NextCell: Predicting Location Using Social Interplay from Cell Phone Traces

Daqiang Zhang, *Member*, *IEEE*, Daqing Zhang, Haoyi Xiong, Laurence T. Yang, and Vincent Gauthier

Abstract—Location prediction based on cellular network traces has recently spurred lots of attention. However, predicting user mobility remains a very challenging task due to the fuzziness of human mobility patterns. Our preliminary study included in this paper shows that there is a strong correlation between the calling patterns and co-cell patterns of users (i.e., ***co-occurrence in the same cell tower at the same time***). Based on this finding, we propose NextCell—a novel algorithm that aims to enhance the location prediction by harnessing the social interplay revealed in cellular call records. Moreover, our proposal removes the assumption held in previous schemes that binds locations of cell towers to concrete physical coordinates, e.g., ***GPS coordinates***. We validate our approach with the MIT Reality Mining dataset that involves 32,579 symbolic cell tower locations and 350,000 hours of continuous activity information. Experimental results show that NextCell achieves higher precision and recall than the state-of-the-art schemes at cell tower level in the forthcoming one to six hours.

Index Terms—Location prediction, social networks, cell towers, social interplay

1 Introduction

The rise of mobile devices and wireless communication has opened an avenue to explore individual location information. Mobile devices act as sensors that continuously acquire location information and other user digital traces [1]-[3]. In a broad sense, these traces reflect user mobility patterns in the form of cellular calls, short messages and other cell-ID related data logs. Out of the digital traces, location information at cell tower level is critical to many telecommunication network services, such as resource allocation [4]. Furthermore, with the prior knowledge of user location, telecommunication infrastructures can enable more accurate network services than ever.

A variety of schemes on location prediction have been proposed, e.g., ETP [5], Wherever [6], Necklace [7], HAMM [8], and TV [9]. These schemes appear to agree on several consensuses. Firstly, due to the privacy and security concerns, real-world and large-scale user mobility traces may not be available. Thus, most location prediction schemes were not tested using real-world mobility traces. Secondly, users exhibit spatial and temporal regularities. Every user periodically visits certain locations, e.g., *home and offices*. Users, however, comply with the mobility patterns only in a very loose manner. As a result, prediction schemes based on spatial and temporal regularities would not yield high-level precision. As uncovered in [10], the predictability of these schemes is bounded. Thirdly, some efforts [11], [12] have shown that social relationships have correlation with user's mobility pattern. However, they all require other data sources to identify user relationships for location prediction, e.g., *getting user relationship information from Facebook in* [13].

We examine the MIT Reality Mining dataset, which is one of most popular cellular traces, involving 1,000,000 GSM traces and 112,508 cellular calls. We find that the call patterns between two users are highly correlated with their encounters afterwards. Specifically, we compute the time interval between two consecutive user encounters in two cases: there is a phone call between two consecutive encounters (case 1), and there is no phone call in between (case 2). The time interval between consecutive encounters in case 1 is $\frac{1}{4}$ times of case 2 on average. Consequently, we introduce the notion of **social interplay** to reveal the social relationships embedded in the cellular call records. It can be defined as a convolution between entropy-characterized call strengths and the probability distribution of two users co-locating in the same cell.

Another challenge in using MIT Reality Mining dataset for location prediction is that locations of cell towers are represented by code names (i.e., *symbolic locations*). This makes some existing prediction schemes invalid, such as [6] and [7]. For a cell tower address "5188.41097" from the Reality Mining dataset, "5188" and "41097" are code names of a cell area and a cell tower, respectively. Apparently, they do not support arithmetic or logic operation, which is used to process GPS coordinates for location prediction.

To this end, we propose *NextCell*—a location prediction scheme that leverages the social interplay to enhance the location prediction precision. NextCell comprises two predictors: one is based on periodic behaviors and the other is based on social interplay. It further adopts a self-adaptive learner to aggregate the outputs of these two predictors. It is worthwhile to mention that NextCell uses symbolic cell tower

- *D. Zhang is with the School of Software Engineering, Tongji University, Shanghai 201804, China. E-mail: dqchang@gmail.com.*
- *D. Zhang, H. Xiong, and V. Gauthier are with the Institut Mines-Télécom/Télécom SudParis, CNRS UMR 5157 SAMOVAR, Evry 91000, France. E-mail: dqzhang@live.ca, {haoyixxiong, vincent.gauither}@gmail.com.*
- *L.T. Yang is with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China. E-mail: ltyang@stfx.ca.*

Manuscript received 05 May 2013; revised 16 Oct. 2013; accepted 04 Nov. 2013. Date of publication 19 Nov. 2013; date of current version 16 Jan. 2015.
Recommended for acceptance by Y. Pan.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TC.2013.223

locations for processing. To summarize, the main contributions of this paper are four folds.

- *NextCell* scheme extracts social interplay from telecommunication call records, which links human mobility with cellular call patterns.
- *NextCell* predicts user location by exploiting social interplay on top of the user periodic behaviors. It further presents a self-adjust learner so as to aggregate the outputs from previous two predictors.
- *NextCell* is able to handle symbolic representation of cell tower locations. This abates the complexity of detecting cell tower topology and transforming the code name of cell towers to computable coordinates.
- Experimental results demonstrate that *NextCell* achieves higher prediction precision and recall than the state-of-the-art schemes at cell tower level, where social interplay is crucial for achieving this significant improvement.

The rest of this paper is organized as follows. Section 2 briefly overviews the related work. Section 3 conducts empirical study on social interplay extraction. Section 4 introduces the design and implementation of *NextCell* scheme in detail. Section 5 reports the experimental results. Section 6 concludes the work.

2 Related Work

Several schemes that address the problem of user location prediction based on cellular phone records have been studied in recent years. Generally speaking, these schemes fall into two broad categories—*regularity-based schemes* and *non-regularity based schemes*.

2.1 Regularity-Based Schemes

These schemes take advantage of the temporal and spatial regularities that are exhibited in user daily lives [14]. They foresee user location by detecting periodic patterns in user traces. Periodic Mobility Model (PMM) in [12] was based on an intuition that the majority of human mobility was periodic among a small set of locations. Relying on the time of the day, PMM predicted user location that was in the location set. Various techniques in artificial intelligence and machine learning have been exploited to discover the mobility regularity, such as low-order Markov model [15], *hierarchical clustering technique* [7] and *non-linear time series technique* [16].

Regularity-based schemes implicitly assume that user mobility periodicity is static. However, users exhibit periodic mobility behaviors only in a very loose manner. For example, Nelson goes to his office every working day in the morning, but he may arrive at his office at any time between 8:00 AM to 11:00 AM. Furthermore, most of these schemes do not support location prediction using symbolic cell tower locations. Only based on user behavior periodicity, the predictability of periodicity-based prediction schemes is around 60% [10] on average. The predictability is far away from the theoretic upper bound. We checked the dataset and found it as around 50%. Actually, user movements are also influenced by irregular factors, e.g., social factors and location factors.

2.2 Non-Regularity-Based Schemes

The interaction among emerging research areas such as social networks, mobile computing, wireless communication and ubiquitous computing have fostered a new research direction for location prediction from the social perspective. Social network based schemes postulate that user mobility is partially driven by social relationships. For example, CM [17] introduced the social relationships into the mobility model. HAMM [8] extended the work in CM by incorporating preferred locations into location prediction. On top of HAMM, [11] proposed a new way to extract the contact lists in mobile devices as social relationships. Lards Backstroke et al. [13] predicted user location using user-supplied address data and the connections between members of the Facebook. All these schemes have demonstrated that social relationships play a fundamental role in user mobility.

In recent years, several user behavior studies based on mobile phone traces have been reported. In [3], the authors showed the existence of social relationships and the behavioral similarity among the frequent call pairs. Francisco Calabrese et al. [18] analyzed cellular data and found that 70% of users having reciprocal cellular calls would co-locate at the same cell tower. Both of them, however, neither modeled the social relationships based on call records, nor proposed any specific prediction model. In contrast, Enjoin Chop et al. [12] modeled user relationships as a function of distance that users would travel. Then, they computed the probability using the function that the user would move by the user current location available from checking-in websites. Additionally, there are other types of prediction schemes based on social networks, such as video deployment by microblog-based prediction [19] and online traffic prediction [20]. These schemes are tailored to the prediction of real-time applications using social networks, but they demonstrate the usefulness of social relationships.

To sum up, the aforementioned schemes have exploited external data sources other than the cellular call records to extract the social relationships. For example, the check-in information in Foursquare was used to deduce social relationship in [12], whereas the friendship information in Facebook was used in [13]. In this paper, we extract a kind of social relationships from telecommunication records, and we further use such relationships to enhance location prediction.

3 Empirical Study

To gain further insight, we conducted an empirical research on the social interplay, as well as its influence on user mobility patterns. For the sake of simplicity, we make the following definitions:

- **Call strength from user** u_a **to user** u_b (λ_{u_a,u_b}) is defined as the average call duration per week from u_a to u_b.
- **Call relative entropy between user** u_a **and user** u_b (e_{u_a,u_b}) is defined as:

$$e_{u_a,u_b} = -plog_{10}(p) - (1-p)log_{10}(1-p),$$

where $p = \lambda_{u_a,u_b}/(\lambda_{u_a,u_b} + \lambda_{u_b,u_a})$. The call entropy characterizes the degree of asymmetry between call strengths of u_a and u_b. The entropy would decrease when the call strength of both sides becomes significantly different, or in our definitions, asymmetric.

TABLE 1
Statistics of MIT Reality Mining Dataset

Item	Description
Starting time	09:27:41, 15-Jan-2004
Ending time	14:36:08, 20-Jul-2005
# of users	106
# of faculties	11
# of cell towers	32,579
# of areas	1,027
# of GSM traces	2,667,895
Avg. # of GSM traces per person per day	46.7
# of Mobile calls	112,508
# of mobile contacts	26,965
Logical location	AreaID.CellID
Physical coordinates	not available

- **Social interplay from user u_a to user u_b**[1] (δ_{u_a,u_b}) is defined as $\delta_{u_a,u_b} = \lambda_{u_a,u_b} * e_{u_a,u_b}$. The social interplay identifies the call strength weaken by the asymmetry between the call strength of both directions. This definition is primarily based on the common sense that two users who call each other symmetrically and frequently have strong connection.

Through preliminary experiments, we would like to answer two questions: (1) Do cellular calls affect user mobility? (2) Does the asymmetry of call strengths really exist, and how does social interplay affect user mobility?

We define two kinds of events between two users. These events are: *cellular call event* that stands for a directed call from one to another and *co-cell event* that refers to two users visiting the same cell tower at the same period. We select MIT Reality Mining data owing to its popularity in mobility prediction research community. The dataset consists of 112,508 cellular calls, and 350,000 hours of human behavior information including user location and co-location. Table 1 summarizes the statistics of the dataset.

3.1 Do Cellular Calls Affect User Mobility?

In order to answer this question, we present two concepts given a pair of users,

- **inter-contact time:** the time interval between two consecutive co-cell events of them, no matter mobile calls happen or not.
- **inter-call-contact time:** the time interval between two consecutive co-cell events, during which these two users have at least a phone call with each other. Inter-call-contact time is a special type of inter-contact time, it exists if some phone calls happen between the co-cell events.

These two concepts are illustrated in Fig. 1. Note that the inter-contact time of users having reciprocal calls does not mean "inter-call-contact time". It is the inter-contact time between user pairs who have reciprocal phone calls. Whereas the inter-call-contact time identifies the time duration from their call to contact of a user pair. Consequently, the inter-call-contact time exists only if there is a call between two consequent contacts between a user pair.

Fig. 2 shows the Cumulative Distribution Function (CDF) of the inter-call time, the inter-call-contact time and the inter-contact time among users who have called each other. The

1. In this paper, we use the term "social interplay" and "asymmetric social interplay" interchangeably in the contexts without ambiguity.

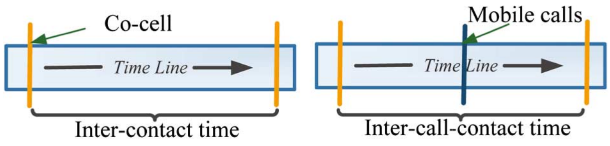


Fig. 1. Inter-contact time vs. inter-call-contact time.

inter-call-contact time is on average $\frac{1}{4}$ times of the inter-contact time. Around 80% of the inter-call-contact time occurs in 2 days, while the same amount of inter-contact time lasts 8 days. This phenomenon explains the common sense that people have higher probability to meet after cellular calls. This is because cellular calls drive call pairs to encounter after calls.

As shown in the comparison between the lines of "inter-contact time" and "inter-contact time of users having reciprocal calls", users with reciprocal calls often encounter more frequently than users without bilateral calls. This indicates that the social interplay considerably affects user mobility afterwards.

3.2 Does the Asymmetry of Call Strengths Really Exist and How Does Social Interplay Affect Human Mobility

This asymmetry is embodied in the call strengths. As shown in Fig. 3, the CDF lines of call pairs (4, 8) and (8, 4) represent the call strengths from the user 4 to the user 8 and vice versa. It clearly shows that the inter-call time of aggregated calls appears certain Poisson property. To be specific, we find that the number of inter-call time counted in disjoint intervals are independent, and the probability of the inter-call time counted in any time interval only relies on the length of the interval. Moreover, the inter-call time is asymmetric for the majority of user call pairs. This asymmetry can also be observed in individual telecommunication records. For a set of users, the incoming degree refers to the number of calls made from this user set to all users, and the outgoing degree denotes the number of calls made to this use set. In general, the degrees of incoming and outgoing calls are unequal with averagely 10% or more discrepancy, shown as node degree in Fig. 4. This indicates that users keep asymmetry in their

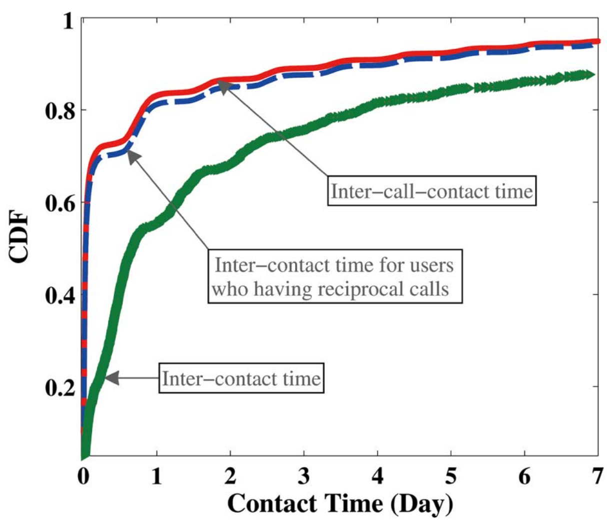


Fig. 2. CDF of inter-call-contact time and inter-contact-time.

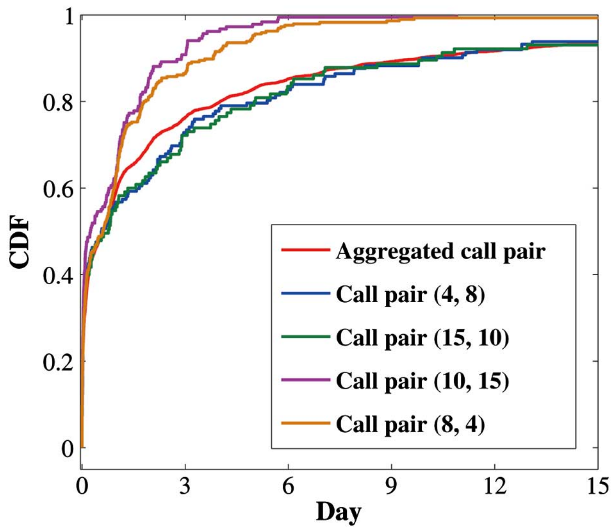


Fig. 3. Inter-call-time with or without direction.

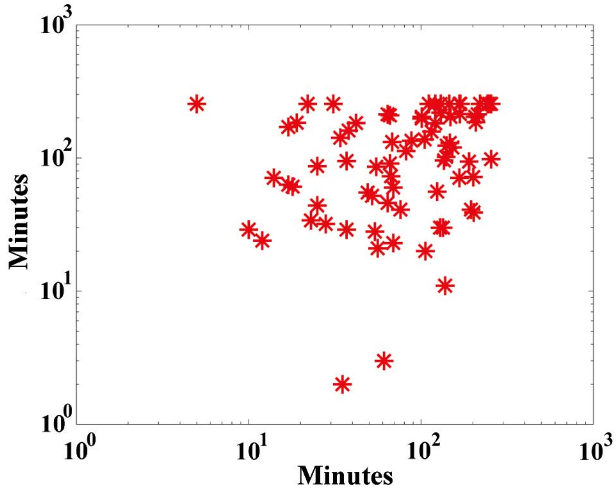


Fig. 5. Reciprocal call relationships revealed in both sides of calls (loglog), where the Pearson correlation coefficient of X and Y axes is 0.4583.

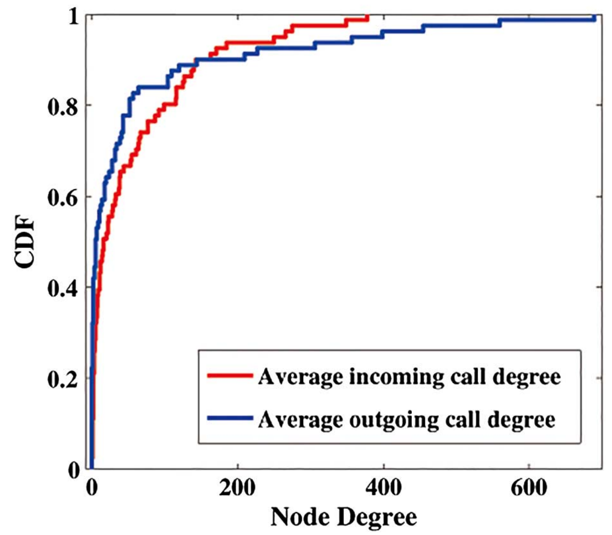


Fig. 4. Node degree diagram.

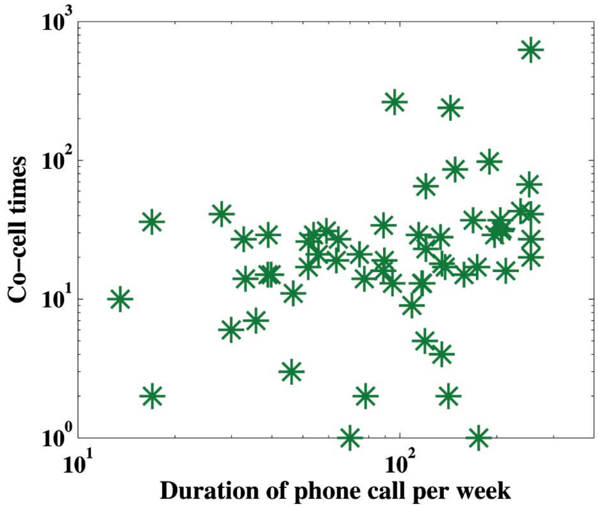


Fig. 6. Cellular calls vs. co-cell events (loglog), where the Pearson correlation coefficient of X and Y axes is 0.3478.

social interplay. Note that the interplay asymmetry in telecommunication records is evident in the diagrams of edge degree, call frequency and call durations. As these diagrams share the same phenomenon with the diagrams of node degree and call strengths, we hereby omit them.

We measure the asymmetry of social interplay for the call logs of both directions in every call pair. We plot the correlation of the average call duration per call pair, illustrated in Fig. 5. The two direction call strengths of calls are correlated as their Pearson coefficient is 0.4583. Note that Pearson correlation indicates the degree of linear dependency between two variables, and 0.4583 usually means a strong positive correlation. Hence, such diagram shows that the call strengths from both directions are positively correlated. This reveals a phenomena that, for each user pair, once the call strength from one-side is large, the call strength from the other side is usually large. Finally, we study in what degree co-cell events are affected by social interplay. Fig. 6 illustrates that the times of co-cell and the average call duration (*per week*) are positively correlated. It shows that the user pairs have strong social interplay may have more chances to co-cell than others.

4 NextCell Scheme: A Novel Location Prediction Scheme

In this section, we introduce *NextCell* scheme that predicts user location from cellular traces at cell tower level. We overview the scheme and then report its design in detail.

4.1 Overview

The goal of *NextCell* is to predict user location at cell tower level in the forthcoming one to six hours. It consists of two subgoals: how many cell towers users will go, and which cell towers they will reach.

Fig. 7 illustrates the architecture of *NextCell* scheme, involving three separate components: *periodicity predictor*, *social interplay predictor* and *self-adjust learner*. *NextCell* uses the first component to handle user mobility regularity originating from spatial and temporal perspectives. It uses the second component to cope with the non-periodic mobility behaviors caused by social interplay. Then, *NextCell* uses an adaptive self-learner to aggregate the outputs of two predictors. The notations used in the design of *NextCell* are given in Table 2.

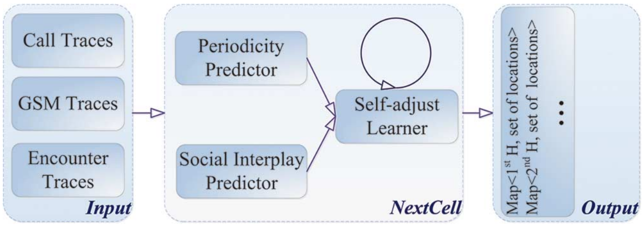


Fig. 7. The design of *NextCell* scheme.

4.2 Periodicity Predictor

This predictor foretells user location by exploiting user behavior periodicity. However, GSM traces in Reality Mining dataset represent cell tower locations using symbols. E.g., *two symbolic locations 5119.40811 and 2412.7131 are two locations extracted from the dataset*. These two locations are code names. Therefore, any arithmetic and logical operation over them is invalid.

To this end, we design and implement a periodicity detector—**Perio** scheme. With an individual user's historical GSM trace as input, Perio method extracts a "time-table" of the user's locations. The time-table maps a time-relevant locations sorted by the probability of visiting. Perio tunes the probability distribution of all possible locations in a time instant through minimizing the Kullback-Leibler Divergence between the location distribution sequences and all observed mobility traces in GSM traces. Given a forthcoming prediction time interval, Perio gives the most probable location in the corresponding list as the prediction result of a user's mobility.

To be specific, *Perio* consists of three steps—detecting periods in user movement, discovering periodic movement behaviors, and forecasting user movement using periodic behavior patterns. In the first step, *Perio* extracts a set of locations as reference locations that are frequently visited by users. To be accurate, it only removes the locations visited once. Then, *Perio* generates multiple periods in the movement through reference locations using a hybrid method of autocorrelation and Fourier transform. In the second step, *Perio* presents a probabilistic model to characterize the periodic behaviors. Thus, it obtains periodic behaviors are statistically generated from partial movement sequences. Finally, *Perio* predicts the user location by using periodic behaviors.

TABLE 2
Notations Used in the Design of *NextCell* Scheme

Name	Explanation
τ	the time interval
u_a, u_b	users a and b
c	a cell tower
C	the set of cell tower
(c, r)	the element pair in the predicted results and r is the probability that the user will visit c
λ	call intensity (i.e the call strength)
Γ	normalized call strength matrix
$\mathbf{I}$	an identity matrix
A^T	the matrix transpose
Δ	the matrix of intransitive social interplay
Θ	the matrix of transitive social interplay
$\bullet\div$	the matrix element-wise division
O^b	the matrix of interval from the last co-cell encounter for each user pair
V_τ	the matrix of each two user pair would encounter within an interval of τ
Λ	the matrix of call intensity each element $\lambda_{i,j}$ referring the call strength from user i to j
L	the number of cell towers in the predicted results
P^τ	the predicted results of periodicity predictor at τ
S^τ	the predicted results of interplay predictor at τ
$P_i^\tau(c,r)$	the element at the i^{th} index of the P^τ

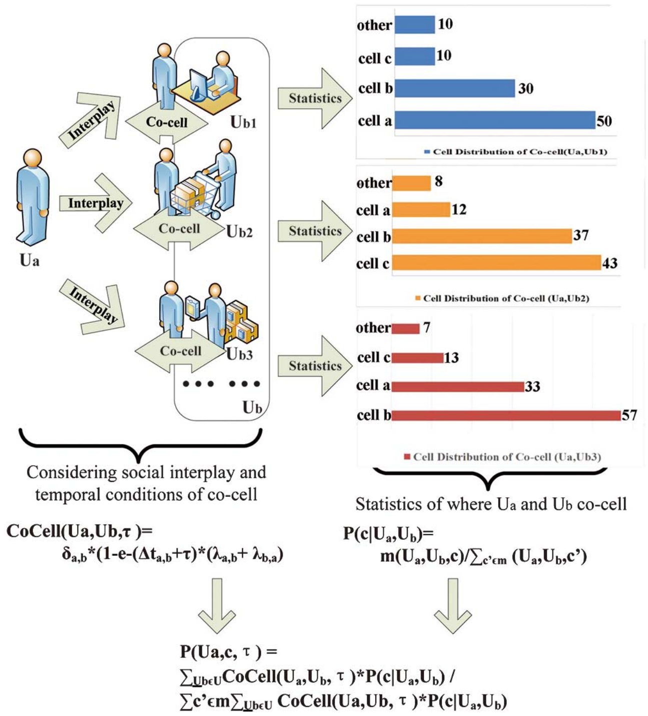


Fig. 8. The framework of social interplay predictor.

4.3 Social Interplay Predictor

This predictor is a component in *NextCell* scheme, which forecasts user location from the social relationship perspective. With all user's GSM traces as input, it firstly discovers the social interplay between arbitrary two users. Then it maps a time-relevant list of possible locations traversed by the user. Such a list is sorted by the probability that the user will move to the corresponding cell tower to meet/co-cell with a friend. Given a forthcoming prediction time interval τ, a user u_a and a full set of all cell towers C, the social interplay predictor makes prediction. To be specific, it will foretells a list of cell towers that might be visited in the forthcoming prediction interval τ. Therefore, the predictor result should be:

$$C_{pred} = \arg\max_{c \in C} P(c|u_a, \tau), \tag{1}$$

where the C identifies the set of all cell towers, and $P(c|u_a,\tau)$ denotes the probability of u_a co-celling with other users at cell tower c at time τ.

In order to compute the above probability, we propose a probabilistic framework shown in Fig. 8. We take two factors into account:

- **Who will co-cell in a given time?** As defined in empirical study, social interplay indicates the possibility that two users co-cell. Hence we propose *social interplay with temporal constraints* ($CoCell(u_a, u_b, \tau)$) to estimate the likelihood that any two users u_a and u_b co-cell at given time τ. $1 - e^{-(\Delta t_{a,b}+\tau)*(\lambda_{a,b}+\lambda_{b,a})}$ identifies the time constraint. We share the same design of time constraint as [21]. Suppose two users have not co-celled for time $\Delta t_{a,b}$, social interplay based predictor tries to compute the possibility that they would co-cell at interval τ in future.

$$CoCell(u_a, u_b, \tau) = \left(1 - e^{-(\Delta t_{a,b}+\tau)*(\lambda_{a,b}+\lambda_{b,a})}\right)*\delta_{a,b}, \quad (2)$$

whereas $\delta_{a,b}$ is the social interplay between user u_a and u_b; which is defined as "call strength weaken by the asymmetry of call directions", it is computed as Eq. 3. To characterize the asymmetry between u_a and u_b, we use the entropy between the call strength of two directions. If u_a and u_b call each other with equal strength, then the entropy gains maximum $-log(0.5) = 0.301$. This shows that u_a and u_b has a strong reciprocal social ties and has more opportunities to co-cell. Otherwise, an asymmetric call relationship will lead entropy to be zero, indicating that even though they may have many calls, they will not co-cell frequently.

$$\begin{aligned} \delta_{a,b} &= e_{a,b}*\lambda_{a,b} \\ &= (-\gamma_{a,b}*log(\gamma_{a,b}) - \gamma_{b,a}*log(\gamma_{b,a}))*\lambda_{a,b}, \\ \textbf{where}\ \gamma_{a,b} &= \frac{\lambda_{a,b}}{\lambda_{a,b}+\lambda_{b,a}}, \gamma_{b,a} = \frac{\lambda_{b,a}}{\lambda_{a,b}+\lambda_{b,a}}, \end{aligned} \quad (3)$$

In summary, Eqs. 2 and 3 reflect the user pairs who have strong social interplay ($\delta_{a,b}$) and have not co-celled for longer time ($\Delta t_{a,b}$) would have higher possibility to co-cell in near future τ.

- **Where will they co-cell?** After knowing who will co-cell at given time, we are particularly interested in where they will probably co-cell. We characterize the probability that users u_a and u_b stay at c when they co-cell by *cell distribution of co-cell* $P(c|u_a, y_b)$, shown in Eq. 4.

$$P(c|u_a, u_b) = \frac{m(u_a, u_b, c)}{\sum_{c' \in \mathrm{C}} m(u_a, u_b, c')}, \quad (4)$$

where C is the full set of cell towers in which users u_a and u_b may co-locate, and $m(u_a, u_b, c')$ represents the times that the two users co-locate at the cell tower c'.

Based on the above two factors, the probability of a user will appear at a cell tower is computed as Eq. 5:

$$P(c|u_a, \tau) = \frac{\sum_{u_b} P(c|u_a, u_b)*CoCell(u_a, u_b, \tau)}{\sum_{c' \in C}\sum_{u_b} P(c'|u_a, u_b)*CoCell(u_a, u_b, \tau)}, \quad (5)$$

where c is a cell tower, u_a and u_b are two users. As shown in Eqs. 2 and 3, the estimation of the *CoCell* (u_a, u_b, τ) including $\delta_{a,b}$ and *temporalconstraints* is key to measure the influence of social interplay. Our framework needs to calculate the social interplay for each of user pairs (u_a, u_b); therefore we propose a three-step matrix computation process as follows:

1. Normalizing the original call strength matrix Λ by Poisson property [22] as Γ,
2. Building the social interplay matrix Δ from normalized call strength matrix Γ,
3. Estimating the CoCell matrix Φ_τ each element of which identifies *CoCell*(u_a, u_b, τ) of a user pair.

$$\Lambda = \begin{bmatrix} 0 & \lambda_{1,2} & \cdots & \lambda_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{n,1} & \lambda_{n,2} & \cdots & 0 \end{bmatrix} \qquad \Gamma = \begin{bmatrix} 0 & \frac{\lambda_{1,2}}{\lambda_{1,2}+\lambda_{2,1}} & \cdots & \frac{\lambda_{1,n}}{\lambda_{1,n}+\lambda_{n,1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\lambda_{n,1}}{\lambda_{1,n}+\lambda_{n,1}} & \frac{\lambda_{n,2}}{\lambda_{2,n}+\lambda_{n,2}} & \cdots & 0 \end{bmatrix}$$

$$\Delta = \begin{bmatrix} 0 & \delta_{1,2} & \cdots & \delta_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n,1} & \delta_{n,2} & \cdots & 0 \end{bmatrix} \qquad \Phi_\tau = \begin{bmatrix} 0 & CoCell(u_1,u_2,\tau) & \cdots & CoCell(u_1,u_n,\tau) \\ \vdots & \vdots & \ddots & \vdots \\ CoCell(u_n,u_1,\tau) & CoCell(u_n,u_2,\tau) & \cdots & 0 \end{bmatrix}$$

Fig. 9. An example of matrix computation.

Fig. 9 gives an example to illustrate the three-step matrix computation process with our probabilistic framework characterized by Eqs. 2, 3 and 5.

4.3.1 Normalizing Call Strengths by Poisson Process

Given that the inter-call time on both sub-processes of a call (e.g., *calls from user 4 to user 8, and calls from user 8 to user 4*) share the Poisson properties. We separate the call processes between every call pair with the total call intensity (i.e., *call strengths*) λ into two directed sub-processes by a factor p. Thus, $p \cdot \lambda$ and $(1-p) \cdot \lambda$ are the intensities of two sub-processes, respectively. We calculate the call strength as the average duration of cellular call per week which is the estimated parameter of Poisson distribution.

Let $\mathbb{F}$ be the function that splits two sub-processes as a matrix Γ under a constraint that the sum of the elements $\Gamma_{i,j}$ and $\Gamma_{j,i}$ is 1, where i and j are the row index and the column index. The matrix Γ is given as Eq. 6, where $\bullet\!\div$ is an element-wise division in matrix (i.e., *given two matrices, the division is performed by each element on a matrix to the corresponding element on another matrix*).

$$\Gamma = \mathbb{F}(\Lambda, \Lambda^T) = \Lambda \bullet\!\div (\Lambda + \Lambda^T). \quad (6)$$

Then, we get the normalized call strengths. Given that the normalized call strengths do not consider the influence of call durations, we incorporate this influence into the measurement of social interplay in the following step.

4.3.2 Building Social Interplay Matrix by Call Relative Entropy

According to the definition in our empirical study, we select the entropy as the basis of measurement. We introduce the operator $\mathbb{E}$ to measure the relative entropy between two probability matrices: Γ and Γ^T. The relative entropy is given as Eq. 7.

$$\mathbb{E}(\Gamma, \Gamma^T) = -\Gamma \cdot \log_{10}(\Gamma) - \Gamma^T \cdot \log_{10}(\Gamma^T). \quad (7)$$

The matrix of social interplay is computed as Eq. 8.

$$\Delta = \mathbb{E}(\Gamma, \Gamma^T) \cdot \Lambda. \quad (8)$$

Algorithm 1. The Design of the Adaptive Learner in *NextCell* Scheme from the t Interval to the $t+1$ Interval.

Input:

$Real^{t+1}$, the list of cell towers users really visit at interval $t+$,

i.e., $\{<c_1, r_1>, \cdots, <c_i, r_i>, \cdots, <c_l, r_l>\}$;

Output:

Pre^{t+1}, a top-L list of predicted cell towers for interval $t+1$,
$= \{ <c_1, r_1>, \cdots, <c_i, r_i>, \cdots, <c_l, r_l> \}$

1 **begin**
2 $cnt = 1$; /*the counter for cell towers*/
3 $\beta = 1$;
4 $\gamma = 1$;
5 /*Get a list of possible cell towers for a user in next prediction interval $t+1$*/
6 **for each** cell tower $c \in P^t$ and S^t **do**
7 $j \leftarrow$ GetCellIndex (P^t, c);
8 $n \leftarrow$ GetCellIndex (S^t, c);
9 $r_{cnt} = \beta * P_j^t(c, r) + \gamma * S_n^t(c, r)$
10 $Pre^{t+1}(cnt) \leftarrow \{c, r\}$;
11 $cnt++$;
12 /* Sort prediction results*/
13 sort(Pre^{t+1}, $''desc''$);
14 /*Get the top-L cell towers*/
15 $Pre^{t+1} \leftarrow$ TruncateList $(Pre^{t+1}(1:L))$;
16 /*Update β, γ for make prediction at $t+2$ interval*/
17 **for each** cell tower $c \in Real^{t+1}$ **do**
18 $j \leftarrow$ GetCellIndex (P^t, c);
19 $n \leftarrow$ GetCellIndex (S^t, c);
20 $j > 0\,?\,\beta = \sqrt{\beta * P_j^t(c, r)} : \beta = 0.0001$;
21 $n > 0\,?\,\gamma = \sqrt{\gamma * S_n^t(c, r)} : \gamma = 0.0001$;

4.3.3 Estimating CoCell from Social Interplay and Temporal Constraint

Recall that the co-cell positively correlates with the times of encounters and the social interplay. We take the temporal constraint into consideration of co-cell events. Suppose there are n users in the system, Eq. 9 computes the temporal constraint matrix (V_τ), where O^b is the matrix of the interval from the last co-cell encounter for every user pair. Each element at a^{th} row and b^{th} column of the matrix V_τ denotes the temporal constraint estimating the possibility that users u_a and u_b co-cell at interval τ in future.

As discussed in Eq. 2, we have the multiply of temporal constraint and social interplay. Therefore, in Eq. 10, we convolute[2] the asymmetric social interplay matrix Θ with temporal constraint matrix V_τ, and then normalize the convolution result as a *CoCell* matrix Φ. Thus, the element of $\Phi_\tau(a, b)$ at the a^{th} row and the b^{th} column is the probability **CoCell**(u_a, u_b, τ).

$$V_\tau = [1]_{n\times n} - \exp([-1]_{n\times n} \cdot (O^b + [\tau]_{n\times n}) \cdot (\Lambda + \Lambda^T)), \tag{9}$$

$$\Phi_\tau = (V_\tau \cdot \Theta) \bullet\!\div ((V_\tau) \cdot \Theta) \times [1]_{n,n} \Big). \tag{10}$$

According to Eqs. 2 and 10, we have proposed an algorithm to calculate the *social interplay with temporal constraint* and *cell distribution of co-cell* for each user pair. Thus far, we have estimated the probability that user moves to a place at an interval driven by social interplay using Eq. 5.

4.4 A Self-Adjust Learner

As we mentioned before, human mobility is affected by individual behavior patterns and co-location behaviors. We need to consider both individual mobility patterns and the social interplay to conduct mobility prediction. To aggregate the two prediction results generated by periodicity predictor and social interplay predictor, *NextCell* presents a self-adjust learner. The learner involves two steps: *Top-L* selection of cell towers and prediction aggregation.

In step one, *NextCell* generates two lists of cell towers P^τ and S^τ from predictors of periodicity and social interplay for the prediction interval τ. Each element in the list is a pair (c, r) denoted in Table 2. In step two, *NextCell* makes the prediction by aggregating the two lists using a boosting technique. It assigns two normalized parameters β and γ for these two lists. Suppose a cell tower appears at the i^{th} and j^{th} place in P^τ and S^τ. Then, the probability of this cell tower that will be visited is computed as:

$$R = \beta \cdot P_i^\tau(c, r) + \gamma \cdot S_j^\tau(c, r). \tag{11}$$

By aggregating all the elements in these two lists, we easily get the top-L recommendations at time interval t. For the time interval $t+1$, we aim to boost the influence of the predictor that correctly predicts, and decrease the influence of the predictor that wrongly predicts. Therefore, we introduce two aggregation parameters, i.e., β and γ, in the proposed self-adjust learner. At the beginning, both β and γ are set to 1, assuming that the contributions of the two predictors are equal. During the prediction process, both β and γ are automatically adjusted according to the relationship between the location ground truth and the predicted results of Perio or social interplay. If the predicted top-L results do not comply with the location ground truth, the corresponding parameter β and γ will be set to 0.0001 (it is a value selected by trial-and-error, it should not be set to 0 because the future value of this parameter is proportional to the current value). If some of the predicted top-L results comply with the location ground truth, the corresponding parameter β and γ will be increased as a function of previous values. Lines 7-15 show the process of reproducing the prediction list of cell towers at time interval $t+1$ moment.

5 System Evaluation

In this section, we conduct a series of experiments to evaluate *NextCell* at cell tower level in the forthcoming one to six hours. Particularly, we would like to answer:

2. We define the element-wise multiply as convolution operator in our predictor design.

TABLE 3
The Overall Precision of *Perio* and *Social Interplay* Schemes on the MIT Reality Data Set, Indicating that the *Perio* Outperforms *Social Interplay*-Based Predictor for 6% More Precision at the First Hour Prediction

Forecast span	***Perio***	***Interplay***
the 1^{st} hour precision	0.5651	0.5021
the 2^{nd} hour precision	0.4855	0.4619
the 3^{rd} hour precision	0.4734	0.4453
the 4^{th} hour precision	0.4505	0.4353
the 5^{th} hour precision	0.4318	0.4149
the 6^{th} hour precision	0.4154	0.3989

- How is the stand-alone performance of our *Social-Interplay-based* predictor and *Perio*? Does *Social-Interplay-based* predictor outperform *Perio* in terms of precision and recall?
- How is overall performance of *NextCell* scheme? Does it work better than the state-of-the-art scheme (*Perio*) or *Social-Interplay-based* predictor in terms of precision and recall?
- How is the performance of *NextCell* scheme in the best/worst cases?
- Does *NextCell* scheme precisely predict the number of cell towers visited by users?
- How much does the asymmetric social interplay affect the prediction precision of the proposed scheme? Shall we use the symmetric social interplay to implement a *NextCell* scheme and evaluate it?
- How does *NextCell* scheme tune the aggregation parameters in the self-adjust learner?

5.1 Metrics

We select two metrics to measure the performance of the *NextCell* scheme—precision and recall. As Eq. 12 defined, $recall_{i,j}$ represents the prediction recall in the i^{th} hour of the user j, where $P_{i,j}$ is the number of the correctly predicted cell towers, $R_{i,j}$ is the number of cell towers that the observed user really traverses in the i^{th} forecast interval, I is the set of forecast intervals (i.e., *{1,2,3,4,5,6}* in experiments).

$$recall_{i,j} = \frac{P_{i,j}}{R_{i,j}}, i \in I. \tag{12}$$

The other is the precision that denotes the ratio of the number of visited cell towers in the prediction result to the total number of predicted cell towers. For example, *NextCell* predicts 5 cell towers for a certain hour, but only 3 towers are visited. Thus, the precision is 0.6. Formally, the precision is given as:

$$precision_{i,j} = \frac{P'_{i,j}}{P_{i,j}}. \tag{13}$$

TABLE 4
The Overall Recall of *Perio* and *NextCell* Schemes on the MIT Reality Data Set, Indicating that the *Perio* Scheme Averagely Achieves 11% More Recall than *Social Interplay*-Based Predictor for First Hour Prediction

Forecast span	***Perio***	***Interplay***
the 1^{st} hour recall	0.5765	0.4664
the 2^{nd} hour recall	0.5180	0.4366
the 3^{rd} hour recall	0.4857	0.4236
the 4^{th} hour recall	0.4625	0.4148
the 5^{th} hour recall	0.4428	0.3953
the 6^{th} hour recall	0.4263	0.3805

TABLE 5
The Overall Precision of *NextCell* Scheme and *Sym-NextCell* (Its Symmetric Interplay Variant) on the MIT Reality Data Set

Forecast span	***NextCell***	***Sym-NextCell***
the 1^{st} hour precision	0.7428	0.6250
the 2^{nd} hour precision	0.6691	0.5570
the 3^{rd} hour precision	0.6289	0.5202
the 4^{th} hour precision	0.5997	0.4936
the 5^{th} hour precision	0.5740	0.4715
the 6^{th} hour precision	0.5510	0.4524

where $P'_{i,j}$ is the subset of $P_{i,j}$ in which each cell tower is visited by the user.

5.2 Experiment Design

To predict use location by behavior periodicity, we present and implement a periodicity predictor *Perio* scheme. The design of the *Perio* is given in Section 4.2. Note that *Perio* scheme does not take social influence into consideration. We select the MIT Reality Mining dataset as the raw dataset. Its statistical information is given in Table 1. Note that we understand that the proposed scheme can be better evaluated by using different dataset, unfortunately due to the privacy concern of Telecom operators, there is still no large-scale mobile phone data available for public access and research evaluation. The MIT dataset is the largest mobile phone dataset we can get, which contains one-year more records of 100+ users. Even without the new dataset, we still expect significant improvement of precision, as the social interplay is a fundamental driver that affects user behaviors.

5.3 Overall Performance

In this section, we compare the *NextCell* with baseline approaches. First of all, *Perio* was selected as a main baseline, since it is the state-of-the-art predictor based on periodical mobility. Furthermore, we are also interested in the performance of social interplay predictor when it works standalone. Therefore social interplay based scheme *Interplay* was also considered. Recap that we find that the social interplay is either symmetric or asymmetric. Thus, to evaluate the influence by asymmetric social interplay, we propose the third baseline approach *Sym-NextCell* i.e., a variant of *NextCell* with symmetric social interplay. The comparison between *Sym-NextCell* and *NextCell* will be discussed in Section 5.6.

We conduct several experiments for all participants to evaluate the overall performance. We use the aggregated precision and aggregated recall for overall performance. They are computed as the average values of the precision and recall for all participants, respectively.

Tables 3 and 4 illustrate the overall average precision and recall values of *Perio* and *Interplay* based scheme in the

TABLE 6
The Overall Recall of *NextCell* Scheme and *Sym-NextCell* (Its Symmetric Interplay Variant) on the MIT Reality Data Set

Forecast span	***NextCell***	***Sym-NextCell***
the 1^{st} hour recall	0.6929	0.6353
the 2^{nd} hour recall	0.6187	0.5685
the 3^{rd} hour recall	0.5789	0.5313
the 4^{th} hour recall	0.5503	0.5046
the 5^{th} hour recall	0.5256	0.4813
the 6^{th} hour recall	0.5036	0.4619

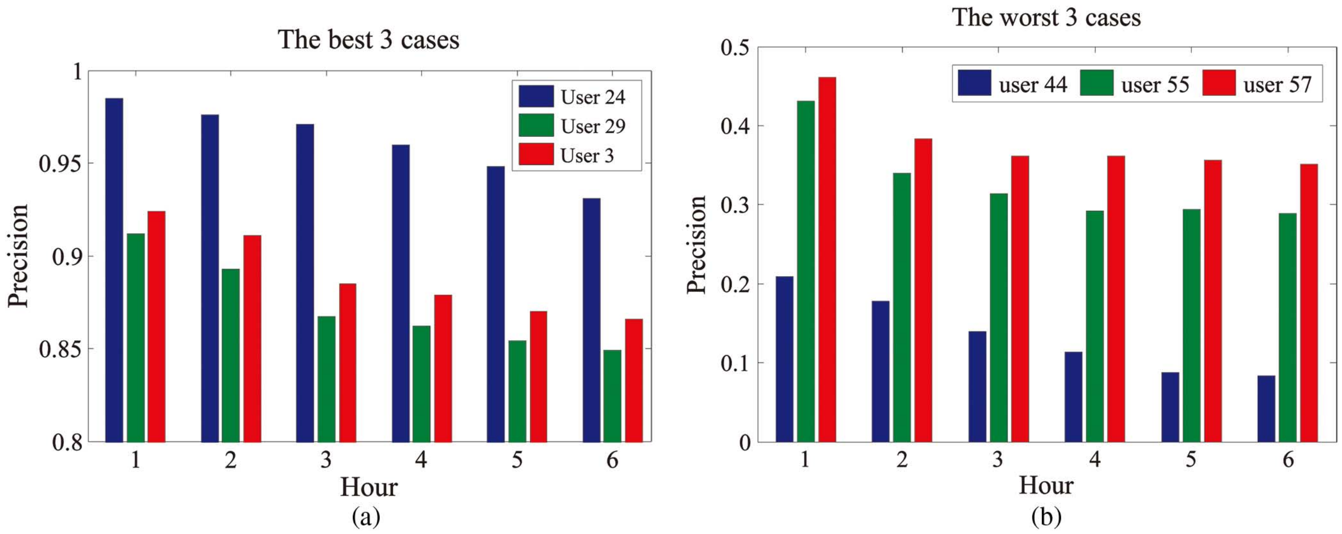


Fig. 10. Case study on *NextCell* scheme: (a) the best 3 cases and (b) the worst 3 cases.

forthcoming one to six hours. These two tables reveal two facts. Firstly, *Perio* scheme achieves 6% more precision than *Interplay* scheme for first hour prediction. Meanwhile, *Perio* obtains 11% more recall on average than *Interplay* for the first hour prediction. However the advantage of *Perio* decreases with the forecast span. This indicates that *Perio* outperforms *Interplay* in terms of prediction precision and recall. Secondly, the values of the precision and the recall for the same prediction hour are close for both *Interplay* and *Perio* schemes. For instance, for the 1st hour, the precision and recall for *Perio* are 0.5651 and 0.5765. This attributes to that a user usually visits one cell tower in an hour, and they also predict one cell tower.

Tables 5 and 6 give the aggregated precision and the aggregated recall of *NextCell* scheme. For all the forecast span, *NextCell* achieves higher precision and recall than *Perio* and *Interplay* schemes. For the 1st and 6th prediction hour, *NextCell* obtains 30%higher precision than both *Perio* and *Interplay* schemes. In contrast, the proposed scheme achieves around 20% recall improvement than the two baseline schemes. To summarize, *NextCell* scheme make improvement over *Perio* with respect to precision and recall. Considering the design of *Perio* and *NextCell* schemes, we easily infer that the improvement is attributed to the social interplay revealed in cell phone logs.

5.4 Best/Worst Cases Analysis

Intuitively, the extreme cases could give us more inspiration to improve the proposed scheme. Hereby, we pick up the top 3 best cases and the top 3 worst cases, and analyze them in detail.

Fig. 10 reports the best cases and the worst cases in *NextCell* experiments where the users get the highest or the lowest precision in the forthcoming one to six hour prediction. In the best cases, *NextCell* reaches averagely 92% precision, up to 96% precision. We check the daily behaviors of the users 24, 29 and 3, and find that they exhibit high level behavioral regularity such that *NextCell* precisely forecasts their locations in the next one to six hours. As a matter of fact, many users in the MIT Reality data set who share the similar behaviors have frequent connections. Thus, it is appropriate to predict their location purely by social interplay.

In contrast, user 44 exhibits a very low level regularity in the worst cases. The precision declines rapidly, because he is very passionate with behavior variation. We have identified 5 cases like user 44. In summary, we conclude that *NextCell* exhibits a precision decrease as the forecast interval increases. Yet, in most time, it declines in a slow and smooth manner.

5.5 Does *NextCell* Scheme Precisely Predict the Number of Cell Towers Visited by Users

The number of cell towers to be traversed in the forecast interval has a significant impact on prediction precision. However, the number of cell towers visited by users varies with the time. Therefore, we choose the aggregated error rate as the metric. Recall that $P_{i,j}$ is the number of cell towers that the user i really passes by in the j interval, while $R_{i,j}$ is the corresponding number of cell towers that *NextCell* predicts. Thus, the error rate is computed as the aggregated ratio for all participants of the absolute difference between $P_{i,j}$ and $R_{i,j}$ to $R_{i,j}$.

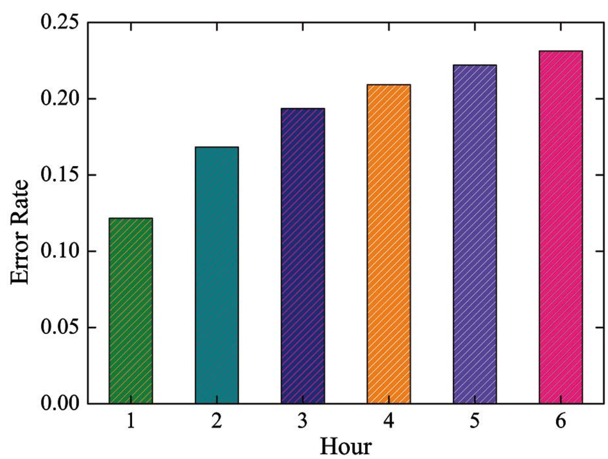


Fig. 11. The aggregated error rates of *NextCell* scheme on predicting the number of cell towers the users may traverse.

TABLE 7
The 1st-6th Prediction Precision for Users 3, 15 and 101

Forecast span	User 3	User 15	User 101
the 1^{st} hour precision	0.8937	0.6218	0.5927
the 2^{nd} hour precision	0.8602	0.4934	0.4339
the 3^{rd} hour precision	0.8596	0.4480	0.3422
the 4^{th} hour precision	0.8582	0.3935	0.3010
the 5^{th} hour precision	0.8666	0.3894	0.2656
the 6^{th} hour precision	0.8247	0.4001	0.2471

We do experiments to check the aggregated error rates for all participants of the proposed scheme. Fig. 11 shows the aggregated error rate CDF diagram of *NextCell*. The x-axis is the prediction hour after the critical calls, and the y-axis is the error rates of the prediction. This figure shows that in the 1st prediction hour, the prediction rate is 87%, indicating that the prediction results are around one cell tower less or more cell towers than the real cell towers that users visit. With the increase of the prediction hour, the error rate shows a slow upward trend, e.g., up to 23.51% for the 6th hour.

5.6 The Influence of Asymmetric Social Interplay on Prediction Precision

Section 3 has identified the asymmetry feature of the social interplay. This asymmetry is measured by call direction and call strength in that direction. In the following, we do not consider call direction in the *Sym-NextCell* scheme. This scheme shares almost everything as *NextCell* scheme except that the computation of the social interplay is simply as 1 when there are calls between user pairs.

By comparing Tables 5 and 6, *NextCell* achieves higher precision and recall than *Sym-NextCell* scheme for all prediction hours. In the 3rd hour, *NextCell* gets 20.90% and 9.06% higher precision and recall than *Sym-NextCell*. The comparison results imply that *NextCell* performs better than *Sym-NextCell* for both recall and precision. Considering that these two schemes are almost the same except the calculation of social influence, we find that the asymmetric social interplay contributes much more to the prediction than the plain call strength. In addition, the findings demonstrate that user relationship are embodied in cellular call records with strong influence on user movement.

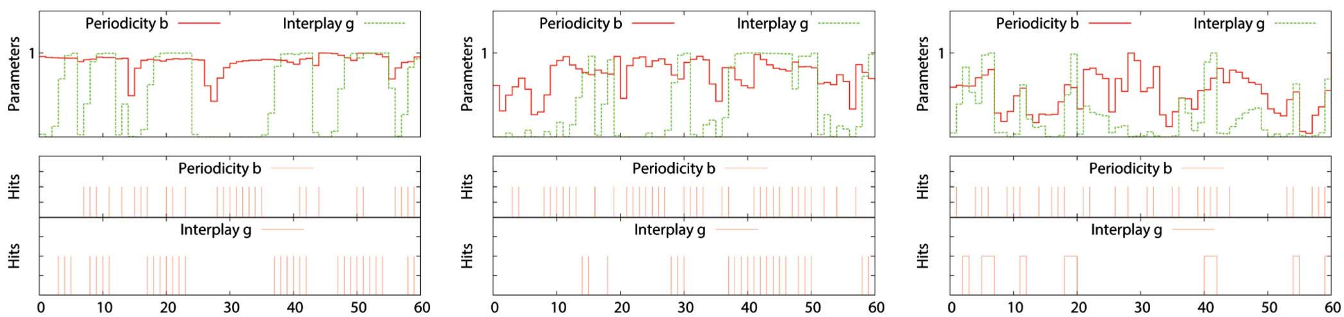
5.7 How Does *NextCell* Tune the Parameters in the Self-Adjust Learner

NextCell involves two parameters in the self-adjust learner for prediction aggregation. This section reports the parameter adaption in experiments.

Apparently, the parameter settings vary with each individual as each user behaviors exhibit high-level individuation. When the ground truth arrives, the parameter tuning activates. When the predictor based on the social interplay correctly predicts user location at the current hour, it has a large probability of making correct predicting at the next hour. Thus, the value of its parameter in the self-adjust learner will be boosted in the next hour. We extract three types of users from the dataset as examples: users 3, 15 and 101. Table 7 gives the overall prediction precision achieved by *NextCell* for the selected users. It indicates that user 3 is the most regular person who exhibits a high level of regularity, while the user 101 is the person who frequently varies his/her behaviors.

The aggregation parameter tuning diagram corresponding to three users 3, 15 and 101 given in Table 7 is shown in Fig. 12. For each user, Fig. 12 provides two sub-figures. In the upper sub-figure, the x-axis is the prediction hour, and the y-axis the parameter value. As the prediction hour increases, we see that the values of β and γ vary dramatically. In the lower sub-figure, the x-axis is also the prediction hour, and the y-axis shows the "hits" corresponding to correct prediction of the specific predictor. A red line is marked for β when the periodicity-based prediction is correct. When red lines are simultaneously drawn for β and γ, both periodicity-based prediction and social interplay based prediction are correct. These two sub-figures imply that user mobility is affected by the mixture of behavior periodicity and social interplay.

We summarize the parameter variation in the following perspectives. Firstly, *NextCell* scheme employs an adaptive mechanism of parameter adaptation shown in Algorithm 1 to achieve high precision. Secondly, social interplay significantly contributes to the forecast of user location. Finally, user behaviors are the convergence of periodicity, social interplay and randomness.



(a) The sensitivity of parameters β and γ for user 3

(b) The sensitivity of parameters β and γ for user 15

(c) The sensitivity of parameters β and γ for user 101

Fig. 12. The performance of *NextCell* scheme with varying the values of fusion parameters β and γ for users 3, 15 and 101. For each user, we provide two sub-figures. One is the value variation of parameters, and the other is hit diagram, indicating when the periodicity-based prediction and the social interplay based prediction correctly predict user location.

6 Conclusion

Location prediction is a hot research topic in recent years, yet it still remains to be fully addressed. We find that the co-cell patterns are highly correlated with the call patterns. Inspired by the findings, we have exhibited the social interplay revealed in cellular calls that can be exploited for user location prediction. We further propose a location predictor scheme called *NextCell*, which predicts user location at cell tower level in the forthcoming one to six hours. NextCell contains two different components: the behavior periodicity based predictor and social interplay based predictor. It also presents a self-adjust learner that aggregates the outputs from the two predictors. Moreover, *NextCell* removes an assumption held in existing schemes that cell tower addresses are associated with physical coordinates. Extensive experimental results in this paper show that *NextCell* can achieve 20% higher prediction accuracy on average.

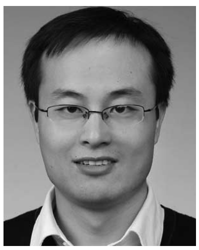
NextCell could be further improved. We plan to investigate the incorporation of short message logs into *NextCell* scheme. We also plan to evaluate the proposed scheme in a wide range of user traces.

Acknowledgments

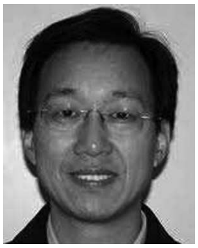
This work is supported by EU FP7 project SOCIETIES (Grant 257493). This work is also supported in part by the National Natural Science Foundation of China (Grant 61103185), in part by Natural Science Foundation of the Higher Education Institutions of Jiangsu Province, China (Grant 11KJB520009), in part by the 9th Six Talents Peak Project of Jiangsu Province (Grant DZ-043), and in part by Major Program of National Natural Science Foundation of Jiangsu Province (Grant BK2011005). The authors would like to thank Dr. Sarah Gallacher for her constructive comments and suggestions.

References

[1] N. Eagle and A. Pentland, "Reality mining: Sensing complex social systems," *Pers. Ubiquitous Comput.*, vol. 10, pp. 255–268, 2006.
[2] N. Eagle, "Behavioral inference across cultures: Using telephones as a cultural lens," *IEEE Intell. Syst.*, vol. 23, no. 4, pp. 62–64, Jul./Aug. 2008.
[3] N. Eagle, A. S. Pentlandb, and D. Lazerc, "Inferring social network structure using mobile phone data," *Proc. Nat. Acad. Sci.*, vol. 106, no. 36, pp. 15274–15278, 2009.
[4] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54–69, Jul. 2005.
[5] L. Backstrom, E. Sun, and C. Marlow, "Evaluating mobility models for temporal prediction with high-granularity mobility data," in *Proc. 19th IEEE Int. Conf. Pervasive Comput. Commun. (PERCOM)*, 2012, pp. 206–212.
[6] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "WhereNext: A location predictor on trajectory pattern mining," in *Proc. 15th ACM Special Interest Group Knowledge Discovery Data Mining (SIGKDD)*, 2009, pp. 637–646.
[7] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, and A. T. Campbell, "Nextplace: A spatio-temporal prediction framework for pervasive systems," in *Proc. 9th Int. Conf. Pervasive Comput.*, 2011, pp. 152–169.
[8] C. Boldrini and A. Pass, "Modelling spatial and temporal properties of human mobility driven by users' social relationships," *Comput. Commun.*, vol. 33, pp. 1056–1074, 2010.
[9] W. J. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling time-variant user mobility in wireless mobile networks," in *Proc. 27th IEEE Int. Conf. Comput. Commun. (INFOCOM)*, 2007, pp. 758–766.
[10] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.
[11] T. Hossmann, T. Spyr, and F. Legendre, "Putting contacts into context: Mobility modeling beyond inter-contact time," in *Proc. 12th ACM Int. Symp. Mobile Ad Hoc Netw. Comput. (MobiHoc)*, 2011, pp. 199–208.
[12] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proc. 17th ACM Special Interest Group Knowledge Discovery Data Mining (SIGKDD)*, 2011, pp. 1082–1090.
[13] L. Backstrom, E. Sun, and C. Marlow, "Find me if you can: Improving geographical prediction with social and spatial proximity," in *Proc. 19th Int. Conf. World Wide Web (WWW)*, 2010, pp. 61–70.
[14] S. Scellato, A. Noulas, and C. Mascolo, "Exploiting place features in link prediction on location-based social networks," in *Proc. 17th ACM Special Interest Group Knowledge Discovery Data Mining (SIGKDD)*, C. Apté, J. Ghosh, and P. Smyth, Eds., Aug. 2011, pp. 1046–1054.
[15] L. Song, D. Kotz, R. Jain, and X. He, "Evaluating next-cell predictors with extensive wi-fi mobility data," *IEEE Trans. Mobile Comput.*, vol. 5, no. 12, pp. 1633–1649, Dec. 2006.
[16] Z. Li, B. Ding, J. Han, R. Kays, and P. Nye, "Mining periodic behaviors for moving objects," in *Proc. 16th ACM Special Interest Group Knowledge Discovery Data Mining (SIGKDD)*, 2010, pp. 1099–1108.
[17] M. Musolesi, S. Hailes, and C. Mascolo, "An ad hoc mobility model founded on social network theory," in *Proc. 7th ACM Int. Symp. Model. Anal. Simul. Wireless Mobile Syst. (MSWiM'04)*, 2004, pp. 20–24.
[18] F. Calabrese1, Z. Smoreda, V. D. Blondel, and C. Ratti, "Interplay between telecommunications and face-to-face interactions: A study using mobile phone data," *PLoS ONE*, vol. 6, no. 7, pp. 114–119, 2011.
[19] Z. Wang, L. Sun, C. Wu, and S. Yang, "Guiding internet-scale video service deployment using microblog-based prediction," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, A. G. Greenberg and K. Sohraby, Eds., 2012, pp. 2901–2905.
[20] B. Zhang, K. Xing, X. Cheng, L. Huang, and R. Bie, "Traffic clustering and online traffic prediction in vehicle networks: A social influence perspective," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, 2012, pp. 495–503.
[21] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: A social network perspective," in *Proc. 10th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2009, pp. 299–308.
[22] A. Holroyd, R. Lyons, and T. Soo, "Poisson splitting by factors," *Arxiv preprint*, arXiv:0908.3409, 2009.



Daqiang Zhang received the BSc degree in management science and MSc degree in computer science from Anhui University, Hefei, China, in 2003 and 2006, and the PhD degree in computer science from Shanghai Jiao Tong University, China, in 2010. From May to October 2006, he was a full-time software engineering at Autodesk Inc., Shanghai. From June 2008 to June 2009, he was a jointly-supervised PhD candidate at Hong Kong Polytechnic University. From September to November 2010, he was an analyst at Goldman Sachs Inc., Beijing. From July 2011 to July 2012, he was a post-doctoral researcher at Institute Telecom, France. Currently, he is an associate professor in the School of Software Engineering, Tongji University, Shanghai, China. His research interests include mobile computing, distributed computing, and wireless sensor networks. He has published more than 60 papers in major journals and international conferences in those above areas, including *IEEE Transactions on Parallel and Distributed Systems*, *IEEE Transactions on Emerging Topics in Computing*, *IEEE Network Magazine*, *ICPP*, *ICC*, and *WCNC*. He received the Best Paper Award from ACCV'2009 and UIC'2012. He is in the editorial board in *European Transactions on Telecommunications* (Wiley publisher), *International Journal of Big Data Intelligence* (Inderscience publisher), *KSII Transactions on Internet and Information Systems* (Korea Society of Internet Information), and *New Review of Hypermedia and Multimedia* (Taylor & Francis publisher). He is a member of CCF.



Daqing Zhang received the BSc degree from China University of Mining and Technology, in 1984, the MSc degree from China Aeronautical Computing Technique Institute, in 1987, and the PhD degree from University of Rome "La Sapienza" and University of L'Aquila, Italy, in 1996. He is a professor in Ambient Intelligence and Pervasive System Design at Institut Mines-Télécom/Télécom SudPais, and CNRS, France. He initiated and led the research in smart home, healthcare/elderly care, and context-aware computing from 2000 to 2007 at the Institute for Infocomm Research (I2R), Singapore. He was the founding head of Connected Home Lab, from 2003 to 2004, the founding head of Context-Aware Systems Department, from 2004 to 2006, and the founding head of IMACA group, from 2006 to 2007, at I2R, Singapore. He is the associate editor in four international journals including *ACM Transactions on Intelligent Systems and Technology*. He also served in the technical committee for top ubiquitous computing conferences such as UbiComp, Pervasive, and PerCom. His research interests include context-aware computing, mobile social networks, ambient assistive living, big data analytics, and urban computing. He has published over 170 papers in referred journals, conferences and books, where his research work on context model and middleware got more than 4000 citations in the past years. He is the winner of the Ten Years CoMoRea Impact Paper Award at IEEE PerCom 2013, the Best Paper Award at IEEE UIC 2012, and the Best Paper Runner Up Award at Mobiquitous 2011. All his research has been motivated by practical applications in digital cities, mobile social networks, and elderly care. In recent years, he has been exploring a new research direction called "Social and Community Intelligence (SCI)", which aims at revealing the individual group behaviors, social interactions, as well as community dynamics by mining the digital traces left by people while interacting with cyber-physical spaces. It is expected that SCI has the potential to push context-aware computing to new frontiers.

Haoyi Xiong received the MSc degree in information technology from the Hong Kong University of Science and Technology and BEng degree in electrical engineering and automation from Huazhong University of Science and Technology, China. He is currently a PhD student supervised by Prof. D. Zhang in ALPS Group of Institut Mines-Télécom, TELECOM SudParis (ex. INT), and Université Pierre et Marie CURIE Paris VI. He is working closely to Dr. V. Gauthier and Prof. Monique Becker. His research interests include mobile crowdsensing, participatory sensing, spatial–temporal pattern mining, and mobility prediction. He has got the Best Paper Award at UIC'2012.

Laurence T. Yang graduated from Tsinghua University, China. He recieved the PhD degree in computer science from the University of Victoria, Canada. He joined St. Francis Xavier University, Antigonish, Canada, in 1999. His current research includes parallel and distributed computing, embedded and pervasive computing. He has published many papers in various refereed journals, conference proceedings, and book chapters in these areas (including around 100 international journal papers such as *IEEE Journal on Selected Areas in Communications, IEEE Transactions on Systems, Man, and Cybernetics, IEEE Transactions on Very Large Scale Integration Systems, IEEE Transactions on Industrial Informatics, IEEE Transactions on Information Technology in Biomedicine, IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Circuit and Systems, IEEE Transactions on Service Computing, ACM Transactions on Embedded Computing Systems, IEEE Systems Journal, ACM Transactions on Autonomous and Adaptive Systems, IEEE Transactions on Vehicular Technology, ACM/Springer Journal on Mobile Networks and Applications, IEEE Intelligent Systems, Journal of Parallel and Distributed Computing*, etc.). He has been involved actively in conferences and workshops as a program/general/steering conference chair (mainly as the steering co-chair of IEEE UIC/ATC, IEEE CSE, IEEE HPCC, IEEE/IFIP EUC, IEEE ISPA, IEEE PiCom, IEEE EmbeddedCom, IEEE iThings, IEEE GreenCom, etc.) and numerous conference and workshops as a program committee member. He served as the vice-chair of IEEE Technical Committee of Supercomputing Applications (TCSA) until 2004, currently is the chair (elected in 2008 and 2010) of IEEE Technical Committee of Scalable Computing (TCSC), the chair of IEEE Task force on Ubiquitous Computing and Intelligence. He is also in the steering committee of IEEE/ACM Supercomputing conference series, and the National Resource Allocation Committee (NRAC) of Compute Canada. In addition, he is the editors-in-chief of several international journals. He is serving as an editor for many international journals. He has been acting as an author/co-author or an editor/co-editor of many books from Kluwer, Springer, Nova Science, American Scientific Publishers, and John Wiley & Sons. He has won several Best Paper Awards [including IEEE Best and Outstanding Conference Awards such as the IEEE 20th International Conference on Advanced Information Networking and Applications (IEEE AINA-06), etc]; one Best Paper Nomination; Distinguished Achievement Award, 2005; and Canada Foundation for Innovation Award, 2003. He has been invited to give around 20 keynote talks at various international conferences and symposia.

Vincent Gauthier received the BS degree in electrical engineering from the University de Bretagne Occidentale, France, in 2002, and the MS and PhD degrees in electrical engineering and computer networks from the University of Paris 6, France, in 2003 and 2006, respectively. He was a guest researcher at National Institute of Standards and Technology, Massachusetts, between 2006 and 2008. He joined the Faculty of Telecom SudParis and the Lab CNRS SAMOVAR (UMR 5157), Evry, France, in 2008, where he is currently an associate professor of the Department of Wireless Networks and Multimedia Services. His current research interests include wireless networks, sensor networks, ad-hoc networks, cross-layer design, self-organization in wireless networks, and cooperative communications. His other research interests include mobility modeling, performance analysis, and queuing theory. He has published more than 40 papers in international journals and conferences, such as *IEEE Transactions on Wireless Communication, IEEE Transactions on Vehicular Technology,* and *IEEE Journal on Selected Areas in Communications*.