A Trajectory-based Recruitment Strategy of Social Sensors for Participatory Sensing

Fei Hao, Mingjie Jiao, Geyong Min, Laurence T. Yang

***Abstract*—Participatory Sensing, a promising sensing paradigm, enables people to collect and share sensor data of the phenomena of interest using mobile devices across many applications, such as smart transportation and air quality monitoring. This article presents a framework of participatory sensing and then focuses on a key technical challenge: developing a trajectory-based recruitment strategy of social sensors in order to enable service providers to identify well-suited participants for data sensing based on temporal availability, trust and energy. To devise a basic recruitment strategy, Dynamic Tensor Analysis (DTA) algorithm is initially adopted to learn the time-series tensor of trajectory so that the users' trajectory can be predicted. To guarantee the reliable sensing data collection and communication, the trust and energy factors are taken into account jointly in our multi-objective recruitment strategy. In particular, the friend-like social sensors are also defined to deal with the emergency during the participatory sensing. An illustrative example and experiment are conducted at a university campus to evaluate and demonstrate the feasibility and extensibility of the proposed recruitment strategy.**

***Index Terms*—Social Sensors, Participatory Sensing, Dynamic Tensor Analysis, Trajectory**

I. Introduction

The popularity of mobile devices and the rapid development of wireless sensing technology advance the emerging of a novel pervasive data sensing paradigm–Participatory Sensing (PS) [1], [2], which allows citizens to sense their surrounding environment voluntarily with their available sensoring devices, *e.g.*, smart phones, and share the information with other citizens through the existing Internet communication infrastructure. Participatory Sensing Systems (PSSs) have tremendous potential in various applications, such as environmental monitoring [3], intelligent transportation [4], and route planning [5]) because they collect sensing data by virtue of the participatory power of ordinary citizens.

The major difference between participatory sensing and traditional sensing lies in that each participant is regarded as a sensor, namely *social sensor*, sensing the surrounding environment to upload data. The analysis ability and mobility of participants in PSSs would greatly reduce the burden on the system and enlarge the geographical coverage of sensing. However, participants as the data collection carriers of the system, are demanded to sense anytime and anywhere, which impedes the wide use of participatory sensing. Furthermore, the participants are mostly interested in or related to the sensing campaign. The number of participants is not considerably large and just allows participatory sensing to be applied in a small range.

Consider a real scenario of Particulate Matter 2.5 (PM 2.5) real-time monitoring in Beijing City. In reality, there are insufficient air quality measurement stations in a city due to the expensive cost of building and maintaining such a station. For example, 35 air quality measurement stations are currently established in Beijing city. Since these stations are stationary base stations with the traditional network coverage mechanism, they cost lots of money and manpower. Generally, an air quality measurement station needs a certain size of land, huge amount of money (about 200,000 USD for construction and 30,000 USD per year for maintenance [6]), human resources to regularly take care of it, and 24 hours per day power consumption. Thus, this fact greatly limits the number of measurement stations. However, we expect to obtain the measured values of air quality in PSSs through the mobile sensing devices held by crowd and further aggregate these values for the purpose of intelligent services supply. In particular, the price of a hand-held PM 2.5 sensing device powered by lithium battery (10 W) is 500 USD. At the worst case, the employers (e.g., environmental protection agency) buy devices for users who are willing to sense the air quality voluntarily/incentively. Roughly, 14,210 users can be recruited with the same cost consumed by the traditional sensing system every year for their participatory sensing campaign. From the sustainability point of view, the PS paradigm is better than the traditional sensing paradigm in terms of both cost and energy.

There exist some prior research work focused on reputation-based, trust-based, and expertise-based recruitment schemas [1], [8]. Different from those existing participants recruitment approaches, this article is to present a holistic recruitment strategy which considers various impact factors to participatory sensing. Therefore, a participatory sensing framework in PSSs is proposed. The proposed framework works as follows: firstly, the historical trajectories of participants are analyzed for extracting the potential social sensors in the sensing layer; secondly, the social sensors are dynamically selected in the sensing layer; thirdly, the social sensors voluntarily/incentively sense their surrounding environment and upload these collected data in the servers. Upon this proposed framework, a trajectory-based recruitment strategy of social sensors which considers the availability, trust and energy of users is devised. To avoid the missing of sensing data, an emergency selection scheme is also proposed to enhance the usefulness of our recruitment strategy of social sensors.

The remainder of this article is structured as follows. Section II presents a participatory sensing framework and provides the problem addressed in PS. A trajectory-based recruitment strategy of social sensors for PSSs is proposed in Section III. Following an illustrative example studied in Section IV, Section V concludes this article.

II. Participatory Sensing Framework and Problem Statement

This sections provides a typical framework of participatory sensing and then presents the problem addressed in this article.

A. A Framework of Participatory Sensing

Figure 1 presents a typical participatory sensing framework in PSSs. This framework is divided into four layers in which different functionalities are enabled. We elaborate the functions and responsibilities of each layer by a bottom-up view approach. 1) **Sensing Layer:** Considering the limited budget and dynamical behaviors of users, a dynamic recruitment strategy of social sensors need to be proposed in this layer in terms of the users' availability, trust value, remaining power of their mobile phones and emergency context. 2) **Data Transmission Layer:** This layer transmits the obtained sensing data to the data center for further processing by the Internet. Before accessing the Internet, mobile users may send the sensing data through either WLAN or Cellular network. 3) **Data Processing Layer:** This layer manages the responsibilities of data aggregation, redundant data filtering, as well as data mining and so on. By processing the obtained sensing data, some relevant services can be provided for application layer and also the recruitment strategy might be adjusted according to these data in sensing layer. 4) **Application Layer:** The various data services in this layer are obtained from the processing layer. For example, the existing services of *Vehicle Navigation System*, *Weather Information* and *Health Tracking* have a wide use in our daily life.

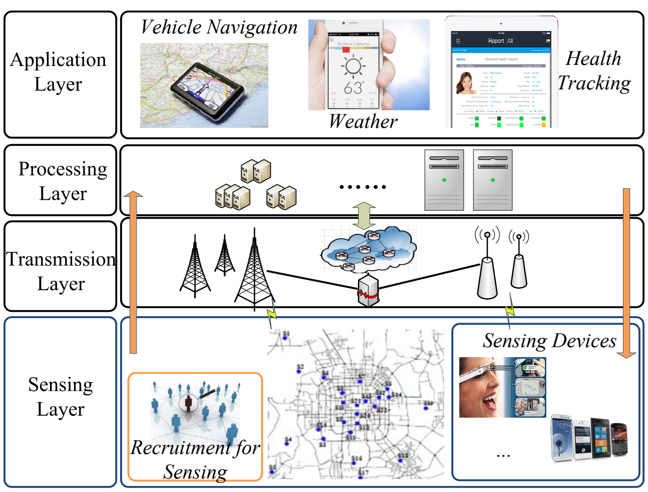


Fig. 1. Participatory Sensing Framework.

B. Problem Statement

In this section, the related definitions in PSSs are introduced at first. Then, the problem statement is described.

- **Social Sensor**: In a PSS, a social sensor is actually a participant who is willing to collect data about a particular phenomenon. Therefore, we use the participant, user and social sensor interchangeably in this paper.
- **Trajectory Similarity**: In the daily life, the behavioral trajectories collected by each participant, more or less have some overlaps between them. In other words, there exists a situation that some of participants appear in the same data collection point (DC-point) at the same time. To quantify this case, the Trajectory Similarity is defined. High similarity of trajectories among some participants implies that they might have some social relations in some extents, such as roommates, classmates or family. In this article, we call the participants who have the similar trajectories as a group of "*Friends-Like Social Sensors*".
- **Participatory Sensing System**: The essence of participatory sensing is data collection and interpretation. Participation requirements allow a campaign organizer (service provider) to recruit the participants that have a certain level of experience or have been available in a certain time-spatial space. Participation metrics include: a) the number of campaigns volunteered for, b) the number of campaigns accepted for, c) the number of campaigns participated in, and d) the number of campaigns abandoned. Individual metrics can be associated with other information about a campaign, such as size, lifetime, and type of sensing required. For example, some potential participants who have been selected for traffic-sensing campaigns in the past 3 months in a certain area.

Generally, a successful participatory sensing campaign is dependent on two main issues: 1) how to build an efficient recruitment strategy of social sensors 2) how to make an incentive mechanism for motivating these social sensors to participate in the sensing campaign based on the recruitment strategy. Our work in this paper focuses on devising a recruitment strategy in which each social sensor is dynamically selected and assigned to a set of DC-points where data should be collected. In this section, we formally describe this problem as follows.

(**Problem Statement**) Given an area and a group of possible social sensors U with their mobile traces, the entire recruitment strategy of social sensors is composed of the following technical aspect: For a campaign $C(G, T)$, with G as the set of the DC-points, and T as the set of time of data sensing and collection determined by the campaign requirements. Thus, the recruitment problem in the recruitment layer is to dynamically select each participant $u \in U$ to any DC-point located in G, such that the location of u is closer to the DC-point in its sensing range g than to that of any others in U. Then, those recruited participants will carry out $C(G, T)$ at required time and location for campaign organizers.

III. A Trajectory-based Recruitment Strategy of Social Sensors

Participation sensing can effectively replace the stationary base stations by recruiting participants. By predicting the trajectory data of participants, it can help us to select the appropriate participants for joining the participatory sensing campaign.

A. Big Picture

The proposed recruitment strategy of social sensor is working within a given monitoring area with M DC-points and N users. As shown in Fig. 2, our recruitment strategy of social sensor contains the following three steps:

Step 1 **Trajectory Data Collection and Tensorization**: The trajectory data of users associated with users, time and location is represented with a tensor $\chi \in \Re^{I_t \times I_g \times I_u}$. We collect the trajectory data within i days. The trajectory data in the i^{th} day is a tensor χ_i. Therefore, the collected data is represented by a time-series tensor $\chi^T = \{\chi_1, \chi_2, \cdots, \chi_i\}$.

Step 2 **Tensor-based Data Training**: The time-series tensor χ^T is trained by Dynamic Tensor Analysis (DTA) approach [7]. Then, an approximate tensor $\widetilde{\chi}$ is obtained.

Step 3 **Prediction and Friends-like Social Sensors Identification**: Based on the obtained approximate tensor, the user's future moving patterns or expected arrival locations are predicted. Further, Euclidean distance is adopted to measure the similarity among their moving patterns of social sensors within a period. Finally, we cluster these users who have high similarity of moving patterns into a group of friend-like social sensors.

Step 4 **Social Sensors Selection**: Generally, we dynamically select the optimal social sensors who can satisfy the availability, trust and energy constraints at each time. According to the historical trajectory, we infer the availability of each social sensor appearing nearby the DC-point using the approximate tensor. During the interaction of participatory sensing, the trust and energy of each social sensor are taken into account timely for adjusting the selection results. In particular, when the emergency happened, such as the lower power of sensing devices, personal urgent affairs and so forth, the participants contained in the group of friend-like social sensors will become the selection candidates. Then, the selected social sensors are stimulated to participate in a given sensing campaign.

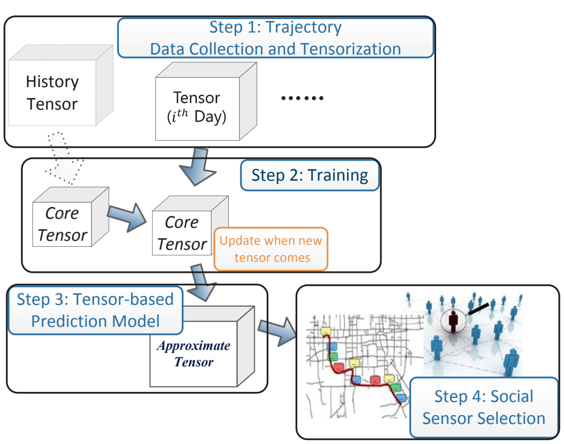


Fig. 2. The overview of our recruitment strategy of social sensors

With this overview of our proposed recruitment strategy of participants, the following sections present the detailed strategy with tensor based DTA algorithm. A tensor, as a type of high dimension matrix which governs the correlations among these dimensions is widely used in many applications. In PSSs, the trajectory data of a certain period is regarded as a type of high-dimensional tensor which is associated with users, time, and location. These dimensions constructed as a tensor are important for representation, processing and storage of PSSs. Hence, this strategy can help us to discover the potential semantic relationships from those data and provide intelligent services for the participatory sensing campaign.

B. Trajectory Data Tensorization and Collection

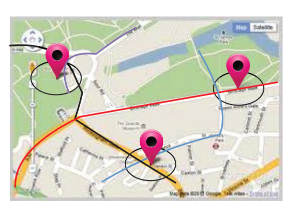
For a given monitoring area, there are N users who are doing their daily activities in this area. In order to tensorize the trajectory data of these N users, we first position and determine the virtual sensing range of those M DC-points, and then collect the daily trajectory data with k time intervals. Apparently, the trajectory data generating from PSSs is mainly composed of three dimensions: *users*, *time*, *locations* (Note that the longitude and altitude of a certain location correspond to a certain pre-partitioned grid). Element of the daily trajectory data can be described as a 4-tuple $a =< T, G, U, V >$ where T is time, G refers to the grid where the users are staying, U denotes a certain user, and V is the element's value. This 4-tuple corresponds to a 3-order tensor as $\chi \in \Re^{I_t \times I_g \times I_u}$, where $\Re$ is defined on the real number domain. I_t, I_g, I_u refer to time, location grids, and users. $I_t \times I_g \times I_u$ denotes the Cartesisan product of each individual domain. The value of each element $x(t_k, g_m, u_n)$ in the 3-order tensor represents the likeliness of user u_n is staying at grid g_m at time t_k which is obtained by GPS-enabled devices.

In other words, a user can only belong to a grid at a certain time. Hence, the constructed tensor including the daily trajectory data is very sparse as shown in Figure 3.

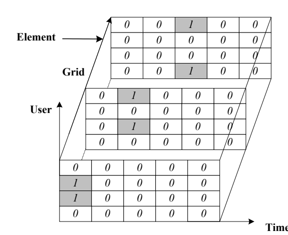
Before the trajectory data training, we construct the time-series tensor $\chi^T = \{\chi_1, \chi_2, \cdots, \chi_i\}$ by collecting i days' daily trajectory data of users.

C. Trajectory Data Training

Dynamic Tensor Analysis (DTA) [7] is an efficient algorithm for dynamically revealing the hidden correlations among the dimensions (e.g., time, users, and locations in PSSs) of



(a) Monitoring Area with Grids



(b) Tensor Representation for Trajectory

Fig. 3. Trajectory Data Tensorization

the tensor. Therefore, we adopt DTA algorithm to analyze the time-series tensor X^T and mine the potential trajectory patterns of users in PSSs.

An initial tensor can be matricized in several modes which are determined by their orders. For example, tensor $\chi \in \Re^{I_t \times I_g \times I_u}$ has three unfolding matrices that can be decomposed into a projection matrix $U_{(d)}$ and an energy matrix $S_{(d)}$ via Singular Value Decomposition (SVD) for the corresponding mode d . Then, a covariance matrix $C_{(d)}$ can be calculated with $U_{(d)}$ and $S_{(d)}$.

The DTA algorithm processes each mode of the tensor continuously. Importantly, the $C_{(d)}$ is updated as $C_{(d)} \leftarrow \lambda C_{(d)} + X_{(d)} X_{(d)}^T$, where $\lambda \in [0,1]$ is a forgetting factor regarded as the predictable information aggregator of time-series data. In other words, the recent time stamps are more important than those far in the past. Then, we decompose the above updated covariance matrix $C_{(d)}$ to obtain the principle eigenvectors that are used for computation of core tensor at the following step.

The aforementioned method for calculating the core tensor is just a training process focused on one time stamp. Clearly, the core tensor is updating dynamically. The calculation of the core tensor is equivalent to learning the historical tensors.

Eventually, an approximate tensor $\widetilde{\chi}$ used in tensor-based prediction model is the product of the core tensor and three metrics.

D. Trajectory Prediction and Friend-like Social Sensors Identification

The approximate tensor $\widetilde{\chi} \in \Re^{I_t \times I_g \times I_u}$ is actually an information aggregator of the results learnt from the previous time-series tensors. Each element value in $\widetilde{\chi}$ denotes the likeliness of a user appearing in a certain grid at a certain time. Therefore, the future trajectory can be predicted by the approximate tensor.

Further, we attempt to identify the friend-like social sensors by using Euclidean distance which powerfully measures the distance between two corresponding points located in the trajectories of the 3-order tensor. In other words, the similarity between any two trajectories of different social sensors can be easily estimated according to the Euclidean distances $sim(\widetilde{\chi}(:,:,u), \widetilde{\chi}(:,:,v)) = \sum_{i=0}^{n}(1/(Dis(\widetilde{\chi}(g_i,:,u), \widetilde{\chi}(g_i,:,v))+1)/n$ where the $\widetilde{\chi}(:,:,u)$ denotes the trajectory of social sensor u in the future; $Dis(X,Y)$ refers to the Euclidean Distance between vector X and Y; and $sim(\widetilde{\chi}(:,:,u), \widetilde{\chi}(:,:,v)) \in (0,1]$.

Based on the above similarity measurement approach, if the similarity between two trajectories of social sensor is greater than a given threshold γ, i.e, $sim(\widetilde{\chi}(:,:,u), \widetilde{\chi}(:,:,v)) \geq \gamma$, then, u and v are regarded as the friend-like social sensors reciprocally.

E. Social Sensors Selection

Since some users may be selected for the settled time and grid, we have a selection strategy to pick up the better volunteers. At each time, we can dynamically select the well-suited social sensors who can satisfy the constraints of

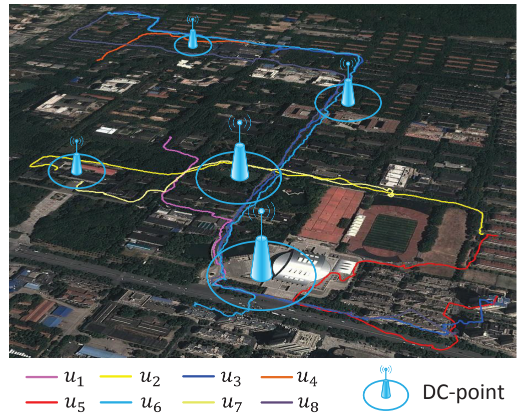
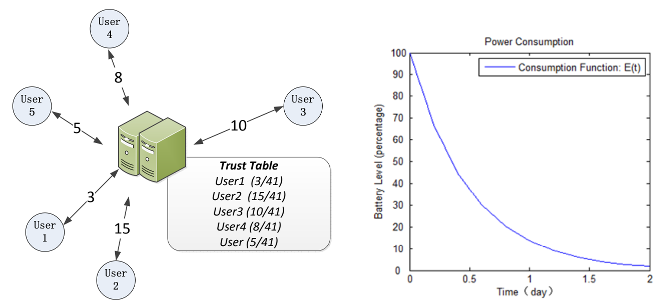


Fig. 4. Trajectory of five typical volunteers and three assistant volunteers on 5 NOV



(a) Trust Evaluation (b) Energy Decay Curve

Fig. 5. Evaluation of Trust and Energy Consumption

availability, trust and energy. For example, the availability can be reasoned by the approximate tensor. We rank the likeliness value of the users on the decrease, and choose the top-k users that can be the potential social sensors. Based on this idea, the following two strategies: *1) basic selection strategy* and *2) multi-objective selection strategy* are devised, respectively.

1) Basic selection strategy: As a basic selection strategy, the availability is a critical factor to be considered. For the targeted grid at a time, the availability based selection strategy of social sensors is dependent on the likeliness ($x^{'}(t,g,u) \in \widetilde{X}$) rank of the users appearing in the targeted area at that time.

2) Multi-objective selection strategy: From the data reliability point of view, the factors of energy and trust which is used to evaluate the reliability of participatory sensing between participants and server [8], are jointly taken into account, then a multi-objective selection strategy is devised and formalized as follows,

$$\begin{array}{ll} \max & \alpha x^{'}(t,g,u) + \beta T(u) + \gamma E(u) \\ s.t. & \begin{array}{rcl} T(u) & \geq & \theta_{trust} \\ E(u) & \geq & \theta_{energy} \end{array} \end{array}$$

On one hand, Figure 5(a) illustrates that if a social sensor candidate in the top-k social sensors list has already participated and been selected in the sensing campaign more frequently, it implies that sensor u has a higher trust value, denoted as $T(u)$. Then we extract the social sensors whose trust values are greater than a given threshold θ_{trust} which is determined by service providers. On the other hand, the service providers wish the possible social sensors u have enough energy remained in their mobile devices, namely $E(u)$. The energy consumption for each participant during the sensing period follows an exponential decay trend [10] $E(t) = E(0)e^{-2t}$ as shown in Figure 5(b), where $E(t)$ indicates the residual energy at time t, $E(0)$ is an initial energy in their mobile devices. Thus, each social sensor should report back the remaining energy information and sensing data. Similarly, another threshold θ_{energy} is adopted for further refining in order to guarantee continuous and reliable data sensing and communication. Note that, the weighted parameters α, β, and γ can be learned with Least-square Method [9]. Particularly, the above multi-objective selection strategy is degraded to the basic selection strategy of social sensors.

In the real world, the emergency occurs inevitably during the participatory sensing, such as the insufficient power of sensing devices and urgent affairs of u, then we can select the friend-like social sensors of u for ensuring the participatory sensing campaign.

IV. An Illustrative Example

In this section, we present an illustrative example of participatory sensing at HUST campus to evaluate its feasibility and effectiveness of the basic recruitment strategy.

A. Setup

For a given monitoring area, we first position and determined the virtual sensing range of those 5 DC-points. We firstly create a public microblog ID and regard it as a sensing data monitoring platform. We collected the GPS location and noise information every 2 minutes between 8:00-8:40 AM from a number of volunteers at our university from SEP to NOV, 2013 according to their social media feedback (texts, images) interacted with our public microblog ID. Thus, these collected daily trajectory data can be constructed as a 3-order tensor which includes users, grids, and time dimensions. The detailed steps can be found at our website[1].

B. Results and Discussions

After data collection, the time-series tensor χ^T of trajectory data is trained by DTA algorithm. For illustration and visualization purpose, we choose the trajectory data gathered by 5 typical volunteers $(u_1, \cdots, u_5)$ between 2 Nov and 8 Nov as shown in Figure 4. In our experiments, these 5 participants are the representative social sensors in 5 groups of friend-like social sensors. Table I shows the expected participants to be recruited at four different times in five grids including pre-deployed DC-points. Clearly, the campaign organizer will recruit users u_1, u_4 as the potential social sensors in grid g_1 at time t_7. The likeliness value reflects the possibility of users who are to be recruited. However, user u_5 in Italic will not be considered in this participatory campaign due to his low likeliness. For example, user u_4 should be first considered as a social sensor in grid g_1 at time t_6 because of the higher likeliness value compared to that of u_1.

[1]http://epic.hust.edu.cn/ps

Campaign $C < grid, time >$	Participants user (likeliness)
$C < g_1, t_6 >$	u_4 (0.963), u_1 *(0.152)*,u_3 *(0.017)*
$C < g_1, t_7 >$	u_4 (1.036), u_1 (0.901), u_5 *(0.036)*
$C < g_2, t_5 >$	u_1 (0.977), u_2 *(0.001)*
$C < g_3, t_{10} >$	u_1 (1.052), u_3 *(-0.102)*
$C < g_3, t_{11} >$	u_1 (0.938)
$C < g_4, t_{10} >$	u_5 (1.131), u_2 *(-0.230)*
$C < g_4, t_{11} >$	u_5 (0.958)
$C < g_4, t_{14} >$	u_5 (0.230), u_1 (0.210)
$C < g_5, t_{15} >$	u_5 (1.118)

TABLE I
Expected Participants to be Recruited in Various PS Campaigns

As mentioned above, trust is one of the considerations in multi-objective selection strategy, is an important personalized factor [8]. Obviously, u_1 is more likely available in grid g_{13} which includes pre-deployed DC-points from time t_9 to t_{11}. If we choose the u_1 as the social sensor at that period, more interactions are benefit to enhancing the trust value that may assist the further selection next time. Since diverse sensing devices (e.g., PM2.5 Sensing Device, Smart Phone) are hold by our volunteers, estimating energy consumption of these devices is becoming a challenge. We will study the multi-objective selection strategy with energy consideration in the future work.

In reality, the PSSs usually suffers several emergencies due to the lower power of sensing devices and personal urgent affairs. Hence, an emergency selection strategy of social sensors is urgently devised for achieving a reliable and accurate participatory sensing campaign. Our proposed emergency selection strategy is to choose the participants contained in the group of friend-like social sensors in the event of emergency. To evaluate the feasibility of the emergency selection strategy, users u_6, u_7, u_8 are taken as the assistant social sensors who might be selected to be the friend-like social sensors. In Figure 4, if user u_3 suddenly terminates his sensing temporarily due to some urgent personal affairs, then the PSS should receive this feedback and select the potential social sensors u_6, u_8 from the group of friend-like social sensors of u_3. Actually, this results are reasonable because users u_6, u_8 and u_3 have the relationships of both *roommates* and *classmates*.

V. Conclusions

To realize a novel sustainable sensing, this article investigates the social sensor recruitment problem in participatory sensing systems. We first present a framework of PSSs. Since the recruitment layer in the proposed framework has not been investigated yet, this paper focuses on the issue of recruitment of social sensors and proposes a trajectory-based recruitment strategy of social sensors for participatory sensing. Specifically, the collected trajectory data of users within a period are constructed as a 3-order time-series tensor. Further, DTA algorithm is adopted to learn this time-series tensor and predict the future trajectory information that supports the basic selection strategy based on availability. Finally, the proposed selection strategy is evaluated with an illustrative example conducted at HUST campus. The proposed framework and associated techniques pave a way for achieving intelligent services in PSSs by the virtue of the participatory power of the selected well-suited participants.

REFERENCES

[1] X.O. Wang, W. Cheng, P. Mohapatra, T.F. Abdelzaher, "ARTSense: Anonymous Reputation and Trust in Participatory Sensing", *In Proc. of Infocom'13*, pages 2517–2525, 2013.

[2] S. Reddy, D. Estrin, M. Srivastava, "Recruitment Framework for Participatory Sensing Data Collections", *In Proc. of Pervasive'10*, pages 138–155, 2010.

[3] V. Kotovirta, T. Toivanen, R. Tergujeff, M. Huttunen, "Participatory Sensing in Environmental Monitoring–Experiences", *In Proc. of IMIS '12*, pages 155–162, 2012.

[4] P. Zhou,Z. Chen, M. Li, "Smart traffic monitoring with participatory sensing", *In Proc. of SenSys '13*, pages 26:1–26:2, 2013.

[5] S. Reddy, K. Shilton, G. Denisov, C. Cenizal, D. Estrin, M. Srivastava, "Biketastic: sensing and mapping for better biking", *In Proc. of SIGCHI'10* , pages 1817–1820, 2010.

[6] Y. Zheng, F. Liu, H. P. Hsieh, "U-Air: When Urban Air Quality Inference Meets Big Data", *In Proc. of KDD'13*, 2013.

[7] J. Sun, D. Tao, S. Papadimitriou, P.S. Yu, C. Faloutsos, "Incremental tensor analysis: Theory and applications", *ACM TKDD*, vol.2, no.3, page 11, 2008.

[8] H. Amintoosi, S.S. KanhereA, "A Trust-based Recruitment Framework for Multi-hop Social Participatory Sensing", *In Proc. of DCOSS'13*, pages 266–273, 2013.

[9] I. Yeo, C.C. Liu, and E. J. Kim, "Predictive dynamic thermal management for multicore systems", *In Proc. ACM/IEEE Design Automation Conference, 2008. DAC 2008*, pages 734–739, 2008.

[10] A. P. Miettinen and J. K. Nurminen, "Energy efficiency of mobile clients in cloud computing", *In Proc. of the 2nd USENIX conference on Hot topics in cloud computing*, 2010.