaussian priors on user and POI latent feature vectors:

$$p(\boldsymbol{U} \mid \sigma_U^2) = \prod_{i=1}^{m} \mathcal{N}(\boldsymbol{U_i} \mid 0, \sigma_U^2 \mathbf{I}) \tag{9}$$

$$p(\boldsymbol{C} \mid \sigma_C^2) = \prod_{j=1}^{n} \mathcal{N}(\boldsymbol{C_j} \mid 0, \sigma_C^2 \mathbf{I}) \tag{10}$$



(a) TGTM-1 (b) TGTM-2

Fig. 1. Graphical model of our 2 methods

Next, through a simple Bayesian inference, the posterior distribution is given by:

$$p(\boldsymbol{U},\boldsymbol{C} \mid \boldsymbol{R},\sigma^2,\boldsymbol{W},{\sigma_U}^2,{\sigma_C}^2) \propto$$
$$p(\boldsymbol{R} \mid \boldsymbol{U},\boldsymbol{C},\sigma^2,\boldsymbol{W},{\sigma_U}^2,{\sigma_C}^2)p(\boldsymbol{U} \mid {\sigma_U}^2)p(\boldsymbol{C} \mid {\sigma_C}^2) =$$
$$\prod_{i=1}^{m}\prod_{j=1}^{n}[\mathcal{N}(R_{ij} \mid f(\boldsymbol{U_i},\boldsymbol{C_j},W_{ij}),\sigma^2)]^{I_{ij}}\prod_{i=1}^{m}\mathcal{N}(\boldsymbol{U_i} \mid 0,{\sigma_U}^2\mathbf{I})\prod_{j=1}^{n}\mathcal{N}(\boldsymbol{C_j} \mid 0,{\sigma_C}^2\mathbf{I}) \tag{11}$$

The log of the posterior distribution over the user and POI latent features is:

$$\begin{aligned}&\texttt{ln}\; p(\boldsymbol{U},\boldsymbol{C} \mid \boldsymbol{R},\sigma^2,\boldsymbol{W},{\sigma_U}^2,{\sigma_C}^2) =\\ &-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n}I_{ij}[R_{ij}-f(\boldsymbol{U_i},\boldsymbol{C_j},W_{ij})]^2-\frac{1}{2{\sigma_U}^2}\sum_{i=1}^{m}\boldsymbol{U_i}^{\boldsymbol{T}}\boldsymbol{U_i}-\frac{1}{2{\sigma_C}^2}\sum_{j=1}^{n}\boldsymbol{C_j}^{\boldsymbol{T}}\boldsymbol{C_j}\\ &-\frac{1}{2}[(\sum_{i=1}^{m}\sum_{j=1}^{n}I_{ij})\texttt{ln}\;\sigma^2+md\texttt{ln}\;{\sigma_U}^2+nd\texttt{ln}\;{\sigma_C}^2]+P\end{aligned} \tag{12}$$

where P is a constant that does not depend on parameters and d is the number of latent features. Maximizing the log-posterior over latent vectors with fixed hyper-parameters is equivalent to minimizing the following sum-of-squared-errors objective function with quadratic regularization terms:

$$E = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}I_{ij}(R_{ij}-W_{ij}\cdot\boldsymbol{U_i}^{\boldsymbol{T}}\boldsymbol{C_j})^2+\frac{\lambda_U}{2}\sum_{i=1}^{m}\parallel\boldsymbol{U_i}\parallel_F{}^2+\frac{\lambda_C}{2}\sum_{j=1}^{n}\parallel\boldsymbol{C_j}\parallel_F{}^2 \tag{13}$$

where $\lambda_U = \sigma^2/{\sigma_U}^2$, $\lambda_C = \sigma^2/{\sigma_C}^2$, and $\parallel\cdot\parallel_F{}^2$ represents the Frobenius norm. A local minimum can be found by applying the gradient descent algorithm to the latent matrices $\boldsymbol{U}$ and $\boldsymbol{C}$ alternatively.

$$\frac{\partial E}{\partial \boldsymbol{U_i}} = -\sum_{j=1}^{n}I_{ij}(R_{ij}-W_{ij}\cdot\boldsymbol{U_i}^{\boldsymbol{T}}\boldsymbol{C_j})\cdot W_{ij}\boldsymbol{C_j}+\lambda_U\boldsymbol{U_i} \tag{14}$$

$$\frac{\partial E}{\partial \boldsymbol{C_j}} = -\sum_{i=1}^{m}I_{ij}(R_{ij}-W_{ij}\cdot\boldsymbol{U_i}^{\boldsymbol{T}}\boldsymbol{C_j})\cdot W_{ij}\boldsymbol{U_i}+\lambda_C\boldsymbol{C_j} \tag{15}$$

For the non-negative property of check-in frequency matrix, we use projected strategy by projecting a negative parameter value to 0 in each updating iteration.

After all the above optimal procedure, missing values in user-POI check-in matrix can be predicted as:

$$R_{ij}{}^* = \boldsymbol{U_i}^{\boldsymbol{T}}\boldsymbol{C_j} \tag{16}$$

To alleviate the potential negative values generated by Gaussian distribution, we have to use a projected strategy. We draw inspiration from [14], and propose an improved model (TGTM-2) which is more suitable for modeling nonnegative values, and it is shown in Fig. 1(b). In this model, we assume the observed check-in frequency R_{ij} in matrix $\boldsymbol{R}$ follows Poisson distribution with the mean Y_{ij} in matrix $\boldsymbol{Y}$. The elements Y_{ij} in matrix $\boldsymbol{Y}$ is defined in the same way as Eq. (8). But U_{ik} and C_{jk} are given the Gamma distribution as the empirical priors.

The generative process of every observed user-POI count R_{ij} is as follows:

1. Generate $U_{ik} \sim Gamma(\alpha_k, \beta_k), \forall k$
2. Generate $C_{jk} \sim Gamma(\alpha_k, \beta_k), \forall k$
3. Generate $Y_{ij} = W_{ij} \sum_{k=1}^{d} U_{ik} C_{jk}$
4. Generate $R_{ij} \sim Poisson(Y_{ij})$

The gamma distributions of $\boldsymbol{U}$ and $\boldsymbol{C}$ are given as follows:

$$p(\boldsymbol{U} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{m} \prod_{k=1}^{d} \frac{{U_{ik}}^{\alpha_k - 1} exp(-U_{ik}/\beta_k)}{{\beta_k}^{\alpha_k} \Gamma(\alpha_k)} \tag{17}$$

$$p(\boldsymbol{C} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{j=1}^{n} \prod_{k=1}^{d} \frac{{C_{jk}}^{\alpha_k - 1} exp(-C_{jk}/\beta_k)}{{\beta_k}^{\alpha_k} \Gamma(\alpha_k)} \tag{18}$$

where $\boldsymbol{\alpha} = \{\alpha_1, \cdots, \alpha_d\}$, $\boldsymbol{\beta} = \{\beta_1, \cdots, \beta_d\}$, $U_{ik} > 0, C_{jk} > 0, \alpha_k > 0$ and $\beta_k > 0$, $\Gamma(\cdot)$ is the Gamma function.

The Poisson distribution of $\boldsymbol{R}$ given $\boldsymbol{Y}$ is given by:

$$p(\boldsymbol{R} \mid \boldsymbol{Y}) = \prod_{i=1}^{m} \prod_{j=1}^{n} [\frac{{Y_{ij}}^{R_{ij}} exp(-Y_{ij})}{R_{ij}!}]^{I_{ij}} \tag{19}$$

where $Y_{ij} = W_{ij} \sum_{k=1}^{d} U_{ik} C_{jk}$, since every W_{ij} for each user-POI pair is fixed, the posterior distribution of $\boldsymbol{U}$ and $\boldsymbol{C}$ given R can be modeled as:

$$p(\boldsymbol{U}, \boldsymbol{C} \mid \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \propto p(\boldsymbol{R} \mid \boldsymbol{Y}) p(\boldsymbol{U} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}) p(\boldsymbol{C} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}) \tag{20}$$

The log of the posterior distribution over the user and POI latent features is:

$$\begin{aligned} \mathtt{ln}\, p(\boldsymbol{U}, \boldsymbol{C} \mid \boldsymbol{R}, \boldsymbol{W}, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \sum_{i=1}^{m} \sum_{k=1}^{d} ((\alpha_k - 1) \mathtt{ln}(\frac{U_{ik}}{\beta_k}) - \frac{U_{ik}}{\beta_k}) + \\ \sum_{j=1}^{n} \sum_{k=1}^{d} ((\alpha_k - 1) \mathtt{ln}(\frac{C_{jk}}{\beta_k}) - \frac{C_{jk}}{\beta_k}) &+ \sum_{i=1}^{m} \sum_{j=1}^{n} [I_{ij}(R_{ij} \mathtt{ln} Y_{ij} - Y_{ij})] + P \end{aligned} \tag{21}$$

where P is a const. Taking derivatives of Eq. (21) with respect to $\boldsymbol{U}$ and $\boldsymbol{C}$:

$$\frac{\partial L}{\partial \boldsymbol{U_i}} = \sum_{j=1}^{n} I_{ij} [W_{ij} (\frac{R_{ij}}{W_{ij} \boldsymbol{U_i} \boldsymbol{C_j}} - 1) \boldsymbol{C_j}] + \frac{\alpha_k - 1}{\boldsymbol{U_i}} - 1/\beta_k \tag{22}$$

$$\frac{\partial L}{\partial \boldsymbol{C_j}} = \sum_{i=1}^{m} I_{ij} [W_{ij} (\frac{R_{ij}}{W_{ij} \boldsymbol{U_i} \boldsymbol{C_j}} - 1) \boldsymbol{U_i}] + \frac{\alpha_k - 1}{\boldsymbol{C_j}} - 1/\beta_k \tag{23}$$

Equation (16) can be used to make predictions until the optimal procedure get convergence.

4 Experimental Analysis

4.1 Description of Datasets

We use part of the datasets kindly provide by [4]. Since Twitter status messages support the inclusion of geo-tags (latitude/longitude) as well as support third-party location sharing services like Foursquare and Gowalla, the spatial coordinates data is crawled using Twitter's API, and the other data was from Foursquare, the detailed data gathering method can be found in [4]. Foursquare is a large-scale location based social network sites. It allows users to check-in at different locations writing textual comments with time stamps and spatial information. In practice, users' check-in behavior may not always rely on temporal, spatial and content influence. For example, some people may go on a business trip and the destination may be very far away from their common active area. In order to clean up the data and remove the outliers which rarely occur, we filter the users who made less than 10 check-ins, and require that each POI should be visited 10 times at least. Moreover, we also constrain that an user should visit each POI at least 3 times. In practice, a large scale POI can have slightly different geographical coordinates. In our experiment, POIs are truncated according to their unique identifier. We get 3373 unique users and 9333 unique POIs after pruning. The pruned dataset contain 31727 check-in records. The density of the user-POI check-in matrix is 0.1008 %.

4.2 Metrics

In our experiments, we split the dataset into two parts : training (80 %) and testing datasets (20 %). We use 3 metrics for evaluating the performance: Normalized Mean Absolute Error (NMAE) , Normalized Root Mean Square Error (NRMSE) [14] and $Precision@K$. $Precision@K$ is a metric designed for top-K POI recommendation. They are defined as:

$$NMAE = \frac{\sum_{r=1}^{N} \mid (R_r - \hat{R}_r)/R_r \mid}{N}$$

$$NRMSE = \sqrt{\frac{1}{N}\sum_{r=1}^{N}\left[(R_r - \hat{R}_r)/R_r\right]^2}$$

$$Precision@K = \frac{\sum_{\boldsymbol{u}} \mid \boldsymbol{R(u)} \cap \boldsymbol{T(u)} \mid}{K}$$

where N is the total number of predictions, R_r is the real rating of an item and $\hat{R}_r$ is the corresponding predicted rating. $\boldsymbol{R(u)}$ is the list of top-K recommended POIs and $\boldsymbol{T(u)}$ are the visited locations of user u. $Precision@K$ w.r.t each user represents what percentage of POIs among the top-K recommended locations has actually been visited.

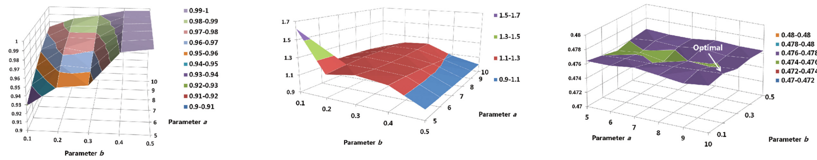
4.3 Impact of Spatial Information, Content and Temporal Influence

Since predictions are made in each subgroup independently, we can know how the models perform in each subgroup exactly. However, we do not know the overall performance with all the data in general. In light of this, after all the independent training and testing procedures in all the subgroups, we compute the metrics with all the testing datasets of all the subgroups as overall performance of the models. Take the metric $NMAE$ as an example, symbol N is the total number of predictions in all the subgroups not only one single subgroup. Other symbols can be derived in the same manner. Experimental results in Sect. 4.3 are the overall performance of the 2 proposed models.

We use grid search and fivefold cross validation to find the optimal parameters of Gamma distribution that achieve best overall experimental performance. The shape parameter of Gamma distribution $\alpha = 100$, and the scale parameter $\beta = 0.05$. The bigger β is, the wider numeric range will the Gamma distribution generate. In our experiment, we find that big β can hurt the model performance.

Parameters a and b are important because they determine to what extent do the textual contents and spatial influence affect the accuracy of prediction. Figures 2 and 3 show 3D-plots of predictive performance of our 2 models. The content parameter a balances the weight of textual contents: the bigger a is, the more we use textual content to make predictions. The spatial parameter b balances the effect of geographical influence, and the bigger b is, the more we take distance into consideration when making recommendation. We can see a clear decreasing trend from Fig. 3, the reason is that a and b affect the weighted matrix W in the exponential form. Although exponential form may generate big values generally, the performance on metric $Precision@K$ is still great. Taking all metrics into consideration, our methods gain the best predictive accuracy when $a = 10, b = 0.2$ for TGTM-1 whereas $a = 9, b = 0.3$ for TGTM-2. We also observe that the performance is more sensitive to spatial influence, slight change of parameter b can make the performance with regard to $NMAE$ and $NRMSE$ fluctuate. This indicates that both textual contents and spatial influence contribute to model performance, but the spatial information affect more. So the geographical attractiveness of POI plays a more important role in making travel decisions.

In our experiment, we find that although the performance with regards to $NMAE$ and $NRMSE$ may be poor, the metric $Precision@K$ is still relatively good (see Figs. 2 and 3). This is reasonable. Intuitively, in the POI recommendation scenario, because the check-in frequencies are in different order of magnitude, we care more about rank of recommendation results other than the accurate missing entry values. This phenomenon also demonstrates the effectiveness that we model the entries in the weighted matrix in exponential manner. Although this strategy may generate relatively big or small predictions, exponential strategy can distinguish users' wide range of check in frequency behavior well. However, this can not be simply applied to traditional rating prediction situation. Since in that situation, the numeric range of rating is fixed in 1–5, which is far more smaller than the range of check-in frequency.



(a) NMAE (b) NRMSE (c) Precision@3

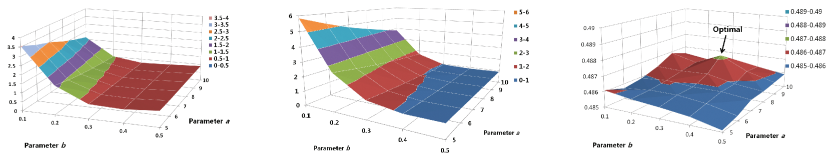
Fig. 2. Performance of TGTM-1 by varying content a and spatial b influence (Color figure online)

Then, we study the effect of temporal influence on our proposed model. In our proposed model, we subgroup the original user-POI check-in matrix into 4 submatrices according to time stamps (i.e., day-of-week and hour-of-day). In order to learn the temporal effect, we conducted the following experiment: using same spatial and textual parameter configuration, compare our complete model (TGTM-1 and TGTM-2) with comparisons (GTM-1 and GTM-2) regardless of the temporal influence. That means the comparisons making predictions using the original matrix without subgrouping strategy. In this experiment, using the above results, we set $a = 10, b = 0.2$ for TGTM-1 and GTM-1. We set $a = 9$, $b = 0.3$ for TGTM-2 and GTM-2. We can see from Fig. 4 that the temporal influence indeed affects the model performance a lot. Compared to the models without considering temporal influence, TGTM-2 and TGTM-1 improve the performance nearly 10 %. The Gamma priors taking time into consideration achieve better predictive performance. This result shows the effectiveness of time-based user-POI subgrouping.

4.4 Comparisons

In this subsection, we present the performance comparison between our models and some state-of-the-art POI recommendation algorithms.

BasicMF: predicts missing values according to $\boldsymbol{P} \approx \boldsymbol{U^T C}$, which only use the user-POI check-in matrix without any other auxiliary information (e.g., geographical, temporal information and textual contents). This one is used as baseline.



(a) NMAE (b) NRMSE (c) Precesion@3

Fig. 3. Performance of TGTM-2 by varying content a and spatial b influence

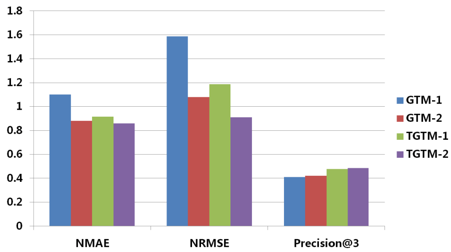


Fig. 4. Effect of temporal influence on our model

GeoCF: [20] takes into account of the spatial influence by assuming a power-law distribution and jointly integrates a user-based collaborative filtering algorithm. The recommendation is a linear combination of the spatial information and user preference.

UPT: [13] takes temporal and geographical influence into consideration, and integrates the auxiliary information into a matrix factorization model. In this approach, time stamps are also used for the purpose of subgrouping.

TLA: [12] incorporates item contents and spatial information into a matrix factorization model, thus it makes predictions based on item contents, spatial influence and check-in frequencies. In this model, spatial information is considered by using a regional level popularity factor rather than geographical coordinates.

TGTM-1 TGTM-2: [Sect. 3.4] use time stamps to do user-POI sub-grouping. In the subgroups, they makes predictions involving check-in frequencies, textual contents and spatial information into a probabilistic matrix factorization model. The user and POI latent feature matrices of TGTM-1 are assumed conforming to Gaussian priors, while the matrices of TGTM-2 are assumed conforming to Gamma priors.

In this experiment, we set $a = 10, b = 0.2$ for TGTM-1 and $a = 9, b = 0.3$ for TGTM-2. From Table 1 and Fig. 5, we can see that TGTM-2 which is applied Gamma prior indeed outperforms TGTM-1 with Gaussian prior. Table 1 also show that UPT, TGTM-1 and TGTM-2 perform different in each subgroup. The reason is that topic interests and geographical distance affect human mobile behavior slightly different when the time scenario change, especially that people tend to visit POIs which match their interests better on weekend in the expense of long distance. Because Basic MF, GeoCF and TLA do not take time information into consideration, we use their overall performance to represent the performance in each subgroup. We can also see that the performance with regard to metric $Precision@K$ is relatively high, the reason is that our data preprocessing procedure filter out the data with little relevance and all the users in testing part do not have too many check-ins. We leave large-scale comparative experiments as our future work. Although all the comparisons perform acceptable, our TGTM-2 model still improve the performance of baseline more than 30 % when $K = 1$.

Table 1. Performance comparison (Dimensionality = 10, K=3)

Subgroups	Metrics	BasicMF	GeoCF	UPT	TLA	TGTM-1	TGTM-2
Weekday work time	NMAE	1.2715	0.9289	1.2834	1.2649	0.9709	0.8542
	NRMSE	3.1386	1.7517	2.6651	3.0738	1.1714	0.9107
	Precision@3	30.09 %	35.24 %	45.53 %	40.16 %	47.69 %	48.75 %
Weekday leisure time	NMAE	1.2715	0.9289	1.2946	1.2649	0.9869	0.8714
	NRMSE	3.1386	1.7517	2.6874	3.0738	1.1869	0.9139
	Precision@3	30.09 %	35.24 %	44.53 %	40.16 %	47.67 %	48.72 %
Weekend work time	NMAE	1.2715	0.9289	1.2982	1.2649	0.9846	0.87
	NRMSE	3.1386	1.7517	2.673	3.0738	1.1929	0.9288
	Precision@3	30.09 %	35.24 %	42.6 %	40.16 %	46.18 %	47.71 %
Weekend leisure time	NMAE	1.2715	0.9289	1.3006	1.2649	0.9804	0.8646
	NRMSE	3.1386	1.7517	2.6885	3.0738	1.1896	0.9189
	Precision@3	30.09 %	35.24 %	43.88 %	40.16 %	45.67 %	46.70 %

BasicMF achieves the lowest accuracy and precision, because this general model does not take any auxiliary information into consideration. For the GeoCF model, because there is no social relationship information in our experimental data, we only take the user-based collaborative filtering factor and spatial influence factor into account. We found GeoCF obtain best performance when spatial influence counts 70 %, whereas user-based collaborative filtering factor counts 30 %. This means that spatial information plays a dominating role in the POI recommendation scenario. GeoCF performs better than BasicMF model, but it is still less accurate than our methods. The reason is that although GeoCF considers more relevant data using collaborative filtering approach and take geographical influence into account, some potential important factors for POI recommendation are still lost. For example, textual content information explicitly shows whether target POI match user's interests. Moreover, temporal information intuitively explains why the user's active range is bigger on weekend. UPT utilizes temporal information to classify original data into 4 subgroups, and integrates spatial influence into matrix factorization model. The experimental results demonstrate the benefits of combining the temporal and geographical influence jointly. TLA takes advantage of textual and spatial information to make POI recommendation. UPT outperforms TLA and GeoCF, which means that the temporal information contributes more to the user's behavior than explicit textual information and inexplicit user preferences inferred using user collaborative filtering method. The probable reason is that, most people work regularly on weekday, and weekday counts more time than weekend. So the topic interest show relatively poor influence.

Table 1 and Fig. 5 also show that our proposed 2 models consistently outperform the comparisons, demonstrating that jointly taking full advantage of textual content, temporal and geographical information indeed improves the performance of POI recommendation. Although some comparisons perform well on $NMAE$ and $NRMSE$, they still can not perform better than our 2 mod-

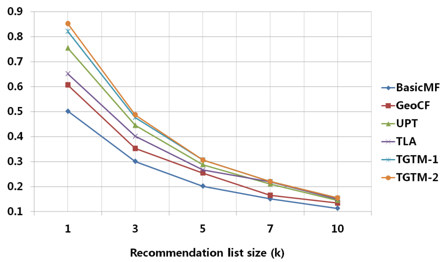


Fig. 5. *Precision@K* performance with different size of recommendation list

els on *Precision@K*. Because they can not handle the big range of check in frequency values suitably, which demonstrates that modeling the geographical influence and content information in exponential form makes our models more flexible for the big range of user-POI check-in frequencies.

5 Conclusions

In this paper, we have presented two probabilistic matrix factorization algorithms to make POI recommendation. There are several advantages of the proposed recommendation methods. First, time stamps are used to subgroup original data for the purpose of deeply understanding users' behavior patterns in different temporal scenarios. Second, the models capture the spatial influence which has been proven to be very important on user's mobile behavior. Third, deeply exploiting textual information meets the criteria of a truly personalized recommender system which is recommending POIs that match users' interests. Last but not least, the proposed approaches model the textual and geographical influence in exponential form, which is suitable for the big range of users' implicit check-in feedback data. Finally, extensive experiments on real data collected through Twitter's API validated the practical utility of our proposed methods.

Our future work includes meeting the online real time needs and conducting large scale comparative experiments to further evaluate the performance.

Acknowledgments. This work was supported by the Key Projects in the National Science and Technology Pillar Program of China under grant No. 2014BAH26F02.

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. J. IEEE Trans. Knowl. Data Eng. **17**, 734–749 (2005)
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2002)

3. Cheng, C., Yang, H., King, I., Lyu, M.: Fused matrix factorization with geographical and social influence in location-based social networks. In: 26th AAAI Conference on Artificial Intelligence, pp. 17–23. AAAI Press, California (2012)
4. Cheng, Z., Caverlee, J., Lee, K., Sui, D.: Exploring millions of footprints in location sharing services. In: 5th International AAAI Conference on Weblogs and Social Media, pp. 81–88. AAAI Press, California (2011)
5. Deshpande, M., Karypis, G.: Item-based top-n recommendation algorithms. J. ACM Trans. Inf. Syst. **22**, 143–177 (2004)
6. Ference, G., Ye, M., Lee, W.: Location recommendation for out-of-town users in location-based social networks. In: 22nd ACM International Conference on Information and Knowledge Management, pp. 721–728. ACM, New York (2013)
7. Gao, H., Tang, J., Hu, X., Liu, H.: Content-aware point of interest recommendation on location-based social networks. In: 29th AAAI Conference on Artificial Intelligence, pp. 1721–1727. AAAI Press, California (2015)
8. Hu, B., Ester, M.: Spatial topic modeling in online social media for location recommendation. In: 7th ACM Conference on Recommender Systems, pp. 25–32. ACM, New York (2013)
9. Koren, Y.: Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 426–434. ACM, New York (2008)
10. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. J. Comput. **42**, 30–37 (2009)
11. Levandoski, J.J., Sarwat, M., Eldawy, A., Mokbel, M.F.: LARS: a location-aware recommender system. In: 28th IEEE International Conference on Data Engineering, pp. 450–461. IEEE Press, New York (2012)
12. Liu, B., Xiong, H.: Point-of-interest recommendation in location based social networks with topic and location awareness. In: 2013 SIAM International Conference on Data Mining, pp. 396–404. SIAM, Philadelphia (2013)
13. Liu, X., Liu, Y., Aberer, K., Miao, C.: Personalized point-of-interest recommendation by mining users preference transition. In: 22nd ACM International Conference on Information and Knowledge Management, pp. 733–738. ACM, New York (2013)
14. Ma, H., Liu, C., King, I., Lyu, M.R.: Probabilistic factor models for web site recommendation. In: 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 265–274. ACM, New York (2011)
15. Liu, X., Aberer, K.: SoCo: a social network aided context-aware recommender system. In: 22nd International Conference on World Wide Web, pp. 781–802. W3C (2013)
16. Salakhutdinov, R., Mnih, A.: Probabilistic matrix factorization. Advances in Neural Information Processing Systems (NIPS 2007), vol. 20, pp. 1257–1264. MIT Press, Massachusetts (2007)
17. Rendle, S., Marinho, L., Nanopoulos, A., Schmidt-Thieme, L.: Learning optimal ranking with tensor factorization for tag recommendation. In: 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 727–736. ACM, New York (2009)
18. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: Grouplens: an open architecture for collaborative filtering of netnews. In: 1994 ACM Conference on Computer Supported Cooperative Work, pp. 175–186. ACM, New York (1994)
19. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. J. Adv. Artif. Intell. **2009**, 1–19 (2009)

20. Ye, M., Yin, P., Lee, W.C.: Exploiting geographical influence for collaborative point-of-interest recommendation. In: 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 325–334. ACM, New York (2011)
21. Yuan, Q., Cong, G., Ma, Z., Sun, A., Thalmann, N.M.: Time-aware point-of-interest recommendation. In: 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 363–372. ACM, New York (2013)
22. Yuan, Q., Cong, G., Ma, Z., Sun, A., Thalmann, N.M.: Who, where, when and what: discover spatio-temporal topics for twitter users. In: 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 605–613. ACM, New York (2013)
23. Zheng, Y., Zhang, L., Xie, X., Ma, W.Y.: Mining interesting locations and travel sequences from GPS trajectories. In: 18th International Conference on World Wide Web, pp. 791–800. ACM, New York (2009)