Probabilistic Space-Time Analysis of Human Mobility Patterns

HA YOON SONG
Department of Computer Engineering
Hongik University, Seoul, KOREA
hayoon@hongik.ac.kr

Abstract: Human mobility models are widely required in many academic and industrial fields. Due to the spread of portable devices with positioning functionality, such as smartphones, ordinary people can obtain their current position or record mobility history. Thus, mobility history can be processed in order to identify human mobility patterns. The human mobility pattern can be analysed in two ways: space and time. Space analysis focuses on a users location, and time analysis emphasises a users mobility on a daily basis. From the raw positioning data of various sets of user mobility, we analysed a personal human mobility model. Each users positioning data set is pre-processed and clustered by space and time. For spatial clustering, we developed a mechanism of clustering with expectation maximisation methodology. For temporal clustering, stay or transition probabilities over a 24 hour period were analysed. We represented the result of the personal human mobility model using the continuous-time Markov chain (CTMC) with spatial transition probabilities over time. We developed a process to construct a personal human mobility model from a persons positioning data set. This personal mobility model can act as a basis for many other academic or industrial applications.

Key–Words: Human Mobility Model, Space-Time Analysis, Location Clustering, Continuous Time Markov Chain, Individual Personal Mobility

1 Introduction

Establishing models of human mobility offer solutions for many open problems in academia and in industry. Precise human mobility patterns can encourage developments in other academic fields, be utilised according to industrial needs and be applied to public affairs. For example, spread patterns of a disease epidemic are affected by human mobility patterns; therefore, location-based services require a precise human mobility model. Additionally, practical mobility models for mobile communication simulation affect the precision of simulation results or the robustness of performance evaluation. In the field of consumer marketing, a detailed human mobility model will generate an improved result. From these human mobility model requirements, we explore the topic of the personal mobility model in this paper. Modern mobile devices are usually equipped with positioning functionalities using GPS [1], GLONASS [2], Galileo [3], etc. Positioning techniques using cellular base stations or crowd source WIFI positioning can be used in combination with other positioning systems; this combination is called a hybrid positioning system. With a set of positioning data collected for each person, human mobility models are derived and represented.

In this paper, we collected three sets of human mobility data. The data sets are named KHU, SHY and LSJ. Positioning data sets in the form of $<$latitude, longitude, time$>$ were collected; the sets comprise human mobility data. Each data set was analysed and the resulting three human mobility models will be presented in this paper. The procedure to determine human mobility models will be discussed in detail. Section 2 shows related researchers and Section 3 discusses major considerations for human mobility modelling. One of the core methods for human mobility modelling will be discussed in Section 4, and the results will be presented in Section 5. In Section 6, we will conclude this paper with considerations for future research.

2 Related Work

Human mobility has various patterns. A rough categorisation can divide human mobility into micro mobility, macro mobility, individual mobility, group mobility, etc. From these various mobility patterns, many academic areas have tried to derive a human mobility model. Psychology, cognitive science, physics and computer science are areas that are eager to study human mobility models, and they require concrete human mobility models.

Psychologists study the psychological factors of human mobility, i.e. the reason for human mobility and selection of transportation methods as they relate to a persons psychology. The latter is considered microscopic and the former is considered macroscopic.

Complex system physics and statistical physics study a governing principle of human mobility and regard human mobility as a real world phenomenon.

Computer scientists require human mobility models for wireless mobile networks, artificial intelligence, robots, unmanned vehicles, etc. Usually, artificial synthetic human mobility models are used in the area of computer science; however, synthetic models have been criticised due to their lack of realism. Computer scientists require realistic and practical human mobility models.

In detail, from the aspect of psychology, the human perception of specific space has been researched, as have the results of perception on human mobility. In other words, knowledge of human mobility is needed to find the destination and route. For example, the actual curve of road and the perceived curve of the road are different, and this difference is a research topic in psychology [4]. Another research topic has been the route of human mobility based on a map. In this research, a straight route is preferred by a human to a curved route [5].

Another topic is transportation methods, such as a car or regional transportation. Factors and frequency between specific regions are another topic. In research before 1973, biographical factors were regarded as important. These factors included place, government, existence of family and members of family, the number of university students in a family and academic careers in the family. However, later research includes additional psychological factors, such as sensation-seeking, which is one factor of the five personality factors in the theories of Hans Jurgen Eysenck. The sensation-seeking scale is designed to measure individual differences in thrill- and risk-taking habits or uninhibited behavioral attitudes towards a novel and unfamiliar situation [6]. Another study examined the effects of human habits on information collected for travel modes, and it found a correlation between the major travel mode selected, the distance to the destination and individual habits [7].

Related research is also found in physics. Random walking is one of the major topics in physics. A random walk is a stochastic process and represents a mobile route of a particle or a wave. In particular, a random walk is used to model mobile fluid particles in statistical physics for the diffusion of heat, sound and light. Levy flight is a specific class of random walk, which represents the moving distance in heavy-tailed law. The heavy-tailed law implies more frequency over short distances and is used for human mobility modelling and earthquake research. Human mobility research in physics states that human mobility takes various forms according to spatio-temporal space. Some related work includes the research of dissemination in human disease epidemics. For example, some contagious diseases such as flu or colds can be blocked once the mobility pattern of humans is identified. For this purpose, dynamic and statistical properties of human mobility must be identified.

Other study on mobility patterns used specially marked bank notes to determine a mobility pattern [8]. A juvenile mobility pattern using the GPS faculty of mobile phones is another example of space-time mobility pattern research, which determines the location of a juvenile person at a specific time [9]. Another research showed that human mobility distribution follows power law from a distance from a central location using the mobility pattern of cellular phones. In this way, human mobility can be predicted up to 93% [10]. A virus on mobile phones can be used for the identification of human communication patterns by tracking the dissemination of the virus [11].

Computer science is a field of prolific research on human mobility models. In order to figure out human mobility patterns, research is done for disease epidemics and virus dissemination over a computer network [12][13][14].

In the Mobile Ad Hoc Network (MANET), human mobility is an important issue, as the performance of MANET is affected by precise human mobility. Statistical characteristics of human mobility patterns greatly affect MANET protocol. In particular, contact time (CT), inter-contact time (ICT), link duration, and lifetime of routing path are dependent on human mobility models [15][16]. For the performance improvement of mobile wireless sensor network protocols, mobility and velocity of mobile nodes are precisely determined by using an enhanced Monte Carlo localisation method [17].

A heterogeneous model, which combines distance metrics and frequent visiting locations, is presented [18]. The distance from the central location follows power law, i.e. higher probabilities in short distance than farther distance geometrically and a fractal pattern for frequented places is generated. These two characteristics are mutually related in human mobility patterns.

In the group mobility model, the intended mobility pattern of a node clearly follows the group leader. In addition, there is an attraction force between nodes within a group and a repulsion force between nodes in different groups. The interaction between the two forces determines the group mobility pattern [19]. Groups can be divided into tight groups and loose groups. Loose groups have a relatively high repulsion force in member nodes, while both groups have the same intention of mobility.

The wireless ad-hoc network presents its own group mobility model. Each group has its reference centre and group velocity, and parameters for partition prediction are assumed [20].

For the mobility models of wireless mobile, an ad hoc, random walk is the simplest model. The random waypoint model, which is a bit more advanced, is used for simulation [21]. However, the random waypoint model is unrealistic since it cannot depict abrupt changes of speed and direction. Another model to compensate for the random waypoint model includes past direction, past speed and past waypoint, and it utilises past values in order to present a more realistic human mobility model [22].

An obstacle-based model, which assumes obstacles in node routes and node detour obstacles, is also presented [23].

Social network theory has been applied to a more realistic human mobility model. In this model, each node can move within a community with a predefined probability and can move within another community with another probability. The probabilities in this model are determined by the social relationship of the node [24]. Sociological orbit presents another human mobility model, which reflects current time and the situation of a node in order to present a human mobility model. The name sociological orbit comes from physics; it is a term used to describe planet mobility or electron mobility around the atom [25]. A human mobility analysis using massive GPS data by progressive clustering is also presented; the purpose of the research is the visualisation of mobility trajectory [26]. A related study, which also utilises trajectory data and data mining, shows the most crowded highway over time and the most frequent mobility of humans over time [27]. Similar research was done in weather forecasting, which utilises spatio-temporal analysis and visualisation [28]. Urban engineering shows a definite interest in human mobility models.

One study presents human mobility using a utility function of multiple factors, such as transportation infrastructure and schedules [29]. Another study utilises covariance structure modelling to show the relationship between human activity and transportation methods and transportation patterns. It showed that the main reason for this relationship comes from personality and family characteristics [30].

3 Considerations

We must consider several characteristics of human mobility before developing a human mobility model construction process. The first characteristic is the space-time nature of human mobility. A human stays in a specific location (space) and transits between locations sometimes (time). The second characteristic is the analysis of mobility data. The core part of positioning data set must be pre-processed. The third characteristic is the consideration of probability-based clustering for human mobility analysis. Finally, we must consider the probabilistic nature of human mobility, which requires a specific probability distribution.

3.1 Space-time nature of human mobility

The first stage of human mobility pattern analysis is doing a spatial analysis to identify human interest locations based on a positioning data set. Spatial pattern analysis by positioning a data set analyses distributions of a spatial pattern and deduces location information. By investigating the distribution or density of positioning data, locations with high density can be identified as frequent places or a residential area. Such locations are expected to have a higher density of positioning data than any other locations. Figuring out a location of the high positioning data density requires a clustering method. However, the result was meaningless with density-only analysis. Spatial pattern analysis must consider the movement of a human, i.e. the velocity of positioning data. Velocity includes a concept of time. Thus, a temporal analysis or a combination of spatial-temporal analysis is introduced. From location and time information of consecutive positioning data, a speed value at a given time can be calculated [31].

With the speed value, positioning data in mobile state and positioning data in stay state can be classified. Different weights can be assigned to positioning data set by its state, then spatial pattern analysis can be re-applied.

With the concept of clustering, several parameters, such as size of clusters and stay time in clusters, can be calculated, and these parameters can affect the weight of the positioning data. For example, positioning points with less than 10 km/hour speed are classified as stable state points, while positioning points with more than 10 km/hour speed are classified as mobile states. However, the time duration is another criterion of state classification. Even in the mobile state, humans can be forced to stop for a while (in case of a stop signal or red light, for example), then continue to move again.

For such situations, once a resumption of movement is detected within a small timeframe, e.g. 10 min, positioning data with a temporary stay are regarded as in a subset of the mobile state.

Therefore, we need space-temporal analysis for human mobility modelling with a bridge of speed values. In order to apply clustering techniques to pattern analysis, an adequate clustering mechanism must be identified in advance. Among the various clustering techniques, probability-based clustering in the form of expectation maximisation clustering has been selected. Therefore, an adequate probability model for creating a human mobility model needs to be investigated.

3.2 Characteristics of data

Several volunteers have collected their positioning data independently. Each positioning data set has its own characteristics, such as duration of collection, positioning device used, patterns of collection, etc. Among the several positioning data sets, we will use three sets that have been identified as meaningful. Two of the data sets were collected consistently over several months. One of the data sets was collected intermittently over more than one year.

The individuals had their own positioning devices and some changed their devices during the collection period. The devices are: dedicated positioning data collection app on iPhone 3GS, dedicated positioning data collection app on iPhone 4S, Garmin GPSMAP 62s [32], Garmin EDGE 800 [33], Garmin EDGE 810 [34] and commercial positioning data collection apps such as sportstracker [35] on iPhone or Android Phones.

Garmin and Android phones only use a GPS-based positioning system, while iPhones have a hybrid positioning system [36] with a combination of three different positioning systems. The iPhone 4S introduced GLONASS. In addition, the interval of positioning data recording irregularly varies from 1 sec to 1 min. GPS only devices cannot obtain positioning data inside a building or underground area. A Garmin device user can set a fixed interval for GPS data collection from 110 sec. However, recent Garmin devices introduced smart collection methodology.

A dedicated positioning data collection app was developed and was used for a while on iPhones. The interval between positioning data recording for the app is totally different. In the stay state, the iPhone app collects positioning data in every user defined interval (e.g. 3 sec) and tries to collect positioning data for every possible movement if it senses movement of the device.

Every volunteer collector has his own pattern of collecting positioning data. The KHU and LJS sets show the continuous collection of positioning data over several months, while the SHY set shows intermittent collection patterns because it starts collecting positioning data just before a mobile stage. It tends to collect positioning data from unusual outdoor activity, which has more points in mobile states. A credential analysis of human mobility patterns must cover the various characteristics of the positioning data set.

The human mobility model construction process was developed to cover the various characteristics of positioning data sets, such as number of positioning data points, collection interval, size of clusters, time spent in clusters, etc. from a multidimensional aspects.

3.3 Probability-based Clustering

In order to build a human mobility model from human mobile trajectories, space-time clustering is required in our research. Clusters identified using clustering techniques represent the location which a person visits frequently or stays at for a longer time. Therefore, it is necessary to select an appropriate clustering method [37]. Clustering algorithms can be divided into four categories: connectivity-based clustering, centroid-based clustering, density-based clustering and distribution-based clustering. Connectivity-based clustering clusters objects with distance connectivity. Centroid-based clustering identifies the centroid of clusters and clusters based on distance to centroids. Among these, k-means clustering is the most well known. Density based clustering identifies data sets based on density of data. DBSCAN is an example. Distribution-based clustering uses probability distribution to find clusters; this requires a distribution model. Among the distribution-based clustering algorithms, expectation maximisation clustering algorithms are used in our research.

Another classification includes partitioning and hierarchical algorithms. Hierarchical methods are divided into bottom-up and top-down methods. The latter refines large clusters and the former aggregates small clusters hierarchically.

For our space-temporal clustering, it was appropriate to use the partitioning method, since human locations are spread over partitions of global areas. Two of the nominative partitioning methods are k-means algorithms in centroid-based and Expectation Maximization (EM) clustering in distribution-based clustering. K-means requires centroid and distance to centroid, usually Euclidean distance, and it is easy to use, although it ignores many properties of clusters other than distance. EM clustering algorithms require a certain probability distribution and parameters for the distribution. It predicts the probability of data belonging to a certain cluster. Based on the prediction, it recalculates parameters. This maximises the expectation of the whole model towards optimal parameters. A probability for data to be included in a cluster is calculated, and, clusters are established according to these probabilities. Usually EM clustering utilises a normal distribution for its probability distribution.

It is determined that EM clustering is a good fit for our purpose because EM algorithms can accommodate various distributions in a users purpose; thus, the space-time nature of the human mobility model can be represented probabilistically.

3.4 Probability density function

For the EM clustering algorithm, a proper probability distribution is required. The usual candidate is normal distribution, i.e. Gaussian distribution, but the result of clustering with normal distribution is abnormal due to the nature of human mobility.

A proper probability distribution for human mobility must reflect power law, since human mobility has a high probability observed within a 12 km distance from the centre of an individuals location, and the mobility over longer distances has the probability distribution of power law [10].

Additionally, our observation is based on the fact that human mobility patterns are usually concentrated in the region of 12 km (residence area) for certain time periods (residence period of human mobility). The transition between resident areas shows power law distribution (transition period of human mobility). Thus, we will introduce power-law distribution, which is similar to exponential distribution. We call the distribution a transformed exponential distribution with a parameter which shows distance of human mobility from the centre point of residence areas, as

$$f(x) = e^{-\lambda x} \tag{1}$$

where λ is a controllable parameter denoting the maximum distance of the cluster, which is usually fixed in a constant value, and x is the distance between the current position of human and the centre of a cluster.

In our approach, we set λ as an inversion of the sum of *the maximum distance of the cluster position data of the stay state from cluster centroid* and *the average distance of positioning data in a cluster from the cluster centroid.* In other words, λ can be calculated as follows: For each cluster C_i, positioning data P_i, can be determined as members of C_i by EM iteration as well as centroid of C_i, $Cent_i$, can be calculated. Among P_i, there exist non-mobile positioning data $P_{i,stay}$, which are classified by speed values; therefore, λ_i can be defined as shown in equation 2 where dist(x,y) stands for pre-defined distance between two positions.

We considered the maximum distance of positioning data sets in the stay state from the cluster centroid and the average distance of positioning data sets from the centroid in order to reflect mobility in a specific cluster.

With stay positions inside a cluster, a maximum distance needed to be introduced. For all positions inside a cluster, an average of the distance to centroid also needed to be considered.

Once we have a stay position inside a cluster, there is a possibility that another positioning data set can become a member of a cluster within the average distance from the stay position. Thus, stay positions play a key role in determining λ_i.

In addition, for positioning data sets in a mobile state, we calibrate distance from the cluster centroid according to the probability density function and time ratio of a specific cluster (TimeRatio). For EM clustering, the probability for positioning a data set to be a member of a cluster has its own weight. Weights for stay positions can be calculated from the transformed exponential distribution, while weights for mobile positions can be determined as products of TimeRatio and raw weights from the transformed exponential distribution. For example, a mobile position in a cluster with a smaller TimeRatio will have smaller weight than another mobile position in another cluster in the same situation even though it has the same probability as another mobile position in another cluster because of the TimeRatio. This calibration mechanism leads to more precise clustering for human mobility modelling.

$$\lambda_i = \frac{1}{max_{P_{i,stay}}(dist(P_{i,stay}, Cent_i)) + average_{P_i}(dist(P_i, Cent_i))} \tag{2}$$

4 Model Construction Process

4.1 Clustering locations

In order to find locations in the form of clusters from the positioning data, four steps are used.

4.1.1 Initialisation

The initialisation step calculates initial parameters, such as the number of initial clusters and the initial size of clusters. The speeds derived from positioning data sets and density of positioning data sets are usually utilised for initialisation.

4.1.2 E-Step

From the initialised parameters of clusters, a probability for a positioning data set to be a member of a specific cluster is calculated by using the probability density function and calibration mechanism of Section 3.4. The maximum and average distances from a positioning data to the centroid of a cluster are major parameters for probability calculation.

In detail, with the location data set

$$X = \{x_1, x_2, \cdots, x_n\} \tag{3}$$

and the cluster set

$$\theta = \{\theta_1, \theta_2, \cdots, \theta_m\} \tag{4}$$

then the weight or probability for a positioning data set to be a member of cluster h is

$$w_h(x) = \frac{w_h \cdot f_h(x|\theta_h)}{\sum_i w_i \cdot f_i(x|\theta_i)} \tag{5}$$

If positioning data set x is found in a moving state, the weight of the moving state is

$$w_h(x) = \frac{w_h \cdot f_h(x|\theta_h)}{\sum_i w_i \cdot f_i(x|\theta_i)} \cdot TimeRatio_h \tag{6}$$

where TimeRatio of cluster c is

$$TimeRatio_c = \frac{t_c}{\sum_i t_i} \tag{7}$$

and t_c denotes the time interval of residence in cluster c, and where $f_h(\cdot)$ is probability density function for cluster h.

4.1.3 M-Step

In this step, we recalculate parameters, such as expectation and variance of each positioning data set to be a member of a specific cluster for maximum likelihood, using weights calculated in E-step.

With the cluster set

$$\theta = \{\theta_1, \theta_2, \cdots, \theta_m\} \tag{8}$$

and when the location data set of cluster A is

$$\theta_A = \{x_{1|A}, x_{2|A}, \cdots, x_{n|A}\} \tag{9}$$

then the mean of cluster A can be calculated as

$$\mu_A = \frac{\sum_i w_i \cdot x_{i|A}}{\sum_i w_i} \tag{10}$$

and the variance of cluster A can be calculated as

$$\sigma_A^2 = \frac{\sum_i w_i \cdot x_{i|A}^2}{\sum_i w_i} - \mu_A^2 \tag{11}$$

4.1.4 Termination condition

E- step and M-step will be repeated until the parameters of each cluster remain unchanged, i.e. the threshold value is used to check if parameters became stable, and those values will converge to maximum likelihood.

The log likelihood is

$$L(\theta) = \sum_{i \in D} log\left(\sum_j^k w_j \cdot f_i\Big(x|\mu_i. \sum h\Big)\right) \tag{12}$$

With this process, clusters representing locations of individual mobility will be calculated and each cluster will be given its own attributes, such as centroid, average radius, variance of radiance, number of positioning data in cluster, etc. The positioning data can be clustered into specific locations; we will call them location clusters.

4.2 Representation of continuous-time Markov chain

Our purpose is to represent human mobility in a formal model. Among the various models, a Markov chain is a good option, since each state of the Markov chain can be mapped onto locations or location clusters, and transition between states can be mapped onto transitions between location clusters clearly. Section 4.1 explains the clustering method used to find clusters of locations. In our approach, this is actually a method to find states of the Markov chain. Continuous-time Markov chain (CTMC) is used, as the staying time at a cluster is parameter for a human mobility model.

The location clusters derived will be regarded as the state of CTMC without any further process.

CTMC $\{X(t)|T > 0\}$ is a continuous time and a discrete value stochastic process with infinitesimal time Δ and satisfies the following equation:

$$P[X(t+\Delta) = j|X(t) = i] = q_{ij}\Delta \quad (13)$$

$$P[X(t+\Delta) = i|X(t) = i] = 1 - \sum_{j\neq i} q_{ij}\Delta \quad (14)$$

where q_{ij} is transition rate from location cluster i to location cluster j.

When Δ converges to 0, the current state of cluster i will transit exponential random time with parameter

$$v_i = \sum_{j\neq i} q_{ij} \quad (15)$$

where v_i is departure rate of location cluster i.

The conditional probability of D_{ij} under condition D_i that a human leaves cluster i in time interval $(t, t+\Delta]$ for cluster j is represented as

$$P[D_{ij}|D_i] = \frac{P[D_{ij}]}{P[D_i]} = \frac{q_{ij}\Delta}{v_i\Delta} = \frac{q_{ij}}{v_i} \quad (16)$$

In CTMC, a human stays at location cluster i in exponential (v_i) time and transits to another cluster. The probability that destination location in cluster j is

$$P_{ij} = \frac{q_{ij}}{v_i} \quad (17)$$

In order to construct CTMC, a transition rate $\{q_{ij}\}$ must be found. From the location clustering in Section 4.1, the number of transitions between location clusters can be found and then conditional transition probability P_{ij} can be calculated. For staying time at each cluster, v_i is calculated and thus $\{q_{ij}\}$ is found. In our approach, the time unit is in 1 min considering the nature of human action.

4.3 Hourly base timed analysis of human mobility

The location clustering shows spatial aspect of human mobility. In order to have temporal analysis of human mobility, a re-analysis of the clustered result in timing base was required. Even though hourly mobility was analysed in this research, any other time interval, e.g. 20 min, can be used as a time base.

After completing the location clustering stage, each positioning data set can have the following extra attributes.

- Location cluster to which the positioning data belong
- Weight for a positioning data set to be a member of the location cluster
- The status of positioning data: stay or moving

For 24 hours per day, each data set in intervals of 1 h will be analysed. The probability of stay for positioning data set P_{stay} and the probability of moving for positioning data set P_{move} will be represented as follows: where N_{stay} is the number of positioning data sets in a stay state, and N_{move} is the number of positioning data sets in a mobile state for a given hour.

The calculation of probability $P_{i,h}$ for a positioning data in a set to be a member of location cluster i at given time h is shown in equation 20.

$$P_{stay} = \frac{N_{stay}}{N_{stay} + N_{move}} \quad (18)$$

$$P_{move} = \frac{N_{move}}{N_{stay} + N_{move}} \quad (19)$$

Based on location clustering, each positioning data set can be re-analysed from a temporal aspect. Thus, the timed analysis can comprise the temporal mobility model across the location clusters.

In Section 5, three mobility sets were analysed and the space-temporal result of the human mobility model will be presented.

5 Results and Discussion

For the verification of our human mobility model construction process, three positioning data sets are selected. Three mobility sets KHU, LSJ and SHYwere voluntarily collected by researchers with various positioning devices. Note that the positioning data set represents a subset of each collectors mobility rather than total mobility. There are several reasons for this, including battery shortage, GPS signal lost inside a building, intentional cut-off by user privacy, user habits, etc. The positioning data set was collected over a period from a few months to several years.

$$P_{i,h} = \frac{number\ of\ positioning\ data\ in\ the\ set\ for\ location\ cluster\ i\ for\ the\ given\ hour}{number\ of\ positioning\ data\ in\ the\ set\ for\ the\ given\ hour} \quad (20)$$

Table 1: Statistical Analysis Result of KHU's Location Clusters

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Centre Position	37.550633816622, 126.924416829235	37.530302415258, 126.737284754725	37.612157810217, 126.726268525124	37.562006698510, 126.984836142539	37.507249050132, 126.744428282204	37.542386420004, 126.727308422318	37.499033670598, 127.026392347701
Std. Dev. of Position	0.001294512255, 0.001344398660	0.003939004009, 0.002612434297	0.000022967529, 0.000084171300	0.000606102734, 0.000528333500	0.000019116018, 0.000018329103	0.001861811056, 0.002804018171	0.000531108111, 0.000569534023
Max Distance	0.922 km	1.285 km	0.021 km	0.152 km	0.02 km	0.73 km	0.137 km
Mean Distance	0.119 km	0.414 km	0.007 km	0.075 km	0.002 km	0.256 km	0.057 km
Time Ratio	0.454	0.4203	0.0059	0.0046	0.003	0.0028	0.0009
# of GPS	36160	36839	496	369	246	274	268
Stay Time (h)	102.251	94.671	1.333	1.052	0.68	0.638	0.224
Location	School	Home	Restaurant (Gimpo)	Shopping area	(Bucheon)	friend	(Gangnam)

$$t \quad \begin{bmatrix} 0 & 0.00391 & 0 & 0.00016 & 0 & 0 & 0.00016 \\ 0.00387 & 0 & 0.00035 & 0 & 0.00018 & 0.00106 & 0 \\ 0 & 0.02500 & 0 & 0 & 0 & 0 & 0 \\ 0.01584 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.02449 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.15652 & 0 & 0 & 0 & 0 & 0 \\ 0.07417 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{21}$$

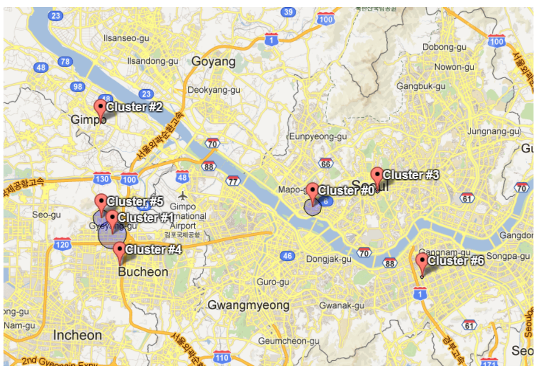

Figure 1: Total location clusters for KHU

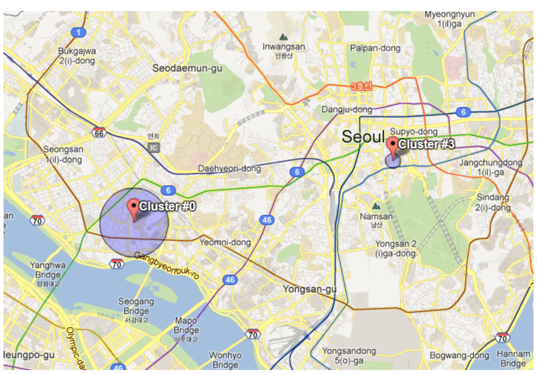

Figure 2: Most frequent cluster of KHU

Figure 1 shows the result of location clustering of the KHU set on Google Maps. Seven location clusters were presented, and TimeRatio is the parameter used to determine the cluster number. The smaller the cluster number, the bigger the TimeRatio.

Figure 2 shows the most frequent cluster, cluster #0, on Google Maps. It is the university to which

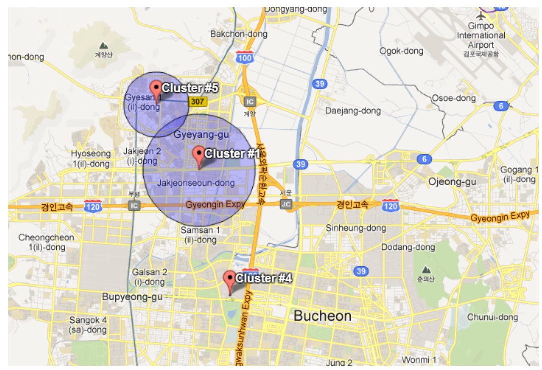

Figure 3: Another frequent clusters of KHU

KHU belongs. Figure 3 shows another frequent clusters of KHU, cluster #1, and neighbouring cluster #5. Cluster #1 is KHU's home. Table 1 represents the statistical results of KHU's location clusters.

Centre position, i.e. the centroid, is a weighted average of positions in a cluster. Std. Dev of Position stands for standard deviation of geographical locations in each cluster and in a form of <latitude, longitude>. Max distance is the maximum distance of positioning data in a cluster to the centroid of the cluster. Mean distance is average distance between member positioning data to the cluster centroid. TimeRatio is the ratio of staying time in a cluster. For example 45.5% of total time has been spent at KHUs home. Count is number of positioning data in each cluster. Stay time (h) is staying time at a location cluster in units of 1 h. Location is verified by a volunteer collector. For example, KHU stayed at location cluster #0, which is his university and the most

Table 2: Space-Time Mobility of KHU

	AM 0 - 1	AM 1 - 2	AM 2 - 3	AM 3 - 4	AM 4 - 5	AM 5 - 6	AM 6 - 7	AM 7 - 8	AM 8 - 9	AM 9 - 10	AM 10 - 11	AM 11 - 12
Stay Prob.	0.921	0.929	0.984	0.996	0.993	0.996	1	1	0.773	0.757	0.856	0.751
Moving Prob.	0.079	0.071	0.016	0.004	0.007	0.004	0	0	0.227	0.243	0.144	0.249
Cluster Prob.	#0: 0.177 #1: 0.768 #4: 0.055	#0: 0.054 #1: 0.931 #5: 0.015	#0: 0.115 #1: 0.882 #5: 0.003	#0: 0.065 #1: 0.935	#0: 0.126 #1: 0.874	#0: 0.285 #1: 0.715	#0: 0.155 #1: 0.845	#0: 0.08 #1: 0.92	#0: 0.015 #1: 0.985	#0: 0.488 #1: 0.355	#0: 0.645 #1: 0.355	#0: 0.720 #1: 0.280
	PM 0 - 1	PM 1 - 2	PM 2 - 3	PM 3 - 4	PM 4 - 5	PM 5 - 6	PM 6 - 7	PM 7 - 8	PM 8 - 9	PM 9 - 10	PM 10 - 11	PM 11 - 12
Stay Prob.	0.907	0.978	0.858	0.984	0.964	0.803	0.865	0.782	0.852	0.825	0.875	0.848
Moving Prob.	0.093	0.022	0.142	0.016	0.036	0.197	0.135	0.218	0.148	0.175	0.143	0.152
Cluster Prob.	#0: 0.960 #1: 0.040	#0: 1	#0: 0.917 #3: 0.034 #6: 0.040	#0: 0.882 #1: 0.103 #6: 0.045	#0: 0.809 #1: 0.112 #3: 0.034 #6: 0.045	#0: 0.750 #1: 0.177 #3: 0.044 #5: 0.018 #6: 0.010	#0: 0.745 #1: 0.220 #3: 0.036	#0: 0.661 #1: 0.328 #5: 0.011	#0: 0.493 #1: 0.418 #2: 0.090	#0: 0.466 #1: 0.476 #2: 0.036 #5: 0.021	#0: 0.428 #1: 0.565 #5: 0.007	#0: 0.293 #1: 0.707

$$\begin{bmatrix} 0 & 0.00764 & 0.00446 & 0.00191 & 0.02547 & 0 & 0.00064 & 0 \\ 0.01201 & 0 & 0.02510 & 0.00218 & 0.01637 & 0.00109 & 0.00109 & 0 \\ 0.01957 & 0.13701 & 0 & 0 & 0.10276 & 0.00489 & 0 & 0.00489 \\ 0.03421 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.32046 & 0.09039 & 0.20542 & 0 & 0 & 0 & 0 & 0 \\ 0.01124 & 0.01124 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.01710 & 0.01710 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.02740 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{22}$$

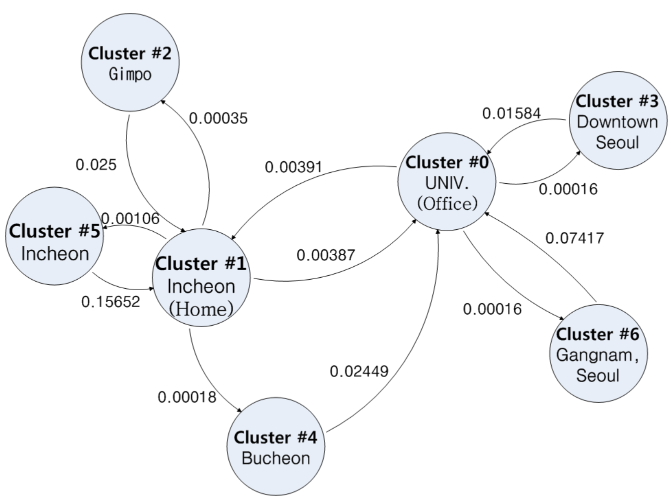

Figure 4: CTMC representation of KHU's personal mobility

The interpretation of KHUs mobility is that he generally visits his home and school, and other location clusters are exceptional. With the criteria of TimeRatio, the regular mobility pattern and exceptional mobility pattern can be classified. Equation 21 shows the transition matrix of KHUs mobility model in CTMC, excluding staying probability.

Figure 4 shows a human mobility model for KHU in a CTMC representation. From our approach, a timed mobility pattern and a general mobility model can be derived. For example, Table 2 shows the 24-hour mobility model of KHU. Another representation, such as cyclic mobility pattern, seasonal mobility patterns and daily mobility patterns can be found with this approach. Table 2 contains hourly information, such as probability of staying in a cluster, mobility, stability, etc. It can be regarded as space-time mobility model in hourly base. For example, between midnight and 1:00 AM, KHU stays at a certain location with a probability of 0.921 or in a mobile state with a probability of 0.079. In case of his stay state, KHU is highly probable at location cluster #1 at that time. In other words, table 2 shows space-time mobility pattern of KHU in hourly base.

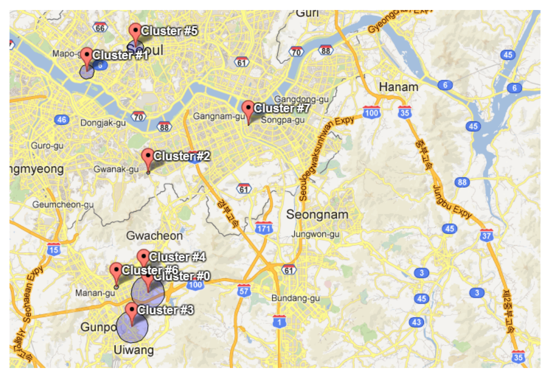

Figure 5: LSJ overall location clusters

Two more mobility models have been constructed. It can be predicted that LSJs mobility model

Table 3: Statistical Analysis Result of LSJ's Location Clusters

	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Centre Position	37.386634977323, 126.981109299404	37.550013906617, 126.924291615841	37.475459415957, 126.981512920091	37.361390070632, 126.966353851471
Std. Dev. of Position	0.005732003779, 0.004366247464	0.001666390074, 0.001289444185	0.000415959390, 0.000311703189	0.007048924810, 0.004712444243
Max Distance	1.497 km	0.682 km	0.145 km	1.498 km
Mean Distance	0.543 km	0.182 km	0.038 km	0.449 km
Time Ratio	0.328	0.191	0.042	0.0367
# of GPS	15,365	5,738	1,856	1,646
Stay Time (h)	26.177	15.270	3.406	2.923
Location	Home	School	Subway station	Friend's home

	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Centre Position	37.400867668205, 126.977248950500	37.567741769992, 126.969634582129	37.390674226339, 126.951644015371	37.510988042066, 127.074913886290
Std. Dev. of Position	0.000296494364, 0.000483393961	0.002888416328, 0.000955257796	0.000362558279, 0.000311703189	0.000082669006, 0.000215435877
Max Distance	0.098 km	0.791 km	0.191 km	0.075 km
Mean Distance	0.051 km	0.177 km	0.082 km	0.013 km
Time Ratio	0.0254	0.018	0.012	0.007
# of GPS	992	847	298	243
Stay Time (h)	2.028	1.482	0.974	0.608
Location	Subway station	Dating place	Subway station	Bus station

Table 4: Space-Time Mobility of LSJ

	AM 0 - 1	AM 1 - 2	AM 2 - 3	AM 3 - 4	AM 4 - 5	AM 5 - 6	AM 6 - 7	AM 7 - 8	AM 8 - 9	AM 9 - 10	AM 10 - 11	AM 11 - 12
Stay Prob.	0.777	0.879	0.981	1	0.998	0.77	0.84	0.673	0.703	0.647	0.805	0.482
Moving Prob.	0.223	0.121	0.019	0	0.012	0.23	0.16	0.327	0.297	0.353	0.195	0.518
Cluster Prob.	#0: 0.685 #1: 0.055 #2: 0.123 #3: 0.057 #4: 0.069 #5: 0.011	#0: 0.402 #3: 0.553 #4: 0.045	#0: 0.521 #3: 0.479	#0: 1	#0: 1	#0: 0.984 #4: 0.016	#0: 0.509 #2: 0.059 #4: 0.010 #7: 0.422	#0: 0.835 #2: 0.021 #3: 0.078 #4: 0.004 #5: 0.015 #7: 0.046	#0: 0.849 #1: 0.030 #2: 0.026 #3: 0.037 #4: 0.049 #5: 0.010	#0: 0.362 #1: 0.381 #2: 0.040 #3: 0.025 #4: 0.003 #5: 0.088 #6: 0.071	#0: 0.192 #1: 0.582 #2: 0.040 #3: 0.025 #4: 0.003 #5: 0.088 #6: 0.071	#0: 0.563 #1: 0.157 #2: 0.032 #3: 0.231 #4: 0.018

	PM 0 - 1	PM 1 - 2	PM 2 - 3	PM 3 - 4	PM 4 - 5	PM 5 - 6	PM 6 - 7	PM 7 - 8	PM 8 - 9	PM 9 - 10	PM 10 - 11	PM 11 - 12
Stay Prob.	0.528	0.48	0.655	0.603	0.967	0.882	0.577	0.547	0.792	0.824	0.729	0.646
Moving Prob.	0.472	0.52	0.345	0.397	0.033	0.118	0.423	0.453	0.208	0.176	0.271	0.354
Cluster Prob.	#0: 0.550 #1: 0.253 #2: 0.131 #3: 0.057 #3: 0.017 #5: 0.034	#0: 0.628 #1: 0.201 #2: 0.120 #4: 0.016 #5: 0.035	#0: 0.040 #1: 0.655 #2: 0.197 #5: 0.108	#0: 0.844 #1: 0.049 #2: 0.038 #4: 0.053 #5: 0.015	#0: 0.718 #1: 0.275 #5: 0.007	#0: 0.448 #1: 0.552	#0: 0.570 #1: 0.329 #2: 0.084 #4: 0.015 #4: 0.003	#0: 0.428 #1: 0.502 #2: 0.008 #4: 0.035 #5: 0.027	#0: 0.643 #1: 0.310 #2: 0.034 #4: 0.009 #5: 0.004	#0: 0.498 #1: 0.237 #2: 0.065 #4: 0.023 #5: 0.026 #6: 0.151	#0: 0.416 #1: 0.429 #2: 0.100 #4: 0.028 #5: 0.027	#0: 0.632 #1: 0.093 #2: 0.108 #3: 0.073 #4: 0.091 #5: 0.002

probability shows a similar mobility pattern to KHUs, while the SHY mobility model is quite different. Figure 5 is a Google Maps representation of LSJs location clusters. Figure 6 shows cluster #0, which has the highest TimeRatio among LSJ's clusters. Cluster #0 is the home of LSJ. Figure 7 shows cluster #1 as LSJ's school.

A transition matrix for CTMC representation of LSJ's human mobility model is presented in equation 22. The transition rate is also dependent on the TimeRatio of each location cluster. An unusual result was found between the size of the clusters and the TimeRatio. Cluster #5 has a bigger distance than any other cluster, while less than 2% of time is spent in cluster #5. This is because LSJ walks in the area of Palaces with his friend, leading to large area in cluster #5; however, LSJ only visited the place once. Comparing the LSJ and KHU mobility models shows a similar mobility pattern because they are university students. Both KHU and LSJ show a similar pattern in that the two most frequented places are home and school. It is assumed that most people frequent home and place of work. Table 3 shows statistical result of location clusters from LSH's mobility model. Table 4 shows space-time mobility pattern of LSJ in hourly base.

SHYs positioning data set was collected beginning in November 2011. He used various position-

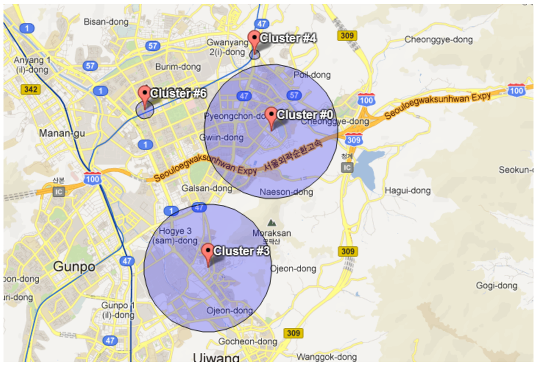

Figure 6: LSJ cluster #0

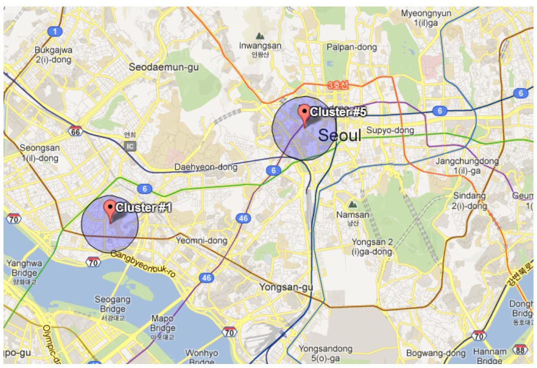

Figure 7: LSJ cluster #1

ing devices, such as iPhone 3GS, iPhone 4S, Garmin EDGE 800, Garmin EDGE 810 and Garmin GSMap 62s. SHY travels frequently, so a research mobility set inside Korea was selected. Figure 9 shows overall location clusters from SHYs positioning data set. SHY's clusters can be partitioned into two categories. One is clusters set around Seoul, Korea which is shown in figure 10. The other is cluster set inside Jeju island, Korea which is shown in figure 11. The zoom-ins of clusters around Seoul including clusters #0 and #1 can be found in in figure 12. In addition, frequent location clusters in Jeju can be zoomed-in as shown in in figure 13.

More than 32 location clusters are found in SHYs positioning data set. Thirty-two location clusters will be represented since the other clusters have a very small TimeRatio and very small stay time. These are regarded as miscellaneous locations. These 32 clusters and transitions between clusters can be represented in CTMC as shown in in figure 14.

KHU and LSJ, as students, show regular mobility patterns, such as commuting between home and school, and have uncommon visits to other locations. However, SHY visits other places more frequently, thus many clusters were identified. SHYs mobility

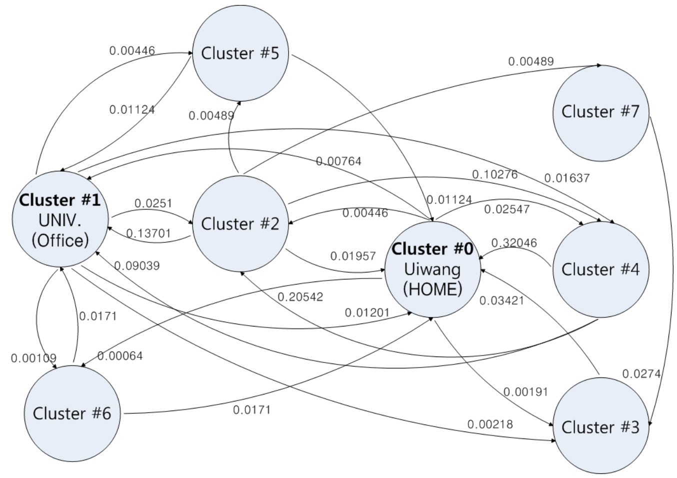

Figure 8: CTMC representation of LSJ's mobility model

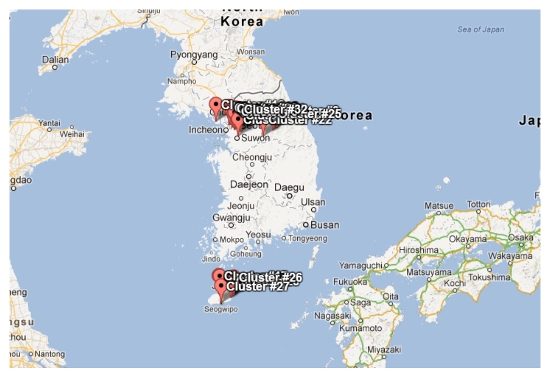

Figure 9: SHY's overall location clusters

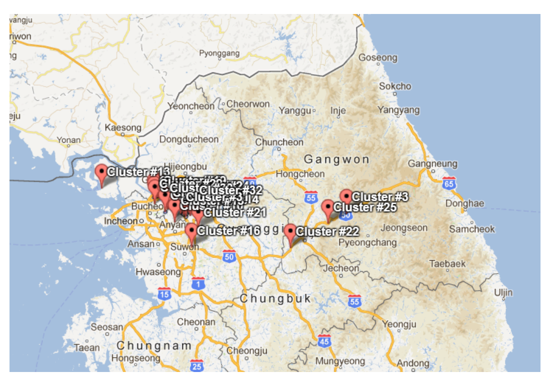

Figure 10: SHY's location clusters around Seoul, Korea

Figure 11: SHY's location clusters around Jeju, Korea

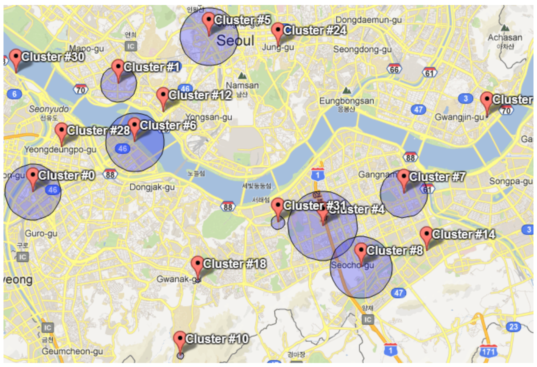

Figure 12: SHY's location clusters including #0 and #1

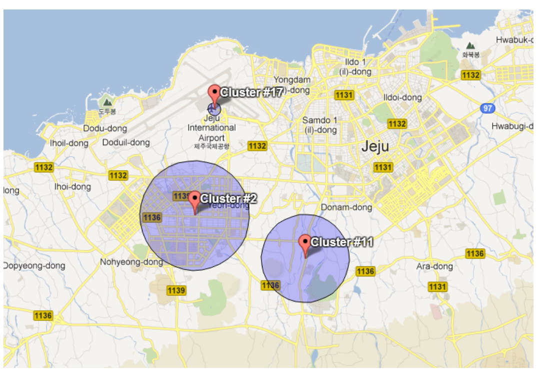

Figure 13: SHY's frequent location clusters in Jeju, Korea

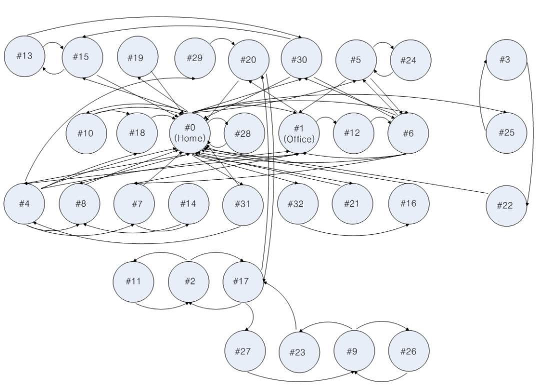

Figure 14: CTMC representation of SHY's human mobility

pattern also contains travel abroad or recreational bicycle riding, which is the reason for various clusters and various location visits. However, SHY also has a mobility pattern between home and school, which is more frequent than other locations, and he has a regular mobility pattern for commuting. One notable phenomenon is SHYs higher TimeRatio at home, which contrasts with KHU and LSJs patterns, which have similar TimeRatios at home and at school.

The numerical details of 32 location clusters can be found in table 5. Transition matrix for SHY is too big to be represented in traditional manner however can be represented in tabular form as shown in table 6. Final result of SHY's mobility analysis can be represented in space-time manner as shown in in table **??**.

6 Conclusions

In this research, a method was devised to construct a personal human mobility model from sets of positioning data. With sets of positioning data collected for specific durations, a space-time identification of certain locations was made by a mobility model construction process that included EM clustering of positioning data sets with proper probability density function and parameters for human mobility modelling. To intuitively represent a human mobility model, CTMC was utilised.

The locations identified by clustering algorithms were mapped onto corresponding states of a Markov chain. The transitions between locations were also mapped onto corresponding transition probabilities of the Markov chain. The timing over locations on a 24-hour basis can be derived from the CTMC representation of a human mobility model; thus, the space-time analysis of a certain model can be accomplished. On this basis, three positioning data sets for three people (SHY, KHU, LSJ) were converted into a human mobility model in order to verify the accuracy of our methodology.

We believe that our research provides a method to establish a human mobility model based on peoples' positioning data set. We found a method to represent a human mobility model over the space-time domain. Perhaps the human mobility model can be used in the area of mobile computing, location based services, surveillance or other related areas, as we can now predict a persons location in spatial-temporal manner. Considering the modern mobile computers usually have positioning functionalities on hand, a location aware computing ability of such computers can

Table 5: Statistical Analysis Result of SHY's Location Clusters

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Centre Position	37.510542645874, 126.883514793096	37.550493028734, 126.923490264870	33.486097721572, 126.488972663256	37.502453969482, 128.231263372909	37.498082298367, 127.018610627529
Std. Dev. of Position	0.002618603791, 0.002854382858	0.001877523216, 0.002330732945	0.003776040732, 0.003583041430	0.002330771012, 0.003295885406	0.004978967776, 0.006749485328
Max Distance	1.431 km	0.878 km	1.315 km	0.894 km	1.495 km
Mean Distance	0.256 km	0.213 km	0.484 km	0.113 km	0.615 km
Time Ratio	0.451	0.172	0.024	0.022	0.019
# of GPS	292019	95838	9842	6540	40894
Stay Time (h)	246.661	94.352	13.223	12.2	10.624
Location	Home	School	Jeju Univ.	Freshman OT	Restaurant

	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
Centre Position	37.567945185807, 126.965715605609	37.528954096954, 126.931048273789	37.509896831252, 127.056515897400	37.482666228215, 127.036576637423	33.428264662410, 126.929795725146
Std. Dev. of Position	0.003219198856, 0.003705098465	0.003750247051, 0.004090337664	0.002539633547, 0.002215850456	0.004442491778, 0.003659221698	0.002880605293, 0.002808822194
Max Distance	1.359 km	1.368 km	1.178 km	1.482 km	0.586 km
Mean Distance	0.395 km	0.404 km	0.265 km	0.401 km	0.318 km
Time Ratio	0.018	0.015	0.010	0.009	0.008
# of GPS	23627	61488	10907	13363	8338
Stay Time (h)	9.858	8.466	5.640	5.215	4.525
Location	Conference	Park	Conference	Meeting	Hotel

	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14
Centre Position	37.450009494659, 126.952288799157	33.477771955830, 126.514768171930	37.540303681501 126.944378388837	37.649658504190, 126.409883476006	37.488851021738, 127.067128533893
Std. Dev. of Position	0.000570231741, 0.000263763432	0.003110346678, 0.003073773188	0.000022446853, 0.000034302481	0.000107570133, 0.000124256412	0.000002399970, 0.000023818912
Max Distance	0.152 km	1.039 km	0.016 km	0.031 km	0.006 km
Mean Distance	0.057 km	0.405 km	0.003 km	0.014 km	0.001 km
Time Ratio	0.004	0.004	0.003	0.002	0.002
# of GPS	634	1778	772	1597	513
Stay Time (h)	2.508	2.396	1.892	1.570	1.390
Location	Seoul Nat'l Univ	Restaurant	Restaurant	Beach	Bar

	Cluster 15	Cluster 16	Cluster 17	Cluster 18	Cluster 19
Centre Position	37.596187680590, 126.796857038599	37.300368376232, 127.078067263383	33.506738594910, 126.493606689729	37.477762051930, 126.960444706845	37.597669774757, 126.828535739283
Std. Dev. of Position	0.004616588250, 0.008025714091	0.000044445525, 0.000076084443	0.000648280999, 0.000361418041	0.000152732387, 0.000314316787	0.001841838468, 0.002308691958
Max Distance	1.044 km	0.029 km	0.142 km	0.098 km	0.478 km
Mean Distance	0.493 km	0.007 km	0.071 km	0.022 km	0.276 km
Time Ratio	0.002	0.001	0.001	0.001	0.001
# of GPS	36074	1092	172	241	1007
Stay Time (h)	1.390	1.088	1.027	0.868	0.667
Location	Bike park	RELATIVE'S home	Jeju int'l airport	Restaurant	Subway station

Table 5: Statistical Analysis Result of SHY's Location Clusters (continued)

	Cluster 20	Cluster 21	Cluster 22	Cluster 23	Cluster 24
Centre Position	37.559099716943, 126.803342397330	37.409833911924, 127.126202625940	37.297884159354, 127.817683623099	33.372331741492, 126.855548449210	37.564213344751, 126.997696706216
Std. Dev. of Position	0.000166443570, 0.000257817505	0.000275052425, 0.000676320386	0.000020265616, 0.000019052684	0.000098015695, 0.000336265300	0.000009575073, 0.000037966743
Max Distance	0.058 km	0.205 km	0.005 km	0.058 km	0.005 km
Mean Distance	0.026 km	0.054 km	0.002 km	0.026 km	0.003 km
Time Ratio	0.0010	0.0009	0.0008	0.0007	0.0007
# of GPS	2169	3993	294	309	837
Stay Time (h)	0.583	0.540	0.480	0.422	0.405
Location	Kimpo int'l airport	Restaurant	Rest area	Restaurant	Evaluation

	Cluster 25	Cluster 26	Cluster 27	Cluster 28	Cluster 29
Centre Position	37.437842789678, 128.095720880899	33.460842115780, 126.933232707191	33.247069272504, 126.568897543964	37.527138873502, 126.896926032165	37.578316987104, 126.796312813848
Std. Dev. of Position	0.000054150602, 0.000037417126	0.000239323964, 0.000731847115	0.007849696500, 0.027381633229	0.000036941818, 0.000005965004	0.000041426305, 0.000088696894
Max Distance	0.015 km	0.162 km	1.088 km	0.005 km	0.013 km
Mean Distance	0.006 km	0.065 km	0.792 km	0.004 km	0.008 km
Time Ratio	0.0007	0.0006	0.0006	0.0005	0.0004
# of GPS	609	277	4220	114	6
Stay Time (h)	0.402	0.361	0.336	0.314	0.26
Location	Student MT	Restaurant	Restaurant	Restaurant	Subway station

	Cluster 30	Cluster 31	Cluster 32
Centre Position	37.554347407970, 126.875649321431	37.499282397306, 126.997779179803	37.538477890923, 127.095232792963
Std. Dev. of Position	0.000013906746, 0.000035270453	0.000897272149, 0.001343433394	0.000171986683, 0.000143806248
Max Distance	0.009 km	0.268 km	0.033 km
Mean Distance	0.003 km	0.147 km	0.021 km
Time Ratio	0.0004	0.0003	0.0003
# of GPS	1124	565	75
Stay Time (h)	0.247	0.183	0.177
Location	Bike park	Vote office	Relative's home

Table 6: Transition Matrix for SHY's Human Mobility Model

Clusters	Transition Prob.	Clusters	Transition Prob.	Clusters	Transition Prob.	Clusters	Transition Prob.
# 0 → # 1	0.00203	# 1 → # 0	0.00548	# 5 → # 1	0.01014	#10 → # 0	0.00665
# 0 → # 4	0.00027	# 1 → # 4	0.00018	# 5 → # 6	0.00169	#10 → #18	0.00665
# 0 → # 5	0.00014	# 1 → # 6	0.00071	# 5 → #24	0.00169	#11 → # 2	0.02086
# 0 → # 6	0.00223	# 1 → # 7	0.00018	# 6 → # 0	0.06103	#11 → #27	0.00695
# 0 → # 8	0.00007	# 1 → #12	0.00018	# 6 → # 1	0.00197	#12 → # 6	0.00881
# 0 → #10	0.00014	# 1 → #20	0.00018	# 6 → # 5	0.00984	#13 → #15	0.01061
# 0 → #15	0.00020	# 2 → #11	0.00504	# 6 → # 7	0.00197	#14 → # 8	0.01198
# 0 → #19	0.00014	# 2 → #17	0.00126	# 6 → #30	0.00394	#15 → # 0	0.01199
# 0 → #21	0.00014	# 3 → #22	0.00136	# 7 → # 0	0.00886	#15 → #13	0.01199
# 0 → #25	0.00007	# 4 → # 0	0.00157	# 7 → #14	0.00295	#15 → #30	0.03596
# 0 → #28	0.00007	# 4 → # 1	0.00157	# 8 → # 0	0.00639	#16 → # 0	0.01532
# 0 → #30	0.00439	# 4 → # 7	0.00157	# 8 → # 7	0.00320	#17 → # 2	0.03244
# 0 → #31	0.00014	# 4 → # 8	0.00157	# 9 → #23	0.00368	#17 → #20	0.01622
# 0 → #32	0.00007	# 4 → #18	0.00157	# 9 → #26	0.00368	#18 → # 0	0.03840
#19 → # 0	0.04992	# 4 → #29	0.00157	#24 → # 5	0.04110	#30 → # 0	0.97143
#20 → # 0	0.02854	#21 → # 0	0.06166	#25 → # 3	0.04144	#30 → # 6	0.01429
#20 → # 2	0.02854	#22 → # 0	0.03470	#26 → # 9	0.04946	#30 → #15	0.01429
#20 → #17	0.02854	#23 → #17	0.03947	#27 → # 9	0.04946	#31 → # 0	0.09077
#28 → # 0	0.05305	#29 → #20	0.06410	#32 → #16	0.09390	#31 → #16	0.09390

Table 7: Space-Time Mobility of SHY

	AM 0 - 1	AM 1 - 2	AM 2 - 3	AM 3 - 4	AM 4 - 5	AM 5 - 6	AM 6 - 7	AM 7 - 8	AM 8 - 9	AM 9 - 10
Stay Prob.	0.723	0.545	0.688	0.664	0.637	0.638	0.411	0.328	0.521	0.424
Moving Prob.	0.277	0.455	0.312	0.336	0.363	0.362	0.589	0.672	0.279	0.576
Cluster Prob.	# 0: 0.826 # 1: 0.003 # 2: 0.006 # 3: 0.029 # 4: 0.105 # 6: 0.009 # 7: 0.018 #18: 0.004	# 0: 0.800 # 1: 0.035 # 2: 0.002 # 3: 0.016 # 4: 0.011 # 6: 0.064 #15: 0.069 #30: 0.003	# 0: 0.709 # 1: 0.053 # 3: 0.023 # 6: 0.005 #15: 0.207 #30: 0.001	# 0: 0.756 # 1: 0.068 # 3: 0.027 #15: 0.147 #30: 0.002	# 0: 0.540 # 1: 0.008 # 3: 0.029 # 5: 0.128 #15: 0.199 #19: 0.007 #20: 0.089	# 0: 0.499 # 1: 0.001 # 3: 0.034 # 5: 0.128 # 6: 0.025 # 7: 0.012 # 8: 0.019 #15: 0.224 #20: 0.067 #19: 0.008 #30: 0.003	# 0: 0.652 # 1: 0.025 # 3: 0.018 # 4: 0.028 # 5: 0.001 # 6: 0.166 # 7: 0.005 # 8: 0.010 #15: 0.085 #21: 0.035 #32: 0.003	# 0: 0.338 # 1: 0.017 # 3: 0.014 # 4: 0.145 # 5: 0.081 # 6: 0.233 # 7: 0.017 # 8: 0.023 #15: 0.094	# 0: 0.343 # 1: 0.212 # 3: 0.018 # 4: 0.045 # 5: 0.005 # 6: 0.194 # 7: 0.003 # 8: 0.008 #15: 0.120 #16: 0.010 #19: 0.023 #21: 0.018 #30: 0.001	# 0: 0.373 # 1: 0.075 # 2: 0.001 # 3: 0.015 # 4: 0.038 # 5: 0.127 # 6: 0.220 # 8: 0.081 #11: 0.021 #16: 0.015 #21: 0.032 #24: 0.002 #30: 0.001

	AM 10 - 11	AM 11 - 12	PM 0 - 1	PM 1 - 2	PM 2 - 3	PM 3 - 4	PM 4 - 5	PM 5 - 6	PM 6 - 7	PM 7 - 8
Stay Prob.	0.593	0.727	0.547	0.43	0.411	0.512	0.549	0.515	0.519	0.645
Moving Prob.	0.407	0.273	0.453	0.57	0.589	0.488	0.451	0.485	0.481	0.355
Cluster Prob.	# 0: 0.737 # 1: 0.067 # 2: 0.009 # 3: 0.019 # 4: 0.053 # 5: 0.024 # 6: 0.062 #16: 0.010 #24: 0.019	# 0: 0.499 # 1: 0.160 # 2: 0.058 # 4: 0.049 # 5: 0.078 # 6: 0.045 # 7: 0.027 # 9: 0.008 #10: 0.008 #11: 0.017 #16: 0.009 #17: 0.002 #22: 0.011 #24: 0.020	# 0: 0.452 # 1: 0.239 # 2: 0.049 # 5: 0.109 # 6: 0.012 # 7: 0.040 # 8: 0.010 # 9: 0.038 #10: 0.005 #11: 0.008 #17: 0.003 #18: 0.002 #22: 0.003	# 0: 0.467 # 1: 0.228 # 2: 0.037 # 4: 0.006 # 5: 0.069 # 6: 0.035 # 7: 0.011 # 9: 0.033 #11: 0.004 #15: 0.090 #18: 0.002 #27: 0.012 #30: 0.002 #31: 0.001	# 0: 0.290 # 1: 0.204 # 2: 0.024 # 4: 0.129 # 5: 0.044 # 6: 0.191 # 7: 0.002 # 8: 0.003 # 9: 0.021 #11: 0.005 #14: 0.005 #15: 0.072 #20: 0.007 #30: 0.003	# 0: 0.404 # 1: 0.191 # 2: 0.033 # 4: 0.066 # 5: 0.042 # 6: 0.098 # 7: 0.020 # 8: 0.049 # 9: 0.038 #10: 0.001 #14: 0.005 #15: 0.035 #20: 0.001	# 0: 0.262 # 1: 0.231 # 2: 0.022 # 4: 0.080 # 5: 0.057 # 6: 0.115 # 7: 0.030 # 8: 0.059 # 9: 0.010 #10: 0.001 #15: 0.052 #17: 0.002 #19: 0.007 #21: 0.025 #23: 0.003 #27: 0.039 #28: 0.002 #31: 0.001	# 0: 0.264 # 1: 0.335 # 3: 0.005 # 4: 0.122 # 5: 0.007 # 6: 0.089 # 7: 0.013 # 8: 0.010 # 9: 0.019 #13: 0.010 #15: 0.046 #19: 0.001 #21: 0.003 #25: 0.002 #27: 0.050 #30: 0.001 #31: 0.012	# 0: 0.342 # 1: 0.282 # 2: 0.008 # 3: 0.011 # 4: 0.057 # 5: 0.005 # 6: 0.201 # 8: 0.046 # 9: 0.007 #10: 0.005 #11: 0.001 #12: 0.002 #13: 0.017 #21: 0.016	# 0: 0.470 # 1: 0.171 # 2: 0.017 # 3: 0.014 # 4: 0.097 # 5: 0.006 # 6: 0.069 # 7: 0.016 # 8: 0.041 # 9: 0.023 #11: 0.003 #12: 0.011 #13: 0.013 #15: 0.036 #16: 0.004 #21: 0.004 #26: 0.005

	PM 8 - 9	PM 9 - 10	PM 10 - 11	PM 11 - 12
Stay Prob.	0.645	0.638	0.747	0.718
Moving Prob.	0.355	0.362	0.253	0.282
Cluster Prob.	# 0: 0.527 # 1: 0.108 # 2: 0.001 # 3: 0.015 # 4: 0.079 # 5: 0.002 # 6: 0.121 # 8: 0.019 #11: 0.004 #12: 0.012 #15: 0.0103 #25: 0.007 #26: 0.001	# 0: 0.512 # 1: 0.063 # 2: 0.008 # 3: 0.005 # 4: 0.118 # 5: 0.012 # 6: 0.044 # 7: 0.115 # 8: 0.020 # 9: 0.021 #11: 0.004 #15: 0.060 #25: 0.013 #26: 0.004 #30: 0.001	# 0: 0.716 # 1: 0.074 # 2: 0.003 # 3: 0.022 # 4: 0.100 # 6: 0.009 # 7: 0.063 # 9: 0.003 #18: 0.002 #25: 0.005 #30: 0.004	# 0: 0.835 # 1: 0.019 # 2: 0.004 # 3: 0.023 # 4: 0.051 # 6: 0.006 # 7: 0.048 #15: 0.012 #30: 0.002

be widely used and our model may help for location dependent applications. For example, an unmanned automotive equipped with personal mobility model can help passenger's computing life more seamlessly.

Acknowledgements: Author gives many thanks to all assitants, especially Mr. Hyunuk Kim, for their help to finish this work.

References:

[1] N. A. of Public Administration, "The global positioning system: a shared national asset: recommendations for technical improvements and enhancements," *National Academies Press*, 2013.

[2] GLONASS. [Online]. Available: URLhttp://glonass-iac.ru/en/

[3] GALILEO. [Online]. Available: URLhttp://ec.europa.eu/enterprise/policies/satnav/index_en.htm

[4] R. W. Byrne, "Memory for urban geography," *The Quarterly Journal of Experimental Psychology*, vol. 31, no. 1, pp. 147–154, 1979.

[5] J. N. Bailenson, M. S. Shum, and D. H. Uttal, "Road climbing: Principles governing asymmetric route choices on maps," *Journal of Environmental Psychology*, vol. 18, no. 3, pp. 251–264, 1998.

[6] K. W. Jacobs and J. C. Koeppel, "The prediction of individual mobility plans." 1973.

[7] B. Verplanken, H. Aarts, and A. Van Knippenberg, "Habit, information acquisition, and the process of making travel mode choices," *European Journal of Social Psychology*, vol. 27, no. 5, pp. 539–560, 1997.

[8] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, no. 7075, pp. 462–465, 2006.

[9] S. E. Wiehe, S. C. Hoch, G. C. Liu, A. E. Carroll, J. S. Wilson, and J. D. Fortenberry, "Adolescent travel patterns: pilot data indicating distance from home varies by time of day and day of week," *Journal of adolescent health*, vol. 42, no. 4, pp. 418–420, 2008.

[10] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[11] P. Wang, M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding the spreading patterns of mobile phone viruses," *Science*, vol. 324, no. 5930, pp. 1071–1076, 2009.

[12] V. Colizza and A. Vespignani, "Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations," *Journal of theoretical biology*, vol. 251, no. 3, pp. 450–467, 2008.

[13] S. Ni and W. Weng, "Impact of travel patterns on epidemic dynamics in heterogeneous spatial metapopulation networks," *Physical review E*, vol. 79, no. 1, p. 016111, 2009.

[14] P. Wang, M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding the spreading patterns of mobile phone viruses," *Science*, vol. 324, no. 5930, pp. 1071–1076, 2009.

[15] F. Bai, N. Sadagopan, and A. Helmy, "Important: A framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 2. IEEE, 2003, pp. 825–835.

[16] B. Zhou, K. Xu, and M. Gerla, "Group and swarm mobility models for ad hoc network scenarios using virtual tracks," in *Military Communications Conference, 2004. MILCOM 2004. 2004 IEEE*, vol. 1. IEEE, 2004, pp. 289–294.

[17] J. Wang, H.-Y. Wang, Q.-H. Gao, and M.-L. Jin, "Enhanced monte carlo localization and tracking algorithm for mobile wireless sensor network," *Dianzi Yu Xinxi Xuebao(Journal of Electronics and Information Technology)*, vol. 32, no. 4, pp. 864–868, 2010.

[18] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "Slaw: A new mobility model for human walks," in *INFOCOM 2009, IEEE*. IEEE, 2009, pp. 855–863.

[19] S. A. Williams and D. Huang, "A group force mobility model," *Simulation Series*, vol. 38, no. 2, p. 333, 2006.

[20] K. H. Wang and B. Li, "Group mobility and partition prediction in wireless ad-hoc networks," in *Communications, 2002. ICC 2002. IEEE International Conference on*, vol. 2. IEEE, 2002, pp. 1017–1021.

[21] E. Hyytia, P. Lassila, and J. Virtamo, "Spatial node distribution of the random waypoint mobility model with applications," *Mobile Computing, IEEE Transactions on*, vol. 5, no. 6, pp. 680–694, 2006.

[22] C. Bettstetter, "Smooth is better than sharp: a random mobility model for simulation of wireless networks," in *Proceedings of the 4th ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems*. ACM, 2001, pp. 19–27.

[23] C. Papageorgiou, K. Birkos, T. Dagiuklas, and S. Kotsopoulos, "An obstacle-aware human mobility model for ad hoc networks," in *Modeling, Analysis & Simulation of Computer and Telecommunication Systems, 2009. MASCOTS'09. IEEE International Symposium on*. IEEE, 2009, pp. 1–9.

[24] C. Boldrini, M. Conti, and A. Passarella, "The sociable traveller: human travelling patterns in social-based mobility," in *Proceedings of the 7th ACM international symposium on Mobility management and wireless access*. ACM, 2009, pp. 34–41.

[25] J. Ghosh, S. J. Philip, and C. Qiao, "Sociological orbit aware location approximation and routing (solar) in manet," *Ad Hoc Networks*, vol. 5, no. 2, pp. 189–209, 2007.

[26] S. Rinzivillo, D. Pedreschi, M. Nanni, F. Giannotti, N. Andrienko, and G. Andrienko, "Visually driven analysis of movement data by progressive clustering," *Information Visualization*, vol. 7, no. 3-4, pp. 225–239, 2008.

[27] F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, C. Renso, S. Rinzivillo, and R. Trasarti, "Unveiling the complexity of human mobility by querying and mining massive trajectory data," *The VLDB JournalThe International Journal on Very Large Data Bases*, vol. 20, no. 5, pp. 695–719, 2011.

[28] S. Lu, R. M. Li, W.-C. Tjhi, K. K. Lee, L. Wang, X. Li, and D. Ma, "A framework for cloud-based large-scale data analytics and visualization: Case study on multiscale climate data," in *Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on*. IEEE, 2011, pp. 618–622.

[29] C. Joh, "Measuring and predicting adaptation in multidimensional activity-travel patterns, bouwstenen, no. 79," *Faculteit Bouwkunde, TU Eindhoven, Eindhoven*, 2004.

[30] T. Jang, "Causal relationship among travel mode, activity, and travel patterns," *Journal of transportation engineering*, pp. 16–22, 2003.

[31] R. W. Sinnott, "Virtues of the haversine," *Sky and Telescope*, vol. 68, no. 2, p. 159, 1984.

[32] "Garmin GPSMAP62s." [Online]. Available: URLhttp://buy.garmin.com/en-US/US/prod63801.html

[33] "Garmin EDGE800." [Online]. Available: URLhttp://buy.garmin.com/en-US/US/into-sports/cycling/edge-800/prod69043.html

[34] Garmin EDGE810. [Online]. Available: URLhttp://buy.garmin.com/en-US/US/into-sports/cycling/edge-810/prod112912.html

[35] Sportstracker. [Online]. Available: URLhttp://www.sports-tracker.com

[36] iPhone. [Online]. Available: URLhttp://www.apple.com/iphone/

[37] C. C. Aggarwal and C. K. Reddy, *Data Clustering: Algorithms and Applications*. CRC Press, 2013.