University of Zurich[UZH]

Department of Geography

Building and Using a Reference Data Set for Transportation Mode Detection

GEO 511 Master's Thesis

Author:
Vera Isler
13-755-905

Supervised by
Prof. Dr. Robert Weibel
Oliver Burkhard

Faculty representative
Prof. Dr. Robert Weibel

Saturday 29th September, 2018
Department of Geography, University of Zurich

Acknowledgements

Reaching the end of this thesis, I would like to express special thanks to:

- Prof. Dr. Robert Weibel and Oliver Burkhard for their supervision, support and collaboration on the thesis and especially on the collection campaign.

- Peter Widhalm and Maximilian Leodolter for the vivid discussions, Skype meetings and collaboration on the collection campaign.

- Christina Röcke and Alexander Seifert from the University Research Priority Program (URPP) "Dynamics of Healthy Aging" at the University of Zurich for lending us smartphones for the data collection.

Abstract

Inferring modes of transportation from motion sensor data such as Global Positioning Service (GPS) or accelerometers has been a topic of interest ever since small sensors have become more available, especially on smartphones. A number of algorithms have been developed for this task, but the lack of a common evaluation data set and corresponding metrics has made the comparison of the performances difficult. Therefore, the first part of this thesis, conducted in collaboration with scientists from the Austrian Institute of Technology, aims at defining a concept for the collection of a reference data set. The conceptual considerations were gathered in a protocol, which was later used for the execution of a collection campaign in Zurich and Vienna. The collection was carried out in a three months period by around 20 participants from the University of Zurich and the Austrian Institute of Technology. Each trip was recorded on three devices (two smartphones and one uTrail device) and annotated simultaneously by the participant on a smartphone application.

The second part of the thesis gives an example of how the collected data could be used and looks at the performance of different classifier on point-wise data versus segmented data. The data was split up using three different methods. First a segmentation by the labels given during the collection, second a first-passage time approach and third using the Behavioral Change Point Analysis by Gurarie et al. [20]. The resulting feature vectors were classified using a random forest, a support vector machine and a K-nearest neighbor approach. The result show, that the classifiers performed generally better on the segmented data. However, further research in this field is necessary and of interest.

Contents

List of Figures

List of Tables

Glossary

AIT Austrian Institute of Technology. 7, 12, 22, 23, 34, 36, 40

BCPA Behavioral Change Point Analysis. 7, 8, 54, 55, 57, 61, 62, 68–70

CSV Comma-separated Values. 37
CV Cross Validation. 60

FPT First-Passage Time. 7, 8, 54, 55, 57, 61, 62, 66–68, 70

GIS Geographic Information System. 16, 71
GIUZ Geographical Institute at the University of Zurich. 35–37, 40
GPS Global Positioning Service. 3, 8, 10–12, 15–20, 24–26, 34, 35, 37, 40–42, 50–54, 56, 70–72
GSM Global System for Mobile Communications. 10, 17

IMT Individual Motorized Transport. 28
IMU Inertial Measurement Units. 34, 35, 37, 51, 73

KNN K-Nearest Neighbors. 7, 8, 59, 60, 63, 65–69

MOASIS Mobility, Activity and Social Interaction Study. 35
MOT Mode of Transportation. 7, 8, 32, 45, 46, 48, 49

RTO Research and Technology Organization. 22

SVM Support Vector Machine. 8, 58, 63–69

UZH University of Zurich. 23, 36

WLAN Wireless Local Area Network. 10, 16, 17

1 Introduction

With the advent of widely available smartphone sensors such as GPS, Global System for Mobile Communications (GSM), Wireless Local Area Network (WLAN), accelerometer, magnetometer or gyroscopes and corresponding applications to collect their records, automated transportation mode detection has been gaining more and more interest. What was originally collected in time-consuming, costly travel surveys, can now potentially be recorded passively by everyone who has a smartphone. Many algorithms for various combinations of sensors and features have been written and tested over the past two decades to automatically recognize transportation modes and the results sound promising.
Detecting the modes of transportation aims to answer the question of how and with which means people travel [36]. This is relevant for a multitude of applications and contexts such as transportation and mobility studies, health and fitness monitoring or personal impact or exposure monitoring. Some examples are given below.
The first thing that comes to mind are fitness trackers. These come as wristbands, watches combined with a smartphone application or just an application on its own and collect GPS and/or acceleration data together with heart-rate, temperature or blood pressure measurements, depending on the device. In this case, a mode of transportation can e.g. be running versus a slow jog versus walking or adding additional work out gear such as cycling. These kinds of applications can be used in a short time frame e.g. for a single workout session or over longer periods to track a person's general motivation to move. While many of these applications are mainly used by individuals for their own interest, data collected in this way can also be used to assess the general fitness or health of a larger population.
An example of personal impact monitoring is the calculation of a person's CO_2 footprint based on the modes of transportation they have chosen over a certain time period. This way, awareness and potentially the motivation to keep this footprint lower, can be increased. On the other hand, knowing how people move in space can show patterns of how e.g. epidemics spread over a city and where the chance of infection is highest.
The application this project is focusing on is transportation and mobility science. Public transportation systems e.g. can benefit immensely from the knowledge gained by detecting transportation modes. Patterns of which modes are used most at which times during the day can be detected, transfer times can be adjusted to the people's needs or lines added or removed according to their utilization. Additionally, demographic and socioeconomic behavior can be characterized and analyzed as mentioned by Bolbol et al. 2012 [4].
Studies which use information concerning transportation mode choices have been con-

ducted for a long time, often using surveying methods such as house visits or telephone calls. This way of collecting data has proven very time-intensive, costly and often not very accurate. Remembering the exact use of transportation modes in retrospect is difficult and the motivation to be precise is often low.
To improve the collection of this type of travel survey data, the advent of GPS played a key role. Tracking the paths of people and showing these to them on a map can help them to remember where they were when and how they got there. But still, the effort invested by participants and researchers is high and there are many error sources left. Additionally, every time new participants are acquired, they need to go through the same process of explaining all of their movement patterns.
This thesis sets its focus on the subject of transportation science. This topic is most related not only to transportation mode detection but also to many spatial issues which are located in the field of geography. However, the data which has been used for training and testing is inconsistent and makes the comparison of different approaches impossible. Therefore, the aim of the first part of this thesis is a project planning and carrying out a data collection campaign to achieve a balanced reference data set. This data set is suggested to researchers of transportation mode detection to test and evaluate their classification algorithms on. For the data collection project, the institute of geographic information systems of the University of Zurich (UZH) collaborated with a research group of the Austrian Institute of Technology (AIT). With Peter Widhalm (senior scientist) and Maximilian Leodolter (PhD candidate) a concept was worked out to ensure a good balance of trip combinations and segments with a chosen set of modes of transport available in the chosen collection regions. The results were incorporated in a campaign protocol. This protocol is designed in a way that it can be used for further data collection in different regions or countries or with a different focus to ensure the possibility of expanding the reference data set further in a controlled manner, depending on the purpose of the research.
In a second step, part of the collected data was used to compare how classification algorithms perform with different types of input. Three classifiers were each tested with point-wise and segmented data. The segmentation algorithms used are a segmentation by labels, a First Passage Time Approach and a Behavioral Change Point Analysis Approach.

1.1 Aims of Thesis

As there have been many studies conducted in the field of automated transportation mode detection, the first part of this thesis focuses on bringing these studies together with a definition of a scheme for a benchmark data set. The second part of the thesis tries to find out how segmentation algorithms developed for animal trajectories work for human GPS data for classification. These two parts will be covered working on the following sub-goals:

– identify important considerations for the evaluation of transport mode classifiers
– construct and define a scheme for the collection of a reference data set
– apply the concept on a data collection campaign in Zurich (and Vienna)
– apply three segmentation algorithms
– compare the outcomes of each segmentation algorithm on three classifiers

Research Questions

– What are the requirements for an evaluation data set for transportation mode detection?
– How can such a benchmark data set be implemented?
– How does segmentation influence the outcome of classification?
– Is the classification on segmented data better than the point-wise classification?

1.2 Thesis Structure

After an introduction to the topic, Chapter 2 describes the background including necessary definitions, explores how the thesis is situated in the current research and attempts to identify the research gaps to be filled. Chapter 3 focuses on the project of building a reference data set conducted in collaboration with a team of the AIT. This chapter first explains the motivation and necessity of this project and how it fits into the research topic. In a next step, a theoretical background of the conceptual considerations concerning the actual data set and how it can be used for evaluation is given. A sub-chapter about protocol and methods gives an in depth view of how the collection campaign was conceptualized, organized and eventually carried out concentrating on which modes of transportation, modalities and devices were chosen, who participated and where the collection took place. The section set-up describes how the data was handled and processed to ensure a good organization.
Following the protocol, an exploratory analysis of the collected data is carried out and the results are compared to a similarly collected data set by Zheng et al. [50]. Here, a closer look is taken at the spatial distribution, the distribution of speed per mode of transport and the summary statistics of the collected data.

Chapter 4 builds on the data collected in the project described in the previous chapter and aims at experimenting with different algorithms for the segmentation of human movement data, followed by the classification on three statistical learning methods. The algorithms used are a segmentation by label, a segmentation based on first-passage time analysis and a Behavioral Change Point Analysis, followed by a random forest, a support vector machine and a K-nearest neighbors classification. The section background explains how the topic is of interest. It is followed by the scientific context of change point detection in movement data. The data which was used, is presented, followed by the implementation of the methods. In the sub-chapter results, the outcomes are shown and they are discussed in the last section of this chapter. To sum up the outcomes of the thesis, Chapter 5 draws a conclusion including a summary of the outcomes and ideas for potential future work.

2 Background

2.1 Definitions

2.1.1 Modes of Transportation

A mode of transportation is the way a person is getting from A to B. We make a distinction between active and passive modes. Active modes are considered walking, running, skating, riding a bike or a scooter whereas passive modes consist of public and private motorized transportation such as buses, trams, trains or undergrounds and cars or motorcycles respectively. The list of modes of transport can be extended almost infinitely and depending on the granularity of the chosen modes of transport, the performance of the classification varies strongly. The granularity explains how detailed the mode detection is carried out in reference to the modes of transport. A high granularity would include different types of cars or means of public transport, whereas a low granularity might group all motorized vehicles to one class.

2.1.2 Trip Segmentation and Modality

Looking at previous work, different definitions for the terms around the use of transportation are used without an overall, clear definition. For the purpose of this project, certain definitions were agreed on, based on the work of Nitsche et al. [30] and are explained in table 1. Many studies used the term "transition" or "transition period" for the time when a change of mode of transport occurs [11, 23, 37, 43, 47]. In this project, this term is not used, as this type of occurrence is annotated as a segment of type walking.

A trip can consist of one or more modes of transport which makes it either uni-, bi- or multimodal. In our definition, a trip is bimodal if it consists of two modes accompanied by a walking segment before the first mode, between the two modes and after the second mode.

Table 1: Definition of some important terms.

Term	Definition
Trajectory / Track	Sequence of GPS points leading to a segment
Tour	Total travel to a place and back
Trip	Travel from A to B, can consist of multiple modes.
Segment	Time window with the same mode of transport in a trip. A new segment starts with the change of mode.
Break/ Change point	Point where a mode change takes place. Divide a trip into segments.

2.1.3 Segmentation

As classification can be carried out point- or segment-wise, a definition of the term segment is necessary. For a point-wise classification, every point in a time-series is classified individually, whereas in a segment-wise classification, the data is split up in segments and a feature vector is calculated for each segment. These feature vectors are classified subsequently. A segment can either be chosen as a time-window or a segmentation according to the modes of transport can be pursued. A perfect segment includes all the points of the same mode of transportation in a row and its limits are so-called change or break points. These ideally occur exactly when the mode was changed. In this project, segments are considered to be only the partition by mode of transport or an approximation thereof and not by time-window.
In this project, change point detection and break point detection are used interchangeably and mark a step of the segmentation process. After having found potential break points, the data is split up or segmented at these points. Many algorithms have been developed for change point detection, not only for GPS trajectories but for any type of time-series data (as for example seen in computer vision and video processing [35, 17]).

2.1.4 Classification

Supervised learning algorithms can be divided up into two subcategories, regression and classification. While a regression problem has a quantitative response variable, the response variable for classification is qualitative or categorical [24]. As the task of transportation mode detection has the variable modes of transport as the response variable, we are working with a classification problem. To use a classification algorithm, two data sets are needed, one for training the classifier, another for testing it. The more data is available, the better the classifier can learn the patterns and the more it improves in recognizing the modes. Learning can be carried out supervised or unsupervised.

In a supervised learning problem, each observation has a response measurement for the predictor [24], giving the training data a ground truth which can later be used for evaluation. Working with humans, the advantage is, that they can be asked what they are doing and which modes they are using over a given time span. This way, a ground truth data set can be collected, where all the modes are known for the exact time windows. The aim is to fit a model which can accurately predict the response for the observations of the data used for testing [24].
In an unsupervised learning on the other hand, the response variable in the training data is missing. This makes the problem more challenging and bases the outcome on relationships between the observations or variables [24]. An example of a unsupervised learning problem is the distinction of different behavior in animal trajectories where the animal was equipped with a tracking device but there is no knowledge as to what it was actually doing while carrying the device.

2.2 Scientific Context

With the advent of small sensor technology and smartphones with built in sensors, the data collected with those, has become more and more accessible and can be collected by everyone without any disturbance of their everyday life and it seems only reasonable to use these to the advantage of transportation mode detection. Next to GPS traces, we can now use localization via WLAN or cell phone networks as well as acceleration measurements or gyroscope and magnetometer data. While these additional measurements may not help a single participant in remembering the modes they used, it opens up a field of opportunities for automated mode detection. Using classification algorithms, computer programs can learn the characteristics of each mode of transportation and assign the according labels to the data.
To be able to use the collected data, an automatic detection of the means of transport the person uses is of importance. This has led to a variety of studies concerning automated transport mode detection over the past 20 years by a multitude of researchers.
Depending on the focus of their work, different modes of transport, sensors, collection approaches and classification algorithms have been implemented. The overall goal of their work is to create a classifier which is able to automatically detect the correct mode of transport at any time as accurately as possible. While some focused more on the choice of sensors and how smartphones can be used in this context (see [37, 46, 45, 30, 23]) others had a stronger focus on detection with low power consumption [47] or on combining the sensor data with Geographic Information System (GIS)

information (see [40, 43, 44]). Overall, the focus of most recent studies lies on the classification, often leaving little space for their explanation of mode of transport choice or for which context the results of their classifier can later be used.
In general, a distinction can be made in the sensor selection and in the handling of the data chosen by the researchers. The sensors which are typically used are location and motion sensors. For the location the global positioning system GPS is used, examples of motion sensors are accelerometers, gyroscopes or magnetometers. Additionally, there are a few groups using other information as for example GSM, bluetooth or WLAN information. This results in rather low granularity and is only suitable for a distinction of grouped modes of transport such as public transport versus active transport [29].
Here, the related work will be divided into sensor differences of GPS, accelerometer and the combination of both and how they have been applied in research. For each section, a table shows the research groups, the modes of transport they considered and the overall accuracy of their results. This is a measure found in many studies, which sounds promising but does not make the approaches comparable. Depending on the choice and number of modes and the granularity of the respective clusters, achieving a high overall accuracy can be easier or more difficult. Training a classifier on three modes of transport with very distinct characteristics (for example walking, bicycle and car) is more likely to give good results than a comparison of 10 different modes, where some of them are very similar (such as distinguishing between cars and buses). However, overall accuracy is the only measure found in most publications.

2.2.1 GPS

Table 2: Studies using only GPS.

Authors	Modes of Transport	Overall Accuracy
Biljecki et al. 2013 [3]	Walk, bicycle, train, underground, car/tram/bus, boat (ferry and sailing), aircraft	91.6%
Bolbol et al. 2012 [5]	Walk, bicycle, bus, car, train, underground	88%
Byon et al. 2009 [7]	Car (arterial roads and highways), bus (arterial roads and highways), streetcar, walk	82%
Gonzales et al. 2008 [16]	Walk, bus, car	91.23%
Maeenpaeae et al. 2017 [25]	Walk, bicycle, bus, car	88.05%
Schüssler et al. 2009 [38]	Bike, car, urban public transport, rail	-
Stenneth et al. 2011 [43]	Walk, bicycle, bus, car, stationary, above ground train	93.42%
Tsui and Shalaby 2006 [44]	Walk, bicycle, bus, auto	91%
Zhang et al. 2011 [48]	Walk, bicycle, car, bus, tram, train	93%
Zheng et al. 2008 [51]	Bicycle, bus, car, walk	-

GPS was the first choice of sensor for automated transportation mode detection. From its components, interesting features such as speed, change in heading an even acceleration can be calculated. With the first portable GPS devices developed in the late 1990's and early 2000's a large new field of research possibilities was opening up. Many publications [3, 5, 7, 16, 25, 38, 43, 44, 48, 52] used only GPS measurements to implement their learning algorithms. Using GPS information has the advantage that the collected data not only gives information about the speed and the modes which are conducted but also shows the tracks of people. This way, not only the question of how people travel can be answered but knowledge about where can be collected as well, making it interesting for further analysis in transportation science. GPS trajectories are widely collected and used and have proven to be very well suited for transport mode detection.

Table 3: Studies using GPS and accelerometer.

Authors	Modes of Transport	Overall Accuracy
Das, Winter 2016 [9]	Bus, train, tram, walk	-
Ellis et al. 2014 [11]	Bike, riding in vehicle, walking, sitting, standing	92%
Feng et al. 2013 [13]	Walk, cycle, run, motorcycle, bus, coach, car, lightrail, metro, tram, train	95%
Manzoni et al. 2010 [26]	Walk, cycle, bus, car, motorcycle, metro, train	82.14%
Martin et al. 2017 [27]	Walk, bike, car, bus, rail	96.80%
Nitsche et al. 2012 [30]	Walk, bike, motorcycle, car, bus, electric tram, metro, train, wait	79%
Reddy et al. 2010 [37]	Stationary, walk, run, cycle, motor	94%
Shah et al. 2014 [40]	Non-motor, car/bus, train/rail	90%
Widhalm et al. 2012 [45]	Walk, cycle, motorcycle, car, bus, tram, aboveground train, subway	-
Xia et al. 2014 [46]	Walk, bike, motorized transport, stationary (2 subtypes: stay + wait)	96.31%

The disadvantage of GPS data is the dependency on an unobstructed view on the satellites [23] and the high power consumption for the collection device. If there is no signal available, no data exists to classify. This is especially problematic in indoor environments and tunnels.
Different foci are seen in the literature, ranging from the sample size and collection frequency [5], the unsupervised classification of data without labels [üssler2009] or the concentration on segmentation over classification seen by Biljecki et al. [3]. Each group of scientists has collected their own set of data with the modes of transport they deemed most important as seen in table 2.
The studies used a variety of classifiers, ranging from expert system approaches [3], to decision trees [25] and support vector machines [48, 5]. The overall accuracies calculated from their results are between 82 and 93 percent.

2.2.2 GPS and Accelerometer

Referring to the argument that retrieving GPS information is rather power-consuming and can be erroneous due to urban canyons or tunnels, some groups [13, 26, 37, 9, 11, 27, 30, 45, 46] decided to combine the GPS with accelerometer data as seen in table 3. While

Feng et al. and Manzoni et al. [13, 26] detected a wide range of transportation modes dividing motorized transport into different categories, Reddy et al. [37] concentrated on only five (stationary, walking, running, cycling and motorized transportation). Ellis et al. [11] on the other hand, concentrated their work more on the active modes of transport, coming from a background of physical activity research. Using acceleration measurements with this focus, allows for more detailed distinction concerning very similar modes of transport such as different types of walking.
Coming from a transportation science background, Das et al. [9], Nitsche et al. [30] and Widhalm et al. [45] are more interested in the modes of transport traditionally used in urban traffic, such as public transportation, walking, cycling and driving a car.
The classification algorithms used included a Bayesian Belief Network [13], decision trees [26] and a combination of decision trees with a first order Hidden Markov Model [37]. The overall prediction accuracies were similarly high as in the studies using only GPS data, ranging between 82% [26] and 95% [13].
The clear advantage of using more than one sensor is the riddance of gaps in the track, giving more information about movement in indoor environments. The combination with GPS data allows geographic localization, which leaves the opportunity of adding geographic information for potentially better classification. GPS can be collected at a lower frequency to decrease energy consumption, as the acceleration measurements make up for this lack with a very low need for power.

2.2.3 Accelerometer

Table 4: Studies using only accelerometer.

Authors	Modes of Transport	Overall Accuracy
Hemminki et al. 2013 [23]	Walk, bus, car	80%
Shafique et al. 2014 [39]	Walk, bicycle, car, train	99.8%
Yu et al. 2014 [47]	Stationary, walk, run, bicycle, vehicle	90.66%

As summarized in table 4, Hemminki et al. [23], Shafique et al. [39] and Yu et al. [47] argue that acceleration measurements are sufficient for inferring modes of transport. As [23] states the advantages as very low power consumption, the independence of any external signal sources, resulting in a continuous transportation behavior monitoring, and the highly detailed information about phone movement. Additionally, the privacy concern of respondents can be reduced, as the location information is not tracked [39].

In comparison to other work, the groups using only accelerometers seem to have smaller sets of modes of transport. While [23] only distinguishes between three modes of transport (walk, bus and car), [39] leaves out buses but adds bicycles and trains, whereas [47] adds stationary and groups all passive modes into one mode of transport called vehicle. This simplifies the inference considerably, as vehicles such as bus and car, which tend to show very similar characteristics and are difficult to distinguish, are grouped. Whether or not this explains the rather high overall accuracy of the results, cannot be found in the literature.
The classifiers applied to these data sets were support vector machines [47], a random forest approach [39] or decision trees in combination with a first order Hidden Markov model [23].

3 Building a Reference Data Set

3.1 Motivation

As seen in Chapter 2.2, a number of research projects regarding transport mode detection have been conducted using various sensors and methodologies. What all of them have in common is a lack of comparability with regard to the data used for the classification performance and how their results were evaluated. In each publication, new algorithms have been developed and tested with little consideration of what has been done before and how it could be improved. For this topic, there is no standardized evaluation procedure and no publicly available benchmark data set. Most of the data was collected by very few people (e.g. [11, 43]) or documented insufficiently like in the work of Zhang et al. [48], where there is no information concerning the time frame of the collection or the number of participants involved. The most common approach towards building a classifier is using a subset of the collected data for training and another for evaluation. To obtain unbiased evaluation results however, the data used for evaluation should be independent from the training data, depending on the aims of the classification. If the classifier is built for a specific city or situation, data from the same background can be used for training and testing, making sure that data of different users is taken to avoid overfitting. Independent data can be achieved for example by a standardized data set for testing. Various studies are implying a lack of suitable ground truth for both training and testing of mode detection algorithms [41, 25, 36]
To improve this situation, the establishment of a common reference data set used for the evaluation of the classification based on the data the researchers themselves have collected is a crucial step. Therefore, this project, a collaboration of University of Zurich (UZH) and Austrian Institute of Technology (AIT), aims at identifying important considerations for the evaluation of transportation mode classifiers and providing a general guideline for data collection. Additionally, we have collected such a reference data set and will make it publicly available for further use, together with all the general requirements and guidelines. By doing so, we hope to encourage the community to compare the performances of their data and algorithms based on our benchmark data set.

3.1.1 Collaboration with AIT

For this project the department of Geography at the University of Zurich collaborated with the Austrian Institute of Technology. The AIT is Austria's biggest Research and Technology Organization (RTO) and works on many projects concerning infrastructure worldwide [31]. As a RTO, it is a good development partner for the industry while also

working closely together with the academic world [31].
One of its research groups, Mobility Data Collection and Analysis, has been working on projects concerning transportation mode detection over the past years [45, 30] giving them experience in dealing with and collecting motion data with smartphones. As we found our common goal of collecting a benchmark data set for the evaluation of transport mode detection, a collaboration emerged, combining the interests and knowhow of the two research facilities. Namely, Oliver Burkhard, PhD student, and I, master student, both from University of Zurich (UZH), were working with Peter Widhalm, senior scientist, and Maximilian Leodolter, PhD student, from AIT.
The collaboration benefited not only from the combined knowledge and the exchange of ideas, arguments and points of view but also from the outcome of the data collection. Together, a framework for an evaluation data set was established, including all the necessary considerations regarding modes of transport, modalities, devices and the set-up of the collection campaign. In a next step, a set of evaluation metrics will be put together. Being able to collect in two large European cities broadens the applicability of the outcome. The two cities Vienna and Zurich differ in size, population, topography and transport system which makes them interesting for comparison and gives data, which is less prone to overfit for the particularities of one city. This variety can be improved in a further step by other research facilities adding their data sets following the same scheme. As Sim et al. [42] state in their research concerning the usability of benchmarks, opinions and experience of different parts of the community are necessary to understand what is needed of a benchmark. The content of the following sections will be part of a publication, which is currently in the working and includes considerations and ideas by all of the people involved.

3.1.2 Benchmark Data Set

A benchmark is defined in the Oxford English Dictionary as "a standard or point of reference against which things may be compared" [34].The term benchmarking was first used in the context of computer science, where methods were developed to make the performance of computer systems, information retrieval systems and related technologies comparable [42]. In this research field, the interest lies in optimizing the performance of computational power or speed and benchmarks come in the form of software and definitions of standards. Defining a benchmark is overall a community effort, which is necessary to validate research results and compare the work of different research groups or facilities [42]. This is of interest in many fields of research as techniques can only be improved by a research community if they are reproducible, accessible and comparable. Sim et al. define a list of requirements for successful benchmarks in their paper on "Using benchmarking to advance research: a challenge to

software engineering" including accessibility, affordability, clarity, relevance, and scalability [42].
Looking at the field of transport mode detection, the time for developing a benchmark to evaluate the results which have already been accomplished or are yet to be found, seems right. The field has been established over the past 20 to 30 years and has had time to proliferate diverse approaches and solutions leaving space to experiment. According to [42], this time is necessary to develop a variety of tools and techniques which can later be compared. In our case, a first step in benchmarking will be done by establishing the frame work for a reference data set for the evaluation of classification algorithms used for transportation mode detection and collecting two such data sets.

3.2 Scientific Context

As seen in the scientific context of transportation mode detection, many groups have studied the ways of automating the process. In this section, a closer look into the data used in these studies is taken to illustrate the need for a benchmark data set for transport mode detection.
In recent studies, the data needed for classification is mostly collected individually by the research groups or by people instructed for the task [7, 13, 27, 50, 46]. While most groups annotated the data they used for training and testing, making the task a supervised learning problem, Schüssler and Axhausen [38] decided to use already available GPS data without any labels. As their outcomes could not be evaluated using available ground truth, Schüssler and Axhausen decided to use the Swiss Micro Census for mobility and transportation for evaluation. The Swiss Micro Census is a collection of statistical data, carried out every five years, about the travel behavior and reasons for travel in the Swiss population [33]. The advantage of their work is, that they were able to collect longer trip chains and full travel days without burdening the participant with trying to remember what they did at which point in time [38]. The disadvantages, however, are, that there is no suitable way of evaluating the results, as there is no ground truth available for comparison. In their study, Schüssler et al. [38] used the micro census as a basis to validate their results, revealing a similar trend in the micro census and the GPS study. This might be an interesting approach for large GPS data sets without labels but cannot cope with the accuracy of a supervised classifier. Potentially, a classifier trained on annotated data from the same city could be used to carry out the classification of the data with better comparable outcomes.
The advantage of collecting your own data is the reliability of it, since it is collected in controlled settings with the features and specifications most important to the re-

spective study. Such specifications include choice of sensors and devices, modes of transport, device handling and how to annotate the collected data. Classification of transport modes is a typical supervised learning problem which rises the need for well annotated, controllable ground truth. Since transportation systems and vehicles vary between cities and countries, individual data sets for certain settings can be appropriate.

However, there are disadvantages to the individual data collection as well. Many reported studies are using a subset of their test data for training (such as [50, 16, 25, 27]), which is prone to overfitting and should be avoided by using independent test data. The documentation of data collection methods and execution is incomplete in many cases [26, 27, 13, 7], lacking important information regarding e.g. the amount of data per mode of transport or the people who collected the data.

Additionally, as stated in a review of transport mode detection approaches based on smartphone data by Nikolìc and Bierlaire [29], most data sets and sample sizes used in mode detection studies are comparably small, which questions the statistical significance of the results.

Taking a closer look at six examples of transport mode detection studies [7, 11, 13, 27, 52, 46], difficulties found in the employed data sets will be shown and solutions considered.

Byon et al. [7] use a data set consisting of 60 hours of GPS data, divided up into 15 hours per mode of transport collected in the Toronto region. The modes of transport used are car and bus (both in downtown areas, arterial roads and highways), streetcar and walking. The data collection was carried out in strict setting with prescribed routes recorded at different times of the day to get data for rush hour as well as regular traffic. In the publication, no specifications on who collected the data, what type of GPS loggers were used or how the loggers were handled during the collection, are given.

A research group from the University of California San Diego [11] on the other hand, documented everything meticulously. 150 hours of GPS traces and acceleration measurements were collected by two trained research assistants in San Diego. Each person was carrying 13 devices, 12 of which were attached to wooden boards, carried in a backpack, to ensure consistent orientation. It is unclear, whether the 150 hours refer to absolute collection time of all 13 devices summed up, or whether the two research assistants spent 150 hours collecting. The 13th device, an accelerometer, was attached to the hip of the participant. Since Ellis et al. [11] come from a public health background, the modes of transport they chose (bike, vehicle, walking, sitting and standing) give more detailed distinction between active modes than passive modes. The protocol is well documented and explains all of their considerations.

In the third example, Feng and Timmermans from the Eindhoven University of Technology [13] have recorded GPS traces for a variety of urban modes of transport (walk, cycle, run, motorcycle, bus, coach, car, light-rail, metro, tram, train) in London with a frequency of 1 Hertz for GPS and 10 Hertz for acceleration, with an accelerometer-enabled GPS device called MobiTest GSL. While they give a number of available data records for training, no specification on the amount of data collected per mode or how the devices were carried is given. The data was collected by a small number of volunteers and annotated via activity-travel diary.
In contrast to the strict settings of [11], Martin et al. [27] collected data in a closer to real-life situation, letting six students carry their phones in ordinary ways (in their hand, pocket or bag), recording GPS (at 1 Hertz frequency) and accelerometer (at 5 Hertz frequency) measurements. About 100 hours of annotated data in the modes of transport walk, bike, car, bus and rail was collected.
A much bigger amount of data was collected by the Chinese research group of Zheng et al. [50] who work with Microsoft Research Asia. They asked 65 people to record GPS traces over a period of 10 month. Most of the data was collected in China with some additional trajectories in the United States, South Korea and Japan. Each participant was payed per kilometer, a total of over 2000 hours of data were collected and annotated. The data is well documented and shows the duration distribution of the used modes of transport (bicycle, bus, car, walk).
The last example of a data set collected solely for the purpose of automated transport mode detection was conducted by a group from the Institute of Remote Sensing and Digital Earth in Beijing [46]. In this case, 12 minutes per mode of transport (walking, biking, motorized transport and stationary) were recorded with smartphones carried in the participant's jacket pocket to avoid the influence of arm motions on the sensors. GPS was sampled at 1 Hertz and acceleration at 50 Hertz.
These six examples were chosen from a large number of reported studies working on the topic of transport mode detection as they show the differences in data acquisition very well, ranging from choice of mode of transport and sensors to amounts of data and recording frequencies. Comparing the results of these studies is difficult because all of them used different data with different modes of transport and different evaluation methods. This makes it complicated to say how well an employed algorithm would work with independent data, even if the independent data was recorded in similar settings as proposed in the study for the training data. Having a benchmark test data set would give the overall research area more consistency, making results comparable and therefore making space for the improvement of already existing algorithms for segmentation and classification.
Similar projects of developing a benchmark data set have been conducted in very dif-

ferent fields of video object segmentation by Perazzi et al. [35] and movement change detection by [17]. As in the case of transport mode detection, a diversity of data sets were available and in use, but a standardized and widely adopted evaluation methodology was missing [35]. Perazzi et al. stress that the data needs to be of very high quality, reliable and well-annotated. Additionally, it should be publicly available to ensure its use in the community to achieve their goal of providing researchers with the tools to evaluate their methods and advance research in the field of video object segmentation [35]. Goyette et al. [17] mention the same problems, adding the fact that the use of own data by research groups makes fair comparison almost impossible. Even though their research differs from the topic of this work, all of these observations can be applied here as well, which leads to the conceptual considerations for a benchmark data set in transport mode detection.

3.3 Conceptual Considerations

Inference of transport modes can be of interest for various applications, leading to a range of factors, which should be considered in a benchmark data set. If the transport mode detection is applied for a mobile application, for example, the requirements might differ from the requirements a transportation scientist has for his data. This calls for a data set which covers a variety of situations and applications without concentrating too much on the claims of a single user. As the aim of this project is to build a scheme for an evaluation data set and later a set of corresponding evaluation metrics this should be taken into consideration. Additionally, the scheme is to be applied to two data collection initiatives in Vienna and Zurich. For a well-balanced reference data set as close as possible to a real-world scenario, a variety of ideas have to be collected and weighed up against each other.
Transport modes in the real world do not generate straight forward location or motion signals but are connected to many superimposed factors (table 5). Depending on the chosen mode these can include driving style, road condition (e.g. pavement, speed limits), traffic conditions (e.g. urban traffic, traffic jams, open road, highway), device handling (how the participant interacts with the device during the trip and how the device is carried) and for the active modes the health of the participant, which determines how fast they walk or react and what the purpose of the trip is (e.g. strolling at a leisurely pace versus hurrying towards a destination). These superimposed factors are not random but correspond to specific traits of the data generating process and need to be taken into consideration when planning a collection campaign. All of these factors should be varied to avoid correlation with the transport modes to recognize. The best

Table 5: Factors.

Factors	Individual Motorized Transport (IMT)	Public Transport	Active MOT
Driving style	x	x (of driver)	
Road condition	x		x (bike)
Traffic condition	x	x (bus)	x
Device handling		x	
Age/Health			x
Trip purpose			x

case scenario is an independent variation of all factors, to obtain a data set with all possible combinations for a validation data set.

The same is relevant for the duration and the modality of each trip. A trip should consist of differing amounts of segments which should have varying durations. As a balance needs to be kept, predefined trips include (among others) these factors. The necessary durations and number of segments depend on the context of the work. With a focus on transportation in urban environments, shorter durations and changes between modes of transport are of interest but as a benchmark data set the data should include content for variable context on different levels [35]. A choice has to be made, whether or not to include trip chains, meaning longer trips consisting of multiple modes with possible breaks, potentially over the course of a whole day.

The next important consideration is the choice of transportation modes. This is, as well, highly dependent on the application context and has shown different directions in published research. Prelipcean et al. [36] have divided up the contexts in three groups, location-based services, human geography and transportation science. The first group incorporates a close to real-time detection as for example seen in Reddy's et al. work [37] where the goal is the information of the user about the real time location of relevant buses. The use of annotated trajectories in human geography lies in enriching the data with domain specific semantics such as recognition of patterns within a person's day to day behavior [36]. An example is presented by Ellis et al. [11] who focused on the recognition of physical activities, searching for possible connections between everyday behavior and obesity. Das et al. [9] on the other hand, were interested in understanding travel behavior with the background of urban planning, focusing more on the interests of transportation science. A standardized evaluation scheme needs to be flexible enough to be of use for different taxonomies and still enable comparability between the approaches. Even with a concentration on one of the three mentioned topics, a big span of possibilities for modes of transport choice is left. Additional to the chosen transport modes, subtypes of modes need to be taken into consideration. Examples are types of cars, a differentiation between bicycles and e-bikes or different peculiarities of walking. These traits do not need to be distinguished for most applications but should be varied in a data set to avoid overfitting for a particular type.

Looking at the choice of transport modes and their subtypes, one can see that these may strongly vary not only in different contexts but also in different settings, for example being on a regional or country scale. Urban traffic has other traits than transportation on the country-side and public transportation may be organized in another way in different countries. Another difference can be found in the topography, the size and the road conditions of a region. While most European roads might offer good infrastructure and easy advancement, cities in other countries might experience much slower traffic with rather bumpy roads. Often, a classifier is calibrated for a certain region, as it will never be used globally. An evaluation set however, should cover as much variety as possible. Therefore, this project offers data collected in two cities and gives the option of adding more data, collected following the same scheme, in other regions and gives the potential to adjusted it, to fit the particular context of a study. The first two collection campaigns serve both as first possibilities for classification evaluation and as exemplary data for further campaigns.

To ensure that all the influences on the recorded signal listed above are met, the data collection needs to take place in strictly controlled settings. A protocol has been developed defining combinations of patterns including the modes of transport, the duration of the trip segments, which subtypes are used, the carrying positions and how the participant is to behave during the trip. The factors in this protocol are limited to a manageable number and can be expanded according to need. As the collection was carried out in controlled settings, scripts need to be written, giving clear, yet simple instructions. To ensure enough variety and diversity in the data, some aspects need to be left to the participants and the circumstances (e.g. unforeseeable traffic jams, running to catch a tram or an abnormally long stop in a bus), to get data which represents real-life as closely as possible, yet as controlled as necessary.

The protocol needs to be documented accurately, to keep the outcome reproducible, reliable and usable. The same is true for the data collected in the campaigns. All parameters should be documented as precisely as possible and their benefit for the benchmark as a whole should be explained to motivate other research groups to participate in the project of making transport mode detection approaches more comparable.

3.4 Protocol and Methods

The collection campaign follows a strict protocol defining the modes, modalities, device handling, position and number of devices, as well as a regional constriction to collect the data. This was constructed with the aim of building a close to real-life, balanced benchmark data set which allows to evaluate and compare methods which are designed to technologically support the collection of a region's mobility behavior. As in any data collection campaign, compromises had to be found between an ideal data set and a reasonable and accomplishable solution. The specifications are given in the following sections.

3.4.1 Modes of Transport

The focus of the modes of transport is set on the topic of transportation science with the aim of collecting a data set suitable for the needs of transportation scientists in urban environments. Their interest lies in generating reliable statistics on how people use transportation means in their daily life [36]. With suitable algorithms, a modern form of travel surveys can be achieved, which facilitates further data collection and research in this field. The requirements for modes of transport in this field is suitable for many other applications such as mode detection for location based services or for emission calculations. The modes of transport were chosen accordingly, focusing on public transportation, motorized individual transport and the active modes walking and biking. For public transport all available means for both cities were chosen with the exception of boat travel. This mean of transport was decided against because it is mostly used in a more leisurely fashion and not to reach places in day to day activities. Looking at the motorized individual transport, it was decided to narrow it down to collecting car data. This is the most common means of transport and widely used in both cities. It could be argued that motorcycles should have been added and it is imaginable that this type of data could be added in a further data collection. The last common way to get around are the active modes walking and biking. In European cities riding bicycle has been gaining popularity as mentioned by Menghini et al. [28] and is encouraged in current plans for urban travel behavior change, making it an important research topic for urban mobility [28]. Walking has two basic functions, first of all as a mode of transportation to get to a place and secondly as a transition period between two stages of a trip. It was decided not to give a separate label to stationary transition periods as it has been done by [37, 23, 40, 43] but to label transitions as a walking segment. This is reasonable, because transitions are often connected with a short walk, e.g. to the next public transportation stop or from the stop to a car. A transition period does not necessarily have to differ at all from a walking segment and

detecting the mode as one, shows more clearly, where walking is involved. This might be of use for example for the planning of public transportation systems, to realize, how much time is necessary to get from one vehicle to the other.
For more specific tasks such as physical activity recognition, data for other activities such as skating, running or hiking can potentially be added at a later stage.

3.4.2 Modalities

The data was collected in two different types of scripts. Unimodal scripts include only one mean of transport, whereas bimodal scripts include two means plus walking segments before and after the trip as well as in between the two means of transport. Technically, this could be regarded as a multimodal trip, but as walking is the most common thing people do when changing modes, we do not count it as a mode in this case. If one of the two modes is walking, however, it is not initiated or ended by another walking segment. This decision was made to keep the variety of trip durations and modalities without having the scope get out of hand. Unimodal scripts provide the possibility to collect longer periods of time with clearer instructions and therefore more control.
In the unimodal scripts the participant was asked to behave in a certain way, giving us a controlled setting to see how specific movements, velocities or road conditions influence the recorded signals. During the bimodal trips on the other hand, the participants had to concentrate only on using the right mode in the right time frame and annotating their trips exactly. How to behave during the modes was left to the participant to broaden the variety of behaviors in the data. Using this scheme, it was possible to keep the complexity for the participants bearable, yet keeping the goal of a close to real-life data set in mind.

Unimodal Scripts

Table 6: Unimodal Scripts.

Public Transport	
Means of transport	Metro
	Bus
	Tram
	Train
	Commuter Train
Start of journey	Sitting
	Standing
Instructions	Change your seat once
	Stand up
	Sit down on an empty seat
	Walk around for 1min
Active MOT (bicycle + e-bike)	
Means of transport	Bicycle
	E-bike
Instructions	Drive fast without any stops
	Drive in an urban environment
	Drive on a gravel road
	Drive uphill
	Drive downhill
Active MOT (walking)	
Means of transport	Walking
Instructions	Sustained harmonic walking
	Running
	Strolling
Motorized Individual Transport	
Means of transport	Car diesel
	Car petrol
	Electric Car
	SUV
Instructions	Drive on highway (min 100km/h)
	Drive in traffic jam
	Drive in an urban environment
	Drive between 60 and 90km/h

For all unimodal scripts a duration of 10-15 minutes was decided. This corresponds to an average trip time in a real-life situation. Similar values have been used for example by Xia et al. in 2014 [46]. The means of transport were divided up into public transport, active modes bicycle and e-bike, active mode walking divided up into harmonic walking, strolling and running and motorized individual transport. For each of these, instructions were formulated covering everyday situations. For the active modes and the motorized individual transport the focus was mainly set on speed and the environment, making sure to include all types of behavior seen in everyday life. In public transport the behavior in the vehicle is changed up to document these influences on the signal and to make a potential classifier attentive to such issues as they would occur in real life. The combinations are shown in table 6. The participants were asked to start the recording when already conducting the mode and end it before stopping. An exception were the bicycle trips, where the trip was started shortly before starting to ride to avoid dangerous situations for the participants. For the trips conducted in cars, participants were asked to pair up to ensure save driving without distractions.
Using all the possible combinations, a total of 51 unimodal scripts were written and for Zurich, 45 were later conducted, leaving out the mode of transport subway as there are no subway lines available in Zurich.

Bimodal Scripts
Next to the unimodal scripts, a number of bimodal scripts were written and collected. Since people often use more than one mode of transport to get somewhere, it is appropriate to collect reference data for multimodal trips as well. It was decided that the focus on our data set would be set on bimodal instead of multimodal trips, adding walking elements at the beginning, in between the modes and at the end of the trip. This is suitable because this combination of two modes with walking segments in between is very often conducted for getting to and from every day tasks. Two approximate durations were chosen. Either a segment is classified as short (trip can take between one and five minutes) or as long (with trip durations of longer than 5 minutes). As a strict specification would have been impossible to keep and a variation was in our interest, this is a good compromise. These requirements keep the task manageable and covers common real-life situations when moving around in an urban environment. For the later classification it is also of great importance because the algorithms do not only have to be able to detect which mode was conducted but also the breakpoints where the mode was changed. With the addition of the walking segments, a trip consists of up to five segments. An exception are the trips were walking is one of the two means of transport and is not introduced and followed by another walking segment. These trips consist of three segments only. This gives us another variation in the data set. In

table 7 the means of transport and the durations of each trip-leg can be found. A few combinations were removed because of their uncommon nature, namely bike-bike and walk-walk in all combinations of durations. Depending on the transport system of a city or region, the scheme needs adjustment. For Zurich for example, all combinations with the mode of transport subway were not carried out. This left us with a total of 102 bimodal scripts.

Table 7: Bimodal Scripts.

Means of transport	Metro
	Train
	Bus
	Tram
	Commuter Train
	Bicycle
	Walk
	Car
Duration per mean of transport	Short (1-5min)
	Long (>5min)

3.4.3 Devices

For all scripts, the collection was carried out using three devices, two smartphones and one uTrail tracking device. The master device where the data was annotated is always carried in the participants hand (or in a device holder or the back pocket of the participant's pants when bicycles are used). The participants are allowed to interact with their master device during the trips as they would during a regular train ride or walk to give set of data as realistic as possible. The first of the slave devices is carried in the front pocket of the participant's pants and the other in their backpack or bag. Whether the first was a uTrail or a smartphone is noted and changed up for the different participants to give a balanced result. Using this scheme, we expected to get a variation in sensor signals for the three most common carrying positions.

For the data collection, two different types of devices are used. First, a custom made device for collection sensor data called uTrail Personal Tracker and second, various types of smartphones where the data was accessed using an application developed by the AIT [31]. Both devices collect GPS (at a frequency of 1 Hertz) as well as Inertial Measurement Units (IMU) data (at a rate of 50 Hertz). The IMU data includes accelerometer, magnetometer and gyroscope measurements. The frequencies were chosen according to the state of the art method seen in most reported studies. Since the frequencies are rather high as often seen in recent studies [46, 45], the possibility of lowering the frequency remains (as suggested by [25]). This way, different experiments with the data can be carried out. The focus of the IMU data is on acceleration data as this has proven to be a good descriptor of transportation modes [48, 10, 23]

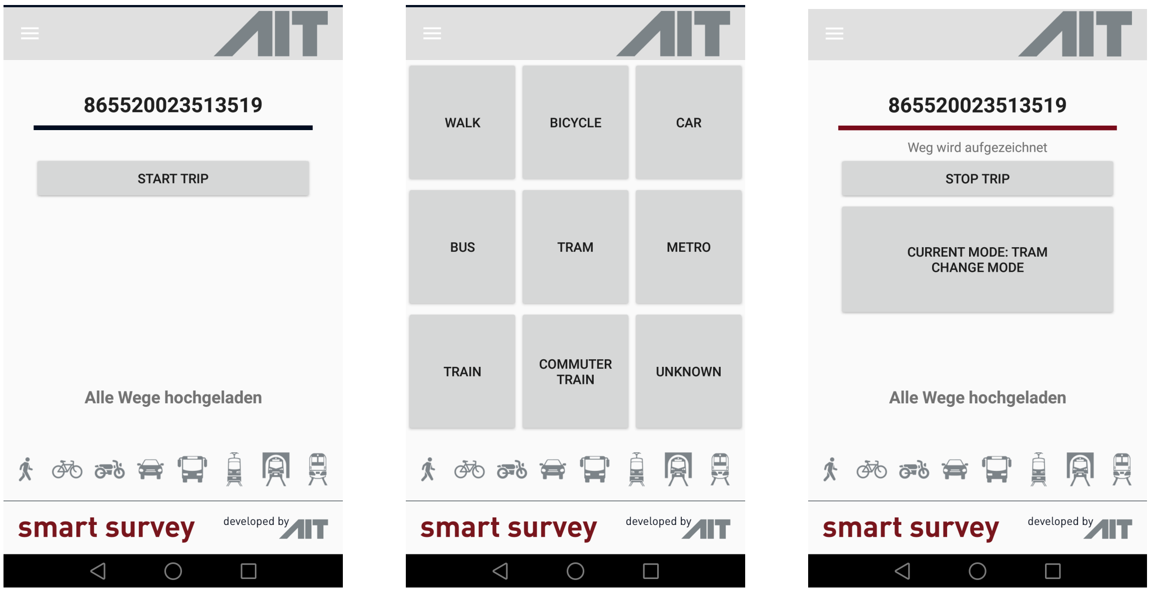
uTrail Personal Tracker

Figure 1: uTrail Personal Tracker device.

The uTrail device (figure 1) is a mobile sensor developed for the Geographical Institute at the University of Zurich (GIUZ) to be used in a variety of research projects. It measures GPS points, accelerometer, magnetometer and audio recording data. Since audio recording was not required in this task, it was deactivated.
In addition to its use in this project, the uTrail is the main data collection device in the Mobility, Activity and Social Interaction Study (MOASIS) conducted at GIUZ [2]. The advantage of the device lies in its long run-time (with a GPS frequency of 1Hz up to 1 week battery life) and a 6-axis accelerometer and magnetometer with a frequency range of 1-200 Hz [32]. Additionally, the device requires no prior technical prerequisites by the participants [14]. The uTrail devices give us the possibility to collect data as an addition to the smartphone data. uTrail data can be extracted from the device using a USB connection and custom made software (Personal Tracker Control Center [32]).

Smartphone
Next to the uTrail devices the sensor data of smartphones running on the android operating system was used. As each participant was asked to carry two smartphones at the

time, one was usually the participant's own device whereas the other was provided by the psychology department of UZH.
To access this data and store it according to our needs, the application AIT Smart Survey was used. This application has been developed by AIT for web-based transport mode detection and has been adapted for the purpose of this project [6]. The application gives the user the option to record trips and annotate them with the respective label per mode of transport as seen in figure 2. Additionally, the users are asked to write comments, helping with exact labels or potential mistakes they made. Figure 3 shows the steps necessary to collect a trip.



(a) Step1 (b) Step2 (c) Step3

Figure 2: AIT Smart Survey Application Home Screen.

3.4.4 Participants and Collection Areas

As there was a need for well-instructed, reliable participants, the number of potential participants was limited. Around ten people from each research facility (GIUZ and AIT) were recruited. This way, we were able to give careful instructions concerning the device handling, how to plan the trip and how to carry it out and to ensure well-labeled data. The relatively small number of participants is a compromise of collecting as clean as possible and losing part of the generalization, as we realized how difficult it can be to follow the scripts exactly without proper instruction. The participants were not chosen according to any special traits such as age or physical activity but are all mobile and in good health. For specific research fields such as the connection of age

and mobility as seen in the MOASIS project at GIUZ [14] the range could potentially be adapted.
The data was collected in the two urban regions of Vienna and Zurich. The focus was set on the urban areas, leaving city grounds only for longer car or train trips. This reflects the day-to-day travel behavior of a person living in or close to a city, potentially commuting to and from it for work or in their free time. This setting was chosen as mobility research is often conducted in urban areas where the network is very dense and the organization of the infrastructure is challenging.

3.4.5 Set-up

As we were using three devices per participant per scripts, data collection and storage was a challenge. The AIT group provided a smartphone application for the participants, a backend to ensure efficient uploading to the server, organizing the data and conducting a plausibility check and a frontend to validate whether the participants and we were content with their collection, meaning whether or not the standards concerning the accuracy of the time frame, the modes of transport and the device handling was as high as possible. This is visualized in figure 3. For each trip, the data is collected using three devices. On the master smartphone, the AIT Smart Survey is running and the participant is asked to add labels during the trip and to write the identification number of the script into the first comment section. The other smartphone is running the same application but in an unknown transportation mode. The third device, a uTrail, is turned on at the beginning of the collection period and records continuously. Data from the AIT Smart Survey is uploaded to the server directly whereas the data from the uTrail devices has to be extracted manually and uploaded to the server separately. All three sets of data are matched using the start and end time of the master device. This way, the labels can be applied to all three sets and they all have the same length. Once the data is uploaded and clipped, the pieces are converted to Comma-separated Values (CSV) format using a file for each, GPS traces and IMU measurements, for every device. Each file is put into a folder matching the script it belongs to and all the master files are checked for plausibility.
The plausibility check includes a number of hard and soft constraints. A violation of a hard constraint occurs if there is no script ID or the number or order of the transport modes does not correspond to the script ID. Soft constraints cover cases where the duration of one or both trip segments does not correspond to the requirements in the script and whether there are more than the one necessary comment for the script ID. The check has three possible outcomes, ”keep”, ”repair” and ”discard”. If both the soft and hard constraints are met, the category of the file is ”keep”. If only the soft

constraints are violated, it is categorized as ”repair” and if either the hard or both soft and hard constraints are violated, ”discard” is assigned. The result of the plausibility check is saved as a separate file in the script folder.
After all the files are in their respective folders the participants are asked to validate their collected data in a web application (figure 4). After gaining access to their data via user-assigned password, each trip can be viewed and a decision as to what should happen to the trip can be met. If the plausibility checked has resulted in ”keep” and the participant agrees, this option is chosen. Another option is for the participant to check ”repair” and to add a comment stating what is wrong with the trip. This way, simple mistakes as e.g. having clicked on the wrong transport mode in the smartphone application can easily be fixed manually. If the trip has violated constraints and cannot be fixed, ”discard” is chosen.

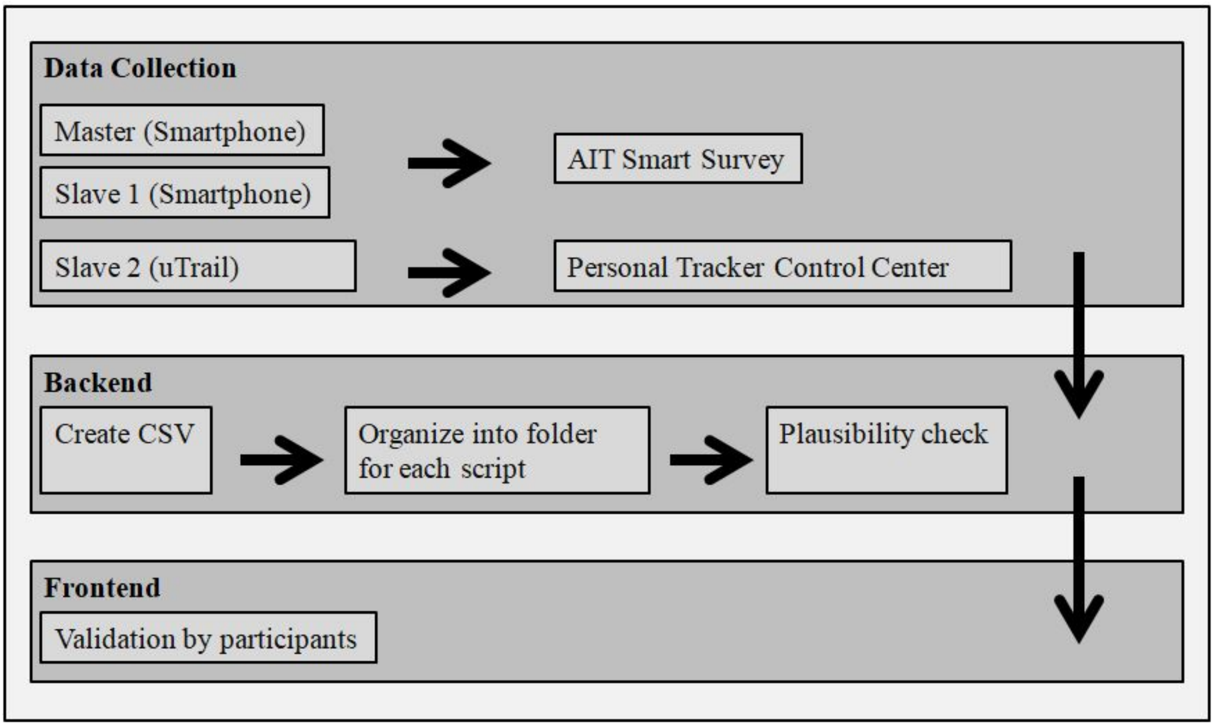


Figure 3: Workflow of data collection.

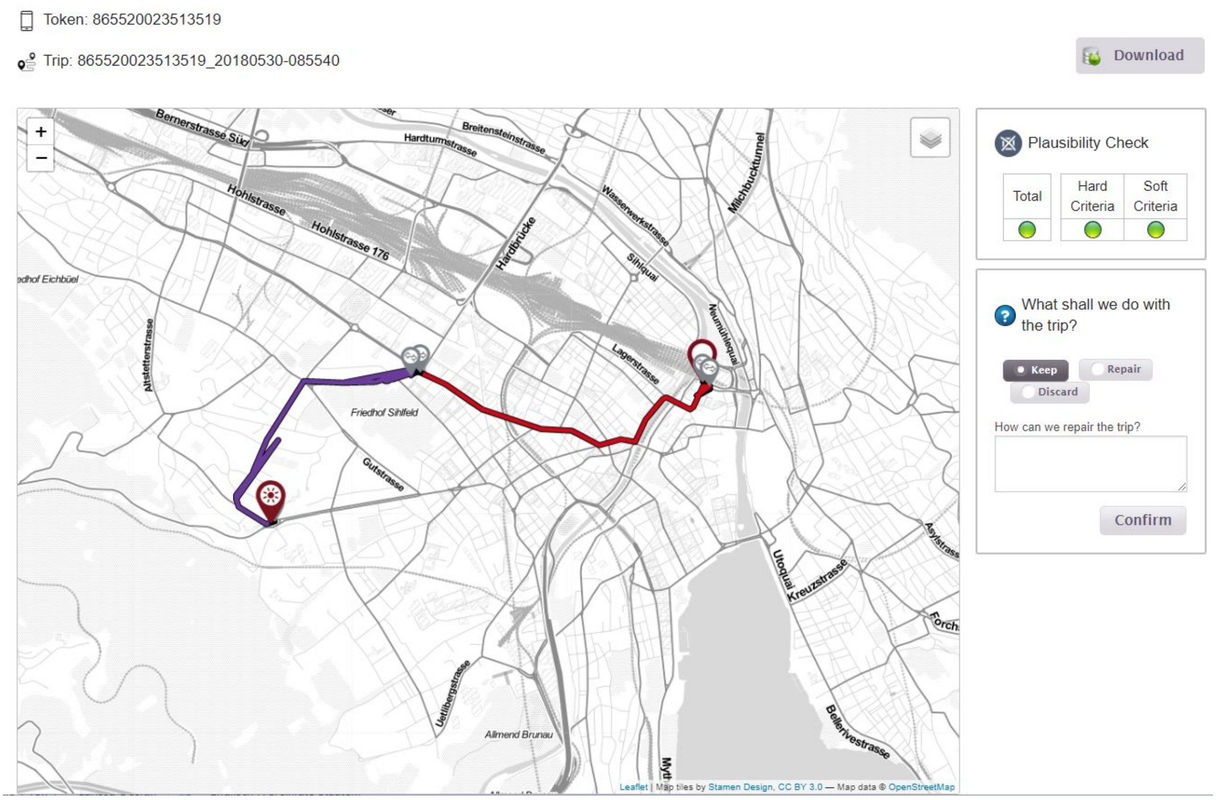


Figure 4: Web application for validation of collected data.

3.4.6 Scripts

To organize the collection campaign, it was decided to generate scripts for each trip we wanted to carry out. A script contains the script number (with an indication of whether the trip is uni- or bimodal), the modes of transport and depending on the modality, the durations for each mode for the bimodal scripts or further instructions for the unimodal scripts. Examples can be seen in figure 5. The scripts were kept as simple as possible to facilitate the execution. Additionally to the scripts, each participant was given a set of overall instructions including a short project description, the workflow of the collection campaign and how to use the web frontend.

Start the trip when already conducting the mode or right before you start (bike)

Please record and label the following trip:

Script	Mode of Transportation	Starting Position	Instructions	Duration
u28	E-BICYCLE	NA	Drive on a gravel road	10-15min

(a) Sample of unimodal script

Please record and label the following trip:

Script	Mode of Transportation	Duration
b10	TRAM	Short (1-5min)
	BUS	Short (1-5min)

(b) Sample of bimodal script

Figure 5: Examples of instructions given to participants.

3.4.7 Data Collection Campaign

The data was collected by 12 people from GIUZ and 10 people from AIT over a period of about two months (May and June 2018). Participants were asked to install the AIT Smart Survey application on their own phone and use it as the master device. Each participant was given a uTrail device, one or two smartphones, as necessary, and a number of print out scripts to carry out. The planning of their trips was left to the participants, asking them only not to start next to their home or workplace to ensure their privacy. All data was made anonymous, identifying the participants only via the token of the master device and the time frame this device was used in. As three devices were used per person, the participants were instructed to carry the master device in their hand and the slave devices according to the labels on the devices to ensure a balanced collection of all three positions.

3.5 Exploratory Analysis

In this section, the bimodal GPS part of the collected data from Zurich will be analyzed visually and statistically. It was decided to look only at the bimodal GPS data due to time constraints and to make a comparison with another data set collected by Zheng

et al. [52, 51, 50] possible, as explained below. A closer look was taken at the spatial distribution, the sensor signals of speed, and summary statistics of the trips. As each trip consists of the data of three devices, a comparison within the trips will be made to see how the device type and its handling influence the data.
Next to the exploratory analysis of our own data, the same analysis will be carried out for another data set collected for the purpose of transportation mode inference. To show the quality of our collected data and how it is useful in its function as a benchmark data set, a subset of an annotated data set by Zheng et al. from Microsoft Research Asia in 2008 [52, 51, 50] has been chosen. This data collection campaign was conducted as part of the GeoLife project started by Zheng et al. The GeoLife project focuses on visualization, effective organization, fast retrieval and deep understanding of GPS tracks via web application [52].
Zheng's et al. data contains the GPS tracks of 65 participants over a time span of 10 months and covers 18 cities (mostly in China) having tracked over 30'000 kilometers [51]. They covered seven modes of transport, namely bus, taxi, walk, subway, train, bicycle and car, which is more than in our data set but includes all the modes of transport chosen by us. For this comparison, a subset of the data available on the microsoft webpage [49] is used. The following analysis gives no overview over all of the data collected by Zheng et al. but merely shows how a portion of it compares to the data collected for this project.

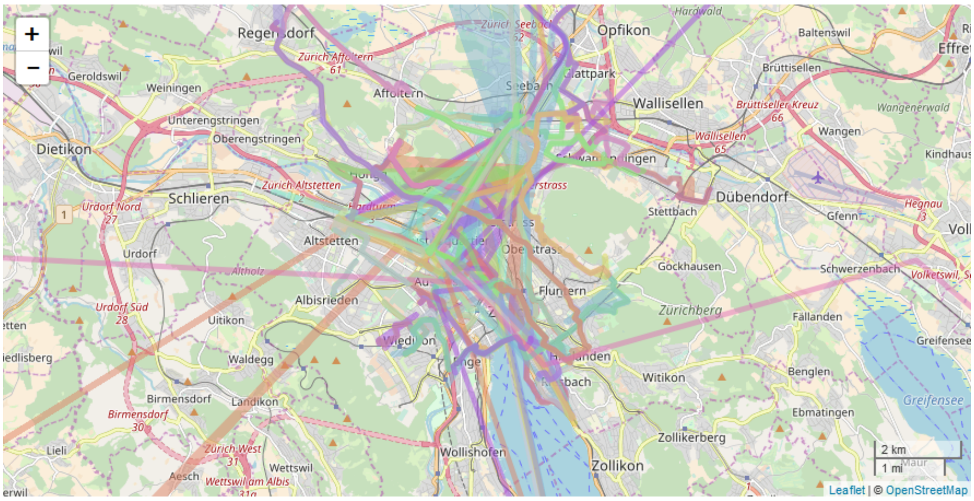
3.5.1 Organization and Pre-Processing

The first step towards exploring is data loading and cleaning, removing GPS outliers and attributes which are not necessary for the purpose of this thesis. As seen in figure 6(a) GPS data collected by smartphones can have significant errors, which need to be removed prior to further use. The errors can have multiple sources including scattering through urban canyons or uncertainties due to lack of satellites in view [13]. Urban canyons occur because of the reflection of the GPS signal on the walls of large buildings which influences the signal on the street in between whereas the satellite view can be obscured because of tunnels, dense tree cover or inside buildings [37]. Both of these problems are of great importance in cities and while they do disturb the signal, they can also be part of what makes the signal particular for a certain modes of transport. When inside a train for example, our devices are not able to pick up any GPS signals, which can be interpreted as a characteristic of that mode of transport. GPS outliers where removed by using three times the interquartile range. This algorithm was applied twice to remove not only outliers of one GPS point, but also the points where the device recorded two points at a unreasonable location. Yet, as seen in figure 6(b), not all outliers were found by this method. For further use, a more suitable method might

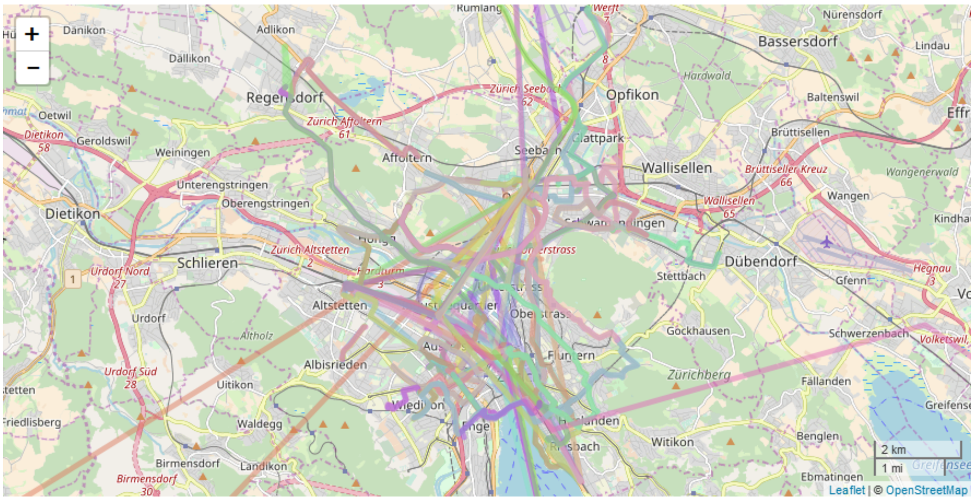
be taken into consideration.
For our data, as well as the data collected by Zheng et al., the annotations are stored in other files than the movement data. Therefore, the two files were reconnected using the time interval given in the annotation files and the labels were applied to each GPS location. In the annotation files, the duration per mode of transport, the distance between the points and the turning angles were calculated for further analysis.

3.5.2 Spatial Distribution

The focus of our data collection campaign was set on urban areas and their surroundings. While most of the public transportation is taking place in cities, longer train rides and car rides on highways and country roads mostly take place outside city boundaries. This can be seen in figure 6. This distribution was guided by the requirements in the scripts but since no exact trips were defined, it seems to follow a natural dispersal. This is connected to the availability of modes of transport and infrastructure. Densely built and populated cities are a hot spot of transportation and the systems are more developed than in rural areas. However, as participants were asked for example to drive 10 minutes on a highway, leaving the city was inevitable and anticipated.
Most research is conducted in this type of urban context because this is where the population lives, moves and where transport systems need to be efficient, working and practical. Transport mode detection is challenging in this environment because the modes of transport are slowed down, forced to stop because of traffic lights or congestion and cannot develop the signal patterns they potentially would without disturbances. Even so, these kinds of signals are part of a mode of transport and should be considered when attempting inferring the transportation modes. The goal of our benchmark data set was a well distributed, balanced set of trajectories. Looking at the map of all the trips conducted in Zurich, the dispersal within the city is satisfactory, with a higher amount of data collected in the north and west of the city. This can be traced back to the participants who are presumably more familiar with the neighborhoods surrounding the university campus.
The data of Zheng et al. has a focus on Beijing with a few trips collected in other countries such as the United States or Japan [52]. These trips seem like a peculiar choice because the amount of data collected in other countries is not big enough to see local differences and the distribution has such a strong focus on China, that the features of their modes of transport are represented the strongest. In figure 7, a cutout of Beijing can be seen. Less outliers were present that had to be removed.

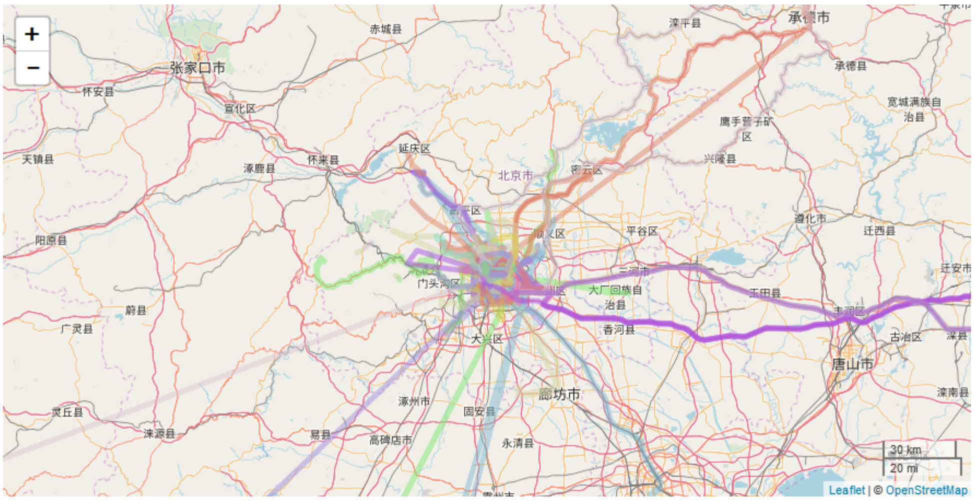


(a) Raw data.



(b) After outlier removal.

Figure 6: Cutout maps of bimodal data collected in Zurich.



(a) Raw data.

(b) After outlier removal.

Figure 7: Cutout maps of bimodal data collected in Beijing.

3.5.3 Analysis of Speed

In tables 8 and 9 the summary statistics of the variable speed can be seen. All of the velocities measured in Zurich are rather low, which can be explained by the urban environment. Public transport cannot reach higher speeds over a longer period of time because of the traffic, traffic lights, pedestrians and stops. The highest median velocity was reached by car segments. As only the bimodal data was analyzed, no trips on highways and only a few segments on country roads were included. Figure 8 shows, that most of the collected data is allocated in a speed range of zero to 25 kilometers per hour. Another, considerably smaller peak can be found at around 40 kilometers per hours. As the speed limit within the city of Zurich is 50 km/h, this might show

the velocity most often reached by trams, buses and cars when they are not slowed down by traffic lights, stop or traffic in general. It can be assumed that the velocity distribution would change considerably if the duration and distances of the trips had been longer and the surrounding of the city had been included more.
The data collected by Zheng et al. shows higher maximum speed especially for the modes bus, walking, bicycle and car. Surprisingly, the maximum speed of walking is higher than the one for running, which might be do to an erroneous label or or some forgotten outliers. Same can be said for the maximum speed of a bus with 214 kilometers per hour. The median speeds correspond more or less with our collected data and seems appropriate for the modes of transport. In figure 9 a similar pattern as in figure 8 can be seen. Buses seem to be slowed down less often than in our case. The peak seen in the Zurich data at around 40 km/h, can be seen in the data by Zheng et al. at about 65km/h. Further studies regarding the speed limits in Beijing have not been taken into account.

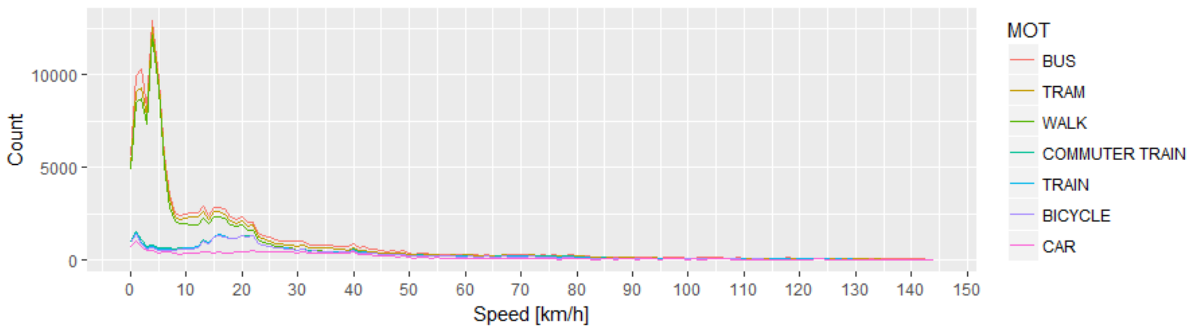


Figure 8: Distribution of speed per MOT.

Table 8: Summary statistics of speed distribution per MOT for Zurich data in km/h.

Mode of Transport	Minimum	1st Quantile	Median	3rd Quantile	Maximum
Bus	0.00	0.30	4.35	10.69	55.07
Tram	0.00	0.44	3.50	9.12	48.40
Walk	0.00	0.00	0.75	1.50	18.87
Commuter Train	0.00	0.00	1.09	16.51	82.25
Train	0.00	0.00	4.66	23.00	116.71
Bicycle	0.00	4.03	5.56	11.88	38.81
Car	0.00	1.34	6.98	13.57	79.29

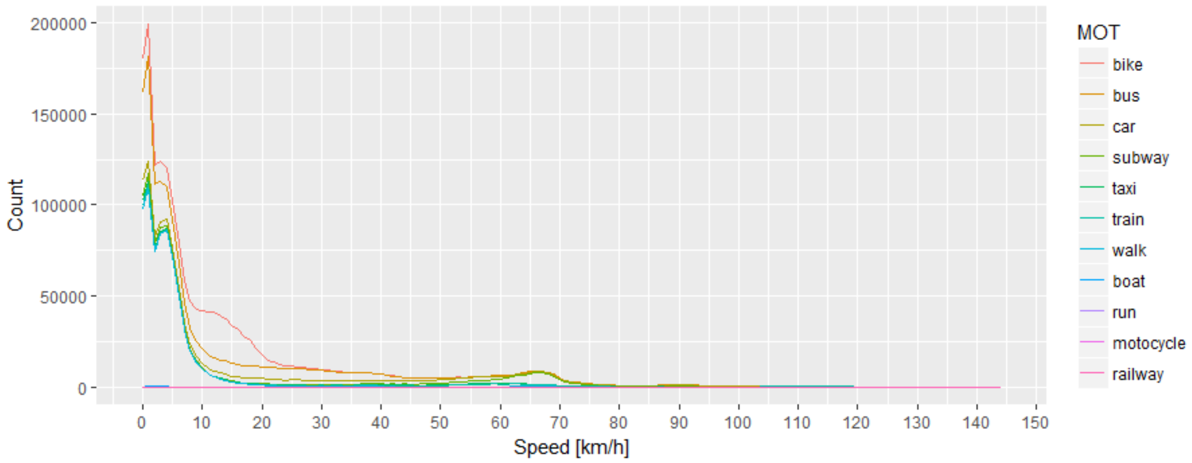


Figure 9: Distribution of speed per MOT for data by Zheng et al.

Table 9: Summary statistics of speed distribution per MOT for Zheng's et al. data in km/h.

Mode of Transport	Minimum	1st Quantile	Median	3rd Quantile	Maximum
Bus	0.00	0.39	2.24	6.80	214.48
Walk	0.00	0.21	0.85	1.47	41.97
Train	0.00	10.87	16.08	20.70	48.64
Bicycle	0.00	1.67	3.14	4.20	73.25
Car	0.00	2.06	6.83	12.97	111.37
Subway	0.00	11.14	17.15	18.63	43.02
Taxi	0.00	0.48	4.55	10.92	33.99
Boat	0.06	0.51	0.66	0.86	2.84
Run	0.07	1.11	1.73	2.43	9.77

3.5.4 Comparison of Three Devices

In figure 10 an example of a bimodal trip (b66) can be seen. In this trip, the instructions were to combine a short car ride with a long walk. The breaks can be seen very well in each device, starting with a short walking segment, followed by the instructed car ride and a longer walking segment in the end. In this scenario, device 1 is the master device, a smartphone carried in the participant's hand. Device 2 is the second smartphone, in this case carried in the participant's bag and device 3 is a uTrail personal tracker carried in the participant's pocket.
One can see that the device carried in the participant's hand shows a lot more noise

than for example device 3. The uTrail device shows the most stable speed measurements with only a few points were the signal was possibly lost.

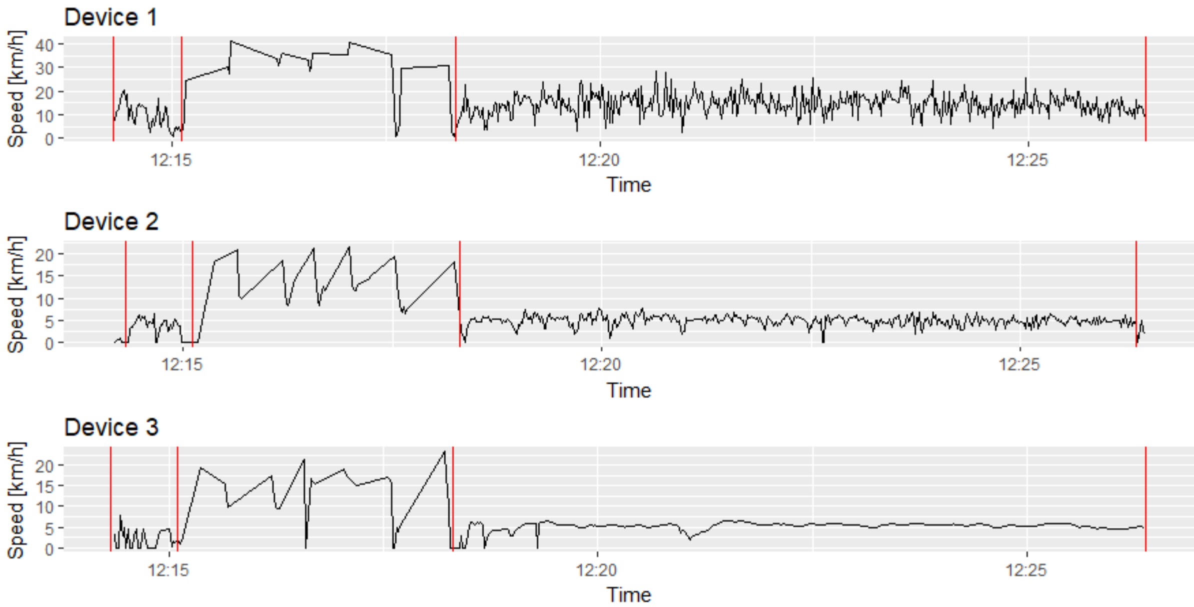


Figure 10: Overview of speed on the example of trip b66.

3.5.5 Distribution of Individual MOT

In this sub-chapter, the distribution of the modes of transport over the whole collection time will be analyzed. All the trips were split up into single segments of different modes of transport and their duration was added up. In figure 11, it can be seen that, in Zurich, much more walking data was collected than any other mode of transport. This can be explained by the fact, that each trip was started, linked and ended by a walking segment. Therefore, most walking segments are rather short, between one and five minutes. As the instructions were given to collect either segments shorter or longer than 5 minutes, most segments last no longer than 15 minutes. An exception are the few train and car rides, which are seen in blue and pink in figure 11.
Tabel 10 shows the amount of data collected in Zurich in duration and distance. Distance-wise, most data was collected for the mode of transport car, followed by walking, bicycles and buses. As walking is generally a very slow way of getting around, the most time was invested in this mode.

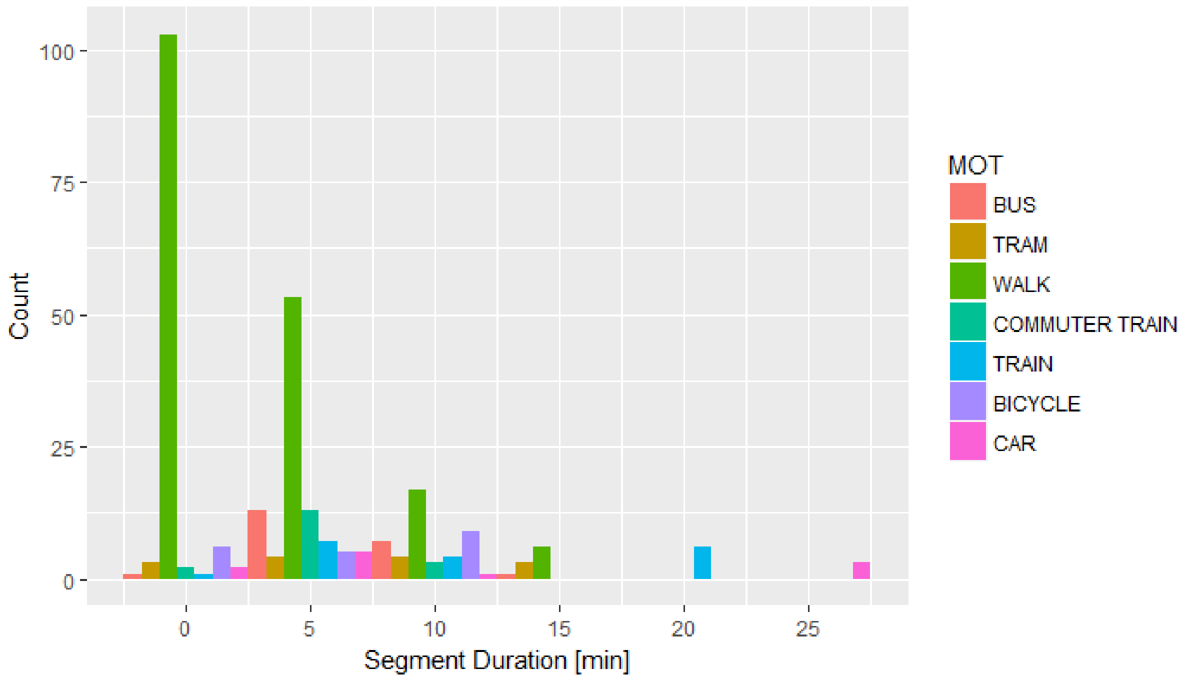


Figure 11: Distribution of segment duration per MOT in Zurich.

Table 10: Amount of data collected per MOT.

Mode of Transport	Amount of Data [min]	Distances per MOT [km]
Bus	129.52	79.82
Tram	129.20	65.59
Walk	636.14	88.69
Commuter Train	119.99	40.44
Train	208.12	40.43
Bicycle	126.20	86.49
Car	230.96	206.47
Total	1580.13	607.93

For the data collected by Zheng et al., the durations of the segments were generally much higher. A few of the segments lasted more than 50 minutes and were cut off in figure 12 to make it more clearly represented. A similar trend as in our data is visible with most segments lasting between 5 and 15 minutes. The longer the segment, the less segments were recorded. Overall, walking and biking were the two modes of transport collected most frequently with more longer subway rides around 30 minutes.
Zheng et al. collected a much bigger amount of data compared to us. As mentioned

in the beginning of the chapter, only a subset of all the data was analyzed here and already we can see a collection of about 500 hours for buses alone.

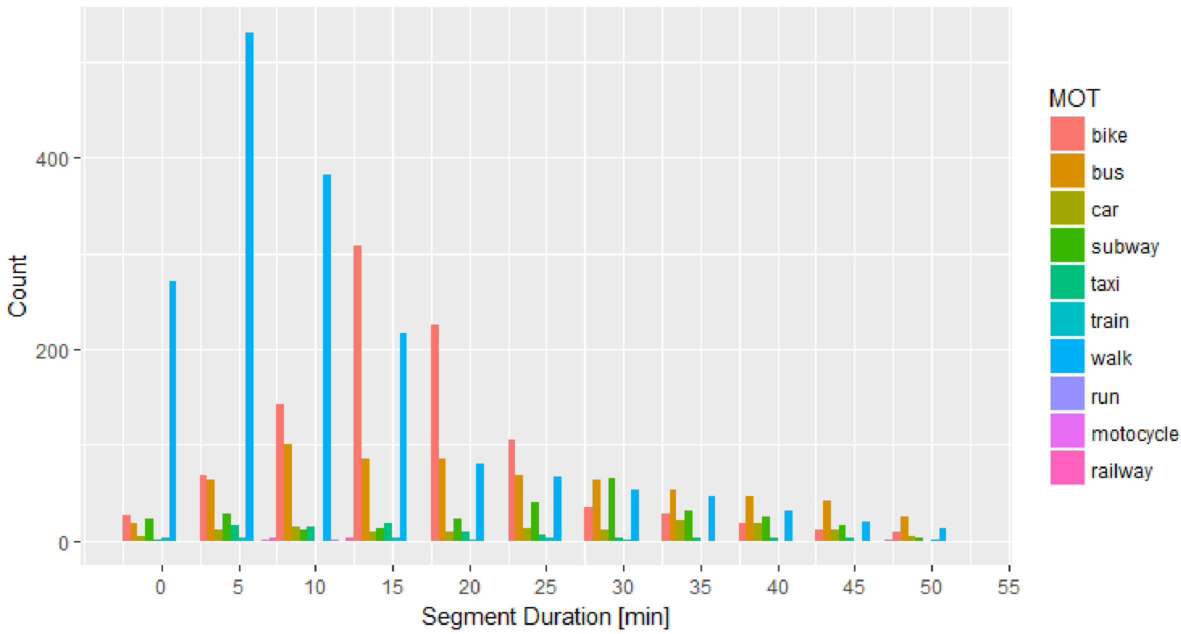


Figure 12: Distribution of segment duration per MOT in Beijing cut off at 50 minutes.

Table 11: Amount of data collected per MOT in Beijing.

Mode of Transport	Amount of Data [h]	Distances per MOT [km]
Bus	529	6'784.67
Taxi	23	10'205.70
Walk	857	24'734.39
Subway	153	3'582.40
Train	42	1'612.84
Bicycle	385	2'682.80
Car	369	39'332.26
Total	2'357	88'935.06

3.5.6 Comparison with Labeled Data Set by Zheng et al.

Overall, the two data sets have many similarities regarding the collected data and its specifications. Zheng's et al. data is documented well but is missing certain information which is necessary for the user. An example is the documentation of modality and how the collected trips were planned. The data is organized by user, making it difficult to extract for example data from a specific location or finding a subset with only unimodal trip data. Additionally, there is no documentation of how the labels were applied. This makes the data less reliable for further use and legitimates the need for a well-documented data set such as the data collected for this project. The more information about the data is available, the more it can be trusted and used in other projects, which is essentially the goal of a publicly available benchmark data set. Of course, Zheng's et al. data has not been collected as a reference data set but as a start of a larger collection of annotated GPS tracks, publicly available for further research in transportation mode detection and presents a good initiation for further collection campaigns.

4 Segmentation and Classification

This chapter shows an example of an approach of working with the data gathered in the collection campaign. As the time frame of this project did not allow for further research, the approaches and results shown are by no means complete and could be further improved by a more thorough pre-processing and the optimization of the features and parameters used. The term classification used in the following section always refers to a supervised classification problem unless otherwise mentioned.
As mentioned in Chapter 2.1, a classification can be carried out point-wise or segment-wise. In this chapter, an overview will be given over the two approaches, followed by the aims of the implementations. In a further step, three segmentations will be introduced and their implementation will be illustrated. In the sub-chapter classification, random forests, support vector machines and the K-nearest neighbors algorithms will be introduced.

4.0.1 Point-wise Classification

Point-wise or real-time classification is interesting for applications in the research area of location-based services, where the mode of transport should be predicted as quickly as possible to return to the customer. However, this approach is rarely chosen, as there are only very few variables available from raw GPS or IMU data, making a prediction more difficult. However, this can be extended through feature engineering. Additionally, this type of classification often shows very abrupt changes in mode, returning a single point of something predicted to be a car ride, framed by obvious walking points. To prevent this, and to gain additional information, near-real-time solutions, working with time-windows, are often implemented. The data is partitioned into time-windows or moving time-windows and for each segment, features can be calculated to improve the classifier.
According to Zheng et al. the length of a segment determines how well the features calculated from it describe a mode of transport [52]. Therefore, a segmentation with break points as close as possible to the real breaks without a lot of additional breaks in between, should be able to help the classification in a later stage.
Another approach is a point-wise classification followed by a smoothing or filtering process. In this way, consistent modes can be achieved [40, 11].

4.0.2 Segment-wise Classification

For purposes where there is time to analyze the data after the collection, segment-wise classifiers have been proven to be useful. For this approach, the data is partitioned into segments corresponding as closely as possible the actual annotated segments, a feature vector is calculated for each segment and in the last step, the classification is carried out on the feature vector. The advantage of this approach lies in the potential of calculating a number of features for each segment. The better a segment can be characterized, the better it will be recognized at a later stage. Additionally, each prediction of a segment is consistent, leaving no single points of different modes inside a segment. This type of classification has been carried out in a number of publications (cf. [3, 11, 25, 26, 39, 43, 45, 47, 48]).

4.0.3 Aims

The project aims at implementing three different segmentation algorithms and compare how they work on three different classifiers. The predictions of each combination, together with the prediction of a point-wise classification, will be evaluated through comparison with the annotation of the ground truth data.
The segmentation algorithms will be tuned visually to find suitable parameters.

4.1 Data

For the segmentation and classification experiments, the data from our collection campaign in Zurich is used. For segmentation, only trips with multiple modes of transport are of interest. For this reason, the bimodal part of our data will be used. A further restraint was given by the time frame of this project, therefore, only the GPS data will be considered. The data used for this project has not yet been sorted completely, therefore, not every trip is covered by three devices and some trips might occur more than once. While these conditions are not ideal, an example of using a data set, which can be analyzed in different ways, using only the parts of the data necessary for the research task, is given.
After loading and filtering by number of modes, the data is presented as a list of trips, each trip containing the location data of the 3 devices. Outliers were removed as shown in chapter 3.5. A column for the annotation was added to each data frame, giving each GPS location a corresponding label for supervised classification.

4.2 Segmentation

Segmentation is the process of finding suiting break or change points and splitting the data into segments according to these. Another word for segmentation used in related work is step detection. As mentioned before, segmented data gives the advantage of calculating features for each segment, characterizing the data more precisely. Here, only segmentation aiming at finding the breaks between the modes of transports and not a segmentation into time windows will be considered.

This has been done in a number of publications as for example seen in Aminikhanghahi et al. [1], Biljecki et al. [3], Schüssler et al. [38], Zhang et al. [48] or Zheng et al. [52].

Usually, break points are found by finding periods with low speed, low values in heading change or high acceleration values as a change of mode is mostly related with a walking segment or a so-called transition period with speed near or equal to zero. This type of threshold segmentation can be seen in the publication by Biljecki et al. [3] who explain their requirements for a segmentation algorithm as a detection of sudden changes. They found potential transition periods where stops (points with speeds lower than 2 kilometers per hour) longer than 12 seconds occur or when there was a shortage in signal. This threshold admittedly fails in the case of very fast transitions as for example a person running to catch a tram [3]. In our case, this system would also fail in the case of trains and commuter trains, as there is a severe signal shortage in these modes of transport. This has not been discussed further by Biljecki et al.

A similar approach was pursued by Zhang et al. [48]. They defined their threshold as a period where there are only small changes in position (distance change for 5 seconds less than 5 meters), small speed values (5 seconds with less than 0.5 meters per second) and large magnitudes in heading change (larger than 100 decimal degrees in 5 seconds) [48].

Schüssler et al. combined a speed threshold (lower than 0.01 meters per second over a time interval of at least 120 seconds) with a search of bundles of GPS points, where there are only little distances between the measures locations [38]. They defined a radius of 15 meters, calculating the point density from a sequence of 30 points in this radius over a period of 300 seconds.

As mentioned in chapter 4.2, the length of the segment determines how well a mode of transport can be described. To achieve this, a number of algorithms for the segmentation of time series have been developed, including a segmentation by first-passage time, the Behavioral Change Point Analysis developed for animal trajectories, a K-means clustering approach, the PELT (pruned exact linear time) algorithm or the BPMM (Bayesian partitioning of Markov Models) algorithm [12, 21].

4.2.1 Segmentation by Label

The data collected in this project was annotated during the process, giving us exact and reliable labels for every point in the time series. These can be used for a perfect segmentation of the trips. This approach is not a reasonable way of finding segments, as a segmentation algorithm's function is to find breakpoints without knowing anything about the actual annotation of the data. However, this segmentation was carried out to later use on different classifiers and observing how a classifier improves (or declines) in performance when given an optimally segmented set of data.

4.2.2 First-Passage Time Segmentation

FPT is originally an approach for detecting how much time an animal spends within a given area along a path [12]. The term first-passage time describes the time required for an animal to cross a circle of a given radius. This scheme can be applied for humans as well. Not only do we move in space similarly to animals (even though in a more organized manner), but when using different transportation modes, the lingering in smaller areas is present as well. An example is the use of public transportation. When waiting for a bus, one might spend a longer time in the same radius without much movement. When entering the bus however, the vehicle is moving relatively quickly with only short lingering around traffic lights or bus stops. This way, FPT can be used to find changes in behavior along a path and in a further step be used for segmentation. FPT can be calculated using the adehabitatLT package provided by Calenge et al. [8]. After the calculation of the FPT, an algorithm developed by Lavielle et al. in the adehabitatLT package can be used to find the optimal breaks [8].

4.2.3 Behavioral Change Point Analysis

BCPA is a likelihood-based change detection algorithm developed by Gurarie et al. in 2009 [20]. As the authors come from a background in biology, the algorithm was developed to identify shifts in the behavior of free-ranging animals in the wild, using GPS movement data. This type of data is generally not annotated, has a rather coarse resolution and can have a considerable amount of gaps for example, if the observed species spends part of its time under water. Our data on the other hand, is annotated and has a high resolution. Gaps occur in certain modes of transport such as train or commuter trains as well as in tunnels, train stations or in urban canyons. Looking at the differences and similarities of the two types of movement data, one wonders how BCPA performs with human instead of animal data.

BCPA identifies significant changes in movement parameter values by sweeping an analysis window over the time series and finding the most likely change points while

simultaneously testing which parameters have changed at this change point [18]. As a response variable, persistence velocity is usually chosen as it has proven to be fairly robust. For an unknown number of behavioral states (or in this case modes of transportation), the likelihood for mean, standard deviation and auto correlation within a given state are obtained. In a next step, potential breaks are found by determining, which combination of the three parameters (mean, standard deviation and auto correlation) best describes a separation of the data. By sweeping a window of a fixed size across the time series, all the change points, models and values of the estimated parameters can be collected [21].
The algorithm is available as a package for R [19] and has two possible output options: either a smooth BCPA output, which returns the average over all the estimated parameters and the location of all the change points or a flat output which finds the most frequently chosen change points, clustering the ones which are close to each other. Within each section, the three parameters are estimated and the location of the change points is recorded [18].
Using the information given about the best breaks, the data can be partitioned into segments. For each segment, the label which occurs most often is chosen.

4.2.4 Implementation

All three segmentation methods have a list of data frames (one entry for each trip) as their input variable. The output is always a list of segments, which can later be used for the calculation of the feature vector.
For the segmentation by labels, a simple split command was used to divide up all the trips into segments and store the outcome in a list of segments. This was a comparably easy task, which will not be discussed further.
For the FPT algorithm, all the data frames were converted to ltraj objects according to the instruction of the adehabitatLT package [8]. Ltraj objects handle a trajectory as a sequence of steps instead of locations and include the calculations of descriptive parameters for these relocations. This package was developed by Calenge as an analysis tool for animal movement [19] and includes implementation tools for FPT calculations. FPT was calculated with different radii to find the best suited breaks.
The last segmentation method implemented on the data is the BCPA algorithm. As mentioned before, it was developed by Gurarie et al. [18] and can be used for finding break points in movement data. The input parameters for BCPA include next to the movement data of the trips, a window size, the window step and the sensitivity factor K. For K, the value 1 was chosen as advised by Gurarie et al. [18] to keep the number of segments low, whereas the other three parameters were tested on a subset of the data and evaluated visually. The parameters used in the final segmentation are found

in table 12. The split factor used for calculation in this algorithm is persistence velocity. After the calculation of the window sweep, the results were used to find the best breaks and splitting the data accordingly.

Table 12: Segmentation Parameters.

Algorithm	Parameter	
BCPA	Windowsize	60
	Windowstep	2
	K	1
	Threshold	5
FPT	Radius	3

4.2.5 Features and Feature Vector Calculation

For each segment, a feature vector is calculated using a number of statistical measures to describe the segment as accurately as possible. This vector is later used for the classification. Which features are calculated and which variables are used to do so is of great importance for the next steps – the choice is wide. The variables available from the collected data are location (longitude/latitude), speed, bearing, altitude and a time stamp. For this project the following features were selected and will be explained in table 13.

Table 13: Features for feature vector.

Speed, turning angle	median, inter quartile range (IQR), variance, 25% and 75% quantile
Additional Features	accumulated travel distance (path and linear), total duration of segment, rate of change

Robust statistics were chosen, to avoid giving too much weight to outliers. Therefore, the median is calculated instead of using the mean of each segment and instead of using the minimum and maximum, the 25% and 75% quantiles were used. These in addition to variance were calculated for the variables speed and turning angle. Instead of using the variable bearing given by the GPS data from the smartphones, the turning angle was calculated from the coordinates and its statistical measures used for the feature vector.

Next to the statistical measures a number of additional features were calculated. The ratio of speed to elevation change shows for example how the mode of transport bicycle tends to speed up when going down where as e.g. a car is restricted by the speed

limits. The resulting label for the overall segment was applied using the label with the highest occurrence over the whole segment.
After the segmentation calculations, a feature vector was generated for each segment, leading to a data frame with one row per segment.

4.3 Classification

Classification algorithms have the aim of predicting qualitative responses as accurately as possible. For the particular problem of transportation mode detection, the response variable is the mode of transport whereas the predictors are either the measurements acquired by the data collection or the features calculated for each segment.
For the statistical learning in this project, three classification algorithms were chosen. First, a random forest approach, secondly a support vector machine and lastly a K nearest neighbors classifier. All three have been used for the task of transportation mode detection numerous times and have proven effective.
Each classifier is used multiple times with different input data. Once it is used pointwise, once for the optimally segmented data using the annotation, once for the segments obtained by FPT and lastly for the BCPA segments. This way, 12 combinations can be tested and evaluated. In the next sections, the classifiers will first be explained in detail, followed by their implementation in R.

4.3.1 Random Forest

Random forest algorithms are based on the idea of decision trees, where the predictor space is stratified into a number of simple regions by a sequence of decisions based on the variables given to the algorithm. The data is first split into two parts, based on the best predictor and the best cut point, where the next decision is taken and so on (see figure 13).
For a random forest, a number of decision trees are built on bootstrapped training samples. Each time a split in a tree is considered, a random sample of m predictors is chosen as potential split candidates from the full set of predictors [24]. Typically, m is chosen as the square root of the full number of predictors. With this method, the classifier does not consider a majority of the predictors at each split. The advantage is, that strong predictors do not have such a big influence which leads to a larger reduction in variance. After each tree has been formed, the random forest gets a class vote from each and then the classification is carried out using majority vote [22]. Through this decorrelation of the trees in the random forest, the average of the resulting trees has a lower variance and is hence more reliable [24]. Random forests are very popular classi-

fiers because they have proven to be fairly accurate and moderately interpretable [22]. Additionally, the effect of overfitting is seldom seen in random forest classification, as the classifiers are not so sensitive to variance [22].

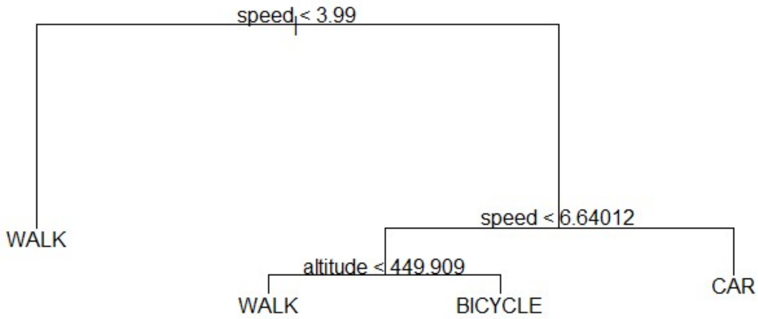


Figure 13: Example of a simple decision tree with two predictor variables.

4.3.2 Support Vector Machines

Support vector machines is an approach based on the concept of separating hyperplanes and the interrelated support vectors. A hyperplane is a flat subspace of dimension p-1 in a p-dimensional space and whereas a separating hyperplane divides up a feature space into two classes (see figure 14) [24]. If the data can be perfectly separated using a hyperplane, there will be an infinite number of possibilities of placing said hyperplane. To get the optimal separating hyperplane a maximal margin classifier is used. As seen in figure 14, the margin is defined by the data points closest to the separating hyperplane. These points are called support vectors.
As this only works if the classes are linearly separable, another solution is needed for cases where the features overlap and no separating hyperplane exists. In this case, a soft margin can be used, allowing for a few training observations to be misclassified to achieve a better result when classifying the remaining observations [24]. This type of learning algorithm is called support vector classifier. The tuning parameter C decides, how many observations are allowed to violate the margin. The higher C is chosen, the wider the margin is and the more biased the classifier turns out. If C is small, however, the classifier is highly fitted to the data, which leads to lower bias but high variance.
A support vector classifier can only be used for linear problems with two classes to distinguish, support vector machines were developed for non-linear decision boundaries. A SVM is a classifier which results from enlarging the feature space by using kernels such as polynomials, splines or radials, leading to more flexible decision boundaries [24] as seen in figure 15. This way, better training-class separation can be achieved [22].

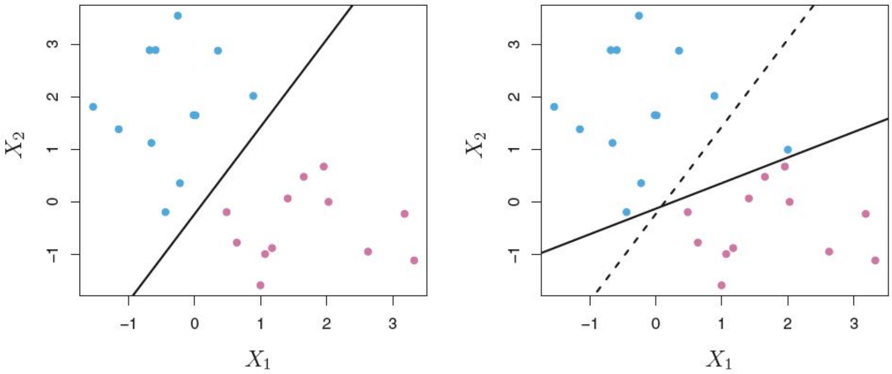


Figure 14: Two classes of observations with a maximal margin hyperplane. On the right side an additional blue observation, leading to a dramatic shift in the maximal margin hyperplane (solid line). Source: [24].

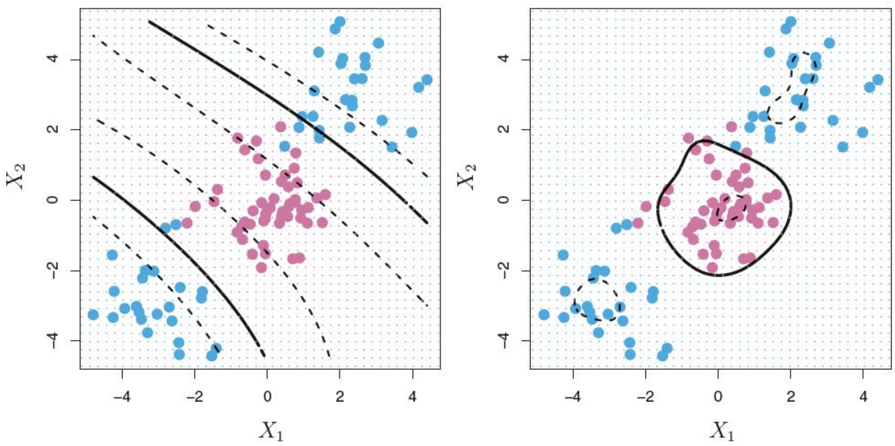


Figure 15: Application of a polynomial kernel of degree 3 on the left side and a radial kernel on the right side, dividing up the classes. Source: [24].

4.3.3 K-Nearest Neighbors

In the KNN method, a positive integer K is chosen, which determines the K points in the training data closest to a test observation. In the next step, the conditional probability for the class is calculated as the fraction of points which are represented most in the K nearest neighbors as seen in figure 16. After doing so for all points in the test set, a decision boundary can be drawn [24].

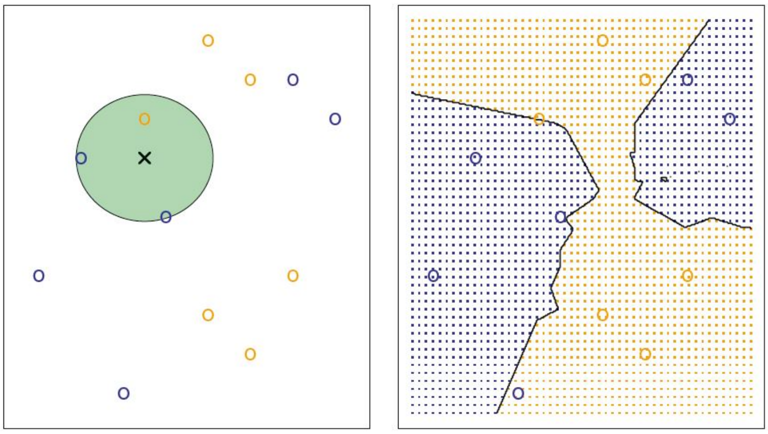


Figure 16: A KNN approach using K=3. The black x will be assigned to the blue group, since the majority of its K-nearest neighbors are blue. Source: [24].

4.3.4 Cross-Validation

For classification, the data is divided up into training and test data. After the model is trained on the first part of the data, the response on a new observation is predicted using the test data. This approach requires large data sets and the training error often underestimates the test error dramatically [24]. The idea is to maximize both, training and test data, which cannot be done if every data point is only used once. Therefore, a K-fold cross-validation approach is implemented. The data set is randomly divided up into K groups of approximately equal size. Each fold is, one after the other, treated as a test set, while the rest of the folds are used to fit the model. Like this, both, training and test set can be enlarged. After carrying out the procedure k times, each time with another fold used as the validation set, we have k prediction error rates in the form of misclassified observations [24]. These can be averaged to give the K-fold Cross Validation (CV) error rate. The error rate for classification problems is described as the number of misclassified observations. The choice of K depends on various factors. If K=N (where N is the number of observations) is chosen, the cross-validation estimator is less biased for the true prediction error, but a higher variance occurs because the training sets are very similar to one another [22]. Additionally, the computational cost is much higher than if a lower K is chosen. By choosing very low K, the cross-validation has a lower variance. Depending on how the performance of the learning method changes with the size of the training set, bias could become a problem [22]. Over the years, 5- or 10-fold CV have proven themselves as a good compromise for bias and variance and are applied widely [22].

4.3.5 Implementation

For each learning method, two classifiers were used, one for point- and one for segment-wise classification. All of the classifier were built using the caret package provided by Kuhn et al. [15]. This package was developed for classification and regression training and includes a number of possible models.
For the point-wise classification, the variable speed and turning angle were used. For each point, the three classifications were carried out using 10-fold cross validation. Because each trip was collected using three devices, it was ensured that all devices of a trip are in the same fold to avoid overfitting. The parameters used for each classifier can be seen in table 14.
The same was implemented for the feature vectors of the segmented data. Here, the number of variables was much higher, including all the features explained above. The predictions gained, were projected back to the point-wise data frame, to enable comparison between the different predictions. For each classifier the predictions were made for later evaluation and comparison.

Table 14: Classification Parameters.

	RF		SVM	KNN
Point-wise	mtry	1	kernel	k
	nodesize	30	gamma	
	ntree	1000	cost	
Segment-wise	mtry	4	kernel	k
	nodesize	30	gamma	
	ntree	1000	cost	

4.4 Results

4.4.1 Segmentation

For the BCPA and the FPT segmentation, the parameters were chosen visually, experimenting on a few examples. In figures 17 and 18 the blue lines represent the actual break points from the annotation, while the red lines show the break points derived from the algorithm. The segmentation algorithm was implemented on every trip, resulting in a list of segments per trip and mode of transport. The label was chosen, using the most often occurring label in the segment.
Figure 17 and 18 show that the segmentation achieved by neither algorithm is optimal.

While some of the breaks were set almost perfectly, others are far off. All of the trips were segmented using the same parameters as seen in table 12.

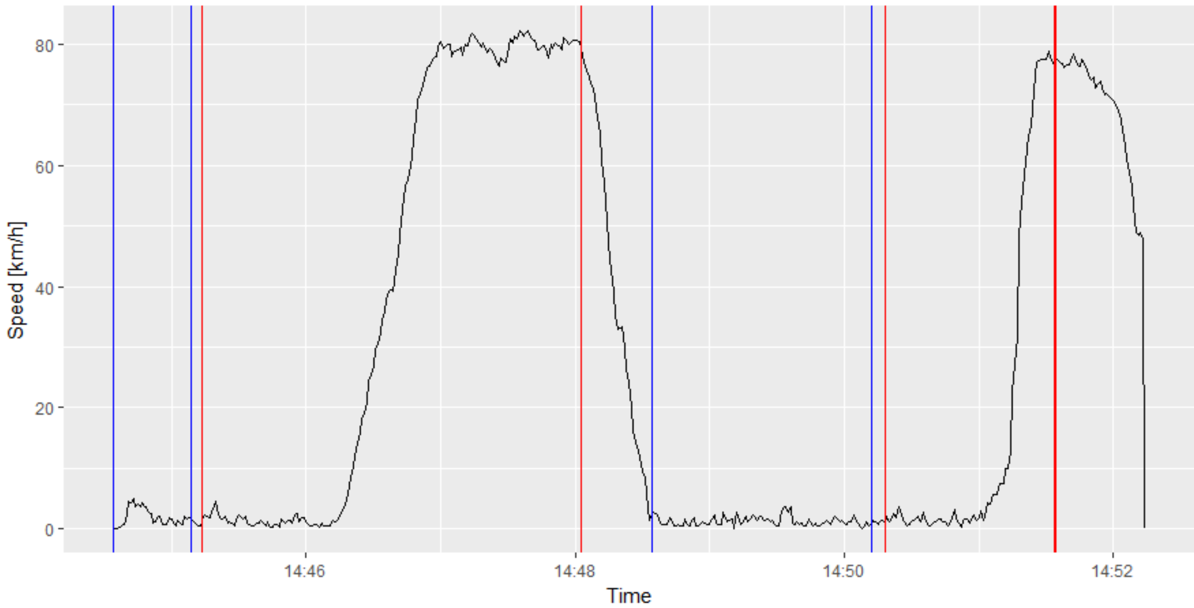


Figure 17: Example of segmentation by BCPA on trip b23. Blue lines represent the actual breaks and red lines show the break points found by the BCPA algorithm.

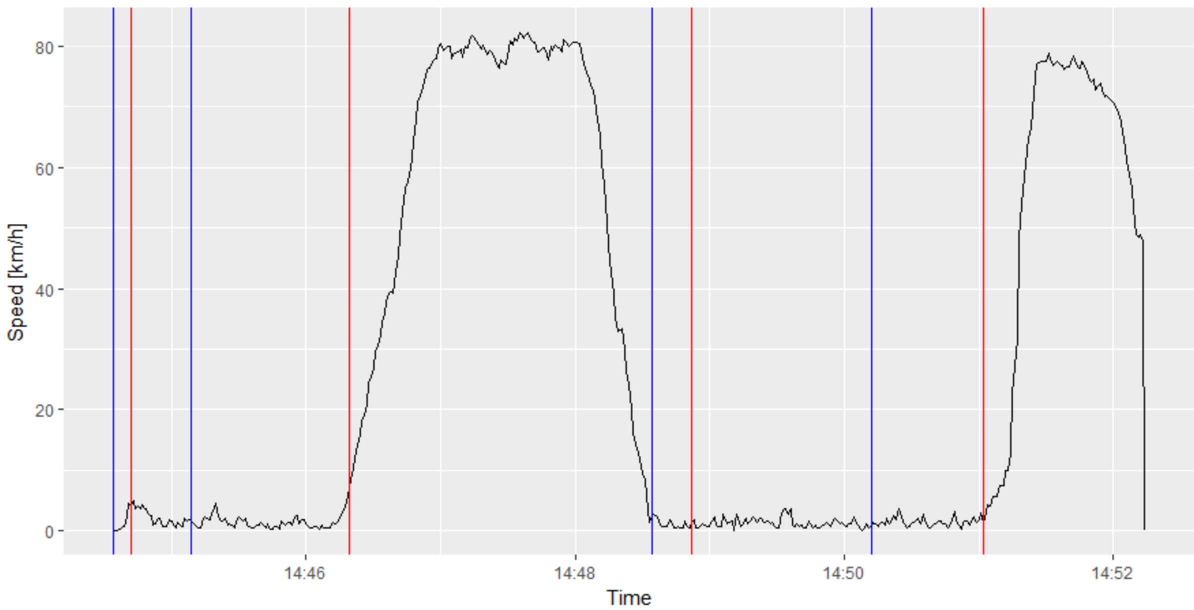


Figure 18: Example of segmentation by FPT on trip b23. Blue lines represent the actual breaks and red lines show the break points found by the FPT algorithm.

4.4.2 Point-Wise Classification

The point-wise classification was carried out using only two variables, speed and turning angle. As the KNN and the SVM classifiers could not handle the amount of data, the results here are only for the random forest classifier. The overall accuracy of the random forest classifier is 62.3 percent. As seen in table 15, walking and trains are the only mode of transportation detected with a relatively high accuracy, followed by bike, commuter trains and cars. The detection accuracy of buses was the lowest.
The classifier struggled mostly with detecting a difference between cars and buses, and trains and commuter trains. Interestingly, the transportation mode train has a rather low precision value but a comparably high recall value. This shows that many points were predicted to be of class train but were not actually in that class, but many of the points which did belong to the train class were actually discovered. The opposite can be seen for the class walking, where very few points were assigned to the wrong class, but only 72 percent of the walking points were discovered.

Table 15: Point-wise segmentation with classification by random forest.

(a) Confusion matrix.

	car	walk	bike	bus	tram	com.train	train
car	9128	7393	3533	1227	931	56	144
walk	421	57889	189	63	240	0	0
bike	2653	1830	6338	372	801	0	2
bus	4018	5194	1749	144	397	0	0
tram	2882	5076	2209	253	600	0	0
com.train	128	1264	19	5	15	199	192
train	441	1354	23	10	16	245	631

(b) Statistics by class.

	car	walk	bike	bus	tram	com.train	train
Precision	0.41	0.98	0.53	0.01	0.05	0.11	0.23
Recall	0.46	0.72	0.45	0.07	0.20	0.40	0.65
Balanced Accuracy	0.67	0.85	0.70	0.49	0.56	0.69	0.81

4.4.3 Classification with Segmentation by Label

The classification of the data segmented by labels achieved the highest overall accuracy with 76.5 percent using a SVM classifier. The two other classifier performed only slightly worse, with accuracies of 75.9 percent for the random forest and 73.5 percent for the K-nearest neighbors classifier.

As in the point-wise classification, walking was detected the best in these scenarios as well with an accuracy of up to 96 percent as seen in table 16. Another mode of transport, which was detected very well is bicycle. All three classifiers had a rather high success rate for the inference of bicycles as seen in tables 16, 17 and 18. Surprisingly, the SVM classifier did a lot better at recognizing trains as well as commuter trains, even though the distinction between the two is still very difficult.

Table 16: Segment-wise classification by random forest.

(a) Confusion matrix.

	car	walk	bike	bus	tram	com.train	train
car	18625	520	1307	1925	2827	0	0
walk	136	64758	286	367	516	41	0
bike	1563	1599	10792	34	34	0	0
bus	5243	525	0	2011	5048	0	0
tram	3171	445	0	3027	6600	0	0
com.train	192	238	0	190	687	130	154
train	760	545	4	419	746	182	0

(b) Statistics by class.

	car	walk	bike	bus	tram	com.train	train
Precision	0.74	0.98	0.77	0.16	0.50	0.08	0.00
Recall	0.63	0.94	0.87	0.25	0.40	0.37	0.00
Balanced Accuracy	0.78	0.96	0.92	0.58	0.67	0.68	0.49

Table 17: Segment-wise classification by KNN.

(a) Confusion matrix.

	car	walk	bike	bus	tram	com.train	train
car	13071	1271	3795	4379	2632	0	56
walk	0	65737	0	0	0	0	367
bike	310	476	11746	941	549	0	0
bus	2632	1994	0	5086	3115	0	0
tram	2028	4334	0	3028	3853	0	0
com.train	142	766	0	268	205	204	6
train	779	718	0	336	665	158	0

(b) Statistics by class.

	car	walk	bike	bus	tram	com.train	train
Precision	0.52	0.99	0.84	0.40	0.29	0.13	0.00
Recall	0.69	0.87	0.76	0.36	0.35	0.56	0.00
Balanced Accuracy	0.79	0.93	0.87	0.65	0.64	0.78	0.49

Table 18: Segment-wise classification using SVM.

(a) Confusion matrix

	car	walk	bike	bus	tram	com.train	train
car	13685	1074	537	1860	3010	0	5038
walk	0	64910	180	300	50	62	602
bike	672	191	12455	153	550	0	0
bus	3212	921	0	5638	3056	0	0
tram	3835	414	269	3244	5362	0	119
com.train	0	305	0	72	205	292	717
train	240	650	0	128	206	66	1366

(b) Statistics by class.

	car	walk	bike	bus	tram	com.train	train
Precision	0.54	0.98	0.89	0.44	0.40	0.18	0.51
Recall	0.63	0.95	0.93	0.49	0.43	0.70	0.17
Balanced Accuracy	0.77	0.97	0.96	0.72	0.68	0.84	0.58

4.4.4 Classification with FPT Segmentation

After the segmentation using the First-passage time algorithm, the classifiers perform worse than with the segmentation by label. The overall accuracies are 61.3 percent for random forest, 59.1 percent for KNN and 52.2 percent with the SVM classifier.
Again, as seen in tables 19, 20 and 21, cars were often mistaken for buses, whereas buses were classified as either walking or cars. With this segmentation, the inference of trams seems more difficult with accuracies not exceeding 52 percent. Many instances were classified as car, walking or even bus.
SVM was the only classifier detecting any commuter train or train points. This is possible because all the confusion matrices are based on the point locations from the original data set and not on the feature vector of the segments. This way, a number of points were lost in the prediction because of the inability of the classifier to work with missing values.

Table 19: FPT segmentation with classification by random forest.

(a) Confusion matrix.

	car	walk	bike	bus	tram	com.train	train
car	24927	4902	450	2110	2401	0	0
walk	7122	72771	3530	2919	4235	0	0
bike	3617	1406	11459	723	1446	0	0
bus	8179	3675	623	2947	2345	0	0
tram	5220	6892	989	2219	1577	0	0
com.train	363	1736	271	349	39	0	0
train	1737	1917	222	70	76	0	0

(b) Statistics by class.

	car	walk	bike	bus	tram	com.train	train
Precision	0.72	0.80	0.61	0.17	0.09	NA	NA
Recall	0.49	0.78	0.65	0.26	0.13	NA	NA
Balanced Accuracy	0.71	0.79	0.81	0.59	0.52	NA	NA

Table 20: FPT segmentation with classification by KNN.

(a) Confusion matrix.

	car	walk	bike	bus	tram	com.train	train
car	22669	4877	1912	3803	1529	0	0
walk	7227	75515	3576	1707	2551	0	0
bike	5136	1861	9343	712	1599	0	0
bus	8882	3776	2385	1332	1294	0	0
tram	3949	8060	2325	1808	728	0	0
com.train	648	1822	103	0	185	0	0
train	1684	2229	55	0	54	0	0

(b) Statistics by class.

	car	walk	bike	bus	tram	com.train	train
Precision	0.65	0.83	0.50	0.07	0.04	NA	NA
Recall	0.45	0.77	0.47	0.14	0.09	NA	NA
Balanced Accuracy	0.68	0.80	0.71	0.52	0.50	NA	NA

Table 21: FPT segmentation with classification by SVM.

(a) Confusion matrix.

	car	walk	bike	bus	tram	com.train	train
car	10177	14841	2507	4844	2421	0	0
walk	3525	73439	3228	4436	5687	262	0
bike	2850	3778	10534	747	742	0	0
bus	7375	4696	709	576	4413	0	0
tram	1806	9362	902	2800	2027	0	0
com.train	70	2474	0	18	78	57	61
train	186	3502	54	0	55	225	0

(b) Statistics by class.

	car	walk	bike	bus	tram	com.train	train
Precision	0.29	0.81	0.56	0.03	0.12	0.02	0.00
Recall	0.39	0.66	0.59	0.04	0.13	0.10	0.00
Balanced Accuracy	0.62	0.71	0.77	0.47	0.52	0.55	0.49

4.4.5 Classification with BCPA Segmentation

The classifiers perform slightly better on the BCPA segments than on the FPT segments. The overall accuracy for the random forest is 63.5 percent, 61.2 percent for the KNN and 62.25 percent for the SVM classification. The random forest classifier performed best. Tables 22, 23 and 24 show similar results concerning the distinction of modes of transport. Again, the modes walking and bike were identified with the highest accuracy. However, looking at table 22, both, precision and recall for the mode bike are rather low. Not many points, which should have been recognized as biking were found and of the ones that were found, a big part were actually false positives.

Table 22: BCPA segmentation with classification by random forest.

(a) Confusion matrix.

	car	walk	bike	bus	tram	com.train	train
car	25484	4467	923	2208	2553	11	42
walk	4773	77482	4099	1410	2662	38	3
bike	4891	1737	10488	310	1214	0	0
bus	9458	3755	444	1388	2708	0	0
tram	5378	3679	1255	2904	3670	0	0
com.train	438	1315	0	456	252	0	286
train	1694	1730	297	120	16	153	0

(b) Statistics by class.

	car	walk	bike	bus	tram	com.train	train
Precision	0.71	0.86	0.56	0.08	0.22	0.00	0.00
Recall	0.48	0.82	0.6	0.18	0.28	0.00	0.00
Balanced Accuracy	0.70	0.84	0.78	0.53	0.60	0.49	0.48

Table 23: BCPA segmentation with classification by KNN.

(a) Confusion matrix.

	car	walk	bike	bus	tram	com.train	train
car	21890	5682	3708	1907	1561	0	4
walk	5719	80964	1837	1344	603	0	0
bike	6629	1622	8975	472	942	0	0
bus	10662	4601	485	571	1434	0	0
tram	6596	5540	1137	2681	932	0	0
com.train	707	1654	238	0	148	0	0
train	1315	2016	343	171	120	0	45

(b) Statistics by class.

	car	walk	bike	bus	tram	com.train	train
Precision	0.63	0.90	0.48	0.03	0.06	NA	0.01
Recall	0.41	0.79	0.54	0.08	0.16	NA	0.92
Balanced Accuracy	0.66	0.84	0.74	0.49	0.54	NA	0.95

Table 24: BCPA segmentation with classification by SVM.

(a) Confusion matrix

	car	walk	bike	bus	tram	com.train	train
car	20414	4874	2229	3810	3235	0	190
walk	5828	79061	1636	1388	1759	350	445
bike	5006	1213	10065	660	1315	0	381
bus	7827	3674	797	2604	2838	0	13
tram	5637	4086	1462	2919	2734	0	48
com.train	770	1334	0	100	190	245	108
train	1960	1591	45	46	104	68	196

(b) Statistics by class.

	car	walk	bike	bus	tram	com.train	train
Precision	0.59	0.87	0.54	0.15	0.16	0.09	0.05
Recall	0.43	0.83	0.62	0.23	0.22	0.37	0.14
Balanced Accuracy	0.66	0.85	0.78	0.57	0.57	0.68	0.56

4.5 Discussion

Looking at the results of the segmentation algorithms FPT and BCPA, one can see, that these might not be the optimal choice for the segmentation of human movement data. While the algorithms performed very well on some trips, the break points in others did not match the labels set during the data collection. The chosen parameters were analyzed visually, which could potentially be improved by statistical evaluation metrics. It is assumed however, that the effect on the overall segmentation would be minimal. FPT for example, has only one free parameter, which might be able to find breaks between certain modes of transport but neglect others. BCPA on the other hand, leaves a lot of room for optimization but as described by Gurarie et al. the BCPA tool was foremost developed for exploratory analysis for animal movement and might not give the exact change points necessary for transport mode detection [20]. The segmentation results in Chapter 4.4 show, that human behavior in an urban environment with all the mechanic movement of vehicles, the loss of signal in indoor environments and trains and very similar records for different modes of transport, does, in fact, differ significantly from the GPS records collected by animals.
The segmentation results in figure 17 and 18 are interesting to consider because it becomes clear to the reader, how much better humans still are at distinguishing things visually than computers. Looking at the speed curves, breaks seem obvious. However, the task needs to be carried out very quickly and cost-effective, making algorithms for segmentation necessary and desirable.
Overall, the classifiers performed better on the segment-wise data, which can be explained by the fact, that a much bigger feature space was available in these cases. For the point-wise classification, a larger feature space could be calculated using moving windows and feature engineering. This might influence the performance of the classifier significantly.
The results of the performances of all classifiers show, that certain modes of transport are more easily detectable than others. The variables recorded and calculated for the walking segments, for example, differ considerably from motorized modes or even other active modes such as biking. In the case of the data used in this project, it can also be argued that there was most training data available for this mode, however, walking seems to be detected rather accurately in many publications [37, 5, 25].
Another mode which was generally well inferred is biking. Biking shows relatively constant velocities, usually higher than the ones recorded for walking and has less extreme turning angles than walking. The other mode biking was often confused with is the class car, possibly because of the many slow car segment present in our data.
Difficult to distinguish are the modes train and commuter train as well as bus and

car. The choice to record both trains and commuter trains was made for the reference data set to give a future users the option of distinguishing the two. However, it was clear from the beginning, that, if used for training a classifier, the distinction would be difficult based on only GPS data. In the case of the data collected in Zurich, only rather short distances were covered in trains and commuter trains respectively. Therefore, the characteristics might not be as distinct as they might be in longer segments. Both modes vary only slightly, because of different train models, their velocity and the amount of stops they make. By additionally using the accelerometer recordings, a better distinction would be possible. An other option would be to combine the two modes into one, which would potentially improve the classifier.

Similar difficulties can be seen for the distinction of cars and buses. As most of the data was collected in an urban setting, the traffic was often slowed down by traffic lights, speed limitations and a lack of long, straight street where higher speeds can be reached. Therefore, car recordings look very similar to buses and even trams, making a distinction difficult when using only GPS data. An idea for improvement would be the use of GIS, adding for example the stops for public transportation.

The results show some examples of how transportation mode detection can be carried out. However, it becomes clear, that the data collected for this project might be more useful for its original purpose, the evaluation of algorithms already present or developed in the future. The collected data shows a great variety of transportation modes, devices and how they were handled. This is an advantage for the evaluation as it presents data close to real life and applicable in many different contexts.

5 Conclusion

5.1 Summary

Transportation mode detection has become more and more interesting for science with the development and improvement of small sensor technology. This project, conducted in cooperation with the Austrian Institute of Technology, shows, how different works, conducted over the years, can be brought together and made comparable with the help of a benchmark data set. The topic has been analyzed thoroughly to find the most important parameters to consider and include as much as necessary to gain various approaches with different aims.
After designing a collection campaign, the collection was carried out in Zurich and Vienna with the help of volunteers, collecting GPS and acceleration data with smartphones and uTrail devices.
In a second part, an example of how the collected data could be put to use is given. Three segmentations were implemented followed by three classification algorithms to see, how well the classifiers perform on the data. The results of the segmentation and classification are not perfect but have a big potential for improvement. If for example, acceleration were included, the parameters optimized or more features calculated the performance of the classifiers would supposedly increase.

5.2 Contributions

5.2.1 Data Collection Campaign

Building the structure for a benchmark data set and implementing these is challenging and leaves space for a big amount of suitable solutions. As said by Sim et al., a benchmark is only useful if it invites fellow researchers to participate and is used widely because of its applicability [42]. Looking back at the collected data and the considerations, the discussions and ideas behind it, it can be said that a good starting point was set. The data was collected in controlled settings with precise annotation and organized in a way it can be easily accessed and used. The data set can and should be broadened to fit all purposes and research areas it could be useful for and to keep it up to date.
It is clear, that this data set is a beginning towards a benchmark for transportation mode detection. In a next step, evaluation metrics and application examples need to be developed and published.
During the data collection, it became more and more clear how difficult it is to collect

data in laboratory quality in the real world. However, this difficulty can be seen as an advantage, as the behavior we have while moving around will never be controlled and shows a variety of signals and characteristics and it is therefore desirable to find this peculiarities in the collected data. Collecting this amount of data is time consuming and asks for participants who were instructed very well. This can be challenging as we were working on a voluntary basis and time was scarce.

5.2.2 Segmentation and Classification

This project provided the option of using our own data for segmentation and classification. As the concept of the data collection was worked out in a small team and collected by only a few participants, the inspection interpretation of the results could be carried out in an informed manner. The difficulties which presented themselves during data collection were known and were able to be taken into consideration. Additionally, a new approach in this field of research was the use of segmentation algorithms developed for the segmentation of animal trajectories in this thesis. While the outcomes might not be very good so far, a potential in adjusting these types of algorithms for human movement data is present.

5.3 Outlook

Looking at the achievements of this project, the potential for further research is still considerable. Concerning the collected data, everyone can contribute with more data, either for other regions or on the other hand with a more specific approach towards their field of research. Furthermore, evaluation criteria for transport mode detection need to be defined, and application examples developed.

Now that this data sets are available, a number of implementation and analysis possibilities emerge. Classification could be extended by using the IMU data as well, comparing different models and using for example the Zurich data as a training set, testing on the data collected in Vienna. Another set of interesting questions regard the unimodal data collected for this project. The unimodal data shows a variety of behavior in public transportation and different types of individual transport such as running, harmonic walking and strolling for the walking mode, bicycles and e-bikes and different types of cars. These data can be analyzed with regard to the effect the variations have on the overall mode of transport and how much this influences classification results. Furthermore, the position, the devices were carried in could be taken into consideration.

As the data was collected with the aim of making it part of a benchmark for transporta-

tion mode detection, the main aim is to use it to re-evaluate classification algorithms proposed in literature as an independent test set. These can be compared and show how well which algorithm performs in comparison with others and where a potential for improvement exists. To do so, the collected data can be adjusted to the user's need, for example by playing with the granularity of the modes of transport or by dividing the trips into unimodal segments for further investigation.
Looking at the utilization of the data, segmentation algorithms still pose a major research gap. Finding change points in human movement data has not been thoroughly researched so far and is potentially very interesting for further classification. As seen in Chapter 4.2, algorithms for animal tracks are a feasible starting point and can be developed further to achieve better results. Here, the acceleration data could be taken into consideration as well, broadening not only the feature space but potentially also the segmentation possibilities.

References

[1] Samaneh Aminikhanghahi and Diane J. Cook. A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51(2):1–29, 2016.

[2] Pia Bereuter, Michelle Fillekes, and Robert Weibel. Analyzing GPS and Accelerometer Data in the Study of the Mobility, Activity and Social Interaction of Older Adults. *Visually-supported Computational Movement Analysis (VCMA 2016)*, (June), 2016.

[3] Filip Biljecki, Hugo Ledoux, and Peter van Oosterom. Transportation mode-based segmentation and classification of movement trajectories. *International Journal of Geographical Information Science*, 27(2):385–407, 2013.

[4] Adel Bolbol, Tao Cheng, Ioannis Tsapakis, and Andy Chow. Sample Size Calculation for Studying Transportation Modes from GPS Data. *Procedia – Social and Behavioral Sciences*, 48:3040–3050, 2012.

[5] Adel Bolbol, Tao Cheng, Ioannis Tsapakis, and James Haworth. Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification. *Computers, Environment and Urban Systems*, 36(6):526–537, 2012.

[6] Norbert Brändle, Peter Widhalm, Markus Ray, and Magnus Åhlander. Smart Survey – A Turnkey Solution for Automatically Identifying Travel Modes and Trip Purposes of People with Smartphones 1. In *Proceedings of the 13th International Conference on Location-Based Services*, volume LBS2016, pages 14–16, 2016.

[7] Young Ji Byon, Baher Abdulhai, and Amer Shalaby. Real-time transportation mode detection via tracking global positioning system mobile devices. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 13(4):161–170, 2009.

[8] C. Calenge. The package adehabitat for the r software: tool for the analysis of space and habitat use by animals. *Ecological Modelling*, 197:1035, 2006.

[9] Rahul Das and Stephan Winter. Detecting Urban Transport Modes Using a Hybrid Knowledge Driven Framework from GPS Trajectory. *ISPRS International Journal of Geo-Information*, 5(11):207, 2016.

[10] Somayeh Dodge, Robert Weibel, and Ehsan Forootan. Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, 33(6):419–434, 2009.

[11] Katherine Ellis, Suneeta Godbole, Simon Marshall, Gert Lanckriet, John Staudenmayer, and Jacqueline Kerr. Identifying Active Travel Behaviors in Challenging

Environments Using GPS, Accelerometers, and Machine Learning Algorithms. *Frontiers in Public Health*, 2(April):1–8, 2014.

[12] Per Fauchald and Torkild Tveraa. Using First-Passage Time in the Analysis of Area-Restricted Reports. *Ecological Society of America*, 84(2):282–288, 2003.

[13] Tao Feng and Harry J.P. Timmermans. Transportation mode recognition using GPS and accelerometer data. *Transportation Research Part C: Emerging Technologies*, 37:118–130, 2013.

[14] Michelle Fillekes, Pia Bereuter, and Robert Weibel. Comparing GPS-based Indicators of Spatial Activity to the Life-Space Questionnaire (LSQ) in Research on Health and Aging. Technical Report September, 2016.

[15] Max Kuhn. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan, and Tyler Hunt. *caret: Classification and Regression Training*, 2017. R package version 6.0-78.

[16] Paola A. Gonzalez, Jeremy S. Weinstein, Sean J. Barbeau, Miguel A. Labrador, Philip L. Winters, Nevine Labib Georggi, and Rafael Perez. Automating Mode Detection Using Neural Networks and Assisted GPS Data Collected Using GPS-Enabled Mobile Phones. *World Congress on Intelligent Transportation Systems*, 2008.

[17] Nil Goyette, Pierre-Marc Jodoin, Fatih Porikli, Janusz Konrad, and Prakash Ishwar. changedetection . net : A New Change Detection Benchmark Dataset. Technical report, 2012.

[18] Eliezer Gurarie. Behavioral Change Point Analysis in R : The bcpa package, 2013.

[19] Eliezer Gurarie. *bcpa: Behavioral change point analysis of animal movement*, 2014. R package version 1.1.

[20] Eliezer Gurarie, Russel D. Andrews, and Kristin L. Laidre. A novel method for identifying behavioural changes in animal movement data. *Ecology Letters*, 12(5):395–408, 2009.

[21] Eliezer Gurarie, Chloe Bracis, Maria Delgado, Trevor D. Meckley, Ilpo Kojola, and C. Michael Wagner. What is the animal doing? Tools for exploring behavioural structure in animal movements. *Journal of Animal Ecology*, 85(1):69–84, 2016.

[22] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The Elements of Statistical Learning. *The Mathematical Intelligencer*, 27(2):83–85, 2001.

[23] Samuli Hemminki, Petteri Nurmi, and Sasu Tarkoma. Accelerometer-Based Transportation Mode Detection on Smartphones. Technical report, 2013.

[24] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning with Applications in R*, volume 102. 2006.

[25] Heikki Mäenpää, Andrei Lobov, and Jose L. Martinez Lastra. Travel mode estimation for multi-modal journey planner. *Transportation Research Part C: Emerging Technologies*, 82:273–289, 2017.

[26] Vincenzo Manzoni, Diego Maniloff, Kristian Kloeckl, and Carlo Ratti. Transportation mode identification and real-time CO_2 emission estimation using smartphones: How CO_2GO works. Technical report, 2011.

[27] Bryan D. Martin, Vittorio Addona, Julian Wolfson, Gediminas Adomavicius, and Yingling Fan. Methods for real-time prediction of the mode of travel using smartphone-based GPS and accelerometer data. *Sensors (Switzerland)*, 17(9):1–20, 2017.

[28] Gianluca Menghini, Nelson Carrasco, Nadine Schuessler, and Kay Axhausen. Route choice of cyclists in Zurich GPS-based discrete motion models. *Arbeitsberichte Verkehrs- und Raumplanung*, 544, 2009.

[29] Marija Nikolíc and Michel Bierlaire. Review of transportation mode detection approaches based on smartphone data STRC 17th Swiss Transport Research Conference Review of transportation mode detection approaches based on smartphone data. In *17th Swiss Transport Research Conference*, number May, pages 1–20, 2017.

[30] Philippe Nitsche, Peter Widhalm, Simon Breuss, and Peter Maurer. A Strategy on How to Utilize Smartphones for Automatically Reconstructing Trips in Travel Surveys. *Procedia – Social and Behavioral Sciences*, 48:1033–1046, 2012.

[31] N.N. AIT. https://www.ait.ac.at/#!/.

[32] N.N. Personal Tracker User's Manual. Technical report, Athens, Greece, 2017.

[33] N.N. Mikrozensus Mobilität und Verkehr (MZMV). https://www.bfs.admin.ch/bfs/de/home/statistiken/mobilitaet-verkehr/erhebungen/mzmv.html, 2018.

[34] N.N. Online OED. https://en.oxforddictionaries.com/definition/benchmark, 2018.

[35] Federico Perazzi, Jordi Pont-Tuset, Brogan T. McWilliams, Luc Van Gool, Mitchel E. Gross, and Alexander Sorkine-Hornung. A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 724–732, 2016.

[36] Adrian C. Prelipcean, Gyözö Gidófalvi, and Yusak O Susilo. Transportation mode detection – an in-depth review of applicability and reliability applicability and reliability. *Transport Reviews*, 37(4):442–464, 2017.

[37] Sasank Reddy, Min Mun, Jeff Burke, Deborah Estrin, Mark Hansen, and Mani Srivastava. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks*, 6(2):1–27, 2010.

[38] Nadine Schuessler and Kay Axhausen. Processing Raw Data from Global Positioning Systems Without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board*, 2105:28–36, 2009.

[39] Muhammad Awais Shafique and Eiji Hato. Use of acceleration data for transportation mode prediction. *Transportation*, 42(1):163–188, 2014.

[40] Rahul C. Shah, Chieh-yih Wan, Hong Lu, and Lama Nachman. Classifying the mode of transportation on mobile phones using GIS information. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing – UbiComp '14 Adjunct*, pages 225–229, 2014.

[41] Li Shen and Peter R. Stopher. Review of GPS Travel Survey and GPS Data-Processing Methods. *Transport Reviews*, 34(3):316–334, 2014.

[42] Susan Elliot Sim, Steve Easterbrook, and Richard C. Holt. Using benchmarking to advance research: a challenge to software engineering. *25th International Conference on Software Engineering, 2003. Proceedings.*, pages 74–83, 2003.

[43] Leon Stenneth, Ouri Wolfson, Philip S. Yu, and Bo Xu. Transportation mode detection using mobile phones and GIS information. *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems – GIS '11*, page 54, 2011.

[44] Sheung Tsui and Amer Shalaby. Enhanced System for Link and Mode Identification for Personal Travel Surveys Based on Global Positioning Systems. *Transportation Research Record*, 1972(1972):38–45, 2006.

[45] Peter Widhalm, Philippe Nitsche, and Norbert Brändle. Transport Mode Detection with Realistic Smartphone Sensor Data. *International Conference on Pattern Recognition (ICPR)*, (ICPR):573–576, 2012.

[46] Hao Xia, Yanyou Qiao, Jun Jian, and Yuanfei Chang. Using smart phone sensors to detect transportation modes. *Sensors (Switzerland)*, 14(11):20843–20865, 2014.

[47] Meng-Chieh Yu, Tong Yu, Shao-Chen Wang, Chih-Jen Lin, and Y. Chang Chang. Big data small footprint: The design of a low-power classifier for detecting transportation modes. *Proceedings of the VLDB Endowment*, 7(13):1429–1440, 2014.

[48] Lijuan Zhang, Sagi Dalyot, Daniel Eggert, and Monika Sester. Multi-Stage Approach To Travel-Mode Segmentation and Classification of GPS Traces. *ISPRS –*

International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVIII-4/(October):87–93, 2012.

[49] Yu Zheng. GPS Trajectories with transportation mode labels. https://www.microsoft.com/en-us/research/publication/gps-trajectories-with-transportation-mode-labels/, 2010.

[50] Yu Zheng, Yukun Chen, Quannan Li, Xing Xie, and Wei-Ying Ma. Understanding transportation modes based on GPS data for web applications. *ACM Transactions on the Web*, 4(1):1–36, 2010.

[51] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. Understanding mobility based on GPS data. *Proceedings of the 10th international conference on Ubiquitous computing – UbiComp '08*, (49):1–10, 2008.

[52] Yu Zheng, Like Liu, Longhao Wang, and Xing Xie. Learning transportation mode from raw gps data for geographic applications on the web. *Proceeding of the 17th international conference on World Wide Web – WWW '08*, pages 247–256, 2008.

Appendix

Appendix A: Collection Campaign Instructions

Project Description

In this project, the University of Zurich and the Austrian Institute of Technology are working together to make an important step in mobility research. We will collect a benchmark dataset that will be accessible to the scientific community. Thus, scientific contributions to the topic of ”Reconstruction of paths covered by the analysis of sensor data” become internationally comparable. For other communities (Image Recognition, Activity Recognition, Time Series Analysis) such benchmark datasets have already been published. Now we close this gap for our area of application.

Workflow

Recording

Three devices are used for the recordings. The first smartphone (the master device) should be carried in the hand / quickly accessible to your hands and serve as an annotation device, i.e. on this device each change from one mode of transport to the next is annotated in the app.
The other two devices (slave devices) are according to stickers either in the front pocket or in a bag or backpack. All three devices need to be turned on, smartphones must have location access enabled and run the AIT Smart Survey app. Trip recording has to be started by pressing a button in the app on all smartphones! After that the slave devices can be put in the pocket or bag, respectively, and do not require any further user input until the end of the trip.
After the trip has been started on all smartphones, push the Change Mode button on the master device, select the first mode of transport and enter the Script ID in the comment field.
Each change between modes of transport is again recorded on the master device using the ”Change Mode” button and the choice of the next mode of transport. Please press the Mode Change Button when:

- You step in/ out of the public transport vehicle.
- You start/ end the engine of your car
- When you are still sitting on your bicycle/ E-bicycle right before you start/ end your bicycle trip.

Make sure to press Change Mode and then Walk as soon as you step out of a public transport vehicle to change to another mode.
At the end of the trip stop recording on all devices by pushing stop trip in the app.
You should be careful to choose your routes such that they do not start and stop at the same place. Respecting your privacy you must not record trips while travelling your everyday routine trips (dont start or finish recordings at any locations that are personally related to you: home, work, etc.).
In the event that anything went wrong during recording a trip (e.g. extraordinary events such as a car breakdown, or met a friend by coincidence associated with a conversation and stay, ...), or you only recorded a trial trip, you can delete this trip in your records before you upload all your records, via the button "Upload Now".
After you have recorded several trips, you can upload them to our server. This always happens for all trips currently stored on your device. For a faster upload, a Wi-Fi connection is beneficial.

Frontend

After all trips have been uploaded, you need to check them on the website. You can login with your credentials. Please find your master token and get access to all your recorded trips. On the map you can see where and how you moved. In the background, a check was already made whether your trip can match the respective script.
Your task is to select and confirm:

- Keep if you are satisfied with the recorded trip,
- Repair... if you made a mistake that we can easily fix (e.g. you accidentally clicked "BICYCLE" instead of "WALK"). In this case you need to tell us in the comment field the required changes,
- Discard if the trip contains serious errors.

FAQ

Why do I have to confirm my trips?
We want to make sure the trip was recorded as described in the script. We can only use confirmed trips for research purposes.

Why am I not allowed to record during my daily routines?
In this research project, we investigate typical transport mode specific patterns in the sensors. We have no interest in personal information, and in particular, no interest in learning mobility profiles of individuals for which one could learn and associate typi-

cal addresses for a person.

How to carry my devices?
The smartphone with which the paths are to be annotated should be held in the hand. The other two devices (whether uTrail or another smartphone) should be carried according to the label on the device.

When do I have to turn on the slave devices?
The uTrail Personal Tracker will have to be turned on at least 10-15 minutes before its actual use, as it has a cold start phase during which a GPS signal cannot yet be received. You may switch off the device as soon as you have stopped the trip on the master device.
If a smartphone is used as a slave device, you should start the trip in the AIT Smart Survey App shortly before the actual start of data collection and be terminated shortly after the actual end of the trip again.

What is a Master/ Slave device?
The master device is the smartphone with which you choose the means of transport, enter the script ID and possibly additional comments. There is only one master device per user.
The slave devices are the other devices, either uTrail and/ or other smartphones. Slave devices need to be turned on and for the slave smartphones you also need to start and end the recording with the Smart Survey App.

I cannot find the transport modes E-Bicycle and E-Car, in the App?
For special transport and variations thereof (E-Bicylce, E-Car, SUV, Petrol / Diesel Car) no extra buttons are provided in the app. Please select the most similar ones for this type of transport. So for E-Bicylce choose Bicylce and for E-Car, SUV, Petrol / Diesel Car please choose Car.

Where do I enter the script ID?
Upon starting the trip you must choose the first mean of transport. On the next screen you enter the script ID in the field for notes. The script ID only needs to be entered on the master device. In addition to the script ID, please leave NO comment in the comment field for the first mean of transport.

Is the order of the transport modes in the script important?
No. It doesnt matter which of the two modes of transport in the script you use first.

It is important that both trips be made and to stick to the respective time frame per transport mode.

I cannot see my trips online. What can I do?
Check if you have already uploaded the files of all your smartphones. The processing of new trips takes some time and is executed only once per hour at XX:20. Please be patient and wait until you see them online. If they are still not online, please contact us via mail.

Appendix B: Software

Software	Description
R 3.4	Language and environment for statistical computing and graphics.
R-Studio 1.1.423	Free and open-source IDE for R; used as the main tool for the processing and visualization of the collected data and all calculations for segmentation and classification.
JetBrains PyCharm Community Edition 2017.2.3	Free and open-source IDE for Python; used for the organization and pre-processing of the collected data.
Overleaf	Online LaTex editor; used to write the thesis.

The R scripts developed for this project are not included in this thesis but can be acquired from the author Vera Isler (vera.isler@uzh.ch), if there is interest.

Personal Declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Zurich, 28 September 2018

Vera Isler V. Isler