Sensor Fusion for Semantic Place Labeling

Roman Roor[1], Jonas Hess[2], Matteo Saveriano[2], Michael Karg[1] and Alexandra Kirsch[3]
[1]*BMW AG, Munich, Germany*
[2]*Department of Electrical and Computer Engineering, TU München, Munich, Germany*
[3]*Department of Computer Science, Eberhard Karls Universität Tübingen, Tübingen, Germany*

Keywords: Semantic Place Labeling, Context, Sensor Fusion, Feature Extraction, Classification, Machine Learning, Smartphone, Autonomous Driving Vehicles, V2V, V2X, GPS.

Abstract: In order to share knowledge about road situations vehicle-to-vehicle (V2V) communication is used. Autonomous driving vehicles are able to drive and park themself without driver interactions or presence, but are still inefficient about the drivers needs as they don't anticipate the users' behaviour. For instance, if a user wants to stop for quick grocery shopping, there is no need looking for long term parking in far distance, a short-term parking zone near the grocery shop would be adequate. To enable autonomous cars to make such decisions, they could benefit from awareness of their drivers' context. Knowledge about a users' activities and position may help to retrieve context information. To be able to describe the meaning of a visited place for user, we introduce a variant of semantic place labeling based on various sensor data. Data sourced by, e.g. smartphones or vehicles, is taken into account for gathering personalized context information, including Bluetooth, motion activity, status data and WLAN, and also to compensate for potential inaccuracies. For the classification of place types, over 80 features are generated for each stop. Thereby, geographic data is enriched with point of interest (POI)-information from different location-based context providers. In our experiments, we classify semantic categories of locations using parameter optimized multi-class and smart binary classifiers. An overall accuracy of 88.55% correctly classified stops is achieved using END classifier. A classification without GPS data yields an accuracy of 85.37%, demonstrating that alternative smartphone data can largely compensate for inaccurate localizations based on the fact of 88.55% accuracy, where GPS data was used. Knowing the semantics of a location, the provided context can be used to further personalize autonomous vehicles.

1 INTRODUCTION

With today's high penetration of smartphone devices – capable of accurately monitoring movement, location, communication, and information consumption – and comparable low effort internet access, mobile communication has become ubiquitous. Furthermore, the latest vehicle generations of several manufacturers are capable of sensing data and access the internet. Hence, in the last few years, vehicle-to-vehicle (V2V) and vehicle-to-everything (V2X) communication has become more important than before. A person can have several reasons visiting a location at different daytimes. For instance, a multi-floor building houses supermarkets and restaurants. Around midday on working days a person is likely to have one hour lunch, however, in the evening a person might be more likely to visit this place for quick grocery shopping. Furthermore, the vehicle can tell other drivers about estimated departing times, to avoid unsystematic parking space searching. The understanding of user behaviour and anticipation of next user actions is of high relevance for autonomous driving vehicles. One highly researched sub-field of intelligent vehicles is the automatic generation of recommendations based on user preferences and user habits. For instance, a built-in digital assistant can recommend a parking space depending on the predicted duration of stay: a free of charge parking garage in close distance to a building for lunch or a paid short-term parking zone nearby for grocery shopping.

With today's development progress of cars and smartphones, an unprecedented amount of data can be captured and processed, providing direct measurements of human behavior and the surrounding environment (Dashdorj and Sobolevsky, 2015), and offering an enormous potential of better understanding users context. One step towards assessing the user context is the semantic identification of a user's whereabouts, e.g., the user's home, workplace, preferred restaurant.

Roor, R., Hess, J., Saveriano, M., Karg, M. and Kirsch, A.
Sensor Fusion for Semantic Place Labeling.
DOI: 10.5220/0006365601210131
In *Proceedings of the 3rd International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2017)*, pages 121-131
ISBN: 978-989-758-242-4

A purely location based identification (using Global Positioning System (GPS)) of place types the user visits yields unsatisfying results in some cases. For example, in areas with a dense accumulation of different types of places, even a slightly inaccurate localization might lead to false conclusions. Thus, we propose a framework to record a comprehensive number of sensor and state values of smartphones. This framework does also predict a person's semantic location context taking the recorded sensor and state data into account. In the further text, we will call the classification of the meaning of a place visited by a user *semantic place labeling*. In order to precisely classify places, descriptive features for each place type are extracted by feature selection algorithms. The multi-class classification problem has also been divided into a set of 2-class classification problems, producing an ensemble of smart binary classifiers.

Since our framework processes sensitive user data, privacy concerns are justified. The focus of this work is of a technical background. Thus, we will not cover research questions about data privacy concerns, but we encourage for further research about this topic.

Summarizing, the main contributions of this paper are as follows: (a) a novel, comprehensive set of features (user behaviour and environmental features) for classification of place types; (b) a novel methodology based on smart binary classifiers to solve the multi-class classification problem with intelligent preselected features; (c) duration-specific smart binary classifiers for exploiting inter-feature correlations.

2 RELATED WORK

The Nokia Lausanne data collection campaign (LDCC) dataset is the basis of the Mobile Data Challeng (MDC), a challenge for students with different tracks. One MDC task was semantic place prediction (Laurila et al., 2013). Since the LDCC dataset isn't fully labeled, the predicted meaning of a place couldn't truly be validated, but only estimated. The winner of the semantic place prediction task achieved a 10-fold cross-validation accuracy of 75% using Gradient Boosted Trees classification (Kiukkonen et al., 2010). Zhu et al. focused on generating as many features as possible and let their algorithm decide about the most relevant features (Zhu et al., 2012).

Microsoft Research Asia released a dataset collected by 178 participants in Beijing, China. The unlabeled dataset, called *GeoLife*, was logged by GPS loggers the users were equipped with (Yu Zheng, 2011). Based on this dataset, Ghosh et al. have developed the *THUMP* framework to analyze large GPS traces, cluster trajectories using geographic and semantic information to identify different categories of people regarding the theory that people move with intent (Ghosh and Ghosh, 2016). Further, the authors in (Lung et al., 2014) show that next location prediction, using the same dataset, can be improved using behavior semantic mining. In (Bar-David and Last, 2014), the authors show a context-aware location prediction algorithm trained and tested on the *GeoLife* dataset. Due to missing labels and incomplete information (GPS only) this dataset doesn't fit our needs for semantic place labeling.

In 2004, the Massachusetts Institute of Technology (MIT) launched a data collection challenge called *Reality Mining Project* at their campus using Nokia 6600 phones running a logging app, that is capable to record GPS, Bluetooth, cell tower IDs, and application usage. This dataset is not as comprehensive as the LDCC dataset, unlabeled, and not widely spread in science (Eagle and Pentland, 2006). Another dataset containing only GPS data is *INFATI*. Jensen et al. have equipped 24 cars – mainly located in Aalborg, Denmark – with GPS logging equipment for two months in 2001 (Jensen et al., 2004). Like the aforementioned datasets, this also does not fit our needs.

Other studies on semantic place labeling so far (Reddy et al., 2010; Consolvo et al., 2008; Arase et al., 2010; Bouten et al., 1997; Perrin et al., 2000; Junker et al., 2004; Preece et al., 2009; Berchtold et al., 2010; Ravi et al., 2005; Bao and Intille, 2004; Chang et al., 2007; Farringdon et al., 1999; Kern et al., 2003; Mantyjarvi et al., 2001; Stikic et al., 2008; Zinnen et al., 2009; Lester et al., 2005; Siewiorek et al., 2003) are mostly based on unlabeled data or on a small number of sensor and state data. The field of physical activity recognition based on accelerometer sensor data is well researched (Consolvo et al., 2008; Arase et al., 2010; Berchtold et al., 2010; Bao and Intille, 2004; Farringdon et al., 1999; Kern et al., 2003). Accuracies of physical activity recognition could be achieved up to 90% (Reddy et al., 2010; Preece et al., 2009; Ravi et al., 2005; Bao and Intille, 2004; Chang et al., 2007; Mantyjarvi et al., 2001), but the current average smartphone has more sensors built in than only a accelerometer. Thus, our research focuses on exploiting as many sensor and state values our algorithms need to detect the semantic of a place for a user using sparse data to not unnecessary drown the battery. In contrast to previously mentioned publications, we focus on semantic place labeling, which it is not well researched, instead of activity recognition.

Figure 1: Our self developed *Mobility Companion* App for Android based smartphones used for ground truth data collection. The timeline view shown of identified stops for selected date extended with labeling possibilities for places and transportation modes.

3 DATA

A logging application for Android based smartphones was developed and distributed via Google Play Store to a diverse range of users, co-workers and friends of the authors, who agreed to participate in our data logging challenge. A broad range of smartphone sensors and status are recorded:

- accelerometer
- Bluetooth (MAC, bond state, name, type, class, connection)
- Google activity recognition API
- GPS
- phone status data (airplane mode, Android version, cell service, phone model, phone plug, plug status, ringer mode)
- wireless local area network (WLAN) (BSSID, SSID, capabilities, frequency, level)

In addition to automatic sensor logging, the participants of our experiment were required to label all visited places and, albeit not relevant for this case, commutes as shown in fig. 1. While labeling places, semantic descriptions of the corresponding locations can be selected by the participants. The user can choose between *home, education, work, friend & family, hotel, restaurant, nightlife, grocery store, shopping, sport medical, leisure, transportation infrastructure* and *other*.

The collected app data is stored in our central database server. To lower the barriers for users to participate in this data collecting challenge, the *Mobility Companion* app should behave inconspicuous while running as a background task on the users' phone. One important criteria is to not drown the battery more than necessary. To achieve this, several battery saving strategies were implemented. Sparse data recording due to a low logging frequency in combination with geofences can reduce the battery usage to under 10%. Most sensor values are logged once between every 45 seconds and 15 minutes. The logging frequency is adapted by our power saving algorithm based on the user's behaviour.

Data collection happened in a period over 183 days in 2016. Over 19 users have contributed their logged data. Although the majority of users are located around Munich, Germany, the recorded data exhibits stops amongst a variety of countries, e.g., Hong Kong and Philippines. We consider the data of a user as valid if a minimum of 30 labeled stops of the corresponding user is collected, which is roughly equivalent to movements of one week.

In total, 1852 labeled stops (see tab. 1) of a total duration of 6700 hours were eligible for classification of which 90% contain motion activity data, 58% Bluetooth data, 88% WLAN data and 99% phone status data (e.g. ringer- and airplane mode). Places of types *grocery store* and *shop* were not distinct enough for many data collectors. Thus, we decided to merge instances of these two labels into *shop*. Instances from place types *hotel*, *Leisure* and *medical* are underrepresented – only 3% of all instances – and could not be used for classification, to avoid over- or underfitting.

Table 1: Distribution of stops. Instances of place type *grocery store* and *shop* are merged into *shop* due to ambiguity of terms. Instances from place types *Hotel*, *leisure* and *medical* are ineligible for classification due to being underrepresented.

Place Type	**Instances**	
home	707	38%
education	107	6%
work	344	19%
friend & family	237	13%
restaurant	177	10%
nightlife	59	3%
shop	85	4%
sport	81	4%
transportation infrastructure	55	3%
Total	**1852**	**100%**

Table 2: Extracted features per stop grouped by category. Based on the data logged by the *Mobility Companion* app and additional data sources, all of the listed features are generated and used for classification.

	Feature per stop
activity	absolute duration {in vehicle, on bicycle, running, still, tilting, unknown, walking}
	relative duration {in vehicle, on bicycle, running, still, tilting, unknown, walking}
	predominant activity
	second most predominant activity
	activity index {current, preceding, succeeding}
	frequency of activity change {current, preceding, succeeding}
settings & status	average cell service signal strength
	has been plugged in
	predominant plug type
	predominant ringer mode
	ringer mode has been changed
	share of time {airplane mode, cell service available, unplugged}
	share of time plugged {AC, USB }
	share of time ringer mode {normal, silent, vibrate}
stop & time	absolute duration of {cluster this day, stop}
	is stop after shop closing time
	is workday
	predominant {preceding, succeeding} place type
	share of time {airplane mode, cell service available, unplugged}
	time of day as middle of stop
	total share of night time spent at this cluster
	total share of time spent at this cluster
WLAN	average network type {overall, strongest networks only}
	average network type of connected network
	connected to educational network
	educational network nearby
	number of unique {BSSIDs, SSIDs} nearby
	share of time connected to a WLAN network
Bluetooth	detected devices, share of type {audio video, computer, health, imaging, misc, networking, peripheral, phone, toy, uncategorized, wearable}
	most connected type of Bluetooth device
	number of unique Bluetooth devices nearby
	share of time connected to a Bluetooth device
geographic	distance to nearest {railway, road, road or railway}
	is close to {railway, road}
	most likely place type based on POI
	POI-probability of place type {home, education, work, friend & family, restaurant, nightlife, shop, sport, transport infrastructure}

Table 3: Every activity type as possible return value of Google's activity recognition API is mapped to an activity index and linked to an activity group in respect to its movement intensity. The values are designed by us to reflect the activity's motion intensity. Thus, it is possible to calculate an average activity index for each stop based on the user's activities.

Activity type	**Index**	**Activity group**
still	0	Non-translational movements
tilting	1	
walking	4	Translational movements
running	7	
on bicycle	9	
in vehicle	10	
unknown	–	–

4 CLASSIFICATION FEATURES

4.1 User Behaviour

4.1.1 User Activity

Physical user activity can be characteristic for a place type. For instance, at work or in a restaurant one is less likely to move than in the gym or a shop. Android has a built in Google application programming interface (API) for activity recognition, which yields a probability distribution over activity categories: *still*, *tilting*, *walking*, *running*, *on bicycle* and *in vehicle*. Several features were extracted based on the determined activity categories, see tab. 2 *(activity)*.

The activity categories were mapped to an activity index ranging from 0 to 10 reflecting the activity's motion intensity as shown in tab. 3. Generally, the more translational the activity is the higher is the activity value.

The *frequency of activity changes* is calculated as a function of number of changes between non-translational movements and translational movements, divided by a fixed interval of 30 minutes. The frequency changes feature can help to determine place types with usually a high activity, e.g., shop, in contrast to low activity place types, e.g., restaurant. Due to battery saving strategies the activity recognition API is not recorded continuously, but up to every 45 seconds.

4.1.2 Smartphone Settings and Status

The way a smartphone is used can give indications about whereabouts. Typically the smartphone settings correlate with place types, for instance, active airplane mode at home during nighttime and silent ringer mode

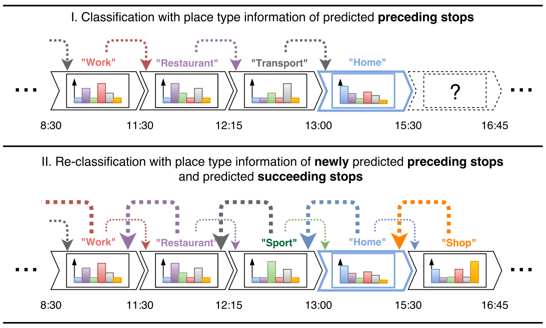

Figure 2: Information about predicted preceding place types can improve classification accuracy. To further improve accuracy preceding place types are reclassified once using information of predicted succeeding place types.

at work. For each stop, the features listed in tab. 2 *(settings & status)* are extracted for classification.

4.1.3 Stop and Time

Over the course of the day, many stops are visited in which the sequence is often not random. *Friend & family* is often visited after work and leaving a shop is followed by the place type *home* in most cases. To embrace such correlations, the *predominant place type in the 2 hours before the stop* and the *predominant place type in the 2 hours after the stop* are calculated. Once succeeding place types have been classified, classification information can be retrieved and used to reclassify preceding place types to improve accuracy. Reclassification is limited to 1 iteration in this case. This concept is illustrated in fig. 2. For practical reasons, information about the preceding and succeeding place types are available from the beginning but falsified along confusion matrices of the classifier, since all stops are potentially subject to false classification. For the preceding place types, a confusion matrix is used that originates from a classification without any knowledge about preceding and succeeding place types. For the succeeding place type a confusion matrix is used that originates from a classification with falsified knowledge of preceding place types and no knowledge about succeeding place types. This way, we simulate uncertainty about succeeding place types. As in a real world scenario, succeeding place types are unknown before visit.

With respect to time information, additional features are used. Spatial and temporal information usually correlate, for instance, it's normal to be at home during nighttime. To assess such relations, features as listed in tab. 2 *(stop & time)* are extracted.

4.2 Environment

4.2.1 WLAN

Due to a high penetration of WLANs, the existing WLAN access points (APs) infrastructure is used, for instance, for GPS localization improvement. Such networks have assigned a service set identifier (SSID), the broadcasted name of the network, and a basic service set identifier (BSSID), an unique identifier of the AP. Thus, we extract *absolute number of unique BSSIDs and SSIDs* as features.

Furthermore, WLAN are differentiated between private and enterprise networks by the type of authentication, the number of APs and used frequencies. WLANs in e.g. households often consist of one AP. In contrast, companies run networks consisting of several APs with different BSSIDs, but the same SSID and support Extensible Authentication Protocol (EAP).

In 2003 the *eduroam* initiative started and aims to give students free WLAN access around the world. Only WLANs in educational facilities are emitting *eduroam* as SSID. Regarding this fact, there is a high likelihood the place is of educational nature if an *eduroam* SSID is detected. Applying all these rules the features listed in tab. 2 *(WLAN)* are extracted.

4.2.2 Bluetooth

In reference to WLAN features, Bluetooth devices in the immediate vicinity are scanned to extract the features listed in tab. 2 *(Bluetooth)*.

4.2.3 Cell Service

At some places there can be very characteristic cell service levels. To investigate if this also applies for place types cell service features are extracted (see tab. 2 *(settings & status)*).

4.2.4 Geographic Environment

In conjunction with Foursquare and Google Places – two comprehensive POI providers – information about nearby places can be exploited. Every detected stop is a result of several closely aligned coordinate pairs. A cluster shape (determined through all coordinate pairs within a certain range that were detected during a stop), and a centroid are calculated for every stop. The centroid is used to query Foursquare and Google Places for POI within a range of 50m (average localization inaccuracy within buildings).

POI-based probability will be derived for each place type. First, places that are not opened throughout the stop's duration are excluded. Second, a weight w_k

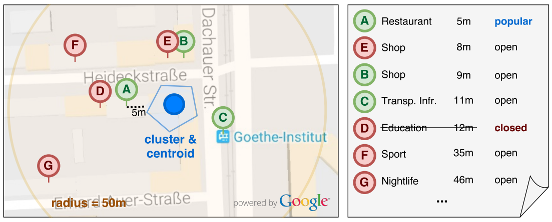

Figure 3: POI around a distance of 50m around the stop cluster shape's centroid are queried from Foursquase and Google Places, symbolized as red and green places. Distance to the stop's cluster and popular times are taken into account in the probability calculation while deriving *POI-based probability*. POIs outside the opening hours are excluded.

is calculated for each place type k using eq. 1, taking into account each place's distance to the cluster area in a quadratic sense and whether they were popular during the stops time frame, as specified by Foursquare.

$$w_k = \sum_{i=1}^{n_k} \frac{1+\beta}{\alpha \cdot {dist_{k,i}}^2} \quad , \tag{1}$$

where β is an additional popularity and α is an additional distance factor. In this case, we set $\alpha = 2$ and $\beta = 0.5$ for popular else $\beta = 0$ for unpopular times. The distance between the stop's cluster shape and a POI i of place type k is expressed by $dist_{k,i}$, as depicted in fig. 3. Finally, to calculate the POI-based probability for each place type p_k eq. 2 is used and reduced by a correction factor γ.

$$p_k = \frac{w_k}{\sum_k^N w_k} \cdot \gamma \quad , \quad \text{where} \quad \gamma = 1 - \left(\frac{dist_{min}}{dist_{max}}\right)^2 \tag{2}$$

To avoid that even distant POI receive an unrealistic, high probability, γ ranges from 0 to 1 and adjusts the probability distribution for situations where no POIs are found in defined maximum distance $dist_{max}$ to the cluster shape. It is calculated by the distance of the overall nearest detected place $dist_{min}$ and the maximal possible distance $dist_{max} = 50m$.

Most probable POI-based place type as additional feature is derived from overall POI-probabilities of place types. Towards *transportation infrastructure* place types, features are generated exploiting Google's Roads and Overpass Rails. All extracted geographic features are listed in tab. 2 *(geographic)*.

5 IMPLEMENTATION & EVALUATION

The database and the features described in the earlier sections are used for evaluation of different classification algorithms and strategies. For training and testing of each classification model, we apply 10-fold cross-validation. Unless otherwise described, default parameter settings of algorithms from Waikato Environment for Knowledge Analysis (Weka)[1] are used for evaluation.

5.1 Multi-class Classification

A multi-class classification setup is the direct application of a classifier on the dataset with optional feature selection prior to the classification. We computed results using 23 different classifier algorithms implemented in Weka. Fig. 4 shows results of the 6 best performing algorithms. Trying to obtain best results, Weka's default parameters of END, Rotation Forest, Logit Boost and Random Comitee were tuned. These algorithms are ensemble classifiers that include a learning subsystem which is iteratively adapted with experience. Hence, to tune the algorithms we subsequently raised the iterations from 10 (default) to over 500. Best results, also in respect to computation time, were achieved with 500 iterations.

A general feature selection can be performed additionally prior to classification. Here, the PCC is used as measurement of relevance, and calculated as weighted average of class-specific correlation between features and place types. All features that carry a correlation value below 0.05 were removed, leaving only features that were related to one or multiple place types. However, the results in fig. 4 show minimal or even a negative improvement of accuracies. This demonstrates the algorithms' intrinsic ability to assess features' relevance. The 3 most distinguishing features per place type, according to PCC, are shwon in tab. 4.

5.2 Binary Classification Setup

For a very individualistic setup regarding feature selection a binary classification model is trained. The END classifier uses a similar concept to break down multi-class problems into a set of binary classification problems in terms of performance improvement. In this setup, PCC information between each feature and place type are used. For each class, features with a PCC value below a specific threshold are removed to avoid potential wrong classification. After feature selection for each place type, the classes are classified individually.

For each of those classifications, all instances of the respective other classes are grouped into an opposing class. Then, the classifier is trained and tested with the target class and the opposing class. Thereby,

[1] http://www.cs.waikato.ac.nz/ml/index.html; Version 3.8.0

Figure 4: Benchmarks of best performing classifiers in a multi-class setup using implementations of Weka with default parameter settings and 10-fold cross-validation, but adjusted iterations to 10 and 500 iterations. The LMT algorithm has a self-optimized number of iterations, the exact number is unspecified. Simple Logistic is by default well-tuned and uses 500 iterations. Feature selection prior to classification shows no significant impact.

for each tested instance, a probability value is given to the target class. The same instance is tested with each class and the class with the highest probability is chosen as predicted class. This is how a multi-class problem is transformed into a set of binary-class problems.

The PCC threshold affects the results as shown in fig. 5 for the top classifiers. Thereby, a performance improvement is achieved in case of the Random Committee and Rotation Forest. Logistic Model Trees (LMT) and Simple Logistic obtained lower accuracies.

5.3 Binary Classification Setup with Duration-specific Model Building

The binary setup described before is limited in the way that features are evaluated in one-dimensional view as only correlations between features and place types are considered, while correlations between features are neglected. For instance, a user only stopping at a gym is less likely to plug in his phone for charging and a short stop at a friend's place usually has a higher activity index than a long stop.

In respect to this assumption, a duration specific model is built consisting of two classifiers for each place type, one for shorter and one for longer stops. They are trained in parallel with the same instances and features except for those whose inner correlation values with the stop's duration are above a specified threshold. Such features are trained only to either one of the classifiers, respectively to the stop's duration. Thus, a trade-off for the gained training-potential is the decrease of training instances for duration-specific features, which can result in increased effects of under- and over-fitting of the duration-specific models.

Also, the classification accuracy depends on the specified threshold value for the PCC between each feature and the stop's duration. To ensure different results, as with a purely binary setup, the value has to be lower than 1.0. Otherwise, the features would be trained not duration-specific. A fix correlation threshold of 0.05 for correlation between features and place types is used as this value achieves the overall best results in the purely binary setup, as discussed earlier.

Fig. 5 shows percentage of correct classifications for the top performing classifiers. In consequence of low threshold values, training features for some of the place types are thinned out and lead to poor classification performance. Overall, compared to the multi-class setup, the accuracy of the Rotation Forest and the Random Committee shows significant improvements. For all other classifiers compared to the multi-class setup, this setup obtains slightly lower accuracies. Compared to the binary setup, this setup achieves similar classification accuracies.

6 DISCUSSION

Besides the best performing classifier END with an overall accuracy of 88.55%, several other classifier yielded similar results. Representatively, the performance of the classifier END is discussed in this section. Classification results of the classifier END are shown in in tab. 5.

With the proposed setup, place types called *home* and *work* can be clearly distinguished, resulting in a true positive rate (TPR) of 0.989 and 0.93, a false positive rate (FPR) of 0 and 0.011, and a precision of 1 and 0.952 respectively. Similar values apply for place type *education*, for instance, FPR is also above 0.93. One reason for clearly distinguishing this place types are the features about *eduroam*, which are only present at educational places.

In contrast, a slightly lower FPR occurs for place

Table 4: Top three most distinguishing features per place type, measured by the Pearson product-moment correlation coefficient (PCC).

Place type	**Feature**	**PCC**
home	total share of time spent at this cluster	0.91
	total share of night time spent at this cluster	0.79
	absolute duration of *cluster this day*	0.67
education	educational network nearby	0.76
	connected to educational network	0.59
	average network type *overall*	0.46
work	number of unique Bluetooth devices nearby	0.63
	distance to nearest *road*	0.52
	is stop after shop closing time(1)	0.45
friend & family	total share of time spent at this cluster	0.30
	is workday	0.22
	total share of night time spent at this cluster	0.21
restaurant	POI-probability of place type *nightlife*	0.36
	total share of time spent at this cluster	0.31
	absolute duration of *cluster this day*	0.31
nightlife	POI-probability of place type *nightlife*	0.23
	relative duration *unknown*	0.21
	frequency of activity change *current*	0.18
shop	POI-probability of place type *shop*	0.53
	relative duration *unknown*	0.30
	activity index *current*	0.27
sport	POI-probability of place type *sport*	0.28
	share of time connected to a WLAN network	0.22
	total share of time spent at this cluster	0.20
transport infrastructure	POI-probability of place type *transp. infrastr.*	0.35
	distance to nearest *railway*	0.29
	activity index *current*(2)	0.25

(1)Shown instead of third most correlated feature 'distance to nearest *road or railway*' due to its strong contextual overlap with 'distance to nearest *road*'.
(2)Shown instead of third most correlated feature 'is close to *railway*' due to its strong contextual overlap with 'distance to nearest *railway*'.

type *friend & family*. Especially instances of *sport* appear to be difficult to distinguish from *friend & family*, as they are the most misclassified place type for *friend & family*. One reason is the low amount of distinguishing sport-features that are available in all instances. Investigating this case, it became clear that most of the participants don't take their smartphone to the gym or secure it in the locker. Hence, sensor values are comparable to place type *friend & family*.

Similarly, *restaurant* shows a relative high TPR and has by far the highest FPR as well as the lowest precision. This indicates a large overlap between features of *restaurant*-stops and instances of other classes. Depending on the user's situation, stays at a restaurant differ significantly in respect to activity profile, duration and time of visit. The purpose of visit can be, for instance, a small meal in a hurry, a coffee only, or a long dinner with subsequent drinks. Logged sensor values are also highly dependent on the type of restaurant. While having an extensive lunch or dinner at a normal restaurant the user is mostly sitting, in contrast to a fast food restaurant with standing tables only. Depending on the type of restaurant, the activity profile will differ significantly. Moreover, restaurants exist numerously and prevalently and are, in contrast to other place type, registered at POI providers such as Foursquare and Google Places. This can affect classification of other place types as *restaurant*, since in immediate distance often a place of type *restaurant* is detected. This may happen, because the users' home is above a restaurant, for stops in a multi-floor building that houses a restaurant above or below users' position, at place types where it is a common part like sport-related places or inaccurate recognized user position. The generated geographic features would take this information into account and potentially suggest a restaurant context. It becomes clear that several factors can be the reason for this high FPR and low precision.

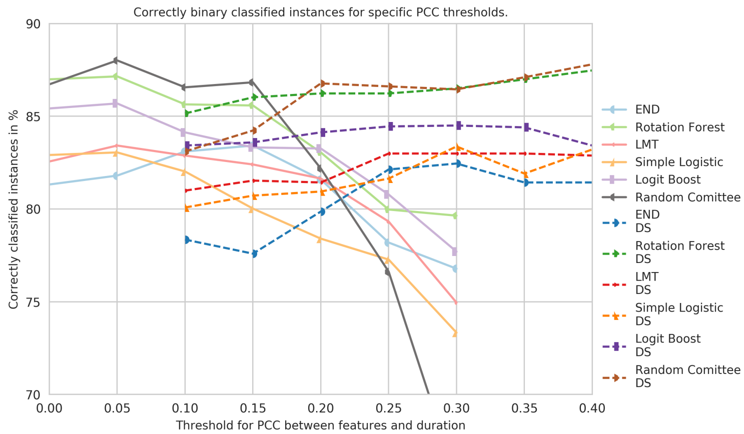

Figure 5: Classification results of top binary classifiers and binary duration-specific (DS) classifiers. Chart shows correctly classified instances in % for specific PCC thresholds. Solid lines indicates binary classifiers results. Dashed lines indicates binary duration-specific results.

Table 5: Classification results of best-performing classifier END with 500 iterations and 10-fold cross-validation.

Place Type	TPR	FPR	Prec.	ROC	F-Meas.
home	0.989	0.000	1.000	0.998	0.994
education	0.935	0.005	0.926	0.995	0.930
work	0.930	0.011	0.952	0.995	0.941
friend & family	0.844	0.024	0.840	0.985	0.842
restaurant	0.842	0.054	0.623	0.974	0.716
nightlife	0.441	0.002	0.867	0.971	0.584
shop	0.718	0.018	0.663	0.977	0.689
sport	0.691	0.005	0.875	0.979	0.772
transp. infrastr.	0.527	0.009	0.630	0.974	0.574
Weighted Avg.	0.886	0.012	0.894	0.990	0.885

Class *nightlife* contains up to 30% miss-classified instances whereat places of type *restaurant* are the evident majority. In addition to the classification difficulties for place type *restaurant*, there is a contextual difficulty. Often, there is a fine line between a nightlife location, such as a bar, and a restaurant. *Nightlife* locations often offer small dishes and restaurants drinks, two reasons for visiting early and staying late respectively. Hence, the classification as either one of them lies solely in context of visit. Due to these circumstances the MDC treated these two place types as one (Laurila et al., 2013). The relative high precision value of 0.867 suggests that distinguishing features exist, since nightlife is classified correctly, both place types can still be treated separately.

The instances of place type *sport* are often miss-classified, due to a low amount of distinguishing features recorded by our app. As mentioned before, the smartphone is often stored in a locker while user is actively moving. In many cases, no sports-related POI is found at sports location. As shown in fig. 6, almost all instances that contain significant POI-probability

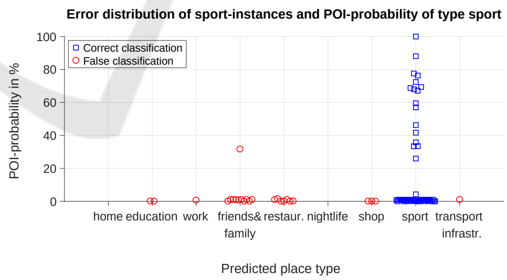

Figure 6: Distribution of classified instances of place type *sport* for the respective location-based POI-probability, this type's most correlated feature. In cases where the POI-probability is near zero (almost no sports-related POI recognized) instances of *sport* often miss-classified.

for type *sport* are classified correctly. Vice versa, almost all wrongly classified instances did not contain a significant POI-probability. The reason because disproportionally many stops are misclassified as *friend & family* is likely related to the fact that environments and behaviors at places of friends can be relatively feature-less as well with similar motion activity.

In the case of *transport infrastructure* almost the

Table 6: Comparison of additional benefit of feature groups w.r.t. accuracy adding or removing specific feature groups for model building with END. General stop and time features are taken as basis, yielding 81.05% accuracy.

Feature Group	**Stop & time feat. and feat. group**	**All feat. except feat. group**
User activity	81.59%	87.91%
Settings & status	82.56%	88.17%
WLAN	84.23%	87.91%
Bluetooth	81.75%	**88.39**%
Geographic	**86.88**%	85.37%

half of the respective instances are classified correctly, due to the low number of distinguishing features, mainly driven by geographic information and user motion activity. As this is the class with the lower amount of instances, the comparable low TPR can be interpreted as a cause of underfitting. Hence, more training data is necessary for a reliable classification.

Additionally, the benefit of each feature group is compared w.r.t. accuracy. The same setup of END is used with feature group *stop & time* as basis, yielding 81.05% accuracy. In contrast, using all features an accuracy of 88.55% can be achieved as mentioned before. The results in tab. 6 shows the significant contribution of geographic data. While adding only this feature group to the base feature group *stop & time* an accuracy of 86.88% can be achieved. Furthermore, it shows that non-geographic features are able to largely compensate for an outage of geographic features, e.g. in situations where an accurate localization isn't possible. Taking every feature group into account except Bluetooth, shows the minimal additional benefit this feature group adds to accuracy.

7 CONCLUSION

Our research shows how semantics about users' whereabouts can be derived based on various sensor and state values. This procedure, called semantic place labeling, is essential for context inference of everyday user situations. Especially for autonomous driving vehicles, knowledge about the users' context is useful in order to anticipate the users' behaviour.

Hence, we have shown how to use our framework to incorporate distributed sensor and state data and classify types of places depending on users' actions. Semantic place labeling can distinguish different place types even at the same location, for instance, housed in a multi-floor building or even wrong logged coordinates due to inaccurate localization, and derive their semantics. The framework design has its focus on extensibility, so every sensor and state source can be attached in order to gain more insight about user intentions and to achieve a higher place type classification accuracy.

Due to lack of freely available data sets, a highly convenient Android based app for tracking of users' behavior, environment and ground truth annotation was developed (Kiukkonen et al., 2010; Laurila et al., 2013; Yu Zheng, 2011). A sizeable amount of valid data was collected and submitted by 19 participants over a time span of 183 days. In this research, over 80 features with mixed relevance to place types are generated per stop. Several classifiers were compared. In the classification, up to 88.6% of test instances are correctly classified across nine place types by END.

The evaluation has shown that even inaccurate location data can be compensated with remaining features, yielding an accuracy of 85.37%. As far as a comparison can be drawn to related approaches, the yielded prediction results are more accurate and classifiers are generalizing better on less routine place types than any other known approach (Zhu et al., 2012; Montoliu et al., 2012; Ghosh and Ghosh, 2016; Bar-David and Last, 2014).

The achieved classification accuracy of 88.6% maybe not sufficient for autonomous driving vehicles. Over 11 out of 100 actions of an autonomous driving car based on anticipation of the users intentions can still be false. Any false anticipated user intentions can annoy the user and should be reduced to a minimum. Hence, the next study should investigate what minimum accuracy is needed to gain the users trust.

Future work will focus on (a) improvement of the Mobility Companion app w.r.t. usability and power consumption, (b) extension of the model, making use of new features – for instance, knowledge about user-user relation and relations between user activities and local events – and data sources – for instance, distributed sensors (like vehicle sensors) – and (c) further logging and publishing of our annotated dataset.

REFERENCES

Arase, Y., Ren, F., and Xie, X. (2010). User activity understanding from mobile phone sensors. In *Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing-Adjunct*, pages 391–392. ACM.

Bao, L. and Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. In *International Conference on Pervasive Computing*, pages 1–17. Springer.

Bar-David, R. and Last, M. (2014). Context-aware location prediction. In *International Workshop on Modeling Social Media*, pages 165–185. Springer.

Berchtold, M., Budde, M., Gordon, D., Schmidtke, H. R., and Beigl, M. (2010). Actiserv: Activity recognition service for mobile phones. In *International Symposium on Wearable Computers (ISWC)*, pages 1–8. IEEE.

Bouten, C. V., Koekkoek, K. T., Verduin, M., Kodde, R., and Janssen, J. D. (1997). A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Transactions on Biomedical Engineering*, 44(3):136–147.

Chang, K.-H., Chen, M. Y., and Canny, J. (2007). Tracking free-weight exercises. In *International Conference on Ubiquitous Computing*, pages 19–37. Springer.

Consolvo, S., McDonald, D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., LeGrand, L., Libby, R., et al. (2008). Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1797–1806. ACM.

Dashdorj, Z. and Sobolevsky, S. (2015). Characterization of behavioral patterns exploiting description of geographical areas. *CoRR*, abs/1510.02995.

Eagle, N. and Pentland, A. S. (2006). Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268.

Farringdon, J., Moore, A. J., Tilbury, N., Church, J., and Biemond, P. D. (1999). Wearable sensor badge and sensor jacket for context awareness. In *Wearable Computers. Digest of Papers. The Third International Symposium on*, pages 107–113. IEEE.

Ghosh, S. and Ghosh, S. K. (2016). Thump: Semantic analysis on trajectory traces to explore human movement pattern. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 35–36.

Jensen, C. S., Lahrmann, H., Pakalnis, S., and Runge, J. (2004). The infati data. *arXiv preprint cs/0410001*.

Junker, H., Lukowicz, P., and Troster, G. (2004). Sampling frequency, signal resolution and the accuracy of wearable context recognition systems. In *Wearable Computers, Eighth International Symposium on*, volume 1, pages 176–177. IEEE.

Kern, N., Schiele, B., and Schmidt, A. (2003). Multi-sensor activity context detection for wearable computing. In *European Symposium on Ambient Intelligence*, pages 220–232. Springer.

Kiukkonen, N., Blom, J., Dousse, O., Gatica-Perez, D., and Laurila, J. (2010). Towards rich mobile phone datasets: Lausanne data collection campaign. *Proc. ICPS, Berlin*.

Laurila, J. K., Gatica-Perez, D., Aad, I., Blom, J., Bornet, O., Do, T. M. T., Dousse, O., Eberle, J., Miettinen, M., Liao, L., Fox, D., and Kautz, H. (2013). From big smartphone data to worldwide research: The Mobile Data Challenge. *The International Journal of Robotics Research*, 26(6):119–134.

Lester, J., Choudhury, T., Kern, N., Borriello, G., and Hannaford, B. (2005). A hybrid discriminative/generative approach for modeling human activities. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pages 766–772.

Lung, H.-Y., Chung, C.-H., and Dai, B.-R. (2014). Predicting locations of mobile users based on behavior semantic mining. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 168–180. Springer.

Mantyjarvi, J., Himberg, J., and Seppanen, T. (2001). Recognizing human motion with multiple acceleration sensors. In *Systems, Man, and Cybernetics, IEEE International Conference on*, volume 2, pages 747–752. IEEE.

Montoliu, R., Martínez-Uso, A., Martínez-Sotoca, J., and McInerney, J. (2012). Semantic place prediction by combining smart binary classifiers. In *Nokia Mobile Data Challenge Workshop.*, volume 1.

Perrin, O., Terrier, P., Ladetto, Q., Merminod, B., and Schutz, Y. (2000). Improvement of walking speed prediction by accelerometry and altimetry, validated by satellite positioning. *Medical and Biological Engineering and Computing*, 38(2):164–168.

Preece, S. J., Goulermas, J. Y., Kenney, L. P., Howard, D., Meijer, K., and Crompton, R. (2009). Activity identification using body-mounted sensorsa review of classification techniques. *Physiological measurement*, 30(4):R1.

Ravi, N., Dandekar, N., Mysore, P., and Littman, M. L. (2005). Activity recognition from accelerometer data. In *AAAI*, volume 5, pages 1541–1546.

Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., and Srivastava, M. (2010). Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks (TOSN)*, 6(2):13.

Siewiorek, D. P., Smailagic, A., Furukawa, J., Krause, A., Moraveji, N., Reiger, K., Shaffer, J., and Wong, F. L. (2003). Sensay: A context-aware mobile phone. In *ISWC*, volume 3, page 248.

Stikic, M., Van Laerhoven, K., and Schiele, B. (2008). Exploring semi-supervised and active learning for activity recognition. In *12th International Symposium on Wearable Computers*, pages 81–88. IEEE.

Yu Zheng, Hao Fu, X. X. W.-Y. M. Q. L. (2011). *Geolife GPS trajectory dataset - User Guide*.

Zhu, Y., Zhong, E., Lu, Z., and Yang, Q. (2012). Feature engineering for place category classification. *Mobile Data Challenge 2012*.

Zinnen, A., Blanke, U., and Schiele, B. (2009). An analysis of sensor-oriented vs. model-based activity recognition. In *International Symposium on Wearable Computers*, pages 93–100. IEEE.