MAPS: A Multi Aspect Personalized POI Recommender System

Ramesh Baral
School of Computing and Information Sciences
Florida International University, Miami FL, USA
Email: rbara012@fiu.edu

Tao Li
School of Computing and Information Sciences
Florida International University, Miami FL, USA
Email: taoli@cs.fiu.edu

ABSTRACT

The evolution of the World Wide Web (WWW) and the smart-phone technologies have played a key role in the revolution of our daily life. The location-based social networks (LBSN) have emerged and facilitated the users to share the check-in information and multimedia contents. The Point of Interest (POI) recommendation system uses the check-in information to predict the most potential check-in locations. The different aspects of the check-in information, for instance, the *geographical distance*, the *category*, and the *temporal popularity* of a POI; and the *temporal check-in trends*, and the *social* (friendship) information of a user play a crucial role in an efficient recommendation.

In this paper, we propose a fused recommendation model termed **MAPS** (**M**ulti **A**spect Personalized **P**OI Recommender **S**ystem) which will be the first in our knowledge to fuse the categorical, the temporal, the social and the spatial aspects in a single model. The major contribution of this paper are: (i) it realizes the problem as a graph of location nodes with constraints on the category and the distance aspects (i.e. the edge between two locations is constrained by a threshold distance and the category of the locations), (ii) it proposes a multi-aspect fused POI recommendation model, and (iii) it extensively evaluates the model with two real-world data sets.

Keywords

POI Recommendation, Social network analysis

1. INTRODUCTION

The LBSNs, such as, the Facebook[1], the Foursquare[2], the Gowalla[3], and so forth have facilitated the users to share the check-in information of the places of interest. Such check-in information has been the subject of interest to predict the POIs that are most likely to be visited in the future. Albeit, the generic recommendation concept has been used for POI domain (for instance, the Collaborative Filtering (CF) [5], the Content Based Filtering [16], and the Hybrid approaches [13]), its special aspects have motivated the community towards more sophisticated approaches for better results.

The frequency of check-ins varies across different users and places, resulting in the *sparsity* of the *user-location* frequency matrix in comparison to the *user-item* rating matrix in the generic systems. The check-in preference to a near place introduces the *spatial aspect* (the distance to a POI). Though the *social aspect* encourages to incorporate the social tie (*for instance, friendship*), it costs the challenge from the unreliability of check-in information diffusion, as only $\sim 96\%$ of people share $< 10\%$ of the commonly visited places and $\sim 87\%$ of people share nothing at all [14]. The *temporal aspect* depicts the temporal check-in pattern. For instance, the popularity of the *bars* is in the evenings and the nights. Many other relevant factors, such as, (i) the utility of a POI, regardless of the distance, cost, (ii) the popularity of the POI (due to social or other impact), and (iii) the dynamic mobility of a user (trend to visit new places) exist. Although the problem is well explored [1, 4, 11, 15, 17, 18], the incorporation of all the major aspects (*the social, the spatial, the temporal, and the categorical*) is barely explored.

2. RELATED RESEARCH

2.1 Simple similarity based approaches

Yuan et. al [17] used the Tobler's First Law of Geography [10], ("*everything is related to everything else, but near things are more related than the distant things*") and proposed a model with the spatial and the temporal aspect. The similarity between any two users was computed using the cosine similarity of their check-in profiles. The recommendation score for a *user-location* pair was computed as the aggregate of visits count on that location across all the users. The temporal similarity was incorporated by assuming the similarity of check-ins that have the same location and the check-in time. The *willingness* of a check-in was claimed to have an inverse relation to the distance. The social, and the categorical aspects were not well defined.

Ye et. al. [15] fused the social and the spatial aspects in their model. Though the *willingness* factor and the weighted cosine similarity measure was used to compare the user profiles for the recommendation, they didn't incorporate the categorical, and the temporal aspects.

2.2 Graphical approaches

Jin et. al [4] used the personalized PageRank [3] to realize

[1] www.facebook.com [2] www.foursquare.com [3] www.gowalla.com

RecSys '16, September 15-19, 2016, Boston , MA, USA

DOI: http://dx.doi.org/10.1145/2959100.2959187

the problem as a graph with the users as the nodes, and the following/followers relation as the directed edges. The rank of a user with respect to the location and a time range was defined using the personalized PageRank [3]. The topic sensitive factor for the (*user, location (p), time* (t_1, t_2)) tuple was taken as the ratio of the number of check-ins for the tuple to the number of check-ins for the (*location (p),time* (t_1, t_2)) tuple over all the users. Similarly, the rank of a location within a time interval was defined. This model also ignored the geographical, categorical and the social aspects. Wang et. al. [11] had the users and the locations as the graph nodes, the *user-user* friendship edges, and the *user-location* check-in edges. The friendship similarity was realized by starting from the target user and by ranking all the users (*that form user-user link*). This was followed by the ranking of all the places visited by those users. The places with the highest rank and within a given threshold distance from the past visits were recommended to the user. Their model also ignored the categorical aspect.

3. METHODOLOGY

The PageRank [8] approach used the number and the quality of the links to a web page to estimate its importance. It was extended to the Topic-Sensitive PageRank [3] by introducing some bias to the PageRank vector. It incorporated the set of influential or representative (*or additional context relevant attributes*) topics to address the importance of particular topics. For a given query, it identified the most closely associated/contextual topics and such relevant topic-sensitive (biased) vectors were used to rank the documents satisfying the query. The convergence of PageRank is assured only if the graph is strongly connected and aperiodic [7]. This becomes true if we add the damping constant $(1 - \alpha)$ to the rank propagation which improves the quality of PageRank not only by limiting the effect of the rank sinks [2], but also by assuring the convergence to a unique rank vector [3].

The **MAPS** is based on the Topic-Sensitive PageRank, where the *representative topics* are the spatial and the categorical aspects of the LBSN. The rank of a location (l) in the context of a user (u) and the time (t) is influenced by the check-in history of the user (u) at the time (t). For instance, if a user's check-in history has frequent check-ins in Starbucks coffee shop at 2 pm, then it is more likely that she will visit a coffee shop at that time in future. This temporal aspect should be taken care while recommending some coffee (or relevant category) shops to her. If that coffee shop is inaccessible, the user might not be surprised if a nearby cafe is recommended. Such a dual affinity of the time and the location category has motivated the MAPS to incorporate the categorical and temporal bias in the POI rankings.

Given two candidate POIs, suggesting the near one is more relevant [10]. If the check-in history of a user depicts that the check-ins were made within some distance of other check-ins, then introducing the distance constraint might give better recommendation. MAPS uses such check-in trends to incorporate the spatial bias in the location ranking problem.

In **MAPS**, every location is termed as a node of a graph and the bag of *user, time* tuple is considered as an attribute of the location node. The *location-location* edges exist if they have the same category or are located within some threshold distance. It uses the categorical and the spatial bias in its Topic-Sensitive PageRank model. The terms used in this paper are defined in Table -1. The categorical sensitive PageRank for MAPS is defined as:

$$\Pi^c_{t_1,t_2}(l) = \alpha * \beta_{t_1,t_2}(l) + (1-\alpha) * \sum_{(l'.cat = l.cat)} \Pi^c_{t_1,t_2}(l') \quad (1)$$

where $\beta_{t_1,t_2}(l)$ is the categoric sensitive factor, defined as:

$$\beta_{t_1,t_2}(l) = \tau_1 * \frac{\sum\limits_{u \in U} \mid V_{u,t_1,t_2}(l) \mid}{\sum\limits_{u \in U, l.cat = l'.cat} \mid V_{u,t_1,t_2}(l') \mid} + \tau_2 * \frac{\sum\limits_{u \in U, l.cat = l'.cat} \mid V_{u,t_1,t_2}(l') \mid}{\sum\limits_{p \in L, u \in U} \mid V_{u,t_1,t_2}(p) \mid} \quad (2)$$

where τ_1, and τ_2 are constant tuning factors. The relation (1) is somewhat similar to LBSNRank [4] but the equation is specific to our approach.

Terms	Definition
$\Pi^{\mathbf{a}}_{t_1,t_2}(l)$	rank of location l in the time range t_1, t_2 using the aspect $\mathbf{a}$
$\beta_{t_1,t_2}(l)$	categoric sensitive factor of location l in the time range t_1, t_2
$\theta_{t_1,t_2}(l)$	distance sensitive factor of location l in the time range t_1, t_2
$P(u, l, t_1, t_2)$	likelihood of checkin by user u to location l in the time range t_1, t_2
$V_{u,t_1,t_2}(l)$	visits by the user u to the location l, within the time interval t_1, t_2
$dist(l_1, l_2)$	distance between locations l_1 and l_2
U	the users in the dataset
L	the locations in the dataset
$l.cat$	category of the location l
ϵ	the threshold distance
α	the damping factor

Table 1: Terms used in the paper

Similarly, the distance sensitive rank of a location is defined using the following relation:

$$\Pi^d_{t_1,t_2}(l) = \alpha * \theta_{t_1,t_2}(l) + (1-\alpha) * \sum_{(l',l) \in E} \Pi^d_{t_1,t_2}(l') \quad (3)$$

where $\theta_{t_1,t_2}(l)$ is the distance sensitive factor, defined as:

$$\theta_{t_1,t_2}(l) = \gamma_1 * \frac{\sum\limits_{u \in U} \mid V_{u,t_1,t_2}(l) \mid}{\sum\limits_{u \in U, dist(l,l') \le \epsilon} \mid V_{u,t_1,t_2}(l') \mid} + \gamma_2 * \frac{\sum\limits_{u \in U, dist(l,l') \le \epsilon} \mid V_{u,t_1,t_2}(l') \mid}{\sum\limits_{p \in L, u \in U} \mid V_{u,t_1,t_2}(p) \mid} \quad (4)$$

where γ_1, and γ_2 are constant tuning factors. The unified rank is the fusion of the two ranks and is defined as:

$$\Pi_{t_1,t_2}(l) = \xi_1 * \Pi^c_{t_1,t_2}(l) + \xi_2 * \Pi^d_{t_1,t_2}(l) \quad (5)$$

where ξ_1, ξ_2 are tuning parameters for the two aspects.
The likelihood of the check-in for the user i at the location l within the time frame t_1,t_2 is defined as:

$$P(u,l,t_1,t_2) = \Pi_{t_1,t_2}(l) * (\psi_d * \sum_{\substack{l' \in L, \\ dist(l,l') \leq \epsilon}} | V_{u,t_1,t_2}(l') |) + \psi_c * \sum_{\substack{l' \in L, \\ l.cat = l'.cat}} | V_{u,t_1,t_2}(l') |) + \psi_s * \sum_{(u',u) \in friend} | V_{u',t_1,t_2}(l) |) \quad (6)$$

The terms ψ_d, ψ_c, and ψ_s are defined using TF-IDF [9, 12] for each user. For a user u,

$$\psi_d = \frac{n_d}{n}.log(1 + \frac{N}{N_d}) \quad (7)$$

where n_d is the number of visits by the user u that are within the threshold distance ϵ, n is the total visits count by u, N is the number of POIs, and N_d is the number of POIs that are within the threshold distance ϵ from the user's check-in history. For the categorial factor, we use the relation:

$$\psi_c = \frac{n_c}{n}.log(1 + \frac{N}{N_c}) \quad (8)$$

where n_c is the number of visits by the user u to the category c, and N_c is the number of POIs with the category c. Similarly,

$$\psi_s = \frac{n_s}{n}.log(1 + \frac{N}{N_s}) \quad (9)$$

, where n_s is the number of visits by the user u in common to her friends, and N_s is the number of visits in common to the friends for all the users $u \in U$.
Based on the aspects we considered, we have analyzed the performance of three different models, the categorical link based model (CLM) (defined in Eqn. (1) and (2)), the spatial link based model (SLM) (defined in Eqn. (3) and (4)), and the fused model MAPS (defined in Eqn. (5)).

4. EVALUATION

4.1 DataSet

We used the Weeplaces and the Gowalla dataset [6], which was collected from the popular LBSNs Gowalla and the Weeplaces. The Weeplaces dataset has 7,658,368 check-ins from 15,799 users over 971,309 different locations. The Gowalla dataset has 36,001,959 check-ins from 319,063 users over 2,844,076 locations. These datasets were well defined and had the attributes relevant to the context of the problem, such as, (i) the location category, (ii) the geospatial co-ordinates, (iii) the friendship information, and (iv) the check-in time. After avoiding incomplete records, the 5 most checked-in categories (and their check-in count) were: (i) Home/Work/Other: Corporate/Office (437,824), (ii) Food: Coffee Shop (267,589), (iii) Nightlife:Bar (248,565), (iv) Shop: Food& Drink:Grocery/Supermarket (161,016), and (v) Travel: Train Station (152,114) for Weeplaces, and (i) Corporate Office (1,750,707), (ii) Coffee Shop (1,063,961), (iii) Mall (958,285), (iv) Grocery (884,557), and (iv) Gas & Automotive (863,199) for the Gowalla dataset. The *work* or *home*-related category (Home/Work/ Other:Corporate/Office) was popular from 6 am to 6 pm, with the highest check-ins (42,019) made at 1 pm. Similarly, the **bars** had highest of 21,806 check-ins at 2 am and the lowest check-ins (15,209) at 5 am. Most of the check-ins were at 12 pm - 6 pm and were either in **Home** or **Work** related categories. The Figure -1 illustrates the inverse relation of the distance to the check-in frequency. It was obtained by plotting the distance between the chronologically sorted consecutive check-ins of each user and the likelihood of the users' check-in in that distance (for ease, the distance was rounded to four decimals). The check-ins centralized within some distance (the dense patches within 0.5 km) illustrate the willingness to near places.

Models	Precision	Recall	F-Score
Weeplaces Dataset			
Ye et. al [15]	0.02417	0.00095	0.00183
LBSNRank [4]	0.08496	0.00063	0.00125
Wang et. al [11]	0.01818	0.00052	0.00101
CLM	0.00428	0.00024	0.00045
SLM	0.09085	0.00799	0.01468
MAPS	0.29769	0.01039	**0.02008***
Gowalla Dataset			
Ye et. al [15]	0.03000	0.00120	0.00230
LBSNRank [4]	0.40900	0.00300	0.00600
Wang et. al [11]	0.10600	0.00200	0.00392
CLM	0.00633	0.00154	0.00247
SLM	0.25350	0.00973	0.01874
MAPS	0.35400	0.03100	**0.05700***

Table 2: Average Performance of MAPS in Weeplaces and Gowalla dataset

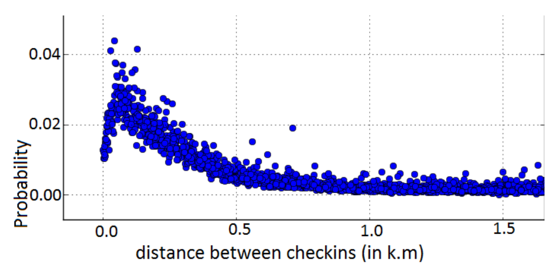

Figure 1: Impact of distance to check-in trend in Weeplaces dataset (*similar trend of Gowalla dataset is not shown due to space constraint*)

4.2 Results

A 5 -fold cross validation with top N (5, 10, 15 and 20) recommendation scores was used for the precision (P), the recall (R) and the F-score (2*P*R/(P+R)) metrics.
We used $\alpha = 0.85$ and the convergence was detected when the rank scores of the nodes were not changing anymore. For each model, the tuning parameters were selected from the random trials conducted with three set of parameters ((0.25:0.75), (0.5:0.5), and (0.75:0.25)). The categoric model performed best when $\tau_1 = 0.75$ and $\tau_2 = 0.25$, and for distance model it was when $\gamma_1 = 0.75$ and $\gamma_2 = 0.25$. Similarly,

Models	Precision@N	Recall@N
Ye et. al [15]	@5= 0.0303 @10= 0.0230 @15= 0.0191	@5= 0.0008 @10= 0.0009 @15= 0.0011
LBSNRank [4]	@5= 0.0853 @10= 0.0848 @15= 0.4090	@5= 0.0006 @10= 0.0006 @15= 0.0030
Wang et. al [11]	@5= 0.0449 @10= 0.0414 @15= 0.1060	@5= 0.0014 @10= 0.0020 @15= 0.0022
MAPS	@5= **0.2440** @10= **0.3050** @15= **0.3360**	@5= **0.0045** @10= **0.0092** @15= **0.0310**

Table 3: Precision@N, Recall@N of MAPS against other studies

among the three set of parameters the unified model performed best with the categorical aspect weight of 0.25. The comparative performance of different models is illustrated in Table -2. The observed difference was statistically significant at 95% confidence level. The Table -3 lists the average metrics across the top 5, 10, and 15 recommendation scores.

5. CONCLUSION AND FUTURE WORK

We analyzed the check-in data based on (a) the categorical, (b) the social, (c) the spatial, and (d) the temporal aspects. This multi-aspect recommendation model with reasonable performance was a significant contribution to the relevant area. Our next task is to analyze the same model against several other contexts, datasets and problem domains.

Acknowledgements

The work was supported in part by the National Science Foundation under grants CNS-1126619, IIS-1213026, and CNS-1461926, the U.S. Department of Homeland Securitys VACCINE Center under Award Number 2009-ST-061-CI0001, and a gift award from Huawei Technologies Co. Ltd.

6. REFERENCES

[1] R. Baral, D. Wang, T. Li, and S. C. Chen. Geotecs: Exploiting geographical, temporal, categorical and social aspects for personalized poi recommendation. In *Information Reuse & Integration, 2016. IRI'16. IEEE International Conference on*, page In Press. IEEE, 2016.

[2] S. Brin, R. Motwani, L. Page, and T. Winograd. What can you do with a web in your pocket? *IEEE Data Eng. Bull.*, 21(2):37–47, 1998.

[3] T. H. Haveliwala. Topic-sensitive pagerank. In *Proceedings of the 11th international conference on World Wide Web*, pages 517–526. ACM, 2002.

[4] Z. Jin, D. Shi, Q. Wu, H. Yan, and H. Fan. Lbsnrank: personalized pagerank on location-based social networks. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 980–987. ACM, 2012.

[5] J. J. Levandoski, M. Sarwat, A. Eldawy, and M. F. Mokbel. Lars: A location-aware recommender system. In *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*, pages 450–461. IEEE, 2012.

[6] X. Liu, Y. Liu, K. Aberer, and C. Miao. Personalized point-of-interest recommendation by mining users' preference transition. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 733–738. ACM, 2013.

[7] R. Motwani and P. Raghavan. *Randomized algorithms.* Chapman & Hall/CRC, 2010.

[8] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: bringing order to the web. 1999.

[9] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523, 1988.

[10] W. R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic geography*, pages 234–240, 1970.

[11] H. Wang, M. Terrovitis, and N. Mamoulis. Location recommendation in location-based social networks using user check-in data. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 374–383. ACM, 2013.

[12] H. C. Wu, R. W. P. Luk, K. F. Wong, and K. L. Kwok. Interpreting tf-idf term weights as making relevance decisions. *ACM Transactions on Information Systems (TOIS)*, 26(3):13, 2008.

[13] D. Yang, D. Zhang, Z. Yu, and Z. Wang. A sentiment-enhanced personalized location recommendation system. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, pages 119–128. ACM, 2013.

[14] M. Ye, P. Yin, and W. C. Lee. Location recommendation for location-based social networks. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 458–461. ACM, 2010.

[15] 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.

[16] H. Yin, Y. Sun, B. Cui, Z. Hu, and L. Chen. Lcars: a location-content-aware recommender system. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 221–229. ACM, 2013.

[17] 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.

[18] J. D. Zhang and C. Y. Chow. Geosoca: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 443–452. ACM, 2015.