Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 44 No. 8
August 2023

POI Recommendation through Content Preference Location Aware Collaborative Filtering

[1]venkata Praveen Kumar Vuppala,
Research scholar, Department of CSE, Annamalai University, Annamalai, Tamil Nadu, India.
vvpraveenkumar@yahoo.com

[2]Dr. S. Pazhaniraja,
Research Guide, Assistant Professor, Department of CSE, FEAT, Annamalai University,
Indiapazhanirajan.sambandam@gmail.com

[3]Dr. Indraneel Sreeram,
HOD & Professor, Department of Cyber security & IOT, St. Ann's College of Engineering and Technology,
Chirala, India,
sreeram.indraneel@gmail.com

Abstract
The focus of this article is on recommending point (places) of interest (POIs) to a user based on the user's present location and previous check-in behaviors. To solve this issue, we separate user preferences for POI content from user preferences for POIs themselves. The final POI suggestion is obtained by combining the projected content rating with the predicted location rating of the POI. We test this strategy with data from Yelp and Foursquare. This technique outperforms the competition and works well in "new city" situations if the user has not rated any of the POIs in the current region.

Keywords: Points of Interest (POI), Content Preference, Location Aware, Collaborative filtering

1. Introduction

Location-Based Social Networks (LBSN) like Yelp and Foursquare, allow user to "check-in" to specific Points Of Interest (POIs) like hotels/restaurants/museums using electronic device. Rating and comments on POIs after visiting by users, as well as visitors can review this POI for earlier selections to visit. The accessibility of evaluation information and Location-Based Social Networks service opens many recent investigation concepts and issues [17] in academies as well as industries [20]. Analyzing users nature, studying action patterns [12, 4], and different applications in the real-life [5, 23,27]. A new application to emerge other things recommends appropriate POIs when a user travels to a location or place [14, 8, 22]. The main variation with standard recommendation system is gap among the user's place and the place of Point Of Interest affects whether the client accepts Point Of Interest recommendation. For instance, even if the user adores Chinese food very much, recommending a Chinese restaurant in Beijing to a user who is now visiting New York City will fail.

Motivation: Two important POIs are assumed for every description: content and location. Location defines geographic location, like latitude and longitude of Point Of Interest. Content defines data of POIs as well as their functionality. Features (e.g. categories of artwork in museums), user-applied tags, and user-generated comments. After visiting the POI, user can rate POI, selects few parameters, generate tags and reviews. The content is explained a set of words to generate or selects by every user of Point Of Interest. Considering the noticed rating information and content of POI, the issue is to identify the user visiting place that given a similarity among user interests and POIs content as well as distance among POIs and to recommend appropriate POIs of customer present region.

The instant solutions are to segment the noticed rating by geographic locations (such as cities) and use the ratings data for that region to create a regional recommender for every region. When user navigates to location, that region's recommendation is used to recommend POIs. This solution is very simple but has many drawbacks. (1) Limiting evaluation data to a single region accelerates "data shortages" and makes evaluation predictions is difficult. (2) Knowledge transfer is lacking due to the upcoming "new town" issue. As an example two user's x and y will visit more general POI in hometown A. Also assume x visits the town B, not y. If y first visit location B, which is a "new city" to y, city B has no observational data of 'y', so B's recommender mistakenly believes that x and y have little common interest. Assume the "new city" issue will be different from "cold start" issue. All POI in the "new city" is present in these systems, but "new" to the particular users who has unrated POIs in particular place. (3) If there are unsuitable POIs for the user's current region, this technique tends to suggest irrelevant POIs. This is due to collaboration filtering in particular place will be disadvantage for collaborative filtering in different places.

Now, several theme- based models [11, 8, 22] incorporate POI content and location data. The primary aim will derive visiting probability of POIs/locations depended on observation check-in as well as user-generation information like tweets [8] and recommendation POIs/locations with the highest estimated probability to the user. Assuming the users' check-in patterns as mentioned earlier, the methods will not perform at user's visits place at present and that will not had many check-in days. In this case, these methods have the highest probability of estimating her POI in "hometown" where the user has visited in the past, so that her POI is recommended, but such her POI is currently unsuitable for users visiting another city. A completed explanation is given in Section 2.

The insight to differentiate between user's preference regarding POI content and user's preference regarding POIs themselves. The formers reflect extended period of interest implemented over a user's lifetime and are independent of current user interests. The latter is the short one. -Duration is limited by POI location and user. Importantly, long-term content preferences are transferable across locations. For example, users who love art have been to museums at earlier tend and they visit unseen city. They explain user content preference from rating data noticed in every geographies, where user's are located, and explain a user region preference based on user current location. When users travel to new city, content-enabled POI in recent place is detected by their location-independent content settings. Particulars and contribution are concluded:

Contributions:

Content Preference Location Aware Collaborative Filtering (CPLACF) addresses long-period content preference as well as short-period user locations preference. As long-term through location-independent content recommendation mechanism, learned from users attribute rating matrix derived from rating information's, and utilize recommendation mechanism to rate users for content in predicted POIs. The model short-term place preference through place recommendation and utilize to detect ratings for POI places. Specifically, subdivide the entire geographic region into several local places (such as cities) and create a "virtual user" for each user i and region k.

User/POI	I	II	III
A	4	4	-
B	5	2	5
C	5	-	4

Fig. 1 User – POI Matrix (M)

POI	FEATURES	REGION
I	F1, F2	R1
II	F1, F3	R2
III	F2, F4	R1

Fig 2. POI – Feature Matrix

User /Feature	F1	F2	F3	F4
A	4	4	4	

B	4	5	2	5
C	5	5		4

Fig 3. User Feature Matrix (M_1)

The model 'i' visiting (i,k) region 'k', as well as find place recommenders from a POI rating matrix of "virtual users" whose entries ((i,k),j) are the "perceptual ratings" of the place. POI 'j', by virtual-users assuming 'I' visits k. Both recommenders are learned through collaborative filtering of matrices. POI's final predictive rating is explained from content predictive and location predictive rating.

Examples for matrices are observed in Figure 2. As (A,II) 4 in actual rating matrix (a) shows 4 ranking for POIs II that user A observed if users visited POI II in area r2. At virtual-users POIs ranking matrix (c), ((A, r1), II) 2 denotes perceptual ranking from II to 'A', assume users are at r1. The rating perceives less for noticed rating due to the users at various regions than POI. When generating the perceptual rating ((A, r1), II), noticed rating (A, II) 4 and the gap among POIs II as well as user's present region r1 are taken into account. Therefore, ((A, r1), II) is the output for transfer of noticed understanding for the remaining domains. Let region r2. II be the user's current region, that is r1, considers effect of gap.

The techniques are primarily not similar for adjusted distance on collaborative filtering scores. Filtering POIs circles or few distances with penalties function. This is due to distance adjustments till prioritizing nearby POIs and require more effort from the user, like setting circle radius. As an example, consider restaurants with both vegetarians in a location is a short distance from user present city, then content recommenders will rank the restaurant which high rating, while location recommenders has fixed distances threshold, then matching application fails to detect restaurants without non-vegetarian or vegetarian restaurants [13].

The application is having some new characteristics. (1) Place and content recommendations are based on CPLACF, which is key element in addressing the information conservation issues mentioned above. (2) The technique explicitly models i visits any place k on virtual users of location-related matrix utilized for location recommenders device, so that Address the 'New City' issue. (3) If user's present location does not have a suitable POI for the method which recommends POI from neighboring regions. In contrast, local promoters tend to recommends not relevant POI in particular instance. (4) This approach is simple to perform. The basic matrix factorization approach as well as pre-processing/post-processing step to complete factorization approach is needed. It evaluates the approach using two Location-Based Social Networks service records, Yelp as well as Four-square. This analysis observes that this approach attains significant improvements by new approaches as mean cross-rank and recall.

The remaining part of the paper is described as follows: Section. 2 discuss related work. Section 3 explains method and explains its parameters. Section 4 shows experimental observations. Finally, the paper is concluded.

2 Related Work

The Studies investigated place/activity recommendation systems utilizing path such as Global Positioning System. In [28], it is assumed that interest in a place and user experiences at that place are mutually reinforcing, and Hyperlink Induced Topic Search (HITS) -like algorithm is proposed for learning the results. Similar to hub/authority ratings, site interest and users experience are global. It is not particular for Site-User Pairs. The location-activity matrix is obtained from Global Positioning System using activities of mining correlation, location features extraction, and location recommends specific activities for specific locations. To recommend locations, this method requires the user's activity to be specified as question. At this approach, user's preference corresponding for activity are implicitly included for previously noticed information, as well as POIs recommendations are based on user present place or city [26].

At [21], the geographic impact of locations is designed by a naive Bayesian algorithm as well as power-of-distance distributions, assume that the distance of POIs pair visits the users are not dependent. Given the set Li of POI visits by users i, possibility of visiting POI 'j' is modeled by conditional $P[j \mid L_i]$ possibility evaluated by j$\in L_i$ [P d $(j,]j^x]$) , where the distance probability j$\in L_i$ [P d $(j,]j^x)$] is assumed to follow power-law distribution. Cheng et al. [3] Using customers entered data, they create users location-matrix containing the check-in frequency of each user at each location, adopt a multicentre Gaussian model to calculate possibility of customers entry nature, and calculate the spatial Model as well as detects the not known frequencies for users places which never been entered. In [11], two kinds of location recommenders are suggested utilizing the theme system. One is depended on the user has earlier visits. The other is related to user's present locations. None of work model related with places, characteristics and users ratings.

Hu and Ester [8] described spatial topics system by including users reviews and movement by geographic topic system in which region as well as topic are unused elements related by every post randomly. The system detects locations of specific documents based on gathering of geo tagged post. This makes it more likely that the location is close to a user location which is frequented earlier. Liu et al. [14] described geographic probability factor model that considers latent preferences, user mobility, and object demand. The system map has user's potential regions which are determined by the POIs users had visits earlier and sets center of that region as user's location.

The described probabilistic and thematic modeling is depended on probabilistic distribution model of users' POI based-observational information, thus giving the priority to POI at user will regularly visits the regions (ex: hometown). Those models are unsuitable for users traveling to new regions which they have not generated posts or check-ins

With models of virtual user the methods recommends POI particularly to i travels to k. For As when users have examined into his POI in Beijing in earlier and for present visits the New York City, as she unchecked the POIs, system which suggest a match for his POI in New York City. However, probabilistic and thematic models tend to agree. Suggest Beijing POI, which is recommended due to highest evaluated probability.

Yin et al. [22] to integrate article content and user visit history, it introduces a theme model called LCA-LDA. Like, they are focused on providing relevant recommendations to user's visits a specific place. By observing every POIs as term, the system known the topic distribution for all users as well as every regions (e.g. cities), users visiting a region, recommends venues derived from learned distribution for that region. Their system didn't consider distance and geo location of POIs in different regions, which are significant factors in modeling. In [2], Evaluation data is subdivided by regions and location categories, "local experts" are obtained based on standard information for regions as well as categories. As in section 1, splitting evaluation information by region accelerates the sparsely of the data and decreases degree of CPLACF. This system uses geospatial as subdivided regions (to model geographic influences), but collaborative filtering is implemented on evaluated information for every region.

3 Methodology

Here it introduces the system. First, let's talk about data representation and task. This system generally contains two elements: Content and Location Recommender. Fig 2 shows structure of the data representation.

Preliminaries

It has I user, J POI, and F characteristics (term) for POI. Users i will provide few mathematic ratings (say 1–5) for many POI j which he visits. The monitored rating information is represented in an 'I X J' user-POI matrix M (Fig. 2a), at which the entry rij shows monitored rating of POIs j given by users i incase it is not reviewed POIs j, rij is not defined. If rating POI j , user i will alternatively selects few characteristics of j , Tij , to show the parameters are considered. If users i didn't rated POI j , It assumes Tij=Ø. The content of POI 'j' is defined as Tji= Tij . Every POIs also had geographical locations, i.e., latitude and longitude. It assumes, the operation will get back the gap between two geographic regions.

As, POI is the hotel, rating can be 1 to 5 star as well as characteristics of hotel will be clean, services, rating, comforts, etc. When users i provides five-star rating for hotels j and definitely choose clean and comforts, rij= 5 and Tij = {cleanliness, comfort}. This defines the users cleanliness and comfort of hotels. Incase user i rate hotel j still choose without features, it assumes that the user totally choose all characteristics, i.e., Tij contains all features of j. Here, Tij consists of explicit and implicit features. Alternatively, a feature could be a term or tag created for POI in a review, and T_{ij} is a gathering of words or tags for POI j generated by users i.

A POIs job is to recommends the highest n POI that match the user liking and are close to present city. The element detect user's ratings unranked POI based on derived rate of data, content, and POI region as well as user present place. Solution first predicts POI content ratings and POI location ratings separately and migrates the rating to explain complete POI ranking. These steps are described below:

Predicts the content preference

Initially, it builds a CON that predicts user's ratings for POI features. Hotel cleanliness is user preferences for POI content are estimated by summarizing predictive ratings for all POI features. To predict POI feature ratings, it employ features-centric recommendations of [24], that implements CPLACF to item parameters rather than the items itself. Initially, it transforms actual users-POI matrix '*M* into *I* x *F* user-feature matrix $M_{1'}$, every row shows user and every column presents feature '*f*' [24]. An entry '*(i, f)*' in M_1 is gradual ranking on features *f* by POI *j* which contains *f* as well as rating by users *i*:

$$g_{if} = agg\ (\{r_{ij} \mid f \in T_{ij} \text{ and } r_{ij} \text{ is defined}\}) \quad (1)$$

Here, agg(X) back to average of values in X and incase vacant then it is unexplained. Hence, gif noticed average ranking on 'f' by 'I'. As it extracts unused user's vector 'ui' for 'I' and latent features vector v f for feature 'f'. This will be completed by implementing standard matrix factorization [10] on M1. The aim is to decrease regularized squared error loss among 'gif' and detected ranking uiT v f , as in [16]. For extensiveness, it includes model of matrix factorization as describes

$$E = \frac{1}{2}\sum_{i,f} c_{if}\left(g_{if} - u_i^T v_f\right)^2 + \frac{\lambda_u}{2}\sum_i \|u_i\|^2 + \frac{\lambda_v}{2}\sum_f \|v_f\|^2 \quad (2)$$

where 'λ_u and λ_v' are regularized parameters; c_{if} is binary indicator which equals to one incase '*i*' is chose features *f* (g_{if} is noticed) and equals to zero. Considering gradient of *E* interms of elements u_i and v_f, we get

$$\frac{\partial E}{\partial u_i} = \lambda_u u_i - \sum_f \left(g_{if} - u_i^T v_f\right) v_f \quad (3)$$

$$\frac{\partial E}{\partial v_f} = \lambda_v v_f - \sum_i \left(g_{if} - u_i^T v_f\right) u_i \quad (4)$$

A local minimum of '*E*' is founded by repeatedly upgrading the entire variables by step equivalent to negative gradient depends on value by '*η*':

$$U_i^{K+1} = u_i^k - \eta \frac{\partial E}{\partial u_i^K} \qquad V_f^{K+1} = V_f^K - \eta \frac{\partial E}{\partial V_i^K} \quad (5)$$

This repeated method ends till convergence. The result is a latent user vector u_i for every 'i' and a feature vector 'v_f' for every features.

The detected ranking of *i* on content of POIs '*j*' is evaluated by aggregate theprediction ranking $u_i^T v_f$ complete feature '*f*' in T_j:

$$r_{ij} = agg\ \ u_i^T V_f . f \in T_j \quad (6)$$

We chose agg(X) is average of values of X because it is easy and performs betters in trails. Remaining alternatives of aggregation functions are also feasible, like average difference from average rating for every objects reviewed by target audience or regression-based weights that matches ratings on multiple criteria [7].

Predicting Location Preferences

The next significant feature in POI recommendation is user region preferences on the POIs. Create LOC to forecast this user's place preferences, depending on POI locations and the user's location. This system dependence divides a given K geospace into K regions. Regions cover adjacent locations and each region has a centroid. The regions should be "small enough" so that the gap among two regions of two cities is approximated by the distance among the centers of the two cities. As, in New York City, locations are cities that corresponds to communities such as Manhattan, Brooklyn, and Queens. The choices of region definition is frequently depends on applications. The system predicts that the region division is described.

Assume user is in region 'k', to estimate user rating for region of POIs 'j'. However, 'j' is unnecessary in region 'k'. Obviously, the score will unmatched noticed score r_{ij} generates if 'i' visits POIs 'j'. In this case, users as well as POIs are within the equal regions. If users is in region 'k', the user's ranking for j region will be decreased as distance from k to j.

This evaluate the distance-adjusted rating by $l_{(i,k)\,j}\,\frac{r_{ij}}{dis(k,i)}$ where $dis(k, j)$ defines the normalized distance between center of region k and POI j:

$$\mathrm{dis}(k,j) = \begin{cases} 1 & \text{if } j \text{ is in region } k \\ 1 + \dfrac{\|c(k) - c(\sigma(j))\|_2}{MIN} & \text{otherwise} \end{cases}$$

(7)

where $\sigma(j)$ is region index of POIs j, $c(x)$ is center of region x, and MIN is the minimum pairwise distance of region centers. Intuitively, $l_{(i,k),\ j}$ is observed rating r_{ij} adjusted by normalized distance between regions of POIs j and region of users i . This adjustment is compatible with Tobler's first law of geography [18], that defines that tendency of users for POIs is inversely proportional to geographic gap among users and POIs. Besides that users are possess by content of POIs, she/he can be low caution by geographical gaps. Because of this purposes for including the rating r_{ij} in distance-adjusted rating. If r_{ij} is unobserved, $l_{(i,k),\ j}$ is unexplained. Note if POIs j is in similar region k as user i , the distance influences disappeared and $l_{(i,k),j}\ r_{ij}$.

It define $(I\ X\ K)\ X\ J$ matrix M_2 (observe Fig. 2c) with $l_{(i,k),j}$ as follows. users i and region k, and create column j for all POIs j. The approach in row (i,k) and column j labeled ((i,k),j) consists distance-adjusted score l(i,k),j (incase explained). It will be instinctively, every row (i,k) restorative virtual user that is an case of user i visits k, and an entry ((i,k),j) is her POI j Simulate the perceptual evaluation at the position of 1 User.

After constructing M2, will conclude the unoccupied scores in M2 by implementing matrix factorization on M2 to known the latent users vector uik of virtual user (i,k) and the latent location vector vj of POI j. Objective functions to minimize is

$$E = \frac{1}{2}\sum_{i,j,k} c_{ij}\left(l_{(i,k),j} - u_{ik}^T v_j\right)^2 + \frac{\lambda_u}{2}\sum_{i,k}\|u_{ik}\|^2 + \frac{\lambda_v}{2}\sum_{f}\|v_j\|^2$$

(8)

The evaluation of u_{ik} and v_j is equal to earlier subsection. The detected locations rating by user i at k for POIs j is calculated by inner product of u_{ik} and v_j:

$$\hat{l}_{(i,k)} = u_{i,k}^T v_j$$

(9)

Recommending POIs

It is believed that content recommenders capture the extendable period of preference, and location recommenders capture the short-period preference. Hence, initially generates a pair of POI which matches the user attraction and next clear them by region. Given integer n with positive, it want to recommends her n most interests POI for uses present visits k. Reminder these POI are unnecessary in k. This is completed in four steps as described in the diagram. 3. (i) For every POIs j, predicts user i's content rating for j using the formula 6. Sorted all POI by detecting content rating. (ii) Chooses the highest m number of her POI from the ranking record as seeds (m > n). Her POIs in seedset are taken to match the user's content interests and her POIs not in the seedset will not recommend for users. (iii) Sort POI in seedset as per the detected locations score calculated in Equation 9. (iv) Recommendation of the highest n POI in sorted lists. The POI contains better favorite places that match the user's interests. This method is called Content Preference Location Aware Collaborative Filtering (CPLACF).

Discussion

The CPLACF model has many characteristics. Initially, recommend her POI which is not necessary from user's current region. Typically, if customer's current regions have POI by predictive content that has a high score. The POI will includes seedset and determined for recommenders. Incase present location didn't had POI by highly detected content scores, her POI in different places with highly detected content scores are incorporate in seedset. The alternative of seed set size 'm' determines comparative weight of content preferences and region preferences. A large 'm' results in a flexible content preference that favors local POI within the user's exact region, small m favors POI by a stronger content preference, hence, local POI by a weaker content preference are the seed set may not enter. Therefore, m will be utilized to authority for trade-off among the content and location preferences.

The techniques suppose the geospace is divided into K region, so location of POI by the location which will be accurate by region centers. The Area granularity affects site modeling calculations and accuracy. A finer granularity results in more accurate modeling of POI locations, but results in a large number of locations K and a large matrix M_2. The tiny and medium-sized geographic areas such as cities and states, K in 100 is usually sufficient. For large geospatials such as countries or the world, K can be large to keep each region small. This can be space and computationally expensive. The solution will cover larger geographic area. Divide the country into small and medium sub-spaces. The state and implement the system for every sub-space separately. The system formed sensible. This is because a user visiting a state will typically consider her POIs in the same state, and the lack of data problem is less acute in states (as in smaller counties).

The system will address "new city" issue with unspecified precautions. If user i is visiting cities k for initial period and hence it is unevaluated her POIs for that cities. By users viewing point, city 'k' is "new city". The system model is the framework through M_2 virtual users (i,k) by perceptual ratings $l_{(i,k),j}$ of POIs 'j'. As users had reviewed few of her POI in some city, no special handling is required to invite user i to visit region k.

4 Experimental evaluation

Here, it evaluates the effective of Content Preference Location Aware collaborative filtering system described in Section 3. Intially, let's talk about datasets, comparison methods, and metrics.

Data sets

The experiments employ datasets Yelp1 and Foursquare2. As datasets are earlier utilized in [8] for recommenders calculation. Yelp records include 45,981 users, 229,906 reviews, 11,537 of his POI, and user's review of POI. The pre-processed comments by extracting ends and rare terms are less than hundred comments and utilizing last 8519 keywords as functions. POI's functionality or content contains of every keywords included in POI's rating. The Foursquare dataset consists of 20,784 users, 153,477 reviews, 7711 POIs, and tweets posted by users when they check into POIs. Similar to Yelp, it extracts 1,377 features after pre-processing the tweets. To divide gap into region, because of its simplicity, it implement K-Means clustering method for all POI location. Other advanced clustering or tessellation techniques [1, 15] are achievable.

Methods For Comparison

Comparison of the techniques that implement or received code from authors:

Probabilistic Matrix Factorization (denoted PMF): The technique ignores POI content and location information, treats POIs as items, and applies PMF to customer object review utility matrix [16]. PMF generalizes PMF model as latent user vector 'ui N(0, λ−u 1 ID)', latent item vector v j N(0, λv−1 ID), and user-item rating rij N(uiT v j , εij−1). The user prediction 'i' ranking the item 'j' is expressed by uiT v j

Partitioned PMF (denoted PAR): In this model, it simulate a location-aware version of Probabilistic Matrix Factorization by creating models for every location utilizing a user-object matrix 'M' that contains review information within regions. Record M has a column for each POIs, and also facilities within present regions had review information. As described in section. 3.5, each region uses only local assessment data, so the method didn't supports CPLACF. For users visits a region, recommend her POIs for the user, who utilizes a recommender which created for that region.

Collaborative Topic Regression (denoted CTR): This is factorization matrix for topic models implemented to item characteristics [19]. As dataset, items are POI. LDA is applied to features of POI j to known latent topic vectors 'θ_j', merged into Probabilistic Matrix Factorization model to enclose latent item vectors searching by setting $V_j \sim N(\theta_j, \lambda_v^{-1} I_D)$. This method does not consider the location of POIs.

Spatial Topic Model (denoted STM): This is POI-recommended productive theme system that considers the spatial and textual features of POI [8]. Spatial topic model provides global recommendations without allowing user present region.

Location Content-Aware model (denoted LCA): This is another productive theme system for POI recommendations that incorporates content and location data [22]. Location content-aware provides individual recommendations depending on user's current location, but distance is not taken into account when modeling location.

Location recommender (denoted LOC): This is place recommenders from Section 3.3, ignoring POI content information and performing Content Preference Location Awareal collaborative filtering based on place ratings. When user i visits region k, this recommender recommend the top n POI sorted by prediction reviews in equation (9).

Content recommender (denoted CON): This is the CON from Section 3.2, ignoring POI location information and performing Content Preference Location Awareal collaborative filtering based on feature ranking. When users recommendations are high and POI rated by prediction score in Equation (6).

Content Preference Location Aware Collaborative Filtering (denoted CPLACF): The POIs Recommends from Section 3.4 and combines predictive reviews for LON and CON devices. The default, size of seed m utilized by this recommendation is limit upto 200.

Probabilistic Matrix Factorization, Partitioned PMF, and Collaborative topic regression presented as a baseline which treats POI as frequent elements. Spatial topic model and Location content-aware are the current new recommendations for POI. Location recommender, Content recommender, and content preference location aware collaborative filtering are explained from the method in Section 3 Consider POI location, content, as well as location. Each latent factor models default to D 10 dimensionality of the latent vector. It varies 'D' from ten to fifty to examine capability. It uses grid searching for obtaining optimal value of the regularized parameters λ_u and λ_v that searches features. The features for topic modelingare 'α 50/D,β' 0.010 and the learning rate is η = [0.0010, 0.010, 0.10, 1]. The code for the geographic stochastic factor system described [14] was not available and not included.

Evaluation Metrics

Perform 10-fold cross-validation on both datasets. H. For every fold, 90% of the noticed scores are selected as train set M, and the testing set T consists of positive scores for different 10 %. If your Yelp rating is 1to 5, positive ranked is less than 4star (because average rating for dataset is about 3.6). If Four square has a rating of 0/1, positive rating 1. This is because every user is normally checking in her POIs for specific region test set assumes the geographic centers as user present region if recommendations are implemented.

At common, recommended systems of predicting users ratings for POIs will be calculated using error-based and rank-based metrics. Therefore, topic model-based systems, Spatial topic model and Location content-aware predicts the chances of visits the POIs. Error-specific metrics like RMSE or MAE are unsuitable chances will incomparable for calculation. As it uses a rank-based metric as well as follow same process [6]. As every system explained in 4.2, initially trains a recommendation based on ratings noticed in M.

After the every testing instance '*(i, j)*' in '*T*' as '*i*' rate POIs '*j*' *as* positive:

(i) Use a recommender program to detect the rating of POIs j and each different POIs not rated by users. (ii) Rank every POI as per their detected rating to create a ranking. The rating '(i,j)' of POIs 'j' in record. The results correspond to when POIs 'j' precede remaining POI. Rank'(i,j)' = 1. (iii) Select n highest rated POI from the record, where n is number of recommended POI. Hit if Rank(i, j) ≤ n. Otherwise an error will occur. Based on the hit concept, consider two rank-based score metrics.

Mean Reciprocal Rank (MRR) is average of reciprocal ranks of the testing instances (i, j) in T , expressed as

$$MRR= (1/|T|)\ \frac{1}{rank\ (i,j)} \qquad (10)$$

A high number of MRR is sensible as the small value of rank'(i, j)' a better rank results for positive rating.

Recall@n is designed for top-n recommendation: As described, for every testing instances (i, j) in T , if rank(i, j) ≤ n, there had on hit, that the positive rating POIs j are recommenders to users. The part of the testing cases in T that have a hit is expressed by recall.

$$Recall@n = \frac{\#hit}{|T|} \qquad (11)$$

where #hit is number of hits in T's testing cases. High recall means that her POIs with more upvotes in T will be recommended to users. Other metrics utilized in [6] is Precision@n, this is similar to $\frac{\text{Recall@n}}{\text{n}}$. Therfore iall systems will determined for same *n*, it is enough to determined Recall@n.

Results

MRR. Table 1 shows the MRR for every method the K=10 region. PMF didn't had better performance because it utilizes the ranked data which avoids features and location data. The Collaborative topic regression and Spatial topic model related to content give better performance than Probabilistic Matrix Factorization. Spatial topic model also determines his POI locations, still the development is not unlimited on the Foursquare dataset due to Spatial topic model global recommendations didn't considers the user present locations. Location content-aware reaches best outputs by models and distribution of topics in various places and providing individual recommendations according to user present location. Therefore, Partitioned PMF implements fairly well, but fails if local referrer didn't have a sufficient score on the New City test (Section 4.5). It especially high MRR for CPLACF considered to Partitioned PMF, Location recommender, and Content recommender proposes the content and location, CPLACF actually improves the implementation of POIs recommendations. For easy representation, CPLACF is represented to CRCF in the following graphs.

Table 1 MRR (mean ± standarderror). K = 10

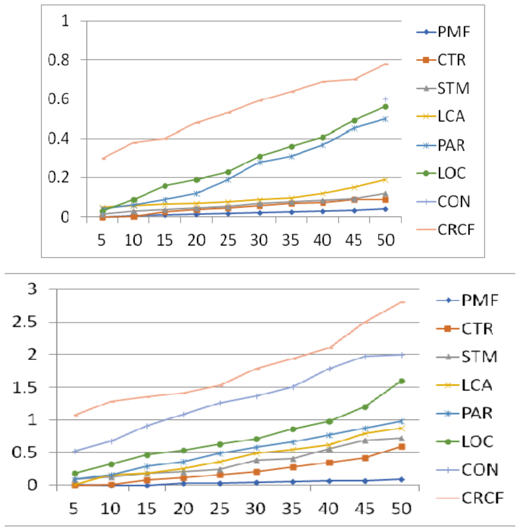
Methods	Yelp	Four-Square
Traditional Recommendation Algorithms		
PMF	0.0078 ± 0.0003	0.0092 ± 0.0004
CTR	0.0107 ± 0.0004	0.00112 ± 0.0006
Spatial Recommendation Algorithms		
STM	0.0102 ± 0.0003	0.0139 ± 0.0006
LCA	0.0159 ± 0.0007	0.0272 ± 0.0011
PAR	0.0254 ± 0.0010	0.0282 ± 0.0012
Our Proposed Algorithm		
CPLACF	**0.0690** ± 0.0014	**0.0521** ± 0.0016

Bold value indicates the bestperformer.

The later observations for performances of top two evaluters, Content Preference Location Aware Collaborative Filtering and Partitioned PMF. The testing case of 97% for the higher ten POIs recommends by the Partitioned PMF from customers current location, which is about 30% for Content Preference Location Aware Collaborative Filtering. The remaining recommends POIs within the areas. As, if Chinese restaurant j is a positive rating testing case '(i, j)', CPLACF will recommend user i's 4 restaurants in his current area and 6 restaurants in other neighborhoods. The hotels are the theme of Chinese cuisine. Therefore, Partitioned PMF recommended the ten hotels in user's present area, of which one was near to the Chinese cuisine. It has high Mean reciprocal rank for CPLACF because he has high POIs 'j' with highest rating '(i,j)' for PAR. Since PAR recommendations depends only on rating data within user i's region, it is prone to lack of data and thus he is prone to mis-evaluation of POIs. For every testing case '(i, j)' in T, hit j is always user i's current region, because it actually visits POIs 'j'. If her favorite POIs are not in user's current location, PAR is likely underperforming. CPLACF understands the user's interests and location restrictions by taking neighborhoods where appropriate.

Recall@n. Figure 4 describes Recall@n. Content Preference Location Aware Collaborative Filtering was clearly completely successful. PAR shows next better evaluation, followed with Location content-aware model. Location recommender, Collaborative topic regression, and Spatial topic model show almost same performance, but Probabilistic Matrix Factorization had enough performances. The purpose for CPLACF has better recall is the similar reason for more Mean reciprocal rank explained. H. T gives a higher percentage of hits for testing case '(i, j)'. By the perspective of every element of Content Preference Location Aware Collaborative Filtering, Content recommender is found to be superior to other algorithms in terms of content consideration, and content preferences model is available. Location recommender, yields similar results for different algorithms, by using CON produces various rankings of POIs (otherwise POIs would maintain little change after re-ranking). The result is a higher CPLACF recall value than these two algorithms. Level rankings are explained in Section 3.4. In other words, CPLACF's superiority stems from a combination of the user's location preferences and content preferences.

Effects of measurement of seed 'm'. It described in section 2. In 3.5, Content Preference Location Aware Collaborative Filtering utilizes a seed sizes 'm' to describe the top 'm' POI rated the content preference which participates in last rating for city preference. Increasing 'm' obtains higher compatibility for content preference and gives huge importance for city preference. Table 2 shows impact of seed measurement 'm' on the Mean reciprocal rank and recall of Content Preference Location Aware Collaborative Filtering. The small n, the low m is, the best recall@n is reached.



(a)
(b)

Fig. 4 Recall@*n* versus *n* (the *x*-axis). *K* = 10.

a Recall@*n* at Yelp. **b** Recall@*n* at Foursquare

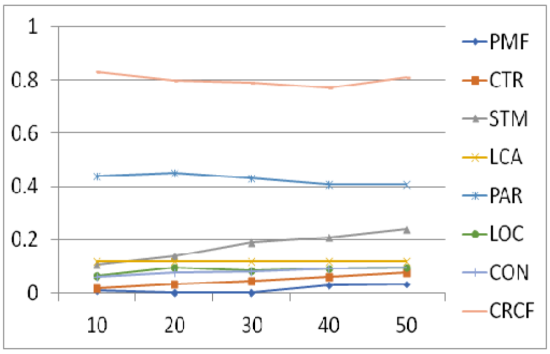
Table 2 Performances of CPLACF for different seed sizes *m*

Seed size *m*	MRR	Recall@5	Recall@10	Recall@50
YELP				
50	0.0585	0.0681	0.0786	0.1118
100	0.0682	0.0826	0.0993	0.1463
200	**0.0690**	0.0927	0.1145	0.1754
300	0.0644	**0.0941**	**0.1195**	0.1913
400	0.0597	0.0898	0.1191	0.1984
500	0.0566	0.0833	0.1183	**0.2045**
FOURSQUARE				
50	**0.0603**	**0.0932**	0.1212	0.1627
100	0.0588	0.0893	**0.1284**	0.1972
200	0.0521	0.0752	0.1193	**0.2279**
300	0.0452	0.0636	0.1017	0.2266
400	0.0409	0.0565	0.0897	0.2211
500	0.0373	0.0492	0.0819	0.2113

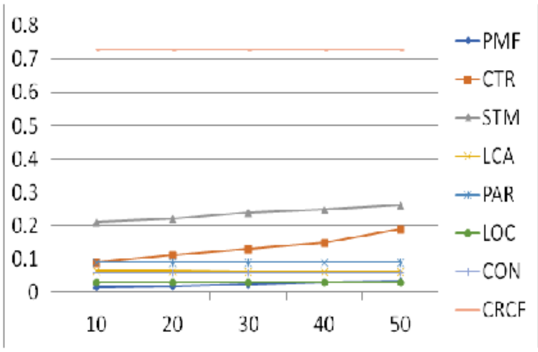
As larger 'm' and minor 'n' tends to includes several Point Of Intersts that are not well prioritized in location prioritization, resulting in poor LOC performance. For large datasets more than thousands of POIs, settles m to several hundred may be better option.

The impact of dimensionality D. Observe implementation for various models in terms of ‘D’. The outputs are shown in Figure 5. It is showed on the x-axis. Where D= ‘10 to 50’, it can see every model achieves constant evaluation and the execution time increases linearly. As D increases, more latent factors can be derived, so the run time increases. The Yelp dataset contains huge data (every comment contains huge terms than tweets in Foursquare dataset), content recommendation, Content recommender obtains higher, Followed by topics model algorithms like Location content-aware and Spatial topic model. However, Foursquare datasets with little content information, location recommendation LOC takes the more time because the score is replicated K times.

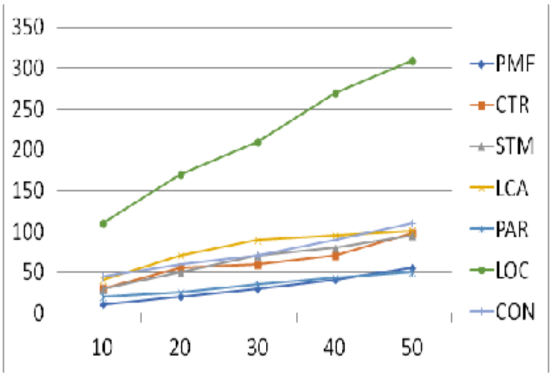
The effect of the number of regions *K.* Also, examine the performance of CPLAF and PAR with respect to the number of regions K. The result is shown in Fig. 6

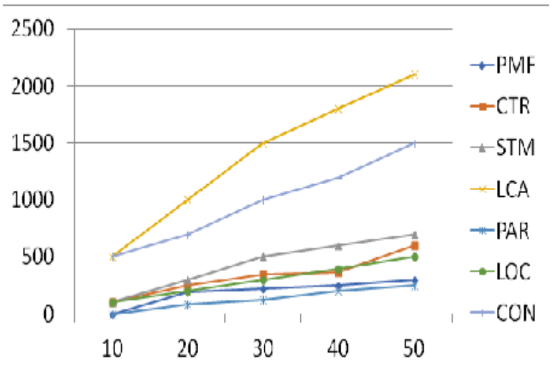


(b)



(a)



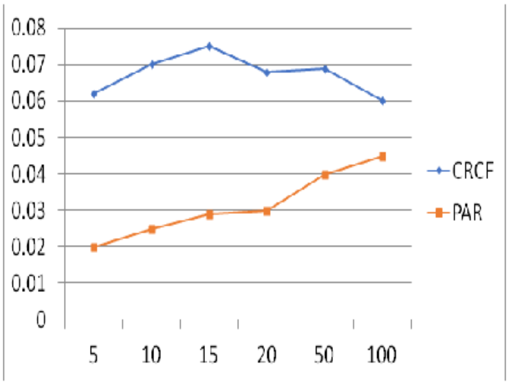


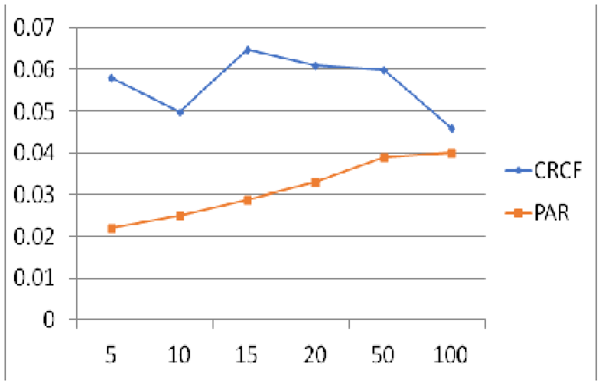
(C) **(d)**

Fig. 5 Performance versus *D* (the *x*-axis). *K* 10. a Recall@5 at Yelp. b Recall@5 at Foursquare. c Runtime (s) at Yelp. d Runtime (s) at Foursquare

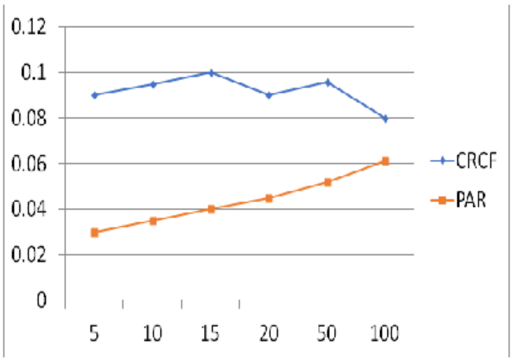
where K represents on x-axis. CPLACF exceeded PAR for K observed. CPLACF implementation will be impacted to some extent and as K increases, PAR performance gradually improves, except for Recall@50. This results comes from T's testing case (i,j). ‘I’ actual visits POIs ‘j’. So users ‘I’ and POIs ‘j’ are every time it is similar to itsdomain. Large ‘K’ creates low region by few POI and Partitioned PMF works as favor the POI, so large ‘K’ enhances hit probability. However, the larger K presents user to visit a 'new city', and the less POI user has in every region. Incase hits will not be in user's domain, PAR's performance is less than desirable.

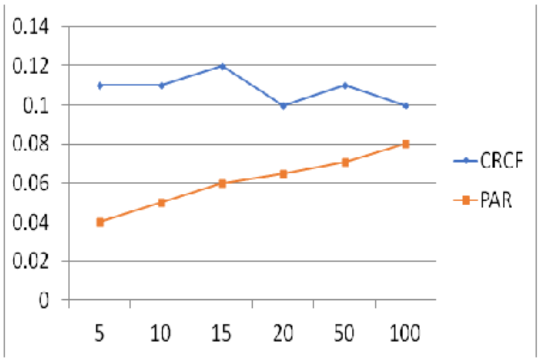
Since the matrix M2 consists of ‘K’ virtual user for every actual customer, count of locations ‘K’ affects training time of CPLACF's locations recommenders (see Table 3). Since the run time of matrix factorization is equivalent to count of noticed scores in matrix, and M2 consists of distance-adjusted the score copied for every location, working time of factorization depended on M2 is equivalent to ‘K’. CPLACF implementation is constant at various ‘K’, so it recommends that you do not need a larger K. As described in section. In 3.5, we can divide a large geographic area as many tiny and medium geographic areas as well as the model will be implemented for every small and average geographic area separately. Hence, there is no need to increase K with geospatial size. As ‘K’ didn’t impact study time for CPLACF's content recommendation program.



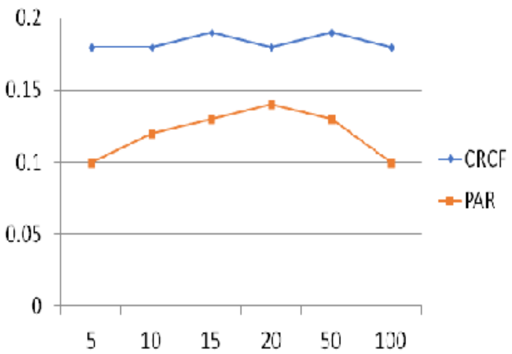


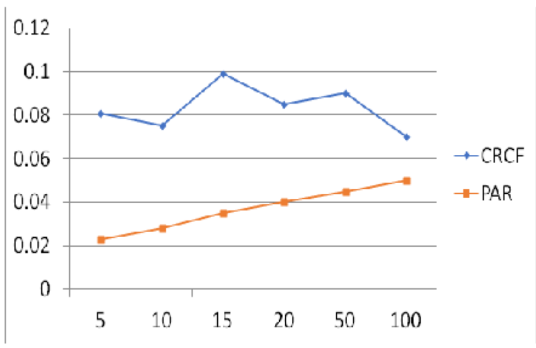
(a) **(b)**



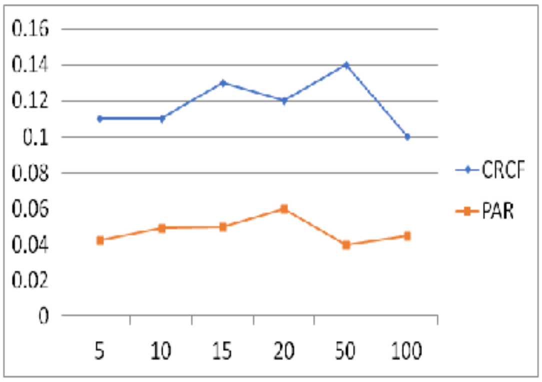


(c) **(d)**

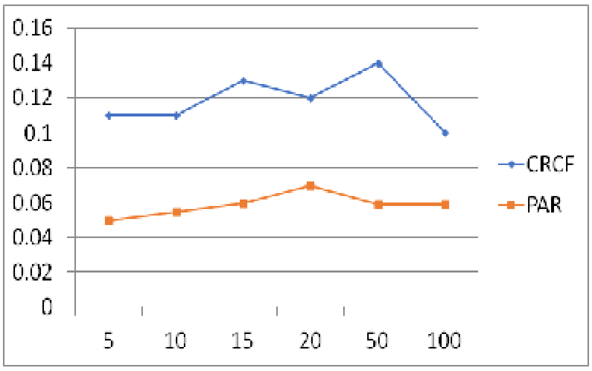




(e) **(f)**



(g)



(h)

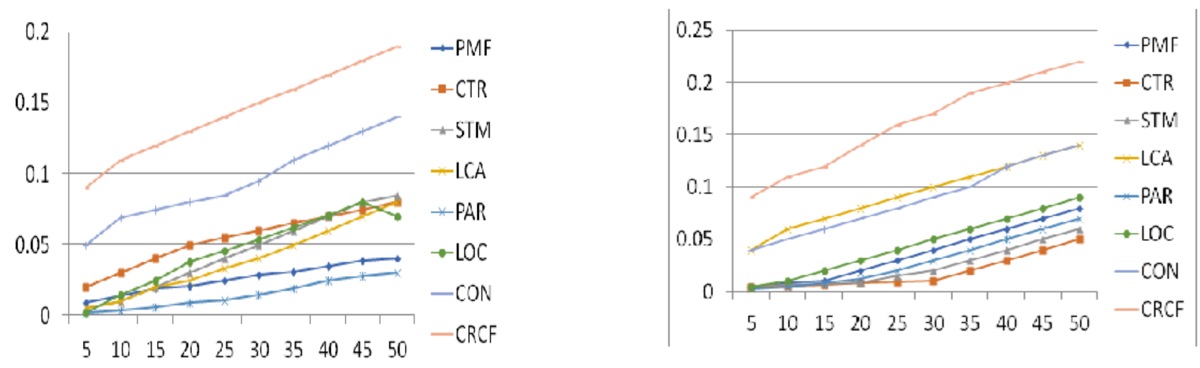
Fig. 6 Performances for different numbers of regions *K*. **a** MRR at Yelp. **b** MRR at Foursquare. **c** Recall@5at Yelp. **d** Recall@10 at Yelp. **e** Recall@50 at Yelp. **f** Recall@5 at Foursquare. **g** Recall@10 at Foursquare.**h** Recall@50 at Foursquare

Table 3 Runtime of learning the location recommender of CPLACF(min) versus *K*

K	5	10	15	20	25	30
Yelp	1.46	2.88	4.31	5.97	14.85	29.43
Foursquare	1.04	1.98	2.91	3.85	9.55	18.87

“New City” Testing

The protest of system is that it can address "new city" problems, i.e. urban problems. Recommendations for users visiting location or regions at initial period and don’t has previous rate the data for that location. To assess claims, divided POI into K ten locations for K-means clusters, set regions indices, and selected ninety percentages of examined reviews as training sets and remaining as testing sets. This makes true for all users whose testing data regions didn’t appeared in trains the information. Especially, users evaluate nine locations as "home cities" as well as treats remaining "new cities". Note that "hometown" is not really the user's hometown, but a region (or even one) that the user has already visited.



(a) **(b)**

Fig. 7 Recall@*n* versus *n* (the *x*-axis) in "new city" testing. *K* 10. **a** Recall@*n* at Yelp. **b** Recall@*n* at Foursquare

A "new city" defines to a region users didn't visits previously. As a 10-fold cross-validations can be performed to calculate implementations for every model for users visits the "new city".
Figure 7 shows recall for every systems in the " new city " test. Comparing with Figure 4, it can see that all methods except PAR perform as both common and " new city " tests. PAR shows a gradual drop in memory that is surprises due to partitioned PMF requires part of the data trains, whereas other models use the entire data trains to generate system. In New City settings, the users didn't rated POIs in location of the test information, so user of partitioned PMF in these regions. Since partitioned PMF is a locations-aware versions of Probabilistic Matrix Factorization, there is an obvious performance hit for better customers. Remaining models that employ entire data trains the benefits of the CPLACF. A "new city" test, founds the straight splitting of the information and creation of the regional model were poor. Also, it's not realistic to recommend only users POIs within the current region like PAR.

Summary

Observed examinations on Location-Based Social Networks services information supports the contention that CPLACF can help remove data shortage vulnerabilities while considering location effects, leads for the high accurate rating POI. The excellent evaluation of described system supports the model to distinguish between long-term user content preferences and short-term user region preference as well as various plans for customizing preference. And also supported the approach.

5. Conclusion

They distinguish between users content preferences for POIs transferable via location and location preferences for POIs that are depended. The difference allows method for the user preferences, therefore the addressing "new city" issue. The combination of kinds of preference provide recommendations for user POIs to users to travelling for recent regions if users are unrated POIs. This model was CPLACF which uses evaluation information by locations to train all users.
Therefore, similarities in preferences in one region can make recommendations in remaining cities. These are key elements for overcoming information scarcity.
Future work will investigate better aggregation strategies for feature scores for content recommendation devices and exponential distance adjustment scores for location recommendation devices. [9] Validation preprocessing can also utilizing concepts and recommenders will be enhanced by acquiring the full Bayesian method [25].

References

1. Adrienko N, Adrienko G (2011) Spatial generalization and aggregation of massive movement data. IEEE Trans Vis Comput Graph

17(2):205–219
2. Bao J, Zheng Y, Mokbel MF (2012) Location-based and preference-aware recommendation using sparse geo-social networking data. In: GIS, pp 199–208
3. Cheng C, Yang H, King I, Lyu MR (2012) Fused matrix factorization with geographical and social influence in location-based social networks. In: AAAI, vol 12. pp 17–23
4. Cheng C, Yang H, Lyu MR, King I (2013) Where you like to go next: successive point-of-interest recommendation. In: IJCAI, AAAI Press, pp 2605–2611
5. Cheng Z, Caverlee J, Kamath KY, Lee K (2011) Toward traffic-driven location-based web search. In: CIKM, pp 805–814
6. Cremonesi P, Koren Y, Turrin R (2010) Performance of recommender algorithms on top-n recommenda-tion tasks. In: RecSys
7. Fuchs M, Zanker M (2012) Multi-criteria ratings for recommender systems: an empirical analysis in thetourism domain. Springer, Berlin
8. Hu B, Ester M (2013) Spatial topic modeling in online social media for location recommendation. In:RecSys, pp 25–32
9. Hu M, Liu B (2004) Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining, ACM, pp 168–177
10. Koren Y, Bell R, Volinsky C (2009) Matrix factorization techniques for recommender systems. Computer42(8):30–37
11. Kurashima T, Iwata T, Hoshide T, Takaya N, Fujimura K (2013) Geo topic model: joint modeling ofuser's activity area and interests for location recommendation. In: WSDM, pp 375–384
12. Leung KW-T, Lee DL, Lee W-C (2011) Clr: a collaborative location recommendation framework based on co-clustering. In: SIGIR, pp 305–314
13. Levandoski JJ, Sarwat M, Eldawy A, Mokbel MF (2012) Lars: a location-aware recommender system. In: ICDE, IEEE, pp 450–461
14. Liu B, Fu Y, Yao Z, Xiong H (2013) Learning geographical preferences for point-of-interest recommen-dation. In: KDD, pp 1043–1051
15. Rösler R, Liebig T (2013) Using data from location based social networks for urban activity clustering. In: Geographic information science at the heart of Europe, Springer, pp 55–72
16. Salakhutdinov R, Mnih A (2008) Probabilistic matrix factorization. In: NIPS, pp 1257–1264
17. Sang J, Mei T, Sun J-T, Xu C, Li S (2012) Probabilistic sequential POIs recommendation via check-in data. In: GIS, pp 402–405
18. Tobler WR (1970) A computer movie simulating urban growth in the detroit region. Econ Geogr 46:234–240
19. Wang C, Blei DM (2011) Collaborative topic modeling for recommending scientific articles. In: KDD,pp 448–456
20. Ye M, Yin P, Lee W-C (2010) Location recommendation for location-based social networks. In: GIS, pp458–461
21. Ye M, Yin P, Lee W-C, Lee D-L (2011) Exploiting geographical influence for collaborative point-of- interest recommendation. In: SIGIR, pp 325–334
22. Yin H, Sun Y, Cui B, Hu Z, Chen L (2013) Lcars: a location-content-aware recommender system. In:KDD, pp 221–229
23. Yuan J, Zheng Y, Xie X (2012) Discovering regions of different functions in a city using human mobilityand POIs. In: KDD, pp 186–194
24. Zhang C, Wang K, Lim E-p, Xu Q, Sun J, Yu H (2015) Are features equally representative? a feature- centric recommendation. In: AAAI
25. Zhang C, Wang K, Yu H, Sun J, Lim E-p (2014) Latent factor transition for dynamic collaborative filtering.In: SDM, pp 452–460
26. Zheng VW, Zheng Y, Xie X, Yang Q (2010) Collaborative location and activity recommendations with gps history data. In: WWW, pp 1029–1038
27. Zheng Y, Zhang L, Ma Z, Xie X, Ma W-Y (2011) Recommending friends and locations based on individual location history. ACM Trans Web 5(1):1–44
28. Zheng Y, Zhang L, Xie X, Ma W-Y (2009) Mining interesting locations and travel sequences from GPS trajectories. In: WWW, pp 791–800.