

Acquisition and Use of Mobility Habits for Personal Assistants

Lukas Nack, Roman Roor, Michael Karg, Olga Birth, Alexandra Kirsch, Bastian Leibe, Markus Strassberger

▶ **To cite this version:**

Lukas Nack, Roman Roor, Michael Karg, Olga Birth, Alexandra Kirsch, et al.. Acquisition and Use of Mobility Habits for Personal Assistants. IEEE International Conference on Intelligent Transportation Systems (ITSC), 2015, Las Palmas, Spain. hal-01693643

HAL Id: hal-01693643
https://hal.science/hal-01693643

Submitted on 26 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Acquisition and Use of Mobility Habits for Personal Assistants

Lukas Nack*, Roman Roor†, Michael Karg†, Olga Birth†, Alexandra Kirsch‡, Bastian Leibe*, Markus Strassberger†
*RWTH Aachen University, Templergraben 55, 52062 Aachen, Germany
Email: lukas.nack@rwth-aachen.de, leibe@vision.rwth-aachen.de
†BMW Group Forschung und Technik, Hanauer Str. 46, Munich, Germany
Email: {roman.roor, michael.karg, olga.birth, markus.strassberger}@bmw.de
‡Human-Computer Interaction and Artificial Intelligence
University of Tuebingen, Sand 14, 72076 Tuebingen
Email: alexandra.kirsch@uni-tuebingen.de

***Abstract*—With large parts of human population increasingly living in big cities, the mobility behavior of humans is about to change faster than ever before. Not only convenience and increasing ecological awareness lead to more intermodal mobility behavior, also the rise of new mobility options like car- or bike sharing are becoming more and more common. Wide distribution of smartphones and the on-trip availability of high-speed Internet let users inform themselves about a vast variety of mobility options. This information overload can overburden users who often have the simple wish to conveniently travel from A to B. Digital Mobility Assistants ease the burden of selecting the best mobility option for a particular user by incorporating the users' habits and preferences and providing relevant information at just the right time. To enable such intelligent assistance, we propose to create personalized mobility models that include not only information about habitual trips and destinations, but also allow for the detection of preferred travel modes. Our system is specifically designed to use sparse sensor data from mobile devices, such as smartphones, to offer an adequate balance between battery-life and data quality.**

I. Introduction

Today's population experiences a growing trend of urbanisation. For the first time in history, in 2007 more people were living in cities than in rural areas. The United Nations Population fund expects the number of people living in metropolitan areas to rise up to 5 billion by 2030 [1]. Providing mobility to such a vast amount of people living together in narrow space not only poses challenges to city planners and mobility providers. It also confronts inhabitants with challenges when planning their daily mobility. The variety of mobility options including private car, public transport, car- and bike-sharing as well as taxis and private shuttle companies offer services each of which might be the best option in a specific situation. While a broad selection of offers is generally advantageous for customers, it makes it hard to decide which mobility option is the best one for a specific customer in a specific situation. The pure display of all available information without accounting for a user's needs can overburden users, leading to frustration and impairing the advantage of the large range of mobility options.

Digital Mobility Assistants are designed to easy the burden of selecting the appropriate mobility option for a specific user in a specific situation by incorporating personal profiles to display only information that is relevant to the user.

Our daily mobility is, in fact, characterized by a deep-rooted regularity [2]. They found that a majority of trips are habitual trips that are carried out at the same times to the same places and it is possible to predict up to 93% of such trips. To enable these predictions, we propose a new method to generate personal mobility profiles by the aggregation of various sensor data and the application of different algorithms. Such mobility predictions might, for instance, include the most likely destination of an upcoming trip as well as likely departure times. It is also possible to draw conclusions about mobility habits and identify trips that are repeatedly executed to the same destinations around the same time.

Based on information about trips of a user in near future, a digital mobility assistant can, for instance, display information only about the preferred mode of transportation on the preferred route. Such information might be the current traffic situation on the trip in case the car is preferred, or the departure times of the preferred subway of a person. Thus, the user is presented with relevant information instead of being overwhelmed with information he/she is not interested in.

In this paper, we present a framework that aggregates spatio-temporal data from mobile devices, for instance smartphones, in the backend. It uses such data to generate models of mobility habits to enable predictions about upcoming destinations, departure times and the preferred transportation mode of a person. The framework is specifically designed to work on sparse smartphone data to achieve a decent battery life while maintaining data quality that is accurate enough for our predictions.

II. Related Work

The prediction of human mobility behavior has been explored from a variety of facets in research. Krajnik et al. show how human behavior can be forecast for a short period of time after an action was initialized. In [3], Pentland et al. predict short term human behavior based on automobile drivers' subsequent actions from their initial preparatory movements. They model human behavior using a Markov Dynamic Model

(MDM) which is based on Kalman filters sequenced together by a Markov chain. In [4], Liao et al. build hierarchical Markov models to learn and infer a person's daily movements. Their approach is based on several levels of abstraction in order to bridge the gap between raw GPS measurements and high level information. Supervised learning is used to infer the transport mode and means of transportation in order to predict popular destinations of a person. Furthermore, Liao et al. describe how to extract a person's activities from GPS traces using Conditional Random Fields (CRF) [5].

An improved approach towards understanding human mobility is proposed by Zheng et al. [6]. He and his team extracted new features out of GPS logs, such as heading change rate (HCR), stop rate (SR) and velocity change rate (VCR). They use the data in addition to their previous work [7] to train a Decision Tree-based inference model. The new features result in eight percent improvement in inference accuracy over previous results.

The approaches mentioned so far focus on learning and inferring user actions from GPS traces logged with mobile phones. But the extracted features mentioned by Zheng et al. are well suited to distinguish between bus, car and walking as well. To achieve a very high accuracy in transport mode detection (over 90%), Geographic Information System (GIS) information is used [8] in many cases. Hence, it is not only possible to detect the transportation mode, such as walking, bike, car, bus, etc. but furthermore information about which particular bus the user is riding. Gonzales et al. [9] even try to understand human mobility patterns. Studying the trajectory of 100,000 anonymized mobile phone users, they found a high degree of temporal and spatial regularity. It particular, it exists a high probability to return to a few highly frequented locations, such as home or work. To be able to understand human mobility patterns and predict trips, the recognition of unknown behavior and anomalies is promising. Liao et al. [4] show how to recognize unknown behavior and anomalies, and Ma et al. [10] propose a method to detect anomalies from GPS logs in real-time for traveling individuals without the need for a previously trained model. Their approach relies on trajectories to be modeled as a discrete-time series of axis-parallel constraints ("boxes") in the 2D space. The incremental comparison between two trajectories where one trajectory has the current movement pattern and the other is a norm can be calculated according to similarity between two boxes. Biagioni et al. describe how to detect anomalies between day routines based on day clustering and how to predict future travel [11]. There is no need for supervised learning, instead a modified similarity algorithms was applied to clustering days using location traces.

In this work we use approaches like described by Liao et al. to infer human mobility by tracking the user and training a HMM [4]. To gain more knowledge about locomotion of the user we use a HMM for travel mode detection instead of a Decision-Tree based inference model like proposed by Zheng et al. [7]. Hence, enables us to enrich the user's mobility with informative knowledge and provide pervasive computing systems with more context information. Taking this into account our framework can predict user's future mobility behaviour regarding temporal information as well as transportation mode information about the next trip.

III. Sourcing of Mobility Data

To be able to extract spatio-temporal mobility patterns from GPS logs we need information about the daily mobility from subjects. Three persons were asked to track their daily life and allow us to analyze their movement information. The intention is to not equip the subject with additional expensive GPS logging hardware, but rather use already owned telemetry of the user, i.e. built-in smartphone sensors (GPS, acceleration, pressure, etc.). On the one hand we might not get the most possible accurate data but on the other hand we can check our approach against available real life data because a) the subject does not have to carry additional hardware and can perform all daily routines as usual and b) then we can evaluate what is the possibility to use the approach with average users.

The mobility data is therefore collected from smartphones using a commercial application. It saves location readings every 45s on average. Three different persons recorded their outdoor movements over a period of several months resulting in 415 hours of data. The data is mostly located in the area of Munich, Germany. In total, more than 13,000 km are collected and divided in 1,700 movement segments. Places, at which the user stays for a longer period of time, are extracted from the movements. The mobility movements are manually labelled with the utilized travel modes which consists out of car, bus, train, tram, subway, walking, running, cycling, plane, and Pendelbus. The latter is a private shuttle bus service which is used frequently by the test users.

A GPS log p_k represents data recorded with the commercial application mentioned above using a smartphone. The format of the log is $p_k = \{lat, lon, t\}$ where: *lat* represents the latitude; *lon* represents longitude; and *t* represents the timestamp of the sensor report

The measurement units of the GPS log attributes are: latitude (*lat*) and longitude (*lon*) are in decimal degree; and the timestamp of the GPS log (*t*) is in ISO 8601 combined date and time format.

A GPS trace list D is a sequence of GPS logs, $D = \{p_0, \ldots, p_N\}$, where the timestamp in the sequence strictly increases.

Additionally, road and public transportation network data is extracted from the OpenStreetMap.org[1] project, see Fig. 1.

IV. Model Generation

To enable our envisioned predictions of mobility behavior, we use different models to capture spatio-temporal patterns in our mobility data. Therefore, we use a preprocessing step to segment our data between places where the person moves and places where he/she is standing still. Based on this segmentation, we can create relations between places the user visited and cluster resulting trips according to different features like weekdays, travel mode or the route the user has taken.

A. Location and Trip Clustering

In a first clustering step, we partition our raw mobility data, which consists of tuples of timestamp, latitude and longitude.

[1] http://www.openstreetmap.org

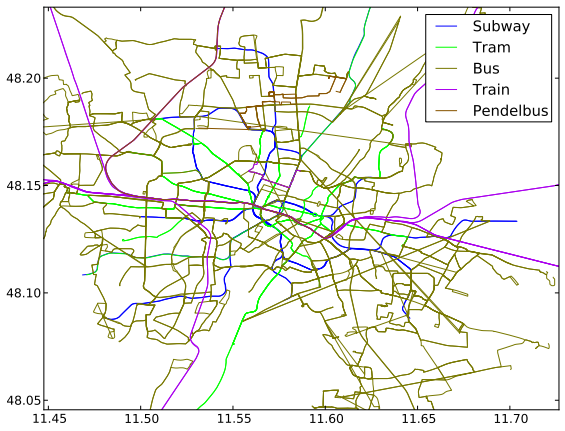


Fig. 1. Geographic Information System (GIS) data is extracted from the OpenStreeMap.org project. The public transportation networks from the inner city of Munich can been seen which is used for the feature extraction by using a map matching method.

Therefore, the data is segmented into parts where the person is moving (moves) and parts where the person is standing still (stops). One challenge when using smartphone-based location data arises from the inaccuracy of the localization measures. Even when a person does not move, the estimated location might vary due to uncertainties in the location estimation. This uncertainties are emerged from a combination of different algorithms which use a variety of data sources, such as GPS, GSM or WiFi. While GPS offers the most accurate location data, it is often not available for places where a person stays for an extended period of time, e.g. inside buildings. To account for such variances and to distinguish between stops and moves, we use the CB-SMoT algorithm.

CB-SMoT [12] operates in two phases. In the first phase, stops are extracted from the recorded location data. Here, a variant of DBSCAN [13] is used. At the beginning, all points are set to be unprocessed. Iteratively, for each unprocessed location p_k the Eps-neighborhood N_{Eps} over the time-sorted point list $D = \{p_0, \ldots, p_N\}$ is computed such that:

$$\begin{aligned} N_{Eps} = &\{p_m \in D | (\sum_{i=m}^{k-1} dist(p_i, p_{i+1})) \leq Eps\} \cup \\ &\{p_n \in D | (\sum_{i=k+1}^{n} dist(p_{i-1}, p_i)) \leq Eps\}, \end{aligned} \tag{1}$$

where $p_i = (x_i, y_i, t_i)$ with t_i holds the timestamp of the location and $t_0 \leq t_m < t_k < t_n \leq t_N$. x_i and y_i are the longitude and latitude, respectively. The parameter Eps determines the maximum distance between p_k and its neighbors on the point list.

The point p_k is considered to be a core point with respect to Eps and $MinTime$ if $|t_n - t_m| \geq MinTime$. $MinTime$ describes the minimum duration the user has to stay at one location to be count as stop. Therefore, by setting both parameters, Eps and $MinTime$, we only allow a maximum average speed of $Eps/MinTime$ for a stop. This speed limit should be set in such a way that both, slow walking segments and the inaccuracy of measurements at stops, are detected correctly. We choose values of $75m$ for the Eps parameter and a minimum time of 3 min a user should stay at the same stop. This corresponds to a speed limit of 1.5 km/h.

If point p_k satisfies the conditions of a core point, a new stop $S = (R_S, \Delta_S)$ is created where R_S describes the region of the assigned locations as topologically closed polygon and Δ_S is the visiting duration. Furthermore, for all points in N_{Eps} the Eps-neighbourhood is computed and the resulting, unprocessed points are added to S. In a last step, the points in S are marked as processed and the procedure is repeated for the remaining unprocessed points in D.

The second phase of CB-SMoT finds movements and merges stops. In general, a user visits specific places very frequently, e.g. work and home. To build our prediction and habit detection on these frequently visited places rather than on each stop itself, CB-SMoT clusters the resulting stops from the previous step. Therefore, we compare each stop S_i with already found places. If the region R_{S_i} of S_i intersect with cluster C_k, the stop is assigned to the cluster. Otherwise a new cluster is created. For each sub-list of points which is not assigned to a stop, a movements is generated.

We obtain a number of clusters and a number of transitions between those clusters, which represent single trips of a person. In a second clustering step, we create start-destination relations between places which correspond to a user's trips. This simple clustering basically just categorizes trips according to their start and destination locations. For instance, all trips from a person's home location to the person's workplace represent one cluster of trips.

B. Travel Mode Detection

After obtaining movement segments from the raw location data, transportation modes are estimated from the resulting trip data. We distinguish between both motorized and non-motorized travel modes, including car, bus, subway, train, tram, Pendelbus, plane, walking, running and cycling.

A point-based approach is taken for the travel mode detection, i.e. instead of predicting the most probable transportation mode for each movement segment, we predict the most probable mode per movement location. As described in section IV-A, a movement segment is constructed between two consecutive stops at which the user stays for a minimum period of time. However, when changing from one transportation mode to another, a person does not always stop in between. For example, after arriving at a station via train, the user might directly walk to her car and continue travelling to her home. Such a behavior can be correctly classified using a point-based approach but not with a segment-based detection.

For each point of a movement segment, a set of features is extracted using a GIS of the surrounding environment and the raw location data. Similar to Zhang et al. [14] we use a set of basic and advanced features. The basic features consist of velocity, distance, velocity- and heading change. All these features are computed by comparing the current point with the last processed location of the movement. The more advanced features use GIS information. For each public transportation mode, we compute the distance from the current point to the nearest transportation line of this type.

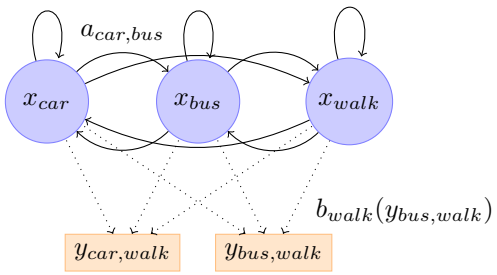


Fig. 2. The HMM detects the travel modes of a location point. For a better visualisation, only 3 states are used (car, bus, walk) which can emit 2 discrete symbols $y_{car,walk}$ or $y_{bus,walk}$. a_{ij} is the probability to transition from state x_i to state x_j. $b_j(y_k)$ is the probability to emit symbol y_k in state x_j.

When a person is riding a bus, locations of the location trace should be close to a bus line network. The other public transportation lines, e.g. a train line, should be farther away from the trace points. A special case should be considered when a user travels by an underground transportation mode, e.g. subway. Here, the location updates are most likely dropping in their frequency since smartphones are not able to receive GPS signals underground. Using GPS alone, the only recorded locations are probably at the entry and the exit of the subway. Using additional localization sources, i.e. GSM and WiFi signals, it is possible to receive location updates at subway stations given there are WiFi networks in range. Hence, the distance of a point to the next subway line does not provide as much information. The distance from each point to the next subway station proves useful in this case.

The framework for travel mode detection we created consists of two parts. In an offline step, features are extracted from a manually labeled ground truth dataset (see section III). A first guess of the travel mode builds up on a supervised learning method. We choose a random forest [15] to construct the inference model using a training set of a series of movement segments. Since the provided features are optimized to describe different motorized travel modes, additional features are used to compensate the lack of information of non-motorized modes. In particular, we benefit from Androids build-in Activity Recognition. This framework provides low cost methods to extract additional travel modes which include walking, running, cycling, and general driving.

A Hidden Markov Model (HMM) is constructed using the training set of movements. Therefore, the observation states of the HMM are both the predictions of the Random Forest and the Activity Recognition results. The hidden variable is assembled by the final travel mode prediction for the location points. See Fig. 2 for a visualisation.

Let $Y = (y_1, y_2, \dots y_L)$ be the observation sequence of length L where $y_t, t \in (1, L)$ is the observation of one point. Hence, y_t is a tuple of the output of the random forest and the Activity Recognition. Since we are interested in the joint probability $P(x_t, y_{1:t})$ where x_t is the hidden variable of point t and $y_{1:t} = (y(1), \dots, y(t))$, we use the Forward Algorithm to filter the HMM. Both, the transition matrix and the emission matrix, can be extracted directly from the ground-truth dataset and are given by $P(x_t|x_{t-1})$ and $P(y_t|x_t)$, respectively.

In an online step, the features of motorized travel modes are extracted from a set of movement locations. A first prediction is obtained using the random forest. The first guess together with the activity recognition result are the input data for the HMM. In a last step, the HMM provides the final travel mode prediction for the movement segment.

A simplified HMM is constructed for evaluation. It uses only the observations emitted by the random forest without the activity recognition. The observations are no tuples but the discrete predictions obtained from the random forest. Therefore, a main task of the simplified HMM is to smooth the observations. Using a point-based prediction model, the results can oscillate between two or more travel modes in one movements segments. For instance, when a user is driving by car and crosses a bus line segment every now and then, she might get the observation sequence $(car, car, bus, car, bus, car)$. Using the simplified HMM the prediction is smoothed and results in the correct sequence $(car, car, car, car, car, car)$. The transition matrix and the emission matrix can be obtained as explained above.

C. Trip clustering

In this third preprocessing step, we account for the departure times of specific trips from one location to another using a specific mode of transport on a specific weekday. We define a trip as a relation between a start cluster and a destination cluster that are generated as described in sectionIV-A. Trips are clustered according to their start and destination clusters, their mode of transportation and the weekday of the trip. Accounting for weekdays is important, since the regularity of most persons mobility strongly varies depending on the current day being a workday or weekend. To this end, we distinguish each specific weekday from Saturdays, while we treat Sundays and holidays the same. One cluster of trips might, for instance, consist of various trips that were done by car from the cluster "home" to the cluster "work" on Mondays.

For each trip cluster, we can now generate statistical models about common departure times. Since a set of observed departure events in the continuous domain of time can in most cases not be described by parametrized functions, we need a way of estimating the probability density function of a departure event over time in a non-parametric way. We therefore use a variant of Kernel Density Estimation [16] to describe a probability density function of the random variable using a Gaussian kernel function. Such an exemplary probability density is illustrated in figure 3. Here, the probability p of a departure from a specific cluster to a destination is plotted over time t. The function is generated by the Gaussian kernels of the observed data points t_i.

The estimated density is represented by the average of all kernel functions averaged over the number of data points N as illustrated in equation 2.

$$P(x) = \frac{1}{N} \sum_{i=1}^{N} G(t_i) \tag{2}$$

$G(t_i)$ here describes the Gaussian kernel function.

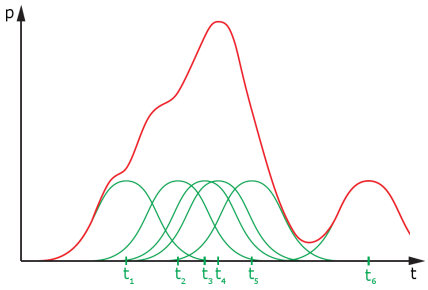


Fig. 3. An exemplary gaussian kernel density estimation over departure times of a specific trip generated by the observations t_i.

V. Destination Prediction and Habit Detection

In this section we describe how destination prediction and habit detection is done by our framework.

A. Destination Prediction

The models generated so far can be used to predict likely next destinations of a person given a location and optionally a time as an input. Such predictions can be useful for Digital Mobility Assistants to filter information that is to be presented to the user. The destination prediction uses information about the current location as well as the current time as input and outputs a weighted list of likely next locations. Our implementation is based on a combination of heuristics that are merged to return a confidence value for each location cluster. The destination prediction considers the following factors:

- **Frequency of visits:** When a prediction is triggered from one location, only locations that have been approached from the requested locations are considered as likely destinations. The higher the frequency of past visits (from the requested location) is, the higher is the impact of the specific destination.
- **Day of week:** Each type of day is treated differently regarding the prediction. We distinguish each weekday individually and group sun- and holidays.
- **Departure time:** For each cluster, a mean departure time is measured. We therefore perform a local search in the probability density function described in section IV-C.

We use this combination of heuristics due to good experiences in previous projects, but alternative prediction methods like Bayesian Networks [4] or Conditional Random Fields [5] also proved to offer reliable predictions. An overview about different methodologies and common accuracies can be found in the work of Herder et al. [17].

B. Habit Detection

Based on the statistical models explained in section IV-A, we can not only create predictions about likely next location, but also extract patterns in the users' daily mobility. We therefore partition the data into tuples of days and analyze trip clusters with respect to human habits. We therefore calculate a habits value to express on how many days since data recording a specific trips has been made. For instance, if a person traveled from work to the gym on six out of ten days, the habit value for this trip would be 0.6. Using such a habit value, we can group habitual trips according to their habit-strength. First experiments showed that for our participants, the detection of commutes from home to work obtain high habit values, as well as trips to canteens, or similar. We can also draw conclusions about common departure times, arrival times and durations of habitual trips based on the corresponding trip clusters.

The filtering of trips according to their habit value enables us to generate an overview about habitual trips sorted by their strength. Such information could, for instance, be used to predict trips that a user will do some when during his/her day, thus supporting him in planning his/her day. In contrast to the destination prediction introduces in section V-A, this predictions do not depend on a specific trigger and enables predictions farther into the future than only the next most likely trip. The habit detection presented so far only represents a first approach into habit detection and will be subject of future work.

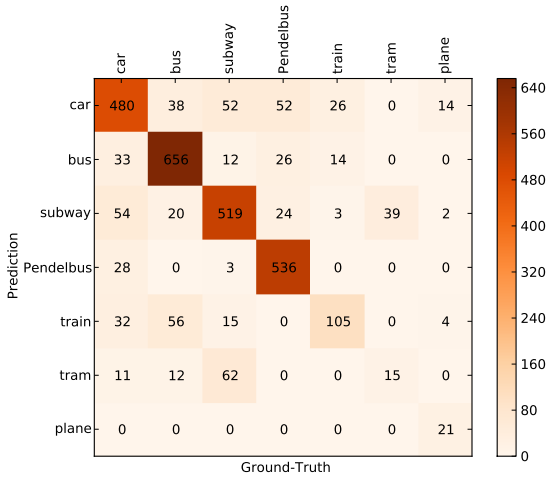


Fig. 4. The confusion matrix displays the performance of the travel mode detection. The simplified HMM is trained by an pseudo-random subset of training data. Another i.i.d. subset is used for the evaluation. The evaluation data is obtained by 10-fold cross-validation using around 3,000 data points. It is visible that most movement points lay on the diagonal and are therefore classified correctly. The simplified HMM achieves an accuracy of 90.9%.

VI. Evaluation

In this section we evaluate the results of our system against the ground truth.

A. Travel Mode Detection

The validation of the travel mode detection is based on the data collected by the commercial smartphones application. To keep the amount of different transportation modes balanced, pseudo-randomly, independent and identical distributed (i.i.d.) subsets are used for training and testing. The subsets are assembled in a way to keep the amount of different travel modes equally distributed inside each set. Since the application does not provide access to the Activity Recognition data, the preliminary evaluation uses the simplified HMM model for the travel mode detection. Therefore, only motorized travel modes are considered.

TABLE I. AVERAGE PERCENTAGE OF CORRECT PREDICTIONS OF OUR DESTINATION PREDICTION. THE LEFT COLUMN INDICATES THE PERCENTAGE OF CORRECT PREDICTIONS OF THE NEXT LOCATION OF A PERSON. THE RIGHT TABLE INDICATES SITUATIONS WHERE THE NEXT VISITED LOCATION OF THE TEST PERSON WAS WITHIN THE MOST LIKELY THREE PREDICTED DESTINATIONS.

	Most likely	Within first 3
Overall	57.7 %	75.2 %
Weekday	62.1 %	76.4 %
Weekend	48.8 %	71.2 %

A total of 3,000 location points are employed and divided into training and test data according to a 10-fold cross-validation. In each run, a random forest is trained. Afterwards, the Hidden Markov Model smooths and corrects some of the misclassified travel modes using the same set of training data.

The prediction model of the random forest achieves an average accuracy of 89.8% on the cross-validation. However, the travel mode detection performance can even be increased slightly using the HMM which provides an average accuracy of 90.9%. Therefore, it increases the travel mode detection performance by around 1%. Nevertheless, the introduction of the HMM provides foremost a method to handle Activity Recognition data of the Android platform which is not included in the current evaluation.

A confusion matrix with a detailed detection performance can be seen in Fig. 4. All travel modes which are located on the diagonal, are detected correctly while all other points are misclassified by either the random forest or the HMM.

B. Destination Prediction

In an initial evaluation step, we estimated the correctness of our destination prediction explained in section V-A. For this evaluation, we used six persons of our dataset and split their trip data into mutually exclusive training sets of six weeks, while three weeks data were used for testing. Generally, the average percentage of correct predictions varied between 73 % and 45 % depending on the different persons. We also investigated, in how many cases the correct prediction is within the most likely three destinations, where the correctness varied between 94 % and 63 % in this case. We also distinguished between weekdays and weekends and averaged percentages of correct predictions are shown in table I.

We assume the high variance in the prediction rates being caused simply be the different way of lives of our test persons. While some of our participants are strongly driven by a high amount of regular events (commutes, kindergarden, ...), others experience a lower amount of regularity in their daily life. Since our dataset is rather small to allow for a general conclusion about the average accuracy of our destination prediction, such an evaluation is left for future work.

VII. CONCLUSION

In this paper, we introduced a framework to generate models of habitual human mobility behavior using spatio-temporal data from mobile devices. The framework includes an estimation of likely next locations of a person as well as corresponding departure times and the preferred mode of transportation. We evaluated our approach, focusing on the evaluation of travel mode detection. We showed how to distinguish between different motorized modes of transportation even when only sparse location data from mobile devices is available. Future work will include the clustering and prediction of different routes or lines for single trips using one specific mode of transportation as well as further research in the detection of mobility habits of users.

REFERENCES

[1] G. Martine, A. Marshall *et al.*, "State of world population 2007: unleashing the potential of urban growth," in *State of world population 2007: unleashing the potential of urban growth*. UNFPA, 2007.

[2] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[3] A. Pentland and A. Liu, "Modeling and Prediction of Human Behavior," *Neural Computation*, vol. 11, no. 1, pp. 229–242, 1999.

[4] L. Liao, D. J. Patterson, D. Fox, and H. Kautz, "Learning and Inferring Transportation Routines," *Artificial Intelligence*, vol. 171, no. 5, pp. 311–331, 2007. [Online]. Available: https://www.cs.rochester.edu/~kautz/papers/gps-tracking.pdf

[5] L. Liao, D. Fox, and H. Kautz, "Extracting Places and Activities from GPS Traces Using Hierarchical Conditional Random Fields," *The International Journal of Robotics Research*, vol. 26, pp. 119–134, 2007.

[6] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on GPS data," *Proceedings of the 10th international conference on Ubiquitous computing - UbiComp '08*, no. 49, p. 312, 2008.

[7] Y. Zheng, L. Liu, L. Wang, and X. Xie, "Learning transportation mode from raw gps data for geographic applications on the web," in *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008, pp. 247–256.

[8] L. Stenneth, O. Wolfson, P. S. Yu, and B. Xu, "Transportation Mode Detection using Mobile Phones and GIS Information," *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 54–63, 2011.

[9] M. C. González, C. a. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns." *Nature*, vol. 453, no. June, pp. 779–782, 2008.

[10] T.-S. Ma, "Real-time anomaly detection for traveling individuals," *ACM SIGACCESS Conference on Computers & Accessibility*, pp. 273–274, 2009.

[11] J. Biagioni and J. Krumm, "Days of our lives: Assessing day similarity from location traces," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7899 LNCS, pp. 89–101, 2013. [Online]. Available: http://www.cs.uic.edu/pub/Bits/JamesBiagioni/biagioni-umap13.pdf

[12] A. T. Palma, V. Bogorny, B. Kuijpers, and L. O. Alvares, "A clustering-based approach for discovering interesting places in trajectories," in *Proceedings of the 2008 ACM Symposium on Applied Computing*, ser. SAC '08, 2008, pp. 863–868.

[13] M. Ester, H. peter Kriegel, J. S, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise." AAAI Press, 1996, pp. 226–231.

[14] L. Zhang, M. Qiang, and G. Yang, "Mobility transportation mode detection based on trajectory segment," *Journal of Computational Information Systems*, vol. 9, no. 8, pp. 3279–3286, 2013.

[15] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[16] M. Rosenblatt *et al.*, "Remarks on some nonparametric estimates of a density function," *The Annals of Mathematical Statistics*, vol. 27, no. 3, pp. 832–837, 1956.

[17] E. Herder, P. Siehndel, and R. Kawase, "Predicting user locations and trajectories," in *User Modeling, Adaptation, and Personalization*. Springer, 2014, pp. 86–97.