RESEARCH   **Open Access**

CrossMark

Personal destination pattern analysis with applications to mobile advertising

Osama O. Barzaiq[†] and Seng W. Loke[*†]

*Correspondence:
s.loke@latrobe.edu.au
†Osama O. Barzaiq and Seng
W. Loke contributed equally
to this work
Department of Computer
Science and Information
Technology, La Trobe
University, Kingsbury Road,
Melbourne, Australia

Abstract

Many researchers expect mobile advertising to be the killer application in mobile business. In this paper, we introduce a trajectory prediction algorithm called personal destination pattern analysis (`P-DPA`) to analyse the different destinations in various trajectories of an individual, and to predict a trajectory or a set of destinations that could be visited by that individual. The `P-DPA` algorithm works on an individual level. Every destination-pattern analysis is related to the self-history and the personal profile of a targeted individual, not on what others do. In addition, we developed a prototype system called SmartShopper. SmartShopper is a personal destination-pattern-aware pervasive system for mobile advertising in (outdoor and indoor) retail environments. The predicted destinations from the `P-DPA` algorithm will be used by SmartShopper to generate a list of relevant advertisements adapted to the personal profile of previous destinations of a targeted individual. We tested the destination prediction accuracy of the `P-DPA` algorithm with a synthetic dataset of a virtual mall and a real GPS dataset.

Keywords: Personal destinations pattern analysis, Mobile advertising, Human mobility, D-trajectory prediction

Introduction

The significant rapid growth of mobile advertising is making advertising companies consider mobile phones as a new platform for advertising revenue [1, 2]. One of the strategic places for advertising is retail environments, as most mobile phone users visit a mall to purchase products or services, and many researchers are exploiting different methods to generate a personalized list of advertisements that could capture the interest of a mobile phone user [3–5].

The aim of this paper is to propose an approach for generating a list of relevant advertisements for a targeted individual by first predicting the set of destinations that could be visited by that individual. Here, a D-trajectory is a sequence of destinations visited by a user, not necessarily the actual physical path/walk taken by the user when visiting the destinations. Hence, the term D-trajectory is used to indicate that we predict a sequence of meaningful destinations that could be visited by a targeted individual, rather than simply a path. Consequently, D-trajectory is a subsequence of the actual physical path/walk of the targeted individual. We investigate an approach that uses only *self-histories* rather than the histories of others for predictions, i.e., rather than predicting the behaviour of

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 2 of 29

a person based on what others typically do as in other work, we predict based only on what the person him/her-self did. The key contributions of this paper are:

- a novel approach to predict a set of destinations that could be visited by a targeted individual using only the personal history of that individual,
- a new perspective on generating a list of advertisements for users, i.e. by using the predicted set of destinations and the personal profile of a targeted individual,
- highlighting the use of self-histories for destination prediction, rather than group-histories (as in typical clustering style prediction approaches), i.e., we predict based on what the person typically does, rather than what people typically do,
- highlighting the notion that taking into account the day of the week is crucial in capturing human routines and regularity, i.e. to predict what a person will do on a day, say its a Tuesday, it can be more useful to look at what the person typically do on Tuesdays, rather than what the person does on all days of the week, and
- highlighting two kinds of behavioural regularity, in terms of the number of destinations visited and the actual destinations visited.

It must also be noted that our algorithm predicts a D-trajectory of a person for a day, given a history of D-trajectories of the person; the history is a set of D-trajectories, one D-trajectory for each day in the person's history. That is, we are not simply predicting the next destination given visited destinations so far, as in other work.

In order to illustrate the application we have in mind, we developed a prototype system (SmartShopper) and a destination prediction algorithm called personal destination pattern analysis (or P-DPA, for short). SmartShopper uses the P-DPA algorithm to predict the set of destinations that could be visited by a shopping mall visitor. Then, Smart-Shopper will use the predicted destinations and the personal history of the targeted mall visitor to generate a list of advertisements that aims to capture the interest of that visitor with high probability. We tested the prediction accuracy of the P-DPA algorithm on a synthetic dataset of an indoor mall and on a real GPS dataset of outdoor D-trajectories for nine persons (the GPS dataset was collected by the LifeMap system of [6]). The experimental results show the good prediction accuracy of the P-DPA algorithm.

The rest of this paper is organised as follows. Section reviews related work. Section presents our D-trajectory prediction algorithm. Section has the details of the synthetic dataset of the virtual mall,and the details of the GPS dataset of [6]. Section shows the experimental results. The details of the prototype system, SmartShopper, are discussed in "System prototype" section. "Conclusion and future work" section concludes the paper with future work.

Related work

Asahara et al. [7] have developed a prediction method that uses a mixed Markov-chain model (MMM) to predict the next location of an individual. In their prediction method, they categorize individuals into groups, assuming that individuals in each group are having similar behavior. Accordingly, they build a mixed Markov-chain model for every group, and every MMM is using one unobservable parameter. The unobservable parameter is used to determine which mixed-model of which group of individuals

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 3 of 29

should be used to generate the transition matrix. They tested the prediction accuracy of their MMM on a physical dataset that were collected during a big event conducted in a famous shopping mall in Japan. For 90 min, the trajectories of 691 participants in the experiment during the big event had been recorded. After that, the recorded trajectories had been divided to 10 sets, 9 sets were used for training purposes and only one set was used to test the prediction accuracy of MMM. For the prediction accuracy test, they built 190 mixed Markov-chain models and the average accuracy was 64 % [7]. However, recording the participants' trajectories during a particular event for a specific period of time could result in significant similarity between most of the recorded trajectories. This is because that the movements of the participants were mostly according to the designated path of the big event of the mall. Moreover, they used 190 mixed Markov-chain models to achieve the prediction accuracy of 64 %, whilst one Markov-chain model with one transition matrix for the 691 participants was able to achieve a prediction accuracy of 45.6 %. The first process of dividing the participants into groups and the second process of building 190 mixed Markov-chain models would take time, and time is also required to predict the next location of an individual.

Kolodziej et al. [8] developed an algorithm using an activity-based continuous-time Markov jump process, which is an extension of the AMPuMM algorithm of [9]. In their algorithm, instead of using the discrete time set used in AMPuMM [9], they used a continous-time Markov process to build a model of the users movements. Their algorithm was able to predict the next location of a mobile user within a range of 115–250 m from the correct location of that user. Nevertheless, when the prediction algorithm is only using a set of four locations and the predicted location is that far from the correct location, the prediction algorithm needs a number of improvements to handle a bigger set of locations and to be able to provide a more accurate predictions. Furthermore, our `P-DPA` algorithm is tested using synthetic dataset of a virtual mall that has 72 stores and a real GPS dataset for nine persons with a number of locations between 109 and 448.

Gambs et al. [10] developed an algorithm called, n-MMC. The n-MMC algorithm predicts the next place of an individual by using the Mobility Markov-chain (MMC) model and the last two visited locations. In their algorithm, they built a transitions matrix that has the probabilities of all the different transitions of an individual to place *k* after visiting place *i* and place *j*. The n-MMC algorithm managed to achieve a prediction accuracy ranging from 70 to 95 %. However, n-MMC predicts the next place of an individual among only a small set of three possible locations: home, work, and other. Furthermore, if the next place predicted by the n-MMC algorithm for a targeted individual was Other, it was not mentioned by the researchers how such a prediction could be used to generate and send useful and relevant information to that individual. Moreover, in their algorithm, n has a constant value of 2 which represents the two previously visited locations but our prediction algorithm, `P-DPA`, is using a version of a transition matrix that has a memory size of *n*, which means it has the probabilities of visiting different destinations after visiting *n* number of locations by a targeted individual, where *n* is determined by analysing the dataset.

Another group of researchers designed a framework called LASA, Location Aware Shopping Advertisement, that uses an ontology based formulation of clients and products profiles to generate a list of ads related to the selection history of a targeted client

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 4 of 29

[5]. LASA uses the shopping mall's Wi-Fi access points to detect the current location of a client, and once the client's location is detected, it will send a list of titles of different products available in all stores in the coverage area of the access point that discovered the client's current location. Nevertheless, the coverage area of an access point may include a big number of stores which could result in generating a long list of products' titles from those stores. Furthermore, a number of access points could detect the signal of a client's mobile phone which could affect the accuracy of determining the current location of that client. Consequently, failing to discover the correct location of a targeted client will hinder the process of determining the list of products' titles that should be sent to that client. In addition, our `P-DPA` algorithm predicts a set of stores that could be visited and then generate a list of ads for those stores instead of just generating a possibly long list of ads for every store in the discovered area of a targeted client as in LASA.

Kim et al. [3] developed a system called, AdNext, that uses Bayesian networks to build a transition matrix for the shopping mall clients in order to predict the business type of the next location that a targeted client could visit. AdNext uses the business type of the last two locations visited by a targeted client to predict the business type of the next location for that client. In addition, AdNext detects the current location of the targeted client by using the shopping mall's access points. Based on the discovered location of the targeted client and the predicted business type, AdNext will start generating a list of ads for every store that lies under the predicted business type and it is also in the area of the detected location. However, predicting the exact next store that could be visited by a client is significantly different from predicting the business type of the next location. Predicting the business type of the next location could result in generating many ads, because such a list will include ads from every store that falls under the predicted type of business. Furthermore, the knowledge of the business type of the last two visited locations is crucial for AdNext to be able to predict the business type of the next location. Consequently, if a targeted client has a plan to only visit one or two stores, then, AdNext will not be able to make any predictions. Instead of predicting the business type, our prediction algorithm, `P-DPA`, predicts a set of stores that could be visited by a targeted client, and then, uses the predicted stores to generate a list of advertisements for the targeted client. In addition, `P-DPA` algorithm can predict *the first store to visit* by a targeted client while AdNext will need to wait for the targeted client to visit at least two stores in order to be able to make any predictions.

The P-DPA algorithm

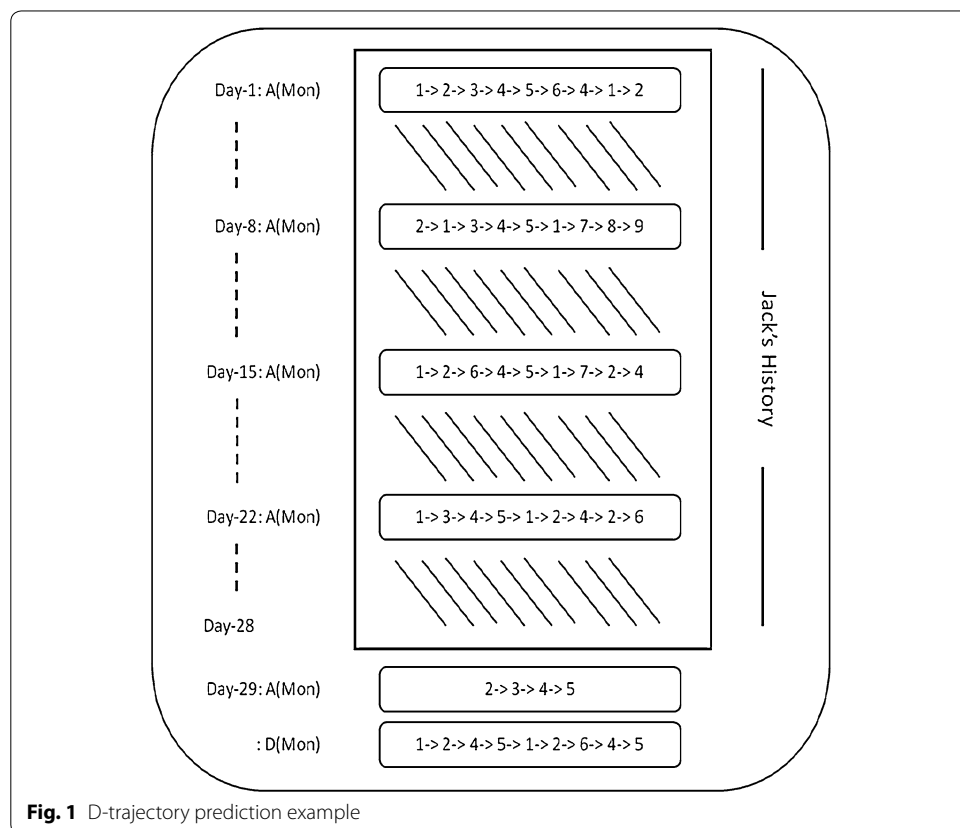We developed a trajectory prediction algorithm called `P-DPA`, personal destination pattern analysis. The `P-DPA` algorithm is predicting what we call a D-trajectory, which is a trajectory of destinations. The D-trajectory is a subsequence of the actual walking trajectory of a targeted individual (containing meaningful destinations, meaningful as determined by the domain).

Different from other work, note that our algorithm does not predict based on what others do (e.g., in collaborative/clustering prediction style approaches), but based purely on the self-history of the individual him/her-self. The `P-DPA` consists of the following steps.

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 5 of 29

- *Targeted analysis period and targeted trajectories.* The `P-DPA` algorithm will analyse up to a history window of 2 years of the different trajectories made by a targeted individual on the same day of the week—e.g. to predict a D-trajectory for a Wednesday, we look at the D-trajectories for all Wednesdays in the last 2 years. We observe that same day of the week is a useful selection filter for the history to be used (we also experimentally show this later).

- *Dynamic prediction length.* It is necessary to determine the length of the predicted D-trajectory. The dynamic trajectory length is used by the `P-DPA` algorithm to determine the number of destinations that are most likely to be visited by a targeted individual on a specific day. The dynamic length will be determined by finding the number of destinations that were visited by an individual on the majority of days that are similar to the day of his/her current activity. For example, if the current activity of individual(x) is occurring on Sunday and the number of destinations that are often visited by individual(x) on all previous Sundays is 7, then the length of the predicted D-trajectory by `P-DPA` is seven destinations.

- *The initial distribution matrix. D* is an initial distribution matrix that contains the probabilities for every destination to be the first destination to visit by individual(x). The matrix is computed from the self-history of individual(x).

- *The transition matrix. R* is a transition matrix that contains the probabilities of different transitions of individual(x) from one destination to another. *R* will be used as a transition model for individual(x), and it's built from the set of the targeted analysis trajectories. The `P-DPA` algorithm is using two versions of the transition matrix. The first version of the transition matrix has a memory size of one, which means that the transition matrix has the probabilities of visiting different destinations after visiting the current location of individual(x). The second version of the transition matrix has a memory size of *n*, which means it has the probabilities of visiting different destinations after visiting *n* number of locations by individual(x), where n is obtained by analysing the dataset.

- *D-trajectory prediction.* The following steps show how `P-DPA` predicts the D-trajectory that is most likely to be used by individual(x) in the current day of his/her current activity.

  1. Once individual(x) starts leaving his/her house or parking his/her car in a shopping mall, `P-DPA` will use the current date and location of individual(x) to fetch the records of his/her various trajectories for a period of up to 2 years that precede the current date.

  2. Then, `P-DPA` will filter and analyse only the trajectories used on all *selected historical days*, i.e. days that are of the same day of the week as the day of the current activity (e.g. if the current day whose D-trajectory is to be predicted is Monday, then only the trajectories on Mondays in the history are selected and used).

  3. From the fetched trajectories, the dynamic prediction length (i.e. length of the D-trajectory to be predicted) will be determined by finding the number of destinations that were visited by individual(x) on the majority of the selected historical days.

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 6 of 29

4. Then, P-DPA will build the initial distribution matrix *D* that contains the probabilities for every destination to be visited first by individual(x), and it will also build the transition matrix *R* that contains the probabilities of different transitions of individual(x) from one destination to another.

5. The destination that is most likely to be visited *first* by individual(x) is the destination that has the highest probability in *D*.

6. The next destination to be visited by individual(x) is the destination that has the highest probability in *R* to be visited after visiting the current destination.

7. We use the dynamic prediction length to determine the number of times required to repeat the previous step to find all the remaining destinations of the predicted D-trajectory.

Algorithm 1 shows how P-DPA predicts the D-trajectory of individual(x), and Fig. 1 shows an example of a predicted D-trajectory by the P-DPA algorithm. Assuming that we have an individual named Jack, and he is about to leave his house, and the current day is Monday, then, the P-DPA algorithm will focus and analyse D-trajectories that were recorded on previous Mondays. From Fig. 1, the current day is *Day-29, A(Mon)* is the actual trajectory of Jack, and *D(Mon)* is the predicted D-trajectory by the P-DPA algorithm.

Fig. 1 D-trajectory prediction example

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 7 of 29

Algorithm 1 P-DPA Algorithm

1: $user \leftarrow getUser()$
2: $currActivityDate \leftarrow getCurrentActivityDate(user)$
3: $dayName \leftarrow getDayName(user, currActivityDate)$
4: $A \leftarrow getTargetedAnalysisPeriod(user, currActivityDate)$
5: $B \leftarrow getTargetedAnalysisTrajectories(A, dayName)$
6: $dynamicPredictionLength \leftarrow getMostFrequentNumberOfVisitedDestinations(B)$
7: $D \leftarrow buildInitialMatrix(B)$
8: $R \leftarrow buildTransitionMatrix(B)$
9: $indivRecentTrajectory \leftarrow getRecentIndivTrajectory(currActivityDate)$
10: $firstDestinationToVisit \leftarrow getDestinationOfHighestProbability(D)$
11: $predictedTrajectory.add(firstDestinationToVisit)$
12: $indivCurrentDestination \leftarrow firstDestinationToVisit$
13: **for** $i = 2$ **to** $dynamicPredictionLength$ **do**
14:     $nextDestination \leftarrow getNextDestination(R, indivCurrentDestination)$
15:     $predictedTrajectory.add(nextDestination)$
16:     $indivCurrentDestination \leftarrow nextDestination$
17: **end for**

Datasets

In this section we will discuss the details of the two datasets that we used to measure the prediction accuracy of the P-DPA algorithm. Using seed information from real individuals, we generated a synthetic dataset that has 3 years of different indoor trajectories and various shopping activities inside a virtual mall. The second dataset is a real GPS dataset of outdoor trajectories for nine persons, which is collected by the LifeMap system of [6] (over roughly 8 months). Note that for our work, the important aspect of the dataset to be considered is the size of the history and the behaviour variability across clients, rather than the number of clients [though there should be sufficient variability in the clients to demonstrate that our system can work with different types of client behaviours (i.e. different extents of behavioural (ir)regularity)]; hence, the nine clients in that dataset and the different generated synthetic behaviours we have were more than adequate to demonstrate the applicability of our algorithms across a range of behaviours (from extremely regular to very irregular).

The virtual mall dataset

We built a simulation system that consists of a virtual mall and a number of virtual clients. The purpose of the simulation system is to generate a synthetic dataset that consists of an adequate number of records of various shopping activities, which could allow us to test the prediction accuracy of the P-DPA algorithm.

The virtual mall

The simulation system consists of a virtual mall, called V-Mall. Table 1 shows the components that we used to build V-Mall. We used a significant number of components in order to make V-Mall very close to a mall in the real world.

Shopping behavior model

After the initial construction of V-Mall, we prepared a survey of 14 questions to assist in understanding the shopping behavior of different individuals and then developed a shopping model for every one of the participants in the survey. The survey's questions are divided into two groups. The first group consists of questions intended to help

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 8 of 29

Table 1 V-Mall components

Attribute	Description
Year	3 years (2010, 2011 and 2012)
Month	12 months (January–December)
Day	7 weekdays (Monday–Sunday)
Visiting hour	8 visiting hours (9 a.m., 10 a.m., 11 a.m., 12 p.m., 1 p.m., 2 p.m., 3 p.m. and 4 p.m.)
Gate	4 gates (gate01, gate02, gate03 and gate04)
Store	72 stores (using names existing in the real world)
Store size	3 sizes (small, medium and big)
Visiting duration	6 visiting durations inside a store (5, 10, 15, 20, 25 and >25 min)
Category	17 categories (using names existing in the real world)
Category type	2 types (limited purchase times, multiple purchase times)
Item	The number of items/products depends on the two factors: the store size and the category type
Price range	3 ranges (low, medium, high)

`V-Mall` understand the relationship between the participant and `V-Mall`, to obtain information as follows:

1. How certain are you about visiting `V-Mall` at least one time in any month of the 12 months?
2. How certain are you about the number of visits that you usually make to `V-Mall` in any month of the 12 months?
3. How certain are you about visiting `V-Mall` at least one time on any day of the week?
4. How certain are you about starting your visit to `V-Mall` by using any gate of the 4 gates of `V-Mall` as your favorite gate?
5. How certain are you about starting your visit to `V-Mall` at any of the following times (9 a.m., 10 a.m., 11 a.m., 12 p.m., 1 p.m., 2 p.m., 3 p.m., 4 p.m.) on any day of the 7 days in a week?
6. How certain are you about the number of stores that you usually visit on any day of the 7 days in a week?

The second set of questions is designed to evaluate the relationship between the participant and different stores in `V-Mall`, e.g., to get information such as:

1. How certain are you about visiting a store of the 72 stores?
2. How certain are you about visiting any of the 72 stores at least one time every month?
3. How certain are you about visiting any of the 72 stores at least one time on every day of a week?
4. How certain are you about the time that you usually spend in any of the 72 stores on different days of a week? (Use: 5, 10, 15, 20, 25, >25 min)
5. How certain are you about using any of the 72 stores as your "first store to visit" in your different visits to `V-Mall`?
6. How certain are you about buying at least one item from each of the 72 stores in every visit you make to `V-Mall`?

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 9 of 29

7. How certain are you about the number of different items that you usually buy from each of the 72 stores in every visit you make to `V-Mall`?

8. How certain are you about the chances of visiting a new store in every trip you make to `V-Mall`? (New store is a store that have not been visited before)

Note that the idea of the survey is not only to see if the participant visits any particular store or uses a particular gate or visits at a particular time, but for those questions where the answer is positive, for each store/gate in the mall or time slot, we also seek to find out the probability of visiting it and which times are more likely.

We carefully selected three individuals to participate in the survey, after an initial observation that they would likely provide diverse (in terms of regularity of visiting stores) stereotypical behaviors. The participants have the following characteristics: a married female that has two children, and two single males. Note that the number of participants is not as significant as we aim towards distinguishable stereotype behaviors, and the survey is only meant to provide seed data for data generation. The collected answers are used by our simulation system to generate a shopping behavior model for every one of the participants. Such models will help `V-Mall` to generate a synthetic dataset of various purchasing activities that could be performed by each stereotype participant in `V-Mall`. The system generated 3 years of different trajectories and shopping activities for each one of the participants. Then, the `P-DPA` algorithm will analyse only the personal history of a specific individual, which is different from other collaborative/clustering prediction style approaches.

Clients creation and dataset generation

After building the shopping behavior models of the participants, `V-Mall` will start generating various shopping activities through the following settings:

- *Virtual clients creation.* `V-Mall` will create 5 virtual clients. Three virtual clients will be created based on the seed data obtained from the shopping behavior model of each one of the survey's participants (using questions described earlier). Those virtual clients will have the following tags: *client-M4*, *client-M5*, and *client-F8*. In addition, `V-Mall` will create another two virtual clients using a shopping behavior model developed completely by `V-Mall`. The first virtual client will be created as a client with *non-specific and irregular shopping behavior*, and it will be tagged *client-S1*. Having an extremely irregular shopping behavior means that such client can visit any number of stores on any day and at any time with absolutely no specific reason/pattern. The fifth virtual client will be created as a client with *an extremely regular shopping behavior*, and it will be tagged *client-S2.* This client has an extremely regular behavior because he is always visiting the same stores on the same days and at the same times. Figure 2 shows a sample of different trajectories conducted by client-F8 in `V-Mall`. The purpose of creating 5 virtual clients with three different types of shopping behaviors is to generate a synthetic dataset that consists of various shopping activities with 3 main stereotypical shopping behaviors, irregular, regular and extremely regular, which allows us to cover the range of possible trajectories better than even real data (which may not exhibit the range of stereotypical behaviours

Fig. 2 Client-F8 trajectories in V-Mall
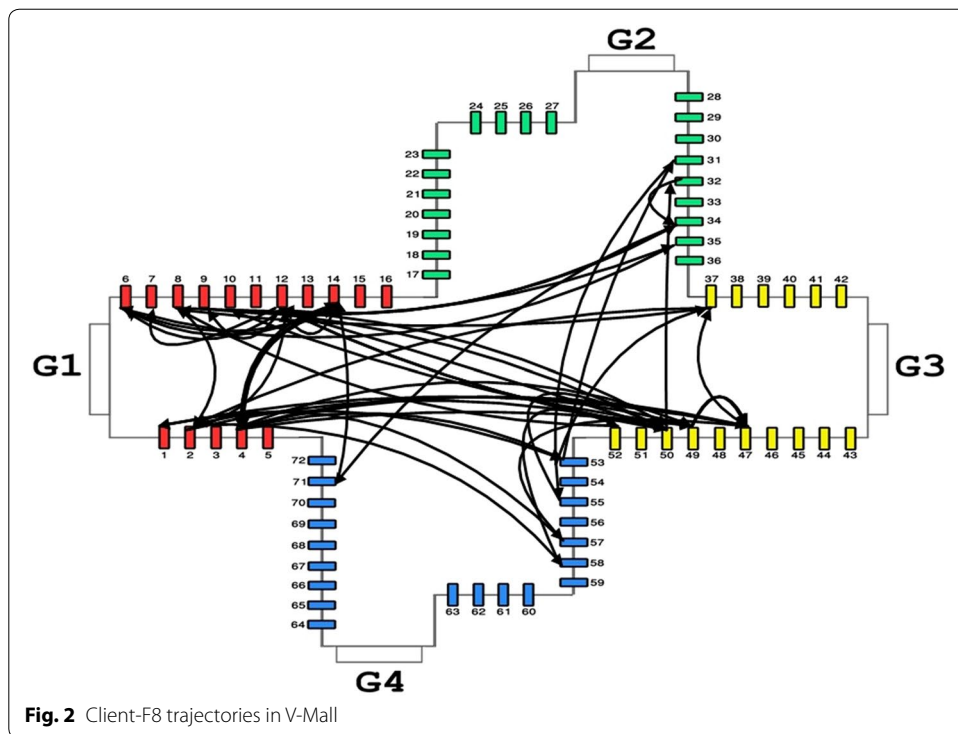
needed for this study). The different transitions of *client-S1* and *client-S2*, as shown in Figs. 3, 4, represent two extreme ends and it can be intuitively observed that the shopping behavior of any individual will be between these two ends (in terms of regularity rather than actual stores visited). Therefore, having three or more respondents to our survey will not change the fact that the shopping behavior of any individual will be between the two extreme ends, irregular and extremely regular.

- *Generation period.* V-Mall will generate various shopping activities and trajectories for a period of 3 years for every virtual client. 2 years of the generated activities will be assigned to a training set, and the activities of the 3rd year will be used to evaluate the prediction accuracy of the P-DPA algorithm. After the completion of the generation step, we will have a dataset that consists of 3 years of various shopping activities for five virtual clients in V-Mall. Table 2 describes the sizes of the contents of the virtual mall dataset. For this dataset, each destination is a visited store and each D-trajectory is a sequence of visited stores in a visit to the shopping mall. Note that for five clients, there are five datasets of D-trajectories, one for each client.

GPS dataset

D-Trajectories based on a real GPS dataset was used to evaluate the prediction accuracy of the P-DPA algorithm. The GPS dataset contains fine-grained mobility data from mobile phones of 12 persons during a period of two months in Seoul, Korea, but only the records of *nine persons* were extractable.[1] This dataset was collected by the LifeMap[2] system of [6]. Table 3 describes the sizes of the contents of the D-trajectory dataset

---

[1] Data for SUB-5, SUB-6 and SUB-11 are not extracted.

[2] CRAWDAD dataset yonsei/lifemap(v.2012-01-03), http://crawdad.cs.dartmouth.edu/yonsei/lifemap, Jan. 2012.

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 11 of 29

Fig. 3 Client-S1 transitions in V-Mall

Fig. 4 Client-S2 transitions in V-Mall

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 12 of 29
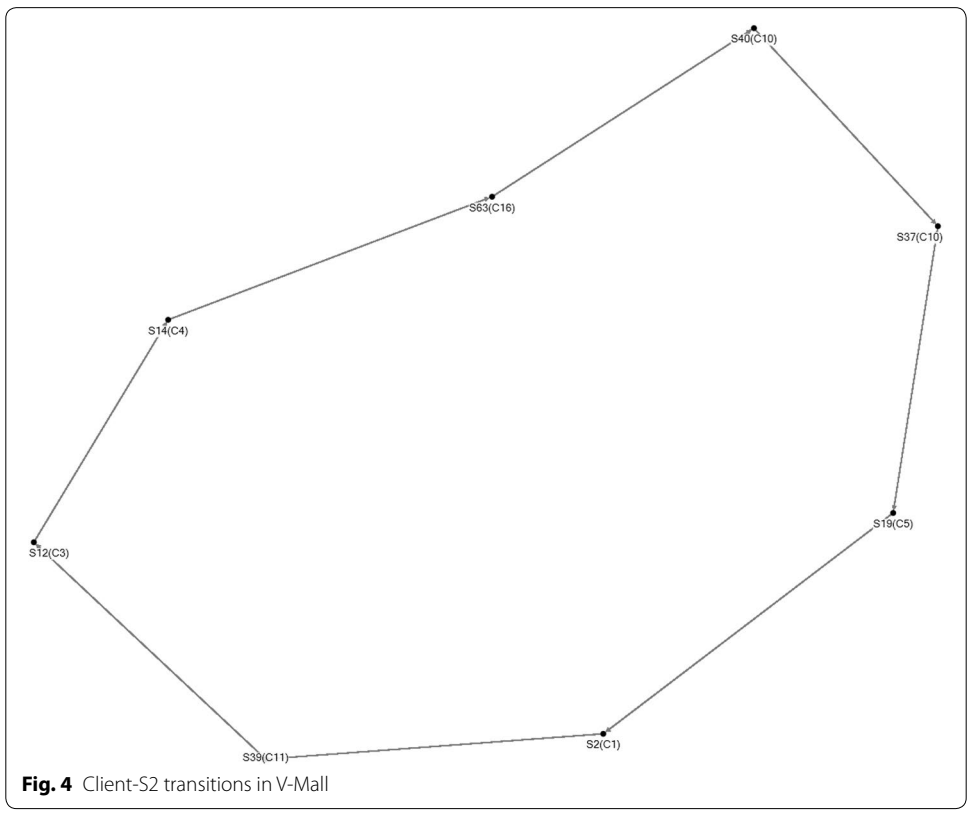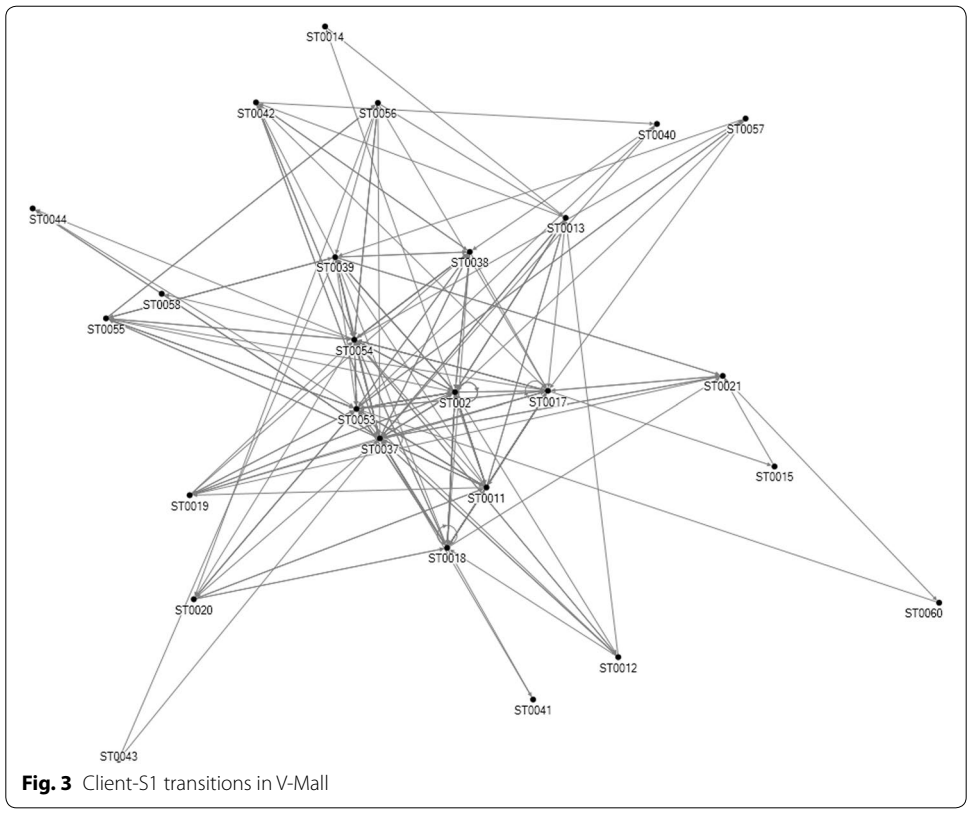
Table 2 V-Mall dataset stats

Client ID	Visits	D-trajectories	Visited stores
Client-S1	376	376	63
Client-S2	627	627	8
Client-M4	537	537	15
Client-M5	859	859	19
Client-F8	947	947	40

Table 3 GPS dataset stats

Subject ID	Days	D-trajectories	Locations (>2min)
SUB-1	133	133	414
SUB-2	197	197	220
SUB-3	270	270	312
SUB-4	191	191	208
SUB-7	130	130	345
SUB-8	249	249	346
SUB-9	191	191	201
SUB-10	47	47	109
SUB-12	361	361	448

derived from this GPS dataset. It should be noted that, for this dataset, we define a destination as a location with stay time >2 min, and a D-trajectory is a sequence of destinations of an individual in 1 day. Note that for the nine clients, there are nine datasets of D-trajectories, one for each client/subject.

Evaluations and results

In this section, we will measure the prediction accuracy of the `P-DPA` algorithm by using a transition matrix with two different memory sizes. In addition, we developed a Markov-chain algorithm for predicting the D-trajectory of a targeted individual. The Markov-chain algorithm is used only for comparison with the `P-DPA` algorithm.

P-DPA with two memory sizes

- *P-DPA_M1* The `P-DPA` algorithm uses a transition matrix that has the probabilities of visiting different destinations after the current location of individual(x). Hence, suffix _M1 refers to a memory size of one destination.
- *P-DPA_MN* Suffix _MN with the `P-DPA` algorithm indicates that the algorithm uses a transition matrix that has the probabilities of visiting different destinations after visiting *n* number of destinations. For example, if individual(x) visited location(i), location(j) and location(k), then, the `P-DPA` algorithm will predict the next destination that has the highest probability to be visited after those three destinations.

Markov-chain algorithm with no memory

Markov-chain algorithm is used in several human trajectory approaches to find the next location of an individual [8–11]. Our Markov-chain algorithm consists of the following steps:

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 13 of 29

- The training set. 2 years of different trajectories from the V-Mall dataset and 60 % of the GPS dataset of individual(x) will be used by Markov-chain algorithm for training purposes.
- Predefined prediction length. Different lengths will be used in the process of predicting the D-trajectory of individual(x). Suffixes _3, _5 and _7 with Markov-chain algorithm indicates that the algorithm will be used to predict a trajectory with a length of 3, 5 and 7, respectively. For example, MC_5 means that Markov-chain algorithm will predict a trajectory of five destinations.
- The initial distribution matrix π that contains the probabilities of visiting every destination by individual(x).
- The transitions matrix T that has the probabilities of different transitions of individual(x) from one destination to another. The training set of individual(x) is used to build π and T.
- The D-trajectory prediction process includes the following steps:

  i. The first destination to be visited by individual(x) is the destination that has the highest probability in π.
  ii. The Markov-chain algorithm does not keep a memory of the last visited destination(s) when making a prediction for the possible next destination. Hence, the next remaining destinations on the predicted trajectory will be found by using the following formula:

$$\pi^i = \pi^{i-1} * T \tag{1}$$

  The next destination to be visited by individual(x) is the destination that has the highest probability in π^i.
  iii. Based on the predefined prediction length, the previous step will be repeated to find the remaining destinations of the predicted D-trajectory.

Measuring prediction success

Note that while we predict a sequence of stores (a D-trajectory), the evaluation we provide in the rest of the paper focuses on precision and recall measures (analogous to information retrieval since relevance of ads relies on this more) rather than order (order is considered elsewhere for the shopping mall scenario in [12]).

We are measuring the prediction accuracy of the P-DPA algorithm as follows:

$$Precision = |D \cap A|/|D| \tag{2}$$

$$Recall = |D \cap A|/|A| \tag{3}$$

$$F - measure = 2 * ((Precision * Recall)/(Precision + Recall)) \tag{4}$$

where A is the actual D-trajectory (i.e., its corresponding set of destinations) used by individual(x), and D is the D-trajectory (i.e., its corresponding set of destinations) predicted by the P-DPA algorithm.

As mentioned earlier, a synthetic dataset of indoor D-trajectories and a real dataset of outdoor D-trajectories (derived from a GPS dataset), altogether involving 14 subjects/

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci. (2016) 6:17*

Page 14 of 29

clients (five in a virtual mall and nine GPS based), were used to evaluate the accuracy of a `P-DPA` predicted D-trajectory compared to the actual trajectory used by the subject/client.

Dynamic prediction length vs. fixed length

Figures 5 and 6 show how using dynamic prediction length versus using a fixed length resulted in a significantly accurate prediction of the number of destinations that could be visited by a targeted individual. Figure 5 shows the length prediction accuracy for every virtual client in the V-Mall dataset. By using a dynamic prediction length, the `P-DPA` algorithm was able to predict the number of destinations with an accuracy of 75.94 % with client-S1, 92.99 % with client-M4, 82.30 % with client-M5 and 100 % with client-S2. On the other hand, using a fixed length with Markov-chain algorithm resulted in different accuracy values. The Markov-chain algorithm with a fixed length of 3 was able to achieve a length prediction accuracy of 49.49 % with client-S1, 13.73 % with client-S2 and 9.53 % with client-F8. Using a fixed length of 7 helped the Markov-chain algorithm predict the number of destinations with an accuracy of 43.32 % with client-M4 and 81.04 % with client-M5.
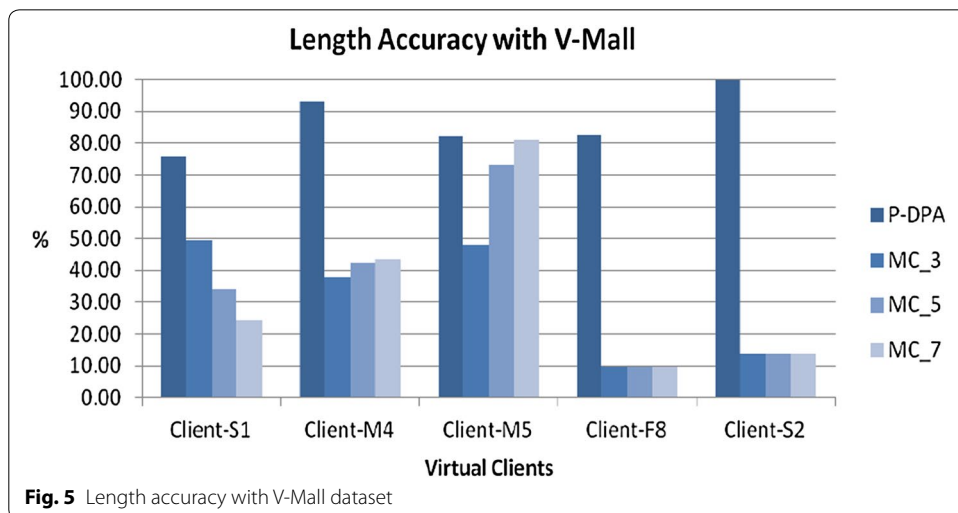
Fig. 5 Length accuracy with V-Mall dataset
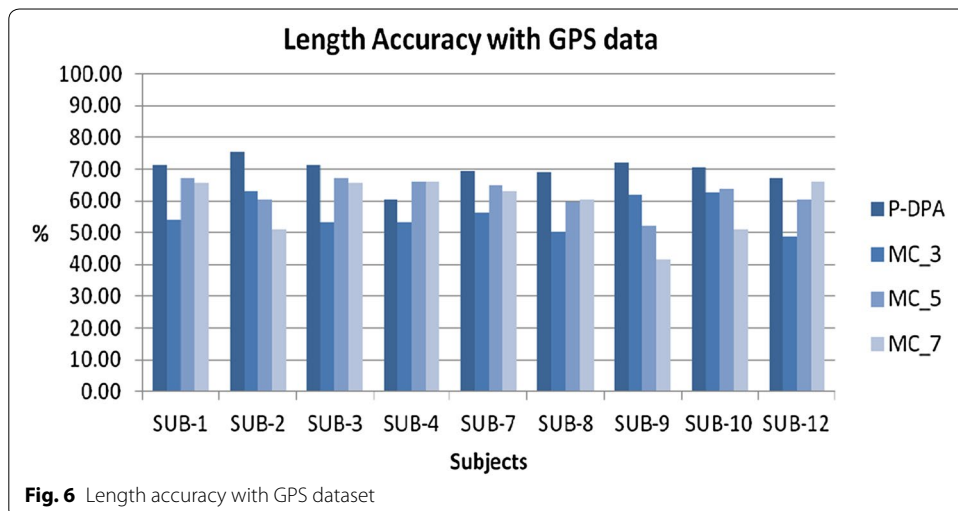
Fig. 6 Length accuracy with GPS dataset

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 15 of 29

Figure 6 shows the length prediction accuracy for every subject in the GPS dataset of [6]. In comparison to the Markov-chain algorithm with different fixed length values, the `P-DPA` algorithm was able to achieve the highest prediction accuracy regarding the number of destinations that could be visited, for every subject. The `P-DPA` algorithm was able to predict the length of the D-trajectory with an accuracy of 71.13 % with SUB-1, 75.57 % with SUB-2, 71.46 % with SUB-3, 60.37 % with SUB-4, 69.42 % with SUB-7, 68.88 % with SUB-8, 71.90 % with SUB-9, 70.67 % with SUB-10 and 67.23 % with SUB-12. On the other hand, the Markov-chain algorithm with a fixed length of 3 was able to achieve a length prediction accuracy of 63.16 % with SUB-2 and 61.76 % with SUB-9, while using a fixed length of 7 resulted in an accuracy of 60.56 % with SUB-8 and 66.08 % with SUB-12. With a fixed length of 5, the Markov-chain algorithm predicted the number of destinations with an accuracy of 67.23 % with SUB-1, 67.08 % with SUB-3, 66.16s % with SUB-4, 65.00 % with SUB-7 and 63.67 % with SUB-10.

Prediction accuracy with V-Mall dataset

Figure 7 shows the Precision value with every virtual client in V-Mall dataset. With client-S1, who has an extremely irregular behavior, Markov-chain algorithm with a fixed length of 3 achieved the highest Precision value of 18.15 %, while `P-DPA_M1` achieved 12.90 % and `P-DPA_MN` achieved 16.80 %. With virtual clients, client-M4, client-M5 and client-F8, who have a regular behavior, the `P-DPA` algorithm was able to achieve the highest precision values over the Markov-chain algorithm with different fixed lengths. With client-M4, the `P-DPA_M1` achieved a Precision value of 83.09 % and the `P-DPA_MN` achieved a Precision of 82.38 %, while the Markov-chain algorithm with a fixed length of 3 achieved 31.76 %. The highest Precision value with client-M5 was achievable by `P-DPA_MN` and `P-DPA_M1` with Precision of 64.79 and 64.21 %, respectively. The Markov-chain algorithm with a fixed length of 3 was only able to obtain a Precision of 39.01 % with client-M5. Both prediction algorithms were struggling to obtain accurate predictions with client-F8, who is a mother of two children. With client-F8, `P-DPA_MN` achieved a Precision of 50.03 %, `P-DPA_M1` achieved 48.92 % and Markov-chain
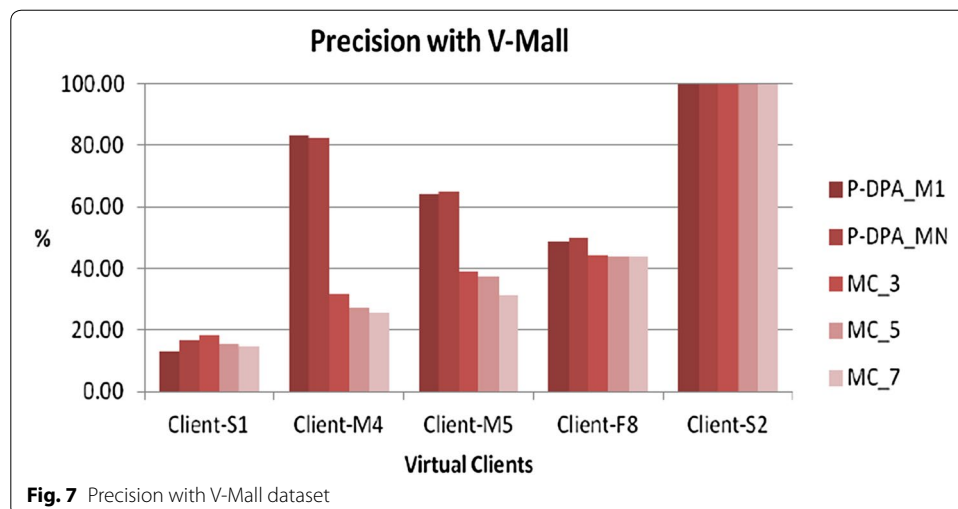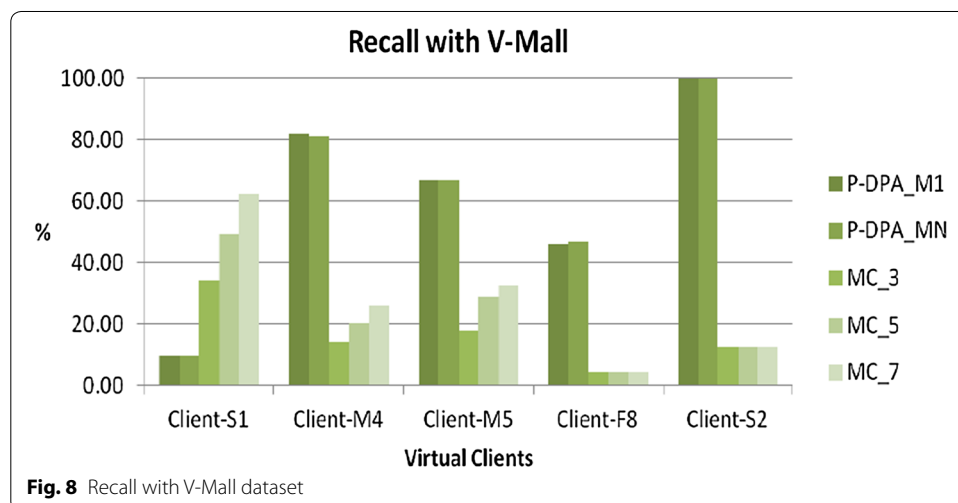

Fig. 7 Precision with V-Mall dataset

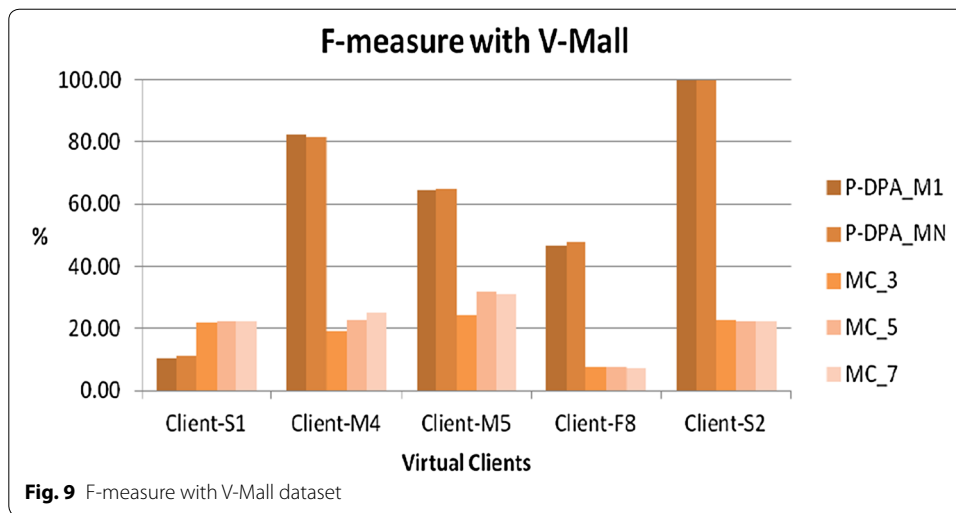Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 16 of 29

achieved 44.20 % using a fixed length of 3. The extreme regularity of client-S2 helped both algorithms, P-DPA and Markov-chain, to achieve a Precision value of 100 %.

Figure 8 shows the recall value with the virtual clients in V-Mall. The extreme irregularity of client-S1 was a significant obstacle for P-DPA, and hence, P-DPA_M1 achieved a recall value of 9.57 % and P-DPA_MN achieved 9.69 %, while the Markov-chain algorithm with a fixed length of 7 was able to achieve the highest Recall of 62.31 %. With client-M4, client-M5 and client-F8, the P-DPA algorithm achieved the highest Recall values. With client-M4, P-DPA_M1 achieved a Recall value of 82.09 % and P-DPA_MN achieved 81.24 %, where the Markov-chain algorithm with a fixed length of 7 achieved only 25.89 %. P-DPA_M1 and P-DPA_MN achieved a Recall value of 66.94 % with client-M5, and Markov-chain using a fixed length of 7 achieved 32.61 %. The main challenge for both algorithms was client-F8. With that virtual client, P-DPA_MN achieved a Recall of 46.97 % and P-DPA_M1 achieved 45.99 %, while Markov-chain with a fixed length of 3 could not achieve more than 4.32 %. Even though client-S2 has an extremely regular behavior, Markov-chain algorithm achieved only a Recall of 12.43 % using a fixed length of 3, while both P-DPA_M1 and P-DPA_MN achieved a recall value of 100 %.

Figure 9 shows the prediction accuracy of both algorithms, P-DPA and Markov-chain, through the F-measure values with every client in V-Mall dataset. The highest F-measure value of 22.26 % with client-S1 was achievable by the Markov-chain algorithm with a fixed length of 5, while P-DPA_M1 and P-DPA_MN could not obtain more than 10.29 and 11.26 %, respectively. On the other hand, with clients, client-M4, client-M5 and client-F8, the P-DPA algorithm achieved the highest F-measure values over the Markov-chain with different fixed lengths. P-DPA_M1 achieved F-measure value of 82.33 % and P-DPA_MN achieved 81.51 % with client-M4, while the Markov-chain algorithm with the use of a fixed length of 7 could not achieve more than 25.25 % with the same client. With client-M5, P-DPA_M1 and P-DPA_MN achieved F-measure values of 64.49 and 64.76 %, respectively, whereas Markov-chain with a fixed length of 5 achieved only 31.91 %. Even with the significant shopping history of client-F8, who is a mother of two children, P-DPA_M1 was able to achieve F-measure value of 46.75 % and P-DPA_MN achieved 47.87 %, while Markov-chain with a fixed length of 3 obtained only 7.69 %.
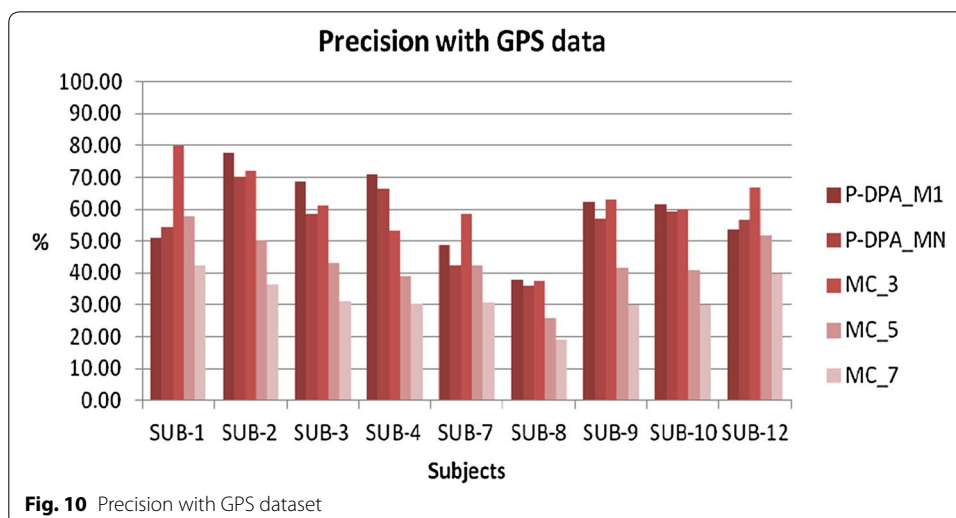

Fig. 8 Recall with V-Mall dataset

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 17 of 29

Fig. 9 F-measure with V-Mall dataset

The last virtual client is client-S2, who has an extremely regular behavior. With that client, `P-DPA_M1` and `P-DPA_MN` was easily able to achieve a F-measure of 100 %, while F-measure value of 22.56 % was the highest value that the Markov-chain algorithm achieved with a fixed length of 3.

Prediction accuracy with GPS dataset

Figure 10 shows the Precision values with every subject in the GPS dataset of [6]. With SUB-1, Markov-chain with a fixed length of 3 achieved a precision value of 79.87 %, but with a different fixed length, there was a significant drop in the precision value. For instance, with the same subject, SUB-1, MC_5 achieved a precision of 57.74 and 42.28 % with MC_7, while `P-DPA_M1` and `P-DPA_MN` using a dynamic length achieved Precision values of 51.06 and 54.21 %, respectively. `P-DPA_M1` with SUB-2 achieved the highest precision value of 77.78 %, and Markov-chain with a fixed length of 3 achieved 72.03 %, while `P-DPA_MN` comes in third place with a precision of 70.26 %. The highest

Fig. 10 Precision with GPS dataset

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 18 of 29

precision values with subjects, SUB-3, SUB-4, SUB-8 and SUB-10, were achievable by `P-DPA_M1` with Precision values of 68.51, 70.85, 38 and 61.70 %, respectively. On the other hand, Markov-chain with a fixed length of 3 achieved the highest precision values with SUB-7 and SUB-12. MC_3 achieved a precision of 58.33 % with SUB-7 while `P-DPA_M1` obtained only 48.81 %. With SUB-12, `P-DPA_MN` achieved a precision of 56.80 % and `P-DPA_M1` achieved 53.66 % whereas MC_3 achieved the highest Precision value of 66.64 %. With SUB-9, both algorithms managed to achieve a high precision value, with `P-DPA_M1` achieving a Precision of 62.27 % and MC_3 achieving 62.90 %.

Figure 11 shows the Recall values with every subject in the GPS dataset. With SUB-1, `P-DPA_M1` achieved a Recall value of 47.04 % and `P-DPA_MN` achieved 37.25 % while Markov-chain with a fixed length of 7 achieved the highest Recall value of 59.49 %. MC_7 also achieved the highest Recall value of 77.03 % with SUB-2 whereas `P-DPA_M1` achieved a Recall of 74.08 % and `P-DPA_MN` achieved 62.17 %. On the other hand, `P-DPA_M1` was able to achieve the highest Recall with SUB-3, SUB-4 and SUB-8, with Recall values of 60.17, 52.54 and 40.16 %, respectively. With those subjects, MC_7 achieved a Recall value of 48.06 % with SUB-3, 48.35 % with SUB-4 and 18.89 % with SUB-8. With SUB-9 and SUB-12, the highest Recall values of 75.21 and 47.16 %, respectively, were achievable by MC_7 while `P-DPA_M1` achieved only a Recall value of 56.23 % with SUB-9 and 43.59 % with SUB-12. The Recall values with SUB-7 were very close, with MC_7 achieving a Recall of 43.90 % and `P-DPA_M1` achieving 42.85 %. With SUB-10, both algorithms managed to achieve almost identical Recall, with MC_7 achieving a Recall value of 65 % and `P-DPA_M1` achieving 64.60 %.

Figure 12 shows the prediction accuracy of both algorithms, `P-DPA` and Markov-chain, through the F-measure values with every subject in the GPS dataset. The highest F-measure value of 57.92 % with SUB-1 was achievable by Markov-chain algorithm with a fixed length of 3, while `P-DPA_M1` and `P-DPA_MN` could not obtain more than 46.17 and 41.45 %, respectively. On the other hand, with subjects, SUB-2, SUB-3, SUB-4, SUB-7, SUB-8, SUB-10 and SUB12, the `P-DPA_M1` achieved the highest F-measure values of 72.51, 60.26, 55.39, 43.15, 38.18, 59.95 and 45.52 %, respectively. With those subjects and
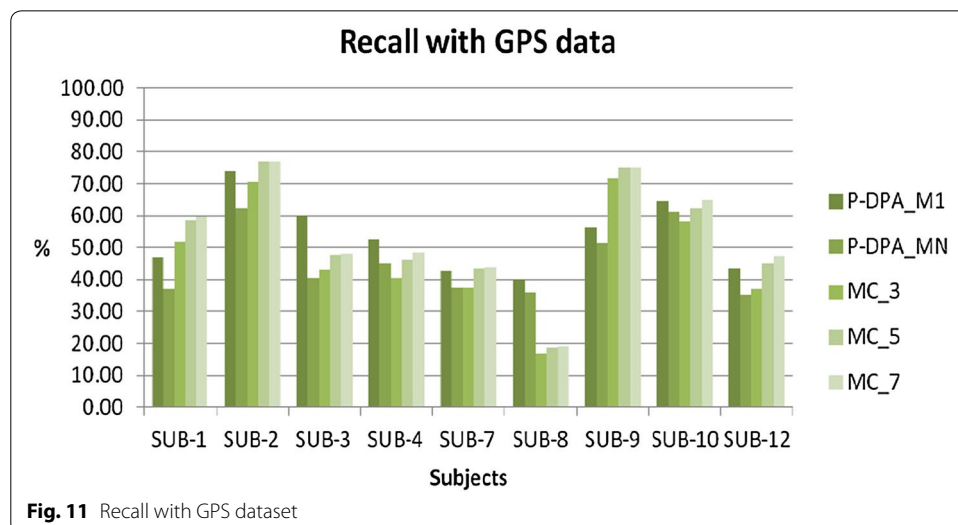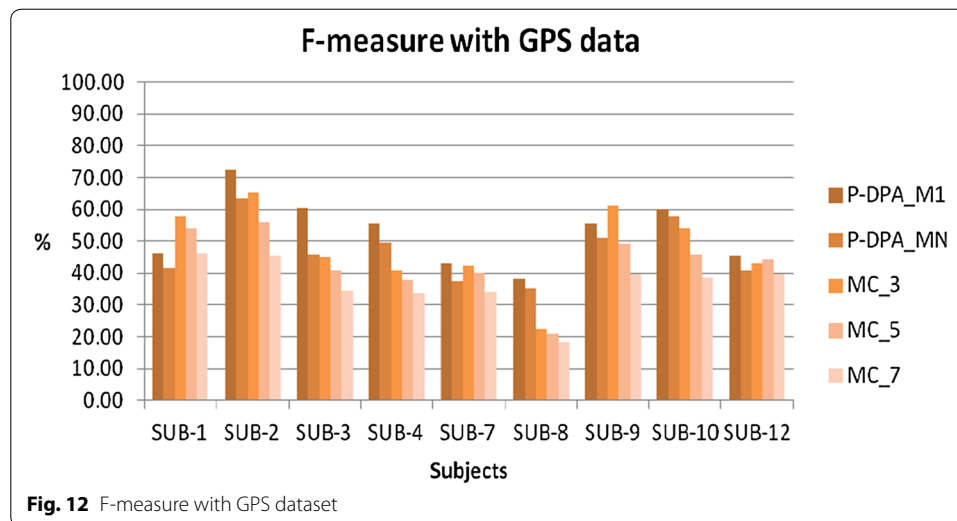

Fig. 11 Recall with GPS dataset

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17
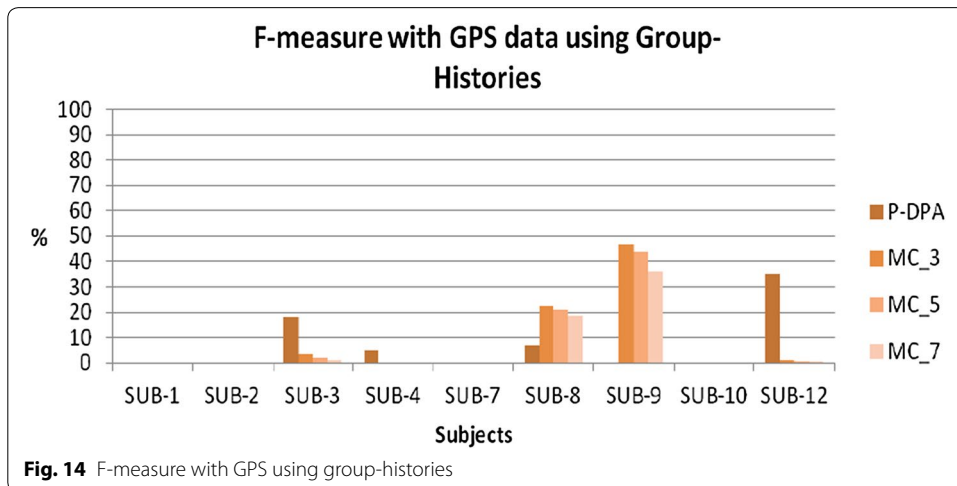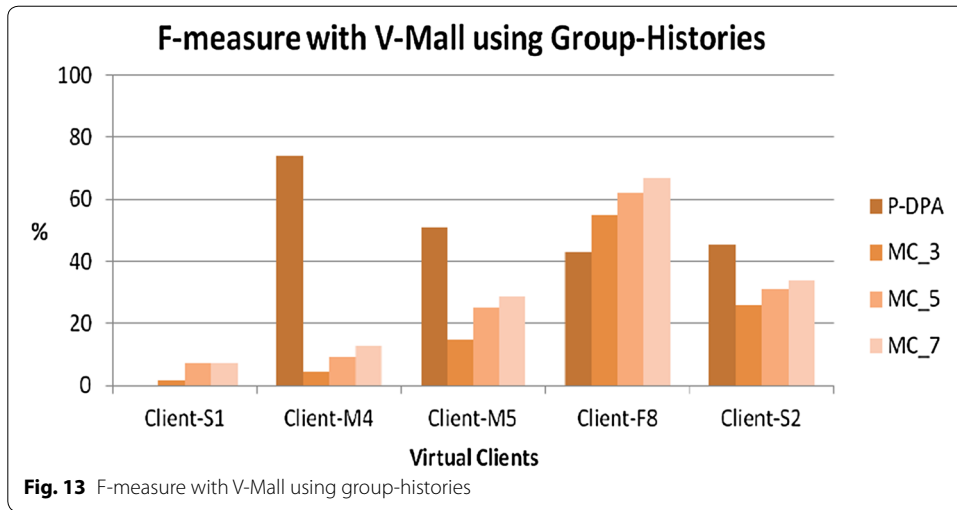
Page 19 of 29

Fig. 12 F-measure with GPS dataset

in the same order, MC_3 achieved recall values of 65.21, 45.14, 40.69, 42.35, 22.24, 54.10 and 43.10 %. With SUB-9, the first place was reserved for MC_3 with F-measure value of 61.24 %, then `P-DPA_M1` in second place with F-measure of 55.63 %, and the third place was for `P-DPA_MN` with F-measure of 50.96 %.

Self-histories vs. group-histories

In comparison to the prediction accuracy results using Self-Histories in Figs. 9 and 12, Fig. 13 shows the F-measure values that represent the prediction accuracy of both algorithms, `P-DPA` and Markov-chain, using group-histories style instead of self-histories with the virtual clients in V-Mall dataset. From Fig. 13, we can see a significant drop in accuracy when using group-histories with the `P-DPA` algorithm. With client-S1, `P-DPA` achieved F-measure of 0 % in comparison to 10.29 % when using Self-histories. With the `P-DPA` algorithm and Group-Histories, the drop in F-measure values can be seen as follows: from 82.33 to 74.16 % with client-M4, from 64.49 to 50.85 % with client-M5, from 46.75 to 43.16 % with client-F8, and from 100 to 45.32 % with client-S2, who has an extremely regular behavior. Also, the prediction accuracy of Markov-chain algorithm with different fixed lengths and by using group-histories suffered from a serious drop in accuracy. With Markov-chain algorithm and group-histories, the drop in F-measure values was as follows: from 22.26 to 7.18 % with client-S1, from 25.25 to 12.89 % with client-M4, and from 31.91 to 28.88 % with client-M5. However, MC_7 and with the use of group-histories was able to achieve a significant increase in F-measure value from 7.38 to 66.87 % with client-F8, who is a mother of two children. In addition, MC_7 achieved another increase in F-measure value from 22.42 to 33.66 % with client-S2.

Figure 14 shows the F-measure values that represent the prediction accuracy of both algorithms, `P-DPA` and Markov-chain, using group-histories instead of self-histories with the subjects in the GPS dataset. From Fig. 14, we can see a significant drop in accuracy when using group-histories with the `P-DPA` algorithm. With the `P-DPA` algorithm and Group-Histories, the drop in F-measure values can be seen as follows: from 46.17 to 0 % with SUB-1, from 72.51 to 0 % with SUB-2, from 60.26 to 17.85 % with SUB-3, from

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 20 of 29

Fig. 13 F-measure with V-Mall using group-histories

Fig. 14 F-measure with GPS using group-histories

55.39 to 4.71 % with SUB-4, from 43.15 to 0 % with SUB-7, from 38.18 to 7.08 % with SUB-8, from 55.63 to 0 % with SUB-9, from 59.95 to 0 % with SUB-10, and from 45.52 to 35.15 % with SUB-12. In addition, with Markov-chain algorithms and Group-Histories, the significant drop in F-measure values was as follows: from 57.92 to 0 % with SUB-1, from 65.21 to 0 % with SUB-2, from 45.14 to 3.41 % with SUB-3, from 40.69 to 0 % with SUB-4, from 42.35 to 0 % with SUB-7, from 61.24 to 46.77 % with SUB-9, from 54.10 to 0 % with SUB-10, and from 44.15 to 1.12 % with SUB-12.

The results show that there is, in this case, greater value in predicting based on what the individual him/her-self did in the past rather than what other people might typically do.

Monday for monday vs. no preselected days

The `P-DPA` algorithm focuses on activities and trajectories that occurred on previous weekdays that are of the same day of the week as the current predicted weekday of a targeted individual. For example, if the current weekday is Monday, then, the `P-DPA`

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 21 of 29

algorithm will focus on recorded trajectories that occurred on up to 2 years of *all previous Mondays*. Figures 9 and 12 show the good prediction accuracy that P−DPA managed to achieve when using this approach. In comparison, Figs. 15 and 16 show the significant drop in F-measure value when using recorded trajectories for up to 2 years of *all previous days*. With V-Mall dataset, there was a drop in F-measure value by 2.19 % with Client-S1, 15.96 % with Client-M4, 6.56 % with Client-M5 and 3.35 % with Client-F8, as shown in Fig. 15. With the GPS dataset, the drop in F-measure value was between 3.30 and 11.85 %, a drop by 4.36 % with SUB-1, 11.72 % with SUB-2, 7.71 % with SUB-3, 11.85 % with SUB-4, 3.30 % with SUB-7, 9.24 % with SUB-8, 4.10 % with SUB-9, 3.60 % with SUB-10 and 11.25 % with SUB-12, as shown in Fig. 16.

The results show that selecting histories of the same day of the week as the day whose D-trajectory is to be predicted helps, rather than simply using the history of all days.
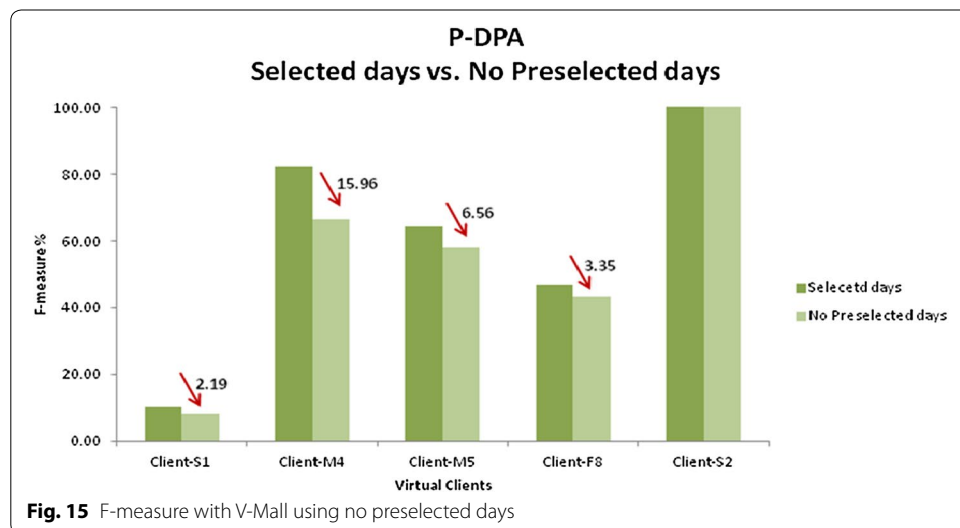

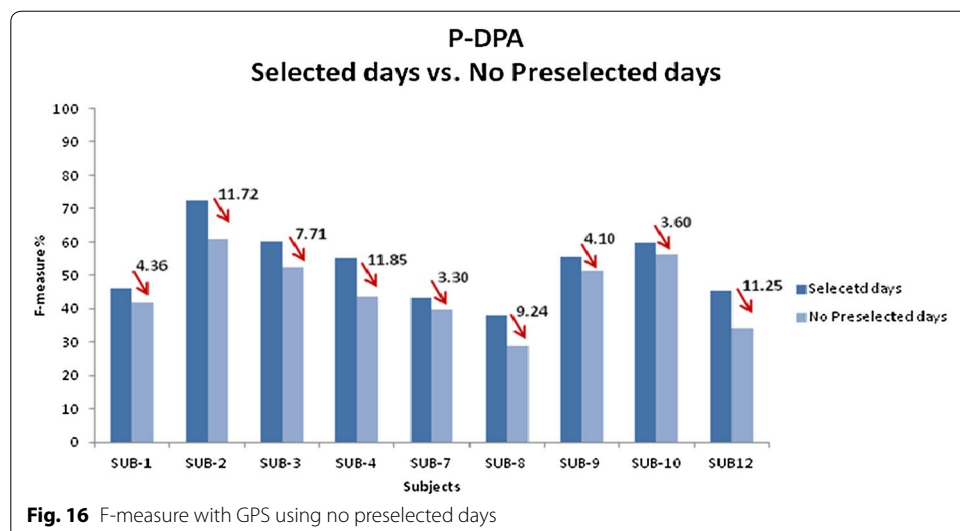**Fig. 15** F-measure with V-Mall using no preselected days

Fig. 16 F-measure with GPS using no preselected days

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 22 of 29

Discussion

Based on the experimental results, we make the following observations.

- Compared to the use of a predefined length, using a dynamic length to predict the number of destinations that could be visited by a targeted individual allowed the P-DPA algorithm to achieve good prediction accuracy with most subjects and virtual clients, as shown in Figs. 9 and 12.

- Figure 17 shows the average F-measure with all virtual clients in V-Mall dataset. The P-DPA algorithm was able to achieve the highest average F-measure of 60.77 %, while with Markov-chain algorithm, the best average F-measure of 21.71 % was achievable when using a fixed length of 7. Figure 18 shows the average F-measure with all subjects in the GPS dataset of [6]. The P-DPA algorithm achieved the highest average F-measure of 52.97 % and Markov-chain algorithm with a fixed length of 3 achieved an average F-measure of 48 %. The problem of showing only the average F-measure is that it does not show exactly how accurate the predictions of a prediction algorithm are for each individual. For example, as shown in Fig. 17, the average F-measure of 60.77 % achieved by the P-DPA algorithm does not explicitly show how the prediction algorithm was able to achieve a significant F-measure value of 100 % with client-S2 or a low F-measure of 10.29 % with client-S1. In addition, with the GPS dataset, the P-DPA algorithm achieved a high F-measure value of 72.51 % with SUB-2 and low F-measure of 38.18 % with SUB-8, but both results are hidden inside the average F-measure value of 52.97 %, as shown in Fig. 18.

- In our work, we focus on the self-history of an individual in order to produce predictions related only to that individual. We also show the accuracy of our prediction algorithm with every single individual. Accordingly, in order to have a better idea of why with some individuals the results were significantly better and why with others our algorithm was struggling to produce an accurate prediction, we measured the behavior regularity of every individual and show how such level of regularity could affect the prediction accuracy of our P-DPA algorithm, positively or negatively. Williams et. al. [13] defined regularity as repeated activity over time. For example, in their approach, the behavior of a targeted individual can be described as highly regu-
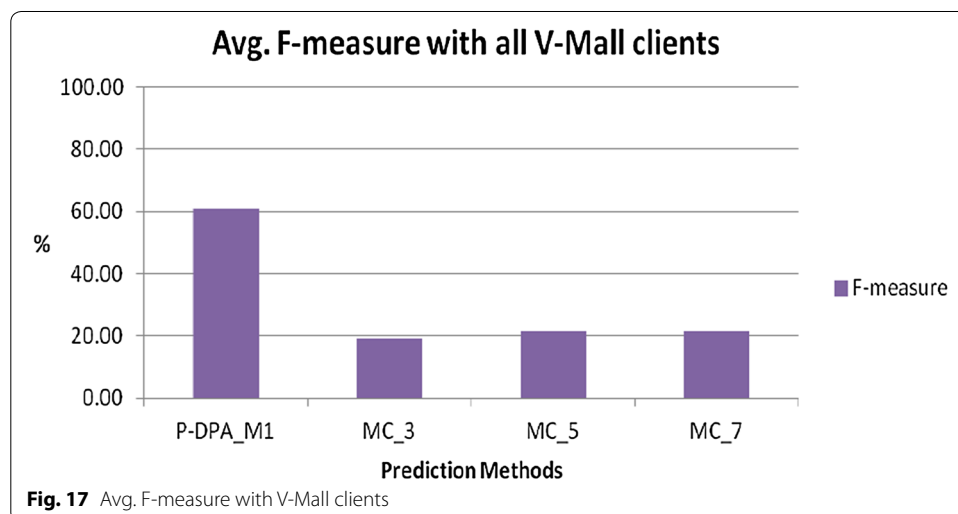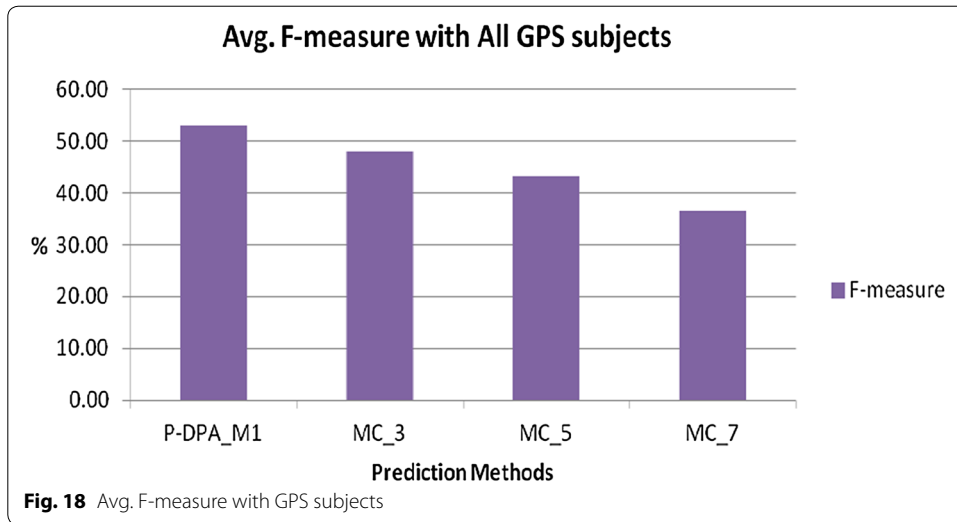

Fig. 17 Avg. F-measure with V-Mall clients

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 23 of 29

Fig. 18 Avg. F-measure with GPS subjects

lar when that individual is visiting a location at very similar times each week. They tested their approach on three datasets and they found that the majority of individuals in those datasets are deemed to have *high irregular behavior*. The work in [13] found that the three datasets have fine-grained scale and two of the datasets are at a city-wide scale, and hence, high irregular behavior for the majority of individuals might have been an expected result by them.

Here, we use a different approach to measure the behavior regularity of a targeted individual. The level of regularity for a targeted individual is assessed by using two types of regularity. The first type of regularity is *Destination regularity*, which is about measuring the tendency of a targeted individual to visit a *similar set* of destinations on the same day of the week, e.g. Monday this week is similar to Monday next week. Accordingly, we are trying to find an answer to the following question: what are the chances of visiting the same *set* of destinations that were visited on the last same day of the week?

We use the following formula to measure the destination regularity of a targeted individual.

$$\mu(W_k) = \frac{1}{N-1} \sum_{i=1}^{N-1} ((W_k(L_i) \bigcap W_k(L_{i+1}))/|W_k(L_i)|) \tag{5}$$

where

- $W_k \in W = \{Mon, Tue, Wed, Thu, Fri, Sat, Sun\}$
- $W_k(L_i) = $ *a set of locations visited on previous* W_k
- $W_k(L_{i+1}) = $ *a set of locations visited on next* W_k.

The second type of regularity is *Length Regularity*, which is about measuring the tendency of a targeted individual to visit the *same number* of destinations on the same day of the week. Consequently, we are trying to find an answer to the following question:

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 24 of 29

what are the chances of visiting the same *number* of destinations that were visited on the last same day of the week?

We use the following formula to measure the length regularity of a targeted individual.

$$\mu'(W_k) = \frac{1}{N-1} \sum_{i=1}^{N-1} abs(|W_k(L_{i+1})| - |W_k(L_i)|) \tag{6}$$

$$\sigma(W_k) = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (abs(|W_k(L_{i+1})| - |W_k(L_i)|) - \mu'(W_k))^2} \tag{7}$$

where $\sigma(W_k)$ is the standard deviation of length difference in all trajectories of subject(x) on W_k.

These two types of regularity, destination and length, were chosen because of their direct impact on our `P-DPA` algorithm. The process to predict the number of destinations that could be visited by a targeted individual will be affected by the level of regularity of that individual regarding the length of his/her trajectory from one weekday to another similar weekday. In addition, the `P-DPA` algorithm is using an initial distribution matrix and a transition matrix. Both matrices will be affected by the level of regularity of a targeted individual regarding the set of destinations that he/she tends to visit from one weekday to another similar weekday.

Figures 19 and 20 show the average destination regularity value for every subject in the GPS dataset for every weekday, and Fig. 21 shows the average and standard deviation for length difference regularity for the same subjects of the GPS dataset. The `P-DPA` algorithm managed to achieve F-measure value of 72.51 % with SUB-2 and F-measure of 38.18 % with SUB-8. Both subjects have a fairly substantial level of regularity regarding visiting similar set of destinations, as shown in Fig. 19. However, Fig. 21 shows that SUB-2 managed to have good average and standard deviation values regarding the regularity of visiting similar number of destinations on the same day of the week, but SUB-8 has the worst average and standard deviation values on most weekdays compared to
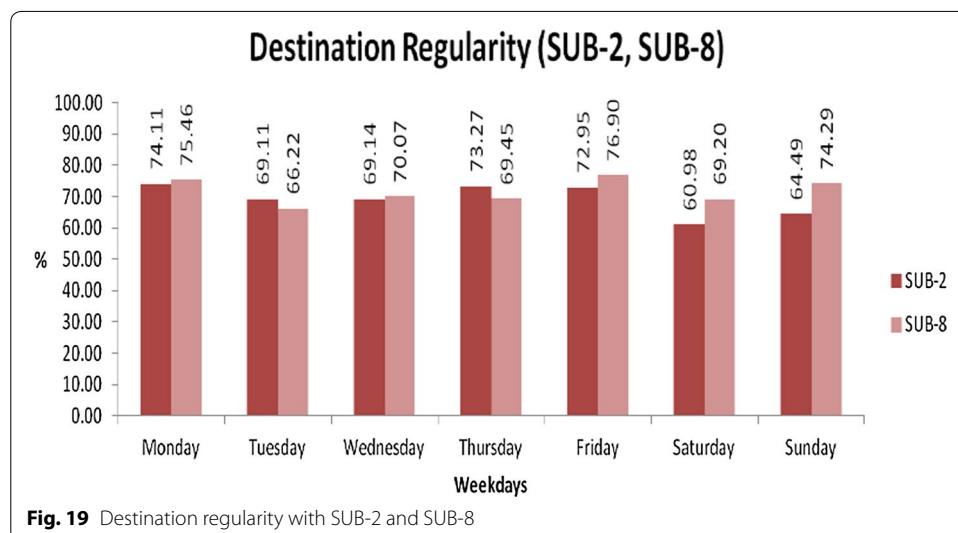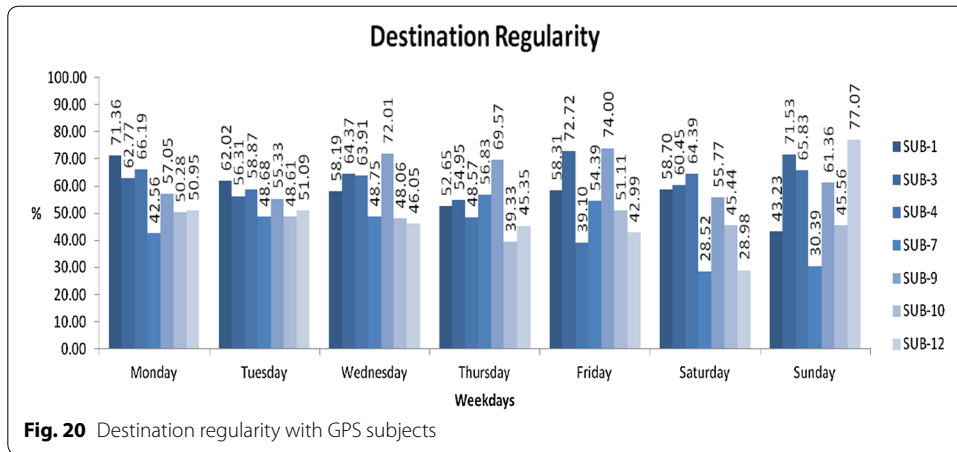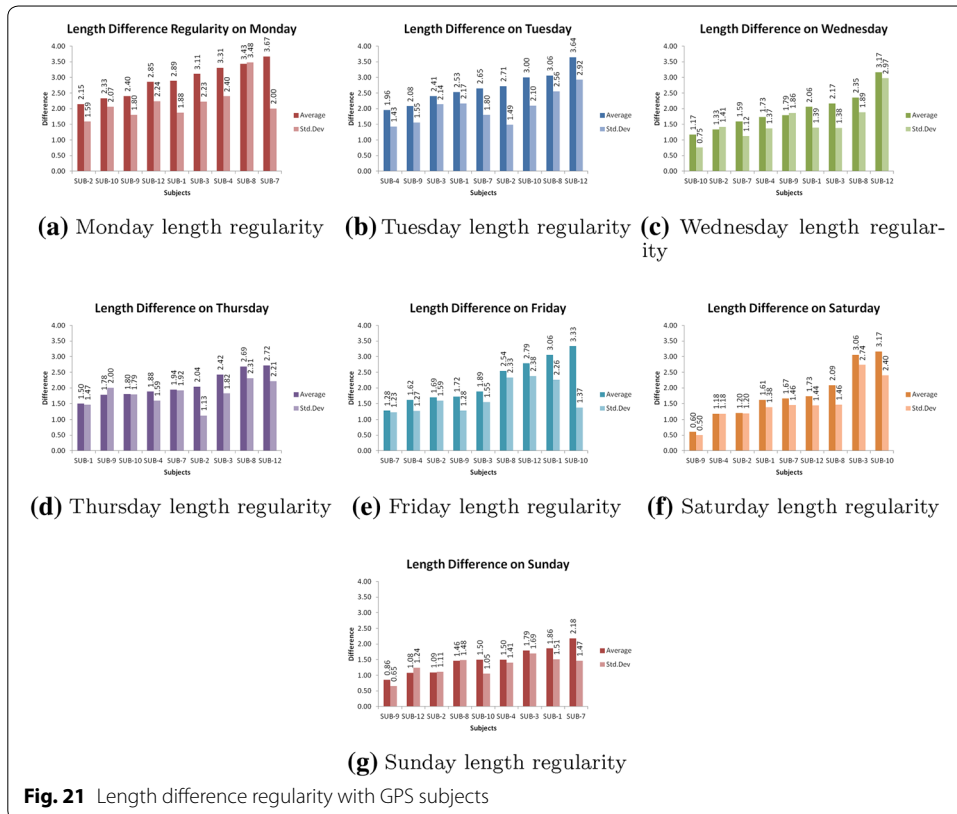

Fig. 19 Destination regularity with SUB-2 and SUB-8

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 25 of 29

Fig. 20 Destination regularity with GPS subjects

(a) Monday length regularity **(b)** Tuesday length regularity **(c)** Wednesday length regularity

(d) Thursday length regularity **(e)** Friday length regularity **(f)** Saturday length regularity

(g) Sunday length regularity

Fig. 21 Length difference regularity with GPS subjects

other subjects. Consequently, the good level of regularity of SUB-2 in both types, destination and length, helped `P-DPA` algorithm to achieve F-measure of 72.51 % whereas the behavior of SUB-8, who has a high level of regularity regarding the set of destinations and a low level of regularity regarding the number of destinations, hindered the `P-DPA` algorithm from capturing a distinctive D-trajectory pattern, which resulted in a low F-measure value of 38.18 %. The behavior regularity level for the remaining subjects can be assessed using the same two types of regularity, which helps explain why with

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 26 of 29

some individuals the results were significantly better while, with others, our algorithm was struggling to produce accurate predictions. `P-DPA` successfully exploits regularities when present.

System prototype

We used the Android platform to develop SmartShopper, a personal destination-pattern-aware pervasive system, in order to test the interaction between a client's mobile phone and the server of `V-Mall`. Figure 22 shows the welcome screen of SmartShopper.

Figures 23 and 24 show the `P-DPA` based predicted set of stores that could be visited by the targeted client and the prediction processing time, which is 0.14 seconds.

Then, the client can click on any of the predicted stores and a list of advertisements will be generated and sent to him/her, as shown in Fig. 25. Figure 26 shows the details of a chosen ad by the targeted client.

Conclusion and future work

In this paper, we have introduced a trajectory prediction algorithm, `P-DPA`, that uses the self-history of a targeted individual in order to predict the set of destinations that could be visited by him/her. We show that self-histories can provide reasonable predictions of future destinations, which can be exploited for ads. In future work, extensive analysis of the behavior of an individual is required in order to be able to identify which other behavior attributes can affect the prediction accuracy of `P-DPA`, positively or negatively. Other personal profile information can be integrated to possibly improve the predictions such as past histories of transactions or purchases and current situations of the user.

Fig. 22 SmartShopper welcome screen

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17
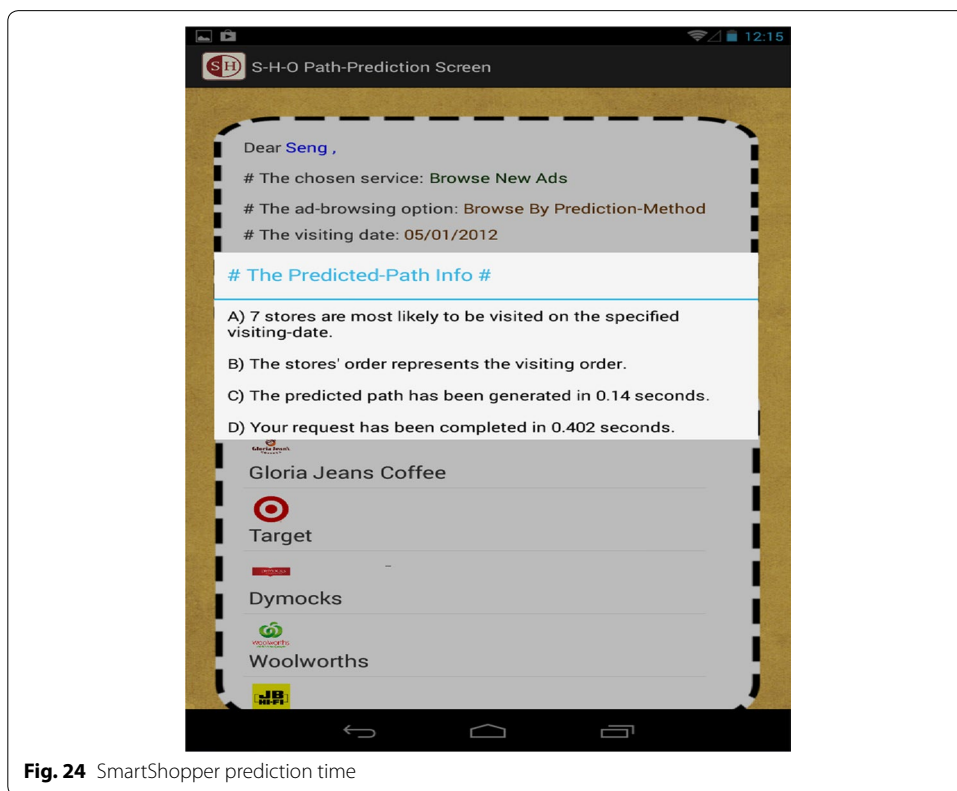
Page 27 of 29

Fig. 23 SmartShopper predicted stores

Fig. 24 SmartShopper prediction time

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 28 of 29

Fig. 25 SmartShopper ads-list

Fig. 26 SmartShopper ad details

Barzaiq and Loke *Hum. Cent. Comput. Inf. Sci.* (2016) 6:17

Page 29 of 29
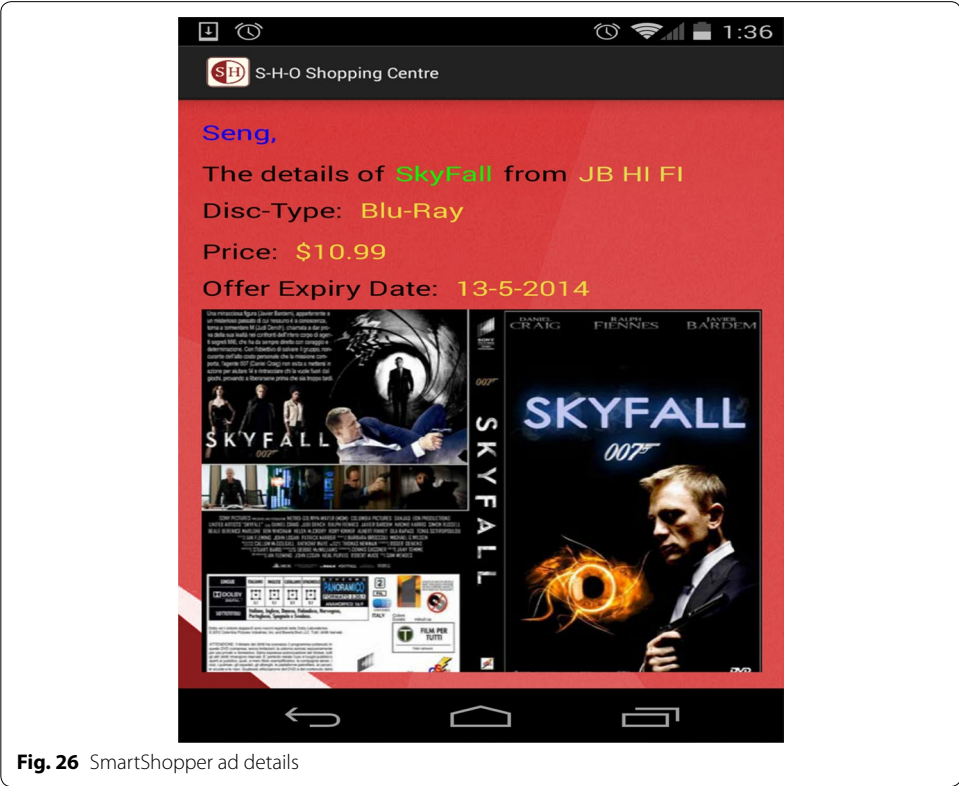
We think that our approach of using personal/self-histories, rather than the combined histories of many people, can provide more personalised ads, and will address issues of variations of behavior across individuals. Further work will experimentally compare and contrast predictions that use the self-history of what a person typically does with predictions that use the combined histories of what people typically do in other scenarios.

Authors' contributions
Both coauthors contributed significantly to the research and this paper, and the lead author is the main contributor. Both authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

References
1. Dhar S, Varshney U (2011) Challenges and business models for mobile location-based services and advertising. Commun ACM 54(5):121–128
2. Krumm J (2011) Ubiquitous advertising: the killer application for the 21st century. IEEE Pervasive Comput 10(1):66–73
3. Kim B, Ha JY, Lee S, Kang S, Lee Y, Rhee Y, Nachman L, Song J (2011) Adnext: a visit-pattern-aware mobile advertising system for urban commercial complexes. In: Proceedings of the 12th workshop on mobile computing systems and applications. ACM pp. 7–12
4. Sánchez JM, Cano JC, Calafate CT, Manzoni P (2008) Bluemall: a bluetooth-based advertisement system for commercial areas. In: Proceedings of the 3rd ACM workshop on performance monitoring and measurement of heterogeneous wireless and wired networks. ACM pp. 17–22
5. Liapis D, Vassilaras S, Yovanof GS (2008) Implementing a low-cost, personalized and location based service for delivering advertisements to mobile users. In: 3rd international symposium on wireless pervasive computing 2008. ISWPC 2008, IEEE pp. 133–137
6. Chon Y, Shin H, Talipov E, Cha H (2012) Evaluating mobility models for temporal prediction with high-granularity mobility data. In: IEEE International Conference on pervasive computing and communications (Percom) 2012. IEEE pp. 206–212
7. Asahara A, Maruyama K, Sato A, Seto K (2011) Pedestrian-movement prediction based on mixed markov-chain model. In: Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems. ACM pp. 25–33
8. Zhang D, Xia F, Yang Z, Yao L, Zhao W (2010) Localization technologies for indoor human tracking. In: 5th international conference on future information technology (FutureTech), 2010. IEEE pp. 1–6
9. Kolodziej J, Khan SU, Wang L, Min-Allah N, Madani SA, Ghani N, Li H (2011) An application of markov jump process model for activity-based indoor mobility prediction in wireless networks. In: Frontiers of information technology (FIT). IEEE pp. 51–56
10. Gambs S, Killijian MO, del Prado-Cortez MN (2012) Next place prediction using mobility markov chains. In: Proceedings of the first workshop on measurement, privacy, and mobility. ACM p. 3
11. Mathivaruni R, Vaidehi V (2008) An activity based mobility prediction strategy using markov modeling for wireless networks. In: Proc of the world congress on engineering and computer science 2008 WCECS. Citeseer pp. 379–384
12. Barzaiq O, Loke SW, Lu H (2015) On trajectory prediction in indoor retail environments for mobile advertising using selected self-histories. In: Proceedings of the 12th IEEE international conference on ubiquitous intelligence and computing (UIC 2015)
13. Williams MJ, Whitaker RM, Allen SM (2012) Measuring individual regularity in human visiting patterns. In: Privacy, security, risk and trust (PASSAT), 2012 international conference on social computing (SocialCom). IEEE, pp 117–122