Outgoing and Neighbouring Check-ins behaviors in location-based recommendation

Manrui Li*, Second author's name* and Third author's name*
*School of Data and Compuyer Science
Sun Yat-sen University, Guangzhou, P. R. China, 510000
Email: marywbrown1@gmial.com, SecondAuthor@Email.com,ThirdAuthor@Email.com

Abstract—*****In the abstract, we need to state the research background, motivation, the proposed method, experiment results, as follows***. Notice that the abstract part should be written after the other parts have been finished.*** One sentence to state the background. One sentence to state the motivation. A few sentences to describe the proposed method. One sentence to state the experimental setting. One sentence to state the experimental results.**

Keywords*—*Keyword one; Keyword two; Keyword three; ***

I. Introduction

The early recommender systems always used the information of traditional websites to promote the sales volume of the online consumption, while the recent recommender systems began to focus on recommending the offline consumption by using the information provided through integrated devices. The Online location-based social networks(LBSNs)are representative. Because the information of real-time locations becomes easier to be acquired by GPS, LBSNs are widely researched and applied, such as Foursquare, Gowalla, Facebook Places, in which users can share the experience in physical world by checking-in with a Point-Of-Interest(POI). LBSNs can help people discovering the interesting places to do outside activities, especially when people are in urban, where the POIs are considerable quantity and various and hard to choose for users.

Though LBSNs can benefit users outdoor activities and bridge the gap of physical world and online social network, LBSNs encounter more challenges than traditional recommender systems. 1).**Users transfer preference with the time period.** In traditional recommender systems, relatively speaking, users' preference is comparatively stable in a short time period. However, the users' preference can be totally different in a day of different period in LBSNs, for example, users always prefer a food-relevant location at lunch time while they may prefer a bar-relevant location at night, which means LBSNs are more correlative with time period. 2).**The limited effect of social influence.** Traditional social relationship has limited effect in LBSNs for the reason that almost 90% user's overlap check-ins to his/her friends' check-ins is less than 20% in LBSNs according to (Cho,Myers,and Leskovec 2011). In other word, we should explore more appropriate social relationship in LBSNs to improve the accuracy of recommendation. 3).**Extreme sparse check-in data.** An individual user usually check in a limited numbers of locations for two reasons. On one hand, though there are numerous locations in a city, users only check-in a small section for the restrict of distance. On the other hand, users prefer to check-in their favorite locations repetitively, which leads to the number of check-ins in different location is limited for a single user.

A. Related Work

The related research work in POIs recommendation is mainly focus on three aspects: geographical check-in behaviors [], social influence [],and temporal patterns [].

Geographical influence is a vital feature of location-based recommendation, for it's the mainly difference between POIs recommender system and traditional recommender system. [**?**] proposed a home-work two centers model, in which the probability distribution over the locations of a user at different time is simply the mixture of home and work location distributions. [**?**] proposed a Multi-center Gaussian Model(MGM) via modeling the probability distribution of a users check-in on a location, and utilized the inverse distance rule and incorporate multi-center geographical influence into the fused MF framework. As for social influence, [**?**] defined three types of friends including social friends, location friends and neighboring friends,then took the union of the three to expand the social network of users. [**?**]used the probabilistic matrix factorization with social regularization (PMFSR) in POIs recommendation to learn the latent features of users and POIs. With regard to temporal patterns, [**?**] explored the periodical feature of time in the location type classification by assuming people prefer different activities at different time, and divided time of day into 24 hours to establish recommender system for each hour. [**?**] explored temporal property of successive check-ins and they discovered that successive check-ins had a remarkable correlation. Besides, sentiment analysis techniques and content-based model always applied in POIs recommendation to explore the explicit preference information from tips in [].

B. Our Work

We divided users' check-in behaviors in two categories, Neighbouring Check-in Behavior(NCB) and Outgoing Check-in Behavior(OCB), and established different models respectively for the two kinds of check-in behaviours. Considering people often decide a place where they can have a series of recreational activities when they go outside, for example, a place they can go shopping and dinner and movies in a successive time. So we proposed Business Circles Model(BCM) to predict the business zones that users probably prefer to go firstly, and then recommend the locations in the zones. Besides, we clustered users into different communities by their interests and incorporated overlapping community information as regularization terms into MF framework. Finally, we consider the temporal information and integrated time parameters in the final loss functions.

The remainder of the paper is organized as follows. Section II describes in detail the ****. Experimental results are reported in section III. We conclude our paper in section IV.

II. MODELS

This is the first part you can write as soon as the algorithm is implemented. That is, you can write this part now.

Describe the algorithm here, one major part per paragraph.

A. Business Circles Model(BCM)

Assume that there are m users $U(u_1, ...u_m)$ and n check-in locations $L(l_1, ...l_n)$, we use k-means algorithm to cluster all the check-in locations and find p business circles $C(c_1, ..., c_p)$.

Then we give the rating r_{ik} of user u_i on business circle c_k if the circle contains check-in locations which have been rated by u_i already. There are two indicators have high correlation with the preference of user on c_k as followed:

1)The average daily quantity of check-ins in business circles.

2)The average rating of user u_i on locations which are in the area of business circles c_k.

So we define the r_{ik} as:

$$r_{ik} = \alpha \ \frac{\sum_{l_j \in c_k} r_{il_j}}{N_{il_j \in c_k}} \cdot \frac{1}{r_{max}} + (1-\alpha) \ \frac{N_{l_j \in c_k}}{max(N_{l_j \in c_k{}'})} \quad (1)$$

The first term indicates average rating of u_i on all the check-in locations, which are in the business circle c_k. N_{i,l_j} is the amount of check-ins that u_i rated on locations in c_k,and $\frac{1}{r_{max}}$ is multiplied to normalize the result in the range of $[0, 1]$. In five class marking system, $r_{max} = 5$. And the second term is the average daily quantity of check-ins in c_k, $max(N_{l_j \in c_k{}'})$ is the maximum check-ins on business circles, $c_k{}' \in C$. The term is also in the range of [0,1]. α is a parameter which will be determined value via experiment.

However, the above function can only rate the business circles which contains the locations that users have been check-in before, so we use these the low-rank matrix factorization approach to predict the other ratings $\widehat{r}_{ik}$. The users ratings on business circles form a $m \times p$ rating matrix $R' = [r_ik]$, where r_ik is the rating of u_i on c_k that we calculate by formula (1).

The objective function E_c is:

$$E_c = \frac{1}{2}\sum_{i=1}^{m}\sum_{k=1}^{p} I_{ik}(r_{ik}{}' - u_i{}^T c_k)^2 + \frac{\lambda_u}{2}||U||_F^2 + \frac{\lambda_c}{2}||C||_F^2 \quad (2)$$

The predict rating of user u_i on c_k is:

$$\widehat{r}_{ik} = u_i{}^T c_k \quad (3)$$

Considering the preference of users on business circles also related with the distance from user home to locations, we introduce distance factor as:

$$P_{r_{ik}} = dist(i, k) \times u_i{}^T c_k \quad (4)$$

$P_{r_{ik}}$ is the final predict rating ,$Dist(i, k)$ is the distance from $u_i's$ home to business center C_k.

Since the check-in probability and the distance from home to the corresponding location obey a power law distribution [**?**],we define the distance function as the following:

$$dist(i, k) = \frac{d_{ik}^{\gamma}}{d_{ik'_{min}}^{\gamma}} \quad (5)$$

d_{ik} is the distance that user u_i's home from business circle c_k, and d_{ik} is the minimum distance from home to business circles. γ is the parameter of power law distribution which can be learned by maximum likelihood estimation.

B. Overlapping Users' Communities

(Cho,Myers,and Leskovec 2011) proposed that traditional social relationship has limited effect in LBSNs for the reason that 90% user's overlap check-ins to his/her friends' check-ins is less than 20% in LBSNs. Sparse dataset and too many check-in categories lead to that result,for example,user A prefer food-related, bar-related and park-related locations while user B prefer food-related, art-related and sports-related locations, their similarity is probable low even though their preference in food-related locations are similar.

In order to solve that problem we proposed ? model to divide users into q overlapping communities by their check-in categories, denoted by $M(m_1, ...m_q)$. ([**?**]) proposed MFC model to incorporate the overlapping community information as regularization terms into the MF framework. Based on MFC model,our model introduce the temporary factor to improve the model more suitable for location-based recommendations.

Different from MFC model, in our model, we consider that users have distinct interests on diverse communities at various time, for example, u_i prefer a restaurant-related location more than a bar-related location at 11:00 13:00, but totally opposite at midnight.

The community regularization term is:

$$\frac{\lambda_h}{2}\sum_{i=1}^{m}\sum_{h=1}^{q} I_{ih} Z_{iht} \sum_{U_f \in M_{ih}} S_{if}||U_i - U_f||_F^2 \quad (6)$$

where M_{ih} is the users in the same community m_h with u_i as shown in figure 1, I_{ih} equals 1 if u_i belongs to c_h or equals 0. Z_{iht} is the preference of u_i on community m_h at time t.

C. Temporal Characters

Z_{iht} is the preference of u_i on community m_h at time t as mentioned above. The most popular way to find the temporary characters is 24 hours partition approach which divide a day into 24 hours [], however, this approach can not consider the weekly period or festival day. So we define time t by three attribute $t(t^1, t^2, t^3)$, t^1 is the hours of the day , t^2 means the day is week day or weekend day, and t^3 stands for if the day is a festival day or not. Then we use Bayes equation to calculate the probability that what categories of community m_h that users will check-in ar time t. The object function as:

$$P(Z_h|t) = \frac{P(Z_h)\prod_j P(t^{(j)}|Z_h)}{\sum_h P(Z_{ih})\prod_j P(t^{(j)}|Z_{ih})} \tag{7}$$

D. Outgoing Check-in Behavior And Neighbouring Check-in Behavior

Users consuming behavior will be changed when they go outside, so does users' check-ins. Firstly, users probably have higher consumption when they are outside than daily cost and the types of consumption are also different. Then, users always have a series check-in behaviors when they are going out in a consecutive time. Besides, the target user who can give you a good advice is different, people prefer to accept the suggestion of user who have the similar track with them when they consume near the home or working place while they are willing to take advice from users who have the same interests with them when they are outside.However, the traditional recommendations can't distinguish these two behavior, so we proposed Outgoing and Neighbouring Check-ins model based on Matrix Factorization(ONMF).

The home place pl_h and working place pl_w can be estimated by using the existing approach in [], then we find an appropriate distance d_{lim} from check-in locations to $pl_{h(w)}$ to divide the dataset into neighbouring check-in data and outgoing check-in data.

We use two loss functions to describe the outgoing check-in behavior and neighbouring check-in behavior to recommend locations. In outgoing situation, we introduce business circles model and integrate the rating $P_{r_{ik}}$ of u_i on c_k and overlapping community information into MF framework as followed:

$$\begin{aligned} E^O = &\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n^O} I_{ij}(r_{ij} - P_{r_{ik}} u_i^T l_j^O)^2 + \frac{\lambda_u}{2}||U||_F^2 + \frac{\lambda_{l^O}}{2}||L^O||_F^2 \\ &+ \frac{\lambda_h}{2}\sum_{i=1}^{m}\sum_{h=1}^{q} I_{ih} Z_{iht} \sum_{U_f \in M_{ih}} S_{if}||U_i - U_f||_F^2 \end{aligned} \tag{8}$$

The superscript O stand for the data comes from the Outgoing check-in data.

As for neighbouring check-in situation, we utilize the geographical character to explore the track similar friends who probably both have the same living area and similar preference. And the object function is:

$$\begin{aligned} E^N = &\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n^N} I_{ij}(r_{ij} - u_i^T l_j^N)^2 + \frac{\lambda_u}{2}||U||_F^2 + \frac{\lambda_{l^N}}{2}||L^N||_F^2 \\ &+ \frac{\lambda_h}{2}\sum_{i=1}^{m}\sum_{h=1}^{q} I_{ih} Z_{iht} \sum_{U_f \in M_{ih}} Soc_{if}||U_i - U_f||_F^2 \end{aligned} \tag{9}$$

The superscript N stand for the data comes from the neighbouring check-in data.

Soc_{if} contains both the geographical similarity G_{if} and interests similarity S_{if}, which is defined as:

$$\begin{aligned} d_{if} &= \sqrt{|L_{h_i} - L_{h_f}|^2 + |L_{w_i} - L_{w_f}|^2} \\ {G_{if}}^{-1} &= \frac{d_{if}}{min\ d_{if}} \\ Soc_{if} &= \beta\ G_{if} + (1-\beta)\ S_{if} \end{aligned} \tag{10}$$

L_{h_i} is the home location of user u_i, and L_{w_f} is the working place location of u_i's friend u_f.

III. Experiments

After doing enough experiments, this part is the second part you can write.

A. Experimental Setting

Accuracy measurement.

Run the experiment on what type of machine/what type of cloud.

What methods have been compared.

B. Data Collection

Where/when/how to collect the data.

Data characterization, may require one table to describe them.

C. Result Analysis

Case study: what you have discovered?

Performance comparison: compare the proposed method with the existing methods in terms of some performance measurement here.

Time complexity analysis: compare the proposed method with the existing methods in terms of some time complexity performance measurement here.

IV. Conclusion

Conclusion can be written after the other parts have been finished.