Peacox – Persuasive Advisor for CO2-reducing cross-modal trip planning

Project Reference: 288466

FP7-ICT 2011: 6.6 Low carbon multi-modal mobility and freight transport)

Project Duration: 1 Oct 2011 – 30 Sept 2014

D 4.2

Implementation and pretest of the trip purpose detection

ETHZ

Author(s)

Lara Montini (ETHZ)

Nadine Rieser (ETHZ)

Date: 28/02/2014

Dissemination level: (PU, PP, RE, CO): PU

Project Co-Funded by the European Commission within the 7th Framework Program

Table of Contents

1. Introduction

This deliverable provides an overview over the current status of the work in Work Package 4 "Automated Travel Mode and Trip Purpose Detection". The deliverable is focused on Task 4.3 Development and implementation of trip purpose imputation. The goal of this task is to implement the detection of trip purposes using temporal and spatial patterns observed in the GPS data. That includes timings and durations of activities and frequencies of visited places.

In this deliverable we introduce trip purpose identification using random forests, a machine-learning approach based on decision trees. Analysis is done using GPS and accelerometer data collected by 156 participants, taking part in a one-week travel survey in Switzerland completed in 2012.

Results show that random forests provide robust trip purpose classification. For ensemble runs, a share of correct predictions between 80 and 85 % was achieved. Different setups of the classifier are possible, and sometimes required by the application context. The classifier's training set and its input variables (feature set) are defined in various ways. Four relevant setups are tested here.

2. Related Work

Two main groups of trip purpose imputation routines can be found in the literature:

- rule-based systems: they rely mostly on the position of the activity, its timing, and GIS data on land uses.
- machine learning approaches: there the focus is more on the activity itself and less on position.

One of the first rule-based trip purpose systems introduced by Wolf et al. (2001) is using land use data to narrow down possibilities to three purposes. In a second step, duration and time of arrival are used to refine classification. Overall, 67 % of trips were correctly identified. Most other rule-based approaches (e.g. Moiseeva et al., 2010; Bohte and Maat, 2009; Stopher et al., 2008) base the imputation solely on the location. Trip purpose is assigned by comparing the activity location to land use databases, points of interest and addresses provided by participants before or during the travel survey. In Wolf et al. (2004) trip purpose is imputed for clusters of activities. They included socio-demographics in the decision process, which was deterministic, but a probabilistic approach along the lines of a Bayesian network was proposed.

The machine learning approaches are mostly decision trees (e.g. Lu et al., 2012; Deng and Ji, 2010; Griffin and Huang, 2005). They are mostly based on variables computed for single activities. Deng and Ji (2010) report accuracies between 70 and 96 % on a rather homogeneous set of participants and Lu et al. (2012) achieve 60 to 73 % accuracies depending on the trip purpose. Liao et al. (2007) achieve good results (80 - 85 % accuracy) using hierarchical conditional random fields. When comparing accuracies of the reported classifications, it has to be considered that the level of detail for trip purposes differ and that the data sets used for classification are also very different in size as well as in the homogeneity of participants.

Here, random forests (Breiman, 2001), another machine learning algorithm, are used for trip purpose classification. Random forests have been successfully applied in different transport related classification problems (e.g. Ali et al., 2012; Greenhalgh and Mirmehdi, 2012; Rodrigues et al., 2012; Morerira et al., 2005). Similarly to the problem at hand, Wu et al. (2011) use random forests to classify activities based on GPS data. They distinguish between indoor, outdoor static, outdoor walking and in-vehicle travel. Stenneth et al. (2011) use a random forest classifier to predict transport mode from GPS and accelerometer tracks, overall more than 93 % of observations were correctly predicted.

The remainder of the deliverable is structured as follows. First, the employed method, including tuning of parameters, and the available data are described. In the subsequent results section, a base setup as well as some comparison setups are reported on. Some conclusions and an outlook on further research round off the deliverable.

3. Method

The method described in this section is based on multi-day GPS and accelerometer data for survey respondents living in the same region. To exploit the multi-day nature of the data, activities are clustered into locations. Clustering is done for single persons but also for the complete set of activities as several respondents might frequent the same public locations. Classification variables, called features, can then be derived for location clusters. To impute trip purpose the following steps are performed:

1. extraction of activities and their locations,
2. clustering of locations,
3. computation of features, and
4. learning and applying the classifier.

First, the data set used for classification is described, followed by the calculation of activity locations and their clustering. Next, an overview of the type of used features and how they are selected is given. Finally, a description of the chosen classification method, random forests, concludes the section.

3.1 GPS Data Set

The data set used for evaluation was collected in and around the city of Zurich in Switzerland and is described in more detail in Montini et al. (2013). Each of the 156 respondents collected approximately one week of second-by-second GPS and accelerometer data using a dedicated GPS device. Respondents were randomly selected from an address pool and they are also reasonably representative for the study area. Respondents were asked to correct an automatically generated travel diary including transport mode and trip purpose on the survey homepage. As the quality of these corrections varies, all were double checked by the survey team. Since the participants were asked for their home and work addresses, the trip purposes being home, working and mode transfer could mostly be imputed by survey personnel. All other purposes are only available if respondents filled in the diaries. The very first activity of the survey period is removed for each participant, because – due to the cold start problems and unclear first handling of the device – the GPS signal can be very far off the actual location.

In total, 6938 valid activities including a 28.54 % share of mode transfers were reported. The other surveyed purposes are being home (22.33 %), working (11.82 %), shopping and services (9.61 %), recreational activities (9.77 %), picking up or dropping off s.o. (2.77 %), business, that is, work-related activities outside work place, (3.32 %) and finally other activities (2.62 %). The remaining 9.22 % of activities have no reported type. Therefore, this data was left out when training and testing the classifier. For the clustering on the other hand, this data are used, as even if the type is unknown, the location and duration provide additional information.

3.2 Activity Location Calculation

Activities are specified at least by start and end times and the GPS and accelerometer points recorded during this interval. Deriving these activities is either done by survey respondents or by an automatic activity detection module of a GPS processing framework. To assign each activity a location in space a representative coordinate is calculated from all GPS coordinates during the activity. Coordinates in our application are given in the metric National Swiss Grid representation. Using a metric grid is important when aggregating coordinates. To compare different aggregation approaches, only home activities were used for testing. Home locations are most suitable as they are known for all participants and they are visited several times within the survey period, which is important for subsequent clustering. Three approaches were tested:

- the unweighted point cloud centroid (which is the mean position of all points),
- median x and y coordinates (calculated separately) and,
- the coordinate with highest density (density was computed as number of coordinates within a radius of 20 meters around the candidate coordinate).

For comparison, distances between the calculated aggregated coordinates and the geocoded home coordinates were computed. Representations were categorized as wrong if the distance was higher than a given threshold. Figure 1 shows the erroneous coordinates for different distance thresholds. It can be seen that the centroid performs worst, the median coordinate is already much better, possibly because the influence of outliers is reduced. The best representation is achieved by the coordinate with the highest density. In this case for a threshold of 50 meters more than 90 % and for 100 meters already 95 % of home locations are well represented. Therefore, in the subsequent clustering, the coordinate with the highest density was used.

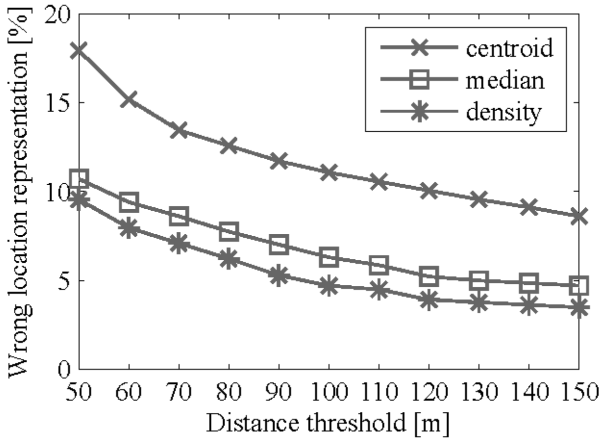


Figure 1 Comparison of centroid, median coordinate and densest coordinate as representation for an activity location.

3.3 Clustering

The total number of locations visited by respondents is not known. Therefore, clustering has to allow for as many clusters as needed. A well-established approach that fulfills this condition is hierarchical clustering (Hastie et al., 2009). Agglomerative hierarchical clustering starts with each activity as one cluster. In each step the pair of clusters that is closest is merged into a new cluster. Clusters are merged as long as they are closer than a given cutoff distance. If two clusters consist of several activity locations, the distance between clusters is not straight forward. Thus, different linkage criteria were tested:

- single linkage: the distance between the nearest possible points of two clusters
- complete linkage: the maximum distance between two points of two clusters
- average linkage: the average of all distances of all possible point pairs of two clusters

To determine optimal clustering, a location should be represented by one cluster and this cluster should only consist of activities of this location. These two outcomes were computed for all home locations, in order to compare the different linkage criteria. Home locations were used, because they are most reliably known and they occur more than once per person. For each person in the data set the following procedure is executed:

1. cluster all available activities
2. count how many clusters contain home activities
3. the cluster with most home activities is assigned to be the main home cluster
4. count how many different locations are part of this main home cluster

For different cutoff distances, the mean values for all persons are calculated and plotted in Figure 2. The optimum would be at 1 for both, the number of clusters per home locations and the number of locations per main cluster. None of the clustering runs reaches the optimum value for either measure and it is unclear which of the two measures is more important. Overall, all linkage criteria perform similarly. Average linkage creates more compact clusters than single linkage, but is not as restrictive as complete linkage (Hastie et al., 2009). Therefore, average linkage was chosen for all subsequent analysis.

To decide on the best cutoff distance for the average clustering, the complete classification system was run with different cutoff distances. For each case 100 different forests were learned, the resulting mean accuracy as well as the 5th and 95th percentile are shown in Figure 5. The best results in terms of accuracy were achieved with a cutoff distance of 100 meters. Therefore, this cutoff distance threshold is used in subsequent analyses.

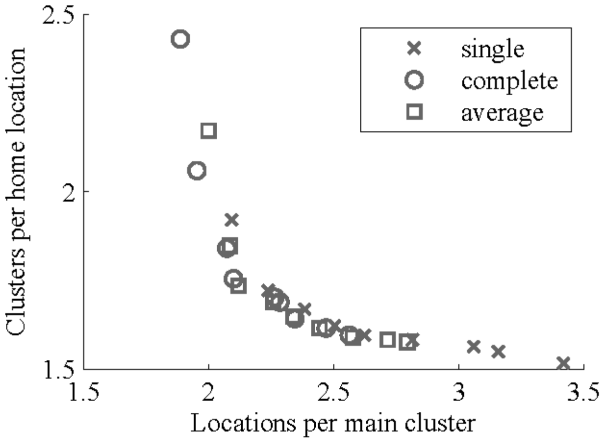


Figure 2 Pareto efficiency of the mean number of clusters that cover home activities and the mean number of different locations within the main home cluster for several cutoff distances.

3.4 Feature Computation

Features we use for trip purpose imputation can be divided into three groups: specific to persons, activities and clusters (per person as well as per data set).

Appropriate person-based data are usually collected in the socio-demographic questionnaire of a travel survey. Some examples are age, education level, income and mobility tool ownership. Moreover, the home address of respondents is often known, and sometimes even more addresses such as work place or favorite shopping centers are surveyed as well. For applications other than travel surveys, such data are probably optional and more sparse.
Simple activity-centered features are, for example, duration and start time. Using GPS data and the derived location representation, distances to important places such as home or work can be calculated. If a travel diary is available or can be automatically generated, the

modes used to get to and leaving the activity can also be used. Further, during walking a very distinct accelerometer pattern can be observed. Therefore, the percentage that is spent walking during an activity can be computed. The underlying idea is that during shopping one might walk more than at home.

Cluster-specific features are associated with the location and not with single activities. For clusters that were computed for single persons, examples are statistical aggregates of the activity features such as mean and standard deviation of the duration. Clusters can also be created for the complete data set, which allows to extract the number of persons knowing a location. However, this feature has to be treated carefully as it is data set dependent, e.g. for a very homogeneous group working at the same university, work is a place everybody knows, for a more diverse data set of a region, train stations are more likely to be known to several people.

3.4.1 *Feature Selection*

An importance measure was used as basis for the selection of the features best suited for trip purpose imputation. Furthermore, it was ensured, that all feature-groups were represented. When building a random forest, the feature importance can be directly computed from the out-of-bag observations, that is the observations that were not used to construct a tree. The main idea is, that a feature is more important if the change of its value in an observation causes a misclassification. Therefore, for each feature the increase of misclassification is determined when the value of this feature is permuted in the out-of-bag observations. Hastie et al. (2009) show that this importance measure shows reasonable ranking but the distribution of the importance tends to be spread more uniformly than with other importance measures.

For the problem at hand, around 40 potential features were tested. The seventeen selected features are listed in Figure 4(a), where features of the same group are depicted in the same color. Of those seventeen features the following are categorical features and treated as such in the classification process: day of week, education level, marital status and the transport modes. All other variables are continuous. Features that were not selected on the basis of the out-of-bag feature importance are mainly person-based features. Namely, all items related to mobility tools like possession of a driver’s license or public transport season ticket were at the bottom of the importance ranking. The most important features are the activity-based features. The flag for weekday or weekend and the day of week, were not among the most important, but they are always available and reliable. Therefore, one of them, the slightly better evaluated day of week, was kept. Contrary to that, trip duration before the activity is left out, because feature importance was low and it's computation is costly and the value is uncertain, especially in the context of uncorrected travel diaries. The modes before and after the activity, having the same drawbacks, are used anyway, as they are evaluated to be much more important.

3.5 Classifier: Random Forests

For trip purpose imputation, decision trees, that is a set of rules learned by a machine and executed in a given order, were already used with good success (e.g. Griffin and Huang, 2005; Deng and Ji, 2010; Lu et al., 2012). Using a random forest, that is a set of decision trees also called ensemble of trees, is therefore, a natural step. Random forests were introduced by, and is a trademark of Breiman (2001).The underlying concept is comprehensible, and more importantly it performs well in a variety of problems. They are also very popular, as they are easy to train and tune (Hastie et al., 2009). Breiman (2001) showed that random forest do not overfit even if more trees are added. A further advantage is that good results can be maintained even if data are missing, as they are estimated internally (Breiman and Cutler, 2013).

Technically, random forests work as follows. Each decision tree in the ensemble has one vote that counts for classification. The class with most votes, is the classification result. In a regular decision tree a data set is split using the feature that results in the best split. Using the same data to learn a tree, results in the same tree. But, in a random forest different votes are needed, and correlation between trees should be reduced to obtain best classification. To achieve that, on the one hand, each tree is learned from a different subset of the training data. On the other hand, at each split in the tree a random subset of features is considered. Each tree is fully learned, that is splits will be created until all training data are correctly classified. For the results presented in this deliverable the random forest implementation of Matlab (MathWorks, 2012), called TreeBagger, was used. For the PEACOX application, it was tested to use this MATLAB functionality from within the Java framework, unfortunately the performance is very bad. Therefore the trip purpose detection implemented as part of the processing framework POSDAP (2012), is based on the random forest implementation of the WEKA project (Hall et al., 2009).

3.5.1 *Tuning and Stability Analysis*

The main parameter that can be used for tuning random forests is the number of features (m) that should be randomly selected for each split decision. Figure 3 shows the out-of-bag error for a run using 17 features in total. It can be seen, that performance is similar for up to 7 randomly selected features at each split and then starts to drop slightly, as correlation between the trees starts to increase. For all runs with seventeen features $m = 7$ is selected. The recommended default value is the ceiling value of the square root of the number of features, which is $m = 5$ in the case of 17 features, as the default value performs well, all runs with less features use their respective default, which is $m = 4$. Furthermore, for all runs 500 trees are learned per classifier, as it can be seen that the error stays stable even if more trees are added.

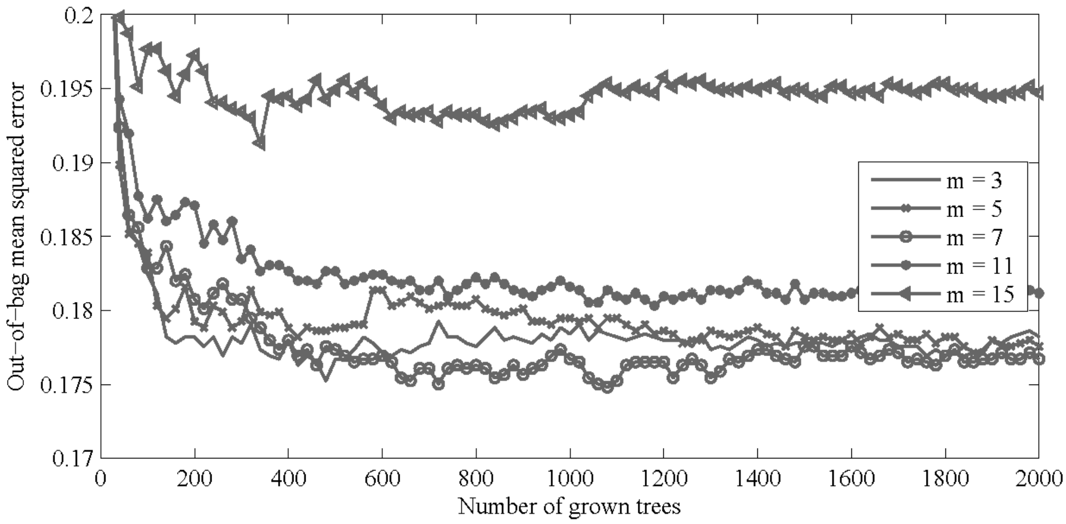


Figure 3 Out-of-bag error for different numbers of randomly selected features (m).

4. Results

The tuning parameters of the classification procedure were discussed in the previous section. In this section, we analyze the accuracy and performance of the classifier. Therefore, different situations are compared:

1. A base situation for comparison is learned on a random subset of activities with 17 input features.
2. The features set is reduced, and changes in the classifier performance are discussed.
3. Activity-based classification is compared to location-based classification.
4. Finally, the performance per person is analyzed.

4.1 Base Setup

For the base setup, random forests are trained on randomly selected 75 % of all activities per activity type. The remaining data are used to test the performance. In total, 100 runs with 500 trees were employed. The mean accuracy, that is the share of correctly classified observations of these runs is 82.3 %. The highest accuracy reached is 84.4 %. The respective confusion matrix is reported in Table 1 together with the recall and precision values of each trip purpose. Recall and precision are defined per trip purpose as follows:

$$\text{recall (trip purpose)} = \frac{\text{correct classifications of a trip purpose}}{\text{all actual activities with this trip purpose}}$$

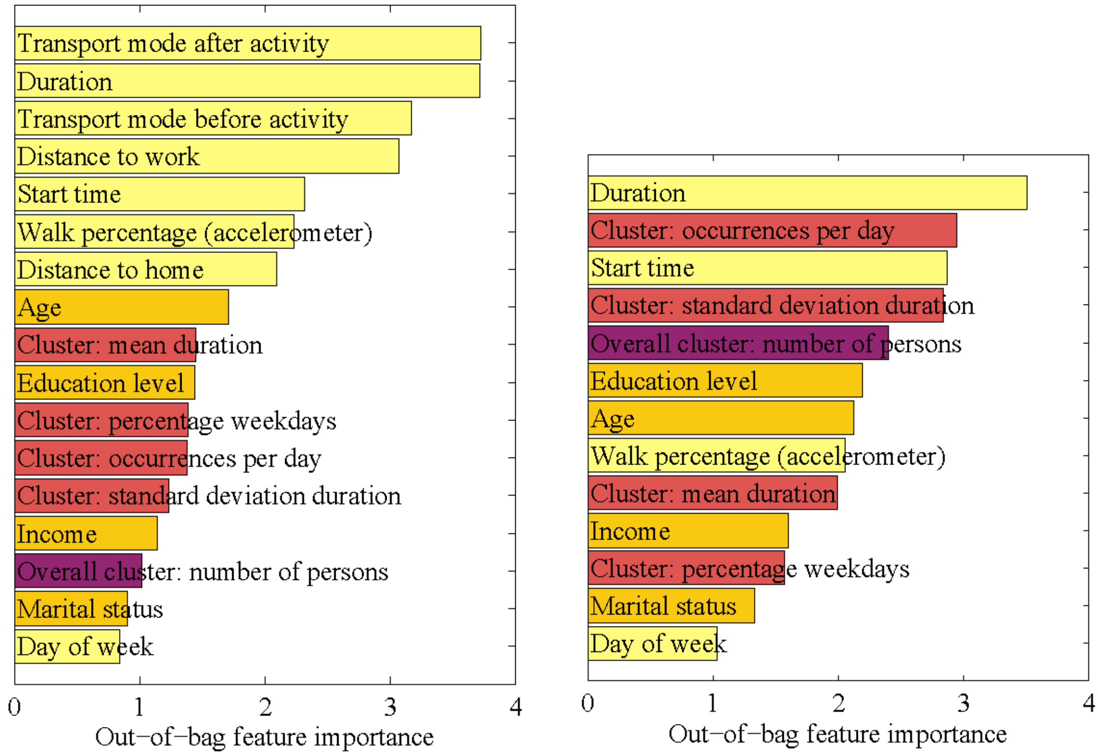
$$\text{precision (trip purpose)} = \frac{\text{correct classifications of a trip purpose}}{\text{all activities predicted to have that trip purpose}}$$

It can be seen that mode transfers, being at home and work / education are recalled best. Almost all mode transfers (99 %) are specified as such and being home is recalled in 97 % of occurrences. Importantly, also the precision of the classifier for these three classes is good with around 85 % for working, 91 % for mode transfers and 93 % for being home. Shopping and recreational activities are reasonably recalled with 74 % and 68 % respectively. The remaining trip purposes (pick-up / drop-off, business and other), that do not occur that often, are not well recalled (between 40 and 20 %) but there are also not too many misclassifications as pick-up / drop-off or business, as shown by high precision values of 83 %.

Table 1 Confusion matrix: random forest with 500 trees for two different feature sets, in each case the run with highest accuracy out of 100 runs is reported

		Prediction								
		Mode transfer	Home	Work / education	Shopping / service	Recreation	Pick-up / drop-off	Business	Other	Recall
17 features. Overall accuracy 84.4 %.										
Truth	Mode transfer	490	1	1	2	0	1	0	0	**99.0 %**
	Home	5	374	4	1	2	1	0	0	**96.6 %**
	Work / education	8	5	177	6	8	0	1	0	**86.3 %**
	Shopping / service	13	3	3	124	19	2	1	2	**74.3 %**
	Recreation	10	11	7	25	115	0	0	1	**68.0 %**
	Pick-up / drop-off	6	3	1	14	4	19	0	1	**39.6 %**
	Business	7	2	11	9	9	0	19	0	**33.3 %**
	Other	4	1	0	19	13	1	0	7	**20.0 %**
	Precision	**90.6 %**	**93.3 %**	**85.1 %**	**62.6 %**	**69.3 %**	**82.6 %**	**82.6 %**	**69.2 %**	
13 features, leaving out potentially hard to obtain features. Overall accuracy 79.8 %.										
Truth	Mode transfer	479	3	0	9	4	0	0	0	**96.8 %**
	Home	8	369	5	1	4	0	0	0	**95.3 %**
	Work / education	10	8	177	1	7	1	1	0	**86.3 %**
	Shopping / service	58	2	7	74	18	2	5	1	**44.3 %**
	Recreation	17	6	5	19	121	0	0	1	**71.6 %**
	Pick-up / drop-off	21	2	1	6	5	10	2	1	**20.8 %**
	Business	14	1	9	3	8	0	22	0	**38.6 %**
	Other	16	0	5	5	14	1	1	3	**6.7 %**
	Precision	**76.9 %**	**94.4 %**	**84.7 %**	**62.7 %**	**66.9 %**	**71.4 %**	**71.0 %**	**50.0 %**	

This base setup uses 17 features in total. In order to understand which features are important to distinguish trip purposes, their out-of-bag importance is shown in Figure 4(a). The 7 most important features are all activity-based (light yellow colored bars). The cluster or location-specific features (darker red colors) and the person-based features are slightly less important.



(a) 17 features **(b) 13 features**

Figure 4 Mean feature importance. Color coding for different feature types (person, activity, cluster and overall cluster features).

4.2 Performance with a Reduced Feature Set

Among the most important activity-based features are the transport modes used before and after the activity. In the context of travel surveys, when data have been corrected by participants this data are probably of high quality and it is beneficial to use it. Yet, they are not taken over into the reduced set, because in order to get good classification results reliable and accurate input data is needed, which is not always the case for transport modes. Particularly problematic can be cases where the mode is imputed and not double-checked by participants. In addition, distances to home and work are left out of the reduced set, in order to analyze, if these locations can also be imputed when their location is not known beforehand. Mean accuracy of 100 runs was 77.2 % and therefore, lower than the accuracy of the base setup. This was expected, as important features were left out. The confusion matrix of the best run (79.8 % accuracy) is reported in Table 1. Values are generally lower than compared to the base setup but the overall tendencies are similar. Note, that home and work locations are recalled as good as in the base setup with 95 % and 86 % respectively. Therefore, it would be possible to extract these location for each person in a first step, and to use the distance to the home and work location in a second step to improve classification. Recall of shopping and services on the other hand decreases considerably to 44 %. Most misclassifications are predicted as mode transfers, which consequently has a much lower precision than in the base setup.

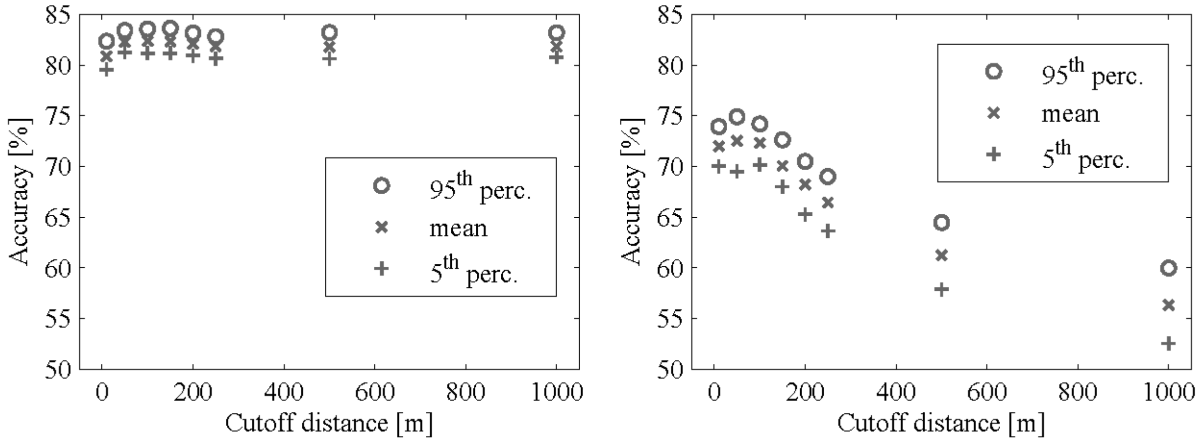
The ranking of feature importance changes slightly compared to the base setup (Figure 4(b)). Especially some cluster-based features (occurrences per day, standard deviation of the duration, and the number of persons knowing a location) gain importance.

4.3 Comparison of Activity-Based and Location-Based Identification

So far, the classification in this paper has been done based on individual activities. However, a lot of the approaches for trip purpose identification in the literature, rely only on the location of the activity, which corresponds more to a location-based identification of trip purposes. To compare these two approaches, an activity-based as a location-based random forest classifier is generated, using the same data source. This data source does not include land use data, which is typically used in location-based approaches, but it could be added to both classifiers if available. To learn the location-based, the data have to be aggregated, that is locations have to be defined, and for each location the trip purpose has to be set, and features have to be computed. The location is represented by the cluster of activities as described in Section 3.3. The trip purpose for this cluster is derived from its activities where trip purpose is known. Most probably, not all activities are of the same purpose, hence, the one that occurs most often is chosen. Naturally, all cluster-based and person-based features are used, activity-based features on the other hand cannot be computed for clusters. The exceptions are the distance to work and home, as all activities in a cluster are per definition nearby and therefore, distances to other locations are still valid.

The classifier is trained on 75 % of all clusters and classification is done for the remaining clusters. For the evaluation, all activities in a cluster are assigned the purpose imputed for this cluster, because the goal is to compare the classification of individual activities.
Figure 5 compares the results for location and activity-based clustering and different cutoff distances. The influence of the cutoff distance is much higher for location-based

identification, which is sensible, as almost all features are cluster dependent. Overall, the performance of the location-based system is worse (mean accuracy at 100 m cutoff 72.3 %).



(a) Activity-based

(b) Location-based

Figure 5 Accuracy for 100 runs with different cutoff distances for hierarchical clustering with average distance.

4.4 Classifier Performance for Different Persons

When generating an automatic travel diary for survey respondents, the classifier is probably learned beforehand using a different data set or at least data from previous respondents. To simulate this situation, a classifier with all 17 features is learned on a subset of persons instead of a subset of activities. For this person-based training the mean accuracy achieved is 78.4 % (5th percentile: 74.3 % ; 95th percentile: 82.1 %). Thus, the mean is 4 % lower than is the case for the activity-based approach (mean: 82.3 %; 5th percentile: 81.2 %; 95th percentile: 83.2 %), also the variation within the 100 runs is higher for person-based training sets. This can be explained, if the variation of accuracies achieved for different persons is high. And indeed, the mean accuracy computed per person was 78.1%, the 5th and 95th percentile are 50 % and 100 % respectively and the standard deviation is 16.6 %. One has to consider that, the quality of reporting for different respondents is very different. For some respondents, only a few and for others more than 100 valid activities were available, still the differences in prediction accuracy are uncomfortably large.

5. Conclusion and Outlook

Random forests perform well with an overall mean accuracy above 80 % for trip purpose imputation on the data set at hand, which is relatively large and very diverse with respect to respondents and purposes. Therefore, trip purpose imputation, trained on this data set, is hoped to be a good basis for classification of other data sets. This should be tested in the future. Trip purposes that occurred more often in the data set also had higher recall values. This might be because random forests tend to favor the classes that appear more frequently in the training data if no good classifier can be built (Speed, 2003). But, the two most frequent activities home as well as mode transfers were expected to give better recall values as they have more distinctive characteristics, e.g., mode transfers are typically very short and home occurs typically every night, thus they are easier to classify.

It was shown in Figure 5 that the activity-based classifier performed better than the location-based classifier. This, as well as Figure 4 is an indication that the use of activity-based features is important for good classification and therefore, these should not be neglected. An important input feature, that we did not use in neither of the two classifiers is land use data, which is usually associated with location-based imputation. We argue that inclusion of this data would improve both classifiers similarly. In the future, land use data should be considered if available, but this can be challenging depending on the application. For example if an application concerns different regions it is less likely that common land use data is found, than for a survey with clearly defined regional boundaries.

Considering applications where only GPS tracks and accelerometer data and no personal information is available, one can assume to get reasonable results with the approach presented in this paper. On the one hand, because person-specific features were less important than those computed for activities. On the other hand, concerning distance to the home and work location, it was shown, that these activities can be derived with good precision. Therefore, to impute a home or work coordinate, one could take all GPS coordinates of a person that are part of a predicted activity and then calculate the densest coordinate for those. A second classifier could then be run, that uses the distance to home and work as input feature.

For the most traditional application in travel surveys, it has to be considered that the variance of results for different persons is rather high. Therefore, when presenting survey respondents with automatic diaries, some respondents will be much more annoyed than others. This large person-dependent differences in accuracy should be investigated in more depth. One possibility to hopefully reduce this variation, is to personalize the classifier, which is the main goal of Task 4.5. The idea is to use a random forest learned on base data that is then personalized by adding trees learned on personal data.

6. References

[1] Ali, S. S. M., N. Joshi, B. George and L. Vanajakshi (2012) Application of random forest algorithm to classify vehicles detected by a multiple inductive loop system, paper presented at the 15th International IEEE Conference on Intelligent Transportation Systems (ITSC), Anchorage, September 2012.

[2] Bohte, W. and K. Maat (2009) Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands, Transportation Research Part C: Emerging Technologies, 17 (3) 285–297.

[3] Breiman, L. (2001) Random forests, Machine Learning, 45 (1) 5–32.

[4] Breiman, L. and A. Cutler (2013) Random forests, webpage, http://stat-www.berkeley.edu/users/breiman/RandomForests/cc_home.htm. Accessed on 29/07/2013.

[5] Deng, Z. and M. Ji (2010) Deriving rules for trip purpose identification from GPS travel survey data and land use data: A machine learning approach, paper presented at the 7th International Conference on Traffic and Transportation Studies, Kunming, August 2010.

[6] Greenhalgh, J. and M. Mirmehdi (2012) Traffic sign recognition using MSER and random forests, paper presented at the 20th European Signal Processing Conference (EUSIPCO 2012), Bucharest, August 2012.

[7] Griffin, T. and Y. Huang (2005) A decision tree classification model to automate trip purpose derivation, paper presented at the 18th International Conference on Computer Applications in Industry and Engineering, Honolulu, November 2005.

[8] Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten (2009) The WEKA Data Mining Software: An Update, SIGKDD Explorations, 11 (1) 10–18.

[9] Hastie, T., R. Tibshirani and J. Friedman (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2. edn., Springer Series in Statistics, Springer, New York.

[10] Liao, L., D. Fox and H. Kautz (2007) Extracting places and activities from GPS traces using hierarchical conditional random fields, International Journal of Robotics Research, 26 (1) 119–134.

[11] Lu, Y., S. Zhu and L. Zhang (2012) A machine learning approach to trip purpose imputation in GPS-based travel surveys, paper presented at the 4th Conference on Innovations in Travel Modeling, Tampa, May 2012.

[12] MathWorks (2012) MATLAB R2012b (Version 8.0), software, http://www.mathworks.com/products/matlab/.

[13] Moiseeva, A., J. Jessurun and H. J. P. Timmermans (2010) Semi-automatic imputation of activity travel diaries using GPS-traces, prompted recall and context-sensitive learning algorithms, Transportation Research Record, 2183, 60–68.

[14] Montini, L., N. Rieser-Schüssler and K. W. Axhausen (2013) Field Report: One-Week GPS-based Travel Survey in the Greater Zurich Area, paper presented at the 13th Swiss Transport Research Conference, Ascona, April 2013.

[15] Moreira, J. M., A. Jorge, J. F. d. Sousa and C. Soares (2005) Trip time prediction in mass transit companies. a machine learning approach., paper presented at the 10th Meeting of the Euro Working Group Transportation, Poznan, September 2005.

[16] POSDAP (2012) Position Data Processing, http://sourceforge.net/projects/posdap/.

[17] Rodrigues, F., F. C. Pereira, A. Alves, S. Jiang and J. Ferreira (2012) Automatic classification of points-of-interest for land-use analysis, paper presented at the 4th International Conference on Advanced Geographic Information Systems, Applications, and Services (GEOProcessing 2012), Valencia, January 2012.

[18] Speed, T. (ed.) (2003) Statistical Analysis of Gene Expression Microarray Data, Chapman & Hall/CRC, Boca Raton.

[19] Stenneth, L., O. Wolfson, P. S. Yu and B. Xu (2011) Transportation mode detection using mobile phones and GIS information, paper presented at the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, November 2011.

[20] Stopher, P. R., C. FitzGerald and J. Zhang (2008) Search for a global positioning system device to measure person travel, Transportation Research Part C: Emerging Technologies, 16 (3) 350–369.

[21] Wolf, J., R. Guensler and W. Bachman (2001) Elimination of the travel diary - experiment to derive trip purpose from Global Positioning System travel data, Transportation Research Record, 1768, 125–134.

[22] Wolf, J., S. Schönfelder, U. Samaga, M. Oliveira and K. W. Axhausen (2004) Eighty weeks of Global Positioning System traces, Transportation Research Record, 1870, 46–54.

[23] Wu, J., C. Jiang, D. Houston, D. Baker and R. Delfino (2011) Automated time activity classification based on global positioning system (GPS) tracking data, Environmental Health, 10, 101.