

APP: Aperiodic and Periodic Model for Long-Term Human Mobility Prediction Using Ambient Simple Sensors

Danaipat Sodkomkham, Roberto Legaspi, Ken-ichi Fukui,
Koichi Moriyama, Satoshi Kurihara, and Masayuki Numao

Institute of Scientific and Industrial Research, Osaka University, Japan,
danaipat@ai.sanken.osaka-u.ac.jp

Abstract. The predictive technique proposed in this project was initially designed for the indoor smart environment wherein intrusive tracking techniques, such as cameras, mobile phone and GPS tracking system, could not be utilized appropriately. Instead, we installed simple motion detection sensors in the experimental space and observed occurred movements at each area. However, the movement data recorded with this setting cannot provide as much information about human mobility as the data from the GPS or mobile phone is capable of. In this paper, we conducted an exhaustive analysis on this specific dataset to determine the predictability of future users' mobility using only this limited dataset and regardless of the predictive technique. Furthermore, we also proposed the predictive technique, named APP, for long-term human mobility prediction that works well on our limited dataset. Finally, evaluation on the real dataset collected inside the smart space over 3 months of movements and daily activities data shows that our model is able to predict future mobility and activities of participants in the smart environment setting with high accuracy even for a month ahead.

Keywords: Human mobility, Smart environment, Long-term prediction

1 Introduction

Understanding and predicting human mobility are crucial components in a number of real world applications. We will mention a few examples here. The PUCK architecture[4] was introduced to intelligently provide reminders in the smart environment since it automatically recognizes habitual activities and then reminds the occupant if he/she forgot some important tasks, such as forgetting to take a medicine after a meal. This could be helpful for participants who have dementia or mind cognitive impairment. Moreover, the ability to predict future locations of people is also an important element in transportation planning [13, 11], bandwidth provisioning in wireless local area network [18], and targeted advertisement dissemination [7].

In our specific example, we have an actual office environment built-in with various sensors and actuators to enable the pervasive computing technology to

control different settings of the environment. Our prototype smart office environment was initially designed to create a working environment that can learn users' behavioral activities and react to these activities smartly. Goals of our developing smart environment is to simplify mundane repetitive tasks, and to make the participant live more comfortable. For example, the smart office that predicts future occupancy of the meeting room and automatically gets electronic facilities in the room prepared right before the meeting. The smart office that predicts participants' needs from their daily activities so that it always has hot coffee ready to be served at the time they need. All of the applications mentioned above require the ability to foresee user's future whereabouts and mobility into far future, as known as the open problem of long-term human mobility prediction.

A smart environment, normally, has sensors installed to sense activities and mobilities of participants inside. Different machine learning algorithms are then employed to explore meaningful information about user behaviors and routines. These information will be later used to build a predictor that foresees users' needs and suggest a proper reaction to them. Therefore, one challenging problem for every anybody who works on smart environments researches would be "*to determine the best approach to observing participants' mobility with the least obtrusiveness while providing enough information to build an accurate predictive model*".

There is apparent trade-off between informativeness and conspicuousness of the sensing technique. For a concrete example, using colored pictures from cameras with a help from image processing technique and semi-supervised classification algorithm, Yu et al. [19] was able to create a system that recognizes people and their positions. Moreover, they were able to map each individual's movement directly into a floor map. From this interesting example, rich mobility information for individual users must be traded with users' discomfort because of surrounding cameras. Apart from camera techniques [1, 3] discussed earlier, mobile phone data [6, 17], GPS [16, 13], and RFID tagging [2, 10] requires users to carry (or put on) the tracking device while inside the environment, which is not feasible in real-world implementation. On the other hand, simple sensors such as infrared distance sensor, ultrasonic distance sensor, and magnetic sensor, are small enough to blend into the environment, and seamlessly observe human mobility inside the environment. It is the case, however, that simple sensors' data is less informative than such high precision sensing technologies and they limit the capability of the mobility predictive model that was built using their data.

Therefore, in this paper, we investigated limits of the predictability over this specific type of mobility dataset. Furthermore, in the latter part of this paper, we present a novel prediction technique, named APP, particularly for the long-term human mobility prediction problem. More specifically, the APP predicts future location of a user at any specific *time frame* in far future, e.g. 21 days from now between 10:00 and 10:05. The prediction is obtained by probabilistic models that compute how likely a certain location will be revisited in the future at the specific

time frame. The prediction part in APP consists of two probabilistic models. Both probabilistic models keep track of visited time stamps, extracts contextual features from each visit (such as what time of the day, what day of the week, or how long after the last visit), and model their relations. The first predictive model, named *Periodic model*, is based on a hypothesis that user keeps visiting some specific locations periodically, such as every three hours, everyday, or every month. Firstly, we analyze periodicity at each location and test this hypothesis. If the hypothesis is accepted, we use the periodic model to predict; otherwise, the second model is used. The second predictive model does not rely on periodicity property in human mobility behavior; instead, it extracts significant patterns of repetitive movements representing user mobility behavior in the past. Hence, it is called the *Aperiodic model*. The aperiodic model postulates that the same (or similar) mobility pattern tends to repeat again at any specific time frame in the future whenever their contextual features are similar. Combining these two models (APeriodic and Periodic, abbreviated to APP) results in a predictor that predicts user's location with acceptably high accuracy and precision, even for a month ahead.

2 Limits of Predictability in Collective Human Mobility

The breakthrough analysis of predictability of human mobility has been studied in [17]. Song et al. explored the limitation in predictability of individual's movements, disregarding quality of the prediction techniques. Despite the difference in user's daily behavior, their analysis over a large population monitored by their mobile phone data shows 93% potential predictability in individual's mobility. In other words, predicting individual's movements can be achieved effectively when history data of individual's movements is available.

When it comes to the situation when historical data of individual's user is not provided, but collective mobility data from multi-users, the fundamental question of predictability on this class of data arises again, i.e. *to what degree is collective mobility predictable?*

2.1 Collective Human Mobility Data

For our experimental smart environment, we have a functioning working space, which includes individuals' cubicle work stations, recreational space and meeting areas, where a total number of 20 graduate students and faculty staff members come to work regularly. Each individual may have different duties, different class schedules and different daily routine, which result in different mobility patterns directionally and temporally. We installed different types of sensors (see figure 5(a)) in the environment to monitor activities and movements that happened inside. Specifically, we used two types of sensors in the experiment. First, infrared distance sensors were mainly used to detect movement at each specific location. Second, magnetic sensors were attached to the hinge of the refrigerator and the oven to observe their usages. All of these sensors are connected

through the lab’s network and continuously feed live streams of mobility data to a database. By employing these ambient sensors, participants needed not to be equipped with a tracking device during the experiment period and can normally move along without any concern of being monitored. We observed visitations and mobility inside the experimental environment over 24 hours a day, for 3 months (precisely 92 days) during the autumn semester. The floor plan in figure 5(b) shows placements of sensors used to capture visitations at each location in the space. The number of interested locations N is 30. We modeled human mobility with two representations for different purposes as follows.

As temporal sequences of repeated visitations. Collective mobility, at each location, is represented with the temporal sequence of repetitive visitations visited by unknown users during the observing period. State of visitation at a moment is denoted by a binary value: either 1 for *visited*, or 0 for *not visited*. For instance, a sequence v_x represents mobility at location x from 00:00 to 23:59, with the sample rate μ of 1 sample per hour.

$$v_x = ((t'_0, 0), (t'_1, 1), (t'_2, 0), \dots, (t'_{23}, 1))$$

when t'_i represents the observed *time frame* from $t_0 + i\mu$ to $t_0 + (i+1)\mu$, and t_0 is the starting time, i.e., $t_0 = 00:00$ and $t'_0 = [00 : 00, 01 : 00]$.

As trajectories. By increasing sample rate of the sensors μ up to 1 sample per 200 milliseconds, we were able to count every visitations. Then, from a temporal sequence of visitations, $((x_0, t_0), (x_1, t_1), \dots, (x_{w-1}, t_{w-1}))$, we linearly searched for each transition point in the sequence where the transition time $t_{i+1} - t_i > 30$ seconds to cut it into smaller sequences that represent trajectories.

Despite unobtrusiveness and simplicity of the ambient sensing method, the obtained data is primarily noisy. To handle noises (such as false triggered events, sensors blocked by obstacles, and simultaneous trajectories from different users) and extract movement trajectories from the collective mobility dataset efficiently, we applied the data mining algorithm, called PrefixSpan [9], to extract only sub-trajectories of length- n that appeared in T more frequently than a certain minimum number of times $support_{min}$ during the experiment.

2.2 Limits of Predictability

Here we evaluated the predictability over the collective mobility dataset using the same methodologies introduced by Song et al. in [17]. Namely, by employing Fano’s inequality [5, 14], we estimated the upper limit of the probability of the destination of a moving user can be predicted correctly given the most recent trajectory and the past collective mobility data.

Let T'_i denote a movement trajectory and let D_i be a destination of T'_i from the observations, $T = ((T'_0, D_0), (T'_1, D_1), \dots, (T'_m, D_m))$. Given a predictive technique $f(T'_i)$ that works well in predicting future location D_i of a moving

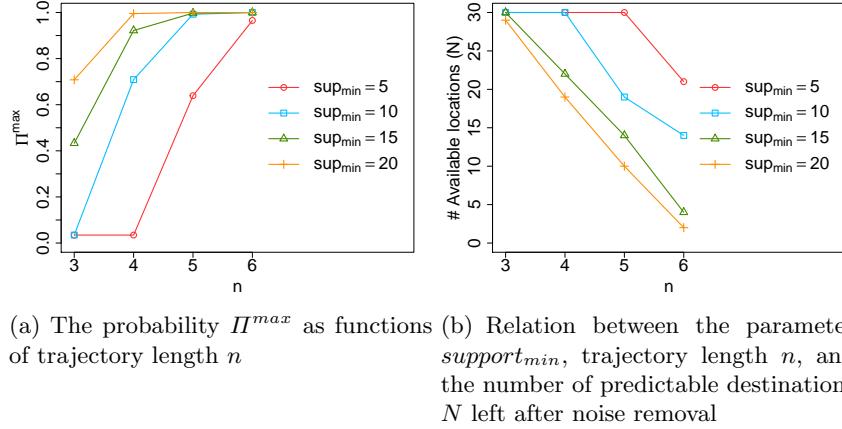


Fig. 1. The predictability of the collective human mobility in the smart environment. The Π^{max} is the upper bound of the probability that a particular predictive algorithm is able to predict user's location correctly using only the collective dataset.

user based on recent length- n movement trajectory T' and a set of length- n trajectories T from historical mobility data, let e denote the event of failed prediction, i.e. $f(T') \neq D$, and let $P(e)$ be its probability. According to Fano's inequality, the lower bound on the error probability $P(e)$ can be found in the following inequality.

$$H(D|T') \leq H(e) + P(e) \log(N - 1)$$

Thus, the probability of predicting correctly, denoted by Π , is $1 - P(e)$. Namely,

$$H(D|T') \leq H(e) + (1 - \Pi) \log(N - 1), \quad (1)$$

where the destination D can take up to N possible locations and $H(e)$ is the corresponding binary entropy which is defined as follow.

$$\begin{aligned} H(e) &= -P(e) \log(P(e)) - (1 - P(e)) \log(1 - P(e)) \\ &= -(1 - \Pi) \log(1 - \Pi) - \Pi \log(\Pi) \end{aligned} \quad (2)$$

The conditional entropy $H(D|T')$ appeared in the inequality (1) quantifies the amount of information needed to predict the destination D given the correlated recent trajectory T' . Given the probability $P(T')$ of the set of past trajectories T containing T' and the joint probability $P(T', d)$, the conditional entropy $H(D|T')$ is defined as follow.

$$H(D|T') = \sum_{d \in D, T' \in T} P(T', d) \log \left(\frac{P(T')}{P(T', d)} \right) \quad (3)$$

Then we calculated the entropy $H(D|T')$ individually for each length n of trajectories in T , and analyzed the maximum potential predictability (denoted

by Π^{max}) or the probability of predicting correctly destination of a user given the collective mobility dataset by solving for the Π^{max} , where $\Pi \leq \Pi^{max}$ in the following equation, according to (1), (2) and (3).

$$H(D|T') = -(1 - \Pi^{max}) \log(1 - \Pi^{max}) - \Pi^{max} \log(\Pi^{max}) + (1 - \Pi^{max}) \log(N - 1)$$

Figure 1(a) shows Π^{max} as functions of n , where n denotes length of the considering trajectories. It is unsurprising that n increases the predictability since longer trajectory gives more evidences to the predictor that help narrowing down search space of the most probable locations. The $support_{min}$ also shows its potential to eliminate unusual trajectories in the dataset, and gives significantly higher potential predictability. However, there is a trade-off between the degree of predictability and the number of predictable locations, as Figure 1(b) shows. High threshold of the minimum support ($support_{min}$) results less number of locations N available to the predictive algorithm to predict.

To summarize from the analysis, despite the fact that the collective human mobility contains cumulative movements and behaviors from different users and seems diverge to the experimenter in the first place, the accurate prediction of user's location is achievable with acceptably high probability. However, this analysis does not provide any clue about the potential predictability in the long-term prediction configuration when the inference of user's mobility cannot rely on recent movement patterns and frequent historical trajectories. Moreover, a number of researches [16, 15] have shown that predictive techniques that work well in the short-term human mobility prediction cannot be extended to the long-term prediction effectively. Thus, in the next section, we studied the possibility to employ the periodicity in human behavior to foresee their mobility in far future instead of directly modeling trajectories.

3 Periodicity in Collective Human Mobility

It can be seen easily even without a guide from data mining tool that most of human activities are periodic to some degree. If a certain action, or movement pattern is repeated regularly with a particular interval τ , and if this behavior is consistent over time, it is certainly predictable with the time period τ . In addition, the probability in predicting the correct location of a user in the future depends on the tendency of such mobility patterns recurring at intervals. Thus we define the *periodicity probability* to formally quantify this property.

Definition 1. Let $P_x(\tau)$ denotes the *periodicity*, which is the probability of a particular event x reoccurring regularly with the constant time interval τ , where τ is a positive integer. Given the temporal sequence, as described in section 2.1, of events from t'_0 to t'_m in which the location x was visited, the periodicity probability $P_x(\tau)$ is defined as follow.

$$P_x(\tau) = P(v_x(t'_{i+\tau}) = 1 | v_x(t'_i) = 1), \quad t'_i \in \{t'_0, t'_1, \dots, t'_{m-1}\} \quad (4)$$

where $v_x(t'_i)$ indicates state of the visitation at x during the time frame t'_i .

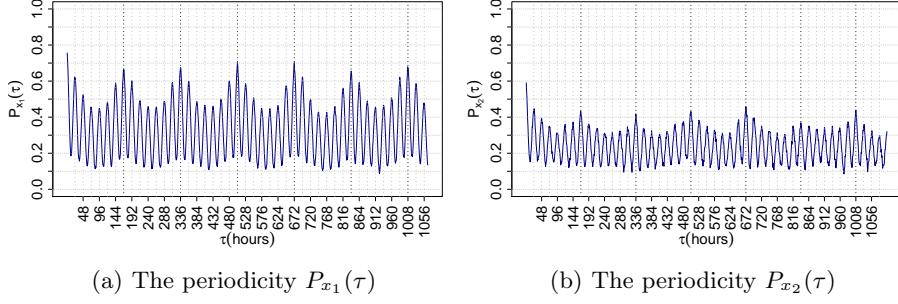


Fig. 2. The periodicity $P_{x_1}(\tau)$ and $P_{x_2}(\tau)$ of repetitive visitations at the location x_1 and x_2 as a function of time period τ .

To find significant periodicity in the collective human mobility, we searched for τ that maximizes the periodicity probability at each location separately. Figure 2 shows two sample locations where periodic behavior can be observed. The small peaks in these plots reveal relatively high probability that these particular locations were visited regularly with the time period τ , when τ are multiples of 24 hours. Moreover, the maximum of predictability probabilities are found at multiples of 168 hours. Undoubtedly, this indicates firm evidences of *daily* and *weekly* behaviors exist in the collective mobility data. With more algorithmic way of finding significant period τ , the Fourier analysis also suggested that $\tau = 24$ hours and 168 hours correspond to two most significant frequencies of $\approx 4.167 \times 10^{-2}$ Hz and $\approx 5.925 \times 10^{-2}$ Hz respectively.

In the next section, we analyze the possibility of the collective human mobility being predictable with the periodic behavioral patterns.

3.1 Predictability of The Periodic Model

Intuitively, the periodicity $P_x(\tau)$ already estimated the precision of a periodic-based predictive model, which is based on a strong assumption of periodically repeated visitations. Hence, the periodicity $P_x(\tau)$ can be considered as a measurement for the predictability of the periodic model. In addition, we also want to provide another predictability analysis employing an academic concept in information theory to the periodic model.

Firstly, we assign the *periodic entropy* to the history data of repetitive visitations at each location to determine the amount of information needed to foresee future visit given records of repetitive visitations in history. At each location x , the periodic entropy is computed as follows.

Definition 2. Given the collective mobility data, the entropy S_x^τ which quantifies the degree of uncertainty of the periodicity $P_x(\tau)$ in the dataset is

$$S_x^\tau = \sum_{\nu \in \{0,1\}} P(v_x) H(v_x(t'_{i+\tau} | v_x(t'_i) = \nu)), \quad t'_i \in \{t'_0, \dots, t'_{m-1}\}, \quad (5)$$

where $P(v_x)$ is the probability of a location x being visited, and the conditional entropy $H(v_x(t'_{i+\tau})|v_x(t'_i) = \nu)$ is

$$H(v_x(t'_{i+\tau})|v_x(t'_i) = \nu) = \sum_{\varphi \in \{0,1\}} P(\varphi|\nu) \log \left(\frac{1}{P(\varphi|\nu)} \right), \quad (6)$$

where $P(\varphi|\nu)$ stands for $P(v_x(t'_{i+\tau}) = \varphi | v_x(t'_i) = \nu)$.

Additionally, let S_{xf}^τ be the entropy of future visitations; namely,

$$S_{xf}^\tau = - \sum_{\varphi \in \{0,1\}} P(v_x(t'_{i+\tau}) = \varphi) \log(P(v_x(t'_{i+\tau}) = \varphi)), \quad t'_i \in \{t'_0, \dots, t'_{m-1}\} \quad (7)$$

Next, we determine the predictability for each location x of the periodic model with the probability $\Pi_{x,\tau}$, which is defined as follow.

Definition 3. Let $\Pi_{x,\tau}$ be the probability that the periodic model predicts times of future visitations at x correctly by always predicting visited at all moments that are $k\tau$ apart from the last visit, for $k = 1, 2, \dots$. Thus the associated entropy $H(\Pi_{x,\tau})$ of the predictability $\Pi_{x,\tau}$ is

$$H(\Pi_{x,\tau}) = -\Pi_{x,\tau} \log_2(\Pi_{x,\tau}) - (1 - \Pi_{x,\tau}) \log_2(1 - \Pi_{x,\tau}). \quad (8)$$

The maximum predictability $\Pi_{x,\tau}^{max}$ can be determined using the Fano's inequality in accordance with (1).

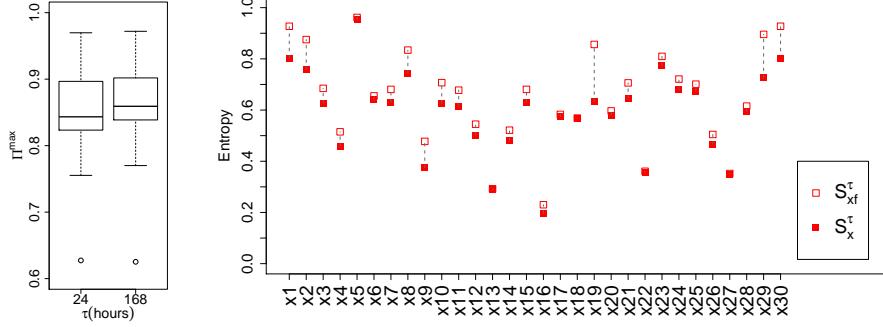
$$S_x^\tau \leq H(\Pi_{x,\tau}) + (1 - \Pi_{x,\tau}) \log_2(N - 1) \quad (9)$$

Because $\Pi_{x,\tau} \leq \Pi_{x,\tau}^{max}$ and $N = 2$ prevents this bound to the binary classification, then the following correction is required.

$$\begin{aligned} S_x^\tau &\leq H(\Pi_{x,\tau}) + (1 - \Pi_{x,\tau}) \log_2(N - 1) \leq H(\Pi_{x,\tau}) + (1 - \Pi_{x,\tau}) \log_2(N) \\ &= -\Pi_{x,\tau}^{max} \log_2(\Pi_{x,\tau}^{max}) - (1 - \Pi_{x,\tau}^{max}) \log_2(1 - \Pi_{x,\tau}^{max}) + (1 - \Pi_{x,\tau}) \log_2(N) \end{aligned} \quad (10)$$

After solving for $\Pi_{x,\tau}^{max}$ in (10), the predictability $\Pi_{x,\tau}^{max}$ determines the upper limit of the probability of predicting future visits of users at location x in far future given an appropriate periodic model (with the time period τ). We evaluated S_x^τ and Π_τ^{max} separately for each location, and the associated distribution of Π_τ^{max} is shown in figure 3(a). Both distributions of the predictability Π_τ^{max} indicate the average predictability over all locations approximately above 80%, in both daily and weekly model. The average predictability of the weekly model is slightly higher and has lower variance than the daily model. One may conclude from the result that the weekly model fit the collective mobility data better than the daily periodic model.

Figure 3(b) shows differences between the periodic entropy $S_x^{\tau=24}$ and the entropy of future visitations $S_{xf}^{\tau=24}$ at each location x . Note that as S_x^τ closer to zero



(a) The pre-
(b) Differences between the periodic entropy and the entropy of
dictability Π_{τ}^{max} future visits at each location.
of periodic model.

Fig. 3. The predictability Π_{τ}^{max} and its corresponding periodic entropy.

and further from S_{xf}^{τ} , future visitations are more likely to depend on previous visitations periodically. Result from this figure clearly suggests that not visiting behavior at every location in our environment is periodic. For instance, the locations $x_7, x_{14}, x_{17}, x_{23}$, and x_{28} were not periodic, while the locations x_1, x_2 , and x_{20} appeared to be more periodic than others. Therefore, the periodicity-based predictive model alone would not work in every location; hence, we have developed the integrated aperiodic and periodic model for long-term human mobility prediction.

4 APP: Aperiodic and Periodic model for long-term human mobility prediction

Our long-term human mobility predictive model combines two predicting paradigms together. The first approach (Periodic approach) employs the periodic property in human mobility to foresee future visits. On the other hand, the second approach (Aperiodic approach) does not rely on the periodicity; instead, it presumes that mobility patterns are similar to the day in the past that has similar features. The APP uses either one of the two approaches to predict human mobility at a certain location x depending on the periodicity probability $P_x(\tau)$ at that specific location x . If $P_x(\tau)$ is more than the user-specific threshold P_{min}^{τ} , then the APP uses the periodic approach. Otherwise, it switches to the aperiodic approach.

4.1 APP: The Periodic Approach

The APP with the periodic predictive approach was designed to foresee times of future visitations at each location in the smart space. To predict future locations

of multi-users, the predictions are computed independently for each location, then all the results are combined together providing a set of locations that are likely to be visited at the specific time in far future.

The fundamental idea behind the prediction is based on the assumption of periodicity. Say, if the visitations at x recur regularly, again and again, with a constant time interval τ , and if this periodic behavior appears consistently over time, then the probability $P_x^\tau(t'_f)$ that the future visitation will occur within the time frame t'_f in the future, when the last visit happened at t'_{m-1} , can be computed as

$$P_x^\tau(t') = P(v_x(t'_{m-1-(k)\tau+\delta}) = 1 | v_x(t'_{m-1-(k+1)\tau+\delta}) = 1), \quad k = 1, 2, \dots, [m/\tau]$$

where $\delta = (f - m + 1) \bmod \tau$.

This simple, yet accurate, predictive approach works well only at certain locations, where users' mobility has apparent periodicity. Otherwise, the periodic approach gets poor prediction because the mobility in those particular locations are not governed by the periodic behavior. To address this problem, we proposed the optional predictive approach contributed to the APP that is independent of the periodic behavior.

4.2 APP: The Aperiodic Approach

In this second predictive technique implemented in the APP, we extract significant patterns of repetitive visitations at each location that happened on different days. Next, the days that have similar visiting pattern are clustered together, resulting groups of *similar days* in the past history that users behaved similarly. Then, we extract contextual features from each group of similar days. In this project, the interested features consist of: (1) what day of week, (2) whether it is a holiday or not. Note that unlimited additional features that might relate to the mobility pattern can be used to characterize the day more comprehensively, such as temperature, traffic, weather condition, or meeting schedule. However, due to the limit in the dataset we have, only these two features are applied.

The intuition that supports this predictive approach is derived from the weekly model in section 4.1, human mobility patterns on the same *day of the week* are likely similar. Additionally, human activities on national holidays are apparently different from normal workdays; so, we need an additional bit to explicitly specify this property. Hence, mobility pattern of *a day* can be modeled individually by the visitations at each location. Recall the temporal sequence v_x in section 2.1, mobility at a certain location x can be represented with a vector:

$$\begin{aligned} d_x &= [v_x, day_{week}, holiday] \\ &= [v_{t'_0}, \dots, v_{t'_{23}}, Sun, Mon, \dots, Sat, Hol] \end{aligned}$$

The day vector d_x consists of 32 bits. The first 24 bits model visitations at location x during a specific time frame of a day, which is divided hourly. The next 7 bits indicate day of week, and the last bit indicates a holiday.

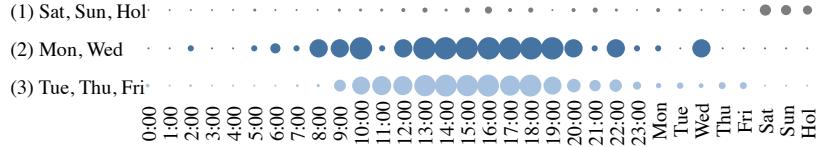


Fig. 4. Three cluster centroids that represent three mobility patterns.

Similarity between two day vectors is basically measured by the Hamming distance [8]. Then the k -means clustering algorithm [12] is applied to a set of day vectors to find clusters of similar days. The parameter k of the algorithm directly implies to the number of different mobility patterns that happened on different days. The centroid of each cluster now represents common mobility pattern that provides predictions (probability) of visitations to be found on any certain days in the future that have similar features. A concrete example of the similar day clusters discovered from the real dataset is shown in figure 4. It is striking that the cluster centroids in figure 4 clearly show 3 different visiting patterns at that particular location. The cluster (1) contains a set of days in the past history when the visitations rarely happened, and the majority of this set are Saturdays, Sundays and the days specified as holiday. On the other hand, the cluster (2) and (3) contain more active days. The days in the cluster (2), which most of them are Monday and Wednesday, have very low visitations records during 11:00-12:00 and 21:00-22:00; moreover, the visitations seems to occur earlier than on the days in the cluster (3). Interestingly, this follows the fact that we have meetings arranged in the experimental space every Monday and Wednesday, and causes the mobility pattern to appear differently to other days.

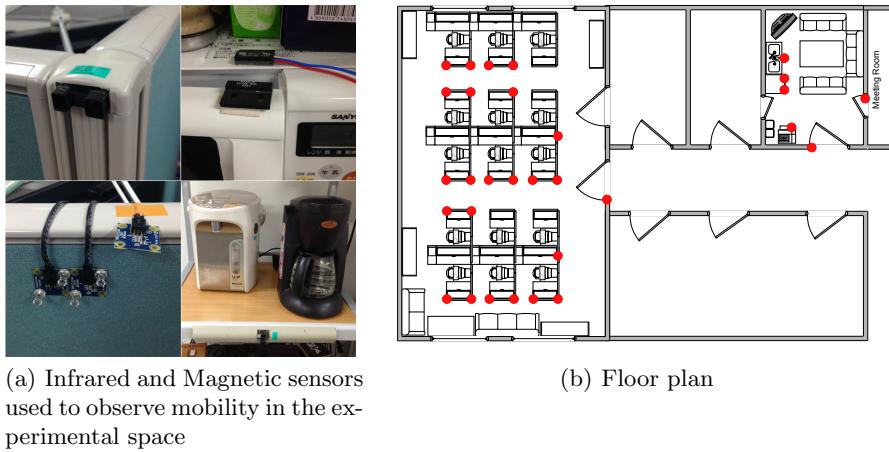
5 Evaluating Prediction Performance

In this section, we evaluated prediction performance of our proposed long-term human mobility predictor on a physical dataset of collective human mobility inside the working environment. As described in section 2.1, the dataset contains 92 days of mutual movements from every participants in the space. Data are collected consecutively 24 hours a day, 7 days a week from ~ 20 users using infrared sensors, and magnetic sensors (figure 5(a)). These sensors were installed at 30 locations over the experimental space to detect activities and mobility at each area. Outline of the space and installed sensors are shown in figure 5(b). Movements and activities committed in the experiment were not scripted beforehand; all actions happened deliberately regarding each individual's routine, work schedule, and needs at that moment.

Firstly, we evaluated the periodic approach for long-term human mobility prediction. Two months of collective mobility data was used to build the predictive model and the remaining 30 days of mobility data was used to test the model. Details of the dataset are summarized in table 1.

Table 1. Human Mobility Dataset

Dataset	Train	Test
Sample rate	hourly	
Number of participants	≤ 20	
Observed locations	30	
Size of data	62 days (1,488 hours)	30 days (720 hours)

**Fig. 5.** (a) Infrared and Magnetic sensors used in the experiment. (b) Placement of the sensors.

5.1 Periodicity and Prediction Performance

We determined relations between the periodicity probability ($P_x(\tau = 24)$ and $P_x(\tau = 168)$) and the prediction accuracy, precision, and recall rate at each location separately. Figure 6(a) and 6(d) exhibit decreasing trend of prediction accuracy when the periodicity probability increased; yet, the periodic predictor gets higher precision and recall rates as the dataset has higher probability of such movements being repeated periodically. Nevertheless, the measurement of prediction accuracy is meaningless to us because the datasets, which contain visitations records at each location in past history, have negative skew. In other words, naive predictor can achieve at least 60% chance of predicting visitations (either “visited” or “not visited”) of users at a specific time frame in the future correctly by always guessing “not visited”. Figure 6(b) and 6(e) show direct relationship between the precision rate and the periodicity probability. Likewise, the recall rates in figure 6(c) and 6(f) show that the datasets with higher periodicity are more predictable than the others. Moreover, when the periodicity probabilities are lower than 0.4, the daily periodic approach (see figure 6(c)) clearly gets poor results. These confirm our hypothesis that the periodic approach alone is not effective in predicting with low periodicity probability.

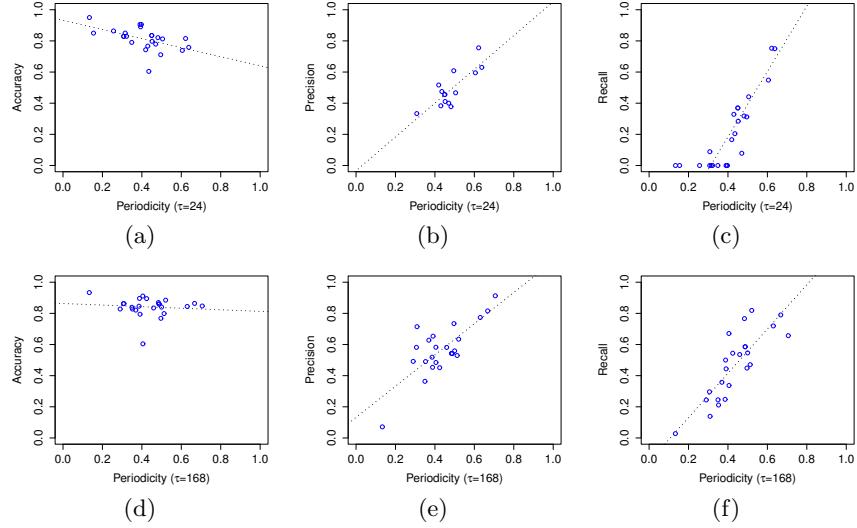


Fig. 6. Periodicity and Prediction Performance.

5.2 Prediction Performance of the Similar-day Approach

The aperiodic part in the proposed predictive technique, *APP*, is implemented with the similar-day predictive approach described in section 4.2. In the previous experiment, the periodic approach underperformed on the datasets where the mobility were not really periodic. Especially in the daily periodic model (see figure 6(a), 6(b), and 6(c)) when most of locations in the experimental space have lower periodicity probability than 0.4. Hence, in this experiment, the aperiodic part of the APP is activated where the periodicity is lower than the minimum threshold $P_{min}^{\tau} = 0.4$, the experimenter-specified threshold.

Figure 7 reveals benefit of implementing the aperiodic part into the APP predictive model. In figure 7(a), the precision rates of the APP, after implementing the similar-day approach in low-periodicity data, are improved significantly, comparing with the periodic approach alone. The precision plots of the periodic approach on the left (periodicity ≤ 0.4) were mostly omitted because the periodic predictor never predicted “*visited*” on those locations, resulting undefined precision rates.

The APP also improves the recall rates, as shown in figure 7(b). It is striking that the recall rates used to get nearly 0.0 in the periodic approach rise up to 0.6 when predicted with the APP predictive technique.

In summary, the aperiodic part in our proposed long-term human mobility predictive technique helps improving the prediction performance especially when the periodicity probability is too low to infer future visitations. However, the similar-day approach that we implemented into the aperiodic part is not effective enough to improve the predictive technique that employs the weekly periodic approach ($\tau = 168$). The reason is that the *day of week* attributes used in the

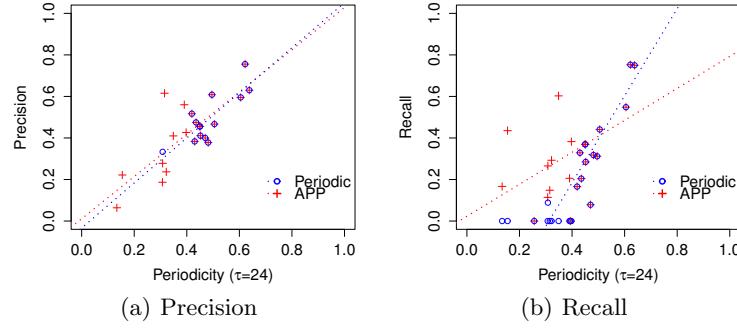


Fig. 7. Prediction Performance of the Similar-day Approach.

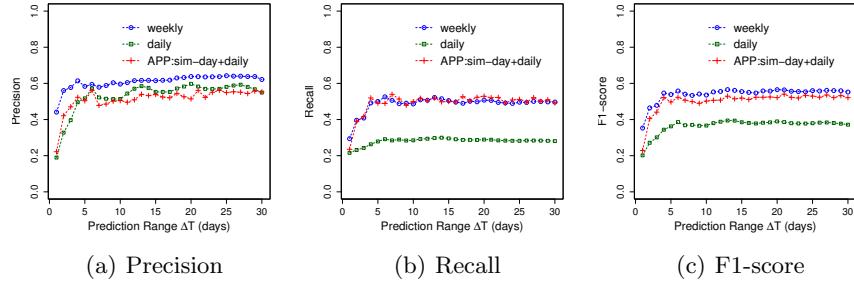


Fig. 8. Prediction Performance on Long-term Prediction.

cluster analysis already corresponded to the weekly basis, and the the *holiday* is not really significant feature since there were few holidays during only 3 months of dataset. So, the implementation of the similar-day approach (aperiodic part) and the weekly periodic approach was not able to achieve much improvement comparing with the weekly periodic predictive approach alone.

5.3 Performance in Long-term Prediction

In this section, we tested the robustness of the APP, Aperiodic and Periodic approach for human mobility predictor, over long range of prediction. The settings of this experiment refer to table 1. The results in figure 8 show the steady prediction performance even when predicting for 30 days ahead. The F1-score, which is the harmonic mean of the precision and the recall rate (in figure 8(c)) summarizes the prediction performance of 3 proposed predictive techniques as follows. Firstly, our collective mobility dataset that seems random in the first place contains enough information to be predicted accurately even in far future. Activities and corresponding mobility in the dataset are likely to be periodic on the weekly basis; hence, the weekly periodic predictive approach alone can get the average F1-score at 0.55. On the other hand, the daily model performs

relatively poor (average F1-score at 0.37) on this dataset because of low periodicity probability on the daily basis. However, after implementing the similar-day approach together with the daily predictive model, the integrated model can achieve average F1-score at 0.52.

6 Conclusions and Future Work

In this paper, awaiting answer to the question about limitations of the predictability of the collective human mobility from simple ambient sensors inside the smart environment has been revealed. We took a challenging decision implementing non-intrusive tracking method. Our choice of tracking method with ambient simple sensors have bold advantages from its unobtrusiveness and simplicity. The limitation, however, is that they cannot distinguish people's identities, such that its data limits a predictive model because it is impossible to create a predictive model individually for each user's mobility pattern.

The predictability analysis reveals potential of building a short-term next location predictive model that predicts next movement of a moving user accurately using only the collective dataset. However, implementing the short-term predictor is outside the scope of this paper. Furthermore, we also discovered acceptably high predictability in long-term prediction by modeling periodic behaviors hidden in the collective mobility data.

Next, we proposed the APP: Aperiodic and Periodic predictive model for long-term human mobility prediction. Results from the experiment shows that performance of the APP predictor improved significantly when predicting in low periodicity situations using the aperiodic approach. However, the aperiodic approach implemented in this project is not effective enough to increase the performance of the weekly periodic predictive approach yet because very limited features can be extracted from the collective dataset. We leave this to be our future work.

References

- [1] Anton Andriyenko, Konrad Schindler, and Stefan Roth. “Discrete-continuous optimization for multi-target tracking”. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE. 2012, pp. 1926–1933.
- [2] Seung-Ho Baeg et al. “Building a smart home environment for service robots based on RFID and sensor networks”. In: *Control, Automation and Systems, 2007. ICCAS'07. International Conference on*. IEEE. 2007, pp. 1078–1082.
- [3] Csaba Beleznai, David Schreiber, and Michael Rauter. “Pedestrian detection using GPU-accelerated multiple cue computation”. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. IEEE. 2011, pp. 58–65.

- [4] B. Das et al. “Automated prompting in a smart home environment”. In: *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*. IEEE. 2010, pp. 1045–1052.
- [5] Robert M Fano. *Transmission of Information: A Statistical Theory of Communication*. 1961.
- [6] M.C. Gonzalez, C.A. Hidalgo, and A.L. Barabasi. “Understanding individual human mobility patterns”. In: *Nature* 453.7196 (2008), pp. 779–782.
- [7] H. Haddadi, P. Hui, and I. Brown. “MobiAd: private and scalable mobile advertising”. In: *Proceedings of the fifth ACM international workshop on Mobility in the evolving internet architecture*. ACM. 2010, pp. 33–38.
- [8] Richard W Hamming. “Error detecting and error correcting codes”. In: *Bell System technical journal* 29.2 (1950), pp. 147–160.
- [9] Jiawei Han et al. “PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth”. In: *ICDE’2001, April* (2001), pp. 215–24.
- [10] Sajid Hussain, Scott Schaffner, and Dyllon Moseychuck. “Applications of wireless sensor networks and RFID in a smart home environment”. In: *Communication Networks and Services Research Conference, 2009. CNSR’09. Seventh Annual*. IEEE. 2009, pp. 153–157.
- [11] J. Krumm and E. Horvitz. “Predestination: Inferring destinations from partial trajectories”. In: *UbiComp 2006: Ubiquitous Computing* (2006), pp. 243–260.
- [12] JB MacQueen. “Some methods for classification and analysis of”. In: (1967).
- [13] A. Monreale et al. “WhereNext: a location predictor on trajectory pattern mining”. In: *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2009, pp. 637–646.
- [14] Nicolas Navet and Shu-Heng Chen. “On predictability and profitability: Would gp induced trading rules be sensitive to the observed entropy of time series?” In: *Natural Computing in Computational Finance* (2008), pp. 197–210.
- [15] A. Sadilek and J. Krumm. “Far out: Predicting long-term human mobility”. In: *Twenty-Sixth AAAI Conference on Artificial Intelligence*. 2012.
- [16] S. Scellato et al. “Nextplace: A spatio-temporal prediction framework for pervasive systems”. In: *Pervasive Computing* (2011), pp. 152–169.
- [17] C. Song et al. “Limits of predictability in human mobility”. In: *Science* 327.5968 (2010), pp. 1018–1021.
- [18] L. Song et al. “Predictability of WLAN mobility and its effects on bandwidth provisioning”. In: *Proceedings of INFOCOM*. Vol. 6. 2006.
- [19] Shou-I Yu, Yi Yang, and Alexander Hauptmann. “Harry Potters Marauders Map: Localizing and Tracking Multiple Persons-of-Interest by Nonnegative Discretization”. In: <http://www.cs.cmu.edu/~iyu/files/cvpr13/cvpr13.pdf> (2013).