POI Recommendation: Towards Fused Matrix Factorization with Geographical and Temporal Influences

Jean-Benoît Griesner
Télécom ParisTech
UMR CNRS LTCI, Paris, France
griesner@telecom-paristech.fr

Talel Abdessalem
Télécom ParisTech, UMR CNRS LTCI, Paris, France
School of Computing, National University of Singapore
UMI CNRS IPAL
talel.abdessalem@telecom-paristech.fr

Hubert Naacke
Sorbonne Universités
UPMC Univ Paris 06, LIP6
Paris, France
Hubert.Naacke@lip6.fr

ABSTRACT

Providing personalized point-of-interest (POI) recommendation has become a major issue with the rapid emergence of location-based social networks (LBSNs). Unlike traditional recommendation approaches, the LBSNs application domain comes with significant geographical and temporal dimensions. Moreover most of traditional recommendation algorithms fail to cope with the specific challenges implied by these two dimensions. Fusing geographical and temporal influences for better recommendation accuracy in LBSNs remains unexplored, as far as we know. We depict how matrix factorization can serve POI recommendation, and propose a novel attempt to integrate both geographical and temporal influences into matrix factorization. Specifically we present GeoMF-TD, an extension of geographical matrix factorization with temporal dependencies. Our experiments on a real dataset shows up to 20% benefit on recommendation precision.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval models; Information filtering—*matrix factorization; kernel density estimation; accuracy measures*

1. INTRODUCTION

The last years have witnessed the emergence of location-based social networks (LBSNs) such as Foursquare, Flickr, Weibo, Facebook places and so on. The success of these LBSNs has promoted the advent of new forms of online services. One of the main goals of these services is to offer to users the possibility to interact with each other and to explore new sets of points-of-interest (or POIs) by sharing their personal experiences and feelings on POIs they have visited in the past.

RecSys '15 September 16 - 20, 2015, Vienna, Austria

DOI: http://dx.doi.org/10.1145/2792838.2799679.

By collecting the mobility records of users, LBSNs constitute a rich and large-scale check-in data source. These data considered as an abundant implicit feedback of the travel experiences of the user, give a significant opportunity to improve POI recommendation performances. POI recommendation is the task of making personalized recommendations of the best POIs fitting the user preferences. Today this task has become an essential component of the LBSN domain since it allows users to have better user experiences, and POI owners as well to get more targeted customers. The traditional way to realize this task is to use classical collaborative filtering (CF) algorithms. CF algorithms are usually distributed into model-based and memory-based approaches. They derive from the check-in data the classical *user-POI rating* matrix in which a rating corresponds to the visit frequency of a user at a given POI.

A recent study showed [7] that weighted matrix factorization was the most adapted method to CF problems with implicit feedback. This method has been exploited and augmented by the authors of [9] to include the geographical influence of POI by the modeling of the spatial clustering phenomenon [14, 15] directly into the factorization process. However LBSN data comes with much more than only geographical information. Notably we have also access to the recorded timestamp of each check-in.

Our work aims at integrating time dependencies into geographical matrix factorization. In this paper, we investigate the idea of augmenting matrix factorization model with both geographical and temporal influences. This leads to the GeoMF-TD algorithm we present in the following.

2. RELATED WORK

Many methods have been proposed to solve POI recommendation. Most of these approaches try to adapt traditional recommendation algorithms to the specific problem of POI recommendation. One important line of research includes matrix factorization models. Matrix factorization techniques have demonstrated since the Netflix challenge [8] to be one of the most accurate recommendation methods many previous works use [2, 12]. Zhang et al. in [16] have proposed Collaborative Location Activity Filtering (CLAF) algorithm for generic recommendation. CLAF is a collective matrix factorization close to the method presented by Singh et al. in [13] based on the exploitation of the correlations existing between the features of the locations and the POIs.

Differently Regularized Matrix Factorization presented in [1] apply CF personalized methods on dimensionally reduced user-POI matrices aiming at minimizing squared regularized errors. In [11] Sattari et al. proposed Improve Feature Combination (IFC), which is based on an extended matrix factorization model that integrates additional data resources before applying Singular Value Decomposition technique to the extended model. It has been proven in several studies that IFC performs better than CLAF in terms of prediction accuracy.

Since each POI comes with a significant geographical dimension, many works have tried to integrate this geographical information into the recommendation model. Recently Ye et al. showed in [14] how to integrate geographical influence with classical CF approaches. More precisely Ye et al. have studied the geographical influence of POI assuming a power-law distribution of the visited POIs. On another hand, Cheng et al. [2] have recently proposed a multicenter gaussian model as a model of the spatial clustering phenomenon. Differently Zhang et al. in [15] proposed a personalized fusion framework based on kernel density estimation of the distances distribution between POIs of each user.

In addition to the geographical dimension, the temporal dimension is another important factor leveraging the accuracy of the model. Exploring temporal dimension into matrix factorization is not a new idea. Recently in [4] Gao et al. have proposed a location recommendation framework with temporal effects (LRT). Specifically they showed how to model two main temporal properties of data (i.e. non uniformness and consecutiveness) with matrix factorization. The experiments conducted showed that LRT outperforms traditional recommendation algorithms.

3. POI RECOMMENDATION IN LBSNS

Let $\mathbf{u} = \{u_1, u_2, ..., u_m\} \subset U^m$ be a subset of users and $\mathbf{p} = \{p_1, p_2, ..., p_n\} \subset P^n$ a subset of POIs. Then let $\mathbf{C} \in \mathbb{R}^{\mathrm{m}\times\mathrm{n}}$ be a user-POI matrix containing m users and n POIs. Value $c_{u,j}$ in $\mathbf{C}$ refers to the visit frequency of user u to the POI i.

3.1 Weighted Matrix Factorization

Basically the goal of matrix factorization is to approximate matrix $\mathbf{C}$ by the product of two matrices $\mathbf{P} \in \mathbb{R}^{\mathrm{m}\times\mathrm{k}}$, and $\mathbf{Q} \in \mathbb{R}^{\mathrm{n}\times\mathrm{k}}$ of latent factors with dimension $k \ll \min(m, n)$ by solving the following classical optimization problem,

$$\min_{\mathbf{P},\mathbf{Q}} \left\| \mathbf{C} - \mathbf{P}\mathbf{Q}^T \right\|^2 + \gamma \left(\|\mathbf{P}\|^2 + \|\mathbf{Q}\|^2 \right) \tag{1}$$

with γ a non-negative parameter to avoid overfitting by controling the capability of $\mathbf{P}$ and $\mathbf{Q}$. Then it becomes possible to approximate the missing value $\widetilde{c_{u,j}}$ in $\mathbf{C}$ by computing the inner product between corresponding latent factors $\widetilde{c_{u,j}} = \mathbf{P}_u\mathbf{Q}_j^T$. However, the application domain of LBSNs is different from traditional recommendation domains. Indeed the check-in datasets in LBSNs provide only indication of *confidence* but no information about *preferences* of users. This property refers to the recommendation problems with *implicit feedback*. Specifically Hu et al. have proven in [7] that weighted matrix factorization (WMF) gives the best results with implicit feedback datasets. Weighted matrix factorization takes into account of the asymmetry existing between *confidence* and *preference* and creates two new variables for formalizing this asymmetry. Then WMF turns the problem of Eq(1) into the following new optimization problem,

$$\min_{\mathbf{P},\mathbf{Q}} \left\| \mathbf{W} \odot \left(\mathbf{R} - \mathbf{P}\mathbf{Q}^T \right) \right\|^2 + \gamma \left(\|\mathbf{P}\|^2 + \|\mathbf{Q}\|^2 \right) \tag{2}$$

where $\odot$ is the element-wise matrices multiplication (i.e. the Hadamard product) and where the only differences with Eq (1) is the presence of the matrix $\mathbf{W}$, and the binary 0/1 matrix $\mathbf{R}$ whose each entry $r_{u,i}$ indicates if user u has visited POI i. The idea of WMF is to assume a minimum confidence for all POI, visited or not. This minimum confidence is encoded within the $\mathbf{W}$ matrix, setted as follows,

$$w_{u,i} = \begin{cases} 1 + \alpha(c_{u,i}) & \text{if } c_{u,i} > 0 \\ 1 & \text{otherwise} \end{cases} \tag{3}$$

where $\alpha()$ is a monotonically increasing function.

3.2 Modeling Geographical Influence

The modeling of geographical influence for POI recommendation in LBSNs has been widely studied in previous works [14, 2, 15, 10]. Recently Lian et al. in [9] have proposed a geographical matrix factorization (GeoMF) to integrate this influence directly into the factorization model of WMF. The idea of the authors was to distinguish for each user the unvisited but interesting POIs among the negative ones. The intuition is that if a user visits a POI without visiting the other closely located POIs then these "ignored" POIs may not be interesting enough for the user. Consequently these POIs become *negative* for the factorization model. This approach divides the space into L even grids $\mathbb{L} = \{g_1, g_2, ..., g_L\}$ and computes for each POI its influence area onto each one of these L grids based on the normal distribution of distances. Specifically they augmented the traditional matrix of latent factors $\mathbf{P}$ and $\mathbf{Q}$ with two matrices of latent geographical factors $\mathbf{X} \in \mathbb{R}^{\mathrm{m}\times\mathrm{L}}$ and $\mathbf{Y} \in \mathbb{R}^{\mathrm{n}\times\mathrm{L}}$. With these new latent factors, Eq(2) is modified as follows,

$$\min_{\mathbf{P},\mathbf{Q},\mathbf{X}} \left\| \mathbf{W} \odot \left(\mathbf{R} - \mathbf{P}\mathbf{Q}^T - \mathbf{X}\mathbf{Y}^T \right) \right\|^2 + \gamma \left(\|\mathbf{P}\|^2 + \|\mathbf{Q}\|^2 \right) + \lambda \|\mathbf{X}\|^2 \tag{4}$$

where λ controls the sparsity constraint over the mobility behavior of each user through the L grids. A row $\mathbf{x}_u$ of $\mathbf{X}$ refers to the activities areas of user u i.e. the distribution of his visit frequencies in each grid g_l of the map, while a row $\mathbf{y}_i$ of $\mathbf{Y}$ refers to the influence area of POI i. More precisely we compute for each POI i and for each grid g_l the gaussian geographical influence i has on g_l:

$$\mathbf{y}_i^l = \frac{1}{\sigma} K\left(\frac{d(i,l)}{\sigma}\right) \tag{5}$$

where $K()$ is the standard normal distribution and σ the standard deviation. With this augmented geographical model we get the recommendation ranking score for user u and POI i as follows,

$$\widetilde{c_{u,i}} = \mathbf{P}_u\mathbf{Q}_i^T + \mathbf{X}_u\mathbf{Y}_i^T \tag{6}$$

One of the most significant advantage of this approach is that it encompasses both preferences of user from latent factors and preferences from geographical factors.

4. GEOMF WITH TIME DEPENDENCIES

The GeoMF model assumes that the space is an isotropic homogeneous space without physical constraints. Especially this model assumes that the influence area of each POI follows a normal distribution fixed in advance and only based on distances over space. However the influence areas of two distinct POIs can be very different in reality by considering different parameters other than the distances. Notably the temporal effects in POIs visit sequences play also a significant role [4]. Particularly these effects can reflect that a POI j can be in the influence area of another POI i but not being really negative.

Following the GeoMF approach, our basic idea is to integrate these temporal influences into the GeoMF model. Actually, we propose to modify the values of the influence area of each POI i through the grid $g_{l \in \mathbb{N}^L}$ to take into account the time spent by a user to go from the POI i to the other POIs collocated in g_l. More precisely, for each POI i, we compute the average time that each user spend to reach j (j is in g_l) from i. We compute this for every user that has at least one check-in at i and another (more recent) check-in at j into g_l. Then, we average the per-user values to get a single value related to POI i. Let $t_i^{g_l}$ be the average time computed between i and collocated POIs existing in g_l. We introduce temporal coefficients $\theta_l(t_i^{g_l})$ as follows,

$$\theta_l(t_i^{g_l}) = \begin{cases} \alpha * \mathbf{y}_i^l & \text{if } t_i^{g_l} > \sigma^i \text{ and } \mathbf{y}_i^l < 0.1 \\ \mathbf{y}_i^l & \text{otherwise} \end{cases} \tag{7}$$

where σ^i refers to the standard variation of time intervals for POI i and $\mathbf{y}_i^l$ has been computed from Eq(5). Then we fuse these coefficients with influence vector $\mathbf{y}_i$ for POI i,

$$\mathbf{y}_i = [\theta_1(t_i^{g_1}), ..., \theta_L(t_i^{g_L})] \tag{8}$$

The idea of these temporal coefficients is to decrease the *negativeness* of potential negative POIs when no user has checked-in them during a certain time. That is why these coefficients let unchanged the influence area value when this value is low. We use these coefficients as a fusion output between geographical gaussian influence over space, and temporal dependencies existing into the dataset.

5. EXPERIMENTS

In our experiments, we compared the accuracy of our approach with GeoMF. This section describes the dataset we used, the evaluation metrics we chose, and the results we obtained.

5.1 Dataset and Experimental Setup

We evaluated the algorithms on check-ins crawled from Gowalla[1] and publicly available [3]. Gowalla was a famous LBSN closed in 2012. Gowalla dataset has already been used in several works on POI recommendation [2, 3, 15]. Table 1 presents the main statistics concerning this LBSN. In order to reduce matrix sparsity in the dataset we keep only users with at least 50 check-ins and for practical purposes we use only check-ins localized in France. Figure 1 presents the spatial distribution of check-ins over the France area. Finally it remains 161 users, 7697 distincts POIs for 12418 distinct check-ins, which is very few but enough for an initial evaluation. Then we organize this dataset as a user-POI matrix.

Number of users	196,591
Number of check-in	6,442,890
Number of social links	950,327
Matrix density	2.9×10^{-5}
Average No. of visited POIs per user	37.18
Average No. of check-ins per POI	3.11

Table 1: Statistics of the Gowalla data set

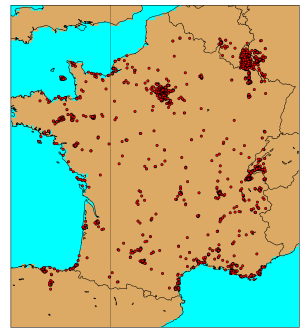

Figure 1: Check-in distribution from Gowalla users during 21 months of the most visited POIs in France.

5.2 Evaluation Metrics

It is traditional for each user $u_i \in U$ to mark off between 20% and 40% of all POIs he has checked-in in the past for testing, while the rest remains for training the model. Basically a recommendation algorithm estimates a ranking score for each candidate POI $i_{cand} \in P$ and returns the top-k highest ranked POIs $p_1, p_2, ..., p_k \in P^k$ as recommendation results for the targeted user. Then we evaluate the recommendation accuracy by finding out how many recommended POIs are effectively present into the test set of this targeted user. More precisely we compute *precision@N* and *recall@N*. The former refers to the ratio of recovered POIs to the N recommended POIs, while the latter refers to the ratio of recovered POIs to the set of previously visited POIs as follows,

$$precision@N = \frac{\sum_{u_i \in U} |TopN(u_i) \cap L(u_i)|}{\sum_{u_i \in U} |TopN(u_i)|} \tag{9}$$

$$recall@N = \frac{\sum_{u_i \in U} |TopN(u_i) \cap L(u_i)|}{\sum_{u_i \in U} |L(u_i)|} \tag{10}$$

where $TopN(u_i)$ represents the set of top-N POIs recommended to user u_i and $L(u_i)$ represents the set of POIs from the test set checked-ins by u_i. We have evaluated *precision@N* and *recall@N* with N ranging from 1 to 20 for precision, and from 1 to 100 for recall. We provide the results we obtained on the average after cross-validation with 5 folds in the next section.

5.3 Results and Discussions

For comparison purpose, we implemented the GeoMF approach using the LibRec Java library[2]. Figure 2 and Figure 3

[1]The Gowalla check-ins dataset can be downloaded here: http://snap.stanford.edu/data/loc-gowalla.html

[2]LibRec library can be downloaded here: http://www.librec.net/

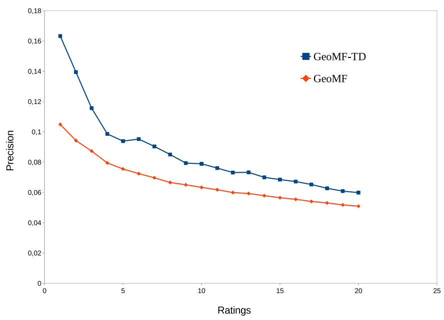

Figure 2: Precision comparison between GeoMF and GeoMF-TD

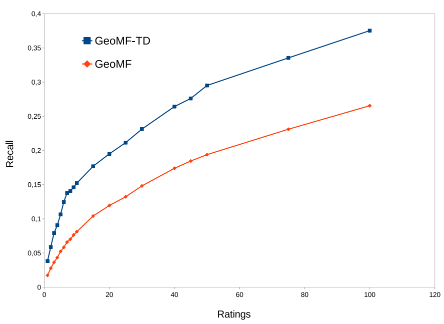

Figure 3: Recall comparison between GeoMF and GeoMF-TD

depict a comparative analysis of respectively the *precision*@N and the *recall*@N results of GeoMF and our approach (GeoMF-TD) with N ranging from 1 to 20 for the precision, and with N ranging from 1 to 100 for the recall. As expected the temporal coefficients we introduce allowed to take into account the temporal dependencies existing between POIs and thus improve the global accuracy. Figures 2 and 3 show an average benefit of 60% for recall and 20% for precision. This overall performance comparison does not integrate the study of the influence of the threshold parameter, but gives promising results for the future.

6. CONCLUSIONS AND FUTURE WORKS

In this paper, we have focused on the problem of POI recommendation in LBSNs. Specifically we have investigated matrix factorization algorithms based on geographical influence. Our goal was to try to leverage the factorization model of GeoMF by considering the temporal influences of POIs checked-ins. To this end we have provided GeoMF-TD algorithm as a first proposal of an extension of GeoMF and we have presented accuracy comparisons. Our experimental evaluation shows that GeoMF-TD presents better accuracy performances than GeoMF.

Considering the preferences of the user will change over time, a future line of work we should investigate will be to take into account of the online integration of user's preferences changes and to capture this evolution into our model. This problem refers to the recommendation *dynamicity* challenge widely studied in recent studies [5, 6]. Additionally, one of our future goal will be to include POI categories into the model, and to cope with the scalability issues.

7. ACKNOWLEDGMENTS

This work has been partly funded by the Télécom ParisTech Research Chair on Big Data and Market Insights.

8. REFERENCES

[1] B. Berjani and T. Strufe. A recommendation system for spots in location-based online social networks. SNS '11, 2011.
[2] C. Cheng, H. Yang, I. King, and M. R. Lyu. Fused matrix factorization with geographical and social influence in location-based social networks. AAAI, 2012.
[3] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: User movement in location-based social networks. KDD '11, 2011.
[4] H. Gao, J. Tang, X. Hu, and H. Liu. Exploring temporal effects for location recommendation on location-based social networks. RecSys '13, 2013.
[5] M. Gueye, T. Abdessalem, and H. Naacke. Technique de factorisation multi-biais pour des recommandations dynamiques. EGC'13, 2013.
[6] M. Gueye, T. Abdessalem, and H. Naacke. Dynamic recommender system : using cluster-based biases to improve the accuracy of the predictions. *Advances in Knowledge Discovery and Management - Volume 5*, 2015.
[7] Y. Hu, Y. Koren, and C. Volinsky. Collaborative filtering for implicit feedback datasets. ICDM '08, 2008.
[8] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. 2009.
[9] 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. KDD '14, 2014.
[10] B. Liu and H. Xiong. Point-of-interest recommendation in location based social networks with topic and location awareness. ICDM'13, 2013.
[11] M. Sattari, M. Manguoglu, I. H. Toroslu, P. Symeonidis, P. Senkul, and Y. Manolopoulos. Geo-activity recommendations by using improved feature combination. UbiComp '12, 2012.
[12] H. Shan and A. Banerjee. Generalized probabilistic matrix factorizations for collaborative filtering. ICDM '10, 2010.
[13] A. P. Singh and G. J. Gordon. Relational learning via collective matrix factorization. KDD '08, 2008.
[14] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. SIGIR '11, 2011.
[15] J.-D. Zhang and C.-Y. Chow. igslr: Personalized geo-social location recommendation: A kernel density estimation approach. SIGSPATIAL'13, 2013.
[16] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. Collaborative location and activity recommendations with gps history data. WWW '10, 2010.