K. A. Hummel and A. Hess, "Estimating human movement activities for opportunistic networking: A study of movement features," World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a, Lucca, 2011, pp. 1-7. doi: 10.1109/WoWMoM.2011.5986468

Estimating Human Movement Activities for Opportunistic Networking: A Study of Movement Features

Karin Anna Hummel and Andrea Hess
University of Vienna, Faculty for Computer Science, Research Group Entertainment Computing
Email: {karin.hummel|andrea.hess}@univie.ac.at

***Abstract*—In mobility-assisted, opportunistic networks, data is disseminated in a store-and-forward manner by means of spontaneously connecting mobile devices. Therefore, mobility itself moves in the center of investigation. Knowledge about movement characteristics of single devices can be used to add realism to random mobility models and to understand the likelihood of communication options.**

This paper contributes to the field of observing movement characteristics of single devices for opportunistic networks by describing movement features and investigating how these features can contribute to human movement activity estimation. Activity descriptions are useful for characterizing the purpose of movement. Additionally, in case movement patterns are uncertain or fragmentary, knowledge about activities may help to estimate average movement characteristics faster. We use activity estimation based on the Naïve Bayes classifier applied to a multi-variate feature set consisting of commonly considered movement features. We investigate the classification success rate experimentally when using all features and when using only a subset of features. Therefore, we conducted a user study collecting real-trip GPS traces labeled by the users. We selected four most frequent urban movement use case activities for classification and achieved a success rate of 80.65%.

I. Introduction

Within the field of opportunistic networking [18], we focus on investigating single device movement characteristics (movement features) to, finally, come up with estimates about preferable forwarding nodes. In this paper, we contribute with providing a method for estimating movement activities based on movement features used to relate movement features to the purpose of a trip.

We have described the method and first results for estimating movement activities by using ”artificial traces”, that are, GPS traces generated artificially by performing movement activities along a movement description (e.g., going to a specific shop and returning) in [9]. Although the results were promising, we could not show the full potential of the approach and extended our work by conducting longer-term experiments producing labeled real-trip GPS traces (i.e., test persons performed daily activities and manually entered the trip purpose). The contribution of this paper is to give quantitative insights in the classification success rate of activities, causes for wrong classifications, and an investigation of the contribution of single movement features to the overall estimation success rate. The work is general in the sense that the approach can be applied to arbitrary movement features and movement activities where the results presented depend on the selection of features, selection of use case activities, and quantization effects used during feature aggregation.

First, we provide a summary of approaches to mobility modeling and movement feature extraction and their application to networking and activity recognition (Section II). Based on an extensive survey of movement features used in the field of mobile modeling, we introduce the most commonly used features meaningful for single devices and extend the list by two simple features, i.e., velocity and direction changes (Section III). In addition to our previous work [9], we add also the *start time* to the set of features as we observed that the daytime has a strong influences on the occurrences of activities. In a descriptive way, we discuss how the movement features may influence opportunistic networking characteristics. In Section IV, we detail the methodology to estimate activities based on a multi-variate Naïve Bayes classification approach. The urban movement activities that have been selected as being major, typical activities are: *Way to Work*, going out for daily *Shopping*, *Evening Activity*, and being a *Tourist* (Section V). Hereby, we selected the most frequent activities reported by the 13 volunteers of the user study conducted. We present and discuss the achieved activity classification success rates based on a data set of 252 trips in Section VI.

II. Related Work on Mobility Modeling and Movement Activity Recognition

For the purpose of capturing and modeling mobility realistically, constraints such as geographic and topological restrictions can be applied ranging from simple grids to detailed maps of urban areas. Another approach is to investigate real-world movement traces and derive patterns by statistical analysis or to find appropriate parameter ranges and distributions for mobility models. Typical results are that well-known distributions can be assumed for characteristics, such as the log-normal or Gaussian distributions for speed, pause time, and the placement of hotspot regions in an area [13]. Another insight has been gained, e.g., in [6], by modeling human trips as power-law flights. The Self-similar Least Action Walk (SLAW) mobility model [14] is based on

978-1-4577-0351-5/11/$26.00

similar statistical patterns combined with a least action trip planning algorithm.

Additionally, algorithms in the field of opportunistic routing and forwarding utilize movement features and patterns. A controlled epidemic forwarding approach where relay nodes are chosen depending on their path history is proposed in [2]. In [17], forwarding nodes are selected by predicting future node contacts based on a 'delivery predictability' metric assuming that a node is suitable for message forwarding to a particular other node if these two nodes meet frequently. In [15], [3], routing decisions are based on similarities in patterns such as meeting frequency and visited locations of destination and forwarding nodes.

Recently, some interest has been raised in movement activity description and estimation, which is a technique used in traffic planning tools and scenario-based mobility simulation, such as the commercially available Legion mobility simulator[1]. Similar to our work, every day activities are considered such as in the Working Day Movement Model which considers a human mobility model based on surveys [5]. Recognition of activities based on real-world mobility data is a different approach. High-level activities can be derived in location-based activity recognition by detecting significant places and inferring the types of the places (such as work, home, at a friend, and parking lot) [16]. In [4], the Reality Mining dataset[2] is analyzed, where an activity type (home, work, other) is derived from a participant questionnaire for each GSM cell association. Particularly, the authors investigate the usefulness of certain time points for predicting future activities.

Related to activity recognition, richer information about mobility behavior can be derived from mobility traces. Recently, modes of transportation have been inferred from captured GPS and accelerometer or cellular data. In [19], the mode detection is based on velocity, whereas in [22], travel time, trip length, and a reference travel time for different types of transportation modes between start and end location is used (provided by GoogleMaps). Similar to these works, we contribute by estimating movement activities, while also extending these results by using a broad set of commonly used movement features.

III. Movement Features and Networking Effects

In related work, movement features are used based on the aim of mobility modeling. For example, in [10], a time-variant community mobility model is proposed based on location visiting preferences and periodical re-appearances at the same location observed in WiFi traces. To the best of our knowledge we were the first presenting an overview of commonly used movement characteristics [9] based on an extended survey. While many related data sets are based on network-related locations (associations with WiFi access points or UMTS cells, contacts based on Bluetooth connections), only few works consider finer grained mobility traces based, e.g., on GPS. As we use fine-granular GPS data, we extended the set of commonly used features by adding *velocity* and *direction changes*. Additionally, *start time* is considered as a simple but very effective feature to distinguish between movement activities.

A. Movement Features

Movement, as a process of changing location over time, can be described from the perspective of the dimensions *space* and *time*. Consequently, movement features can be either assigned to the classes of (i) *spatial features*, (ii) *temporal features*, and (iii) *spatio-temporal features*. The features are detailed and defined as used in this paper.

Spatial Features: Spatial features used to describe single or group movement are:

- **Flight length.** The flight length is defined as the distance between two end points of a path without a pause (compliant to [14]). An additional constraint is, that the distance of each sampled position in between to a straight line connecting the end points is restricted. In this paper, the flight length is calculated as the length (in meters) of a path traveled between two consecutive pause times.
- **Activity range.** The area covering all locations that have been visited is termed the activity range (e.g., as in [12]). Here, the term **mobility range** is used similarly to activity range and is defined as the distance (in meters) of each position within a trip to the center of a rectangle covering the trip. (Note, that other geometric areas can be used similarly.)

Additionally, *prevalence* is defined as the fraction of time a user spends at a given location [21], [12] and treated as a spatial metric determining the spatial distribution of users. A spatial feature which is derived from the basic prevalence metric is, for instance, 'popular location' [10], [13]. For small experimental data sets which do not cover the area sufficiently, this metric is not useful and, thus, will not be used in this paper.

Temporal Features: Temporal features commonly used and used in our study are:

- **Pause time.** The pause time is defined as the duration between two consecutive movement phases [14]. In the experiments presented, a time period is considered as pause time if the velocity is less than 0.5m/s for at least five seconds.
- **Start time.** The start (and end) time allows to capture daytime dependent behavior. In relation to our approach is the use of these times in [13]. In our work, we use the hour of day a trip started to calculate the values of this feature.

[1] http://www.legion.com/

[2] http://reality.media.mit.edu/download.php

Additionally, *persistence* is used as a metric to capture the time a user stays associated with a position or network element, such as an access point. In [21], the session length distribution is derived from persistence. This metric is not used further in this paper, since it is closely related to pause time and we do not consider network associations.

Spatio-Temporal Features: Spatio-temporal features are combinations of spatial and temporal properties, here, we use:

- **Revisits.** Repeated appearances at locations have been widely investigated in related work, such as periodic re-appearances (at the same access point after a certain time gap [10]), the return time, i.e., the time until a node returns to an area after exiting this area, and the hitting time, i.e., the remaining time until a node will reach a location [10]. In this paper, we will use and investigate the **number of revisited positions** within a trip and the **time between revisits**. Positions are assumed to be "the same" position (i.e., a revisited position), if they lie within a configurable position radius. We selected a radius $r = 20$m, which is feasible for GPS accuracy in urban areas. A revisit is only assumed if the moving person has first moved away from the position (here, 50m).

Meetings are often captured in terms of frequency and duration of contacts between nodes. In particular for opportunistic networks, contacts are of major interest to determine forwarding behavior. Metrics used to describe meetings are inter-meeting time, i.e., the time elapsing between two successive contacts (contact periods) for a node pair (see, e.g., the work in [8], [11], [20]), contact time [8], [11], meeting time [11] (i.e., the time until two nodes meet for the first time), (inter-)any-contact time [8], and time distance [20] (i.e., the shortest time a node can pass a message to a particular node). In this paper, we focus on individual movements and do not capture meetings.

Additionally, we use the average **velocity** (i.e., the speed between two positions measured consecutively[3]) and the **direction change** between two points in time describing the smoothness of movement. A direction change (ranging from 0° to 180°, left and right) is calculated after a distance d is reached (here, $d = 20$m), i.e., a direction change is the difference between the current direction and the direction measured d meters earlier on the movement path. (We also experimented with calculating the direction changes after a configurable time interval, but this method led to unwanted artifacts related to velocity.)

B. Movement Features and Opportunistic Forwarding

The challenge in opportunistic networks is to find best carry-and-forward nodes for the dissemination goal (e.g., spreading information over long distances). Based on the general properties of opportunistic networking and forwarding, we use the following forwarding decision metrics and relate them to movement features: (i) connection duration, (ii) number of connection opportunities, and (iii) distance reached. Movement features will influence these decision metrics, however, a strong statement about this relation requires extended investigations. Here, we discuss some likely relations which can result in future hypothesis (which often depend on additional assumptions about average human behavior):

First, the *connection duration* is likely to increase with the movement features pause time and prevalence as well as with the meeting characteristic contact time.

Second, the *number of connection opportunities* is likely to increase with the meeting characteristic contact frequency and, when considering that people tend to crowd around places of interest (hotspots), the number of revisits can also be an indicator for a high number of connection opportunities.

Third, the *distance reached* is likely to increase with the flight length, mobility range, and average velocity. In case fast spreading of information over long distances is a design goal of a dissemination algorithm, nodes with appropriate movement features can be preferred.

[3]Note, that the sampling interval for GPS position measurements is 1s.

IV. Movement Activity Estimation

We refer to a *movement activity* as a composition of movements which can be summarized by a name corresponding to the purpose of the trip of a mobile user or entity.

The set of features C is defined as $C = \{C_1, C_2, \dots C_K\}$, where K is the number of mobility features which may vary depending on the modeling purpose. Further, we define a multi-variate feature vector $V_i = (v_{i1}, v_{i2}, \dots v_{iK})$ where each v_{ik} describes a movement feature category of C_k. An example for one C_k is the average velocity, where $v_j k$ is a speed interval (e.g., walking speed in the range of]0;2]m/s).

We use a *Naïve Bayes* approach to detect a movement activity A_j under a given feature vector V_i observed. With these definitions, Bayes' theorem can be used as follows:

$$P(A_j|V_i) = \frac{P(V_i|A_j)P(A_j)}{P(V_i)}, \text{ where}$$

$1 \leq j \leq n$ and $1 \leq i \leq m$; n is the number of different movement activities and m the number of different movement feature vectors, i.e., the combination of all categories of each feature C_k (note, that the number of features used can be one which reduces the vector to a scalar value).

Once this probability is calculated for all different movement activities A_j, the activity with highest probability can be selected as classification output. In our study, we used the Weka [7] implementation of a Naïve Bayes classifier which showed reasonable good results in comparison to other classifiers provided by Weka, i.e., C4.5 decision tree, support vector, and Bayes Network classifiers.

V. Purpose and Setup of Experiments

In order to demonstrate that activities can be classified along the eight features introduced previously, we performed experiments and evaluated them along the movement activity classification success rate. The purpose of the experiments is, (i) to describe the potential of the classification approach, (ii) to investigate whether the interval configurations of the movement features influences the classification outcome, and (iii) to investigate whether some movement features are more important for the classification outcome than others.

A. Gathering and Preprocessing Data

In the study, 13 test persons were voluntarily tracking daily trips over one month by carrying small GPS receivers with them mostly in Vienna or while sightseeing in other European cities. The participants deliberately decided which trips they want to track and which they want to omit from the study data. The GPS traces are accomplished by notes taken by each person to give semantic information about the trips (labels). This self-reported information includes trip category, mode of transport (walking, running, bike, type of public transport, car), and start/end locations.[4]

In urban areas, GPS signals are disturbed by obstacles such as buildings, trees, and tunnels, which leads to outliers in measurements. Hence, GPS measurement outliers were eliminated by applying a sliding outlier detection method based on velocity values using the Local Outlier Factor (LOF) algorithm [1]. Hereby, a sliding window size of 60 GPS positions has been chosen to take different modes of transport into account (walking, bike, car).

B. Movement Activities Considered

Four different typical movement use case activities are used due to their importance expressed by the number of such trips in our data set:

- **Way to Work.** These trips describe typical movements from home to the workplace and vice versa using various means of transport (public transport, car, walking, etc.) – number of trips: 151.
- **Shopping Activity.** These trips are characterized by a situation where a person starts the trip at home or at the office, visits several shops, and returns – number of trips: 29.
- **Evening Activity.** This activity is quite similar to the reverse "Way to Work" trip in terms of usage of modes of transports, mobility range, etc. However, this movement also includes visits (pauses) at theaters, restaurants, or bars and ends likely late in the evening – number of trips: 41.
- **Tourist Activity.** This movement activity usually describes pedestrian movements in an urban area. The moves correspond to the purpose of sightseeing in a city and taking pictures along the way – number of trips: 31.

Apart from these activities resulting in overall 252 trips, trip categories resulting in only a small number of representatives were reported, which are, thus, not included in the evaluation. For instance, the entire data set contains also 11 sport, 13 visit, 9 walk, and 7 consultation trips.

C. Movement Features Observed

The measured movement features are described along their Empirical Cumulative Density Functions (ECDFs) based the set of 252 experimental trips. Note, that some of the ECDF diagrams show only excerpts depicting the important information of the entire diagram. The following remarkable observations can be stated (see also Figure 1 and 2):

Velocity: Depending on the mode of transport, different velocity distributions are observed. *Tourist* trips show the steepest curve in the lowest (pedestrian) velocities, while the long-distance car trips of this activity type result in a large fraction of high vehicular velocities. *Evening Activity* and *Way to Work* trips both show peaks at about 15 m/s representing non-pedestrian movements.

Direction Changes: The occurrences of direction changes are quite similar among all activity types. While the *Tourist* trips contain a slightly higher number of small direction changes (below 40°), the ECDFs for the other activity types vary marginally but reflect also pedestrian and vehicular movement.

Pause Times: The shortest pause times occur in *Way to Work* trips. 90% of the pause times lie below 67 seconds. The *Evening Activity* trips contain the highest pause time values due to, e.g., restaurant visits (the upper 2% lie above 4 000 seconds).

Flight Length: The *Tourist* trips contain a high number (80%) of short flight lengths ranging up to 230 meters which corresponds to the behavior of tourists walking between sights but also some higher values due to the vehicular cases. Trips of the type *Way to Work* and *Evening Activity* exhibit higher flight length values resulting from using also public means of transport and cars.

Revisiting Positions: The frequency of position revisits allows to describe loops in pathways numerically. Naturally, the number of revisits per position is generally low in the *Way to Work* trips. The *Shopping* ECDF shows a high number of revisited positions leading to the observation that people take most of the time a similar path back within one trip.

Time between Revisits: In *Way to Work* trips the occurrences are mostly (99%) below 160 seconds. Such very short periods between visits of the same position may occur when, e.g., walking back on a street section after getting off a bus. In *Shopping*, 90% amount to less than 1 300 seconds,

[4]To provide anonymity of the participants, we only use and publish aggregated data.

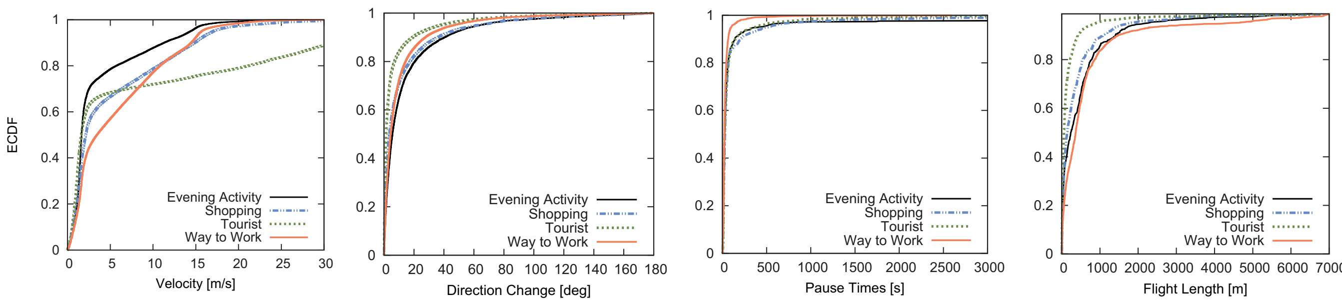

Figure 1. Empirical CDFs of (a) velocity, (b) direction change, (c) pause time, and (d) flight length occurrences.

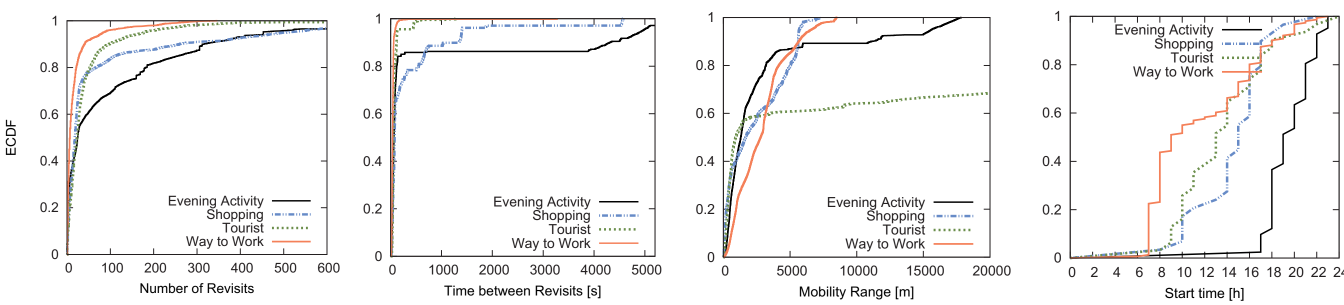

Figure 2. Empirical CDFs of occurrences of (a) number of revisits per position, (b) times between revisits, (c) mobility range, and (d) start time.

while a small number of positions is revisited after more than one hour (up to $4\,500$ seconds). *Evening Activity* shows the highest times between revisits due to pause times at popular places and taking similar routes when leaving these places.

Mobility Range: Smaller upper bounds for mobility range values can be observed in *Shopping* and *Way to Work* trips, although a higher fraction of small distance values can be found in *Shopping* trips. The *Tourist* trips are significantly influenced by long-distance car travels.

Start Time: As expected, the start time occurrences differ significantly for the use cases. 55% of the *Way to Work* trips occur between 6:00 and 10:00, whereas for the second bunch of work trips the start time ranges from 14:00 to 23:00 (supposably, most trips from work to home). The *Evening Activity* trips start as one would expect between 17:00 and 23:00. The curve for *Tourist* trips show steeper ascents in the afternoon, while the *Shopping* curve starts to ascent earlier in the morning growing continuously over the day.

VI. Movement Activity Estimation Results

The classification results are derived by applying the Naïve Bayes Classifier of the open source machine learning library WEKA [7]. The classifier is applied to all GPS trips according to the eight features modeled as a multi-variate vector as introduced in Sections III-A and IV. The trips are classified into the sample movement activities: *Evening Activity*, *Shopping*, *Tourist*, and *Way to Work*.

First, the data set has been divided into a training and a test data set by randomly selecting approx. 50% of the traces in each movement activity type for building a training set and using the remaining data for testing. After training the multi-variate Bayes classifier, it was used to classify the trips of the test set.

A. Defining Categories for Movement Features

Previous to classification into activities, meaningful interval categories were defined for each movement feature (four to five categories). Hereby, meaningful means that we included intuitive and observed knowledge about the data set. (Table I summarizes the intervals per category.)

For example, by defining different categories of velocity, we separate the value range into categories of pedestrian movement, i.e., walking (velocity v of about $0-2$m/s), fast walking or running motion or slowly driving vehicles (v of about $2-8$m/s), moderate driving on city streets (v of about $8-20$m/s), and fast speed vehicles such as trains or cars on a highway ($v > 20$m/s). For the other categories, similar semantics have been considered.

B. Results of the Multi-Variate Classification

The classification using the four to five meaningful categories introduced in Section VI-A resulted in an encouraging overall classification success rate (aka *true positive rate*) of 80.65%. The details about the assignment of trips to each activity label after classification are given in Table II.

The number of trips classified wrongly expose similarities between different movement activities. By qualitatively investigating the experimental data set, we found examples and

Feature	Interval categories				
Velocity [m/s]	[0-2[	[2-8[	[8-20[	≥ 20	
Pause time [s]	[0-180[	[180-600[	[600-1800[	[1800-3600[	≥ 3600
Flight length [m]	[0-300[	[300-600[	[600-1000[	≥ 1000	
Mobility range [m]	[0-300[	[300-1000[	[1000-2000[	≥ 2000	
Direction change [°]	[0-45[	[45-90[	[90-135[	[135-180]	
Number of revisits	[0-5[	[5-10[	[10-20[	[20-40[	≥ 40
Time betw. revisits [s]	[0-300[	[300-1800[	[1800-3600[	≥ 3600	
Start time	[0:00-6:00[	[6:00-12:00[	[12:00-17:00[	[17:00-24:00[	

Table I
INTERVAL CATEGORIES FOR EACH MOVEMENT FEATURE.

	Assigned label			
	A	B	C	D
Evening (A)	16	0	1	3
Shopping (B)	0	8	2	4
Tourist (C)	0	6	8	1
Way to Work (D)	4	0	3	68

Table II
CLASSIFICATION MATRIX OF THE FOUR DIFFERENT MOVEMENT ACTIVITY TYPES.

	Number of features (n)						
	7	6	5	4	3	2	1
Mean	79.74	77.71	76.40	73.76	71.39	67.68	62.30
Stdv	1.32	1.95	2.63	2.93	4.07	4.71	3.65

Table III
MEAN SUCCESS RATE [%] AND STADARD DEVIATION FOR ALL COMBINATIONS OF FEATURE VECTORS OF DIMENSION n.

explanations of wrong classifications. For example, shorter pedestrian *Tourist* trips might be classified as *Shopping* movement activity, while *Way to Work* trips might be classified as *Evening Activity* if taking place at a late hour with pause times.

It can be observed that trips of categories containing a larger number of data instances reach a higher classification success rate, such as *Way to Work* and *Evening Activity* (true positive rate of 90.67% and 80%, respectively). For the movement activities with lower classification success rates, we expect improvements by adding more data instances.

Additionally, to have a closer look at the false positives in the movement activities with a lower classification success rate, we split the data sets for Shopping and Tourist activities into pedestrian and vehicular trips. Most of the *vehicular Shopping* trips are classified incorrectly because they show too many similarities with *Way to Work* trips. Dividing these problematic trips into vehicular and non-vehicular subclasses results in a slightly higher overall classification success rate of 81.30%.

Since we selected the interval categories heuristically, we investigate the impact of interval settings to the classification results. For comparison, we used different interval settings, dividing the parameter range into s intervals of equal sizes. When splitting the value range of every feature into $s = 20$ interval categories, for instance, the rate of correctly classified decreases to 66.39%, while for $s = 10$ categories the rate is 68.03%. Hence, it can be concluded that the way the intervals are selected has a significant impact on the classification success rate. The semantically rich interval definition showed significantly better results than the experiments with interval categories of just equal-sized intervals (fewer intervals showed better results). In addition to this influencing factor, our results depend on the selected activity use cases in the data set. However, when using the same interval settings, the classification approach can be easily applied to an arbitrary set of activity classes.

C. *Fitting Set of Movement Features*

In an additional evaluation step we investigated whether we can reach a similar classification rate (80.69%) when using less features or if some movement features impair correct classification. A smaller number of features in the multi-variate classification vector is further beneficial from a computational point of view by reducing the computational cost.

Therefore, the trips have been classified again by leaving one to seven criteria out leading to combinations of seven to only one single features. The resulting average classification success rates (and standard deviations) are given in Table III. When omitting features during classification, the mean success rate is decreasing with every reduction step from 79.74% to 62.30%, while the standard deviation is increasing. During the detailed observation, we see that still best success rates were achieved in combinations where *number of revisits* was included (followed by *start time*). As a result of our observation, these two features can be considered as important features to differentiate between the movement activities included in the study.

VII. CONCLUSION

We presented and investigated a method to classify human movement trips along movement features using a Naïve Bayes classifier. The classifier has been trained with real-trip GPS data along the movement features of the GPS traces: velocity, direction change, pause time, flight length, number of revisited positions, time between revisits, mobility range, and start time.

We used four different movement activity use cases and classified the recorded GPS trips according to these activities. We achieved an overall classification success rate of up to 80.65%, which is a promising result. The approach is in the sense general that it can be applied to arbitrary movement activities and arbitrary movement features and feature interval settings. However, the results depend on the use cases, features considered and feature interval settings (and the aggregation method). Our observations showed that varying the interval settings, both in terms of interval numbers and arranging of intervals, influences the classification success rate significantly: interval settings based on semantic knowledge (heuristics) yield better results than simple, equal-sized intervals. When searching for the most important movement features for classification, the number of position revisits and the start date of a movement activity were most important.

In future work, we plan to enhance the data set (GPS trips) and to state and evaluate hypotheses how the movement features and movement activities influence opportunistic networking metrics.

References

[1] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. LOF: Identifying Density-based Local Outliers. *SIGMOD Record*, 29(2):93–104, 2000.

[2] M. K. Bromage, J. T. Koshimoto, and K. Obraczka. TAROT: Trajectory-Assisted Routing for Intermittently Connected Networks. In *CHANTS '09: 4th ACM Workshop on Challenged Networks*, pages 9–18, 2009.

[3] B. Burns, O. Brock, and B. Levine. MV Routing and Capacity Building in Disruption Tolerant Networks. In *INFOCOM '05. IEEE Int. Conf. on Computer Communications*, 2005.

[4] D. Choujaa and N. Dulay. Predicting Human Behaviour from Selected Mobile Phone Data Points. In *Ubicomp '10: 12th Int. Conf. on Ubiquitous Computing*. ACM, 2010.

[5] F. Ekman, A. Keränen, J. Karvo, and J. Ott. Working Day Movement Model. In *MobilityModels '08: ACM SIGMOBILE Workshop on Mobility Models*, pages 33–40, 2008.

[6] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding Individual Human Mobility Patterns. *Nature*, 453(7196):779–782, 2008.

[7] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten. The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11(1), 2009.

[8] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot. Pocket Switched Networks and the Consequences of Human Mobility in Conference Environments. In *SIGCOMM 2005 Workshop on Delay Tolerant Networking*, 2005.

[9] K. A. Hummel and A. Hess. Movement Activity Estimation for Opportunistic Networking Based on Urban Mobility Traces. In *Wireless Days (WD), 2010 3rd IFIP*, 2010.

[10] W. jen Hsu, T. Spyropoulos, K. Psounis, and A. Helmy. Modeling Time-Variant User Mobility in Wireless Mobile Networks. In *INFOCOM '07: IEEE Int. Conf. on Computer Communications*, pages 758–766, 2007.

[11] A. Jindal and K. Psounis. Fundamental Mobility Properties for Realistic Performance Analysis of Intermittently Connected Mobile Networks. In *IEEE Int. Conf. on Pervasive Computing and Communications (PerCom) Workshops*, 2007.

[12] J. Kim and A. Helmy. The Challenges of Accurate Mobility Prediction for Ultra Mobile Users. *ACM SIGMOBILE Mobile Computing and Communications Review*, 13(3):58–61, 2009.

[13] M. Kim, D. Kotz, and S. Kim. Extracting a Mobility Model from Real User Traces. In *INFOCOM '06: IEEE Int. Conf. on Computer Communications*, pages 1–13, 2006.

[14] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong. SLAW: A Mobility Model for Human Walks. In *INFOCOM '09: IEEE Int. Conf. on Computer Communications*, 2009.

[15] J. Leguay, T. Friedman, and V. Conan. Evaluating Mobility Pattern Space Routing for DTNs. In *INFOCOM 2006. IEEE Int. Conf. on Computer Communications*, pages 1–10, 2006.

[16] L. Liao, D. Fox, and H. Kautz. Extracting Places and Activities from GPS Traces Using Hierarchical Conditional Random Fields. *Int. Journal of Robotics Research*, 26(1):119–134, 2007.

[17] A. Lindgren, A. Doria, and O. Schelén. Probabilistic Routing in Intermittently Connected Networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 7(3):19–20, 2003.

[18] L. Pelusi, A. Passarella, and M. Conti. Opportunistic Networking: Data Forwarding in Disconnected Mobile Ad Hoc Networks. *IEEE Communications Magazine*, 2006.

[19] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava. Using Mobile Phones to Determine Transportation Modes. *ACM Trans. Sen. Netw.*, 6(2):1–27, 2010.

[20] V. Srinivasan, M. Motani, and W. T. Ooi. Analysis and Implications of Student Contact Patterns Derived from Campus Schedules. In *MobiCom '06: Int. Conf. on Mobile Computing and Networking*, 2006.

[21] C. Tuduce and T. Gross. A Mobility Model Based on WLAN Traces and its Validation. In *INFOCOM '05: IEEE Int. Conf. on Computer Communications*, pages 664–674, 2005.

[22] H. Wang, F. Calabrese, G. DiLorenzo, and C. Ratti. Transportation Mode Inference from Anonymized and Aggregated Mobile Phone Call Detail Records. In *IEEE Int. Conf. on Intelligent Transportation Systems*, 2010.