Volume 25 Issue 04, 2022 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

SIMILARITY BASED PERSONALIZED LOCATION RECOMMENDATION SYSTEM USING HUMAN MOBILITY DIARY AND SPATIO TEMPORAL DATA

Venkata Praveen Kumar V
Research Scholar, Department of CSE, Annamalai University, Chidambaram, Tamil Nadu, India

Dr. S. Pazhanirajan
Assistant Professor, Department of CSE, FEAT, Annamalai University, Chidambaram, Tamil Nadu, India

Dr. S. Indraneel
HOD and Professor, Department of Cyber Security & IOT, St. Ann's College of Engineering and Technology, Chirala, Andhra Pradesh, India

Abstract
Realistic spatio-temporal trajectories for human mobility are being generated and are used in various applications. This article is going to use these spatio-temporal trajectories and is going to present a framework to suggest visiting spot recommendations based on the trajectories of the similar individuals. The above framework works in 3 steps. First step involves in identification of visiting spots, second step involve in identifying the people who visited them and the third step is the recommendation of the visiting spots to the unvisited individuals. The proposed method adds more fields to the Mobility Diary proposed by Luca Pappalardo and provide an algorithm to identify the Visiting spots. Basing on the available historical data of the individuals, we identify the people who can be suggested with above recommendations. If the individual comes in the above category and if he/she has not visited then he will be suggested with the above recommendation. We compared the proposed method with real data and found that the method is generating results in more accurate way.
Keywords: Spatio Temporal Data, Location Recommendation, Point of Interest (POI), Data Mining, Human Mobility.

1. Introduction

Location recommendations are going to be popular in the nearby future with the increase in GPS enabled devices like cell phones, note books, i-phones etc. Instead of people choosing the locations, it will be nice if the devices suggest the best location for us to visit basing on our interests and basing on our GPS history. In our proposed work, Point of Interest (POI) recommendations are suggested by comparing our past spatio temporal routines with that of the similar people's routines.

In general, researchers use user preferences and sequential temporal information to recommend POIs. The earlier work in this area is done basing on Markov Chain model. All

Submitted: 15/09/2022 **Accepted**: 17/11/2022

though all previous models have improved POI recommendation from sequential modeling, none of them compared the spatio temporal routines with similar individual for better recommendations.

We exploit the DITRAS technique proposed by Luca Pappalardo to capture the routines of the similar individuals, we compare the routines and bring recommendations of POI. We have taken care that while comparison, the individual's daily activities are not considered. For Example, an individual may go to gym in the morning, then to office , then to a coffee shop in return to home and finally to groceries shop after refreshing at home. The above locations are a part of the individual's daily routine on week days.

So it is clear that individuals always visit the POIs around work area and home on weekdays and to shopping malls, pilgrims, refresh centers on weekends. Therefore the traces of movement vary differently on different days. Probably work on weekdays and entertainment on weekend.

Individuals visit the POIs that are geographically adjacent to them at normal weekends or when they have less number of holidays and the POIs that are much far when they get more number of holidays. Hence we consider the geographical influence and time factor also to develop the model.

The rest of this paper is organized as follows. In Section 2, we review the related work. In Section 3, we introduce real world algorithm and we introduce our proposed methods. In Section 4, we evaluate our proposed models and in Section 5, we conclude this paper and add possible future work.

2. Review of Related Work

Most of the studies of human mobility are at a stage of collecting data and analyzing the data. Understanding human mobility patterns is important for epidemic prevention, traffic forecasting, urban and transportation planning (Giannotti et al. 2013). A large heterogeneity is found in human mobility by observing the mobile phone dataset. 100000 mobile phones are tracked for six months period and it was found that in contrast with random trajectories, human trajectories follow a high degree of temporal and spatial regularity yielding to simple reproducible patterns (Gonzalez et al. 2008). It was found from the user data base, that the predictability of human mobility is 93 to 80% following a daily routine (Song et al. 2010). Poppalardo considered 10 million GPS trajectories and found that there is power law behavior in waiting times (Pappalardo et al. 2013). Mobile phone networks and social media records are a great source of human movements (barbosa-Filho et al. 2017). Call Detail Records (CDRs) collect geographical, temporal and interation information from mobile phones (Calabrese et al. 2011). When working with CDR user sampling is to be done carefully (Iovan C, et al. 2013). Social relationship can explain 10 to 30% of human Mobility (Choe E et.al. 2011). It is hard to collect data about city life. Rather than conducting surveys, it is easy to collect mobile phone data and analyze it (De Nadai M, et. al. 2016). Mobility models are derived from real traces using the marks the mobile users leave in wireless networks (Hess, et al. 2015).

People select popular places in the city to visit (Hasan s et al. 2013). Mobility scenario is a weighted contact graph. The mobility of an individual is directly proportional to the mobility of the contacts (Hossmann T et al. 2011). Multiple links exists between two users across different social networks forming the online social system a multiplex (Hristova D et al. 2015). Preferred locations, personal agendas and social graphs are the main reasons for human Mobility (Karnshuk D et al. 2011).

Data mining and statistical techniques provide a powerful source for analyzing the human mobility. Clustering of humans were done in the form of workers, students and non-workers (Jiang S et al. 2012).

Cities are being instrumented with digital devices and infrastructure to get huge amount of human mobility traces (Kitchin R et al. 2013). The interest in Spatio temporal human mobility should be in a holistic view (Kopp C et al. 2014).

Small world in Motion (SLAW) uses a bi-dimensional space to generate a set of locations. The distance between these generated locations is heavy tailed. Now a synthetic individual is supposed to move along these generated locations. The synthetic individual starts at a location and randomly chooses the next location. The movement of a synthetic individual is based on the concept of Least Action Trip Planning (LATP). As per LATP, every location has the equal probability for becoming the next location where the probability decreases with the power law of distance (Lee et al. 2012).

Some algorithms have tried to represent the human mobility and tried to represent how individuals move from one location to another. Small World in Motion (SWIM) is an algorithm designed based on location preference. Here, each synthetic individual choose the next destination on a bi-dimensional space. The chosen destination is based on the weight of the location. The weight of a location increases with the increase of popularity and decreases with the increase of distance. The popularity of a location is directly proportional to the number of visitors (Kosta S et al. 2010).

The Exploration and Preferential Return (EPR) does not fix the locations in advance; rather allow it to grow over time. This model works on the two basic principles called Exploration and Preferential Return. Exploration is the process of exploring new locations and Preferential Return is the tendency of the individual to visit to the location that was already visited. Here a synthetic individual has to select the next location by choosing one of the two methods explained above. Initially, the individual performs the Exploration method until all the nearby locations are visited which is called greedy exploration. The probability for the individual to explore decreases as the number of visited locations increase. Now the concept of preferential return comes into limelight. As most of the nearby locations are visited, the visit will be based on preferential return here after (Song et al. 2010).

There are two human mobility models namely exploration model and temporal models. The exploration models accurately reproduce heterogeneity but unable to go through the regularities in human temporal patterns. On the other hand, the temporal models able to reproduce the regular human mobility schedules failing in capturing some global mobility patterns. Luca Pappalardo proposed a model where the heterogeneity of human mobility and regularity of human movement are considered into account. Here d-EPR MD a generative algorithm is proposed. The temporal model of d-EPR MD is based on Markov chain and the spatial mechanism is based on EPR model. He also designed a data-driven algorithm Mobility Diary Learner (MDL). The modeling framework of Diary based Trajectory simulator (DITRAS) is to simulate the spatio temporal patterns of human mobility. DITRAS operates in two steps: the generation of mobility diary and the translation of the mobility diary into mobility trajectory. (Pappalardo L et al. 2017).

1. Data description and Analysis

In this section, we first introduce a Mobility Diary proposed by Luca Pappalardo et a. 2017 and then conduct an analysis on it. A mobility diary captures the temporal patterns of human mobility. A mobility diary contains a list of arrival times and the time spent at each location. Now the mobility diary is converted into mobility trajectory by capturing the spatial patterns of human mobility.

A mobility trajectory contains triples and is defined as T=<(x1, y1, tl),…(xn, yn, tn)>, where xi and yi are coordinates and ti is a time stamp where 1≤i≤n and ti<ti+1.

The mobility trajectory can be further converted into Sampled Mobility Trajectory S(t) = <l1, l2, …,ln> where li is a geographical location obtained by the coordinates (xi, yi).

2. Design

In this section, the proposed DLRS (Diary based Location Recommendation System) is introduced in detail. To the Sampled Mobility Trajectory, we add columns namely visiting spot and day. A value one in the visiting spot column denotes the spot a visiting spot. If the value is zero in the column, it means that the location is not a visiting spot. Let s1, s2, … sm be the number of spots visited by an individual in which v1, v2, … , vn be the visiting spots. A spot si is a visiting spot provided si is not in the daily routine of the individual.

- **Algorithm Find_Visiting_Spots**

Input: S={s1, s2, …, sm}, dataset of spots visited by an individual
Wd=Number of weekdays
K=constant

Output: V={v1,v2, …, vn} where vi is a visiting spot

```
//let k be the number of unique spots from S
    u1=s1;
    c1=1;//where ci represent the count of unique location
    k=1;
```

```
    For i=1 to m //for all the visited locations
        for j=1 to k //for all the unique locations
            if(si==uj) //if visited location is already a unique location
                cj=cj+1//increment the corresponding count
                break;
            end
        end
        if j==k//if visited location is not in unique location list
            k=k+1  //increment the count of unique locations
            uk=si //add visited location to unique location
        end
    end
    th=wd*k //calculate threshold value
    g=1
    for j=1 to k //for all unique locations
        if tk>cj //if threshold value greater than count of unique location
            if g==1 //if it is the first entry of the visiting spot
                v1=uj //add uj as the visiting spot list of the list
                g=g+1 //increment visiting spot count
            end
            else
                vg=uj //add uj to the visiting spot list
                g=g+1 //increment visiting spot count
            end
        end
    end
    return V //return the visiting spot list
end
```

A spot si is a visiting spot vj provided the periodicity of si is less than the threshold value. For example, on a daily routine, an individual may have his routine like gym, office, coffee shop and a shopping mall. For this individual, the gym, office, coffee shop and shopping mall cannot be a visiting spots.

The threshold value for a location si can be obtained by the following formula. The threshold value tv for a location si is calculated as

tv= wd*k

Where wd is the number of working days and k is a constant which varies from sector to sector.

If $\sum si < tv$ where $1 \leq i \leq m$, the location si be converted into visiting location vj. Thus from m locations of an individual, we get n visiting spots. From the m spots visited by the individual, k unique spots are identified by eliminating redundancy from which n visiting locations are identified basing on the threshold value.

The algorithm to identify the similarity of visiting spots is as follows.

Algorithm Suggest_Visiting_Spots

Input: V = {VI1, VI2, VI3,…, VIT} //visiting spots of all the individuals.

Output: R ={R1, R2, R3, …, RT}; //Visiting spot recommendation for each individual

```
For i=1 to T
        for i=1 to T
        for j= 1 to T
                if i==j
                        skip
                else
                        n1=number of visiting spots of VI1
                        n2=number of visiting spots of VI2
                        index=0
                        count=0
                        for k=1 to n1
                                for m=1 to n2
                                        if VI1[k]==VI2[k]
                                                count++
                                        end
                                end
                        end
                        if count>=0.9*n1
                                for m=1 to n2
                                        found=0
                                        for k=1 to n1
                                                if VI2[m]==VI1[k]
                                                        found=1;
                                                        break;
                                                end
        end
                                end
                                if found==0
                                        R[i][index++]=VI[j][m]
                                end
                        end
                end
        end
end
```

return R
end

Using the above methodology, we find the visiting spots of multiple individuals from which we can get huge visiting spots. The pool of visiting spots is stored in a separate array.

Now the visiting pattern of each individual is compared with similar individuals by excluding their daily routines. For each individual in the data set, we check for similarity of routine with every other individual. If T is the total number of individuals in the data set then V is the set of visiting spots for each individual.
V = {VI1, VI2, VI3,…, VIT}

Where VI_i represent the Individual i's visiting spots. The total number of visiting spots varies from individual to individual.

3. Experimental Evaluation

We evaluate the proposed method on the dataset called TIST. This data set includes long-term global scale data collected for 18 months. It contains 33,278,683 check-ins by 266,909 users on 3,680,126 venues (in 415 cities in 77 countries). Those 415 cities are the most checked 415 cities in the world, each of which contains at least 10000 check-ins.

This data set further contain 3 sub datasets in which, the first dataset contain 4 columns namely user ID, Venue ID, time, time zone offset in minutes. The second dataset contains 5 columns namely venue ID, latitude, longitude, venue category name, country code. The third dataset consists of 6 columns namely city name, latitude, longitude, country code, country name and city type.

Precision and recall are the two metrics that are used to evaluate a Location recommendation system. Precision is the ratio of successfully predicted locations to the top N recommendations and Recall is the ratio of the number of successfully predited locations to the top K recomendations.

Actual ⇨ Predicted ⇩	Positive	Negative
Positive	True Positive	False Positive
Negative	False Negative	True Negative

$$\text{Precision}=\frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

$$\text{Recall}=\frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

Comparison with other methods:

We also compare DLRS with other methods that adopt similar ideas. Here we coparet the effectiveness of overall recommendations. The following two figures show the performance of various approaches. All approaches are shown in terms of their bbest performance (i.e. the performance under the optimal parameter settings).

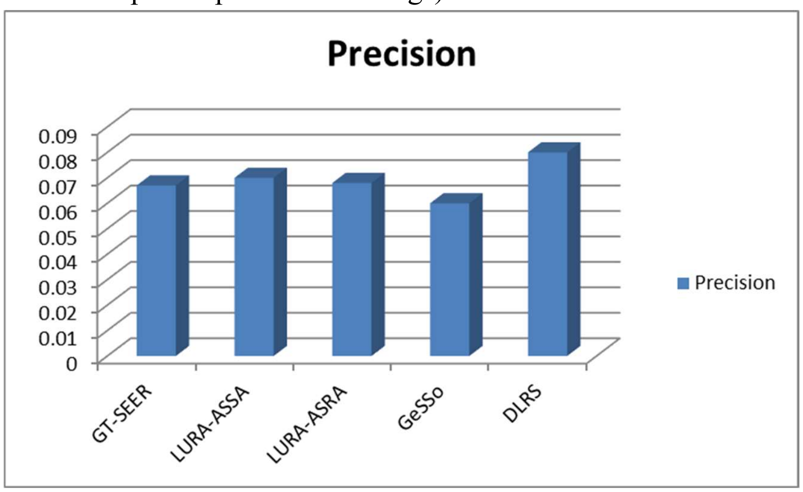

Fig 1. Comparison of DLRS with other popular Location Recommendation Systems with respect to Precision.

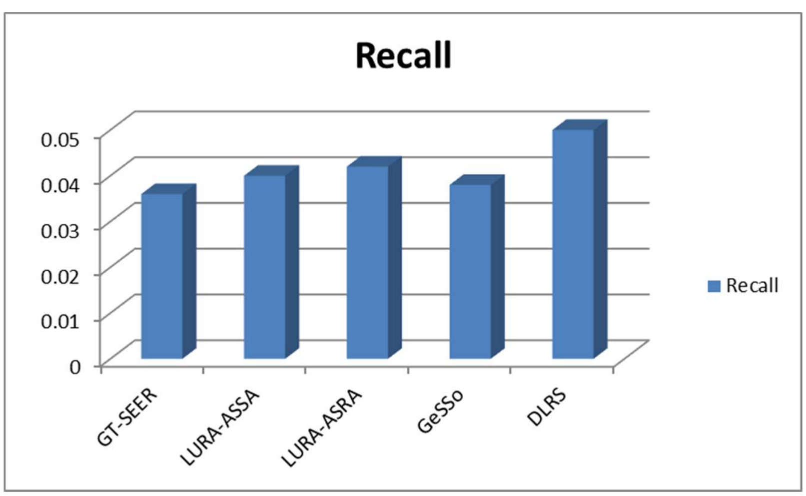

Fig 2. Comparison of DLRS with other popular Location Recommendation Systems with respect to Recall

Geo – Temporal Sequential Embedding Rank for Point-of-Interest (GT-SEER) is an application in location-based social network. In this method geographical influence is added to the temporal sequential Embedding Rank for better results.

LURA is a method that recommends new location based on the principle of Learn-Update-Recommend-Aggregate. Our results are compared with the best sampling and aggregation strategies namely LURA-ASSA and LURA-ASRA.

Geographical and Synthetic Social Influences (GeSSo) uses a kernel estimation approach with a quartic kernel function to model geographical influences.

Our proposed model DLRS is generation results in more accurate way than any other model.

4. Conclusion & Future Scope

In this work, we focus on personalized location recommendation. The results of our experiment prove that the proposed model is far better than the existing models. Existing models mainly concentrated on location based social networks, the similarities with neighbors, friends and social relations. They also concentrated on past check-ins of the user. But our proposed model compares with the movement of the similar individual though they are not in our social network. We believe that the results of this work shall be useful to human community and can also be implemented in upcoming technologies of personalized location recommendation.

The future scope of this model involves in suggesting the amount of time required to visit the proposed location which includes the min time, max time and average time spent by an individual in the proposed location.
The other future scope can be suggesting the location based on time availability of the individual. Like, the system recommends nearby places if the available time is small and will suggest far places if the amount of time available is large. The suggestion will also include the travel time.

• References

[1] Barbosa-Filho H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F, Tomasini M (2017) Human mobility: models and applications. arXiv:1710.00004

[2] Calabrese F, Colonna M, Lovisolo P, Parata D, Ratti C (2011a) Real-time urban monitoring using cell phones: a case study in rome. IEEE Trans Intell Transp Syst 12(1):141–151. https://doi.org/10.1109/ TITS.2010.2074196. ISSN 1524-9050

[3] Cho E, Myers SA, Leskovec J (2011) Friendship and mobility: user movement in location-based social networks. In: Proceedings of the 17thACMSIGKDDinternational conference on knowledge discovery and data mining, KDD'11. ACM. pp 1082–1090

[4] De Nadai M, Staiano J, Larcher R, Sebe N, Quercia D, Lepri B (2016) The death and life of great Italian cities: a mobile phone data perspective. In: Proceedings of the 25th international conference on world wide web,WWW'16, pp. 413–423, Republic and Canton of Geneva, Switzerland, 2016. International WorldWideWeb Conferences Steering Committee. https://doi.org/10.1145/2872427.2883084. ISBN 978-1-4503-4143-1

[5] Giannotti F, Pappalardo L, Pedreschi D, Wang D (2013) A complexity science perspective on human mobility. In: Mobility data: modeling, management, and understanding. pp 297–314

[6] González MC,HidalgoCA, Barabási A-L (2008) Understanding individual human mobility patterns.Nature 453(7196):779–782. https://doi.org/10.1038/nature06958

[7] Hasan S, Schneider CM, Ukkusuri SV, González MC (2013) Spatiotemporal patterns of urban human mobility. J Stat Phys 151(1–2):304–318. https://doi.org/10.1007/s10955-012-0645-0

[8] Hess A, Hummel KA, Gansterer WN, Haring G (2015) Data-driven human mobility modeling: a survey and engineering guidance for mobile networking. ACMComput Surv 48(3):38:1–38:39 10.1145/2840722

[9] Hossmann T, Spyropoulos T, Legendre F (2011) A complex network analysis of human mobility. In: 2011 IEEE conference on computer communications workshops (INFOCOM WKSHPS). pp. 876–881 https://doi.org/10.1109/INFCOMW.2011.5928936

[10] Subbarao Gogulamudi, Dr. V.Mahalakshmi, Dr. Indraneel Sreeram. (2020). To Improve the Efficiency in Sentiment Enlists. International Journal of Control and Automation, 13(02), 324 - 331. Retrieved from http://sersc.org/journals/index.php/IJCA/article/view/8101

[11] Indraneel Sreeram, Venkata Praveen Kumar Vuppala, HTTP flood attack detection in application layer using machine learning metrics and bio inspired bat algorithm, Applied Computing and Informatics, Volume 15, Issue 1, 2019,

Pages 59-66, ISSN 2210-8327,

https://doi.org/10.1016/j.aci.2017.10.003

(https://www.sciencedirect.com/science/article/pii/S2210832717301655)

[12] Venkata Praveen Kumar Vuppala and Indraneel Sreeram, 2018. Wavelet Transformation and Wavelet Network Classifier for ECG Classification. Asian Journal of Information

Technology, 17: 131-141.DOI: 0.36478/ajit.2018.131.141

https://medwelljournals.com/abstract/?doi=ajit. 2018.131.141

[13] Iovan C, Olteanu-Raimond A-M, Couronné T, Smoreda Z (2013) Moving and calling: mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies. In: Springer (ed) 16th international conference on geographic information science (AGILE'13). pp 247–265 https:// doi.org/10.1007/978-3-319-00615-4_14

[14] Sreeram,I., and Vuppala,V.P.K. (2017). Voice Enabled ATMS and Deployment of Wavelets for Recognition and Authentication of Voice Based Inputs. i-manager's Journal on Computer Science https://doi.org/10.26634/jcom.4.4.13417

[15] Jiang S, Ferreira J Jr, González MC (2012) Clustering daily patterns of human activities in the city. Data Min Knowl Disc 25(3):478–510. https://doi.org/10.1007/s10618-012-0264-z

[16] Karamshuk D, Boldrini C, ContiM, PassarellaA(2011) Human mobility models for opportunistic networks. IEEE Commun Mag 49(12):157–165. https://doi.org/10.1109/MCOM.2011.6094021

[17] Kitchin R (2013) The real-time city? big data and smart urbanism. GeoJournal 79(1):1–14. https://doi.org/ 10.1007/s10708-013-9516-8. ISSN 1572-9893

[18] Kopp C, Kochan B, May M, Pappalardo L, Rinzivillo S, Schulz D, Simini F (2014) Evaluation of spatiotemporal micro simulation systems. In: Knapen L, Janssens D, Yasar A (eds) Data on science and simulation in transportation research. IGI Global, Hershey

[19] Kosta S, Mei A, Stefa J (2010) Small world in motion (SWIM): modeling communities in ad-hoc mobile networking. In 2010 7th annual IEEE communications society conference on sensor, mesh and ad hoc communications and networks (SECON). IEEE. pp 1–9. https://doi.org/10.1109/secon.2010.5508278. ISBN 978-1-4244-7150-8

[20] Lee K, Hong S, Kim SJ, Rhee I, Chong S (2012) Slaw: self-similar least-action human walk. IEEE/ACM Trans Netw 20(2):515–529. https://doi.org/10.1109/TNET.2011.2172984. ISSN 1063-6692

[21] Lenormand M, Gonçalves B, Tugores A, Ramasco JJ (2015) Human diffusion and city influence. J R Soc Interface 12(109). https://doi.org/10.1098/rsif.2015.0473. ISSN 1742-5689

[22] Lenormand M, Bassolas A, Ramasco JJ (2016) Systematic comparison of trip distribution laws and models. J Transp Geogr 51:158–169. https://doi.org/10.1016/j.jtrangeo.2015 .12.008. ISSN 0966-6923

[23] Munjal A, Camp T, Navidi WC (2011) Smooth: a simple way to model human mobility.

In: Proceedings of the 14th ACM international conference on modeling, analysis and simulation of wireless and mobile systems, MSWiM ’11. ACM, New York, NY, USA. pp 351–360. https://doi.org/10.1145/2068897. 2068957

[24] Pappalardo L, Rinzivillo S, Qu Z, Pedreschi D, Giannotti F (2013) Understanding the patterns of car travel. Eur Phys J Spec Top 215(1):61–73. doi:10.1140/epjst%252fe2013-01715-5

[25] Pappalardo L, Rinzivillo S, Pedreschi D, Giannotti F (2013a) Validating general human mobility patterns on gps data. In: Proceedings of the 21th Italian symposium on advanced database systems (SEBD2013)

[26] Song C, Qu Z, Blumm N, Barabási A-L (2010) Limits of predictability in human mobility. Science 327(5968):1018–1021. https://doi.org/10.1126/science.1177170

[27] Hristova D, Noulas A, Brown C,MusolesiM,Mascolo C (2015) A multilayer approach to multiplexity and link prediction in online geo-social networks. EPJ Data Sci 5(1):24. https://doi.org/10.1140/epjds/ s13688-016-0087-z. ISSN 2193-1127