Context-aware Point-of-Interest Recommendation Using Tensor Factorization with Social Regularization

Lina Yao, Quan Z. Sheng, Yongrui Qin, Xianzhi Wang, Ali Shemshadi, and Qi He
School of Computer Science, The University of Adelaide
Adelaide, SA 5005, Australia
{firstname.lastname}@adelaide.edu.au

ABSTRACT

Point-of-Interest (POI) recommendation is a new type of recommendation task that comes along with the prevalence of location-based social networks in recent years. Compared with traditional tasks, it focuses more on personalized, context-aware recommendation results to provide better user experience. To address this new challenge, we propose a Collaborative Filtering method based on Nonnegative Tensor Factorization, a generalization of the Matrix Factorization approach that exploits a high-order tensor instead of traditional User-Location matrix to model multi-dimensional contextual information. The factorization of this tensor leads to a *compact* model of the data which is specially suitable for context-aware POI recommendations. In addition, we fuse users' social relations as regularization terms of the factorization to improve the recommendation accuracy. Experimental results on real-world datasets demonstrate the effectiveness of our approach.

Categories and Subject Descriptors

H.2.8 [**Data Management**]: Database Applications—*Data mining*; H.3.5 [**Information Storage and Retrieval**]: General

Keywords

Tensor factorization; social regularization; location based social networks; recommendation

1. INTRODUCTION

With the popularization of mobile devices, wireless networks and location-enabling techniques, location-based social networks (LBSNs), such as Foursquare, Gowalla, and Brightkite, have been attracting millions of users. People are increasingly using LBSNs services to connect with friends, explore places (e.g., restaurants, shops, cinemas etc.), and share their locations via check-in activities, which contain rich clues of users' preference on locations [5, 2, 6]. We take Brightkit check-in data[1] as an example to show some key patterns:

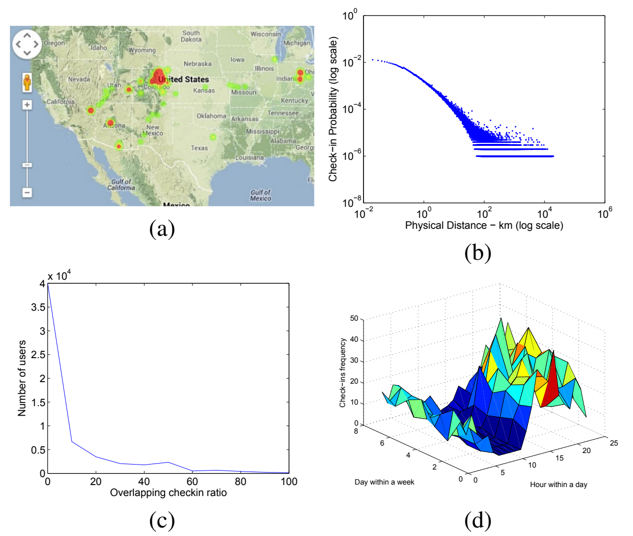


Figure 1: **Illustration of check-in patterns in terms of geographical pattern (a) and (b), friendship impact (c) and temporal pattern of one specific user (d) over Brightkit**

- Check-in activities have shown some interesting geographical characteristics such as spatial clustering patterns (as shown in Figure 1 (a)). Generally, people tend to visit nearby locations of their homes or offices, which are their POIs, and may also be interested in visiting the nearby locations of these POIs, even if they become far away from their homes [3, 11]. In addition, the majority of POIs checked in by the same users tend to aggregate within a certain range of short distances (as shown in Figure 1 (b));
- Users' social relationships have shown impact on checking-in activities. A dominant percentage of users have less than 10% overlapping locations with their friends (as shown in Figure 1 (c));
- Check-in data also show the periodical features depending on the type of POIs. For example, restaurants' peak time may be in the lunch hours, yet nightclubs or cinemas are most probably during nights and weekends (Figure 1 (d)).

Such availability of check-in data with rich spatial-temporal-social information makes it possible to design the context-aware location recommendation applications. In view of such informative patterns in check-in information, many significant works have been con-

[1] http://snap.stanford.edu/data/loc-brightkite.html

SIGIR'15 August 09 - 13, 2015, Santiago, Chile
©2015 ACM ISBN 978-1-4503-3621-5/15/08 ...$15.00
DOI: http://dx.doi.org/10.1145/2766462.2767794.

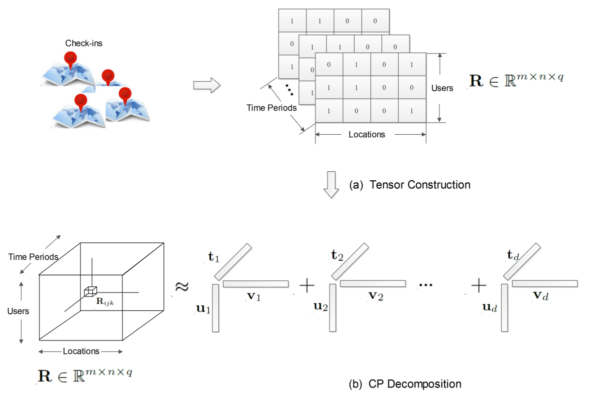


Figure 2: (a) Tensor Construction; (b) CP Decomposition

ducted on POI recommendation by analyzing users' check-in history and social constraints patterns [3, 12, 5]. In particular, Ye et al. [11] apply a power-law probabilistic model to capture the geographical influence among users' check-ins. Zhang et al. [13] consider social and geographical influence from both user and location perspectives, and develop a recommendation model by fusing Kernel density estimation into the matrix factorization framework. Yuan et al. [12] propose a collaborative recommendation model by incorporating temporal information. In a very recent work, Hu et al. [6] propose to improve the recommendation accuracy by incorporating the geographical neighborhood information.

Instead of modeling check-ins as a traditional two-dimensional user-location matrix, in this paper, we address the POI recommendation problem by incorporating the multi-dimensional contextual information of check-in data with high order tensor factorization to uncover the hidden dependency of multi-dimensional contextual information. Karatzoglou et al. [7] also propose a tensor-based context-aware recommendation framework using tucker decomposition. However, they did not consider the internal correlations among the contextual entities. In contrast to their work, we employ three order tensor to uncover the hidden dependency of multi-dimensional contextual information, and further explore the social influence among users, to support more accurate POI recommendation.

This paper makes the following contributions:

- We propose using high order tensor to interpret the multi-dimensional contextual information of check-in data in a compact manner. In particular, we further impose users' social connections as an extra regularization to improve recommendation accuracy.
- We evaluate our proposed method using real datasets, and the experimental results demonstrate the effectiveness of our approach.

Notations. In the rest of this paper, scalars will be denoted by lowercase letters, (e.g., m), vectors by boldface lowercase letters (e.g., $\mathbf{u}$), matrices by boldface capital letters (e.g., $\mathbf{U}$), and tensors by Calligraphy (e.g., $\mathcal{R}$). We will use $\circ$ to denote the outer product, $\odot$ the Khatri-Rao operation, and $||\cdot||_F$ the Frobenius norm.

2. PROPOSED METHOD

The POI recommendation problem in this paper can be defined as: given the historical check-in records of m users $\{u_i\}_{i=1}^m$ on n locations $\{v_j\}_{j=1}^n$ and q timeframes $\{t_k\}_{k=1}^q$, recommending the target users a set of locations that they might be interested in. We model the check-in tensor $\mathcal{R} \in \mathbb{R}^{m\times n\times q}$ via third order tensor, where each $\mathcal{R}_{ijk}$ quantifies users' preference in terms of frequency, i.e., the times that user u_i visits location v_j within time slice t_k. The tensor construction from check-in records $\mathcal{R} \in \mathbb{R}^{m\times n\times q}$ can be defined as the check-in frequency of if user u_i visits location v_j within time frame t_k, where $\mathcal{R}_{ijk}$ denotes user u_i checks in at location v_j within time period t_k (Figure 2 (a)). To address the temporal dependency of users' dynamic check-ins over time, the third order tensor $\mathcal{R}$ can be estimated from check-in records using tensor decomposition techniques, such as the High Order Singular Value Decomposition (HOSVD) and CANDECOMP/PARAFAC (CP) decomposition [4]. In this paper, we employ CP decomposition D-component of rank 1 tensors [8] to characterize the three dimensional check-in records (Figure 2 (b)), the preference of user u_i on location l_j in time frame t_k can be approximated as: $\hat{\mathcal{R}} \approx \sum_{d=1}^{D} \mathbf{u}_d \circ \mathbf{v}_d \circ \mathbf{t}_d$ where $\hat{\mathcal{R}}$ denotes the predicted approximation of $\mathcal{R}$; $\mathbf{u}_d \in \mathbb{R}^m$, $\mathbf{v}_d \in \mathbb{R}^n$ and $\mathbf{t}_d \in \mathbb{R}^q$. We aim at finding the decomposition $\hat{\mathbf{R}}$ that best approximates the original tensor $\mathcal{R}$ to achieve best recommendation results. In particular, we solve the optimization problem by minimizing the squared loss as follows.

$$\begin{aligned} L(\mathbf{U},\mathbf{V},\mathbf{T}) &= \frac{1}{2}\min_{\mathbf{U},\mathbf{V},\mathbf{T}} ||\mathcal{R}-\hat{\mathcal{R}}||_F^2 \\ &= \frac{1}{2}\min_{\mathbf{U},\mathbf{V},\mathbf{T}} ||\mathcal{R}-\sum_{d}^{D}\mathbf{u}_d \circ \mathbf{v}_d \circ \mathbf{t}_d||_F^2 \end{aligned} \tag{1}$$

where $\mathbf{U} = [\mathbf{u}_1, ..., \mathbf{u}_D] \in \mathbb{R}^{m\times D}$, $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_D] \in \mathbb{R}^{n\times D}$, and $\mathbf{T} = [\mathbf{t}_1, ..., \mathbf{t}_D] \in \mathbb{R}^{q\times D}$ are all factor matrices. To avoid overfitting, the regularization terms associated with $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{T}$ are introduced into Equation 1 as:

$$\min_{\mathbf{U},\mathbf{V},\mathbf{T}} \frac{1}{2}||\mathcal{R}-\hat{\mathcal{R}}||_F^2 + \frac{\lambda}{2}(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2 + ||\mathbf{T}||_F^2) \tag{2}$$

where λ is the regularization parameter. We further consider users' explicit social friendships to improve recommendation accuracy, which has been widely used in two-dimensional matrix factorization based recommendation frameworks [9]. The social regularization term can constrain the matrix factorization objective function and indirectly propagate users' preferences.

The Equation 2 can be reformulated as:

$$\begin{aligned} L(\mathbf{U},\mathbf{V},\mathbf{T}) &= \min_{\mathbf{U},\mathbf{V},\mathbf{T}} \frac{1}{2}||\mathcal{R}-\hat{\mathcal{R}}||_F^2 + \frac{\lambda}{2}(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2 + ||\mathbf{T}||_F^2) \\ &\quad + \frac{\alpha}{2}\sum_{i,j=1}^{s}\mathbf{A}_{i,j}\sum_{d}^{D}([\mathbf{U}_{i,d}] - [\mathbf{U}_{j,d}])^2 \\ &= \min_{\mathbf{U},\mathbf{V},\mathbf{T}} \frac{1}{2}||\mathcal{R}-\hat{\mathcal{R}}||_F^2 + \frac{\lambda}{2}(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2 + ||\mathbf{T}||_F^2) \\ &\quad + \frac{\alpha}{2}\sum_{d}^{D}\mathbf{U_{*d}}^T\mathbf{L}\mathbf{U_{*d}} \\ &= \min_{\mathbf{U},\mathbf{V},\mathbf{T}} \frac{1}{2}||\mathcal{R}-\hat{\mathcal{R}}||_F^2 + \frac{\lambda}{2}(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2 + ||\mathbf{T}||_F^2) \\ &\quad + \frac{\alpha}{2}\mathrm{tr}(\mathbf{U}^T\mathbf{L}\mathbf{U}) \\ &= \min_{\mathbf{U},\mathbf{V},\mathbf{T}} \frac{1}{2}||\mathcal{R}-\hat{\mathcal{R}}||_F^2 + \frac{1}{2}\mathrm{tr}[\mathbf{U}^T(\lambda\mathbf{I}+\alpha\mathbf{L})\mathbf{U}] + \frac{\lambda}{2}(||\mathbf{V}||_F^2 + ||\mathbf{T}||_F^2) \\ &= \min_{\mathbf{U},\mathbf{V},\mathbf{T}} \frac{1}{2}||\mathcal{R}-\hat{\mathcal{R}}||_F^2 + \frac{1}{2}\mathrm{tr}[\mathbf{U}^T(\lambda\mathbf{I}+\alpha\mathbf{L})\mathbf{U}] + \frac{\lambda}{2}\mathrm{tr}[\mathbf{V}\mathbf{V}^T + \mathbf{T}\mathbf{T}^T] \end{aligned} \tag{3}$$

where $\mathbf{A}_{i,j}$ indicates the similarity between user i and j, $\mathbf{L}$ is the Laplacian matrix induced from users' social networks matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$; $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is the diagonal matrix whose i-th diagonal element is the sum of all the elements in the i-th row of $\mathbf{A}$, i.e., $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$.

$$\begin{aligned}
\nabla_{\mathbf{U}} L(\mathbf{U}, \mathbf{V}, \mathbf{T}) &= \nabla_{\mathbf{U}} ||\mathcal{R}_{(1)} - \hat{\mathbf{U}}(\mathbf{V} \odot \mathbf{T})^T||_F^2 + \frac{\lambda}{2}\mathbf{I} + \frac{\alpha}{2}\mathbf{L} \\
\nabla_{\mathbf{V}} L(\mathbf{U}, \mathbf{V}, \mathbf{T}) &= \nabla_{\mathbf{V}} ||\mathcal{R}_{(2)} - \hat{\mathbf{V}}(\mathbf{U} \odot \mathbf{T})^T||_F^2 + \frac{\lambda}{2}\mathbf{I} \\
\nabla_{\mathbf{T}} L(\mathbf{U}, \mathbf{V}, \mathbf{T}) &= \nabla_{\mathbf{T}} ||\mathcal{R}_{(3)} - \hat{\mathbf{T}}(\mathbf{U} \odot \mathbf{V})^T||_F^2 + \frac{\lambda}{2}\mathbf{I}
\end{aligned} \tag{4}$$

where $\mathcal{R}_{(1)} \approx \mathbf{U}(\mathbf{V} \odot \mathbf{T})^T$, $\mathcal{R}_{(2)} \approx \mathbf{V}(\mathbf{U} \odot \mathbf{T})^T$ and $\mathcal{R}_{(3)} \approx \mathbf{T}(\mathbf{U} \odot \mathbf{V})^T$. As Equation 3 is not convex, we adopt the alternating optimization strategy to solve the objective function. In particular, we alternately optimize each of the three parameters $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{T}$, while fixing the other two until convergence to find the optimal solution using Equation 4.

3. EXPERIMENTS

3.1 Experimental Settings

Dataset. We select the check-in occurred during January 2010 to September 2010 from the original Brightkite [3], we remove users whose checked-in records are fewer than 5 POIs, and then removed POIs with fewer than 5 users checked in. For each user, we randomly mark off another 20% of POIs as testing data to evaluate the effectiveness of the recommendation methods. We use *Precision@x* and *Recall@x* to evaluate our proposed method (x= 5, 10, 15, 20).

Comparison Methods. We compare the proposed models **TenInt** with the following other methods, which basically consist of two categories: (1) *Non-contexts*: we implement basic non-negative matrix factorization (**NMF**), user-based collaborative filtering (**UCF**) and item-based collaborative filtering (**ICF**) [10]; (2) *Partial contexts*: we develop three models by taking into geographical influence, temporal information and friendship into account, respectively.

- **NMF.** It only considers the 2D user-location matrix. It applies non-negative matrix factorization on user-location matrix to predict the possibility of check-in. The user-location matrix can be decomposed into two lower dimension matrices in this method:

$$\min_{\mathbf{U},\mathbf{V}} \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} (r_{ij} - \mathbf{u}_i^T \mathbf{v}_j)^2 \tag{5}$$

- **User-CF (UCF).** It predicts a user's preferences by taking the preferences of other similar users into account and use *Jacaard similarity* for similarity computation.

- **Item-CF (ICF).** It predicts a user's preferences on a target location by taking his preferences on similar locations into account [10] and *Jaccard similarity* is used for similarity computation.

- Friendship-aware CF (**FA**). The probability of user i checks in location k can be calculated as:

$$R_{ik}^f = \frac{\sum_{j \in F_i} s_{ij}^f \cdot c_{jk}}{\sum_{j \in F_i} s_{ij}} \tag{6}$$

where $c_{jk} = 1$ if user i checked in location k, otherwise 0. s_{ij}^f is computed as:

$$s_{ij}^f = \lambda \frac{|F_i \cap F_j|}{|F_i \cup F_j|} + (1 - \lambda) \frac{|L_i \cap L_j|}{|L_i \cup L_j|} \tag{7}$$

where $F.$ denotes user's friendship set and $L.$ denotes the locations checked in by each user. λ is used to balance the importance of friend impact and impact of shared checked in locations [11]. $\lambda = 0.4$ is the best setting based on our empirical study in this work.

- Geographic-aware CF (**GA**). We use Gaussian Mixture Model (GMM) to capture the geographical clustering influence, where Gaussian center could be user's home, office, or entertainment places like shopping malls or restaurants [1]. The probability that user i visits location k is modeled as below:

$$R_{ik}^g = \sum_{m=1}^{M} \pi_m N(l_k | \mu_m, \sum_m) \tag{8}$$

where l_k denotes location k, which is represented by longitude and latitude coordinates, and m is the number of Gaussian clusters.

- **Time-aware.** To address the temporal influence in users' check-in behaviors, we decompose the time over two dimensions of day (Mon. to Sun.) and hour (1 to 24). The probability of check-in is computed as:

$$R_{ik}^t = \frac{\sum_{l=1}^{L} \mathbb{I}(i, l) \cdot sim(l, k) \cdot f(T)}{\sum_{l=1}^{L} sim(l, k) \cdot f(T)} \tag{9}$$

where l denotes a subset closely associated with user i according to his historical check-in records. $sim(l, k)$ can be computed using Jaccard similarity. $f(T)$ is a temporal adjustment function for each user, which can be computed by using:

$$f(T) = \eta \cdot Pr(k|h) + (1 - \eta) \cdot Pr(k|d) \tag{10}$$

where $Pr(k|h)$ is the probability of check-in at location k, given the h-hour within a day (24 hours one day). $Pr(k|d)$ is the probability of check-in at location k, given the d-th day within a week (7 days one week).

We also compare with a linear model (**LIM**) by integrating these three partial contextual models together. The overall probability that user i would visit location k can be obtained:

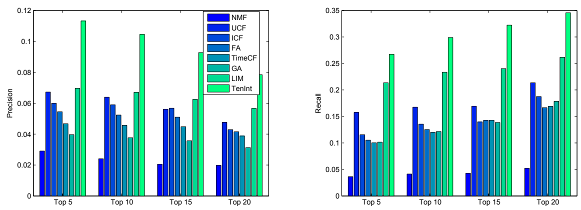
$$R_{ik} = \alpha R_{ik}^t + \beta R_{ik}^f + (1 - \alpha - \beta) R_{ik}^g \tag{11}$$

$\alpha = 0.1$ and $\beta = 0.6$ are the best settings based on our empirical study.

3.2 Results

3.2.1 Overall Comparison

The overall comparison results are shown in Figure 3, from which we summarize two main observations. First, our tensor-based recommendation method significantly outperforms all compared methods (including both non-context aware methods and context-aware methods) in terms of top 5 to top 20 validations. Our method obtains better prediction than the linear model, mostly because the tensor-based factorization can better reveal the hidden information. Second, all context-aware methods (e.g., the linear model fully combined with friendship, spatial and temporal influence) have better performance than the ones without or with only partial context-awareness. The basic matrix factorization method has the worst accuracy, as it only works on user-location matrix and does not integrate any contextual information. To sum up, the results demonstrate the effectiveness of incorporating multi-dimensional contextual information in a unified tensor based approach in improving the recommendation performance.



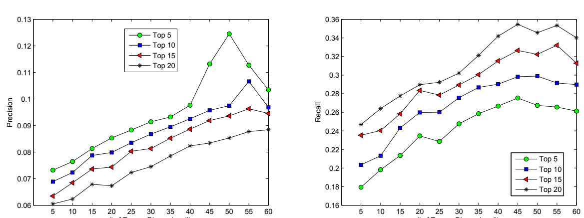
(a) Precision@x (x = 5, 10, 15, 20)

(b) Recall@x (x = 5, 10, 15, 20)

Figure 3: Precision and recall over all methods

3.2.2 Impact of Tensor Dimensionality

As the parameter dimensionality fundamentally determines the number of latent factors involved in the tensor factorization, in this section, we investigate the impact of the dimensionality by varying the value of dimensionality from 5 to 60 with a step size 5. Figure 4 shows the precision and recall at top 5, 10, 15 and 20 under different tensor dimensionality. We observe that the precision and recall keep increasing with larger dimensionality, however they slightly drop when dimensionality reaches around 55. The results reveal a larger dimensionality can effectively uncover information of check-ins and improve the recommendation performance. But when the dimensionality exceeds certain threshold, the performance may degrade because of over-fitting. Larger dimensionality also requires more computational cost. Based on our results, we have set the dimensionality as 50 in above comparison.



(a) Precision@x (x = 5, 10, 15, 20)

(b) Recall@x (x = 5, 10, 15, 20)

Figure 4: Impact of tensor dimensionality

4. CONCLUSION

In this paper, we propose a tensor non-negative decomposition based Point-of-Interest (POI) recommendation approach using users' social constraints as regularization. We model the check-in records as three dimensional tensor and employ the non-negative tensor factorization method to enable effective POI recommendation in a higher dimensional space. Specially, we propose to impose users' social constraints as regularization terms on tensor non-negative factorization to improve the recommendation accuracy. Our proposed method achieves better performance than seven standard baselines in our experimental study.

5. REFERENCES

[1] C. Cheng, H. Yang, I. King, and M. R. Lyu. Fused matrix factorization with geographical and social influence in location-based social networks. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.

[2] C. Cheng, H. Yang, M. R. Lyu, and I. King. Where you like to go next: Successive point-of-interest recommendation. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 2605–2611. AAAI Press, 2013.

[3] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090. ACM, 2011.

[4] D. M. Dunlavy, T. G. Kolda, and E. Acar. Temporal link prediction using matrix and tensor factorizations. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(2):10, 2011.

[5] H. Gao and H. Liu. Data analysis on location-based social networks. In *Mobile Social Networking*, pages 165–194. Springer, 2014.

[6] L. Hu, A. Sun, and Y. Liu. Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 345–354. ACM, 2014.

[7] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 79–86. ACM, 2010.

[8] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[9] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 287–296. ACM, 2011.

[10] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM, 2001.

[11] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 325–334. ACM, 2011.

[12] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann. Time-aware point-of-interest recommendation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 363–372. ACM, 2013.

[13] J.-D. Zhang and C.-Y. Chow. igslr: personalized geo-social location recommendation: a kernel density estimation approach. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 334–343. ACM, 2013.