

Investigating Cyclic Visit Pattern of Mobility Through Analysis of Geopositioning Data

Ha Yoon Song(✉) and Suchan Hong

Department of Computer Engineering, Hongik University, Seoul, Republic of Korea
hayoon@hongik.ac.kr, suchanhong@naver.com

Abstract. Intuitions guide us that there are cyclic patterns for a person to visit a location, and there is a tendency of multiple cycles in visiting patterns. Nowadays, it is possible for a person to collect personal mobility data due to the help of smartphones and other portable devices. These devices collects raw geolocation (or geopositioning) data and the set of geolocation data can be analyzed in various ways. Based on location clusters distilled from raw geolocation data, we can establish mobility model of a person and investigate cyclic patterns of a person to visit location clusters. Based on the aggregate personal mobility models collected over several years, we calculated and analyzed the cluster revisiting time and visualized it as a graph. Regarding geolocation data for location clusters as set of time sequence, number of visiting cluster is measured in a unit of minutes. The number of visits from whole data is normalized in every 15 min. For various geolocation data set of a volunteer, cyclic patterns of a visit are examined in terms of autocorrelation, autocovariance and intervisiting time.

Keywords: Mobility pattern analysis · Mobility modeling · Temporal mobility · Cyclic mobility pattern · Recurrent location visit

1 Introduction

Recently IoT and smart city are emerging with 4IR, commonly known as 4th Industrial Revolution. This also requires a model of human mobility based on individual geolocation information. The effort to create profit by applying the model of human mobility pattern to various industrial environments has been diversified, and basic research on this has become important. Human mobility pattern model become available thanks to smart and IoT devices which can collects large amount of geolocation data. This sort of research will be a help to Location Based Services (LBS).

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (NRF-2019R1F1A1056123).

© Springer Nature Switzerland AG 2019
S. Misra et al. (Eds.): ICCSA 2019, LNCS 11619, pp. 589–602, 2019.
https://doi.org/10.1007/978-3-030-24289-3_44

We can understand the relation between individual location and cluster based on human mobility information model. People are staying in one place or move to another place in a time sequence. We used human geolocation data to analyze revisiting time to the location. In other words, it would be valuable to figure out cyclic visit pattern to a location cluster latent in human mobility model. Objects other than human also have their own mobility pattern. Therefore, we can also examine object mobility pattern. However our research is based on human mobility data since human is one of the most active mobile object.

In this paper, we use two things to conduct time analysis to mobility pattern: (1) geolocation information collected over a period of time. (2) Individual integrated mobility model by using collected geolocation information. The geolocation information is composed of <date, time, UNIX time, latitude and longitude>, and the information of individual integrated mobility model is composed of <cluster number, latitude of cluster center, longitude of cluster center, and cluster radius>. The time analysis on mobility pattern is based on internal cluster data among all geolocation data. A volunteer provided sets of raw geopositioning data, and named the data set as CDY. The volunteer had collected geolocation data by smartphone with an app named Sports Tracker [1] and GPS receiver manufactured by Garmin [2].

The paper is comprised of the following parts. Section 2 introduces existing studies using geolocation data, and Sect. 3 introduces a method of predicting mobility patterns by time series analysis. In Sect. 4, we describe the experimental results on the mobility pattern prediction method. In Sect. 5, we conclude the paper with a summary.

2 Existing Research

There has been research based on geolocation data in many ways. Creating individual mobility model with geopositioning data is one of the base of our research by use of clustering with Expectation Maximization algorithm [3,4]. The EM (expectation maximization) algorithm is a clustering algorithm first proposed by Hartley in 1958 [5] and systematized by Dempster in 1977 [6], and is an algorithm that creates an optimized model through the iterative refinement process after generating the initial model. The expectation maximization algorithm is an algorithm for finding parameters with maximum likelihood and maximum a posteriori through iterative refinement process. Log-likelihood function is used to evaluate the suitability of the model when finding parameter. In the expectation stage, when $\theta^{(t)}$ is given, we define the expectation value Q of the likelihood when a new θ is used. In the maximization step, a new parameter $\theta^{(t+1)}$ that maximizes Q is calculated. The EM algorithm is used as follows. The initial value of the parameter θ is arbitrary set. It stops when $\theta^{(t)}$ is sufficiently converged to some value. When converged, the probability of moving among the generated clusters is generated in a form of Markov chains. A Study on generating group mobility model based on individual mobility model exists [7].

In order to generate a collective mobility model, the distance between the clusters must be obtained. Here, Haversine formula [8] is necessary to find the distance on earth, which is the distance between two points on a sphere given their longitudes and latitudes. The following shows the details of Haversine formula.

$$latitude = (latitude_2 - latitude_1) \times \pi/180$$
$$longitude = (longitude_2 - longitude_1) \times \pi/180$$
$$latitude_1 = latitude_1 \times \pi/180$$
$$latitude_2 = latitude_2 \times \pi/180$$
$$a = sin(latitude/2)^2 + cos(latitude_1) \times cos(latitude_2) \times sin(longitude/2)^2$$
$$c = 2 \times atan2(\sqrt{a}, \sqrt{(1-a)})$$

then the distance between to geopositioning locations is

$$\mathrm{d} = \mathrm{R} \times \ \mathrm{c} \text{ where R is earth's radius (mean radius} = 6{,}371 \text{ km}).$$

Compare the distance and the radius between the centers of the clusters in the two individual mobility models, respectively. If the distance of the cluster center is less than the sum of the radii of the two clusters, a macro and a micro cluster are created. When clustering, the micro cluster collects all location information that is common to both clusters. The macro cluster collects and clusters all location information contained in at least one of two clusters. Based on EM algorithm and Haversine formula, meaningful location clusters were distilled from raw geopositioning data as shown in [4]. The location clusters identified will be the basis for the following chapter.

A method to capture the relation between a human mobility and the social context by data analysis [9] exists but not that much related to our research. There also has been study on extracting human mobility patterns from GPS data. They identified the relation between human mobility data and sociological aspects of human movements.

The other research is regarding a template that decompose seasonal trend for long time series. A method to decompose complex long time series into seasonality, trend, and remainder components. It enables time series irregularity detection and forecasting [10].

Although there have been many studies, to our knowledge, no research has been done on cyclic pattern of object mobility. Our research attempts to find relationship between object mobility over time domain.

3 Temporal Analysis on Mobility Pattern

In this study, we try to analyze the correlation by using temporal data in order to identify cyclic movement patterns through the location data. We try to identify the relationship between time and place based on the characteristics of staying in one place (location cluster) or moving to another place over time. One of the key point is to find out the meaningful sets of duration of cycle in mobility pattern.

3.1 Considerations

In order to determine if there is a correlation between time and place, it is necessary to determine whether there is a certain rule for the time range in the cluster. Therefore, we extracted the clusters with the individual integrated mobility models and confirmed the time of visiting the two most notable clusters. In this aspect, autocorrelation and autocovariance analysis were used to confirm the positional correlation over time. Autocorrelation, $R(t, \tau)$, measures the correlation coefficient with the intervisit time variable (τ) in the same time series. t stands for arbitrary time in the time series. The formula is as follows:

$$R(t, \tau) = E[X(t) \times X(t + \tau)]$$

Autocovariance, $C_X(t, \tau)$, is also a measure of the covariance with the time variable (τ) in the same time series. The formula is as follows.

$$C_X(t, \tau) = Cov[X(t), X(t + \tau)], \text{ or } C_X(t, \tau) = R_X(t, \tau) - \mu_t \mu_{t+\tau}$$

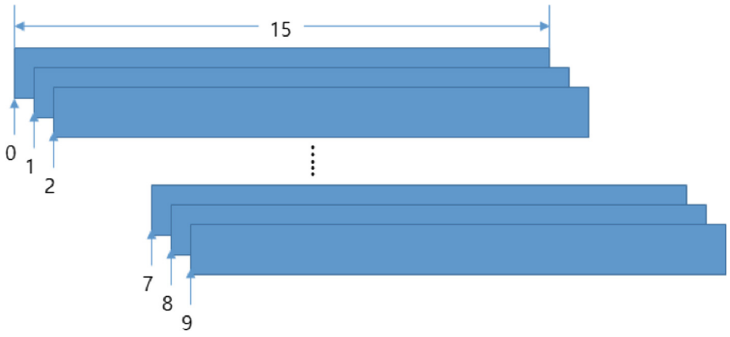


Fig. 1. Concept of time slice

3.2 Timing Analysis Based on Time Slice Window

Use of constant time array is required to obtain autocorrelation and autocovariance. The unit of time slice in our research is in 15 min. In other words, we regarded the data as meaningful in case that people stayed for more than 15 min. Figure 1 shows the concept of time slices.

The window slides over time and the range of time is 15 min. For the first step, our initial interest is the geolocation data for a specific clusters inside a time slice. Over the whole time domain, for every time slices, we calculated autocorrelation and autocovariance by using $E[X(t)]$ and $E[X(t + \tau)]$ as shown in Fig. 2.

We used mobility data from May 2013 to November 2017. Due to large amount of data and computation to calculate correlation and covariance, we used parallel programming technique, as known as GPGPU programming to save computation time. The experiments were done on the PC with Intel i7 CPU with 16 GB RAM and nVidia 980Ti GPU.

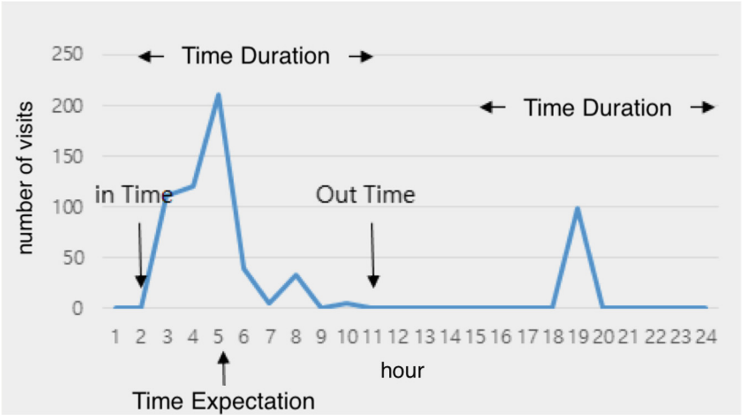


Fig. 2. In time, out time, time expectation with respect to time and number of visit

Table 1. Correlation coefficient

Range of correlation coefficient	Meaning
$-1.0 \leq r \leq -0.7$	very strong negative (−) correlation
$-0.7 < r \leq -0.3$	strong negative (−) correlation
$-0.3 < r \leq -0.1$	weak negative (−) correlation
$-0.1 < r \leq 0.1$	no correlation
$0.1 < r \leq 0.3$	weak positive (+) correlation
$0.3 < r \leq 0.7$	strong positive (+) correlation
$0.7 < r \leq 1.0$	very strong positive (+) correlation

4 Analyzed Results

4.1 Analyzed Results of Cyclic Revisit Time to Location Clusters

Firstly, we analyzed the correlation coefficient values to see whether they are meaningful according to the revisit time. Meaning of the value of correlation coefficient can be represented as shown in the following Table 1.

Each results can be represented in a graphical form and each graphs were drawn for two cases: (1) very strong negative (−) correlation and (2) very strong positive (+) correlation.

Figure 3 through Fig. 16 graphically represent the time which has very strong positive or negative correlation as represented in Table 1. Figures 3, 7, 9, 11, 13 and 15 are related with Cluster 1. Figures 4, 8, 10, 12, 14 and 16 are related with Cluster 2.

The difference between Cluster 1 and Cluster 2 must be compared for further analysis.

Figure 3 shows a correlation graph based on Cluster 1 of CDY for week 1. The horizontal axis represents time in minutes from 0 to 100080 and the vertical axis has a range of −1.0 to 1.0 representing correlation values.

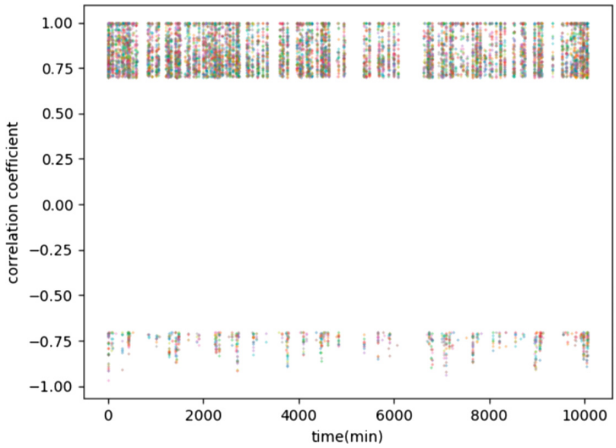


Fig. 3. Correlation graph for Cluster 1 for CDY data of the 1st week

Figure 4 shows a correlation graph based on Cluster 2 of CDY for week 1. Comparing with Fig. 3, correlation pattern is not overlapped and correlation pattern shows sparse frequency than that of Cluster 1. In other words, correlation in cluster 2 is less clear than that of cluster 1. In addition, it is possible to verify that the strong negative correlation coefficient is empty in 10000 to 10080 min of Fig. 4, whereas strong positive correlation coefficient is imprinted. This can be interpreted that Fig. 4 from 10000 to 10080 min have a high correlation.

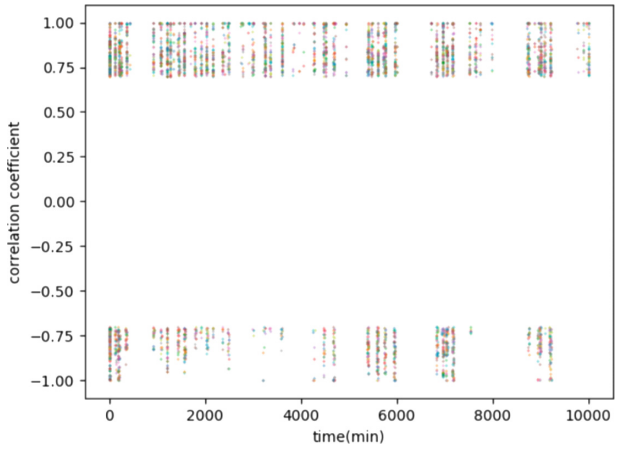


Fig. 4. Correlation graph for Cluster 2 for CDY data of the 1st week

Figure 5 shows a correlation graph based on Cluster 2 of CDY for week 2. All the data were sliced into 10080 min to calculate the correlation coefficient. Here in week 2, very strong positive correlation only shows in 0 to 9000 min. From the result, we can say that probability of revisiting Cluster 2 is high, whereas it is low from 9000 to 10080 min.

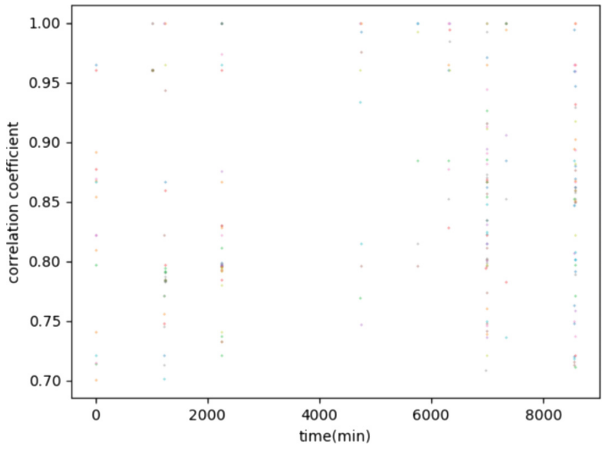


Fig. 5. Correlation graph for Cluster 2 for CDY data of the 2nd week

Figure 6 shows a correlation graph based on Cluster 2 of CDY for week 4. In contrast to Fig. 5, it shows many blanks. Few of very strong positive correlation can be found in 0 to 100 min, 7000 to 7500 min, and around 10000 min. These time domain doesn't have negative correlations which is closed to −1 value. Therefore, it is highly likely to revisit on 0 to minutes, 7000 to 7500 min, and 10000 min on week 4.

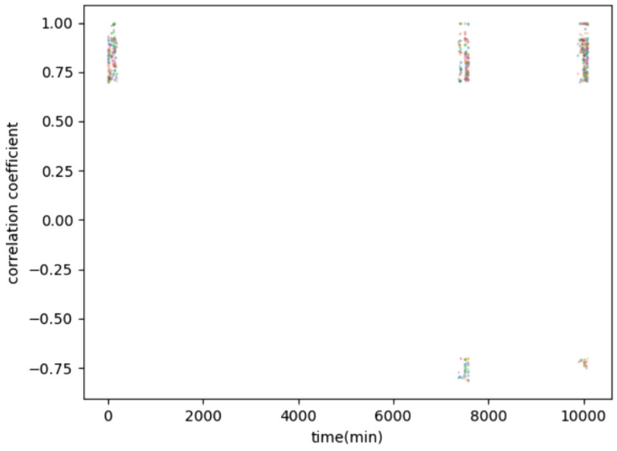


Fig. 6. Correlation graph for Cluster 2 for CDY data of the 4th week

Figures 7 and 8 shows a correlation graph based on Cluster 1 and Cluster 2 of CDY for week 10. By comparison, Cluster 1 shows revisiting time without distinguished intervisit time interval.

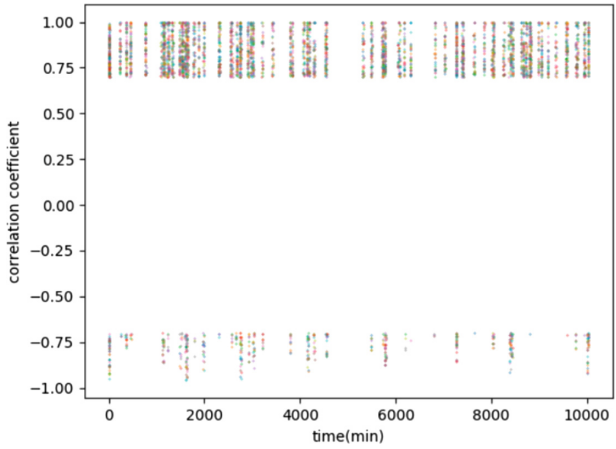


Fig. 7. Correlation graph for Cluster 1 for CDY data of the 10th week

Figure 8, shows similar pattern with Fig. 7. But it shows clear distinct pattern which is divided into 8 parts. Here, it can be said that revisit cycle is around 1 day or 1,440 min.

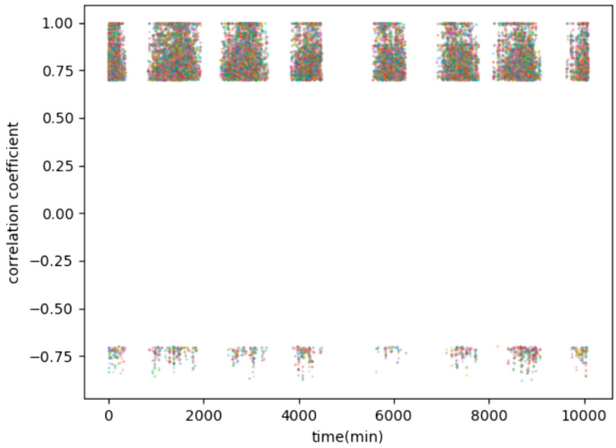


Fig. 8. Correlation graph for Cluster 2 for CDY data of the 10th week

Figures 9 and 10 shows a correlation graph based on Cluster 1 and Cluster 2 of CDY for week 11. Both graphs shows empty space from 7800 and 9300 min. From this result, we can see CDY didn't visit Cluster 1 nor Cluster 2 on 6th day of week 11.

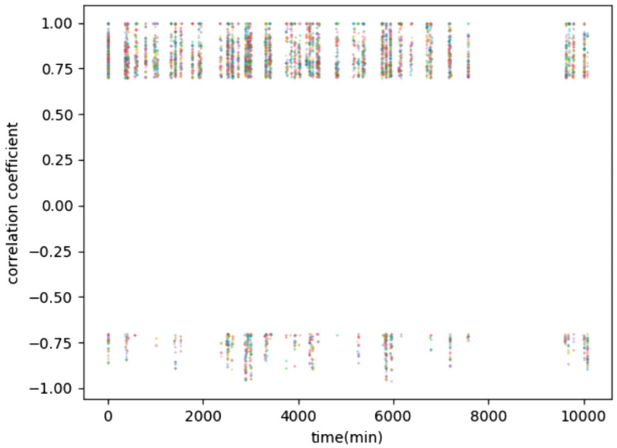


Fig. 9. Correlation graph for Cluster 1 for CDY data of the 11th week

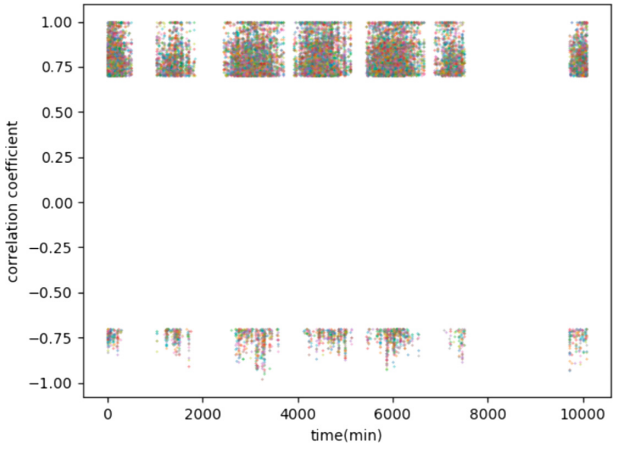


Fig. 10. Correlation graph for Cluster 2 for CDY data of the 11th week

Figures 11 and 12 shows a correlation graph based on Cluster 1 and Cluster 2 of CDY for week 12. Both graphs shows similar pattern where 3500 min to 6000 min is empty. But Cluster 2 shows additional empty space around 2000 min, around 8000 min, and around 9000 min. Here, we can see neither of clusters were visited from 3500 min to 6000 min. And Cluster 2 was not visited around 2000 min, 8000 min and 9000 min additionally.

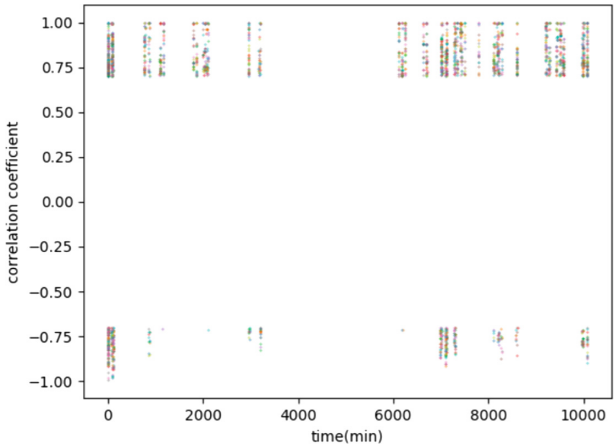


Fig. 11. Correlation graph for Cluster 1 for CDY data of the 12th week

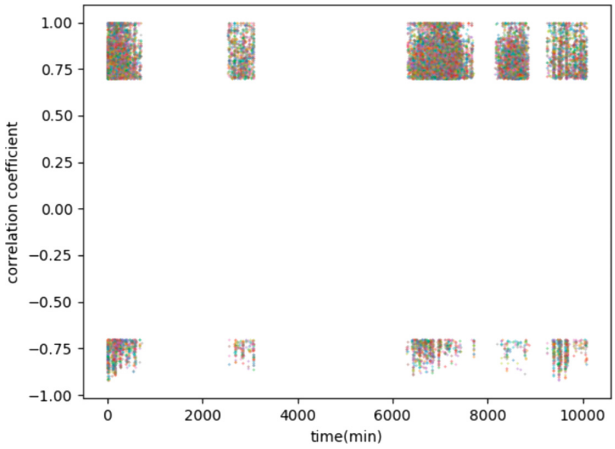


Fig. 12. Correlation graph for Cluster 2 for CDY data of the 12th week

Figures 13 and 14 shows a correlation graph based on Cluster 1 and Cluster 2 of CDY for week 17. In Fig. 13, strong correlation is evenly sparse and it does not have specific pattern. This can be said that CDY will visited Cluster 1 regardless of time in week 17.

Figure 14 differs from Fig. 13. It is divided into six zones in total, with a distinguished strong correlation pattern. Here, CDY is highly likely to visit Cluster 2 in dense zone and not visit in empty zone. There's no empty space between day 2 to day 3 and day 6 to day 7. Here, we can see CDY is mostly likely to stay in Cluster 2. There's some strong positive domain without strong negative correlation. This can be said that it's highly likely to revisit.

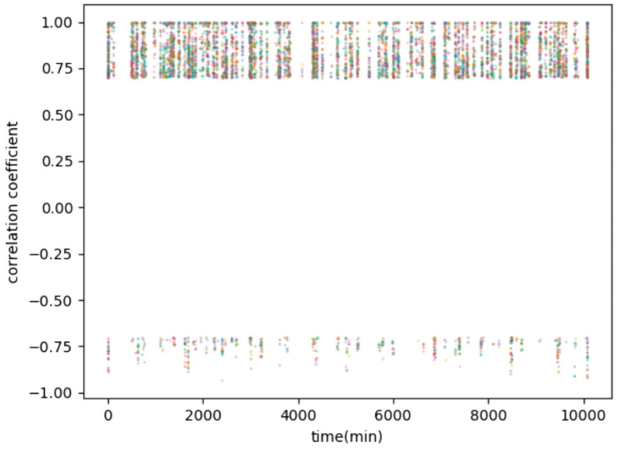


Fig. 13. Correlation graph for Cluster 1 for CDY data of the 17th week

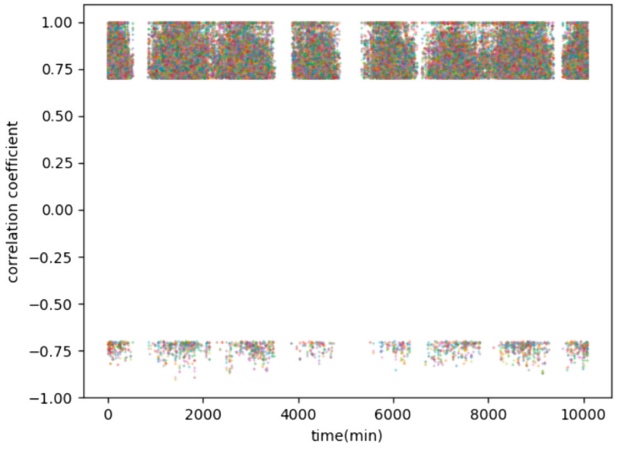


Fig. 14. Correlation graph for Cluster 2 for CDY data of the 17th week

Figures 15 and 16 shows a correlation graph based on Cluster 1 and Cluster 2 of CDY for week 21. Unlike most graphs in other weeks, strong positive correlation can be found in both Figs. 15 and 16. When comparing with week 17, Fig. 15 shows a similar pattern to Fig. 13. But Fig. 16 shows more dense correlation pattern than Fig. 14 without any empty space. As a result, we can see it's highly likely to visit steadily in week 21. We can conclude that CDY is more likely to visit rather than not visit.

In most of the figures, it is notable that the similar amount of time interval for both highly positive correlations and highly negative correlations. It can be concluded that the negative correlation and the positive correlation shows similar pattern. However, in each of the figures, it is highly dense where the correlation

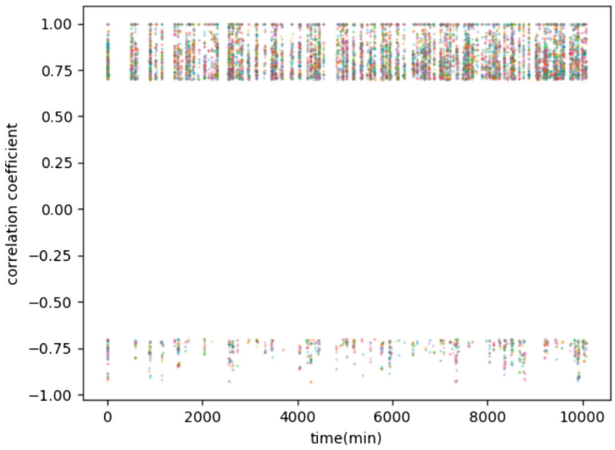


Fig. 15. Correlation graph for Cluster 1 for CDY data of the 21st week

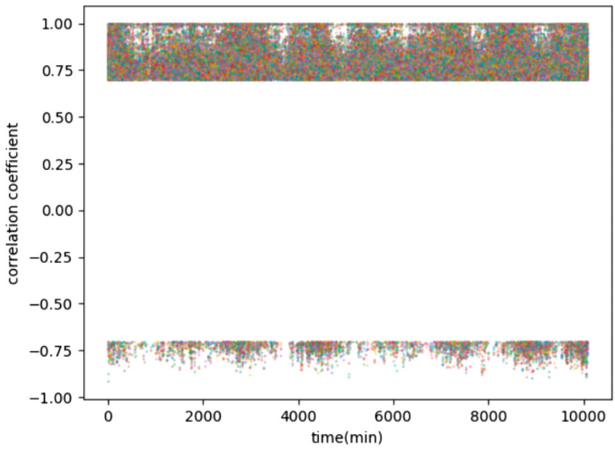


Fig. 16. Correlation graph for Cluster 2 for CDY data of the 21st week

is larger than 0.7 and less than or equal to 1.0, but values close to −1.0 is really sparse, where correlation values are less than −0.7 and greater than −1.0. It can be concluded that positive (+) correlation appears more frequently than negative (−) correlation. Then, it can be judged that the time with positive correlation usually exists strongly and more meaningful than negative correlation. This trend became clear and clear as we proceeded our analysis with wider range of time domain.

Table 2 shows the top 25 results with strong correlation coefficients. Cluster 1 and Cluster 2 are compared with each other and the top 25 are represented in the table. The column cluster 1 and cluster 2 shows the number of visits to the corresponding cluster. Each column has intervisit time τ in the next column. In Cluster 1, the intervisit time τ is in the top 25 from 1 min to 21 min and

Table 2. Top 25 categories with strong correlation coefficients

Rank	Cluster 1	τ (min)	Cluster 2	τ (min)
1	10,332	1	2,196	1
2	3,724	2	961	1,719
3	2,060	3	830	2
4	1,077	4	635	3,301
5	943	9	634	1,328
6	918	11	573	3
7	903	10	477	25
8	879	5	477	15
9	874	8	473	40
10	832	7	466	4
11	824	6	463	47
12	771	12	461	65
13	747	14	460	43
14	736	13	444	63
15	698	15	438	86
16	601	16	434	337
17	578	17	434	45
18	566	20	430	49
19	538	19	428	39
20	530	18	424	58
21	518	10,070	423	117
22	499	21	423	24
23	488	10,078	415	70
24	486	10,077	411	37
25	486	10,069	410	67

includes 10,069 min, 10,070 min, 10,077 min, and 10,078 min. For these 25 cases, the volunteer is more likely to stay after a visit. For Cluster 2, we can see that the intervisit time τ is bigger than that of Cluster 1. For example, Cluster 2 can be interpreted as having many visits after 1,719 min, and having visited again after 3,301 min. As a result, Cluster 2 has clearer revisit times than Cluster 1.

5 Conclusion

In this paper, a temporal analysis of mobility data is presented to determine what form the mobility pattern is in the object mobility model, especially cyclic patterns latent in mobility model. Raw geopositioning data sets were collected

by a volunteer. The concept of time slice window is incorporated. The autocorrelation and autocovariance were calculated by time slice over the collected geopositioning data. The figures showing the graph of the correlation coefficient over time were presented as correlation coefficient between 0.7 and 1.0 which is strong positive, and as correlation coefficient between −0.7 and −1.0 which is strong negative. A strong positive correlation coefficient was found and a strong negative correlation coefficient was simultaneously found as shown in the figures, and the intervisit time can be identified representing cyclic mobility patterns. Through the analysis, we found out the frequency of cyclic revisit for each notable location cluster and also identified the intervisit time to location clusters with high correlations. This sort of cyclic pattern analysis must be useful tool to design various Location Based Service (LBS). For the future research, we need to investigate more persons' mobility data and geolocation data collected by other objects such as Internet of Things (IoT) devices with the same analysis method. There's a other location data collecting app named Swarm [11], which is developed by Foursquare group [12]. This could be used in future research too. Even though we guess the personal mobility pattern shows similar cyclic pattern, however, it may differ from person to person. For the mobility of IoT objects, no results were reported while it must be great interest of related researchers.

References

1. Pheasant, D., Larue, M.: Active sports tracker and method. U.S. Patent Application No 11/609,634 (2006)
2. Garmin. Garmin gpsmap 62s (2015). https://buy.garmin.com/en-US/US/on-the-trail/discontinued/gpsmap-62s/prod63801.html
3. Kim, H., Song, H.Y.: Daily life mobility of a student: from position data to human mobility model through expectation maximization clustering. In: Kim, T., et al. (eds.) MulGraB 2011. CCIS, vol. 263, pp. 88–97. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-27186-1_11
4. Song, H.Y.: Probabilistic space-time analysis of human mobility patterns. WSEAS Trans. Comput. **15**, 222–238 (2016)
5. Hartley, H.: Maximum likelihood estimation from incomplete data. Biometrics **14**, 174–194 (1958)
6. Rubin, D.B., Dempster, A.P., Laird, N.M.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. **39**(01), 1–38 (1977)
7. Song, H.Y., Kim, D.: A method for group mobility model construction. In: Multimedia, Computer Graphics and Broadcasting, Korean Institute of Information Scientists and Engineers (2015)
8. Robusto, C.: The cosine-haversine formula. Am. Math. Mon. **64**(1), 38–40 (1957)
9. Gaito, S., Zignani, M.: Extracting human mobility patterns from GPS-based traces. In: IFIP Wireless Days (2010)
10. Song, X., Wen, Q., Gao, J.: RobustSTL: a robust seasonal-trend decomposition algorithm for long time series. In: IFIP Wireless Days (2018)
11. Sahin, E.: Swarm robotics: from sources of inspiration to domains of application. In: International Workshop on Swarm Robotics (2004)
12. Wiese, J., Lindqvist, J., Cranshaw, J.: I'm the mayor of my house: examining why people use foursquare-a social driven (2012)