Predicting Human Location Using Correlated Movements

Thi-Nga Dao 1, Duc Van Le 2 and Seokhoon Yoon 1,*

1 Department of Electrical and Computer Engineering, University of Ulsan, Ulsan 44610, Korea; daothinga.mta@gmail.com
2 School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore; vdle@ntu.edu.sg
* Correspondence: seokhoonyoon@ulsan.ac.kr; Tel.: +82-52-259-1403

Received: 24 October 2018; Accepted: 2 January 2019; Published: 3 January 2019

Abstract: This paper aims at estimating the current location, or predicting the next location, of a person when the recent location sequence of that person is unknown. Inspired by the fact that the behavior of an individual is greatly related to other people, a two-phase framework is proposed, which first finds persons who have highly correlated movements with a person-of-interest, then estimates the person’s location based on the position information for selected persons. For the first phase, we propose two methods: community interaction similarity-based (CISB) and behavioral similarity-based (BSB). The CISB method finds persons who have similar encounters with other members in the entire community. In the BSB method, members are selected if they show similar behavioral patterns with a given person, even though there are no direct encounters or evident co-locations between them. For the second phase, a neural network is considered in order to develop the prediction model based on the selected members. Evaluation results show that the proposed prediction model under the BSB scheme outperforms other methods, achieving top-1 accuracy of 71.13% and 69.36% for estimations of current and next locations, respectively, with the MIT dataset and 92.31% and 92.03% in case of the Dartmouth dataset.

Keywords: mobility prediction; behavioral pattern; tempo-spatial information; cellular network trace

1. Introduction

Human mobility prediction can be used in many potential and promising applications, such as location-based recommendation services, contagious disease control, patient tracking, geographic profiling, and urban planning [1–5]. In particular, in order to prevent the outbreak of infectious diseases (e.g., malaria and flu), which usually spread due to people’s travels and interactions, estimating the location of persons who are at a high risk of infection is crucial [1,4]. In geographic profiling, in order to strengthen security, law enforcement agencies need to analyze movements of suspects and need to determine their most likely position [5].

If someone’s recent location is known using a global positioning system (GPS) or other localization techniques (e.g., Wi-Fi, cellular-based systems), it is relatively simple to estimate that person’s current or next locations. However, there are cases where someone’s recent location sequences are not available. For instance, the GPS function is required to turn off due to lack of battery power in a smartphone even though people want to share their positions. In addition, there are some cases where data is sparse [6,7] (e.g., call detail records, health record of individuals, and credit card spending data). Moreover, in some cases, it is necessary to estimate locations even when someone may not want to share that information (e.g., geographic profiling of criminals). Therefore, in this work, we consider a...
new problem which is estimating someone’s current or next location, particularly when recent location sequence information is not available.

Specifically, the mobility information of other Persons with Correlated Movements (PCM) with a given person is used to estimate the person-of-interest’s current or future location, based on the fact that the behavior of one individual is largely related to other members. In other words, the current or next location of a given person is estimated using the known position information of other specific people. Although there were several studies on the prediction of human location using social relationships [8–10], those works focused on finding direct social ties between people (e.g., friendships) using encounters or calling traces between them. However, our work, in addition to friendships, considers behavioral similarity between people, which may not be reflected by direct encounters or co-location events between them.

In this paper, we propose a two-phase human location prediction model. More specifically, in the first phase, PCMs with a given person are selected. For this phase, two novel PCM selection methods are proposed: community interaction similarity-based (CISB) and behavioral similarity-based (BSB). In the CISB method, interactions with other community members are considered as well as direct encounters between certain individuals, i.e., the strength of the social relationship between people is measured by not only their direct interactions but also their interactions with other members of society. The motivation is that if two people have a close social tie, then they may have similar patterns of meeting other members of the community. In contrast to the CISB method, which attempts to find PCMs with stronger social ties, the objective of the BSB method is to find PCMs with similar behavioral patterns. As a result, the selected PCMs under the BSB method may not have a strong social relationship with the person-of-interest. In the second phase, motivated by the fact that the movement of an individual is highly correlated with other people, the current or next location of a person is estimated using the position information of selected PCMs. A PCM-based location prediction (PLP) model is developed based on a neural network (NN) [11].

For this paper, two different datasets were used to evaluate the proposed prediction model. The first dataset provided cellular traces collected by the Massachusetts Institute of Technology (MIT) Media Lab [12], which recorded the time a mobile phone was associated with a base station. In a metropolitan area, small and short-range cells that cover distances of a few hundred meters are more popular [13]. Moreover, since people usually carry their phones all the time, the position of a base station can be used to represent a person’s approximate location [10,14–18]. The second dataset is the mobility data extracted from Wi-Fi traces collected in the Dartmouth university campus [19]. A log message was recorded whenever a mobile device associates or disassociates with an AP. Due to the relatively short range of the Wi-Fi technology, an AP can be used to reflect a person’s location [20–22].

Experiments were performed in order to validate the proposed prediction model in various situations. We draw a comparison between the designed PLP model and the baseline one, called most frequent location (MFL), in which the most visited position is predicted as the location of the person-of-interest. Two proposed PCMs selection methods are also compared with three recent existing approaches: encounter frequency-based (EFB) [18,23,24], spatial closeness (SC) [6], and spatio-temporal closeness (STC) [6]. Specifically, EFB is rooted in encounter frequency, which is one of the natural characteristics of a friendship, i.e., two members who more frequently meet each other are likely to be closer friends. In SC, the persons who have the most similar distribution of visited locations to the person-of-interest are selected, meanwhile STC chooses members with the most synchronous movement with that of the person-of-interest.

The main contributions of our work can be summarized as:

- Our work considers a new problem in which human location should be estimated in a challenging situation, i.e., the recent historical location of the person-of-interest is not available. To address this challenge, a two-phase framework with low time and space complexity is proposed. Specifically, for a given person-of-interest, persons with correlated movements (PCMs) are first selected, then
the location of the person-of-interest is estimated based on the spatio-temporal information of selected PCMs.

- Two novel PCMs selection methods are proposed: CISB and BSB. CISB selects PCMs who have the strongest social tie with the person-of-interest. Because socially close people may have similar patterns of meeting other members in the community, CISB finds persons who have similar encounters with other members.
- While the purpose of CISB is to find PCMs who have the strong social relationship with the person-of-interest, BSB aims at selecting PCMs who have similar behavior patterns with the person-of-interest, even though there may be no direct encounters or evident co-location between them. As a result, the selected PCMs under the BSB method may not have strong social tie with the person-of-interest.
- A PCM-based location prediction model, which leverages the recent spatial data of selected PCMs, is developed.
- Experimental results show that the performance of BSB substantially dominates other PCMs selection methods in terms of prediction accuracy. In addition, higher performance can be achieved when a larger number of PCMs are embedded into the model. In particular, a beneficial effect of a large number of PCMs is more clearly observed in case of BSB than other PCMs extraction methods.

The rest of this manuscript is organized as follows. Sections 2 and 3 discuss related studies of human mobility prediction and the two considered datasets, respectively. Then, the proposed framework is described briefly in Section 4. In Section 5, two methods for selecting PCMs in the first phase are presented. Section 6 explains the PCM-based location prediction model in detail. Then, in Section 7, we analyze the evaluation results of the proposed model with different PCMs selection methods. Finally, the conclusions from this work are drawn in Section 8.

2. Related Work

In this section, related studies on different aspects of human mobility are presented and compared with the proposed model.

2.1. Human Mobility Analysis

A number of studies were aimed at revealing human movement characteristics [25–27]. For example, Karamshuk et al. [27] classified the properties of human mobility into three groups: spatial, temporal, and connectivity. With regard to the spatial characteristic, they focused on geographic movement, i.e., how far a person moves and where a person goes. Flight was defined as a Euclidean distance between two consecutive spots visited by the same individual. Temporal features were also considered, e.g., pause-time indicates the time period a person stays at a specific location. Meanwhile, the connectivity property reflects the contact or encounter between two people, e.g., inter-contact time was defined as the elapsed time between two adjacent contacts for a pair of people.

In addition, a few studies considered the effect of social relationships on the human movement [8,28]. Cho et al. [8] observed that people are likely to visit a distant place where a friend stays nearby, based on the datasets of online location-based social networks and cell phone location traces. Meanwhile, short-distance travel is less affected by social ties.

Those studies mainly focused on capturing human mobility characteristics, whereas we plan to design a movement prediction model to estimate a person’s current or future location. Note that embedding the properties of human mobility will be helpful for accurately predicting human movement. Therefore, in this work, a mobility prediction model is proposed, considering human movement characteristics such as encounter frequency and social correlation.
2.2. Human Mobility Prediction

Several studies tried to predict where a person stays given the prior information on historical locations of that person [2,6,18,29–32]. Most of those studies are based on the Markov model. For instance, in order to predict the person’s position in upcoming time slots, Pang et al. [29] proposed a modified Markov model considering spatio-temporal information (i.e., sojourn time and location transition preference). The authors claimed that the modified Markov model achieves higher prediction accuracy than the original Markov model.

Alhasoun et al. [6] constructed a prediction model based on dynamic Bayesian networks with the assumption of knowing the last visited location of the person-of-interest and historical position data of ‘strangers’. Strangers were defined as members who do not necessarily have a social link to the person-of-interest. Authors proposed three methods to determine strangers: temporal closeness, spatial closeness, and spatiotemporal closeness. In case of the temporal closeness approach, the members who have the most similar pattern of communication (e.g., call, sms, and data) to the person-of-interest were chosen. Meanwhile, the spatial closeness method selected the person with the most similar distribution of visited locations to the person-of-interest. The spatiotemporal closeness considered the chi squared test value to compute the closeness between two people $p$ and $q$. Specifically, the contingency table was first constructed where the element at row $i$ and column $j$ represents the number of times persons $p$ and $q$ concurrently stay at location $i$ and $j$, respectively. Then, the chi squared test value was used to measure the degree of association between two persons on the contingency table.

In addition, Zeng et al. [32] first determined the missing data of human trajectories by using the Gibbs sampling algorithm. Then, a high-order Markov chain model was constructed to predict the most likely location of the person-of-interest. Meanwhile, Noulas et al. [2] considered two location prediction models based on linear regression and M5 model trees to address the problem of predicting the person’s next location. Mobility features such as historical visits and temporal information were fed into the prediction model. The authors concluded that combining different mobility features achieved noticeably higher performance than using a single feature approach. Unlike models in [2,6,18,29–32], we design a prediction model which does not require a prior location history of the person-of-interest.

In addition to historical visit information, a number of factors can be used to predict human mobility, such as social friendships, location preferences, and temporal information. Among these factors, the strong relationship between person movement and friends was revealed in some studies [8,33,34]. Consequently, a number of mobility prediction models were designed with the support of friendship information [8–10,18,35,36].

For instance, Cho et al. [8] decomposed human mobility into two parts: periodic and socially correlated movements. The authors demonstrated that short-distance travel is usually affected by periodic mobility, whereas friendship tends to influence long-range movement. They first developed a human mobility prediction model assuming that people’s periodic travels follow a mixed Gaussian distribution of home and workplace. Then, with consideration for social friendships, the probability of being in a location is estimated as a function of the time period during which a friend stays in that location and the distance from the person to his/her friend. Finally, a mobility probability distribution combining periodic and socially correlated movement is formulated using Bayes’ theorem.

Even though certain studies [8,9] did not require a location history for people, their models only work with datasets of geographic locations, e.g., GPS traces. Therefore, to address the limitation of [8,9], our work attempts to design a more flexible and applicable model that does not require a dataset of physical locations. Moreover, our model considers behavioral similarity between people as well as friendship. Note that behavioral similarity may not be revealed based on direct interactions or encounters between people.

In regard to datasets of symbolic locations, among the models considering social correlation features to predict person movements, a few approaches worked with datasets of non-geographic locations [10,18]. In [18], a location prediction model was considered based on two factors: periodic movements and social relationships. Specifically, a Markov-based model was constructed to capture
periodic movements while colocation frequency was used to measure the closeness between people. In order to reflect the impacts of both factors on human mobility, the location prediction model was built where a different weight was assigned to each factor. Meanwhile, Zhang et al. [10] proposed an algorithm called NextCell to predict the future locations of people. A boosting technique was used to combine two predictors that are based on periodic behaviors and social interplay. The periodic behavior predictor considers a probability distribution over locations. Meanwhile, in the social interplay predictor, the probability that two people co-locate at a given time was estimated as a function of phone call features.

Although the above studies [10,18] consider the datasets of symbolic coordinations, those models do not take into account the impact of time features on human mobility which are believed to be an important factor for human mobility prediction [2,8,29]. Meanwhile, our work aims at considering both temporal and spatial information of social friends to design the movement prediction model. Moreover, for predicting the current or future location of a person, the proposed framework also takes into account other human mobility characteristics, e.g., encounter frequency, community interactions, and behavioral similarity.

2.3. Social Community Detection

A social network can be partitioned into different disjoint communities where people in the same community tend to have strong connection and similar behavior. In other words, being able to detect social communities can facilitate the prediction of human movement. Therefore, in our work in order to select PCMs for a given person, behavioral similarities and community interactions are considered, in addition to encounter frequency.

A number of studies were conducted to detect social communities using graph clustering [24,37] where a network was divided into disjoint communities by using clustering techniques. There were several studies on community detection based on the contact history of members in the network, e.g., encounter frequency and duration [38] and the total number of past encounters of a person [39]. Eagle and Pentland represented the behavior of individuals from a set of primary vectors called eigenbehaviors [39]. Then, community affiliation can be inferred by computing the social behavior distances (e.g., the total number of past Bluetooth encounters) between a person and other members of a social circle.

3. Preliminaries

3.1. Dataset

In this work, we consider two different datasets. The first one is called the MIT Reality Mining dataset [12], which was collected during a period of nine months with the attendance of 106 subjects, including students and faculty members of MIT. Since subjects in the MIT dataset are involved in the same university, the social relationships between people exist with a high probability.

The MIT dataset provides cell tower logs including the tower transition events and a set of base stations seen by the participants. In cellular networks, a mobile phone can be within the ranges of several cellular towers. However, the phone is only associated with the tower with the strongest signal. The events of tower transitions are recorded with cell tower ID and a timestamp. Due to the fact that small and short-range cells that cover distances of few hundred meters are more popular in metropolitan areas [13], the cell tower logs can be used to represent human locations [10,14–18]. Hence, this work uses the dataset of cellular traces in order to evaluate the movement prediction model. The subjects in the dataset participated in the experiment in different time periods, and some subjects have no data or very little data [24]. Therefore, by considering overlapping periods and available data, 43 people with sufficient mobility data were selected. \( M \) and \( m \) are defined as the set of chosen people and the cardinality of \( M \), respectively.
The second one, called Dartmouth dataset [19], provides the mobility traces extracted from logs of APs in the Dartmouth university campus. A log message including the timestamp, user ID, and AP ID was recorded when a mobile device connects or disconnects to the AP. Because of the short range of the Wi-Fi technology, human mobility can be represented as a sequence of connected APs [20–22]. For the Dartmouth dataset, a 4-month period from 3 January to 30 April 2004 was considered since during this period the academic campus was relatively consistent [20,40]. Similar to the MIT dataset, persons whose mobility data was provided less than 75% over the experiment period was filtered out. Then, the Dartmouth dataset includes 162 mobile users.

3.2. Location Extraction

In this subsection, the way to extract locations for people is described. In this work, location information based on time slots is used. Note that a mobile device may be connected to several base stations (e.g., cell towers or access points) during a time slot. Therefore, in such cases, the way to determine the locations of people in a time slot is needed.

Let \( \lambda \) denote a threshold for location extraction \((0 \leq \lambda \leq 1)\). By using \( \lambda \), the representative base station with which the phone is associated the most is determined, since a mobile device may connect to several base stations during a time slot. Specifically, if the ratio of the time a person spends at a base station to the time slot length exceeds \( \lambda \), this base station is regarded as the representative human location in the specific time slot. In this work, \( \lambda \) and time slot length are set to 0.5 and 30 min, respectively. In cases where no base station satisfies the conditions for location extraction, or where a mobile phone does not receive any signal during a time slot, the person’s location in that time slot is marked as undefined.

In case of the MIT dataset, positions at which people rarely stay should be pruned. The locations are first arranged in descending order of occurrence frequency in the dataset. Then, a location set that contributes 98% of the cellular traces is selected for use in this work. Let \( L \) denote the set of all symbolic locations in the dataset after the pruning process. In the MIT dataset, the total number of positions remaining after the above pre-processing is 482, i.e., \(|L| = 482\). Meanwhile, in the Dartmouth dataset, there are total 399 APs. Table 1 summarizes the main characteristics of both datasets.

| Table 1. Summary of two datasets after locations and users extraction. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Dataset         | Number of People \((m)\) | Number of Locations \(|L|\) | Duration (Days) | Period of a Day |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| MIT             | 43              | 482             | 75              | 8 a.m. to 12 p.m. |
| Dartmouth       | 162             | 399             | 118             | 8 a.m. to 12 p.m. |

4. The Proposed Mobility Prediction Model

4.1. Problem Definition

Recall that \( M \) and \( L \) are the set of \( m \) members in the entire community and the set of locations, respectively. We take into account the problem of predicting the position of person \( p \) \((1 \leq p \leq m)\) at current or future time \( t \) with the preconditions that the recent locations of person \( p \) are unknown. Assume that the historical positions of \((m-1)\) other members can be observed. Therefore, we can make use of the historical information of the rest of people for predicting where person \( p \) stays at the current or future time.

Let us define \( y_t \) as the location of person \( p \) at time \( t \) and \( R \) as the spatio-temporal information of the rest of members during the historical period. Specifically, \( R = \{ R_j \mid R_j \in L, i \in M, i \neq p, 1 \leq j \leq t \} \). The problem of predicting human location is formulated as follows. The objective is to predict where person \( p \) is the most likely to visit at time \( t \) given the recent information \( R \), or to maximize the following probability:

\[
P(y_t | R)
\]  (1)
4.2. Two-Phase Mobility Prediction Framework

The proposed human mobility prediction framework consists of two phases, i.e., detection of persons with correlated movements (PCMs) and PCM-based location prediction, as shown in Figure 1. Let \( r \) denote the number of selected PCMs. In the first phase, \( r \) from among \((m-1)\) remaining members are extracted as PCMs of person \( p \). Then, given the location information of the chosen members, the current or future position of person \( p \) is predicted.

Figure 1. The proposed framework for predicting the current or future location of person \( p \).

In the first phase, we measure independently the social correlation between person \( p \) and each of the \((m-1)\) other members. In order to choose \( r \) useful PCMs for person \( p \), these \((m-1)\) people are arranged according to their correlation scores, and \( r \) members with the highest scores are selected. The training set is used in the first phase where the input is vector \( x \) that represents positional information of the \((m-1)\) remaining members. Table 2 shows an example of vector \( x \) when the location of an arbitrary person, \( p \), is estimated, where \( x \) consists of spatial and temporal information. More specifically, \( x_{\text{loc}}^1, x_{\text{loc}}^2, \ldots, x_{\text{loc}}^{p-1}, x_{\text{loc}}^{p+1}, \ldots, x_{\text{loc}}^m \) denotes the positions of the \((m-1)\) remaining members, while \( x_{\text{day}}, x_{\text{time}} \) account for the day and time slot indices, respectively. The first phase returns \( r \) PCMs of person \( p \) as the output.

Then, the location prediction phase estimates the most likely current or future position of person \( p \) using the spatio-temporal information of the \( r \) selected PCMs. Each sample of the training set is given by \((z_{i_1}^{i_2}, y_t)\), where label \( y_t \) is location information of person \( p \) at time \( t \), and feature vector \( z_{i_1}^{i_2} \) is spatio-temporal information of PCMs from time slots \( i_1 \) to \( i_2 \) where \( 1 \leq i_1 \leq i_2 \leq t \). In the proposed framework, a conditional probability distribution is estimated with the objective of maximizing probability \( P(y_t|z_{i_1}^{i_2}) \). More specifically, given the location information of PCMs at time slots between \( i_1 \) and \( i_2 \), the position of person \( p \) at time \( t \) needs to be predicted.

In this work, \((t-1) \leq i_1 \leq i_2 \leq t\) is considered, i.e., \((i_1,i_2) = \{(t-1,t-1),(t-1,t),(t,t)\}\). In cases where \( i_1 = i_2 = (t-1) \), the location of PCMs in the previous time slot \((t-1)\) is used to predict the next location of person \( p \) (i.e., estimate person \( p \)'s future location at time \( t \)). Meanwhile, \((i_1,i_2) = (t-1,t)\) indicates that the model estimates the current position of the person using previous and current location information of PCMs. Whereas, if \( i_1 = i_2 = t \), the current position of the person is estimated given the information of PCMs at time \( t \). Hereafter, the model is regarded as estimating the location of a person at time slot \( t \). In the second phase, the PCM-based location prediction model is used to label an unknown sample as one out of \(|L|\) possible locations.
Table 2. Elements of the input vector in the first phase of the model for predicting the location of person \( p \).

| Feature | Description |
|---------|-------------|
| \( x_{i}^{\text{loc}}, 1 \leq i \leq m, i \neq p \) | location of \((m - 1)\) remaining members |
| \( x^{\text{day}} \) | index of day in a week |
| \( x^{\text{time}} \) | index of time slot in a day |

There are several benefits from the proposed framework. Our model does not require a recent location sequence of person \( p \) when predicting the current or future location of person \( p \), which is beneficial in cases where the location information of that person is not available. Note, however, that the location information of person \( p \) is necessary for training parameters of the prediction model.

The proposed framework first selects \( r \) PCMs based on human mobility characteristics, e.g., encounter frequency and behavioral patterns. Then, the location information of only these chosen PCMs is used to predict a person’s location at time \( t \). As a result, by reducing redundant input features of the second phase, the overfitting problem can be mitigated. Moreover, the proposed framework allows for low time and space complexity.

5. Detection of Persons with Correlated Movements

In this section, two methods for selecting PCMs are proposed: community interaction similarity-based (CISB) and behavioral similarity-based (BSB) methods. Each uses a different measurement score to estimate the closeness between people’s movement patterns. Then, the \( r \) PCMs with the highest scores are selected.

5.1. Community Interaction Similarity-Based Method

Recall that the encounter frequency-based (EFB) method only uses direct encounters between certain individuals to estimate their friendship. In contrast, to measure correlation scores between persons \( p \) and \( q \), CISB considers interactions between these persons and other community members as well as direct encounters between them. The CISB method is inspired by the fact that interactions and relationships with other members in the community have a great impact on a person’s behaviors.

The CISB method utilizes the community interaction-based similarity (CIS) tensor shown in Figure 2, which represents the spatio-temporal encounters between people. The 3D CIS tensor, \( S \), consists of \( m \) layers each of which is a matrix, e.g., \( S_{p,q}^{t} \) reflects the interactions between person \( q \) and the rest of the \((m - 1)\) people. More specifically, the \( p \)th row in the matrix, i.e., \( S_{p,q,t}^{t} \), indicates the temporal encounters of persons \( p \) and \( q \). Let \( n_{\text{day}} \) and \( n_{\text{time}} \) denote the total number of days in the overlapping period and the number of time slots per day, respectively. \( n = n_{\text{day}} \times n_{\text{time}} \) indicates the total number of samples per person during the whole period. \( a_{p,q}^{t} \) accounts for the encounter between persons \( p \) and \( q \) in time slot \( t \).

Vector \( S_{p,q}^{t} \), representing the encounters between persons \( p \) and \( q \), is defined as follows:

\[
S_{p,q}^{t} = \begin{cases} 
1, & \text{if persons } p \text{ and } q \text{ are in the same cell} \\
0, & \text{otherwise} 
\end{cases} 
\]

Person \( p \) is not considered able to encounter himself or herself, because tensor \( S \) represents the interactions between a person and the entire community, i.e., \( \forall t, p : S_{p,t,p} = 0 \). The CIS tensor owns a symmetric property, i.e., \( \forall t : S_{p,t,q} = S_{q,t,p} \).
where we consider the meet/min correlation coefficient [41] for measuring the similarity between the places two people with similar behaviors visit at the same time are denoted as correlated locations. The pattern, and hence, the location of a person can be estimated by using the other’s positional information.

Figure 2. A three-dimensional CIS tensor, denoted by $S$.

From the CIS tensor, we reshape every 2D matrix, $S_{i,j,p}$, to a one-dimensional vector with $(m \times n)$ elements, which is denoted by $s^{(i)}$. In order to evaluate the closeness between persons $p$ and $q$, we consider the meet/min correlation coefficient [41] for measuring the similarity between $s^{(p)}$ and $s^{(q)}$, which can be used if the vectors include a lot of $1$ s. Thus, score $d_{p,q}$ between nodes $p$ and $q$ is calculated as follows:

$$d_{p,q} = \frac{s^{(p)^T} s^{(q)}}{\min(\|s^{(p)}\|_1, \|s^{(q)}\|_1)}$$

(3)

where $\|s^{(q)}\|$ is the $L^1$-norm of vector $s^{(q)}$. After obtaining scores between person $p$ and $(m - 1)$ others, $r$ people with the highest scores are chosen as PCMs.

Because of the characteristics of CIS, i.e., $\forall i, p: S_{p,i,p} = 0$, and the dot product in the numerator of Equation (3), when measuring similarity $d_{p,q}$, the interactions between persons $p$ and $q$ vanished unintentionally. For example, in calculating $d_{1,2}$, two vectors $s^{(1)}$ and $s^{(2)}$ are as follows:

$$s^{(1)} = [0 \quad 0 \quad ... \quad 0 \quad a^1_{1,1} \quad a^1_{1,2} \quad ... \quad a^1_{1,n} \quad a^1_{2,1} \quad a^1_{2,2} \quad ... \quad a^1_{2,n}]$$

$$s^{(2)} = [a^2_{1,1} \quad a^2_{1,2} \quad ... \quad a^2_{1,n} \quad 0 \quad 0 \quad ... \quad a^2_{m,1} \quad a^2_{m,2} \quad ... \quad a^2_{m,n}]$$

Since the first $n$ elements of $s^{(1)}$ and the $(n + 1)^{th}$ to $(2n)^{th}$ elements of $s^{(2)}$ are zeros, similarity score $d_{1,2}$ that is estimated using Equation (3) does not consider the encounters of persons 1 and 2.

To avoid this vanishing problem, when calculating $d_{p,q}$, the CIS tensor is temporarily modified as follows: $S_{p,i,q} = S_{p,i,p}$ and $S_{q,i,q} = S_{q,i,p}$. For example, the modified vectors $s'^{(1)}$ and $s'^{(2)}$ corresponding to $s^{(1)}$ and $s^{(2)}$, respectively, can be obtained as follows:

$$s'^{(1)} = [a^1_{1,1} \quad a^2_{1,1} \quad ... \quad a^1_{1,n} \quad a^2_{1,n} \quad a^1_{2,1} \quad a^2_{2,1} \quad ... \quad a^1_{2,n} \quad a^2_{2,n}]$$

$$s'^{(2)} = [a^2_{1,1} \quad a^2_{1,2} \quad ... \quad a^2_{1,n} \quad a^1_{2,1} \quad a^2_{2,1} \quad ... \quad a^2_{m,1} \quad a^1_{m,2} \quad ... \quad a^2_{m,n}]$$

By using the modified CIS tensor, the CISB method takes into account the interactions with the entire community members.

5.2. Behavioral Similarity-Based Method

Persons $p$ and $q$ are said to have similar behavior if they experience a correlated visit behavior pattern, and hence, the location of a person can be estimated by using the other’s positional information. The places two people with similar behaviors visit at the same time are denoted as correlated locations. Note that correlated locations may be different. As seen in Table 3, persons $p$ and $q$ show a lot of similar behaviors even though they do not encounter each other. For example, person $p$ exercises at a gym whenever person $q$ visits a café. It is beneficial to measure the behavior similarity between
persons because the location of a person can be inferred by using information of another person with a correlated behavioral pattern. For instance, the location of person \( p \) at 3 p.m. on day 2 can be predicted using person \( q \)’s location (café). Based on this observation, in the BSB method, persons with the highest behavioral correlation are chosen as PCMs.

Table 3. Behaviors of Persons \( p, q, \) and \( u \).

| Time slot | Day 1 | Day 2 |
|-----------|-------|-------|
| 1 p.m.    | math class | 1 p.m. | math class |
| 3 p.m.    | gym       | 3 p.m. | office A   |
| 5 p.m.    | restaurant B | 5 p.m. | café       |
| 7 p.m.    | tea house | 7 p.m. | restaurant C |
|           |           |       | piano class|
|           |           |       | office A   |

The key difference between the BSB method and friendship-based methods (i.e., CISB and EFB) is as follows. When selecting a PCM, the BSB method does not consider real friendships between people while friendship-based methods extract co-location information to measure the friendships between them. Specifically, in the EFB method, members with the highest number of encounters with a given person are chosen as PCMs. As shown in Table 3, persons \( p \) and \( u \) encounter each other several times (in math class and the laboratory), which indicates a close friendship between them. Therefore, in friendship-based methods, person \( u \) is likely to be chosen as a PCM of person \( p \). Meanwhile, in case of two people with similar behaviors but different correlated locations, a person is not likely to be selected as a PCM of the other person. For example, persons \( p \) and \( q \) have the same behavior of having dinner at 5 p.m. at restaurant B and restaurant C, respectively. Since there is no encounter between people \( p \) and \( q \), in friendship-based methods person \( q \) does not obtain a high enough score to be selected as a person \( p \)’s PCM. On the other hand, the BSB method considers this correlated behavior for choosing PCMs.

In the BSB method, a feed-forward neural network (NN), shown in Figure 3, is constructed to evaluate the behavioral similarity between persons \( p \) and \( q \). \( x_{q,t}^{\text{loc}} \) is defined as the location of person \( q \) at time \( t \), while \( x_{t}^{\text{day}} \) and \( x_{t}^{\text{time}} \) are indices of the day and the time slot at time \( t \), respectively.

The objective is to learn the underlying conditional probability distribution \( f_{\text{BSB}}(x_{q,t}^{\text{loc}}, x_{t}^{\text{day}}, x_{t}^{\text{time}}) = P(x_{p,t}^{\text{loc}} | x_{q,t}^{\text{loc}}, x_{t}^{\text{day}}, x_{t}^{\text{time}}) \), i.e., the probability mass distribution that person \( p \) stays in a specific location during time slot \( t \) given the spatio-temporal information of person \( q \).

Figure 3. Architecture of the neural network for measuring behavioral similarity between persons \( p \) and \( q \) in the BSB method.

Here, the training process is described. First, symbolic location \( x_{q,t}^{\text{loc}} \) is embedded into an \(|L|\)-dimensional indicator vector using one-hot encoding, because there is no ordinal relationship between locations. Similarly, time slot and day indices (i.e., \( x_{t}^{\text{day}} \) and \( x_{t}^{\text{time}} \), respectively) are also
converted to indicator vectors. Then, these vectors are used as input units of the neural network including one hidden layer with 150 logistic sigmoid activation nodes. These input features propagate progressively throughout the network until reaching the softmax output layer, which obtains a conditional probability distribution over locations. Θ is defined as the parameters of the NN consisting of weights, denoted by W, and biases that are randomly initialized. Let η denote the number of samples in the training dataset. After retrieving the conditional probability at the output layer, regularized cost function $J(\Theta)$ using cross-entropy error [42] is calculated as follows:

$$J(\Theta) = -\frac{1}{\eta} \sum_{j=1}^{\eta} t(j)^{T} \log f^{(j)}_{\text{BSB}}(x_{q,\text{loc}}, x_{i,\text{day}}, x_{i,\text{time}}) + \Omega(W)$$

where $f^{(j)}_{\text{BSB}}(x_{q,\text{loc}}, x_{i,\text{day}}, x_{i,\text{time}})$ and $t(j)$ are the estimated output and target vectors of the $j^{th}$ instance, respectively. Note that $\log(.)$ is an element-wise operation. $\Omega(W)$ is the $L^2$-based regularization term. The neural network model is trained to minimize cost function $J(\Theta)$ using a back-propagation algorithm combined with a gradient descent method.

In order to select $r$ PCMs of person $p$, $(m-1)$ NNs are trained separately, where each NN measures the similar behavior pattern between individuals $p$ and $q$ ($p, q \in M, p \neq q$). The prediction accuracy, $\xi_{p,q}$, of the $q^{th}$ NN, which indicates the behavioral similarity of persons $p$ and $q$, is defined as the ratio of the number of correctly predicted instances to the number of test samples. In the BSB method, $r$ PCMs among the $(m-1)$ remaining members with the highest $\xi_{p,q}$ are selected.

It is worthwhile to note that there are alternative ways of selecting a set of PCMs. For example, forward search (FS) [43] begins with an empty set and then progressively incorporates each feature (i.e., PCM) into the set. However, FS has an inherent weakness, i.e., high time and space complexity when a large number of features need to be selected.

Another simple approach that can be considered is that positional information of all $(m-1)$ remaining people is used as input features for a prediction model that labels the location of a person at time $t$ with one of the $|L|$ positions. In our work, this simple method is called ignored feature selection (IFS) since there is no feature selection process. Note that IFS may need to train a model with tens or hundreds of thousands of features, which may result in a long training time, and might suffer from the overfitting problem.

In order to gain insight into computational costs, the number of training parameters for three methods including IFS, FS, and BSB is compared. For IFS, there is no PCM selection phase, i.e., the location of person $p$ is predicted given the positional information of the $(m-1)$ other people. With FS, to choose the first PCM, $(m-1)$ NNs are trained independently. To select the second PCM, a combination of the first PCM and one of the $(m-2)$ remaining members is examined; $(m-2)$ NNs need to be learned, each of which estimates person $p$’s location at time slot $t$ given the contextual information of two examined people. Similarly, to select the $j^{th}$ PCM, $(m-r)$ NNs are used to evaluate the subset of the $(r-1)$ already selected PCMs and one of the $(m-r)$ remaining members. In the FS method, a large $r$ value leads to a significant increase in the number of training parameters.

In the BSB and FS methods, it is assumed that three selected PCMs are used for the location prediction phase, where a neural network with three hidden layers of 500, 150, and 50 neurons is considered. For a fair comparison, the number of hidden neurons for IFS is proportional to that of the BSB/FS methods in the mobility prediction phase with the ratio $(m-1) : r$ because $(m-1)$ and $r$ are the numbers of people whose location information is given to the IFS and BSB/FS methods, respectively.

Table 4 compares the number of training parameters of the three methods, where $l_{\text{BSB}}$, $l_{\text{IFS}}$, and $l_{\text{FS}}$ denote the number of parameters, i.e., weights and biases, in behavioral similarity-based, ignored feature selection, and forward search methods, respectively. When the total number of people, $m$, increases, the three considered methods need to train more parameters. However, the BSB method always learns the least number of parameters. Moreover, the increase rate is quite different in each method, e.g., when $m$ varies from 43 to 1000 people, $l_{\text{BSB}}$ rises $\frac{147.42}{7.0} \approx 21$ times, compared to
1453.7 \approx 24.6 \text{ times with } l_{\text{IFS}} \text{ and } \frac{8933.1}{158.3} \approx 54.6 \text{ times with } l_{\text{FS}}. \text{ In summary, because only location information of PCMs is used to predict human mobility of a given person, the proposed BSB method incurs a much lower computation cost in terms of time and space complexity than IFS and FS, which makes the BSB method more practical, particularly when there is a large number of people.}

Table 4. Comparison of the number of parameters among behavioral similarity-based (BSB), ignored feature selection (IFS), and forward search (FS) methods.

| m     | l_{\text{BSB}} (millions) | l_{\text{IFS}} (millions) | l_{\text{FS}} (millions) |
|-------|---------------------------|---------------------------|---------------------------|
| 43    | 7.00                      | 59.02                     | 158.30                    |
| 100   | 15.36                     | 142.09                    | 878.17                    |
| 500   | 74.06                     | 725.03                    | 2229.0                    |
| 1000  | 147.42                    | 1453.7                    | 8933.1                    |

6. The PCM-Based Location Prediction (PLP) Model

In this section, we propose the PCM-based location prediction (PLP) model based on the neural network to predict human mobility given the movement patterns of r selected PCMs. Figure 4 illustrates the NN-based prediction model where the location of a person at time t is estimated given the information of r PCMs, i.e., \( z_{l_{1:t}}^{i_1} = \{ z_{l_{1:t}}^{\text{loc}}, z_{l_{1:t}}^{\text{day}}, z_{l_{1:t}}^{\text{time}} \} \). Spatial information \( z_{l_{1:t}}^{\text{loc}} \), where \( 1 \leq l \leq r \) and \( i_1 \leq \tau \leq i_2 \), indicates the symbolic location of the r PCMs from time \( i_1 \) to \( i_2 \). Meanwhile, temporal information \( z_{l_{1:t}}^{\text{day}} \) and \( z_{l_{1:t}}^{\text{time}} \) are the indices of the day of the week and the time slot in the day at time \( t \), respectively. Let \( y_{\text{pred}}^{\text{NN}} = f_{\text{NN}}(z_{l_{1:t}}^{\text{loc}}, z_{l_{1:t}}^{\text{day}}, z_{l_{1:t}}^{\text{time}}) \) denote the output of the NN-based predictor, i.e., the most likely location of the person during time slot \( t \). Note that to predict the location at time \( t \) of \( m \) persons in the community, \( m \) corresponding NN-based prediction models need to be trained separately. The remainder of this subsection describes an NN-based predictor.

The function \( f_{\text{NN}}(z_{l_{1:t}}^{\text{loc}}, z_{l_{1:t}}^{\text{day}}, z_{l_{1:t}}^{\text{time}}) \) can be decomposed into three parts, as follows.
1. The contextual information of \( r \) PCMs, i.e., \( z_{11}^{\text{loc}}, z_{12}^{\text{loc}}, \ldots, z_{i,r}^{\text{loc}}, z_{i}^{\text{day}}, z_{i}^{\text{time}} \), is mapped onto real vectors using the one-hot encoding function \( c(\cdot) \), where output is the indicator vector in which only one element is set to 1 while others are 0. Since \( |l| \) is the size of the extracted location set, \( c(z_{11}^{\text{loc}}), c(z_{12}^{\text{loc}}), \ldots, c(z_{i,r}^{\text{loc}}) \) are vectors in \( \mathbb{B}^{|l|} \), where \( \mathbb{B} \) is a set of binary numbers; \( c(z_{i}^{\text{day}}) \) is a day-index vector in \( \mathbb{B}^7 \). Meanwhile, the cellular trace between 8 a.m. and 12 p.m. is extracted and each time slot lasts 30 min; therefore, \( c(z_{i}^{\text{time}}) \) is mapped onto a 32-dimensional binary vector. In addition, grouped time slots are considered. In this case, \( c(z_{i}^{\text{time}}) \) is a three-dimensional vector where each dimension corresponds to one of the three parts of a day, i.e., morning, afternoon, and evening.

2. The NN classifier maps input sequence \( c(z_{i}^{j}) = \{c(z_{i,1}^{\text{loc}}), c(z_{i}^{\text{day}}), c(z_{i}^{\text{time}})\} \) to a conditional probability distribution over locations. Let \( v^{(i)} \) denote an indicator vector in which only the \( i \)th element is 1 and others are 0 (\( 1 \leq i \leq |l| \)). Vector \( t \) is defined as the target vector of the NN classifier. The output of the NN is denoted by vector \( y_{\text{NN}} \), where the \( i \)th element, \( \hat{y}_i \), estimates posterior probability \( \hat{y}_i = P(t = v^{(i)} | c(z_{i}^{j})) \) that the person stays in location \( i \) given the information of \( r \) PCMs.

3. The vector \( v^{\text{max}} = \arg \max_{v^{(i)}} P(t = v^{(i)} | c(z_{i}^{j})) \) that indicates the most likely location of the person is mapped onto symbolic position \( y_{\text{NN}}^{\text{pred}} \) by the inverse function, i.e., \( y_{\text{NN}}^{\text{pred}} = c^{-1}(v^{\text{max}}) \).

Now, the neural network training process that includes feed-forward propagation, a cost function calculation, and a parameter update is presented. First, feed-forward propagation is described. In this work, \( \mu \) is defined as the number of hidden layers, and a neural network is considered with logistic sigmoid activation units denoted by \( \sigma(\cdot) \). Hereafter, the input layer is referred to as layer 0, the \( j \)th hidden layer as layer \( j \) (\( 1 \leq j \leq \mu \)), and the output layer as layer \( (\mu + 1) \). Let \( \Theta = \{W^{(1)}, W^{(2)}, \ldots, W^{(\mu + 1)}, b^{(1)}, b^{(2)}, \ldots, b^{(\mu + 1)}\} \) denote all learning parameters consisting of weights and biases of the NN classifier. More specifically, \( W^{(j)} \) and \( b^{(j)} \) are the matrix of weights and the vector of biases, respectively, that indicate connections from layers \( (j - 1) \) to \( j \); \( h^{(j)} \) is defined as the state vector at layer \( j \), and \( h^{(j)} \) is given by:

\[
\begin{align*}
    h^{(j)} &= \sigma(W^{(j)\top} h^{(j-1)} + b^{(j)}), \quad 1 \leq j \leq \mu \\
\end{align*}
\]

(5)

Note that \( h^{(0)} \) is a vector of input features at layer 0. For the output layer, the unnormalized output vector \( h^{(\mu + 1)} \) is calculated by \( h^{(\mu + 1)} = W^{(\mu + 1)\top} h^{(\mu)} + b^{(\mu + 1)} \).

Then, the output of a softmax function represents the categorical distribution over locations where the \( i \)th element of the output vector reflecting the probability that the person stays at time \( t \) in location \( i \) is calculated as follows:

\[
\begin{align*}
    \hat{y}_i = P(t = v^{(i)} | c(z_{i}^{j})) &= \frac{e^{h_i^{(\mu + 1)}}}{\sum_{j=1}^{|l|} e^{h_j^{(\mu + 1)}}} \\
\end{align*}
\]

(6)

where \( h_i^{(\mu + 1)} \) is the \( i \)th value of unnormalized output vector \( h^{(\mu + 1)} \).

Secondly, the calculation of the cost function is briefly described below. Assume that there are \( \eta \) training instances. Cost function \( J(\Theta) \) using the cross entropy metric and regularization method can be expressed as follows:

\[
\begin{align*}
    J(\Theta) &= -\frac{1}{\eta} \sum_{j=1}^\eta t^{(j)\top} \log \hat{y}_{\text{NN}}^{(j)} + \Omega(W) \\
\end{align*}
\]

(7)

where \( \hat{y}_{\text{NN}}^{(j)} \) and \( t^{(j)} \) are, respectively, the estimated output and target vectors of the \( j \)th instance, and \( W \) denotes the weight matrices. Note that \( \log \hat{y}_{\text{NN}}^{(j)} \) is an element-wise operation. Regularization \( \Omega(W) \) is implemented using the \( L^2 \) parameter norm penalty.

Finally, parameters including weights and biases are obtained by applying a back propagation algorithm, which minimizes the cost value using the gradient descent method with momentum.
Let $\alpha$ and $\gamma$ denote the learning rate of gradient descent and the momentum coefficient, respectively, ($0 \leq \gamma \leq 1$). Let $\hat{\theta}$ denote a training parameter in the set $\Theta$, i.e., $\hat{\theta} \in \Theta$. Then, each $\hat{\theta}$ is updated in the $j^{th}$ epoch as follows:

$$\omega = \gamma \omega + \alpha \frac{\partial J(\Theta)}{\partial \hat{\theta}}$$

$$\hat{\theta}^{(j)} = \hat{\theta}^{(j)} - \omega$$

where partial derivative $\frac{\partial J(\Theta)}{\partial \hat{\theta}}$ is the gradient of the cost function with respect to $\hat{\theta}$, and $\omega$ is the current velocity vector with the same dimensions as parameter $\hat{\theta}$.

7. Evaluation Results and Discussion

In this section, the performance of the human mobility prediction framework is examined under different PCMs detection methods and is compared with the baseline prediction model, most frequent location (MFL). The whole dataset is randomly partitioned into training, validation, and test sets at the ratio of 5:2:3. Recall that in case of the Dartmouth dataset the 118-day period is selected and for each day human mobility from 8 h to 24 h is considered. Since each time slot lasts 30 min, there are total 32 time slots per day. Therefore, the number of samples in the Dartmouth dataset is $118 \times 32 = 3776$. The training, validation, and test sets consist of 1886; 755; and 1133 instances, respectively. The training set is used to fit the model while the purpose of validation set is to determine the appropriate hyper-parameters for the neural network, e.g., the number of hidden layers, hidden units, activation units, and learning rate. Note that only the training set is used in the first phase of the prediction model and the performance of model is estimated based on the test set.

Table 5 presents the setup for the mobility prediction framework. Specifically, the two proposed PC selection methods consisting of community interaction similarity-based (CISB) and behavioral similarity-based (BSB) are evaluated and compared with three recent selection approaches: encounter frequency-based method (EFB) [18,23,24], spatial closeness (SC) [6], and spatiotemporal closeness (STC) [6]. Table 6 shows a list of acronyms which are used in the manuscript.

| Parameters                  | Values                           |
|-----------------------------|----------------------------------|
| PCM selection method        | {SC, STC, EFB, CISB, BSB}        |
| Prediction model            | {PLP}                            |
| Temporal location type      | {pre-loc, cur-loc, pre-cur-loc}  |
| Time slot feature           | {un-grouped, grouped}            |
| Number of selected PCMs ($r$) | {1, 2, 3, 4, 5}                  |
| Top-$k$ accuracy            | {top-1, top-2, top-3, top-4}     |

Table 6. List of Acronyms.

| Abbreviation | Meaning                          |
|--------------|----------------------------------|
| BSB          | Behavioral similarity-based      |
| CISB         | Community interaction similarity-based |
| EFB          | Encounter frequency-based       |
| MFL          | Most frequent location          |
| PCM          | Persons with correlated movements|
| PLP          | PCM-based location prediction    |
| SC           | Spatial closeness               |
| STC          | Spatiotemporal closeness        |

Recall that Alhasoun et al. [6] constructed a prediction model based on dynamic Bayesian networks which leveraged historical location data of selected PCMs to predict the most probable
position of the person-of-interest. However, the prediction model in [6] requires to know the last visited place of the person-of-interest, which is not applicable to our considering problem where the historical position data of the person-of-interest is assumed to be unknown. Note that since there was no existing prediction model designed for the considering problem, the PCM-based location prediction (PLP) model is used in the experiments.

Additionally, we make a performance comparison between our two proposed PCMs selection approaches and counterpart methods in [6]. There were three similarity measurement metrics to select PCMs in [6], and specifically the first method, temporal closeness, compares patterns of communication (e.g., call, sms, and data) between people. Due to the requirement of extra communication information of members, the temporal closeness is not considered as a counterpart method for selecting PCMs in our work. Therefore, we compare the proposed PCMs selection methods with only two approaches in [6]: STC and SC.

For the PLP model, a variety of NN architectures which have a different number of layers, hidden units, and activation functions were examined. Then, among the considered setting, the most appropriate NN architecture was selected by using the validation set. Specifically, there is 1 hidden layer of 150 sigmoid activation units in the NN of the BSB approach. In the NN-based predictor, 500, 150, and 50 sigmoid hidden units are used in the first, second, and third hidden layers, respectively. Weight and bias values are initialized with normal distribution. The $L^2$ regularization coefficient is set to 0.01. Learning rate in the gradient descent and the momentum coefficient are set to 0.4 and 0.1, respectively. Full batch learning with 1500 iterations is examined. Note that only the training set is used to select the $r$ PCMs in the first phase of the prediction models. For example, in case of the Dartmouth dataset, $r$ PCMs are determined by using the training set of 1866 samples which are randomly selected from the total 3776 instances.

Moreover, three types of temporal information are considered. The number of selected PCMs varies from 1 to 5, while top-$k$ accuracy is considered with $k = \{1, 2, 3, 4\}$. In addition, un-grouped and grouped time slot features are compared. The bold text in the second column of Table 5 denotes default values of the prediction model. In this work, performance metrics including average and standard deviation of prediction accuracy are obtained over 5 runs. The results are collected via averaging the entire people.

The most noticeable results are summarized as follows. From the conducted experiments, the performance of designed PLP model outperforms that of the baseline method, MFL. With regard to the PCMs selection methods, the proposed BSB shows significantly better performance than other PCM extraction approaches. In addition, the generalization capability of the proposed framework increases as a larger number of PCMs are embedded into the model. In particular, this increase is more clearly shown in case of BSB.

7.1. Performance Comparison of PCM Selection Methods

In this subsection, the performance results of PCMs extraction methods are discussed with different temporal features of selected PCMs. The location of a person during time $t$ is predicted with the support of PCMs’ positions during the time slot $(t-1)$ (denoted by pre-loc), time slot $t$ (denoted by cur-loc), and both $(t-1)$ and $t$ time slots (denoted by pre-cur-loc). Recall that the time slot lasts 15 min. The obtained results are shown in Figure 5.

First, we evaluate the predictability of the baseline prediction model, most frequent location (MFL). As can be seen in Figure 5, the MFL model achieves much lower performance than the proposed PLP architecture, and specifically leads to top-1 accuracy of 38.62% and 70.92% with MIT and Dartmouth datasets, respectively. Since the MFL model does not use information of PCMs to predict where person $p$ stays at time $t$, the temporal feature of PCMs does not affect the performance of MFL model.
Now, the performance of PLP model is analyzed with different PCM selection methods. In EFB and CISB methods, members with the most similar movement paths are selected as PCMs, i.e., people who stay the most frequently with a given person. The main difference between these two methods is that the EFB scheme weights the link of two people based on direct encounters between them, whereas the CISB method compares community interactions between persons to estimate their social correlation. Even though in some cases the CISB method obtains slightly higher accuracy than the EFB method, the two generally achieve similar performance. This indicates that a social relationship between two persons can be mainly reflected by direct interactions between them.

In contrast, the BSB method aims at selecting PCMs with the most correlated behavioral patterns. Our work shows that discovering behavior patterns helps to significantly improve prediction accuracy of the model compared to the friendship-based method. In fact, for members who work in the same university or company, they tend to have similar lifestyles (e.g., behavioral pattern) owing to the common schedules of the workplace. Therefore, considering behavioral similarity between people can be helpful in selecting better PCMs for predicting people’s locations. Specifically, when embedding pre-cur-loc information of PCMs selected by the BSB scheme into the PLP model, we achieve top-1 prediction accuracy as high as 71.00% and 92.31%, respectively, with MIT and Dartmouth datasets.

In the SC scheme, regardless of encounters between two people, members who have similar spatial distribution, or longevity of visiting places are selected as PCMs. The SC method may choose PCMs who have different frequency or regularity of staying in locations with the person-of-interest. For example, if two people $p$ and $q$ spend 4 h at the library everyday, the SC method is likely to choose person $q$ as a PCM of $p$. However, if person $p$ stays at the library in the whole afternoon while $q$ spends 2 h in the morning and 2 h in the evening, using location information of $q$ may not be appropriate to predict the mobility of person $p$.

Meanwhile, in case of the STC method, two people have a high closeness score if they move in a synchronous manner or their movement are highly dependent. Note that PCMs in both STC and BSB tend to move in a synchronous way with the person-of-interest. However, a fundamental difference between BSB and STC is that the STC method does not consider the temporal information when measuring the similarity between two people. Specifically, the STC approach computes the spatial distribution of person $p$ given the location of person $q$. Meanwhile, the BSB method investigates the spatial distribution of person $p$ given both spatial and temporal information of person $q$. By considering the temporal data when selecting PCMs, the BSB approach is able to measure movement association with both spatial and temporal aspects. As a consequence, the BSB method achieves significantly higher prediction accuracy than the STC one with both datasets.

However, the performance gap between BSB and STC in the Dartmouth traces is noticeably smaller than that in the MIT one. The gap difference may come from different characteristics of human
mobility in two datasets. Recall that the STC scheme does not consider the dependence between human movement and the temporal information. Therefore, if human mobility extracted from a dataset (e.g., the Dartmouth) is not highly related to the temporal data, the STC can achieve better accuracy than the case of strong relationship between human movement and the temporal information.

One should note that BSB, SC and STC approaches do not take into account the actual encounter or friendship between people. The experiment results in Figure 5 indicate that friendship-based methods (EFB and CISB) causes the lower performance than other approaches (BSB, SC, and STC) in which friendship is not considered. In addition, the proposed BSB approach achieves the most accurate prediction among the considered PCMs selection methods. For example, using location data of PCMs at time \((t - 1)\) and \(t\), the PLP model predicts human locations at time \(t\) with top-1 accuracy of 81.64%, 82.57%, 85.92%, 82.07%, and 92.31% using EFB, SC, STC, CISB, and BSB methods, respectively, in case of the Dartmouth dataset.

Also, from the observations in Figure 5, among three considered temporal features, pre-cur-loc and cur-loc lead to quite similar performance in both datasets. For example, in case of the MIT traces, the PLP model achieves 71.13% and 71.00% accuracy in predicting the location of person \(p\) given pre-cur-loc and cur-loc information of PCMs, respectively. As anticipated, using data of PCMs during time \(t\) (cur-loc) can result in a more accurate prediction than the use of \((t - 1)\) (pre-loc).

7.2. Top-k Accuracy

In this subsection, we evaluate the top-\(k\) accuracy of the proposed PLP model when \(k = \{1, 2, 3, 4\}\). Other parameters are set to the default values shown in Table 5. For top-\(k\) accuracy, the model outputs a list of \(k\) location(s) with the highest probability. If the correct location belongs to the list of \(k\) element(s), the prediction is considered accurate.

As expected, Figure 6 shows that the obtained accuracy increases from top-1 to top-4. For example, in case of the MIT dataset, the BSB approach results in 71.00%, 82.17%, 85.53%, and 87.59% accuracy corresponding to top-1, top-2, top-3, and top-4, respectively. In all examined PCM selection schemes, the steepest rising rate is observed when \(k\) changes from 1 to 2. Then, the outcomes tend to plateau.

As can also be observed from Figure 6, there is higher prediction accuracy in the Dartmouth dataset than the MIT one. This observation is attributed to the fact that human movement in the MIT dataset was collected within a wider area than that in the Dartmouth traces. More specifically, all APs in the Dartmouth dataset are located in the university campus while movement trajectories of people in the MIT dataset are not bounded in the school area. Therefore, people mobility in the MIT dataset tends to be more divergent and more difficult to predict than that in the Dartmouth traces.

![Figure 6](image-url)
7.3. Effects of the Number of PCMs

In this subsection, the number of selected PCMs varies from 1 to 5. As seen in Figure 7b, using location information of one PCM (i.e., \( r = 1 \)), the PLP model obtains top-1 accuracy of 86.53% with the BSB method compared with 70.92% in case of the MFL model. Meanwhile, the PLP model with EFB, SC, STC, and CISB approaches results in 76.58%, 76.94%, 79.57%, and 76.70% accuracy, respectively.

It should be emphasized that when adding more PCMs, the performance of the prediction model is generally enhanced. In friendship-based methods (CISB and EFB), this is attributed to the fact that a person usually interacts with multiple PCMs rather than one person during the day. For example, in the morning, person \( p \) encounters co-worker \( q \) at the workplace. At lunch, \( p \) and \( r \) have an appointment at their favorite restaurant. Then, persons \( p \) and \( s \) go to the fitness center to exercise in the afternoon. In other PCMs selection approaches, the performance gain with more PCMs is due to the fact that a person has movement correlation with various PCMs depending on the time of a day. For instance, assume that person \( p \) is a middle-aged man. He may have a similar behavioral pattern with younger co-worker \( q \) during the daytime, whereas after work, he may have behaviors similar to other middle-aged men, rather than with the younger co-workers.

Moreover, the performance gap between the BSB and friendship-based methods (CISB and EFB) becomes larger as the number of PCMs embedded in the model increases especially in the MIT dataset. This is because selected PCMs of CISB and EFB tend to have more similar encounter patterns with person \( p \) than under the BSB method. To show this, KL divergence [44] is considered to measure similarity between temporal encounter distributions of PCMs and person \( p \). Vector \( q \) denotes the probability distribution of encounters between persons \( p \) and \( q \). The \( i^{\text{th}} \) element of \( q \) represents the probability that persons \( p \) and \( q \) encounter each other during time slot \( i \). Vector \( r \) indicates a distribution of encounters between persons \( p \) and \( r \). The length of \( r \) and \( q \) equals the number of samples for each person. Using KL divergence, the difference between the two probability distributions is given by:

\[
\phi_{q,r} = \sum_i q_i \log \frac{q_i}{r_i} \tag{10}
\]

where \( q_i \) and \( r_i \) are the \( i^{\text{th}} \) elements of vectors \( q \) and \( r \), respectively. If the score from KL divergence is small, encounter distributions between PCMs and person \( p \) are close, which indicates similar movement patterns between PCMs.

Assume that a mobility prediction model with \( r \) PCMs is considered. If another PCM with highly similar movement patterns to those of the \( r \) existing PCMs is added at input, the location information of the added PCM would not be helpful in predicting location of person \( p \). As shown in Table 7, in cases where the number of selected PCMs is 5, the average value of KL divergence

![Figure 7. Effects of the number of PCMs on the performance of the prediction model in case of (a) the MIT dataset and (b) Dartmouth dataset.](image-url)
between PCMs’ patterns of encounters with a person-of-interest is 11.01, 11.03, and 15.19 in EFB, CISB, and BSB methods, respectively. We can also observe from Table 7 that KL divergence values of SC and STC approaches are generally higher than friendship-based ones since SC and STC do not consider encounters between people.

Table 7. Comparison of KL divergence values between PCMs selection methods.

| Dataset  | EFB   | SC    | STC   | CISB  | BSB   |
|----------|-------|-------|-------|-------|-------|
| MIT      | 11.01 | 11.66 | 15.07 | 11.03 | 15.19 |
| Dartmouth| 13.59 | 13.63 | 20.32 | 13.76 | 20.44 |

Lower KL divergence values from EFB and CISB methods indicate that they tend to select PCMs with more similar mobility patterns than the BSB method, and accordingly, movement patterns of the selected PCMs under the BSB method are more divergent than those under EFB/CISB methods. This contributes to a large accuracy gap between the BSB and friendship-based methods. The results also agree with the assumption that correlated locations in the BSB method do not have to be the same places.

7.4. Effects of Time Slot Feature Selection

Figure 8 compares the precision of the mobility prediction model with two different time slot features: un-grouped and grouped. For the un-grouped case, 32 time slots in a day are mapped into a 32-dimensional vector. Meanwhile, for the grouped case, one-hot encoding converts a time slot index into a vector in \( \mathbb{B}^3 \) where each dimension indicates one of the three parts of a day.

In addition, as shown in Figure 8, it is clear that grouping time slot feature achieves higher prediction accuracy than with un-grouped features. The results reflect that people tend to spend time with each other for a relatively long period. This becomes more understandable in environments like university campuses or companies, where people usually stay with their colleagues for a significant amount of time during the day. After work, they may enjoy leisure activities with friends in the evening. Another reason for the low performance with the un-grouped case may be the small number of data samples compared with the large input dimensions in the un-grouped case.

Figure 8. Comparison of time slot features in the human prediction framework in case of (a) the MIT dataset and (b) Dartmouth dataset.

8. Concluding Remarks

Human mobility prediction has been a key point for the success of a variety of potential applications. For instance, with the development of device-to-device communication technologies, being able to accurately predict human locations will facilitate the design of an efficient data
routing protocol in opportunistic networks. Moreover, the location prediction model inspires further applications such as urban data mining, location-based recommendation services, and contagious disease control. In addition, location estimation is required even when someone may not be willing to share their locations (e.g., geographic profiling of criminals).

Therefore, in this work, we address the mobility prediction problem in which the current or next location of a person is estimated, especially without requiring the positional history of that person. Since the movement of a person is highly related to other people, a two-phase framework is proposed in which persons with correlated movements (PCMs) with the person-of-interest are determined in the first phase. Then, information of these selected PCMs is leveraged in the second phase to estimate the location of the person-of-interest. For selecting PCMs, the communication interaction similarity-based (CISB) method considers encounter interactions between people, whereas the behavioral similarity-based (BSB) scheme selects PCMs who have similar behavioral patterns, instead of considering co-location between people.

For the considered problem, our approach is robust, because it can reduce overfitting by only using the location information of the selected PCMs to estimate the positions of the person-of-interest. Low time and space computational complexity is also achieved. Furthermore, geographic locations are not required in our model. Experimental results show that the BSB method gains significant improvement over other PCMs extraction approaches. In addition, the performance generally increases when more PCMs were embedded into the designed prediction model. In particular, a large number of PCMs is more beneficial under the BSB method than others.

However, our PCMs selection methods are centralized approaches because of requiring mobility traces of all people. Therefore, the proposed PCMs extraction methods are restricted to small or medium networks. In addition, information assurance needs to be taken into account since mobility data of people is exchanged in the network. As a future work, to address the limitations of the centralized approaches, we plan to design a PCMs extraction method which does not require mobility traces of all people. Furthermore, datasets with more people will be used to examine the proposed prediction model in the future work.

Author Contributions: S.Y. and T.-N.D. designed the framework. The performance evaluation was conducted by T.-N.D. and D.V.L. All authors have written and approved the manuscript.

Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1D1A3B03934617).

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Asgari, F.; Gauthier, V.; Becker, M. A survey on Human Mobility and its applications. arXiv 2013, arXiv:1307.0814.
2. Noulas, A.; Scellato, S.; Lathia, N.; Mascolo, C. Mining User Mobility Features for Next Place Prediction in Location-Based Services. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining, Brussels, Belgium, 10–13 December 2012; pp. 1038–1043.
3. Su, J.; Chin, A.; Popivanova, A.; Goel, A.; de Lara, E. User mobility for opportunistic ad-hoc networking. In Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications, Washington, DC, USA, 3 December 2004; pp. 41–50.
4. Wesolowski, A.; Eagle, N.; Tatem, A.J.; Smith, D.L.; Noor, A.M.; Snow, R.W.; Buckee, C.O. Quantifying the Impact of Human Mobility on Malaria. Science 2012, 338, 267–270. [CrossRef] [PubMed]
5. Rossmo, D. Geographic Profiling; CRC Press: Boca Raton, FL, USA, 1999.
6. Alhasoun, F.; Alhazzani, M.; Aleissa, F.; Alnasser, R.; González, F. City Scale Next Place Prediction from Sparse Data through Similar Strangers. In Proceedings of the ACM KDD Workshop, Halifax, NS, Canada, 14 August 2017.
7. Zhao, Z.; Koutsopoulos, H.N.; Zhao, J. Individual mobility prediction using transit smart card data. Transp. Res. Part C Emerg. Technol. 2018, 89, 19–34. [CrossRef]
8. Cho, E.; Myers, S.A.; Leskovec, J. Friendship and Mobility: User Movement in Location-based Social Networks. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 1082–1090.

9. Sadilek, A.; Kautz, H.; Bigham, J.P. Finding Your Friends and Following Them to Where You Are. In Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, Marina Del Rey, CA, USA, 8–12 February 2012; pp. 723–732.

10. Zhang, D.; Zhang, D.; Xiong, H.; Yang, L.T.; Gauthier, V. NextCell: Predicting Location Using Social Interplay from Cell Phone Traces. IEEE Trans. Comput. 2015, 64, 452–463. [CrossRef]

11. Bengio, Y.; Ducharme, R.; Vincent, P.; Janvin, C. A Neural Probabilistic Language Model. J. Mach. Learn. Res. 2003, 3, 1137–1155.

12. Eagle, N.; Pentland, A. Reality Mining: Sensing Complex Social Systems. Pers. Ubiquitous Comput. 2006, 10, 255–268. [CrossRef]

13. Rappaport, T.S.; Roh, W.; Cheun, K. Mobile’s millimeter-wave makeover. IEEE Spectr. 2014, 51, 34–58. [CrossRef]

14. Zhang, D.; Zhang, D.; Xiong, H.; Yang, L.T.; Gauthier, V. NextCell: Predicting Location Using Social Interplay from Cell Phone Traces. IEEE Trans. Comput. 2015, 64, 452–463. [CrossRef]

15. Bengio, Y.; Ducharme, R.; Vincent, P.; Janvin, C. A Neural Probabilistic Language Model. J. Mach. Learn. Res. 2003, 3, 1137–1155.

16. Eagle, N.; Pentland, A. Reality Mining: Sensing Complex Social Systems. Pers. Ubiquitous Comput. 2006, 10, 255–268. [CrossRef]

17. Rappaport, T.S.; Roh, W.; Cheun, K. Mobile’s millimeter-wave makeover. IEEE Spectr. 2014, 51, 34–58. [CrossRef]

18. Zhang, D.; Zhang, D.; Xiong, H.; Yang, L.T.; Gauthier, V. NextCell: Predicting Location Using Social Interplay from Cell Phone Traces. IEEE Trans. Comput. 2015, 64, 452–463. [CrossRef]

19. Bengio, Y.; Ducharme, R.; Vincent, P.; Janvin, C. A Neural Probabilistic Language Model. J. Mach. Learn. Res. 2003, 3, 1137–1155.

20. Karamshuk, D.; Boldrini, C.; Conti, M.; Passarella, A. Human mobility models for opportunistic networks. IEEE Commun. Mag. 2011, 49, 157–165. [CrossRef]

21. Chang, C.-C.; Hsieh, C.-H.; Lin, C.-J.; Maji, A.; Jakkula, J.;unker, W.; Niu, T.; Wang, X.; Lee, C.-J.; Chua, K.-S. LIBLINEAR: A Library for Large Linear Classification. J. Mach. Learn. Res. 2011, 12, 1139–1142.
31. Eagle, N.; Clauset, A.; Quinn, J.A. Location Segmentation, Inference and Prediction for Anticipatory Computing. In Proceedings of the AAAI Spring Symposium: Technosocial Predictive Analytics, London, UK, 23–26 March 2009; pp. 20–25.
32. Zeng, S.; Wang, H.; Li, Y.; Jin, D. Predictability and Prediction of Human Mobility Based on Application-Collected Location Data. In Proceedings of the 2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Chengdu, China, 9–12 October 2017; pp. 28–36.
33. Backstrom, L.; Sun, E.; Marlow, C. Find Me if You Can: Improving Geographical Prediction with Social and Spatial Proximity. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; pp. 61–70.
34. Hossmann, T.; Spyropoulos, T.; Legendre, F. Putting Contacts into Context: Mobility Modeling Beyond Inter-contact Times. In Proceedings of the Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc ’11), Paris, France, 16–20 May 2011; ACM: New York, NY, USA, 2011.
35. Su, Y.; Li, X.; Tang, W.; Xiang, J.; He, Y. Next Check-in Location Prediction via Footprints and Friendship on Location-Based Social Networks. In Proceedings of the 2018 19th IEEE International Conference on Mobile Data Management (MDM), Aalborg, Denmark, 26–28 June 2018; pp. 251–256.
36. Sepahkar, M.; Khayyambashi, M.R. A novel collaborative approach for location prediction in mobile networks. Wirel. Netw. 2018, 24, 283–294. [CrossRef]
37. Girvan, M.; Newman, M.E.J. Community structure in social and biological networks. Proc. Natl. Acad. Sci. USA 2002, 99, 7821–7826. [CrossRef] [PubMed]
38. Boston, D.; Mardenfeld, S.; Pan, J.S.; Jones, Q.; Iamnitchi, A.; Borcea, C. Leveraging Bluetooth co-location traces in group discovery algorithms. Pervasive Mob. Comput. 2014, 11, 88–105. [CrossRef]
39. Eagle, N.; Pentland, A.S. Eigenbehaviors: Identifying structure in routine. Behav. Ecol. Sociobiol. 2009, 63, 1689–1689. [CrossRef]
40. Kosta, S.; Mei, A.; Stefa, J. Large-Scale Synthetic Social Mobile Networks with SWIM. IEEE Trans. Mob. Comput. 2014, 13, 116–129. [CrossRef]
41. Goldberg, D.S.; Roth, F.P. Assessing experimentally derived interactions in a small world. Proc. Natl. Acad. Sci. USA 2003, 100, 4372–4376. [CrossRef] [PubMed]
42. Golik, P.; Doetsch, P.; Ney, H. Cross-entropy vs. squared error training: A theoretical and experimental comparison. Interspeech 2013, 13, 1756–1760.
43. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. J. Mach. Learn. Res. 2003, 3, 1157–1182.
44. MacKay, D.J.C. Information Theory, Inference & Learning Algorithms; Cambridge University Press: New York, NY, USA, 2002.