Regularized Content-Aware Tensor Factorization Meets Temporal-Aware Location Recommendation

Defu Lian†, Zhenyu Zhang†, Yong Ge‡, Fuzheng Zhang§, Nicholas Jing Yuan♯, Xing Xie§
†Big Data Research Center, University of Electronic Science and Technology of China
‡Eller School of Managment, The University of Arizona
§Microsoft Research, ♯Microsoft Cooperation
Email: {dove.ustc, zhenyu.uestc, ygestrive} at gmail.com, {fuzzhang, nicholas.yuan, xingx} at microsoft.com

***Abstract*—Although weighted tensor factorization tailored to implicit feedback has shown its superior performance in temporal-aware location recommendation, it suffers from three critical challenges. First, it doesn't distinguish the confidence of negative preference for time-dependent unvisited locations from that for fully unvisited ones. Second, discontinuity arises from time discretization, and thus an infinitely large margin may exist between different bins of time. Third, geographical constraints of neighbor locations are not taken into account. To address these challenges, we propose a regularized content-aware tensor factorization (RCTF) algorithm, which exploits three strategies to address the corresponding challenges. First, it introduces a novel interaction regularization; second, it represents each bin of time by a derived feature vector from eigen decomposition of a time-bin similarity matrix, to capture the proximity of neighbor bins of time; third, it encodes geographical information of locations by discrete spatial distributions, so that spatial proximity constraints can be satisfied by simply feeding them into location content. The proposed algorithm is then evaluated for time-aware location recommendation on two large scale location-based social network datasets. The experimental results show the superiority of the proposed algorithm to several competing time-aware recommendation baselines, and verify the significant benefit of three strategies in the proposed algorithm.**

I. Introduction

The advancement and popularity of smart mobile devices make people easier to acquire real-time location information. This development has triggered the advent of location-based social networks (LBSNs), such as Foursquare, Facebook Place and so on. In these LBSNs, users can post their physical locations in the form of check-in and share their experiences at the points of interest (POIs), e.g., restaurants. These behaviors can be optionally synchronized on multiple social networks. Therefore, this emergence has not only led to location-based socializing becoming a new form of social interaction, but also has leveraged *location recommendation* to help people speed up familiarization of the surroundings.

Due to the availability of large-scale mobility history, location recommendation has been an important research topic for a long time, concentrating the aspects of spatial and (or) temporal. Although several existing work has already integrated spatio-temporal modeling with collaborative filtering [1], [2], [3], [4], there is still a lack of generic frameworks. Within such frameworks, not only should implicit feedback characteristics of mobility history be captured, as suggested in [5], [6], but also both temporal and spatial information is only encoded as features for input to achieve the same objective of spatio-temporal modeling.

A recently proposed weighted tensor factorization algorithm tailored for implicit feedback [7] makes it possible to develop such a general framework. This is because it extends weighted matrix factorization [8], [9], which treats the data as an indication of positive and negative preference with vastly varying confidence, and achieves the state-of-the-art collaborative filtering for implicit feedback [10], [5], [11]. However, it still suffers from three challenges and makes it suboptimal for temporal-aware location recommendation. First, this algorithm can not distinguish the confidence of negative preference for time-dependent unvisited locations from that for fully unvisited ones, as illustrated in Fig. 1 ("1" labeled missing entries versus "2" labeled missing entry). For example, if one user visited a restaurant A at noon and a restaurant B in the evening, but didn't pay a visit to a restaurant C at all. Then the confidence of negative preference for the restaurant B at noon could be not distinguished from that for the restaurant C. Second, discontinuity arises from time discretization, and an infinitely large margin may exist between different bins of time. This challenge makes it difficult to simultaneously alleviate data sparsity and distinguish user preference at different moments within the same time bin. Third, geographical constraints between neighbor locations are not imposed, but locations within the same areas share similar geographical attractiveness.

To this end, in this paper, we propose an efficient regularized content-aware tensor factorization framework (RCTF). Within this framework, an interaction regularization between users and locations is suggested, for the sake of efficiently dealing with the first challenge. Although it is also possible to directly distinguish their respective negative preference, it would increase time complexity of the algorithm, preventing it from putting into practical use. To deal with the second challenge, we represent each bin of time by a derived feature vector from eigen decomposition of a time-bin similarity matrix, for capturing the proximity between neighbor bins of time. Thus, the algorithm becomes insensitive to the bin size of time discretization so that a significantly smaller bin size (e.g., 15 min instead of 1 hour) can be chosen. In order to cope with the third challenge, we encode geographical information of locations by discrete spatial distribution over grids of even-size

2374-8486/16 $31.00 © 2016 IEEE
DOI 10.1109/ICDM.2016.76

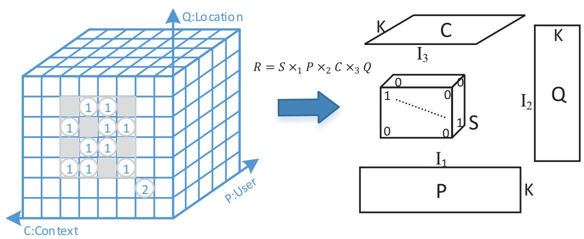

Fig. 1: Tensor Canonical Decomposition. In the left tensor, gray entries of the first user are observed, and "1" labeled entries are time dependent missing, versus "2" labeled fully missing entries.

and show that spatial proximity constraints can be satisfied by simplify feeding them into location content.

Finally, we evaluate the proposed algorithm for time-aware location recommendation on two large scale location-based social network datasets, containing over 4.6M and 2.8M check-ins of 85K and 36K users, respectively. The experimental results show the superiority of the proposed algorithm to several alternative time-aware location recommendation algorithms, including time-aware user-based collaborative filtering [1], collective time dependent matrix factorization with temporal smoothness [12] and with multi-task learning [13]. In addition, the experimental results indicate the significant benefit of interaction regularization, spatial proximity constraints and time-bin representation in the proposed algorithm.

II. Preliminary

In this paper, we study three order tensor factorization models, where a user-time-location tensor $\mathcal{R} \in \{0,1\}^{M \times C \times N}$ is provided. Here there are M users, C time bins and N locations. Each entry $r_{u,t,i}$ in the tensor $\mathcal{R}$ indicates whether a user u has visited a location i within a time bin t. Following common symbolic notation, tensors are denoted by calligraphic upper-case letters, matrices by bold upper case letters, vectors by bold lower case letters, and scalars by lower case letters. Unless specified, vectors mean column vectors.

A. Tensor Factorization for Implicit Feedback

Tensor factorization models are studied in several fields for many years. Its canonical decomposition has been exploited in [7] to extend weighted matrix factorization, and applied for temporal-aware collaborative filtering from implicit feedback. Its superiority to other loss function (e.g., ranking) based factorization models has been shown. Therefore, based on this canonical decomposition, we will propose a regularized content-aware tensor factorization model for time-aware location recommendation. In this canonical model, latent factors in three dimension are learned by optimizing the following objective function,

$$\mathcal{L}_T = \sum_{u,t,i} w_{u,t,i}\big(r_{u,t,i} - \mathbf{1}^T(\mathbf{p}_u \circ \mathbf{c}_t \circ \mathbf{q}_i)\big)^2 + \lambda(\|\mathbf{P}\|_F^2 + \|\mathbf{C}\|_F^2 + \|\mathbf{Q}\|_F^2), \quad (1)$$

where $\mathcal{W} = (w_{u,t,i})$ is a weighted tensor, whose each entry $w_{u,t,i}$ is set to α if $r_{u,t,i} > 0$ and to 1 otherwise. $\alpha \gg 1$ is usually tuned by held-out (cross) validation. Thus, if we have observed a user u has visited a location i within a time bin t, the confidence of user positive preference for locations during this time period is significantly higher than the confidence of her negative preference (i.e., in case of $r_{u,t,i} = 0$).

III. Regularized Content-Aware Tensor Factorization

This paper presents a regularized content-aware tensor factorization (RCTF) framework to address three aforementioned challenges. For the first challenge, rather than assigning lower confidences (e.g., $w_{u,t,i} = 0$) to the first type of negative preference so as to make explicit distinction, we exploit interaction regularization to constrain the dot product between users and locations. Such a strategy plays an important role in reducing time complexity of the algorithm out of the growth of tensor density. For the second challenge, we seek a unique vector representation for time bins, which simultaneously captures the continuity between neighbor bins of time. For the final challenge, we represent geographical information by location content and show that it could equivalently impose spatial proximal constraints.

A. The Objective Function

Assuming the unique representation of a time bin $t \in \{t_1, \cdots, t_C\}$ is $\mathbf{z}_t \in \mathbb{R}^D$ for general consideration and geographical information of location i is encoded as a content vector $\mathbf{y}_i \in \mathbb{R}^L$, the objective function of regularized content-aware tensor factorization is

$$\mathcal{L} = \sum_{u,t,i} w_{u,t,i}\big(r_{u,t,i} - \mathbf{1}^T(\mathbf{p}_u \circ (\mathbf{T}^T\mathbf{z}_t) \circ \mathbf{q}_i)\big)^2 + \beta \sum_{u,i} w_{u,i}(r_{u,i} - \mathbf{p}_u^T\mathbf{q}_i)^2 + \gamma \sum_i \|\mathbf{q}_i - \mathbf{V}^T\mathbf{y}_i\|^2 + \lambda(\|\mathbf{P}\|_F^2 + \|\mathbf{T}\|_F^2 + \|\mathbf{V}\|_F^2), \quad (2)$$

where $\mathbf{T} \in \mathbb{R}^{D \times K}$, $\mathbf{V} \in \mathbb{R}^{L \times K}$ corresponds a mapping matrix from time-bins' feature space and location content to the joint latent space, respectively, being aligned with users and locations. Each $r_{u,i}$ is a 0/1 variable, indicating whether a user has visited a location i and each $w_{u,i}$ indicates a confidence of user preference. Due to the existence of the third term, $\mathbf{q}_i$ in this equation is the addition of $\mathbf{q}_i$ of Eq (1) and the feature effect $\mathbf{V}^T\mathbf{y}_i$, thus absorbing the impact of geographical information.

1) Interaction Regularization: According to [14], human exhibits significant propensity of returning to previously visited locations. Therefore, the confidence of negative preference for time-dependent unvisited locations should be smaller than that for fully unvisited ones. In other words, it is higher likely for the former locations to be positively preferred within this time bin. Explicitly lowering the confidence of negative preference for time-dependent unvisited locations will increase time complexity of tensor factorization. Actually, time-dependent unvisited locations are actually visited ones. Thus their higher preference than fully unvisited ones can be satisfied together with time-dependent visited ones by the regularization term $\beta \sum_{u,i} w_{u,i}(r_{u,i} - \mathbf{p}_u^T \mathbf{q}_i)^2$ in Eq (2), subject to $w_{u,i} \gg w_{u,i'}$ if $r_{u,i} = 1$ and $r_{u,i'} = 0$. Consequently, the confidence of positive preference for visited locations (including both time-dependent unvisited and time-dependent visited ones) will be significantly larger than negative preference for fully unvisited locations. This term is called as interaction regularization since it constrains dot product between users and locations rather than the norm of user latent factors or location latent factors.

2) Seeking Representation of Time Bins: As stated, for dealing with the second challenge, we will seek a unique representation for each time bin of C bins $\{t_1, \cdots, t_C\}$, to capture the continuity between neighbor bins of time. In particular, we construct a similarity matrix between any two time bins and perform eigen decomposition on this similarity matrix. The similarity between two time bins s and t is computed as: $s^T_{s,t} = \exp -\frac{d(s,t)^2}{2\sigma_c^2}$, where $d(s,t)$ is their distance. By varying σ_c, we can control how neighbor bins of time are correlated. Based on the similarity definition, we can construct a similarity matrix $\mathbf{S}_T = (s^T_{s,t})$ of time bins. Since $\mathbf{S}_T$ is a real symmetric matrix, it can be diagonalized. Due to the small number of time bins, we can apply eigen decomposition on this similarity matrix,

$$\mathbf{S}_T = \tilde{\mathbf{Z}} \Lambda \tilde{\mathbf{Z}}^{-1} \tag{3}$$

where $\tilde{\mathbf{Z}}$ is an orthogonal matrix, subject to $\tilde{\mathbf{Z}}^T \tilde{\mathbf{Z}} = \mathbf{I}$, whose each column corresponds to eigenvectors of $\mathbf{S}_T$ and Λ is a diagonal matrix whose diagonal entries are the corresponding eigenvalues. Then we set the representation of each time bin as $\mathbf{Z} = \tilde{\mathbf{Z}} \Lambda^{\frac{1}{2}}$, whose each row is the representation of the corresponding time bin. Due to the uniqueness of eigen decomposition of full rank similarity matrix, we can get a unique representation for each time bin. One benefit of this representation is that each dimension is orthogonal to each other, making the updating formula of $\mathbf{T}$ concise.

3) Spatial Proximity Constraints: It is also possible to apply eigen decomposition on a similarity matrix $\mathbf{S}_G = [s^G_{i,j}]$ between locations for getting representation of geographical information, whose each entry is defined as

$$s^G_{i,j} = \begin{cases} e^{-\frac{d_{ij}^2}{4\sigma^2}}, & \text{if } d_{ij} < \epsilon \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

However, it suffers from computational issues due to the large size of similarity matrix. Spatial proximity constraints can be also satisfied, by imposing graph Laplacian regularization based on the location similarity matrix. However, there are three shortcomings. First, the similarity matrix is of large size, in spite of sparseness, requiring large space for storage. Second, the additions and deletions of locations must dynamically update ϵ neighborhoods of nearby locations. Third, the update of location latent factors is coupled in the alternative least square algorithm, preventing parallel update of latent factors for different locations and making it sensitive to the order of update.

Fortunately, according to [6], location content could refine the similarity between locations in terms of mobility history based on (normalized) dot product. This observation will be validated in Section III-B. Therefore, in this paper, we represent geographical information of locations by discrete spatial distribution over grids. Then spatial proximity constraints could be satisfied as long as the cosine similarity of spatial distribution between locations could be shown to well approximate location similarity in Eq (4). To this end, we first define discrete spatial distribution, and then present one theoretical result.

Definition 1. ***Discrete Spatial Distribution*** *with respect to a location i is defined over a sample space of spatial grids $\{l_1, \cdots, l_G\}$ of even-size, where $Pr_i(L = l)$ is the probability that influent zone of the location falls into a specific grid l. Assuming influent zone density follows normal distribution over a two-dimensional space, the probability is defined as*

$$Pr_i(L = l_g) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{d_{ig}^2}{2\sigma^2}} \Delta A, \tag{5}$$

where ΔA is the area of any grid of sufficiently small size and d_{ig} is the distance between the location i and the grid l_g.

According to this definition, we could represent each location i by a probability vector $\mathbf{y}_i = \{Pr_i(L = l_g)\}$. Then the relationship between dot product of locations' spatial distribution and their Euclidean distance could be shown in the following theorem:

Theorem 1. *When spatial grids are of infinitesimal size, if restricting the influent zones within $r = \frac{\epsilon}{2}$ from each location and assuming the influence at the border zones approaches to zero (i.e., $e^{-\frac{r^2}{\sigma^2}} \to 0$), then $\mathbf{y}_i^T \mathbf{y}_j \approx \frac{\Delta A}{2} s^G_{i,j}$.*

This theorem states that the location similarity matrix $\mathbf{S}^G$ could be approximately decomposed as $\mathbf{S_G} \approx \frac{2}{\Delta A} \mathbf{Y}\mathbf{Y}'$ based on the spatial distribution representation. In practical, grids cannot be of infinitesimal size, thus the approximation error may be larger but much more difficult to analyze.

B. Optimization

All the parameters except $\mathbf{T}$ in Eq (2) can be learned effectively by alternative least square optimization algorithm according to [6]. For the sake of learning $\mathbf{T}$, we introduce a variational matrix variable $\mathbf{C} \in \mathbb{R}^{C \times K}$, subject to a linear equality constraint $\mathbf{C} = \mathbf{Z}\mathbf{T}$, where $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_C]^T \in \mathbb{R}^{C \times D}$. Note that when $\mathbf{Z} = \tilde{\mathbf{Z}} \Lambda^{\frac{1}{2}}$, $\mathbf{Z}$ is a squared matrix, i.e., $C = D$. Based on the method of Lagrange multipliers, we can rewrite the objective function of Eq (2) as follows:

$$\begin{aligned} \mathcal{L} = & \sum_{u,t,i} w_{u,t,i} \big(r_{u,t,i} - 1^T(\mathbf{p}_u \circ \mathbf{c}_t \circ \mathbf{q}_i)\big)^2 \\ & + \beta \sum_{u,i} w_{u,i}(r_{u,i} - \mathbf{p}_u^T \mathbf{q}_i)^2 + \gamma \sum_i \|\mathbf{q}_i - \mathbf{V}^T \mathbf{y}_i\|^2 \\ & + \mathrm{tr}(\mathbf{\Theta}^T(\mathbf{C} - \mathbf{Z}\mathbf{T})) + \lambda(\|\mathbf{P}\|_F^2 + \|\mathbf{T}\|_F^2 + \|\mathbf{V}\|_F^2), \end{aligned} \tag{6}$$

where $\mathbf{\Theta} \in \mathbb{R}^{C\times K}$ is a Lagrange multiplier, and perform alternative least square optimization with respect to all original parameters and two extras.

With respect to optimizing $\mathbf{p}_u$, setting the gradient of $\mathcal{L}$ w.r.t $\mathbf{p}_u$ to zero, we can obtain,

$$\mathbf{p}_u = (\sum_{t,i} w_{u,t,i}(\mathbf{c}_t \circ \mathbf{q}_i)(\mathbf{c}_t \circ \mathbf{q}_i)^T + \beta \sum_i w_{u,i}\mathbf{q}_i\mathbf{q}_i^T + \lambda I_K)^{-1} (\sum_{t,i} w_{u,t,i} r_{u,t,i}(\mathbf{c}_t \circ \mathbf{q}_i) + \beta \sum_i w_{u,i} r_{u,i}\mathbf{q}_i) \quad (7)$$

where $\sum_{t,i} w_{u,t,i}(\mathbf{c}_t \circ \mathbf{q}_i)(\mathbf{c}_t \circ \mathbf{q}_i)^T$ can be efficiently computed based on $\sum_{t,i}(w_{u,t,i}-1)(\mathbf{c}_t \circ \mathbf{q}_i)(\mathbf{c}_t \circ \mathbf{q}_i)^T + (\mathbf{C}^T\mathbf{C}) \circ (\mathbf{Q}^T\mathbf{Q})$, since $(\mathbf{c}_t \circ \mathbf{q}_i)((\mathbf{c}_t \circ \mathbf{q}_i))^T = (\mathbf{c}_t\mathbf{c}_t^T) \circ (\mathbf{q}_i\mathbf{q}_i^T)$.

Therefore, assuming the number of non-zero entries of tensor $\mathcal{R}$ is N_R, the time complexity of sequentially updating latent factors of all users is $\mathcal{O}(N_R K^2 + MK^3)$, since N_R is usually larger than the number of non-zero entries in the matrix $\mathbf{R}$. Due to the independence of update between different users, it is possible to greatly accelerate this procedure based on parallel computing.

With respect to optimizing $\mathbf{q}_i$, we can obtain the update formula by setting the gradient of $\mathcal{L}$ w.r.t $\mathbf{q}_i$ to zero,

$$\mathbf{q}_i = (\sum_{u,t} w_{u,t,i}(\mathbf{p}_u \circ \mathbf{c}_t)(\mathbf{p}_u \circ \mathbf{c}_t)^T + \beta \sum_u w_{u,i}\mathbf{p}_u\mathbf{p}_u^T + \gamma I_K)^{-1} (\sum_{u,t} w_{u,t,i} r_{u,t,i}(\mathbf{p}_u \circ \mathbf{c}_t) + \beta \sum_u w_{u,i} r_{u,i}\mathbf{p}_u + \gamma \mathbf{V}^T\mathbf{y}_i) \quad (8)$$

Based on similar analysis, time complexity of sequentially updating latent factors of all locations is $\mathcal{O}(N_R K^2 + NK^3)$.

With respect to learning $\mathbf{V}$, we can see that it is equivalent to minimizing the objective function $\sum_i \|\mathbf{q}_i - \mathbf{V}^T\mathbf{y}_i\|^2 + \frac{\lambda}{\gamma}\|\mathbf{V}\|_F^2$. Therefore, the solution of $\mathbf{V}$ is

$$\mathbf{V} = (\mathbf{Y}^T\mathbf{Y} + \frac{\lambda}{\gamma}\mathbf{I}_L)^{-1}\mathbf{Y}^T\mathbf{Q}. \quad (9)$$

Since the location-zone influence matrix $\mathbf{Y}$ is a large but sparse matrix, the inversion of $\mathbf{Y}^T\mathbf{Y} + \frac{\lambda}{\gamma}\mathbf{I}_L$ may suffer from computational issues. Therefore, we resort to conjugate gradient descent algorithm for solving this linear equation system. The complexity is $\mathcal{O}(\|\mathbf{Y}\|_0 K \#iter)$, where #iter is the number of iterations of conjugate gradient descent to reach a given threshold of approximation error.

Remark of Eq (9): According to the matrix inversion lemma, $\mathbf{V} = \mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T + \frac{\lambda}{\gamma}\mathbf{I}_L)^{-1}\mathbf{Q}$. Since $\mathbf{Y}\mathbf{Y}^T$ approximates to $\frac{\Delta A}{2}\mathbf{S}_G$, $\mathbf{V}^T\mathbf{y}_i \approx \mathbf{Q}^T(\mathbf{S}^G + \frac{2\lambda}{\gamma\Delta A}\mathbf{I}_L)^{-1}\mathbf{s}_i^G$, where $\mathbf{s}_i^G$ is the i^{th} row of the matrix $\mathbf{S}^G$. Thus, the term $\gamma\mathbf{V}^T\mathbf{y}_i$ in Eq (9) captures the effect of nearby locations, verifying that location content refines mobility similarity between locations, as stated in [6].

With respect to learning $\mathbf{T}$, first setting the gradient of $\mathcal{L}$ with respect to $\mathbf{T}$ to zero, we have $\mathbf{Z}^T\mathbf{\Theta} = 2\lambda\mathbf{T}$. Left multiplying both side with $\mathbf{Z}$ and applying the equality constraint, we have the solutions for $\mathbf{\Theta}$ and $\mathbf{T}$, i.e.,

$$\begin{aligned}\mathbf{\Theta} &= 2\lambda(\mathbf{Z}\mathbf{Z}^T)^{-1}\mathbf{Z}\mathbf{T} = 2\lambda(\mathbf{Z}\mathbf{Z}^T)^{-1}\mathbf{C}, \\ \mathbf{T} &= \mathbf{Z}^T(\mathbf{Z}\mathbf{Z}^T)^{-1}\mathbf{C}\end{aligned} \quad (10)$$

where $\mathbf{Z}\mathbf{Z}^T = \mathbf{S}$ is of full rank and thus invertible. Based on the newer $\mathbf{\Theta}$, we can update $\mathbf{c}_t$ based on

$$\mathbf{c}_t = (\sum_{u,i} w_{u,t,i}(\mathbf{p}_u \circ \mathbf{q}_i)(\mathbf{p}_u \circ \mathbf{q}_i)^T + \epsilon\mathbf{I}_K)^{-1} (\sum_{u,i} w_{u,t,i} r_{u,t,i}(\mathbf{p}_u \circ \mathbf{q}_i) - \frac{\boldsymbol{\theta}_t}{2}) \quad (11)$$

where $\epsilon\mathbf{I}_K$ is placed for numerical stability. Since $\boldsymbol{\theta}_t$ absorbs latent factors of other bins of time by Eq (10), we can deal with the discontinuity problem. Based on the same analysis as above, the time complexity of updating latent factors of all bins of time is $\mathcal{O}(N_R K^2 + TK^3)$, dominating the complexity of updating $\mathbf{T}$ and $\mathbf{\Theta}$, due to the small number of time bins.

IV. Experiments

In this section, we first introduce the datasets and present the preprocessing procedure. We then discuss evaluation protocols and illustrate the assessment metric, followed by experimental results and discussions.

A. Dataset and Preprocessing

We evaluate the proposed algorithm for time-aware location recommendation, where we consider daily dynamics but ignore the inter-days difference, and split a full day into T time periods $\{t_1, \cdots, t_T\}$. Note that we need to consider a cycle property when measuring distance (i.e., time intervals) between time periods. For example, $d(23h, 0h) = 1h$. Such an evaluation is then conducted on two large-scale location-based social network datasets. The first dataset, which was used in [15] and crawled from Gowalla, contains *6,423,854* check-ins at *1,280,969* locations from *107,092* users from *Feb. 2009* to *Oct. 2010*, where each user has *60* check-ins and checks in at *37* locations on average. We select locations which has been visited by at least 5 users and select users who have checked in at least five locations. The statistics of the filtered dataset is shown in Table I. The other dataset is what we crawled from a Chinese location-based social network – Jiepang. Out of privacy concern, users' historical check-ins are not publicly available, but they may be shared as tweets on Sina Weibo (China's Twitter). Thus we crawled *3,464,798* location check-ins at *213,684* locations in Beijing from *55,650* users via Sina Weibo open APIs from *Mar. 2011* to *Mar. 2013*, where each user has *80* check-ins and checks in at *47* locations on average. After applying the same filtering for users and locations, we also show dataset statistics in Table I.

TABLE I: Data statistics

	#checkins	#locations	#users	density
Gowalla	4,656,469	308,957	85,034	1.8×10^{-4}
Jiepang	2,845,018	50,005	36,545	1.6×10^{-3}

B. Evaluation Protocols

Each user's check-in history is sorted in a chronological order, and take the preceding 80% check-in history into a training dataset and the left 20% into a testing dataset. Then testing check-ins, which have already been appeared in the training dataset, will be removed from the testing dataset, since repetitive behavior is usually not the focus of recommendation.

We then trained recommendation algorithms on training check-ins and evaluated them on testing check-ins. Presenting each user within each time period with the top-k candidate locations sorted by their prediction preference, we assess recommendation performance by checking how many of these locations actually appeared in each user's mobility data. Recall at cut-off k is such a performance assessment metric, and has been widely used in location recommendation [16], [17], [5], [11]. Formally, in case of time-aware location recommendation, it is defined as follows:

$$recall@k = \frac{1}{M}\sum_{u=1}^{M}\frac{\sum_{t=1}^{T}|\mathbb{S}_u^{(t)}(k)\cap\mathbb{V}_u^{(t)}|}{\sum_{t=1}^{T}|\mathbb{V}_u^{(t)}|},$$

where $\mathbb{S}_u^{(t)}(k)$ is the collection of top k recommended locations for a user u within a time period t and $\mathbb{V}_u^{(t)}$ is the set of her visited locations within the time period t.

The recommendation algorithms to train include not only the proposed algorithm, **RCTF**, but also the following competing baselines.

- **UTE** is time-aware user-based collaborative filtering [1], which takes temporal effect into account when computing user similarity.
- **WMF** is weighted matrix factorization, proposed for one-class collaborative filtering, tailored to recommendation based on implicit feedback.
- **CTMF-MTL** is collective time-dependent matrix factorization with multi-task learning [13].
- **LRT** is collective time-dependent non-negative matrix factorization with temporal regularization [12].
- **LTCR** is a Location and Time aware Collaborative Retrieval model [18], which leverages Weighted Approximately Ranked Pairwise (WARP) loss with the aim of better top-n ranking results.

After comparing the proposed algorithm with the competing baselines, we then study the effect of interaction regularization, spatial proximity constraints.

C. Parameter Setting

The parameters of the proposed model were individually tuned by held-out validation on the training dataset based on a grid search. Then, on the Gowalla dataset, K is set as 40, α in the weighted tensor $\mathcal{W}$ as 2100, $w_{u,i}$ in the weighting matrix $\mathbf{R}$ as 800 if $r_{u,i} > 0$, the coefficient β of interaction regularization as 100, the coefficient γ of spatial proximity constraints as around 5000, and regularization coefficient of user, time, spatial grid is 0.01, 0.01, and 100, respectively. On the Jiepang dataset, K is also set as 40, but α in the weighted tensor $\mathcal{W}$ as 300, $w_{u,i}$ in the weighting matrix $\mathbf{R}$ as 10 if $r_{u,i} > 0$, the coefficient β of interaction regularization as 10, the coefficient γ of spatial proximity constraints as

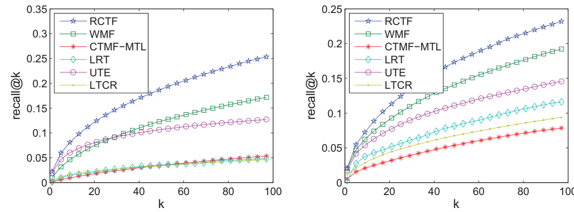

Fig. 2: Comparison with baselines (Gowalla & Jiepang)

2000, and regularization coefficient of user, time, spatial grid is 0.01,0.01, and 500, respectively.

D. Experimental results

1) Compare with baselines: The comparison of the proposed algorithm with the competing baselines are shown in Fig. 2 for the Gowalla dataset (left) and the Jiepang dataset (right). From these two figures, we can have the following three major observations. First, *RCTF greatly outperforms all the baselines on both datasets*. Its superiority to UTE indicates the superiority of factorization models to user-based collaborative filtering. Its advantage over CTMF-MTL and LRT shows the benefit of tensor factorization compared to collective matrix factorization. The major reason lies in its nature encoding for similarity between different time periods due to factorization characteristics, as long as there are sufficient mobility history within these time periods. In contrast, in collective matrix factorization, multi-task learning only enforces time-dependent latent factors of each user to be similar with some discrepancy; temporal regularization incorporates neighborhood-based methodology, making it difficult to balance between defining precise temporal-aware user similarity and addressing the sparsity challenge. Second, *LRT is not as good as CTMF-MTL and WMF on both datasets*, indicating the superiority of weighted loss function in case of recommendation for implicit feedback datasets, in accordance with previous findings [11], [5]. Third, *time-aware user-based collaborative filtering shows good recommendation performance on both datasets*, agreeing with the conclusion in [16], [5], but suffers from computational issues due to user similarity computation, in particular, when taking temporal smoothness into account. Finally, *LTCR doesn't achieve the better top-k ranking performance as expected on both datasets*. One difference from RTCF mainly lies in the loss function, where LTCR exploits WARP while RCTF leverage weighted squared loss. Thus the superiority of RTCF to LTCR illustrates the advantage of weighted squared loss for location recommendation.

2) Benefit of interaction regularization: In order to study the effect of interaction regularization, we show the performance of three algorithms, Regularized Weighted Tensor Factorization (*RWTF*), Weighted Tensor Factorization (*WTF*) and *WMF*, in Fig. 3, for the Gowalla dataset (left) and for the Jiepang dataset (right). RWTF is a simplified version of RCTF, but without taking similarity between time periods into account and without imposing spatial proximity constraints. Compared to RWTF, there is no interaction regularization in WTF. By comparing RWTF with WTF, we can see the benefit of interaction regularization, indicating the effect of distinguishing time-dependent unvisited locations from completely unvisited locations. In other words, in WTF, due to increasing number of local optimums rooting from the severer sparsity challenge [19], it may be much easy for latent factors of users and locations to be caught in local optimums.

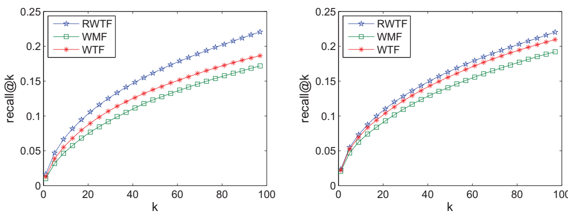

Fig. 3: Effect of interaction regularization (Gowalla & Jiepang)

3) Benefit of spatial proximity constraints: We study the potential benefit of spatial proximity constraints, by comparing RCTF with the counterpart without imposing spatial proximity constraints, i.e., RCTF (w/o geo). The results are showed in Fig. 4 for the Gowalla dataset (left) and for the Jiepang dataset (right). From these two figures, we can observe the significant effect of imposing spatial proximity constraints, in particular on the Gowalla dataset, illustrating the rationality of imposing spatial proximity constraints by feeding spatial distribution into location content. However, the benefit of spatial proximity constraints on the Jiepang dataset is comparatively smaller. The major reason lies in data validity, since locations checked in by Jiepang users don't satisfy strict distance restriction. In other words, they can check in anywhere they love, in particular after some time of usage.

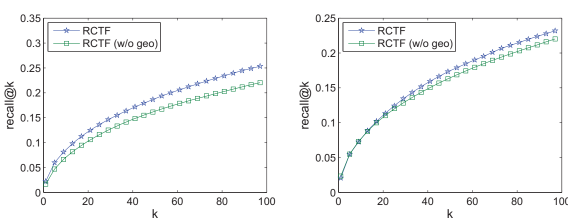

Fig. 4: Effect of spatial constraints (Gowalla & Jiepang)

V. Conclusions

In this paper, we first reveal three critical challenges in weighted tensor factorization tailored to recommendation based on implicit feedback, and then propose an efficient regularized content-aware tensor factorization algorithm, which exploits three strategies for addressing these three challenges. We finally evaluate the proposed algorithm on two large scale mobility datasets. The experimental results not only indicate the superiority of the proposed algorithm to the competing baselines for time-aware location recommendation, but also validate the benefit of these three strategies: interaction regularization, spatial proximity constraints and time period representation, indicating the rationality of introducing these three characteristics.

Acknowledgement. This work is supported by grants from the Natural Science Foundation of China (61502077, 61631005) and the Fundamental Research Funds for the Central Universities (ZYGX2014Z012).

References

[1] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "Time-aware point-of-interest recommendation," in *Proceedings of SIGIR'13*. ACM, 2013, pp. 363–372.

[2] X. Li, G. Cong, X. Li, T.-A. N. Pham, and S. Krishnaswamy, "Rank-geofm: A ranking based geographical factorization method for point of interest recommendation," in *Proceedings of SIGIR'15*. ACM, 2015, pp. 433–442.

[3] H. Yin, X. Zhou, Y. Shao, H. Wang, and S. Sadiq, "Joint modeling of user check-in behaviors for point-of-interest recommendation," in *Proceedings of CIKM'15*. ACM, 2015, pp. 1631–1640.

[4] D. Lian, X. Xie, V. W. Zheng, N. J. Yuan, F. Zhang, and E. Chen, "Cepr: A collaborative exploration and periodically returning model for location prediction," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 1, 2015.

[5] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui, "Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation," in *Proceedings of KDD'14*. ACM, 2014, pp. 831–840.

[6] D. Lian, Y. Ge, F. Zhang, N. J. Yuan, X. Xie, T. Zhou, and Y. Rui, "Content-aware collaborative filtering for location recommendation based on human mobility data," in *Proceedings of ICDM'15*. IEEE, 2015, pp. 261–270.

[7] B. Hidasi and D. Tikk, "Fast als-based tensor factorization for context-aware recommendation from implicit feedback," in *Machine Learning and Knowledge Discovery in Databases*. Springer, 2012, pp. 67–82.

[8] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative filtering for implicit feedback datasets," in *Proceedings of ICDM'08*. IEEE, 2008, pp. 263–272.

[9] R. Pan, Y. Zhou, B. Cao, N. Liu, R. Lukose, M. Scholz, and Q. Yang, "One-class collaborative filtering," in *Proceedings of ICDM'08*. IEEE, 2008, pp. 502–511.

[10] C. Wang and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," in *Proceedings of KDD'11*. ACM, 2011, pp. 448–456.

[11] Y. Liu, W. Wei, A. Sun, and C. Miao, "Exploiting geographical neighborhood characteristics for location recommendation," in *Proceedings of CIKM'14*. ACM, 2014, pp. 739–748.

[12] H. Gao, J. Tang, X. Hu, and H. Liu, "Exploring temporal effects for location recommendation on location-based social networks," in *Proceedings of RecSys'13*. ACM, 2013, pp. 93–100.

[13] Y. Wang, N. J. Yuan, D. Lian, L. Xu, X. Xie, E. Chen, and Y. Rui, "Regularity and conformity: Location prediction using heterogeneous mobility data," in *Proceedings of KDD'15*. ACM, 2015, pp. 1275–1284.

[14] C. Song, T. Koren, P. Wang, and A. Barabási, "Modelling the scaling properties of human mobility," *Nature Physics*, vol. 6, no. 10, pp. 818–823, 2010.

[15] E. Cho, S. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proceedings of KDD'11*, 2011, pp. 1082–1090.

[16] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *Proceedings of SIGIR'11*. ACM, 2011, pp. 325–334.

[17] B. Liu, Y. Fu, Z. Yao, and H. Xiong, "Learning geographical preferences for point-of-interest recommendation," in *Proceedings of KDD'13*. ACM, 2013, pp. 1043–1051.

[18] W. Zhang and J. Wang, "Location and time aware social collaborative retrieval for new successive point-of-interest recommendation," in *Proceedings of CIKM'15*. ACM, 2015, pp. 1221–1230.

[19] N. Srebro and T. Jaakkola, "Weighted low-rank approximations," in *Proceedings of ICML'03*, 2003, pp. 720–727.