Prediction of Human Mobility using Mobile Traffic Datasets with Hidden Markov Model

A report submitted in partial fulfillment of the requirements of

M.Tech. Major Project

by

Anshika Rawal
(15CSE2003)

Under the guidance of

Dr. Pravati Swain

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY GOA

FARMAGUDI, PONDA, GOA-403401, INDIA

MARCH, 2017

Department of Computer Science and Engineering
National Institute of Technology Goa
Farmagudi, Ponda, 403 401, Goa, India

CERTIFICATE

This is to certify that the project report titled **"Prediction of Human Mobility using Mobile Traffic Datasets with Hidden Markov Model"** is submitted by **Miss. Anshika Rawal** bearing Roll No. **15CSE2003** pursuing M.Tech. program in Department of Computer Science and Engineering. The content in this report is a bonafide record of work carried out by him under my suervision.

Examiner 1:

Examiner 2:

Examiner 3:

Dr. Pravati Swain
(Supervisor)

Acknowledgment

First and foremost, I would like to express my heartfelt gratitude and regards to the Director of National Institute of Technology Goa (Prof. Udaykumar Yaragatti) for his unmatched support and blessings.

I express my deepest gratitude to my supervisor, Dr. Pravati Swain, for her constant support and guidance throughout the project. I thank her for constantly motivating and encouraging me towards the successful completion of my project and her guidance enabled me to complete the project with perfection. I can not think of anyone else to thank first, other than reviewers for their constructive and insightful comments. It is hard to express in words, the magnitude of my admiration and respect for Dr. Pravati Swain. I personally feel that Dr.Pravati Swain is the best advisor anybody can get. Her idea of conducting weekly two meetings with their students to evaluate their work and provide proper guidance for future work. I am thankful to her countless number of suggestions that helped me to take my work at better level. I am also thankful for the persons who gave me their valuable advice and support to carryout this work more effective. I express my deep sense of gratitude and reverence to my beloved parents who supported and encouraged me all the time, no matter what difficulties are encountered.

Above all, I would like to thank The Almighty God for the wisdom and perseverance that He has been bestowed upon me during this project work, and indeed, throughout my life.

Abstract

From time untold, humans have developed various forms of communication to express their views. Cellular communications have made it possible to connect every corner of the world. Thus, from this mobile traffic data, we can infer the information about user mobility. The analysis of human location histories is getting attention in present; this is a recently emerged research field. Prediction of human mobility can be addressed by mobile traffic data as it is cost efficient as well as gives remarkable performance. Human mobility prediction is critical to efficient data acquisition and to build a relationship among mobile traffic generated at different locations. This paper proposes a framework for predicting human mobility is based on Hidden Markov Models (HMMs). In this proposed approach, first locations are clustered according to their characteristics, and later trains an HMM for each cluster. The usage of HMMs empowers to deal with spatio temporal data, location characteristics or clusters are unobservable parameters and set of possible location can be visited are visible symbol. The Proposed framework runs successfully on a real time and large mobile traffic dataset.

Contents

List of Figures

Chapter 1

Introduction

Prediction of human mobility possesses a widespread usage in the context of location based services. To predict human location enough amount of work has been done in both short and long timeframes. Now geo positioning techniques are ubiquitous, where some techniques such as GPS made it easy to collect interesting data in various domains. Such systems are becoming popular due to the increasing popularity of devices in the recent years. Most of the previously proposed algorithms are designed on the basis of GPS signal, Radio Frequency (RF) and Wi-Fi, which do not consider any constraints on power consumption. On the contrary, battery usage is a most noticeable issue in case of GPS sensors because GPS sensors having high battery consumption. So these techniques are impractical if continuous movements have to be observed and such system does not deal with uncertainty as well. The knowledge of a person's motion can be easily and excellently be determined by the Mobile data which most of the people are using very frequently and is accessible to all [1]. Mobile traffic data provide required information very easily. Mobile traffic data gives the more coverage than any other technology, the most focused and important point can be made is that the operating cost for this kind of system is virtually negligible. Hence, Mobile traffic analysis technology has rapidly established as a key new source in the field of mobility modeling.

1.1 Mobile Traffic Analysis

As the population is increasing, the more number of people are attracted to it and using the cutting edge of technology every day. The statistics show that a vast fraction of the population is using mobile phones making the indirect effect on the success of this technology. More and more new consumers are using it and the mobile phones required to dynamically interact with network infrastructure, Internet and users logged to the resources to interact with various services [2] like resources and billing management which are associated with georeferenced events. So mobile traffic analysis gives the information related to user mobility and their tendency of movement. Mobile traffic analysis is better than GPS signal reading due to its battery efficiency [3] as well as the spurious signal problem also not there. So user mobility can be analyzed by Mobile traffic dataset as it is an efficient way over traditional methods.

1.2 Mobility Analysis

Mobility analyses deal with the extraction of mobility information from mobile traffic. Mobility is Intended here in its broadest acceptation, and includes generic human movements at both individual or aggregate levels, as well as specialized patterns that concern specific users, e.g., traveling on transportation systems. We also review in following section literature survey on the dependability of mobile traffic data as a source of mobility information. Mobility models are derived from the mobile data which contributes to a various no. of domains such as, human sociology[4], urban planning and traffic engineering. In the following, we review most relevant works that leverage mobile data to study human mobility.

1.3 Human Mobility

Human mobility is a general term that gives the information about human movement and their characteristics. Human mobility patterns produce some fundamental rules

that govern the prediction of next movement. Here, one simulative or mathematical model needs to be constructed to reproduce such patterns, so that prediction can be done with adequate accuracy. The characterization of predicted human mobility from mobile data concerns about:

- To investigate the fundamental things that governs the patterns in movement.
- To propose such mathematical models that is capable to regenerate these patterns.

Visited Locations: Detection of the geographical locations which an individual visited and it represents the crucial and first part by studies that enable mobile data to infer law about human mobility. It can be any location throughout the map of that area which is only bounded by the network coverage. This pin points the area been visited by that individual and can be used as a measure of person's movement from one place to other.
Travel distance: Travel distance: The distance form subsequent location which gives the information of the total traveled distance by an individual has also attracted significant attention as it tells all the physical units such as speed and acceleration.
Spatio-temporal regularity: Results of mobile data analysis show that individuals tend to have regularity in most of their pattern of movements which determines for how long a person been in a spatial locality and helps in the predictability.
Predictability: Due to strong regularity of human mobility patterns this can be inferred that how easy to predict individual's movements which can be such as telling where an individual is visiting frequently or seldom. How many times an individual is going to visit that place and when?
Factors affecting mobility: There are large numbers of factors that affect human mobility laws inferred from strong regulated patterns. An example is that of movement patterns possess diversity in areas with different development levels or topological features.
Models of Human Mobility: There are two broad categories of models of human mobility first describe the movement of individuals and second concerns about the movement of aggregated dynamics of whole populations.

1. Model for Individual's mobility: This indicates the movement of an individual user independently [5].

2. Model for Aggregated mobility: This type of model describes the movement of a large number of people with low spatial granularity.

Chapter 2

Related work

In this section, we review the quite extensive literature on the dependability of mobile traffic data as a source of mobility information. Mobility models are derived from the mobile data which contributes to a various number of domains such as, extracting user's location patterns from location history [4], urban planning [6] and traffic engineering. In the following, we review most relevant works that leverage mobile data to study human mobility. There are various applications of mobile traffic data such as paper [6], focused on detection of urban fabrics which is one specific application of mobile traffic analysis. The urban fabrication of the metropolitan area emphasizes on such aspects as, type of infrastructure, transportation, roads, sports facilities, industrial plants and human activities.

Various techniques have been used to deal with the prediction of human mobility from mobile traffic data. The papers [7], [8], [9] use a model based on frequent patterns of visited locations and data mining techniques. In Paper[10], GPS dataset has been used for predicting human mobility on the basis of Hidden Markov Models (HMMs). The prediction accuracy of 13.85 percent has been achieved by considering regions of approximately 1280 square meters. In paper [11], a model is proposed based on a HMM which is resistant to uncertain location data, as it works with data collected by using cell-towers to localize the users instead of GPS devices, and

reaches good prediction results in shorter times. They have stated that keeping track of location by GPS is not feasible because of its battery consumption and spurious signal problem. However, the impact of time parameter on the predication model is missing. The paper [12] proposes a social relationship based mobile node location prediction algorithm using daily routines. After considering user's dynamic behavior resulting from their different daily schedule, they have proposed algorithm to predict user's mobility in different daily time periods. Then prediction results are amended using other user's location information which has strong relationship to that particular user. In paper [13], Sebastien et al. have addressed the issue of predicting the next location of an individual based on the observations of his mobility behavior over some period of time and the recent locations that he has visited using Mixed Markov Chains (MMC). This work has several potential applications such as the evaluation of geo-privacy mechanisms, the development of location-based services anticipating the next movement of a user and the design of location-aware proactive resource migration.

Chapter 3

Problem Statement

The estimation of probability transition of an individual or aggregate of people from the current location to another location in future, by capturing the mobile traffic data. For the above requirement, this paper presents a probabilistic model to predict human mobility using large scale mobile traffic data. The Hidden Markov Model (HMM) has been used as the prediction model. The proposed framework is divided into two phases where in the first phase, the geographical locations are clustered according to their characteristics such as the highest traffic generated in a particular time period. In the second phase, the proposed HMM will be trained by the generated clusters. The usage of HMMs empowers to deal with spatiotemporal data, location characteristics and possible visited states which are the observable states. Some important basic methodologies are depicted below in context of their usage in proposed approach:

3.1 Agglomerative Hierarchical Clustering

This is a hierarchical clustering approach by initiating with each point as a singleton cluster and then repeatedly merging the two closest clusters until a single, all-encompassing cluster remains, the Algorithm in detail as follows [14]:

Algorithm 3.1: Agglomerative Hierarchical clustering

Let $Y = \{y_1, y_2, ..., y_n\}$ be the set of data points.

- Begin with the disjoint clustering having level L(0) = 0 and sequence number n = 0.
- Calculate the least distance pair of clusters p and q, according to dist[p, q] = min dist[i, j], i is not equal to j and the minimum function is over all pairs of clusters (i and j) in the current clustering.
- Increment: n = n+1. To form the next cluster (n) by merging of clusters (p) and (q) into a single cluster.
  Set the level of this clustering to L(n) = dist[p, q].
- Delete the rows and columns of the corresponding cluster p and cluster q. Then add new a row and column for the newly formed cluster to update the distance matrix, D. The distance between the new cluster, denoted (p,q) and old cluster(k) is defined as: d[(k), (p,q)] = min d[(k),(p)], d[(k),(q)].
- Stop, if all the data points are in one cluster else repeat from step (2)

3.2 Hidden Markov Model (HMM)

Hidden Markov Models (HMMs) [15] are well known approach for the prediction of sequential data. The HMM shown in Figure 3.1 is used to predict future states given the present state. A Discrete time HMM is characterized by λ = (A, B, π) where:

- N number of hidden states: $\{s_1, s_2, ..., s_n\}$ and M distinct observations $\{y_1, y_2, ..., y_n\}$.
- State transition probability distribution A, i.e.,
  $A = a_{i,j}, a_{i,j} = P(q_{t+1} = Sj|q_t = S_i)$ 1 ≤ i, j ≤ N
- Observation symbol probability distribution B, i.e.,
  $B = b_{i,k}, b_{i,k} = P(O_t = y_k|q_t = S_i)$ 1 ≤ i ≤ N , 1≤ k ≤ M.
- Initial state distributiRabineron , i.e., $\Pi = \Pi_i$, $\Pi_i = P(q_1 = S_i)$ 1≤i≤N

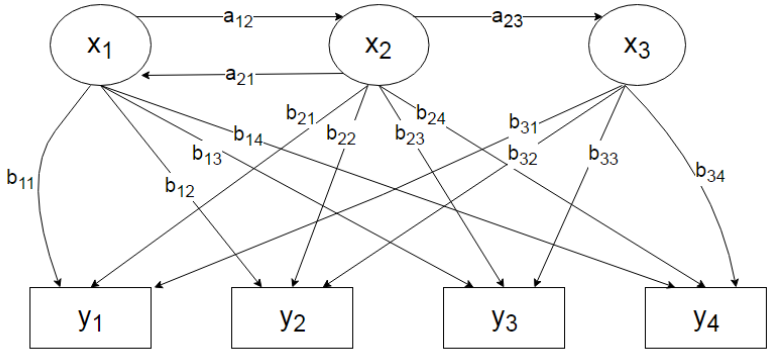

Figure 3.1: Discrete HMM with 3 States and 4 Observation States.

The HMM is a doubly embedded stochastic process, where an observation is a probabilistic function of a state. These are the central issues in HMM:

Evaluation problem: Given the observation sequence $O = \{o_1, o_2, ..., o_t\}$ and a HMM, λ = (A, B, π), What is the probability that the sequence of observation will be generated by given HMM, i.e., P(O| λ).

Reconstruction or Decoding problem: Given the observation sequence $O = \{o_1, o_2, ..., o_t\}$ and a model λ = (A, B, π), how do we choose a corresponding state sequence $Q = \{q_1, q_2, ..., q_t\}$ which is optimal in some sense, i.e., best explains the observations.

Training: Given the sequence of observation $O = \{o_1, o_2, ..., o_t\}$ and, sequence of hidden states, how to adjust the model parameters i.e., state transition probability distribution $A = a_{i,j}$ and observation symbol probability distribution $B = b_{i,k}$ to maximize the probability P(O| λ).

Chapter 4

Proposed Framework

The proposed framework shown in Figure 4.1 contains three phasess which are depicted in a block diagram. This framework runs on call detail records or mobile traffic data which is collected by cellular providers. Then, the methodology proposed by Marco, et al.,[16] is applied with some required modification to classify the large scale mobile traffic datasets into some optimal number of clusters. The cellular users are classified into different clusters based on their cellular traffic generation patterns, i.e., use of cellular data or call-in and call-out activities which are called classification of call profiles. The resultant clusters are the input to the prediction model, are considered as hidden states and the possible visited locations are the observable states of the proposed HMM.

4.1 Mobile Traffic Dataset

The large scale mobile dataset provided by cellular provider presents the mobile traffic volume in terms of outgoing and incoming SMS, voice calls and internet traffic data. In Table 4.1, it is shown all the different kind of attributes and their types which is going to be used in the further analysis of the given dataset.

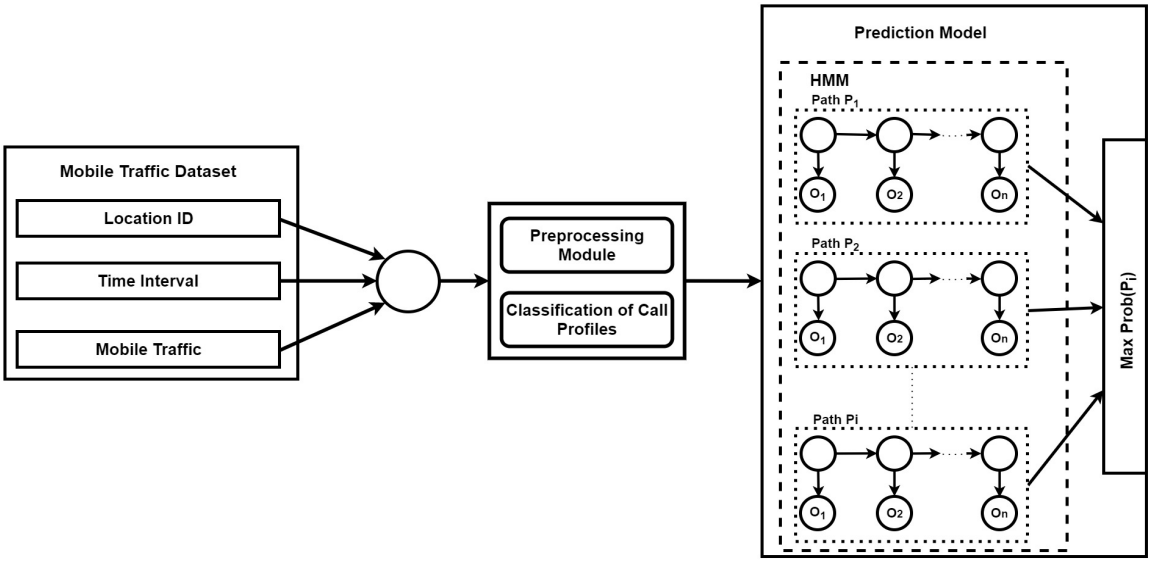

Figure 4.1: Proposed Framework

Table 4.1: Attribute Details

Attribute Name	Attribute Type
Square ID	Numeric
Time Interval	Numeric
Country Code	Numeric
SMS-in Activity	Numeric
SMS-out Activity	Numeric
Call-in Activity	Numeric
Call-out Activity	Numeric
Internet Traffic Activity	Numeric

4.2 Classifying Call Profiles

Figure 4.2 presents the flowchart of the classification of the call profiles where mobile dataset is the input. In the given dataset [17], the attribute called square ID represents the geographical location. One particular city is fitted in a grid and each cell is referred to a particular location associated with unique square ID. The square ID is numbered from the bottom left corner to its right top corner of the grid [17]. There are more than 10,000 square IDs in the dataset. It has been observed that some locations generate very low amount of data which is negligible. To make the proposed

framework more robust, the low traffic generated locations are merged with adjacent location. If the traffic generated by the adjacent location is below the threshold value then repeat the cluster merging process, is called the location filtering. After location filtering the resultant locations are identified by the unique square ID.
After location filtering, the dataset is transformed into a graph $G = (P, E)$ where each location (location is equal to square ID) is represented by vertices $S = \{s_1, s_2, ..., s_n\}$, s_i represents the ith square ID. Each pair of (s_i, s_j) is connected by an edge e_{ij}. Each edge e_{ij} associated with a weight w_{ij},. Here, e_{ij} represents the similarity between the vertices s_i and s_i. There are two different types of traffic volume similarity measure exist [16] such as Traffic volume similarity in equation and traffic distribution similarity in equation (4.1). The weight w_{ij} is calculated using the equation (4.1).

- *Traffic distribution similarity:* It shows the traffic distribution over different location at a particular time period. The weight is calculated as:

if: i== j ***then:*** w_{ij} =1

else:

$$w_{ij} = \frac{1}{\sqrt{\sum_{i,j \in Z} (\frac{v_i^t}{v_i} - \frac{v_j^t}{v_j})^2}}, v_i = \sum_{t \in T, z \in Z} v_i^t \tag{4.1}$$

Here, v_i^t is the volume of traffic at ith location in time period t and v_i represents the total traffic generated.

Clustering: The generated weighted graph in Figure 4.3 is the input to the hierarchal clustering algorithm which gives the optimal number of clusters by using Calinski and Harabasz index or Beale index [16]. Initially, each vertex is assigned to a different cluster Ci and, at each step we calculate the distance between two clusters C_i and C_j, i.e., d_{ij}. Then merge the pair of cluster which possesses the minimum distance. Here, the distance is calculated as :

$$d_{ij} = \frac{1}{\mid C_i \mid . \mid C_j \mid} . \sum_{i \in C_i, j \in C_i} \frac{1}{W_{ij}} \tag{4.2}$$

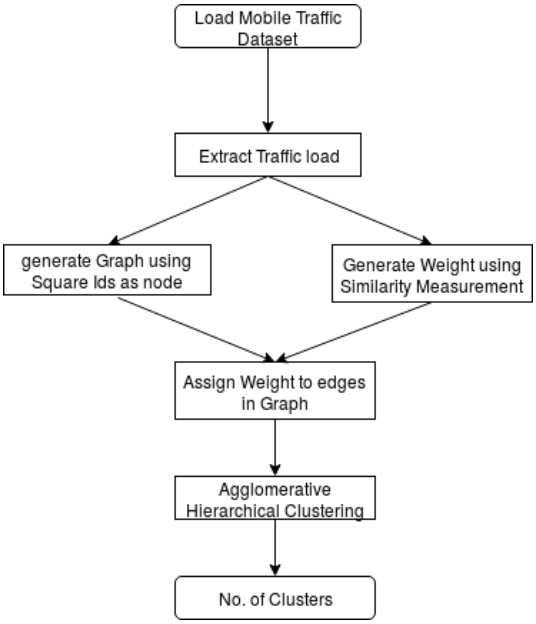

Figure 4.2: Flowchart of Classifying Call Profiles.

Until we get optimal number of clusters, the above merging process is repeated which is presented in Figure 4.4

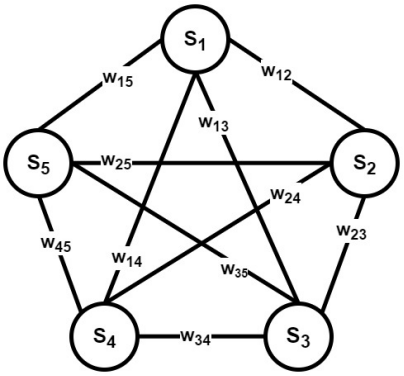

Figure 4.3: Generated Weighted Graph

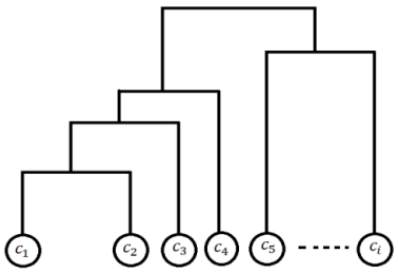

Figure 4.4: Dendrogram of Clusters

4.3 Prediction Model

The last phase of framework is the prediction model, i.e., to predict the human mobility based on the mobile traffic dataset. After classification of call profiles that

is explained in Section 4.2, the prediction model will be applied on resulted optimal number of clusters. Here, the locations are clustered based on the highest traffic generation among all other locations at a certain time period (t). It is noted that a particular location can be considered in more than one cluster based on the traffic generation at different time period. In this proposed prediction model, the whole day is divided into different time intervals because of user's current and future visits are highly related to their daily routines, based on time intervals. Here, HMM is used to capture the user's mobility from their daily routines and concerned with time periods. Unlike the papers [7, 8], only the location has been considered as the parameter for the prediction model. The behavior in a particular time period explains 50 to 70 percent of human movements [12] such as the same set of user's behavior in morning time are different in evening time. Hence, the various time periods should be considered in the prediction model to predict the most possible visited state which is more realistic results. This spatiotemporal dataset can be modeled using HMM, efficiently. HMM is most suited approach for the analysis of time series data, in which the sequences of observation states are assumed to be generated by a Markov process with unobserved or hidden states [18]. In the proposed HMM, clusters which are gained from classification are hidden states and the locations that can be visited are considered as visible states. Each hidden state has probability distribution for transition from one hidden state to another hidden state as well as over the possible location to be visited.Figure 4.5 shows the flow of location prediction, where first block shows the preprocessed dataset and following blocks depict the input sequence to the Hidden Markov Model and generated possible paths.After generating possible paths, which observation sequence possess the maximum probability is selected and gives the next location information.

Figure 4.6 depicts the architecture of the proposed Hidden Markov Model. The random variable C(t) represents the cluster i.e., the set of locations generate the highest traffic at a particular time t, $C(t) = \{C_1, C_2, ..., C_t\}$and $t = \{1, 2, ..., n\}$. Similarly, the random variable Z(t) represents the observed state or possible visited locations at time t, $Z(t) = \{Z_1, Z_2, ..., Z_t\}$ and $t = \{1, 2, ..., n\}$. Each hidden state has probability distribution of transition from one hidden state to another hidden state as well as probability transition from hidden state to the possible visited location.

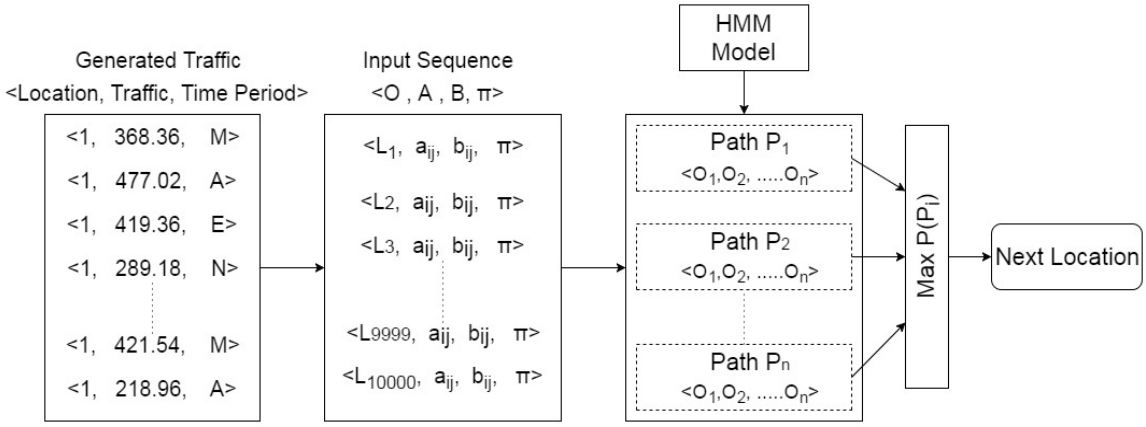

Figure 4.5: Flow of Location Prediction

The proposed prediction model satisfies the Markov property i.e., the process will be in state C_t at time period t which only based on the probability that the process in the state C_{t-1} at time period t-1. Because, the mobile user's mobility are based on their daily routines with corresponding time periods. Similarly, Z(t) depends only on C(t) at time period t.

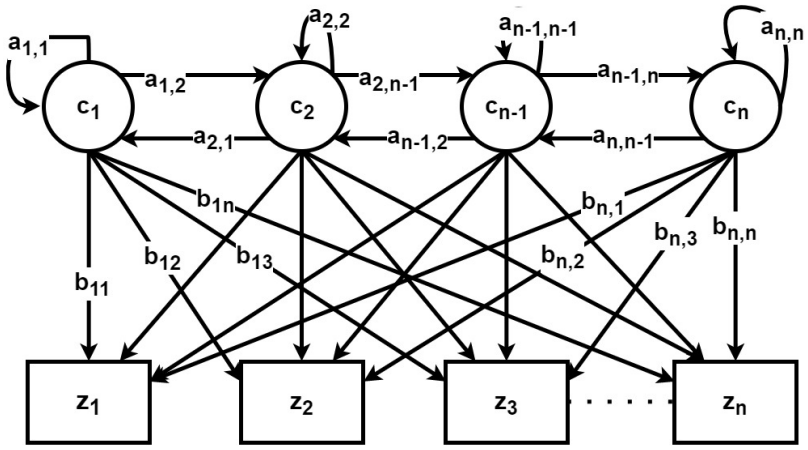

Figure 4.6: Discrete HMM with 4 States and n Observation States.

$$P(Location_{future}) = \alpha\sigma_t^{t+1} P(Z_{t+1} \mid C_{t+1}) P(C_{t+1} \mid c_t).P(C_t \mid Z_t) \qquad (4.3)$$

Here $Location_future$ represents the most probable location at time instant t+1 and Equation (4.3) represents the probability of a location being visited next or at time t+1, where $P(C_t \mid Z_t)$ is a decoding problem can be stated as given a present location

Z_t (visible symbol) how do we choose a corresponding hidden state C_t. the term $P(Z_{t+1} \mid C_{t+1}).P(C_{t+1} \mid C_t)$is an Evaluation problem , where present hidden state is given and we have to compute efficiently the next most probable location (visible symbol).
$P(C_{t+1} \mid C_t)$ is probability of transition between hidden state C_t to C_t+1 and $P(Z_{t+1} \mid C_{t+1})$ is the probability that process in hidden state C_t while the process emit the visible symbol Z_t at time instant t. The gives the probability that the process being emitted Z_{t+1} while in the hidden state C_{t+1} at time instance t+1. α (Distance Factor) is a distance measure between present location and next probable future locations which affects the probability of location which is likely to be visited. As nearer locations are more likely to be visited next,distance factor increase the probability of nearer next location to be visited by multiplying the normalized distance. The whole day has been divided into four periods such as morning, noon, evening and night. In the proposed HMM, there are four hidden states based on the traffic generated for different locations at a particular time period, i.e., C_1, C_2, C_3, C_4 respectively corresponding the time periods.

Chapter 5

Experimental Results

5.1 Dataset and Experiment Setup

The mobile traffic dataset is given by cellular provider, provided dataset contains information about mobile traffic related to 10000 locations. Experiment has been carried out on large dataset (more than thousand locations) which consist mobile data exchange during one week. Here experimental results are shown for 10 locations to avoid the complexity although proposed framework can be applied on vast number of locations as well. Hence, The proposed framework introduced in chapter 4 has been applied to the dataset [17]. Here,we represent the outcomes of the evaluation of proposed framework. First, weighted graph is generated from the mobile traffic generated in each define time period Morning, Noon, Evening, Night. In generated weighted graph, vertices are labeled as location ID and edges are associated with weights which is similarity measure between two adjacent vertices(location ID). After generating graphs, iterative aggregation of graph vertices into a dendogram structure has been done to classify the vertices which represent the location IDs.

Figure 5.1 to Figure 5.4 represents the generated graph corresponding to each time period(morning, noon, evening, night), where nodes are the square Ids and edges are associated with Weight. Weight is assigned according to the equation 4.1, shows

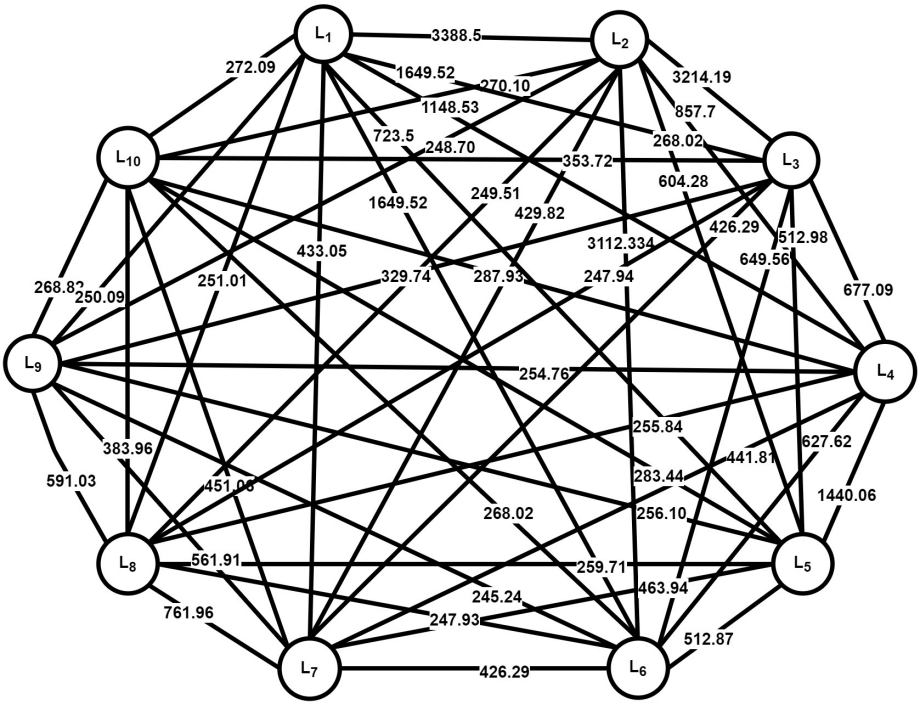

Figure 5.1: Graph of Morning traffic

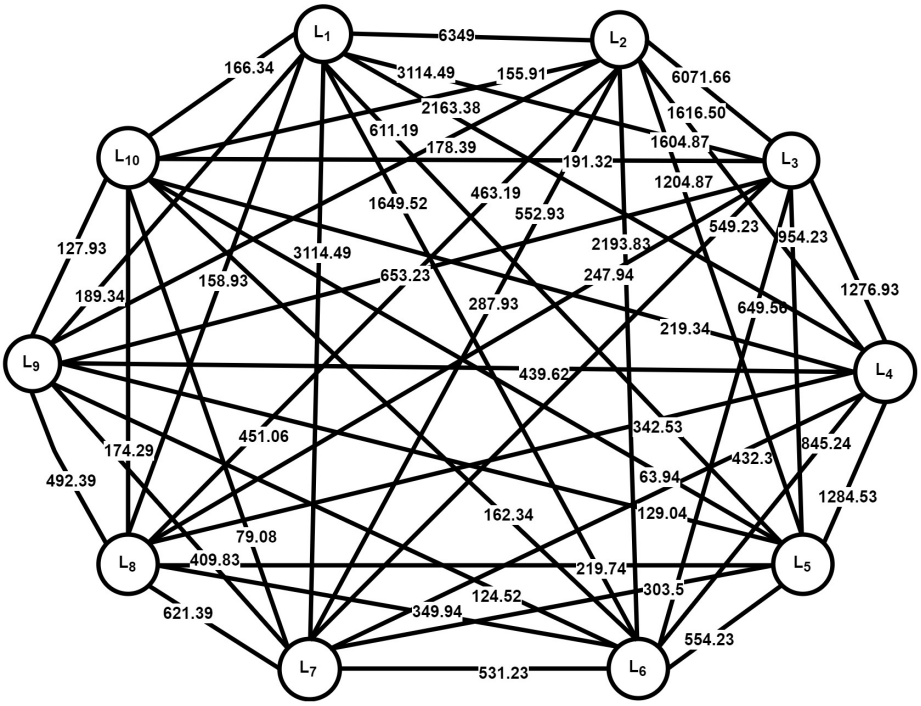

Figure 5.2: Graph of Noon Traffic

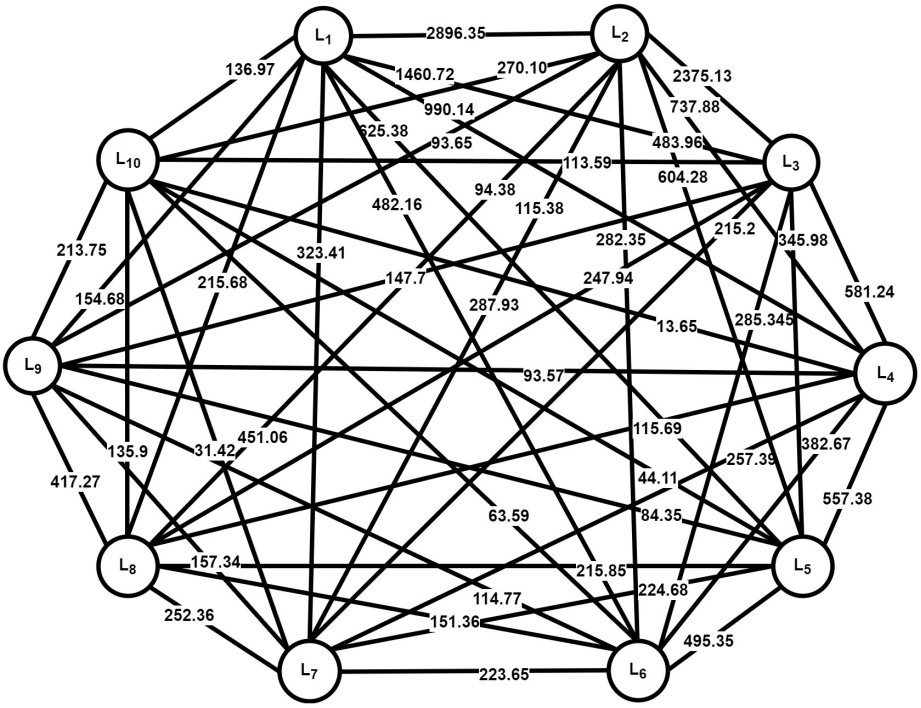

Figure 5.3: Graph of Evening Traffic

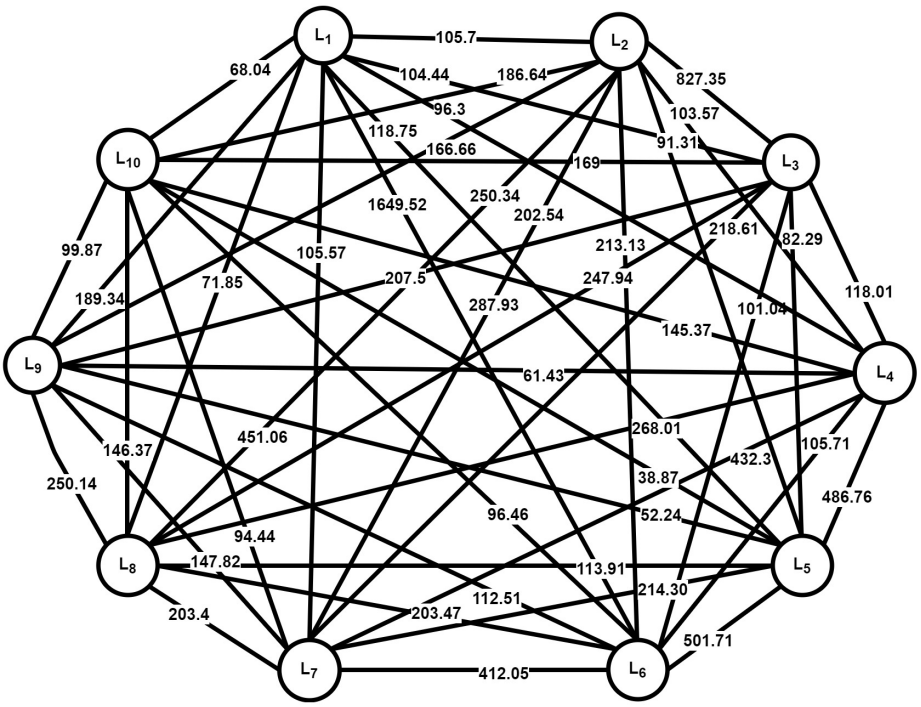

Figure 5.4: Graph of Night Traffic

the similarity in traffic generated by different location IDs(locations) in particular time period such as morning time period. Figure 5.5 to Figure 5.8 shows the dendogram resulted by Hierarchical Clustering Algorithm.

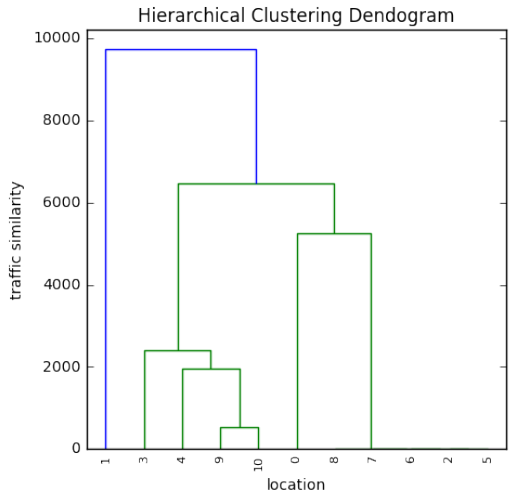

Figure 5.5: Dendogram of Morning Traffic

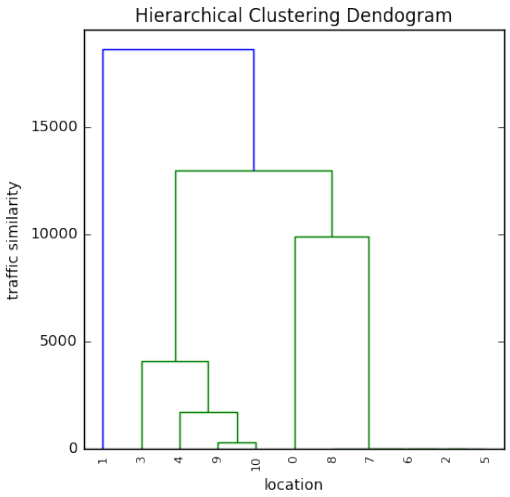

Figure 5.6: Dendogram of Noon Traffic

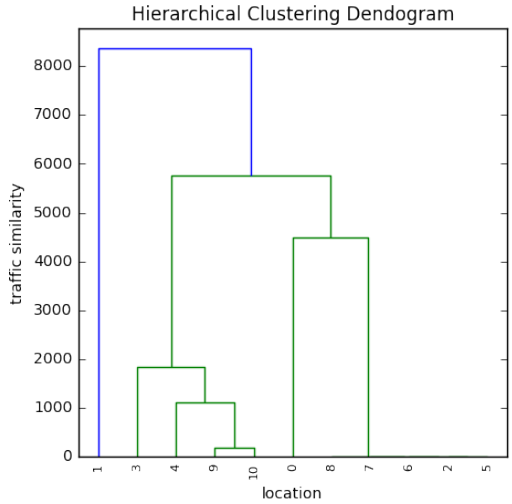

Figure 5.7: Dendogram of Evening Traffic

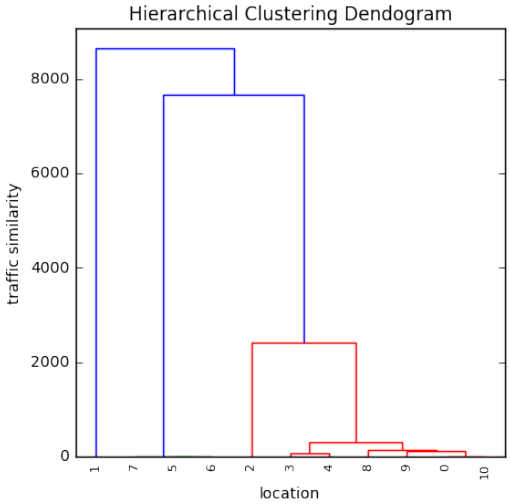

Figure 5.8: Dendogram of Night Traffic

5.2 Prediction results:

Hidden Markov Model is introduced in section 4.3 is serving as prediction model.decoding problems, in a given model and a sequence of visible symbol than what is the most likely state sequence in the model that produced the observations [15]. Decoding problem is solved by viterbi algorithm which a dynamic programming solution to give the maximum likelihood of hidden states. After clustering, which cluster possesses the highest traffic amongst all in particular time is selected for further processing. Here, Hidden Markov Model contains four hidden states corresponding to four time periods of a day and visible symbols are the locations. If observation sequence is given than we can calculate the probablity of given observation is being generated by forward algorithm [15] as well as most probable hidden states sequence can be inferred by viterbi algorithm.

	Present Location	Predicted Location	Probabilities
0	1.0	1.0	0.009485
1	1.0	2.0	0.007104
2	1.0	3.0	0.011885
3	1.0	4.0	0.003058
4	1.0	5.0	0.007965
5	1.0	6.0	0.007898
6	1.0	7.0	0.007898
7	1.0	8.0	0.007898
8	1.0	9.0	0.011493
9	1.0	10.0	0.002124

Figure 5.9: Probabilities of Locations to be Visited Next

In Figure 5.9, second column shows the fixed present state and third column shows all the possible next location to be visited. Probabilities can be inferred from the last column corresponding to the particular sequence. Figure 5.9 shows the probabilities of location to be visited next,where present location is fixed and next location's probability is calculated to state the highest probable future location. α

(distance factor) is defined in section 4.3 is calculated between present location and probable future locations.

	Present Location	Predicted Location	Probabilities
0	1.0	1.0	0.009485
1	1.0	2.0	0.007104
2	1.0	3.0	0.005943
3	1.0	5.0	0.001991
4	1.0	6.0	0.001580
5	1.0	9.0	0.001437
6	1.0	7.0	0.001316
7	1.0	8.0	0.001128
8	1.0	4.0	0.001019
9	1.0	10.0	0.000236

Figure 5.10: Probabilities of Locations to be Visited Next after Introducing α

Figure 5.10 shows that how distance factor carries a noticeable impact on the human mobility. After introducing α , probability of nearer locations significantly increases to be visited next.

5.3 Applications

Traffic load variation in different time period of a day is shown in fig 5.11. traffic load variation gives the information about resource consumption in particular location during particular time period such as morning. So, resource management strategies can be benefited by identifying the underloaded and overloaded locations. Underloaded locations, if a location possess the less consumer(Mobile Traffic) than allocated resources and vice versa for overloaded locations.

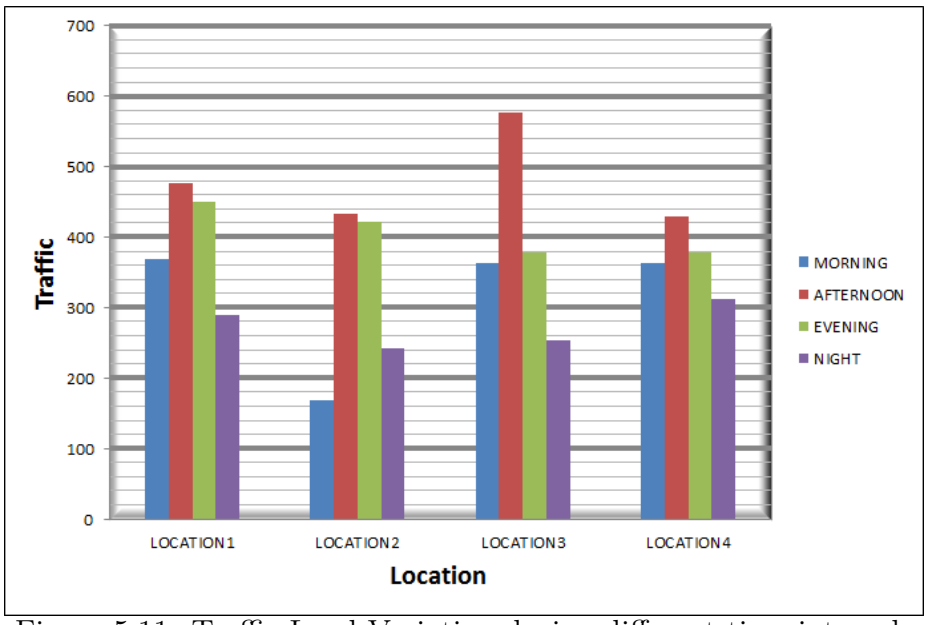

Figure 5.11: Traffic Load Variation during different time intervals

Chapter 6

Conclusion and Future Works

We have proposed a framework with Hidden Markov Model to predict the human mobility using the mobile traffic data. The literature survey on mobile traffic analyses has been presented and, captured the relationship between human mobility with mobile traffic data. The location flittering has been embedded in the existing classification of the call profiles, which gives the number of clusters from the large scale mobile traffic data. After using of hierarchal algorithm, the optimal numbers of clusters are input for the human mobility prediction model. In proposed framework HMMs is going to serve as prediction model.

Bibliography

[1] Anastasios Noulas, Salvatore Scellato, Neal Lathia, and Cecilia Mascolo. Mining user mobility features for next place prediction in location-based services. In *Data mining (ICDM), 2012 IEEE 12th international conference on*, pages 1038–1043. IEEE, 2012.

[2] Ghazaleh Khodabandelou, Vincent Gauthier, Mounim El-Yacoubi, and Marco Fiore. Population estimation from mobile network traffic metadata. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2016 IEEE 17th International Symposium on A*, pages 1–9. IEEE, 2016.

[3] Yohan Chon, Elmurod Talipov, Hyojeong Shin, and Hojung Cha. Mobility prediction-based smartphone energy optimization for everyday location monitoring. In *Proceedings of the 9th ACM conference on embedded networked sensor systems*, pages 82–95. ACM, 2011.

[4] Andrew Kirmse, Tushar Udeshi, Pablo Bellver, and Jim Shuma. Extracting patterns from location history. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 397–400. ACM, 2011.

[5] Ingrid Burbey. *Predicting future locations and arrival times of individuals*. PhD thesis, Virginia Tech, 2011.

[6] Angelo Furno, Razvan Stanica, and Marco Fiore. A comparative evaluation of urban fabric detection techniques based on mobile traffic data. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pages 689–696. ACM, 2015.

[7] Joao Bártolo Gomes, Clifton Phua, and Shonali Krishnaswamy. Where will you go? mobile data mining for next place prediction. In *International Conference on Data Warehousing and Knowledge Discovery*, pages 146–158. Springer, 2013.

[8] Anna Monreale, Fabio Pinelli, Roberto Trasarti, and Fosca Giannotti. Wherenext: a location predictor on trajectory pattern mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 637–646. ACM, 2009.

[9] Juha K Laurila, Daniel Gatica-Perez, Imad Aad, Olivier Bornet, Trinh-Minh-Tri Do, Olivier Dousse, Julien Eberle, Markus Miettinen, et al. The mobile data challenge: Big data for mobile computing research. In *Pervasive Computing*, number EPFL-CONF-192489, 2012.

[10] Wesley Mathew, Ruben Raposo, and Bruno Martins. Predicting future locations with hidden markov models. In *Proceedings of the 2012 ACM conference on ubiquitous computing*, pages 911–918. ACM, 2012.

[11] Disheng Qiu, Paolo Papotti, and Lorenzo Blanco. Future locations prediction with uncertain data. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 417–432. Springer, 2013.

[12] Ruiyun Yu, Xingyou Xia, Shiyang Liao, and Xingwei Wang. A location prediction algorithm with daily routines in location-based participatory sensing systems. *International Journal of Distributed Sensor Networks*, 2015.

[13] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. Next place prediction using mobility markov chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*, page 3. ACM, 2012.

[14] Rui Xu and Donald Wunsch. Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3):645–678, 2005.

[15] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[16] Diala Naboulsi, Razvan Stanica, and Marco Fiore. Classifying call profiles in large-scale mobile traffic datasets. In *INFOCOM, 2014 Proceedings IEEE*, pages 1806–1814. IEEE, 2014.

[17] http://www.dandelion.com.

[18] Omnia Ossama and Hoda MO Mokhtar. Similarity search in moving object trajectories. In *Proceedings of the 15th International Conference on Management of Data*, pages 1–6. Citeseer, 2009.