International Journal of Pervasive Computing and Communications

Comparison of context-aware predictive modeling approaches: Semantic place in inferring mobile user behavior

Tapio Soikkeli

Comparison of context-aware predictive modeling approaches

Semantic place in inferring mobile user behavior

Tapio Soikkeli
Department of Communications and Networking, Aalto University, Espoo, Finland

Received 22 January 2015
Revised 14 April 2015
Accepted 14 April 2015

Abstract

Purpose – The aim of this paper is to empirically examine how to best incorporate such contextual data, such as location or the semantic place of mobile users, into mobile user behavior models. Acquiring such data has become technically easier than ever. The proper utilization of these data leads, hypothetically, to better understanding of mobile user behavior and, consequently, to enhanced mobile services.

Design/methodology/approach – The paper systematically compares, under multiple experimental settings, the predictive performances of models built with three different approaches (pre-filtering, contextual modeling and post-filtering) used for incorporating contextual data into user behavior models. The comparisons focus on by which approach additional semantic place information can be best utilized for making the most accurate inferences on mobile user behavior. Real-life smartphone usage data are utilized in the analysis.

Findings – The results demonstrate that none of the considered approaches uniformly dominate the others across all experimental settings. However, they show circumstance specific differences that need to be aligned with practical use cases for the best performance.

Practical implications – Identifying the most suitable approaches for utilizing the semantic place (and other contextual) data is an important practical problem for electronic service providers, whose offerings are increasingly moving to the mobile domain and thus need to respond to the demands of mobility.

Originality/value – The paper constitutes an initial step toward understanding and systematically evaluating different approaches for incorporating semantic place data into modeling mobile user behavior. Practitioners in the mobile service domain can apply the initial results and academics build upon them with more diverse experimental settings.

Keywords User behavior, Context, Machine learning, Semantic place, User modeling

Paper type Research paper

1. Introduction

Modern connected mobile devices such as smartphones and tablets are able to generate a constant stream of behavioral data related to their users' mobile usage habits, whereabouts and surroundings. It is not a surprise that actors from governments through companies to individual people are seeking ways to leverage these data: governments to keep order and reduce crime; companies to enhance their services to increase profits; and individual people to seek, e.g. health and wellness improvements in the spirit of the Quantified Self movement. A rich variety of use cases for the behavioral and contextual information inferred from these data has emerged, context-aware mobile

International Journal of Pervasive Computing and Communications
Vol. 11 No. 3, 2015
pp. 323-346

1742-7371
DOI 10.1108/IJPCC-01-2015-0003

applications and services not among the least of them. Some of the more straightforward uses include pure GPS (Global Positioning System) coordinates for maps and navigation, whereas, in more complicated cases, contextual information-based modeling and the usage of, e.g., predictive algorithms are required. Examples of these include context-aware mobile recommender systems, whether they recommend apps, services or content; faster app launching (Yan *et al.*, 2012) or profile and battery management based on context-aware prediction; and context-, application- and service-specific adaptation operations (Baladron *et al.*, 2012) for network resource allocation. The basic idea is that by adding contextual information into the system, the system can adapt better to the users' needs. The envisioned benefits from this in the mobile services domain relate to better usability, enhanced quality of experience for the users and, at the end, increased profits for the companies.

Nowadays, collecting behavioral and contextual data is technically easier than ever. For example, if granted permission, a smartphone application can monitor, in addition to the app usage itself, the user's location, cellular and WiFi connections or other nearby devices and send these data to the service provider. Naturally, just accumulating data is not valuable. Suitable approaches are needed to incorporate the data in proper manner into the respective system (e.g. a smartphone application), and then make sure the output is the desired one. Place-related contextual data, such as pure GPS location or the semantic place, have traditionally been at the forefront when considering context-aware mobile devices and services. The reason is that the mobile nature of the devices allows users to use the services in different places, and, on the other hand, the devices themselves have been equipped with location tracking sensors already for many years. Moreover, the place and the meaning of the place for the user are regarded as important determinants of mobile user behavior. Thus, semantic place data collection and inference are rather explored areas (Laurila *et al.*, 2013). Also, applying the semantic place information to context-aware implementations has gained attention, but individual studies have done it in more or less an *ad hoc* manner and in relatively narrow use cases. Systematically examining how the additional semantic place data or the refined information can be best incorporated into behavioral models behind the implementations has gained less attention.

Comparing different incorporation approaches is important not only to the academic community but also to the industry for several reasons. For some time, providers of electronic services have tapped into the usage data (e.g. user demographics, usage frequencies or diurnal usage patterns) of their services to provide, for example, more personalized offerings. The services are, however, moving to the mobile domain at an increasing pace and forcing the companies to update their traditional service offerings to respond to the mobile nature of service usage, such as changing locations or semantic places. Although, moving to the mobile domain is practically inevitable, the services differ in ways and degrees needed for responding to the demands of mobility. Therefore, investigating the strengths and weaknesses of different, e.g. semantic place data; incorporation approaches under different circumstances; and finding the most suitable approach is an important problem for the service providers.

In this work, we compare three different approaches to incorporate semantic place information (refined location data) into models inferring mobile user behavior. In addition, we examine how the models built with the approaches compare with models without any semantic place information. The three approaches are *Pre-filtering (PreF)*, *Contextual modeling (CM)* and *Post-filtering (PoF)* and have been initially introduced in the domain of recommender systems (Adomavicius *et al.*, 2011). The semantic places considered are *Home*, *Office/School*, *Other meaningful place* and *Elsewhere*, and the data utilized in the work are collected directly from smartphone users' devices utilizing the handset-based measurement method (Karikoski, 2012). Thus, we have two main research questions in this work:

RQ1. How models built with semantic place-based *PreF*, *CM* and *PoF* approaches compare performance-wise with each other, when inferring mobile user behavior?

RQ2. How models built with semantic place-based *PreF*, *CM* and *PoF* approaches compare performance-wise with models without semantic place information, when inferring mobile user behavior?

In the article, we answer the research questions empirically by examining two handset-based datasets across a range of different experimental conditions. We build models including semantic place information with all of the three aforementioned approaches, and models without semantic place information. We compare, then, the predictive performance of the different types of models. The article makes the following contributions to contextual mobile user behavior and smartphone usage studies:

- it focuses on how to technically utilize and incorporate semantic place information to mobile user behavior models;
- applies three incorporation approaches (mentioned above) from the recommender system domain to the mobile user behavior domain with relevant modifications;
- systematically compares the approaches under several experimental conditions;
- systematically compares the approaches to an approach that ignores semantic place information; and
- accompanies previous studies in inferring semantic place information from handset-based data.

The article is organized as follows. In the next chapter, we review the related work. In Chapter 3, we formulate the problem at hand by explaining how we understand the semantic place, and set the conceptual basis for semantic place information inference and mobile user behavior modeling. Chapter 4 goes through our experimental setup and Chapter 5 presents and discusses the results. Chapter 6 concludes the article.

2. Related work

Mobile end-user context is considered an important element in examining and modeling the end users' behavior. Thus, a relatively large and diverse body of previous research investigating the effect of different contextual aspects on the behavior of mobile users has emerged. Due to the technical advancements in mobile devices (i.e. the introduction of the smartphone), the research has been able to move from the more traditional survey questionnaire type of an approach, such as in Liang and Yeh (2011) and Xu and Yuan (2009), into utilizing device monitoring or the so called handset-based measurements (Karikoski, 2012). Measuring the users' mobile usage directly from their devices enables large-scale experiments and more objective results. Notable handset-based measurement based studies and data collection efforts include the Reality Mining project described in Eagle and Pentland (2006), the Lausanne Data Collection Campaign and the related Mobile Data Challenge (Laurila *et al.*, 2013) and the Copenhagen Networks Study (Stopezynski *et al.*, 2014). Programmability of smartphones paves the way for researchers to develop and distribute suitable data collection platforms, including the Context-Phone (Raento *et al.*, 2005) and more recently the Funf Framework (Aharony *et al.*, 2011).

These types of handset-based data collection platforms enable the collection of diverse data highly suitable for end-user, context-related studies. One of the important tasks has been inferring various contextual elements directly from the users' data. The places users visit and spend time in, and especially the meanings of these places for the users, are considered to affect different aspects of user behavior. Thus, it is not a surprise that a considerable effort has been put into identifying these semantic places (Bayir *et al.*, 2010; Huang *et al.*, 2012; Isaacman *et al.*, 2011; Soikkeli, 2011; Verkasalo, 2008; Zhu *et al.*, 2013). The methodologies vary from user participation and relatively heuristic models to various machine learning and classification methods. From a theoretical point of view, the inference methods and achievable accuracies are interesting, but the real value of the inferred contextual information materializes when it is combined with some other data or information. A direct next step is to model and predict the users' next places and movements based on historical data (Ashbrook and Starner, 2003; Etter *et al.*, 2012; Mayrhofer *et al.*, 2003; McInerney *et al.*, 2013). Implementations of this type of modeling help, for example, in predicting the spread of diseases and in infrastructure planning. Mobile services are expected to become increasingly more personalized for the users' unique needs, and contextual information, such as location or place, is identified as a key enabler for this (Skillen *et al.*, 2014). For example, personalized Help-on-Demand services (Burns *et al.*, 2012) utilize context-based modeling. Other areas for implementing various context-based modeling in the mobile domain are context-aware content delivery (Lungaro *et al.*, 2011), fast app launching based on context-based prediction (Yan *et al.*, 2012), context-aware battery management (Ravi *et al.*, 2008), context-aware route recognition (Mazhelis *et al.*, 2011) and context-aware mobile recommender systems (Baltrunas *et al.*, 2012), to name a few.

High-level paradigms for user modeling include data mining-based methods and knowledge-based methods (Chen *et al.*, 2014). Also, hybrid methods, such as those by Chen *et al.* (2014), have been introduced. The context-aware predictive modeling approaches considered in this article fall mainly into the paradigm of data mining-based methods, and were first introduced in the domain of recommender systems (Adomavicius *et al.*, 2011). According to Adomavicius *et al.* (2011), pre-filtering means that ratings data for recommendations are filtered based on some relevant context before feeding to the recommender system. Contextual modeling means that the additional contextual information is used inside the recommendation-generating algorithms. Finally, post-filtering means that the recommendations are generated traditionally without the contextual information, but then afterwards modified according to the contextual information. Panniello *et al.* (2014) compares performance differences between the approaches in the recommender system/e-commerce domain. The author is not, however, aware of previous work systematically examining and comparing these types of approaches in the mobile user behavior domain or with semantic place information in any domain.

3. Problem formulation

3.1 What is semantic place?

A person's semantic place differs from a pure location or a general place, in that the place or location has some distinguishable meaning for that particular person. The same location or a general place can have different meanings for different people. For example, a house at a particular location might be a home for one person, but a friend's place for another person. A restaurant might be a recreational place for one person, but a workplace for another person. Generally, people might behave somewhat differently depending on what kind of a semantic place they are in. We believe this assumption applies also for mobile user behavior. For example, the set of applications used, usage session durations, the intensity of usage or the type of content watched or downloaded might vary depending on the semantic place of the user. Evidence of changes of this kind in mobile user behavior has been observed by Karikoski and Soikkeli (2013), Soikkeli *et al.* (2013) and Verkasalo (2008).

The semantic places we consider in this article are: *Home*, *Office/School*, *Other meaningful place* and *Elsewhere*. As a semantic place *Home* is quite self-explanatory. By *Office/School*, we mean the place the user works or studies in. *Other meaningful* refers to a place which does not have the characteristics of a *Home* or an *Office/School*, but which the user still considers a significant place in her life. These kinds of semantic places might be, for example, a friend's place, a parents' place or a place for a hobby. In general, *Elsewhere* refers to something other than a significant place for the user.

In some of the previous mobile user behavior research, semantic place is considered as a part of the end user's context. Several definitions of context appear in the literature in which components such as the location of the user, the identity of the people near the user, objects around the user, interests and emotional status of the user, date, season, the temperature, etc. have been mentioned (Brown *et al.*, 1997; Schilit and Theimer, 1994; Schmidt *et al.*, 1999). One of the most cited and well-accepted definitions of context is provided by Dey (2001). It states that: "Context is any information that can be used to characterize the situation of an entity". Based on this definition, semantic place can also be recognized as context.

3.2 Inferring semantic place information

At least two principal approaches exist for acquiring semantic place information in the mobile domain: ask directly from the user or infer based on some other data. In the "ask directly" approach, we enquire from the users at which semantic place they are currently (currently meaning, e.g. just after using some application). In principle, it is possible to push these questions directly to the users' devices from capable enough handset-based measurement platforms. The inference approach is indirect and requires suitable data (preferably in the form of time-stamped logs), such as GPS coordinates, cell ID data and/or data on surrounding WiFi beacons. The inference problem can be specified by a model:

$$C = f(X_1, X_2, \ldots, X_p) \tag{1}$$

where the dependent variable C is the semantic place information, e.g. as in our case: $C = \{Home, Office/School, Othermeaning ful place, Elsewhere\}$. The independent variables $X_1, X_2, \ldots, X_p$ are based on the additional information we are able to collect, e.g. through the handset-based measurements.

3.3 Mobile user behavior modeling

Let us have a user base $\boldsymbol{U}$ represented by N users. Each user U_i is described by a set of m demographic attributes $\boldsymbol{D}_i = \{D_{i1}, D_{i2}, \ldots, D_{im}\}$ and a set of r application sessions $\boldsymbol{AS}_i = \{AS_{i1}, AS_{i2}, \ldots, AS_{ir}\}$, where each application session AS_{ij} of user U_i is described by a set of p application sessions attributes $\boldsymbol{A}_{ij} = \{A_{ij,1}, A_{ij,2}, \ldots, A_{ij,p}\}$. Finally, we have semantic place (or other contextual) information C associated with each application session AS_{ij}.

Table I illustrates a snapshot of the user table containing demographic, application session and contextual information of user U_i. The user can be described by demographic attributes, such as $\boldsymbol{D}_i = \{UserID, Gender, Age, WorkStatus\}$, by a number of, say, 200 application sessions; $\boldsymbol{AS}_i = \{AS_{i1}, AS_{i2}, \ldots, AS_{i200}\}$, where each application session is described by the application session attributes, such as, $A_{ij} = \{AppName, SessionDuration, PreviousApp, IdleTimeBeforePreviousApp, Timestamp\}$; and by contextual information C which, in the scope of this paper, denotes the semantic place of the user. The structure in general supports any kind, or more than one type, of contextual information.

Now, the basic form of the model used for inferring mobile user behavior is:

$$Y = f(X_1, X_2, \ldots, X_p) \tag{2}$$

where the dependent variable Y is one of the application session attributes A_j, and the independent variables $X_1, X_2, \ldots, X_p$ include all of the demographic attributes $\boldsymbol{D}$ and all of the application session attributes $\boldsymbol{A}$, except the attribute A_j, which was chosen to be the dependent variable. The performance of the model is measured by using suitable performance metrics. For example, the inference model can be a Naïve Bayes classifier built on the data of k similar users $U_1, U_2, \ldots, U_k$ for the purpose of inferring A_j "*AppName*" using all the demographic attributes and all application session attributes, except A_j. Suitable performance metrics are, for example, the model's predictive accuracy or the Area Under ROC (Receiver Operating Characteristic) Curve computed by using cross validation or out-of-sample data for evaluation. Models of type (2) do not consider any semantic place information. Let us define next models into which semantic place information is incorporated based on the *PreF*, *CM* and *PoF* approaches. Let us assume that f produces an n-tuple $(P(a_1), P(a_2), \ldots, P(a_n))$, i.e. an ordered instance value set, which indicates as a probability $P(a_k) = p_k$ the degree to which the instance

Table I. Conceptual data table of attributes used in mobile user behavior modeling

Users U	App session AS	Demographic attributes D			App session attributes A			Context C
U_i	AS_{i1}	D_{i1}	…	D_{im}	$A_{i1,1}$	…	$A_{i1,p}$	…
U_i	AS_{i2}	D_{i1}	…	D_{im}	$A_{i2,1}$	…	$A_{i2,p}$	…
…		…	…	…	…	…	…	…
U_i	AS_{ir}	D_{i1}	…	D_{im}	$A_{ir,1}$	…	$A_{ir,p}$	…
…	…	…	…	…	…	…	…	…

(that is, application session) described by $\boldsymbol{X}$ belongs to class a_k. $a_1, \ldots, a_n$ are the possible values of Y, that is, the possible values of the chosen A_j. Then we can have:

$$Y = \arg\max f_{C=b}(X_1, X_2, \ldots, X_p) \tag{3a}$$

$$Y = \arg\max f(X_1, X_2, \ldots, X_p, C) \tag{3b}$$

$$Y = \arg\max [f(X_1, X_2, \ldots, X_p) \cdot \boldsymbol{w}_{C=b}] \tag{3c}$$

Model (3a) represents the *PreF* approach in which only application sessions associated with a particular value of the context attribute $C = b$ are used for building the model. In other words, the application sessions used in the model are filtered based on the contextual information. For example, if we want to build a model for mobile user behavior at home, then only those application sessions are considered which have $C =$ "*Home*". In model (3b), which represents the *CM* approach, the contextual attribute is considered as one of the independent variables and is used alongside the demographic and application session attributes for inferring Y. In model (3c), which represents the *PoF* approach, $\boldsymbol{w}_{C=b} = (w(a_1), w(a_2), \ldots, w(a_n))$ denotes weights for $a_1, \ldots, a_n$, and it is conditional to context attribute $C = b$. In this work, the weights are calculated as the change in the probabilities of $a_1, \ldots, a_n$ appearing in the application sessions having occurred at all semantic places versus at the certain semantic place b. Additionally, if class value a_1 does not appear at all in b, then $w(a_1) = 0$. The $(\cdot)$ indicates element-wise multiplication, which modifies the original n-tuple proportional to the weights. The modification method is rather conservative. If the original classifier f of (3c) produces instance probabilities close to each other, then the class value appearing relatively more often in b than in all semantic places gets chosen. More importantly, if a class value does not appear at all in b, then it cannot be chosen. This is a combination of the two post-filtering methods (called *weight* and *filter*) utilized by Panniello *et al.* (2014). An extreme approach to post-filtering is deciding beforehand (based on some a priori knowledge) one outcome per semantic place and then weighing the other outcomes to zero, while f is trivial. For example, a user can decide that her ring tone should automatically change to silent when in school and to normal otherwise. This is the simplest method of adapting to contextual information.

4. Experimental setup

4.1 Datasets

For the experiments, we have two different mobile user datasets:

(1) a dataset of $N = 20$ users; and

(2) a dataset of $N = 140$ users.

From now on, we call the first mentioned dataset as *N20 Dataset* and the latter as *N140 Dataset*. The datasets were collected between 2010 and 2012, by using the handset-based measurement method. More detailed information about the method can be found, in Karikoski (2012). In short, the data were collected by using a special purpose software platform provided by a third-party developer. The platform enables collection of a wide variety of smartphone usage data from application usage and network cell IDs to battery levels, for instance. The *N20 Dataset* includes 11,413 application sessions (per user average: 571) from a period of two months and the *N140 Dataset* includes 387071 application sessions (per user average: 2,765) from a period of one and a half years. The users were students and staff of Aalto University, Finland.

Particularly interesting from the semantic place point of view are the cell ID data. The cell ID data are a time-stamped sequence of the IDs of all the cell towers a user's handset has been connected to during the data collection period. In principle, it is known all the time under which cell a user is. In addition to the raw cell ID data, the *N20 Dataset* includes user generated ground truth data where the users have upon request (pop-up questions sent periodically to the device) stated under which semantic place they are (*Home*, *Office/School*, *Other meaningful place* or *Elsewhere*) at a particular time. By combining, time-wise, the cell ID and the ground truth data, we can map the real semantic place classes to the cell IDs.

In the case of the *N140 Dataset*, we do not have the ground truth data and thus it needs to be inferred. For every user, we can calculate how much time the user has spent under a particular cell. Our hypothesis is that the time-spending behavior of a user can be used for classifying the semantic places. For example, if a lot of nighttime is spent at a particular place (under a particular cell), it might be considered *Home*. If the majority of regular working hours are spent at a particular place, it might be considered *Office/School*. In reference to equation (1), Table II shows the dependent variable C which takes as values the semantic places, and the independent variables $\boldsymbol{X}$, which are essentially descriptions of a user's time-spending behavior in a place indicated by a cell ID, and used to classify cell IDs as one of the semantic places. In the case of the *N20 Dataset*, we do not need to infer the contextual information indirectly, as the users have given us this information. However, we can use this data set, by applying machine learning methods, to teach the predictive model f from equation (1) and then utilize the model to classify the cell IDs of the *N140 Dataset*.

The data used for modeling mobile user behavior are described in Tables III and IV at the level of exact attributes and their respective values or value ranges. These data include the demographic data collected via pre-questionnaires from the users, and application session data collected via the handset-based measurements. The applications used are divided into 13 application classes for the *N20 Dataset* and into 10 application classes for the *N140 Dataset*. The classes are somewhat modified from the classes introduced by Smura *et al.* (2009). For the duration attributes which were originally continuous, such as the session durations, idle times (i.e. idle time durations) between sessions, the values are discretized into nominal values. The discretization is

Dependent variable C	*Values/range*
Semantic place (*N20 Dataset*: from user *N140 Dataset*: inferred)	Home, Office/School, other meaningful place, elsewhere
Independent variables X	*Values/range*
Share of time spent in a place	0.0-1.0
Share of time spent in a place during weekdays	0.0-1.0 (only Monday-Friday considered)
Share of time spent in a place during weekends	0.0-1.0 (only Saturday-Sunday considered)
Most time spent in a place during	Night (1 a.m.-9 a.m.), day (9 a.m-5 p.m.), evening (5 p.m.-1 a.m.)

Table II. Attributes used for inferring semantic place data

Demographic data D	Values/range
Gender	Male, Female
Age (years)	23-44
Work status	Full-time, part-time, not working
Device usage	Light, medium, heavy
App session data A	*Values/range*
App class	Browsing, calendar, calling, camera, contacts, games, maps & navigation, messaging, music & audio, photos & gallery, social_network, video
Session duration	Very_short, short, intermediate, long
Previous app	Browsing, calendar, calling, camera, contacts, games, maps & navigation, messaging, music & audio, other, photos & gallery, social_network, video
Session duration of previous app	Very_short, short, intermediate, long
Idle after previous session	Zero, very_short, short, intermediate, long
Idle before next session	Zero, very_short, short, intermediate, long
Weekday	Weekday, weekend
Time of day	Morning_day, day_evening, evening, night

Table III. Attributes used for modeling mobile user behavior: *N20 Dataset*

Demographic data D	Values/range
Device usage	Light, medium, heavy
App session data A	*Values/range*
App class	Browsing, calendar, camera, clock, contacts, games, maps & navigation, messaging, photos & gallery
Session duration	Very_short, short, intermediate, long
Previous app	Browsing, calendar, camera, clock, contacts, games, maps & navigation, messaging, other, photos & gallery
Session duration of previous app	Very_short, short, intermediate, long
Idle after previous session	Zero, very_short, short, intermediate, long
Idle before next session	Zero, very_short, short, intermediate, long
Weekday	Weekday, weekend
Time of day	Morning_day, day_evening, evening, night

Table IV. Attributes used for modeling mobile user behavior: *N140 Dataset*

done by using *equal-frequency* binning (Liu *et al.*, 2002). The *Time of day* attribute divides the day into four six-hour periods, starting from morning-day (7 a.m.-1 p.m.) and ending to night (1 a.m.-7 a.m.).

4.2 Number of user segments

The behavioral models can be built for different units of analysis, i.e. for different user groups. On the other end, we have the most aggregate level where all users together act as a unit of analysis. On the opposite end, we have a single user acting as a unit of analysis. Assuming good clustering methods and proper selection of variables for the basis of clustering, the groups of users become more homogenous when moving from aggregate level toward the single user level. More homogenous data can potentially lead to more accurate predictions. However, moving from the aggregate level to the single user level, the models can run into data sparsity problems, i.e. there are too few data for making accurate predictions.

The data sets are segmented four times, increasing the number of segments every time, to examine the effect of moving from the aggregate level to the single user level. The number of segments varies as follows:

(1) *1_Segment*: One model is built for the whole dataset.

(2) *2_Segments* (*N20 Dataset*) or *4_Segments* (*N140 Dataset*): The dataset is divided into two (*N20 Dataset*) or four (*N140 Dataset*) user segments and one model is built for each segment.

(3) *4_Segments* (*N20 Dataset*) or *14_Segments* (*N140 Dataset*): The dataset is divided into 4 (*N20 Dataset*) or 14 (*N140 Dataset*) user segments and one model is built for each segment.

(4) *20_Segments* (*N20 Dataset*) or *140_Segments* (*N140 Dataset*): One model is built for each individual user.

4.3 Predictive algorithms

Three different Weka (Hall *et al.*, 2009) classifiers were used for building predictive models:

(1) Naïve Bayes (John and Langley, 1995);

(2) *Bayes Network* classifier with K2 search algorithm (Bouckaert *et al.*, 2013); and

(3) *J48* (a C4.5 decision tree algorithm) (Quinlan, 1993).

These classifiers were selected partly because of their popularity and partly because of their relatively fast computation times. The Naïve Bayes classifier operates on relatively restrictive assumptions, but is nonetheless competitive with many state-of-the-art classifiers (Friedman *et al.*, 1997). More complicated Bayes Network classifiers have, less restrictive independency assumptions. C4.5 is often considered a standard benchmark in machine learning. Some initial testing was also done with a few other types of classifiers, including Support Vector Machines and rule-based classifiers. The results were similar, but the computation times much longer in comparison with the chosen classifiers.

4.4 Dependent variables

For the experiments, we have chosen five dependent variables *Y*. The variables, which are the same for both the *N20 Dataset* and the *N140 Dataset*, are:

(1) *Application class* of the application used by the user.

(2) *Duration* of an application session.

(3) *Idle time before the next application session*, i.e. the time the user's device remains unused between the end of usage of current application and start of usage of the next application.

(4) *Day of week* when a particular application session has occurred.

(5) *Time of day* a particular application session has occurred.

The chosen dependent variables are highlighted in Tables III and IV.

4.5 Performance metrics

Two different performance metrics were used in our experiments:

(1) *Predictive accuracy*, which is computed as the ratio between the number of correctly classified cases and the total number of classified cases (Fawcett, 2006).

(2) *Area Under ROC* (Receiver Operating Characteristic) *Curve* (abbreviated often as AUC).

AUC value is equivalent to the probability that a classifier will rank a randomly chosen positive case higher than a randomly chosen negative case (Fawcett, 2006). The performance metrics are calculated by using cross-validation (Kohavi, 1995) for each model.

5. Results and discussion

5.1 Semantic place inference

In this section, we describe the semantic place inference in practice and take a look at some interesting observations along the process. As mentioned earlier, the *N20 Dataset* has the ground truth semantic place data available. In the case of the *N140 Dataset*, however, inference is needed. We utilize the *N20 Dataset* to train model (1) for the *N140 Dataset*. Along the process, we examine the dynamics of the inference model performance across predictive algorithms, the longitudinal length of the input data, individual semantic places and number of user segments through the lens of the *N20 Dataset*.

The ground-truth data are relatively comprehensive for *Home* and *Office/School*, as a user does not have many of these semantic places. In the case of the *Other meaningful place*, the pop-up questionnaire method might have missed some places of this kind of the users. Thus, the *Other meaningful place* ground truth might be somewhat less comprehensive. In the case of *Elsewhere*, the users have, in general, "tagged" only a fraction of places that would fall under this class. In all, the users have tagged 11.5 per cent of the places they have visited. In total, 47 per cent of these places are tagged as *Elsewhere*. From the 88.5 per cent of untagged places, majority is presumably *Elsewhere*-like and rarely (or only once) visited or passed by places. Only a small fraction of time has been spent per an untagged place.

Figure 1 shows predictive accuracy and AUC results for Naïve Bayes, Bayes Network and J48 classifier implementations of model (1). The figure also shows the average results over all the classifiers. The idea behind the semantic place inference method relies on regularities in the users' time spending behavior. We hypothesize that, over longer time periods, some of the irregularities in the users' time-spending behavior smooth out and thus accuracy of the semantic place classification improves. We divide the cell ID data into twelve half-a-week-long sections that cumulatively add up to six weeks of data. On the x-axis, e.g. three weeks of data mean that the classifier operates on a dataset that corresponds to that of users spending three weeks under some cell IDs. Here, six weeks of data correspond to the whole dataset we use for training the classifiers. While training the classifiers, the places are weighed relative to the amount of time spent, i.e. the places where the most time is spent are also the most important to classify correctly to gain a good performance.

Both the prediction accuracy and AUC measures improve as the classifier models get a time-wise longer period of data to operate with. The average prediction accuracy improves from 73 per cent (0.5 weeks of data) to 85 per cent (6 weeks of data), and the average AUC improves from 0.70 (0.5 weeks of data) to 0.89 (6 weeks of data). The best classifier examined here is the Naïve Bayes classifier which, with the full dataset, achieves a prediction accuracy of 87 per cent and an AUC of 0.95.

The prediction accuracy measures shown above are the accuracies of the whole classifications, i.e. all of the semantic places together. The prediction accuracy of a classifier is the same as the weighted average of the *true positive rate* (Fawcett, 2006) of the classes. The AUC measures shown above are the weighted averages of AUCs of every class. Figure 2 shows the true positive rates and AUC measures of each class (i.e. semantic place) per classifier. For these results, we utilized the whole six weeks of data.

Based on the true positive rate (TPR), *Elsewhere* (93 per cent average TPR) and *Home* (91 per cent average TPR) seem to be the easiest semantic places to classify correctly. The average true positive rates of *Office/School* and *Other meaningful place* are 70 and 32 per cent, respectively. Based on the AUC measure, *Elsewhere* (0.96 average AUC) is

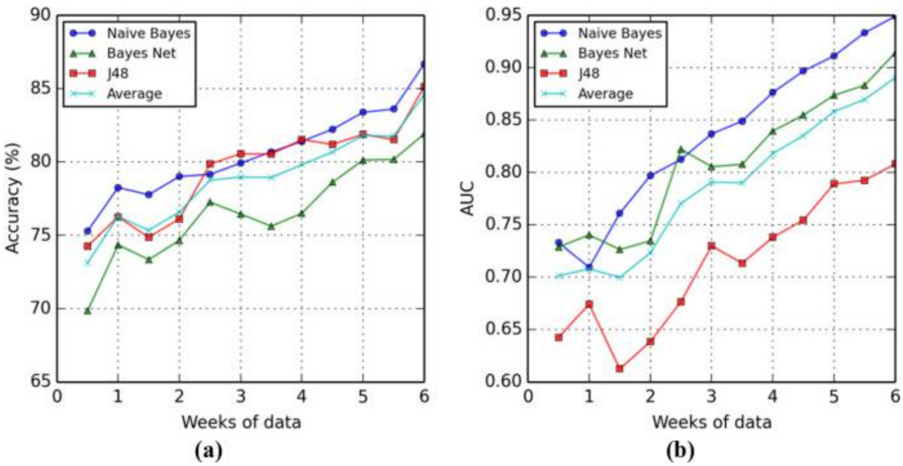


Figure 1.
Accuracy (a) and AUC (b) measures of Naïve Bayes, Bayes Net and J48 classifiers as a function of the amount of data

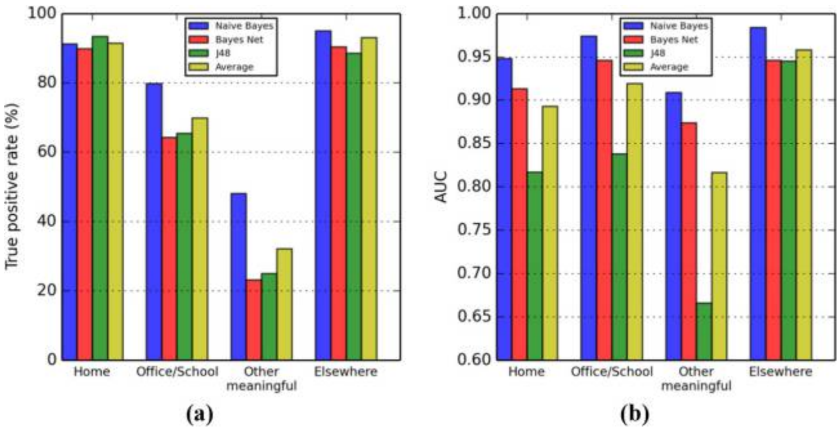


Figure 2.
Accuracy (a) and AUC (b) measures of Naïve Bayes, Bayes Net and J48 classifiers per semantic place

the easiest to classify, followed closely by *Office/School* (0.92 average AUC) and *Home* (0.89 average AUC). *Other meaningful place* has an average AUC of 0.82. Intuitively, the results do not surprise. *Home* should be quite easy to detect based on the users' time-spending behavior (a lot of nighttime), as well as *Office/School* if a user has somewhat regular working or lecture hours during the weekdays. *Elsewhere* differentiates relatively easily from the other semantic places in that the time spent in individual *Elsewhere*-type of places is very limited in comparison with the more "meaningful" places. The *Other meaningful place* proves to be more difficult to classify. First, because of bigger diversity in *Other meaningful*-type of places even in the case of one user and definitely across different users. Second, the quality of the ground truth might be poorer than in the cases of, e.g. *Home* and *Office/School*.

By examining performance differences between the different classifiers in Figure 2, we notice that the Naïve Bayes classifier outperforms the other classifiers in most of the cases. Based on TPR, Bayes Net classifier and J48 perform quite equally. Based on AUC, J48 falls clearly behind the Bayes Net.

The above-described results are based on the whole *N20 Dataset*. In other words, the classifiers try to catch the aggregate or general time usage behavior of the users and classify the semantic places of each user based on this general model. Even though the time usage behavior of the users is relatively homogenous compared, for example, with device usage, the one-model-fits-all approach has its limitations. Figure 3 shows prediction accuracy and AUC results for a Naïve Bayes classifier across the varying number of user segments. The segments were generated by grouping together demographically and time usage-wise (based on, e.g. the share of time spent at *Home*) similar users with the *k*-Means clustering method (Kanungo *et al.*, 2002). For *1_Segment*, the result is based on 1 overall model; for *2_Segments*, it is the average result of 2 models; for *4_Segments*, the average result of 4 models; and for *20_Segments*, the average results of 20 models. The results indicate that the more personalized (i.e. moving from the aggregate level toward the individual level) the models are, the better the performance. From *1_Segment* to *20_Segments*, the accuracy and AUC results rise from 87 to 98 per cent and from 0.95 to 0.99, respectively.

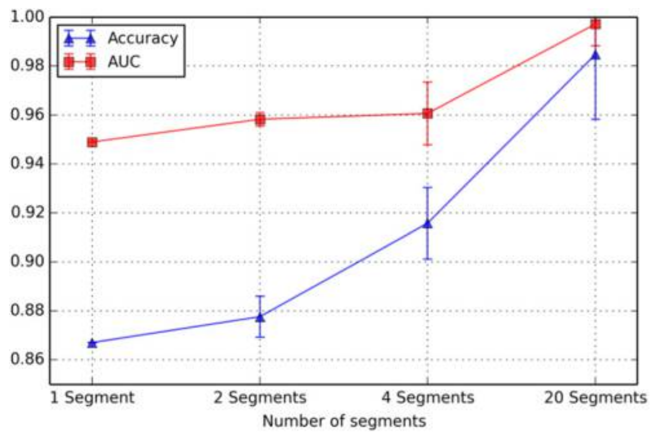


Figure 3.
Accuracy and AUC measures of Naïve Bayes classifier per number of user segments. Error bars show standard deviations of performance between segments

It is logical that the more personalized models yield better results. By following one user a period of time, one eventually gains a good insight on the patterns of place-related time usage of the user. Only a fraction of people live their everyday lives absent any recognizable diurnal routines, but obviously, the routines of one person might differ considerably of the routines of another person. However, clear diurnal routines can easily be observed also on the aggregate level. It is still common that, on average, people sleep at night and go to work during daytime. On the societal level, e.g. commuting and Internet traffic follow recognizable diurnal patterns. Thus, also the aggregate models can reach acceptable performance under suitable conditions.

The goal in this section was to train a classifier model for inferring semantic place information from the *N140 Dataset* for the next section. Based on the above examination, we will use the Naïve Bayes classifier. However, despite the good average performance of the personalized models we have to settle for the aggregate model built from the whole *N20 Dataset*. A model built from the data of a single user is over-fitted for that particular user. An aggregate model is likely to classify better the places of a randomly chosen new user than a model tailored to some other user. It is admittedly possible to use some similarity metric, try to find (from *N20 Dataset*) users similar to the new user (from *N140 Dataset*) and build a classifier model from the data of the similar users to classify the places of the new user. Collaborative filtering (Breese *et al.*, 1998) and user-based Nearest Neighbor (Cover and Hart, 1967) type of methods could be used for this. We, however, leave this type of an approach for future research.

5.2 Comparison of context-aware predictive modeling approaches

The aim of this section is to experimentally examine how models built with semantic place-based *PreF* equation (3a), *CM* equation (3b) and *PoF* equation (3c) approaches compare performance-wise with each other, when inferring mobile user behavior. We also study how these semantic place-powered models compare with models without any semantic place information equation (2). Referring to Chapter 4, our experimental settings include two datasets, four different numbers of user segments, three classifiers, five dependent variables and two performance measures. For the *N20 Dataset*, the semantic place information is given by the users and for the *N140 Dataset*, the semantic place information is inferred as described in the previous section. Given the different experimental settings, we end up with thousands of generated models. Showing the results of each model here in detail is not reasonable and thus we need a more structured way of representing the results. First, we calculate a relative performance difference measure *Diff* for every semantic place-powered model against its semantic place-ignorant counterpart:

$$Diff = \frac{Perf_{spp} - Perf_{spi}}{Perf_{spi}} \quad (4)$$

where *spp* and *spi* refer to semantic place-powered and semantic place-ignorant, respectively and *Perf* to a performance measure (either prediction accuracy or AUC). If *Diff* is positive, the semantic place-powered model outperforms the semantic place-ignorant, and if negative – the other way around. The range of prediction accuracy is [0, 1] whereas the range of AUC is [0.5, 1]. The relative performance difference between these two is, however, commensurable which enables calculating the averages over the relative performance differences of the two performance measures. From now on, when we talk about *Diff*, we mean the average, unless otherwise noted. Second, we divide the representation into *marginal* analysis and *regional* analysis. Marginal analysis summarizes performance differences while varying the number of user segments, dependent variable or semantic place separately. Regional analysis presents the performance differences while the number of segments, dependent variable and semantic place are varied together. This reveals (graphically) more granular performance difference regions (cf. usage circumstances) on the three-dimensional (number of segments, dependent variable, semantic place) space.

5.2.1 Marginal analysis. Figure 4 shows marginal analysis results for models built with the *PreF*, *CM* and *PoF* approaches (*N20 Dataset* on the left and *N140 Dataset* on the right). The dashed line on $Diff = 0$ level represents the semantic place-ignorant approach. Figures 4(a and b) average *Diff*s over classifiers and dependent variables to examine how the approaches compare throughout the number of user segments. Overall, the semantic place-powered models perform slightly better than the semantic place-ignorant models. Also, a slight trend of increasing performance difference is observable as we move from aggregated models toward user-specific models. However, in the majority of cases, the relatively large standard deviations undermine any strong conclusions on the effect of the user specificity of the models. In the case of the *N20 Dataset*, the performances of models built with the different semantic place-powered approaches fit inside each other's error bars. In the case of the *N140 Dataset*, the *PreF* approach falls behind the two others. Figures 4(c and d) average *Diff*s over classifiers and number of segments to examine how the approaches compare throughout the dependent variables. The semantic place-powered models perform better than the semantic place-ignorant models when *Day of Week* or *Time of Day* is the dependent variable. Otherwise, they perform equally or slightly worse. *PreF* falls behind the other two approaches otherwise, except when *Time of Day* (and *Day of Week* in the *N20 Dataset* case) is the dependent variable.

Figures 4(e and f) average *Diff*s over classifiers, dependent variables and number of segments. However, the results are separated between the semantic places. In the case of *PreF*, this division comes naturally because the underlying data fed to the models are filtered by the semantic place. In the case of the other two approaches, each classified instance (i.e. an application session) is checked for, in which semantic place occurred, and the performance metrics are then calculated per semantic place. Models built with *CM* and *PoF* approaches outperform the corresponding semantic place-ignorant models relatively consistently across the semantic places. The largest difference between the two is observed in the case of the *N20 Dataset* and the *Other meaningful place*. Otherwise, the outputs of the two approaches behave similarly. The *PreF* approach results in worse performance, especially for the *Other meaningful place*, and also for *Elsewhere* for the *N140 Dataset*. Also, the standard deviations are larger than in the other two approaches. Overall, the standard deviations in the cases of *Home* and *Elsewhere* are smaller than in the cases of *Office/School* and *Other meaningful place*, indicating more consistent results across the experimental settings for the former two.

Based on the results shown in Figure 4, models built with the *PoF* approach are the only ones outperforming their semantic place-ignorant counterparts consistently. The *PreF* approach leads into more varying results, occasionally underperforming when compared with the respective semantic place-ignorant approach. The *CM* approach lies in between, but is overall close to the *PoF* approach. The three approaches have

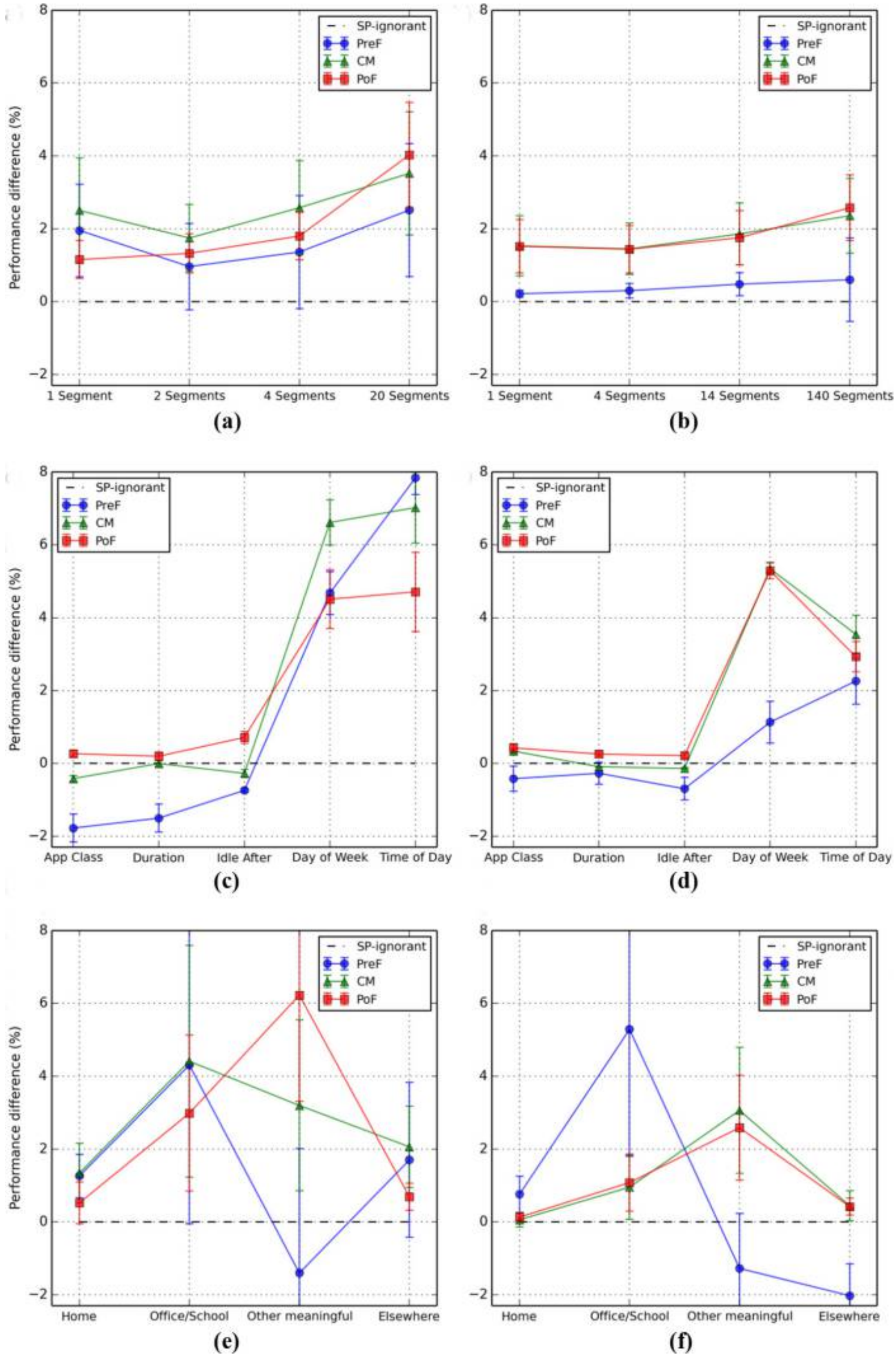


Figure 4.
Performance differences between models built with *PreF*, *CM* and *PoF* approaches. The differences are averaged over classifiers, performance measures, dependent variables (a, b, e, f), number of user segments (c, d, e, f). Figures for *N20 Dataset* are on the left (a, c, e) and *N140 Dataset* on the right (b, d, f). Error bars show standard deviations

somewhat different properties which might contribute to the observed differences. As the *PreF* approach filters data based on the desired semantic place before classification, it is subject to the so-called data homogeneity versus data sparsity bias (Adomavicius *et al.*, 2011). Semantic place specific data are more homogenous, easing the classification, but, on the other hand, few data are available, making the classification harder. For example, the amount data for just the *Other meaningful place* are relatively low, and the mobile usage habits are not so distinguishable, that is more homogenous, compared with the usage habits in general. On the other hand, the usage habits in *Office/School* are relatively homogenous. We can measure inverse data homogeneity (i.e. heterogeneity), e.g. with an *unalikeability* metric (Kader and Perry, 2007), which is a measure of variability for nominal variables. *Diff*s of the *PreF* approach have −0.48 correlation with unalikeability of the respective dependent variables. That is, the less variation in the observed values of the dependent variable in the (filtered) input data, the better performance the classifiers show. As an example, if *N20 Dataset* is filtered by *Office/School*, unalikeability averaged over all dependent variables decreases by 31 per cent when compared with corresponding unalikeability calculated over the whole dataset.

Models based on *CM* or *PoF* do not suffer from data sparsity nor gain from data homogeneity at least in relative terms, since the input data for classification (and the respective unalikeabilities) are the same as for models based on the semantic place-ignorant approach. Consequently, the correlations of *Diff*s with unalikeabilities of corresponding dependent variables are −0.045 and 0.0013 for *CM* and *PoF*, respectively. In the case of *CM*, every instance is accompanied also with the semantic place information. In principle, machine learning models benefit almost always when more independent variables are added into the model. However, based on our results, the benefits are not that high compared with the other two approaches. On the outset, the *PoF* approach is semantic place-ignorant; however, the initial classifications are then modified afterward based on the available semantic place information. The usefulness of the approach lies in the quality of the modification. By utilizing insights from the data itself and/or from domain or user provided knowledge, carefully chosen weights for modifying the original classification can produce good results. On the other hand, however, badly chosen weights can produce useless results.

5.2.2 Regional analysis. The relatively compact form of the marginal analysis benefits the high-level comparison of the *PreF*, *CM* and *PoF* approaches at the expense of revealing some of the more subtle aspects. For example, the good performances of the semantic place-powered models in the cases of *Day of Week* and *Time of Day* might be specific for certain semantic places or certain user segmentations. Figures 5-7 show the regional analysis results for *PreF*, *CM* and *PoF* approaches, respectively. The stars (*) on heatmap cells indicate statistical significance ($p < 0.1$) for that a particular *Diff* is positive, that is, the semantic place-powered model outweighs the corresponding semantic place-ignorant model. The corresponding semantic place-powered and -ignorant models are built for the same units of analysis and thus the two samples are related. The distributions of *Diff*s are not normal, so we use a Wilcoxon signed-rank test (Siegel, 1956) for testing the significance.

The overall picture between the marginal and regional analyses is similar. While inferring *Day of Week* or *Time of Day*, the semantic place-powered models outperform their semantic place-ignorant counterparts. However, this behavior is concentrated, especially on the *Office/School* semantic place and on the *Other meaningful place* to some

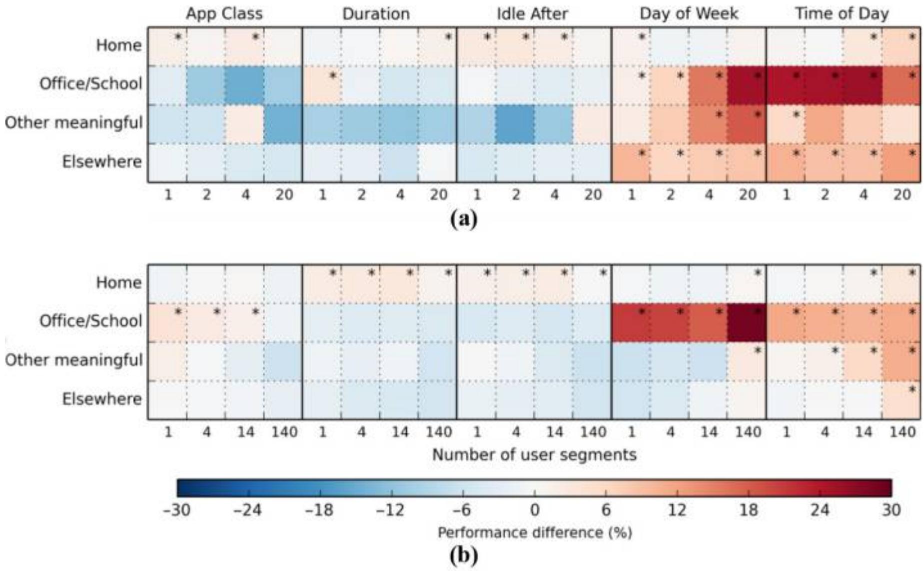


Figure 5.
Performance differences between models built with *PreF* approach and their semantic place-ignorant counterparts

Notes: (a) *N20 Dataset*; (b) *N140 Dataset*; *indicates statistically significant positive difference

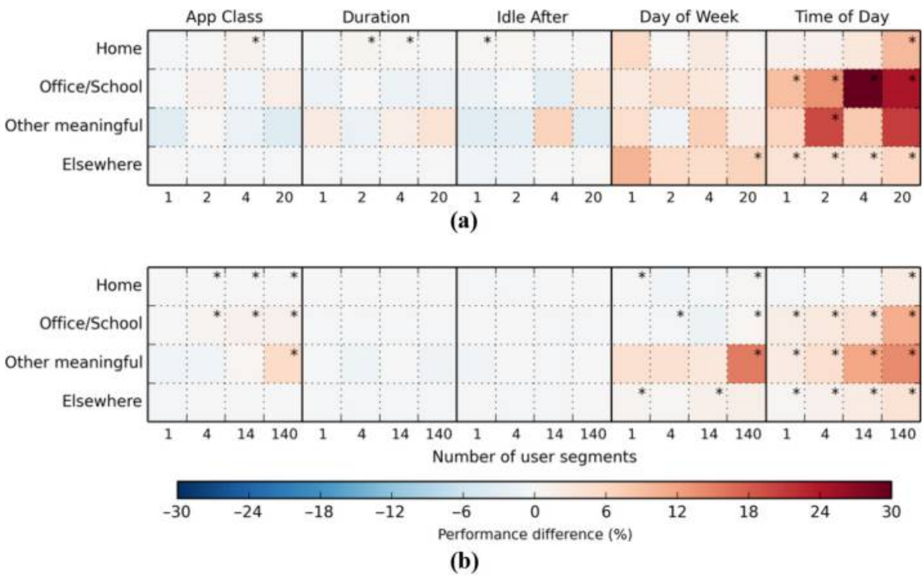
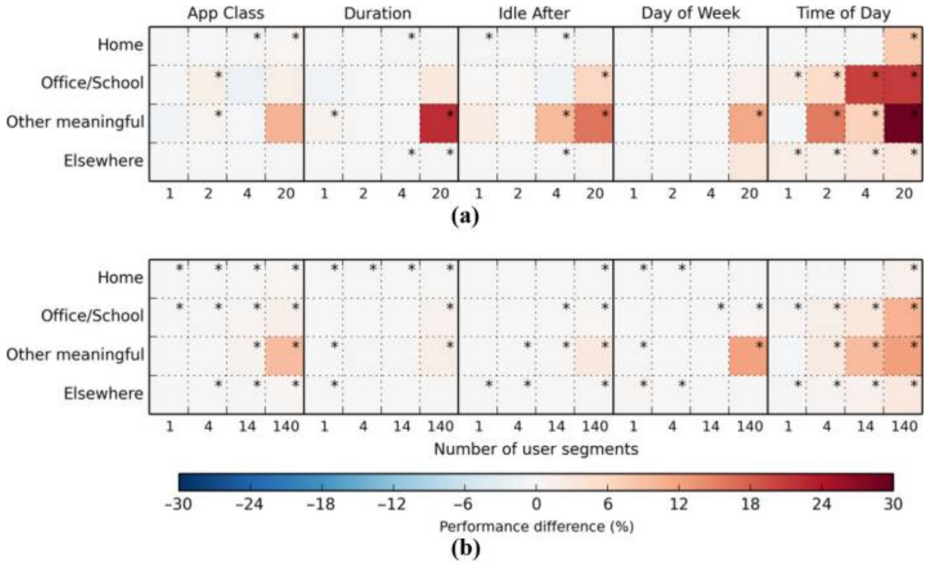


Figure 6.
Performance differences between models built with *CM* approach and their semantic place-ignorant counterparts

Notes: (a) *N20 Dataset*; (b) *N140 Dataset*; *indicates statistically significant positive difference



Notes: (a) *N20 Dataset*; (b) *N140 Dataset*; *indicates statistically significant positive difference

Figure 7. Performance differences between models built with *PreF* approach and their semantic place-ignorant counterparts

degree. The results reflect the fact that the users have some recognizable diurnal routines their time-spending behavior follows. When it comes to the other dependent variables, the semantic place-powered models perform slightly better at *Home* overall, and at *Office/School* in the case of *App Class*. Additionally, the *PoF* approach models perform relatively well at *Other meaningful place*, especially with increasing user specificity (higher number of segments). Some of the earlier results, Karikoski and Soikkeli (2013) and Soikkeli *et al.* (2013), show that application sessions at home are longer, but occur less frequently than in other semantic places. This might reflect in our results in that it is easiest to infer *Duration* and *Idle After* when users are at *Home*. Also, the positive *App Class* results of *Home* and *Office/School* might reflect the earlier results (Soikkeli *et al.*, 2013) which imply that some applications are more semantic place sensitive than others.

With regional analysis, *PreF* reveals even more of its on/off behavior. For certain dependent variables, it works nicely and produces significant positive *Diff*s, whereas for others it performs clearly worse than the semantic place-ignorant approach. This might be a reflection of the mentioned data homogeneity versus data sparsity tradeoff. If semantic place-specific data homogeneity is high, the models perform well, but if the homogeneity is comparable with the whole dataset, the lower amount of data brings performance below $Diff = 0$ level. Compared with *PreF*, *CM* provides more conservative results over the performance differences. It has less and, in absolute terms, smaller $Diff < 0$ results, but also smaller $Diff > 0$ results. With *PoF*, $Diff < 0$ results are practically nonexistent. However, the larger $Diff > 0$ results are only a few, although the Wilcoxon signed-rank test also deems many of the smaller differences significant. *PoF* increases its performance when moving toward user-specific models more than the other two approaches. One reason for this is the post-filtering weights which can be used to drop unnecessary (for a certain semantic place for a certain user) dependent variable values while retaining all the valuable predictive data. For example, a user might have used only a few app classes while in *Other meaningful place*. By weighing the other app classes to zero benefits the inference.

6. Conclusions

In this article, we aimed at answering two main research questions:

RQ1. How models built with semantic place-based *PreF*, *CM* and *PoF* approaches compare performance-wise with each other, when inferring mobile user behavior?

RQ2. How models built with semantic place-based *PreF*, *CM* and *PoF* approaches compare performance-wise with models without semantic place information, when inferring mobile user behavior?

To answer the questions, we examined two handset-based datasets empirically across different experimental settings. For one of the datasets, namely, *N20 Dataset*, we had user provided semantic place information. For the other dataset, namely, *N140 Dataset*, we had to infer the semantic place information. The inference process showed that a reasonable level of accuracy in inferring the semantic places (*Home*, *Office/School*, *Other meaningful place* and *Elsewhere*) is achievable with relatively simple methods. Additionally, the inference accuracy increases as the time period for observing the user's time-spending behavior in different places is increased, and as the inference models become more user specific (personalized for increasingly homogenous user groups).

For *RQ1*, we conclude, based on the empirical results, that the three different context-aware predictive modeling approaches – *PreF*, *CM* and *PoF* – show circumstance-specific differences when inferring the examined aspects of mobile user behavior. Although, among the approaches considered, there are no clear winners that uniformly outperform the alternatives, some approaches provide the best solutions in certain circumstances. While inferring dependent variables such as *Day of Week* or *Time of Day*, all of the approaches lead into increased performance, compared with semantic place-ignorant approaches. With the other dependent variables (*App Class*, *Duration* and *Idle After*), the performance increases are mostly negligible; however, the *PreF* approach shows even decreased performance when compared with the semantic place-ignorant approaches. Differences in performance are visible across the different semantic places. *PreF* is relatively strong in inferring user behavior in *Office/School*, but poor in *Other meaningful place*. The other two approaches are stronger, especially in the case of the *Other meaningful place*. *PreF* works well if the information (semantic place in this work) used for the filtering separates well certain types of user behavior (e.g. daily or weekly mobile usage patterns in this work), but works poorly if the separation falls short while the amount of data is significantly reduced by the filtering. *CM* is the most conventional from the approaches examined in this work. It produces reasonable performance improvements with the lowest amount of additional work. *PoF* performs well if the actual post-filtering part for modifying the initial inferences is sound and works well together with the additional information. Developing the best possible modification method is an art of its own and a clear topic for future research. Future research is also needed on alternative approaches to the three examined here, as well as, on possible hybrid approaches. In short, *PreF* offers a "high risk, high reward" approach, and thus, cannot be applied blindly without knowing the properties of the underlying data. *CM* offers a relatively reliable "safe bet" which, in general, does not fall behind semantic place-ignorant approaches, but rather outperforms them at least slightly. *PoF* has potential for the "best of breed" approach on condition that suitable a priori knowledge is available or can be inferred for constructing the post-filtering weights.

For *RQ2*, we conclude that, when generalized over all the experimental settings, models built with the semantic place-powered approaches outperform models without any semantic place information. However, when examining the results with finer granularity, it becomes visible that semantic place information is able to contribute positively only to the inference of certain aspects of mobile usage, such as when the usage takes place, and, in some more limited cases, the app used and its usage duration and frequency. The experimental settings used in this work are limited, but, nonetheless, highlight the important matter of case specificity. In many context-aware mobile applications, location and semantic place are often, by default, the first contextual information to be added. However, the particular use case defines the usefulness of the additional information. In many cases, some of the more personal traits of the users dominate the mobile user behavior when compared with just being at different semantic places. Of semantic places, home and workplace are considered the most important as they in general set the rhythm of people's daily life and routines and people spend most of their time in these places. Indeed, also in this work, knowing whether the users are at home or at the office/school proved useful in inferring certain aspects of mobile user behavior.

The present work constitutes an initial step toward understanding and systematically evaluating different approaches for incorporating additional semantic place data into modeling mobile user behavior. A cost of acquiring and incorporating any additional contextual information, such as the semantic place, always exists. Thus, it needs to be examined whether applying the information is useful, and by which means it is the most useful. The experimental part of the article is limited by user-wise narrow datasets and the timeliness of the data can be questioned in the fast moving mobile domain. However, we think that the underlying behavioral and data-related aspects the article focuses on are not that time-sensitive. Nonetheless, further research on the different context-aware predictive modeling approaches requires new datasets and new aspects of mobile user behavior to work on. The strengths and weaknesses of the approaches lie in different places, and mapping them comprehensively requires different experimental settings. Also, a more theoretical examination is needed to investigate the reasons behind the observed differences between the approaches.

References

Adomavicius, G., Mobasher, B., Ricci, F. and Tuzhilin, A. (2011), "Context-aware recommender systems", *AI Magazine*, Vol. 32 No. 3, pp. 67-80.

Aharony, N., Pan, W., Ip, C., Khayal, I. and Pentland, A. (2011), "Social fMRI: investigating and shaping social mechanisms in the real world", *Pervasive and Mobile Computing*, Vol. 7 No. 6, pp. 643-659.

Ashbrook, D. and Starner, T. (2003), "Using GPS to learn significant locations and predict movement across multiple users", *Personal and Ubiquitous Computing*, Vol. 7 No. 1, pp. 275-286.

Baladron, C., Aguiar, J.M., Carro, B. and Calavia, L. (2012), "Framework for intelligent service adaptation to user's context in next generation networks", *IEEE Communications Magazine*, Vol. 50 No. 3, pp. 18-25.

Baltrunas, L., Ludwig, B. and Peer, S. (2012), "Context relevance assessment and exploitation in mobile recommender systems", *Personal and Ubiquitous Computing*, Vol. 16 No. 5, pp. 507-526.

Bayir, M.A., Demirbas, M. and Eagle, N. (2010), "Mobility profiler: a framework for discovering mobility profiles of cell phone users", *Pervasive and Mobile Computing*, Vol. 6 No. 4, pp. 435-454.

Bouckaert, R.R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A. and Scuse, D. (2013), "Weka manual for version 3-6-10", available at: www.cs.waikato.ac.nz/ml/weka/documentation.html (accessed 26 July 2014).

Breese, J.S., Heckerman, D. and Kadie, C. (1998), "Empirical analysis of predictive algorithms for collaborative filtering", *14th Conference on Uncertainty in Artificial Intelligence*, *Morgan Kaufmann Publishers, San Francisco, CA*, pp. 43-52.

Brown, P.J., Bovey, J.D. and Chen, X. (1997), "Context-aware applications: from the laboratory to the marketplace", *IEEE Personal Communications*, Vol. 4 No. 5, pp. 58-64.

Burns, W., Chen, L., Nugent, C., Donnelly, M., Skillen, K.L. and Solheim, I. (2012), "A conceptual framework for supporting adaptive personalized help-on-demand services", *LNCS*, Vol. 7683 No. 1, pp. 427-432.

Chen, L., Skillen, K., Burns, W., Quinn, S., Rafferty, J., Nugent, C.D. and Solheim, I. (2014), "Learning behaviour for service personalisation and adaptation", in Wang, X. *et al.* (Eds), *Machine Learning and Cybernetics, CCIS*, Vol. 481 No. 1, pp. 287-297.

Cover, T.M. and Hart, P.E. (1967), "Nearest neighbor pattern classification", *IEEE Transactions on Information Theory*, Vol. 13 No. 1, pp. 21-27.

Dey, A.K. (2001), "Understanding and using context", *Personal and Ubiquitous Computing*, Vol. 5 No. 1, pp. 4-7.

Eagle, N. and Pentland, A. (2006), "Reality mining: sensing complex social systems", *Personal and Ubiquitous Computing*, Vol. 10 No. 4, pp. 255-268.

Etter, V., Kafsi, M. and Kazemi, E. (2012), "Been there, done that: what your mobility traces reveal about your behavior", *Mobile Data Challenge by Nokia Workshop, in conjunction with International Conference on Pervasive Computing*, *Newcastle*.

Fawcett, T. (2006), "An introduction to ROC analysis", *Pattern Recognition Letters*, Vol. 27 No. 8, pp. 861-874.

Friedman, N., Greiger, D. and Goldszmidt, M. (1997), "Bayesian network classifiers", *Machine Learning*, Vol. 29 Nos 2/3, pp. 131-163.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H. (2009), "The WEKA data mining software: an update", *AMC SIGKDD Explorations*, Vol. 11 No. 1, pp. 10-18.

Huang, C.M., Ying, J.J.C. and Tseng, V.S. (2012), "Mining users behaviors and environments for semantic place prediction", *Mobile Data Challenge by Nokia Workshop, in conjunction with International Conference on Pervasive Computing*, Newcastle.

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J. and Varshavsky, A. (2011), "Identifying important places in people's lives from cellular network data", *LNCS*, Vol. 6696 No. 1, pp. 133-151.

John, G.H. and Langley, P. (1995), "Estimating continuous distributions in Bayesian classifiers", *11th Conference on Uncertainty in Artificial Intelligence*, *Morgan Kaufmann Publishers, San Francisco, CA*, pp. 338-345.

Kader, G.D. and Perry, M. (2007), "Variability for categorical variables", *Journal of Statistics Education*, Vol. 15 No. 2.

Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R. and Wu, A.Y. (2002), "An efficient k-means clustering algorithm: analysis and implementation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24 No. 7, pp. 881-892.

Karikoski, J. (2012), "Handset-based data collection process and participant attitudes", *International Journal of Handheld Computing Research*, Vol. 3 No. 4, pp. 1-21.

Karikoski, J. and Soikkeli, T. (2013), "Contextual usage patterns in smartphone communication services", *Personal and Ubiquitous Computing*, Vol. 17 No. 3, pp. 491-502.

Kohavi, R. (1995), "A study of cross-validation and bootstrap for accuracy estimation and model selection", *International Joint Conference on Artificial Intelligence (IJCAI'95)*, *Montreal, Quebec*.

Laurila, J.K., Gatica-Perez, D., Aad, I., Blom, J., Bornet, O., Minh Tri Do, T., Dousse, O., Eberle, J. and Miettinen, M. (2013), "From big smartphone data to worldwide research: the mobile data challenge", *Pervasive and Mobile Computing*, Vol. 9 No. 6, pp. 752-771.

Liang, T. and Yeh, Y. (2011), "Effect of use contexts on the continuous use of mobile services: the case of mobile games", *Personal and Ubiquitous Computing*, Vol. 15 No. 2, pp. 187-196.

Liu, H., Hussain, F., Tan, C.L. and Dash, M. (2002), "Discretization: an enabling technique", *Data Mining and Knowledge Discovery*, Vol. 6 No. 1, pp. 393-423.

Lungaro, P., Viedma, C., Segall, Z. and Kumar, P. (2011), "An experimental framework to investigate context-aware schemes for content delivery", *Vehicular Technology Conference (VTC Fall)*, *IEEE, San Francisco, CA*, pp. 1-5.

McInerney, J., Stein, S., Rogers, A. and Jennings, N.R. (2013), "Breaking the habit: measuring and predicting departures from routine in individual human mobility", *Pervasive and Mobile Computing*, Vol. 9 No. 6, pp. 808-822.

Mayrhofer, R., Radi, H. and Ferscha, A. (2003), "Recognizing and predicting context by learning from user behavior", *International Conference on Advances in Mobile Multimedia (MoMM2003)*, *Wollzeile*, pp. 25-35.

Mazhelis, O., Zliobaite, I. and Pechenizkiy, M. (2011), "Context-aware personal route recognition", *LNCS*, Vol. 6926 No. 1, pp. 221-235.

Panniello, U., Tuzhilin, A. and Gorgoglione, M. (2014), "Comparing context-aware recommender systems in terms of accuracy and diversity", *User Modeling and User-Adapted Interaction*, Vol. 24 Nos 1/2, pp. 35-65.

Quinlan, R. (1993), *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, CA.

Raento, M., Oulasvirta, A., Petit, R. and Toivonen, H. (2005), "Context-phone: a prototyping platform for context-aware mobile applications", *IEEE Pervasive Computing*, Vol. 4 No. 2, pp. 51-59.

Ravi, N., Scott, J., Han, L. and Iftode, L. (2008), "Context-aware battery management for mobile phones", *Sixth Annual IEEE International Conference on Pervasive Computing and Communications*, *IEEE, Hong Kong*, pp. 224-233.

Schilit, B. and Theimer, M. (1994), "Disseminating active map information to mobile hosts", *IEEE Network*, Vol. 8 No. 5, pp. 22-32.

Schmidt, A., Beigl, M. and Gellersen, H. (1999), "There is more to context than location", *Computers & Graphics*, Vol. 23 No. 6, pp. 893-901.

Siegel, S. (1956), *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill, New York, NY.

Skillen, K., Chen, L., Nugent, C.D., Donnelly, M., Burns, W. and Solheim, I. (2014), "Ontological user modelling and semantic rule-based reasoning for personalisation of help-on-demand services in pervasive environments", *Future Generation Computer Systems*, Vol. 34 No. 1, pp. 97-109.

Smura, T., Kivi, A. and Töyli, J. (2009), "A framework for analysing the usage of mobile services", *info*, Vol. 11 No. 4, pp. 53-67.

Soikkeli, T. (2011), "The effect of context on smartphone usage sessions", M.Sc. thesis, Aalto University School of Science, Espoo.

Soikkeli, T., Karikoski, J. and Hämmäinen, H. (2013), "Characterizing smartphone usage: diversity and end user context", *International Journal of Handheld Computing Research*, Vol. 4 No. 1, pp. 15-36.

Stopezynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M.M., Larsen, J.E. and Lehmann, S. (2014), "Measuring large-scale social networks with high resolution", *Plos One*, Vol. 9 No. 4, arXiv:1401.7233 [cs.SI].

Verkasalo, H. (2008), "Contextual patterns in mobile service usage", *Personal and Ubiquitous Computing*, Vol. 13 No. 5, pp. 331-342.

Xu, Z. and Yuan, Y. (2009), "The impact of context and incentives on mobile service adoption", *International Journal of Mobile Communications*, Vol. 7 No. 3, pp. 363-381.

Yan, T., Chu, D., Ganesan, D., Kansal, A. and Liu, J. (2012), "Fast app launching for mobile devices using predictive user context", *10th international conference on Mobile Systems, Applications, and Services*, ACM, *New York, NY*, pp. 113-126.

Zhu, Y., Zhong, E. and Wu, B. (2013), "Feature engineering for semantic place prediction", *Pervasive and Mobile Computing*, Vol. 9 No. 6, pp. 772-783.

About the author

Tapio Soikkeli works as a Researcher and is a Doctoral Candidate at the Department of Communications and Networking at Aalto University in Finland. His research interests include handset-based measurements, mobile communication service usage, mobile end-user context and mobile user behavior in general. Tapio Soikkeli has started his doctoral studies in Fall 2011 and is expected to graduate as a Doctor of Science in Technology in 2015. Currently, Tapio Soikkeli holds an MSc (Technology) from Aalto University. Tapio Soikkeli can be contacted at: tapio.soikkeli@aalto.fi