Adaptive Sensor Cooperation for Predicting Human Mobility

Paul Baumann
Wireless Sensor Networks Lab
TU Darmstadt, Germany
paul.baumann@wsn.tu-darmstadt.de

Abstract
My thesis focuses on the prediction of human mobility. I am interested in gaining a deeper understanding of the factors that influence the performance of human mobility prediction algorithms. The main contributions of my work are: the analyses of different factors that influence the performance of mobility predictors, the design and development of a self-adaptive approach for detecting and recognizing users’ relevant places, and estimating users’ momentary predictability. The latter contribution aims to enable the possibility for the application scenarios to decide how much to trust the provided predictions and mobility data.

Author Keywords
Human mobility, prediction, localization

ACM Classification Keywords
H.m [Information systems]: Miscellaneous.

Expected Contributions
Several algorithms for predicting human mobility, e.g., users’ next relevant place, have been presented in the literature [9, 16, 8]. Many of these predictors estimate users’ future whereabouts by leveraging sequences of places visited by the users in the past. These sequences – to which we also refer to as mobility traces or historical data – can for instance be captured by leveraging users’ mobile phone [6].
In my thesis I focus on gaining a deeper understanding of the factors that influence the performance of human mobility predictors. The main contributions of my work that have been partially already done are: the analysis of different factors that influence the performance of mobility predictors [14, 13], the development of a self-adaptive approach for recognizing users’ relevant places, and providing an indicator about the users’ momentary predictability [12]. In particular, this thesis addresses challenges that we divide into 3 categories.

First, often mobility predictors rely on mobility assumptions [9]. Humans tend to follow simple and reproducible patterns, however, they still differ in their mobility patterns [10]. Thus, no single predictor performs uniformly well [9]. People tend to vary their mobility behavior. For instance, humans’ mobility might vary depending on the time of the day or even change over the time due to, e.g., a new home or workplace. Furthermore, accurately predicting human mobility requires careful attention to a number of important factors: the choice of an adequate predictor, the identification of data features to input into the selected predictor, and the selection of adequate metrics for evaluating its performance. The actual prediction task the predictor is required to solve might differ depending on the specific application scenario. To this end, we perform a holistic analysis of these factors that influence the performance of those predictors [14].

Second, providing location-based services requires knowledge about users’ relevant places. We refer to the definition of relevant places of a user as the set of places “where the user spends a substantial amount of time and/or visits frequently” [3]. In order to recognize these kind of places, a number of techniques have been proposed [5, 3]. We divide these techniques into geometry-based approaches, e.g., GPS, and fingerprint-based approaches that make use of the environmental signatures, e.g., Wi-Fi, Bluetooth, or acoustic. The performance of these techniques, measured in terms of different state-of-the-art metrics, varies from user to user and even depends on the different temporal and spatial factors. For instance, an algorithm that leverages Bluetooth data might achieve the best performance in recognizing home. At the same time another approach that relies on GPS provides better results in recognizing users’ workplace. This leads to the conclusion that no single algorithm nor location data source is sufficient to provide the best performance in all situations across all users.

Third, besides the knowledge about the predicted user mobility and the recognized relevant place, a number of applications would also benefit from the additional information about the confidence of this information. Focusing on the mobility prediction of a specific user, a home automation system can for instance use predictions with a high confidence about the arrival time of the resident to automatically control the heating system. At the same time, predictions with a low confidence might require additional user feedback before the home automation system takes further actions. In the context of our work, we refer to the predictions’ confidence as the users’ predictability. Lin et al. define predictability “as the information-theoretic upper bound that fundamentally limits any mobility prediction algorithm in predicting the next location based on historic records” [11]. Given this definition, Song et al. demonstrate that the average upper-bound for users’ predictability is at 93%. The average users’ predictability has been already studied in the recent published work [2, 11]. However, there are still daily-life situations in which despite high prediction confidence, the prediction will be wrong. Capturing and detecting these situations remains an open challenge. To this end, we focus on users’ momentary predictability, i.e.,
given the current users’ context (e.g., place or time of day), we are interested in answering the question how predictable is the user in this context.

Building upon these challenges, we first describe our contributions and then outline how they might be integrated to form a self-adaptive system for predicting human mobility. We first perform a holistic analysis of the factors that influence the performance of predictors. Among those factors, we consider temporal (e.g., time of day), spatial (e.g., current relevant place), phone contextual (e.g., last caller), and demographic (e.g., age or gender) data. We also consider different prediction tasks, e.g., predicting users’ next place. The goal is to uncover latent mobility patterns and to provide a deeper understanding of why, when, and which factors influence users’ mobility. We expect that our findings will allow configuring predictors, deciding which factors should be considered, and allowing to switch between algorithms depending on the current user context.

Our second contribution aims to provide a hierarchical approach for the identification and recognition of relevant places. To this end, we analyze the performance of different state-of-the-art algorithms that leverage available location data sources, e.g., GPS, Wi-Fi, GSM, or Bluetooth, to detect and recognize relevant places. Building upon our results, we plan to propose an algorithm that learns the aspects that influence the choice of the location data source and automatically decides which data source to choose in a self-adaptive manner.

Last but not least, we will propose an approach to estimate users’ momentary predictability. This value is then derived from the gained knowledge and the available sensor information in the particular context. All the three parts will then be integrated in the self-adaptive system that we introduce in the following section. We then highlight the arising research questions, the already provided contributions, and the planned next steps. We also briefly highlight related work and outline the main differences between our work and existing approaches.

A Self-adaptive System

This section describes our self-adaptive approach, the arising research questions and challenges, and the already provided and planned contributions. The goal behind our approach is to enable on-the-fly switching between different state-of-the-art predictors, algorithms for identifying and recognizing users’ relevant places, and required sensor data. The decision builds upon the provided contributions in this thesis. Besides the information about the selected algorithms and sensor configurations, our system provides an estimation of users’ momentary predictability allowing the envisioned applications to decide how much to trust the provided prediction result and the identified users’ places.

The architecture of our system is depicted in Figure 1 and comprises of four main components: MONITORING PREDICTABILITY, KNOWLEDGE FROM OFFLINE DATA ANALYSIS, ALGORITHMS AND SENSORS SELECTOR, and PREDICTORS. After providing a brief overview of the main components of the envisioned system, we next take a closer look at the first three mentioned components by highlighting the research challenges and questions, and the already provided as well as expected contributions in this thesis.

Monitoring Predictability
The main goal of this component is to estimate users’ momentary predictability. Thus, the research question is:

**RQ: How to estimate the momentary users’ predictability by leveraging the available sensor data?**

For this purpose, we leverage the introduced concept of
predictability by Song et al. [2]. They show that the average theoretical upper-bound for the average users’ predictability is at 93%. Lin et al. [11] extend this study by considering the usage of different temporal and spatial information scales, e.g., different size of time slots for determining users’ current place, to analyze the influence of temporal and spatial factors. Both approaches demonstrate that the right choice of this data does matter and has an influence on the estimation of users’ average predictability. One of the first steps towards the estimation of the momentary predictability has been done by McInerney et al. [7] that applied the concept of the instantaneous entropy (IE). The authors demonstrate a correlation between the estimated values and the usage of a number of smartphone applications, e.g., maps.

In our first step towards answering this research question, we perform an analysis on the IE with the goal to estimate this metric’s performance and to identify the potential room for improvements [12]. To this end, we define a synthetically generated daily mobility schedule that mimics users’ mobility between different relevant places. The main reason for using the synthetical schedule is the ability to compute the theoretical predictability. Furthermore, we define three “unpredictable” situations, i.e., mobility predictions will be wrong in most of these cases. Our results show that IE tends to under- and overestimate users’ predictability, especially, in situations when users change their place.

Given our first results, we are currently developing an estimator for the momentary predictability that builds upon the existing results. To this end, we leverage information about why and how the different temporal, spatial, contextual, and demographic data influence users’ mobility. This information is then provided by the component KNOWLEDGE FROM OFFLINE DATA ANALYSIS and is part of the contributions of this dissertation. From attending the UbiComp 2014 Doctoral School we further expect to get additional feedback.

Knowledge from Offline Data Analysis
This part of our system contains all the knowledge that allows other components to understand how to estimate the users’ momentary predictability. Furthermore, it covers the question which factors, when, and why influence the performance of predictors and algorithms for recognizing users’ places. The main research question (RQ) at this stage is summarized as follows:

RQ: Which sensor data and how needs to be collected in order to estimate users predictability and to select the best-performing predictors?

Accurately predicting human mobility requires careful attention to a number of factors, e.g., the choice of an adequate predictor. The actual prediction task the predictor is required to solve might also differ depending on the specific application scenario. We differentiate between four different prediction tasks: Next-slot Place (NSP) (i.e., relevant place for the next time slot) [1, 14], Next-slot Transition (NST) (i.e., whether the user will change her place in the next time slot – transition occurrence) [8], Next-place (NP) (i.e., next relevant place regardless when the transition will occur) [16], and Residence Time (RT) (i.e., how long the person will stay at the current place) [16, 13].

To this end, we first analyze [14] the influence of temporal and spatial factors on the predictors’ performance. We run an exhaustive evaluation by leveraging the maximum-likelihood approach [2] and 18 different combinations of temporal and spatial features. Our main insights include the demonstration that accuracy as a metric, which is frequently used to evaluate the predictors’ performance, is
not sufficient enough to evaluate predictors’ performance. Furthermore, we show a negative correlation between the predictors’ performance in solving the NPS and NPT tasks. The reason for this observation is the negative correlation between the temporal and spatial factors. The more temporal features, e.g., time of day, are used for the prediction, the better the performance in solving the NPT task. At the same time the worse the performance in solving the NPS task. Finally, we highlight a positive correlation between the number of algorithms that agree on the prediction and the prediction accuracy. This is the first step towards providing the predictions’ confidence.

In our recently completed but not yet published work, we consider phone context (e.g., last caller) and demographic features as well as the Next-place (NP) prediction task. We show that the influence of phone context data on the considered state-of-the-art predictors is negligible. At the same time, considering demographic data, e.g., age or gender, allows improving the performance. Building upon these results, we derive a population model that demonstrates which features and algorithms should be used in order to achieve the best performance for a given prediction task and metric.

In the previously mentioned work, we do not consider the Residence Time (RT) prediction task, but we plan to run similar steps in order to get a better understanding of factors that influence the prediction of residence times. We already took the first step towards this goal [13]. We adapt the concept introduced by Song et al. [2] in order to analyze the predictability of residence times. Our results show that predicting residence time is a harder task than predicting users’ next relevant place. Furthermore, we leverage 8 predictors and demonstrate that these tend to underestimate the residence time more often than to overestimate it.

A further contribution we are working on is to investigate which sensor data should be used in which situation in order to increase the performance in recognizing users’ relevant places. Hereby, we focus on the trade-off between smartphones’ resources usage and the algorithms’ performance. We are currently developing a hierarchical self-adaptive approach that learns the factors, which have the influence on this decision, and switches between the location data sources automatically [15].

In order to perform the offline analysis steps described above, we leverage a number of private and public data sets. Until now, we have based our work on the Nokia MDC [6] and our own collected data set that also includes ground-truth data. Recently, we have received access to the Cambridge data set [4] that we also will consider for our next steps.

Algorithms and Sensors Selector
The main purpose of this component is to combine all the available data in order to make the decision on which predictor and which algorithm for recognizing the relevant places should be used. To this end, the research question (RQ) is summarized as follows:

RQ: How to take advantage of the existing knowledge about the factors that influence users’ mobility in order to decide which predictors and sensor data should be used?

Therefore, we will design and develop an objective function that takes all the available data into consideration with the specific weights. The choice of the weights will then be evaluated towards optimizing the selected metric.

Objective for Attending the Doctoral School
My main motivation to attend the UbiComp 2014 Doctoral School is to present my work to fellow PhD students and senior researchers, to highlight the already achieved results,
and to get feedback for the next steps of my research. I attended UbiComp 2013 and learned from other students that attending the Doctoral School is a great opportunity to get feedback and to exchange ideas with other researchers.

Brief Biographical Sketch
Paul Baumann is a research / teaching assistant and a PhD student at the Wireless Sensor Networks (WSN) Lab, TU Darmstadt, Germany. He joined the WSN Lab in May 2012. His supervisor is Prof. Dr. Silvia Santini. During his PhD study, he has been multiple times for short research stays at ETH Zurich in the group of Friedemann Mattern and was a visiting researcher in 2014 for three months at Carnegie Mellon University in the group of Anind Dey. Paul expects to graduate by the end of 2015. He also has already published in the field on human mobility prediction at UbiComp 2013 [14], MobiCom 2013 [13], and SPAWC 2013 [12].

Acknowledgements
This work has been partially supported by the Collaborative Research Center 1053 funded by the German Research Foundation and by the Priority Program Cocoon funded by the LOEWE research initiative of the state of Hesse, Germany.

References