PREDICTING VENUES IN LOCATION BASED SOCIAL NETWORK

Omar F.Almallah and Songül Albayrak

Department of Computer Engineering,
Yildiz Technical University, Istanbul, Turkey

ABSTRACT

The circulation of the social networks and the evolution of the mobile phone devices has led to a big usage of location based social networks application such as Foursquare, Twitter, Swarm and Zomato on mobile phone devices mean that huge dataset which is containing a blend of information about users behaviour's, social society network of each users and also information about each of venues, all these information available in mobile location recommendation system .These datasets are much more different from those which is used in online recommender systems, these datasets have more information and details about the users and the venues which is allowing to have more clear result with much more higher accuracy of the analysing in the result.

In this paper we examine the users behaviour's and the popularity of the venue through a large check-ins dataset from a location based social services, Foursquare: by using large scale dataset containing both user check-in and location information .Our analysis expose across 3 different cities.On analysis of these dataset reveal a different mobility habits, preferring places and also location patterns in the user personality. This information about the users behaviour's and each of the location popularity can be used to know the recommendation systems and to predict the next move of the users depending on the categories that the users attend to visit and according to the history of each users check-ins.

KEYWORDS

Personalized Recommendations, Location based social networks.

1. INTRODUCTION

The evolution of the mobile phone has led to big usage of the location based social network application such as Twitter, Facebook, Google latitude and foursquare. The location based service (LBS) can be defined as a software service using the location information data to control features and this information service has now a big number of users who is depending on it in different majorities like health, entertainments and personal life. LBS contain services to recognize a location and give the longitude and the latitude of a person or object such as restaurant or the location of friends, LBS allow to the users to track their package and know where is it and in which time it will send or came.

Location based social networks are a type of social network in which the geographic services are the main object in this system which enable the additional social network dynamics such as geo-coding and geo-tagging, according to that the users can visit any location in the world and make geo-tagging about anything the users want to show it like photos, comments and videos. In particular, the geographic services presents it by using three layers first for users, locations and the content layer [1][2]. Here it's sure that the users can exploit information from one layer to another, according to these layer we can calculate the similarity of the information between the users for instance, the places that most of the users can prefers. LBSNs improved the quality of service on: firstly, the recommendation of social events, places, friends and activities, secondly, users behaviour and community detection, finally, personalized recommendation of social events, locations and friends.

Users of location based social network application can records the location that had been visited it, referred to 'check-in'. The check-in is mostly consist of active users, date & time, places and accompanying people. This records of each check-in allows to the users to keep track the place they have been visit [3], also these records allow to the user to receive notifications about there other contact and where they made check-in and they opinion about it. On other hands the users can check the comment of the other users about the places that the user want to visit it these application allow to check everything before going to it. While the user is wondering notification can received about each area passed from directly the program is showing the rate and the comment of other users about this place, before all of these applications also allow to the user to check the place and everything about it from comment and rate and give his own one about that place. The other benefit is to tag the other user who is with the user and show it as a notification in the application. These applications allow to there users to select a list of friends and create their own list of friends as a kind of social networking system [4].

One of the most popular location based social network application is Foursquare. Foursquare is a mobile application which become too much popular as local search and discovery app which is recommending the users to find which places they prefer, also provides recommendations of the locations which is much more close to the users. Foursquare helps and let the users to search and look for different categories such as restaurants, markets, malls, sport places and so on ,also help to look and search for another places by entering the names of the place. Foursquare also displayed the personalized suggestion depending on the time like in the morning will recommend a breakfast places and so on .The recommendations in this app will depend on the history of the users and what they always search and looking for, like if the user is looking always for dinner places it will directly suggest for the user new places for dinner.

In this paper, we present an examination for recommendation system depend on the users behaviours and the popularity of the places and location in different categories that the users of the location based social networks applications are looking for. We provide a prediction system for three cities through Foursquare program to prediction the most popular city and to predict the next place that the user will look for depend on the history of the users and also the categories that the users prefer to attend to go.

2. RELATED WORK

There are two main research related to our work the first one is about recommendations system and the second one is about the prediction system using the history of the user and the prediction system according to the categories.

There are too many number of researches about data mining algorithms that basis on the recommendation system idea [5]. These systems are taking the users performance and chooses such as rating, comments and tags as an input to predict the new move for the user and to show the popularity of the venues, all these is calculating under the term of collaborative filtering. In most of new research work has focus about the dataset which had been taken from the web site, namely movies such as Netflix [6] and music such as Yahoo Music and also another program which give a huge dataset.

Too many literature focus on the idea of using big data in there recommendation system, in [7], they trade on the check-in categories data to model implicit user motion pattern. In there research they focus on the prediction categories according to the user activities, also the prediction according to the location given by the rated category distribution.

In another review, researcher depends on there analysed about user's check-in history also social interaction pattern by using network structures of Foursquare users and venues dataset, also they focus on the venues geographical information and its effect on the users behaviours when he/she choosing the place to better understand the sensitive factors [8].

Also there's research, which evaluate a series of ways and techniques for identification of users from their own check-in information. They applied techniques to analysed the data according to users check-in over time and also the frequency of visit specific location, the techniques was depending on the users identification to analysed the dataset (Trajectory-based Identification, Frequency-based Identification and Measuring the Complexity of the Identification Task) depending on spatio-temporal trajectory emerging for their users check-in. They applied also a hybrid way to exploit both types of information [9].

On other hand, there are researches which depends on users history information to predict the next steps or the next place that the user planning to visit. The idea of these research is about when the users make a check-in by using Foursquare application and showing it by using another social network program such as Twitter. The dataset collected depending on each twit will contain any information about check-in through Foursquare or another location based social network Application [10].

The task of successive personalized point of interest (POI) in LBSNs considered by focusing on how to solve the POI recommendation or prediction by observed two prominent properties in the check-in sequence: personalized Markov chain and region localization. By submitted a matrix factorization method, namely FPMC-LR, to firm the personalized Markov qualifications and the centralized zone. The idea here is not just about personalized Markov chain in the check-in sequence, but also to look around the localized area [11].

For another point according to the interest of the user they focus on the problem of the time aware point of interest recommendation systems, which recommending to the user to visit new places at

a given time according to the users interest. To explain the geographical the temporal influence in the point of interest POI recommendation systems. They suggested Geographical-Temporal influences Aware Graph (GTAG) to calculate and record the check-in also the geographic and temporal influence. In this project to make the Geographical-Temporal influences Aware Graph (GTAG) more effective and more efficient recommendation they developed the performance of it by propagate methods named Breadth-first Preference Propagation (BPP). According to this algorithm the recommendation system will returns results within at most 6 diffusion steps. The recommendation systems results were time aware because of the perception that the users looking for different places at different time [12].

3. DATASET

In this section, we describe the dataset that we are dealing with what it contain for how many check-in made by how may user and the 3 cities that we are comparing it. Then we analysed the dataset by calculating the number of check-in in each venue and find the average and the number of check-in for each location also find the average number of check-in for each user.

3.1. Data Collection

Location based social networks have been very popular subject and this kind of datasets it become very interesting point to attracted millions of people footprint and what they are looking for. The dataset had been collected from foursquare program which is represent and deal with different categories and venues that the users of this program prefer to look for it. In order to study on personalized location recommendations foursquare dataset one of the best option in this time to understand what the users interesting in. Users can access foursquare by using their own phone very easily and show their status by make check-in in any place that the users wants. Through these programs users can write any comment about there status with the rate that want to give for this place, also make share and tags for anyone from there friends which is exist in there contents, also foursquare allow for the users to create there own list about the places that they prefers which it's the best for the users according to what the user is always looking for.

The foursquare dataset that we used in our research showing dealing with different kind of categories in three different cities which they are London, Austin and Dallas in (March and April of 2011) the dataset shows check-in in too many places for different categories according to too many users. Each of cities has different amount of check-in and also check-in in different places in each of the cities for different categories and positions. Each for these cities dataset has its own properties about the number of users, check-in and the categories. For London city dataset, it contain more than 4 millions check-in for more than 100 thousands users in more than 40 thousands different positions and places for different categories and also it contain the longitude and latitude of each venues and for the check-in it's explaining and exposure the date , time for each check-in and showing also the number of the street and the building, these dataset had been taken in 2011 for March and April, according to these dataset there's some user made more than 200 check-in and other less these number until one check-in in different categories ,date and time. According to all these information we can predict the best places also speculate the rush hour and give the best recommendation for the others users.

3.2. Data Analysing

In this part we analyzed our dataset for three cities acount the number of check-in for each city .Also we calculate the average number of check-in with the number of users and with the number of places .In order to have more clear result and in order to make to make the dealing withthis much big dataset much more easily .This calculation for the average for each city according to the number of check-in with the number of users and also with number of places as shown in Table.1

Table.1 Average properties observed on Forsquare over the specific date of the data set:Total number of user(N), places (M) and check-in (C), average number of check-in per user (Cu) and per place (CP)

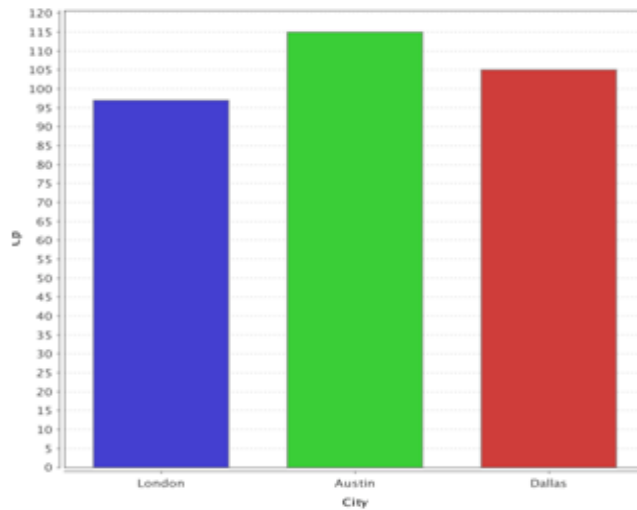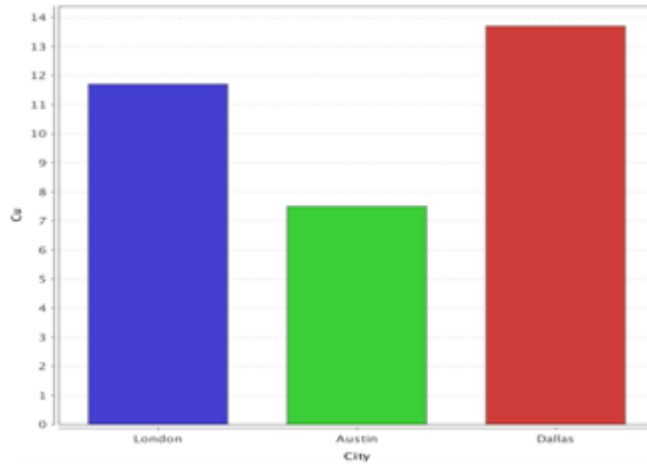| City	N	M	C	Cu	Cp
London	104,076	4,384	4,162,121	11.7	97
Austin	42,122	12,971	1,474,270	7.5	115
Dallas	35,593	15,751	1,637,232	13.7	105

Figure1. Average number of check-in per place

Figure 2. Average number of check-in per user

According to the table 1 we calculate the average number of check-in per place as shown in Fig.1 that the highest number is for Austin city, also the average number of check-in per user Fig.2 that the highest average is for Dallas city.

4. PREDICTION ANALYSIS

Location based social system network can improve the services on Generic recommendation of social activities, events, friends and location, another one is the personalized recommendations of activities, events, friends and users mobility. All the recommendation systems algorithms depend on the dataset in a different way such as following friends algorithms will depend on the friends what they prefer and like the algorithms which depending on the history of the users and according to that will advise the users to what they prefer to visit.

The personalized recommender system depends on the check-in histories of the users. Then, compare the history of users with each other and find the correlation of the performance for the users and suggest to the users new places, events and activities. In particular, the personalized recommender take the advantage of the time that some of the user has visited a place and give a rate or comment on that place and predict for the user unvisited place similar to the history of that user [13].

We applied some technical method on our dataset for the prediction system according to the history of the users and the number of the check-in that the user made in the three city also depending on categories and find the most popular one for the three cities In another hand predict the next place that the user will like to go according to the popularity of the place in our dataset to know the most popular city between Austin, Dallas and London by taking and counting the most high rating location and according to that we predict the popular city.

In this section, we applied a set of algorithms that we examined for the prediction system:

4.1. Visiting popular places

The first one is non-personalized baseline on the rank of the place also on the number of check-in that the users make in that place. We found the most popular places in the three cities:

Table.2 Austin city table for the most popular places

PLACE	CITY	CATEGORY	NO.OF CHECK-IN
Austin convention center	Austin	Convention Center	11,219
Austin bergstrom international	Austin	Airport	9,356
Starbucks	Austin	Coffee Shop	3,170
Seaholm power plant	Austin	Concert Hall	2,768

Table.3 London city table for the most popular places

PLACE	CITY	CATEGORY	NO.OFCHECK-IN
Starbucks	London	Coffee Shop	8,220
Terminal 5	London	Terminal	4,290
Apple Store	London	Electronic	3,301
LondonWaterloo Railway station WAT	London	Train Station	3,278

Table.4 Dallas city table for the most popular places

PLACE	CITY	CATEGORY	NO.OFCHECK-IN
Starbucks	Dallas	Coffee Shop	4,109
NorthPark Center	Dallas	Mall	1,158
Kroger	Dallas	Grocery Store	1,146
AMC Theatre	Dallas	Cineplex	3,278

The previous tables show the most popular places in each of the three cities depending on the number of check-ins in each of these places.

4.2. Attending places by categories

The next method is a content based filtering. The prediction system depend on the most popular categories that most of user attend to visit and the type of the categories that most of the user are looking for in all the three city.

Table.5 Austin city table for the most popular categories

CATEGORY	CITY	NO.OF LOCATION
Bar	Austin	18,839
Hotel	Austin	15,218
Mexican	Austin	11,428
Airport	Austin	9,388

Table.6 London city table for the most popular categories

CATEGORY	CITY	NO.OF LOCATION
Pub	London	36,544
Train Station	London	35,040
Coffee Shop	London	16,328
Bar	London	16,089

Table.7 Dallas city table for the most popular categories

CATEGORY	CITY	NO.OF LOCATION
Home	Dallas	9,756
Mexican	Dallas	7,462
American	Dallas	6,923
Bar	Dallas	6,109

According to the previous tables that it showing the most interested categories in each of the three cities such as the most popular category in Austin city is "Bar" which means all the places under the name of Bar category is get attend more than the other categories.

4.3. Attending Places by User History

This method is depending on the history of each user and giving the next move depending on what that user is prefer to visit and attend to go. Each user has a type of venues like to visit and according to that the program will speculate the next place for the user according to the history of him/her. We applied this method on our dataset for each user in the three cities and predict the next move for each user depending on their history and the number of check-in that the user made in his/her favourite venues.

Table.8 The most popular places in Austin city according to the history of selected users

PLACE	USERID	CITY	CATEGORY	CHECK-IN OFSAMPLE USER
Rista Bar and crill	34713233	Austin	Burgers	201
Parside at lake creek	57861921	Austin	Home	187
Browning Hangar	30935019	Austin	Sculpture	159

Table.9 The most popular places in Dallas city according to the history of selected users

PLACE	USERID	CITY	CATEGORY	CHECK-IN OFSAMPLE USER
Andy's House	208854807	Dallas	Home	231
Amy And Spencer	125855370	Dallas	Home	196
AAA Texas	29010688	Dallas	Oter_Travel	163

Table.10 The most popular places in London city according to the history of selected users.

PLACE	USERID	CITY	CATEGORY	CHECK-IN OFSAMPLE USER
Paddingtion Station PAD	20656631	London	Train Station	245
TFL Bus 100	40846787	London	Bus	178
London Liverpool street Railway	148302559	London	Train_stationl	169

In the previous tables is showing the prediction system for each city depending on the check-in of the users in all the three cities. According to the history of the users on Foursquare program to predict the next place that the user will go to it.

5. EVALUATION

We have evaluated the recommendation system algorithms result and compare it across the predictor, datasets and the three cities. In this section we describe our methodology and the metrics that we used to evaluate the recommendation system quality.

5.1 Methodology and Metrics

In this section we separate the dataset in to train and test in order to calculate accuracy result. We filter the dataset from zero and unknown check-in $C_{i,j}$ value to have more clear result. We use three metrics to calculate and find the quality of these recommendation system's result that we have found.

5.2 Result

We calculate the precision and the recall for each of the three method that we analysed our dataset on for all of the three cities to major the quality of our prediction and analysis system to find the best result as shown in Tables.

Table 11. Austin city the precision, recall and accuracy

Method / Austin city	precision	recall	accuracy
Popular places	0.5742	0.658	0.662
Attending by categories	0.1635	0.0739	0.215
User history	0.1056	0.16	0.1667

Table 12. Dallas city the precision, recall and accuracy

Method / Dallas city	precision	recall	accuracy
Popular places	0.526	0.6173	0.6214
Attending by categories	0.1306	0.0691	0.1625
User history	0.0576	0.09	0.0918

Table 13. London city the precision, recall and accuracy

Method / London city	precision	recall	accuracy
Popular places	0.4866	0.5691	0.5711
Attending by categories	0.259	0.1206	0.223
User history	0.1257	0.24	0.2414

The user will visit the top-N places as the result on the tables .The measured of the recall and the precision were depend on true positives (tp), false positives (fp), and false negatives (fn):

$$precision = \frac{tp}{tp+fp} \quad , \quad recall = \frac{tp}{tp+fn}$$

Two formulas to calculate the prediction system quality are precision and the recall with the accuracy that we applied on it to calculate the quality of prediction system.

According all of the three cities of our dataset the highest accuracy, precision and recall where about the prediction system depend on the popularity of the places.

6. CONCLUSION

In this paper we had described the prediction system focusing on three point in the dataset that we have from Foursquare the history for each of the users and what it will be the next step of them. We focus about attending new places by their categories and which category is the most popular one in our dataset. Also we concentrate counting the most popular place in each of the cities and predict the most wanted one according to the number of check-in that get taken on it. The goal of the three kind of prediction in our paper is to give a recommendation for the other users and the same users of the Foursquare program because this number that we found it's rating point for the users that they can depend on it for the next step of them.

In the term of the future work, we seek to evaluate a recommendation system that can calculate these number and give directly high rate for the prediction by applying another filter to have more specific result with high percentage of quality of prediction system.

REFERENCES

[1]   Needleman, Rafe; Claire Cane Miller; Adrianne Jeffries (3 September 2010). "Reporters' Roundtable: Checking in with Facebook and Foursquare". CNET. Retrieved 8 October 2010.

[2]   "Recommending Social Events from Mobile Phone Location Data", Daniele Quercia, et al., ICDM

[3]   M. C. Gonzalez, C. A. Hidalgo, and A.-L.Barabasi, "Understanding individual human mobility patterns," Nature, vol. 453, pp. 779–782, June 2008.

[4]   C.-Y. Chow, J. Bao, and M. F. Mokbel. Towards Location-based Social Networking Services. In ACM SIGSPATIAL-LBSN, 2010.

[5]   G. Adomavicius and A. Tuzhilin. Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE TKDE, 17(6):734–749, June 2005.

[6]   Y. Koren. Collaborative Filtering with Temporal Dynamics. In Proceedings of KDD '09, pages 89–97, Paris, France, 2009.

[7]   What's Your Next Move: User Activity Prediction in Location-based Social Networks by (Jihang Ye_ Zhe Zhu_ Hong Cheng).

[8]   What's Your Next Move: User Activity Prediction in Location-based Social Networks by (Jihang Ye_ Zhe Zhu_ Hong Cheng).

[9]   It's the Way you Check-in: Identifying Users in Location-Based Social Networks by: Luca Rossi and Mirco Musolesi. School of Computer Science University of Birmingham, UK l.rossi@cs.bham.ac.uk . Mirco

[10]   Location-based Predictions for Personalized Contextual Services using Social Network Data   by: Rui Zhang, Bob Price, Maurice Chu and AlanWalendowski . Palo Alto Research Center Inc.(PARC, a Xerox Company) 3333 Coyote Hill Rd. Palo Alto CA 94304 Rui.zhang@parc.com

[11]   Daniele Quercia, et al., "Recommending Social Events from Mobile Phone Location Data", ICDM 2010.

[12]   Graph-based Point-of-interest Recommendation with Geographical and Temporal Influences by: Quan Yuan, Gao Cong, Aixin SunSchool of Computer Engineering, Nanyang Technological University, Singapore 639798 {qyuan1@e., gaocong@, axsun@}ntu.edu.sg . Copyright 2014 ACM 978-1-4503-2598-1/14/11

[13]   p. symeonides et al. recommender systems for location based social network springersbriefs for electrical engineering.

AUTHORS

Omar Almallah
Computer engineering
I am a master student in yildiz technical university in Istanbul turkey.
Phone :05372675182
Email: omarfiras991@gmail.com

Dr. SONGüL VARLI ALBAYRAK
Associate Professor
YILDIZ TECHNICAL UNIVERSITY
Computer Engineering Department
songul@ce.yildiz.edu.tr