

Analysis and Comparison of Publicly Available Databases for Urban Mobility Applications

Dina Bousdar Ahmed¹, Estefania Munoz Diaz¹

¹Institute of Communications and Navigation, German Aerospace Center (DLR), Wessling, Germany

Abstract

The challenges of multimodal applications can be addressed with machine learning or artificial intelligence methods, for which a database with large amounts of good quality data and ground truth is crucial. Since generating and publishing such a database is a challenging endeavour, there are only a handful of them available for the community to be used. In this article, we want to analyze three of these databases and compare them. We assess these databases regarding the ground truth that they provide, e.g. labels of the means of transport, and assess how much unlabelled data they publish. We compare these databases regarding the number of hours of data, and how these hours are distributed among different means of transport and activities. Finally, we assess the data in each public database regarding crucial aspects such as the stability of the sampling frequency, the minimum sampling frequency required to observe certain means of transport or activities and, how much lost data these databases have. One of our main conclusions is that accurately labelling data and ensuring a stable sampling frequency are two of the biggest challenges to be addressed when generating a public database for urban mobility.

Keywords

Machine learning, artificial intelligence, data mining, smartphone, passenger, dataset, vehicle, localization, detection, means of transport, transport mode.

1. Introduction

Digital technology is becoming a lever to integrate, aggregate and facilitate multimodality in urban mobility. In this article, we follow Klinger's definition of multimodality as "*the situational combination of transport modes*" [1]. Multimodality is a means to achieve cleaner, less polluted cities, as well as to improve the efficiency of passenger flow in cities and foster economic wealth.

In order to enable urban multimodality, it is necessary to address key issues like how to efficiently plan a passenger's trip making use of the different means of transport available locally [2]. In fact, it should be possible to take into account the passenger's preferences, e.g. using a bike-sharing system instead of a bus, or even take into account a passenger's physical limitations.

Some of the challenges to enable urban mobility applications can be addressed with machine learning or artificial intelligence technologies [3, 4, 5]. In this article, we define machine learning as *a set of methods that automatically detect patterns in data, and they are then used to uncover patterns to predict future data, or to perform other kinds of decision-making under uncertainty*

IPIN 2021 WiP Proceedings, November 29 – December 2, 2021, Lloret de Mar, Spain

✉ dina.bousdarahmed@dlr.de (D. Bousdar Ahmed); estefania.munoz@dlr.de (E. Munoz Diaz)

ORCID 0000-0003-4681-394X (D. Bousdar Ahmed)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

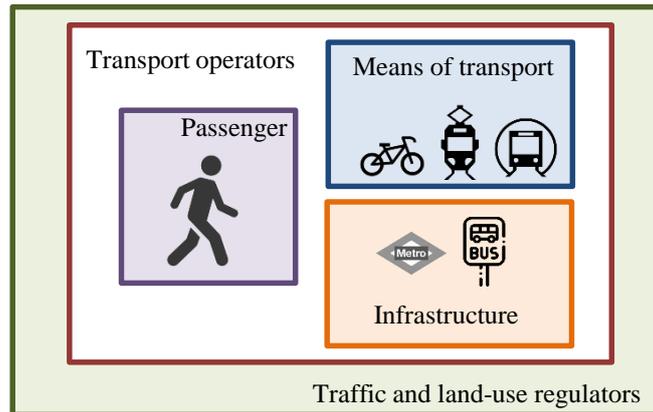


Figure 1: Visualization of the five main elements of the urban mobility ecosystem.

[6]. In contrast, we define artificial intelligence as the area of computer science that aims at creating intelligent machines that work and react like humans. In order to enable machine learning or artificial intelligence in urban mobility applications, the first requirement is the availability of large amounts of data to learn the latent patterns behind the various aspects of urban mobility [7]. Therefore, it is key to invest effort in collecting not only data, e.g., accelerometer, turn rate or GNSS data from smartphones, but also the corresponding ground truth, e.g., labels of a passenger’s choice of means of transport during a multimodal trip.

The generation of public databases with large amounts of accurately labelled data is an expensive and challenging endeavour. Only a handful of research institutions and companies have the means of taking on that endeavour but the majority of the research institutions or companies would benefit from these public databases. In fact, the generation of public databases that become the benchmark for the development of urban mobility applications will enable the comparison of different systems. Public databases also foster the standardization of the evaluation of systems for urban mobility applications; among them, passenger positioning [8] and the detection of the mean of transport.

There exists already some databases with data that could potentially be used for urban mobility applications. Unfortunately, only a handful of these databases are public [9, 10, 11]. Most of these databases are private [12, 13, 10], which hinders the progress in the development of applications for urban mobility.

To the best of our knowledge, there is no work that analyzes and compares publicly available urban mobility databases. We believe that such an analysis and comparison is of interest as some databases target similar applications, e.g. the recognition of the transport mode [9, 10, 11]. The community will benefit from a comparison that exemplifies how one can analyze public databases, identify their strengths and weaknesses and, ultimately, make an informed decision regarding which database to use for the development of an urban mobility application.

The goal of this article is to analyze and compare publicly available databases that contain data which could potentially be used to develop urban mobility applications. With our work, we aim at providing an example of how databases can be assessed regarding different criteria,

e.g., the sensor data in the database, the means of transport or activities considered, the type of ground truth provided or the amount of missing data within the database. More specifically, we have the following objectives:

- define urban mobility applications and the criteria to assess the databases for such applications and
- select publicly available urban mobility databases analyze them regarding the criteria listed above.
- compare the selected public databases and highlight, through our analysis, the key factors that should be taken into account when creating a public database.

2. Overview of Urban Mobility Applications

In this section, we provide first our concept of the urban mobility ecosystem to then summarize some of the key applications in urban mobility. Based on these concepts, we present the criteria that we use to select public databases for urban mobility as well as the criteria to analyze and compare these databases.

2.1. The Urban Mobility Ecosystem and Its Applications

This subsection provides an overview of urban mobility applications. First, we present our concept of the urban mobility ecosystem, see Figure 1, and its five main elements:

- Passenger
- Infrastructure
- Means of transport
- Transport operators
- Traffic and land-use regulators

The passenger is the user of the transport network and transport hubs and thus the central element of the urban mobility ecosystem. Smartphones, which are associated to passengers, play a key role in the urban mobility ecosystem. Smartphones are already ubiquitous in developed countries and are spreading fast in developing countries. Thanks to smartphones, digital technologies for urban mobility are developed with a passenger-centric approach. The passenger makes use of the infrastructure, which is the built environment that supports the operation of the means of transport. The infrastructure compiles transport hubs and stops as well as tracks, catenaries, lamp posts, roads, sidewalks, bicycle lanes and the like.

Means of transport encompass the different vehicles or ways to move from an origin to a destination. We classify means of transport in walking, non-motorized, and motorized vehicles. Both, non-motorized and motorized vehicles, can be divided in shared and private vehicles. Examples of non-motorized vehicles that might be shared are scooters and bicycles. The shared motorized vehicles' offer is as rich as the private variety, e.g. motorbikes, e-scooters, e-bikes and cars among others. Under the public motorized vehicles, the most common are buses, suburban trains, subway and trams.

Transport operators can be classified as well in public and private and they are responsible for designing and making available the means of transport, transport network and the transport hubs to the passengers, taking into account the passengers' needs.

The traffic and land-use regulators are as well part of the urban mobility ecosystem: how land is used and how cities are designed helps to determine the urban infrastructure and what kind of means of transport are most adequate. For instance, single-family homes on large lots increase the need for cars, whereas high-rise apartments with limited parking create the conditions for people to prioritize public means of transport like subways, buses or taxis. Traffic and land-use regulators are also responsible for regulating the operation of the transport operators, thus influencing the design of the means of transport in urban environments.

2.2. Criteria to Select, Analyze and Compare Databases

This subsection introduces the criteria we have used to select, analyze and compare public urban mobility databases. We focus on those databases recorded with smartphones and that contain and label one or more of the following features:

- different carrying modes,
- different means of transport and
- different activities.

In addition, we consider only publicly urban mobility databases that are downloadable at the time of writing this article.

We will analyze the selected databases taking into account the following criteria:

- *Criterion 1:* Labels and positioning ground truth.
- *Criterion 2:* Size of the databases in number of samples and number of users.
- *Criterion 3:* Carrying mode of the smartphone, used means of transport and activities considered.
- *Criterion 4:* Sensors recorded and sampling frequency.

3. Selection and Analysis of Publicly Available Urban Mobility Databases

For this article, we have considered the 17 databases listed in Table 1 in [10], the 21 databases listed in Table VIII in [8], and additionally the databases in [14, 11, 15, 16]. Given the criteria in Section 2.2, we have selected three databases out of the 42 mentioned previously for this article:

- The Geolife Trajectory Dataset [17, 9], which contains GPS data collected from smartphones or GNSS receivers from volunteers in Asia. The volunteers travelled throughout different means of transport and carried out different types of activities, e.g. shopping or sports.
- The Sussex-Huawei Locomotion Dataset [10, 18], which contains sensor data, e.g. from inertial sensors, magnetometers, GNSS etc., from three volunteers that carried out the

tests in the United Kingdom. The trajectories contain data from different means of transport, activities like still, walking or running and the volunteers used four smartphones simultaneously to record data.

- The Transport Mode Detection (TMD) Dataset [11, 19], which contains sensor data, e.g. from inertial sensors and magnetometers, collected with a smartphone by volunteers in Italy. The trajectories expand throughout different means of transport, and include data from standing still and walking.

The selected databases are not only public but also contain data recorded with smartphones and labels of the carrying modes and the activities. The remaining databases have not been selected because they are either not public or cannot be downloaded, do not contain smartphone data or do not contain data or labels of different activities and means of transport.

In the following sections, we analyze each of these databases according to the criteria described above.

3.1. Criterion 1: Labels and Positioning Ground Truth

The ground truth included within a database will condition the type of application for which a database is suitable. In this article, we are interested in applications that detect the mean of transport or the activity of the passenger as well as positioning applications. Therefore, we focus on the analysis of the following type of ground truth: labels of the carrying mode, the means of transport and the activity and positioning ground truth, which is summarized in Table 1.

Table 1

Details regarding criterion 1: labels and positioning ground truth. We specify the type of label and positioning ground truth provided by each publicly available database.

Title	Carrying mode label	Means of transport label	Activity label	Ground truth position
Geolife GPS Trajectory Dataset	-	yes	-	-
Sussex-Huawei Locomotion Dataset	yes	yes	yes	yes (GPS logs)
TMD Dataset	-	yes	yes	-

While all considered databases include labels of the used means of transport, only the Sussex-Huawei Locomotion Dataset and TMD Dataset include activity labels. Therefore, all databases would enable the development of algorithms that detect the carrying mode whereas only the Sussex-Huawei Locomotion Dataset and the TMD Dataset enable the development of activity recognition methods.

Regarding the labels of the smartphone carrying mode, the Sussex-Huawei Locomotion Dataset is the only publicly available database, out of the three considered, that includes labels of the carrying mode. Therefore, this database enables the development of algorithms that

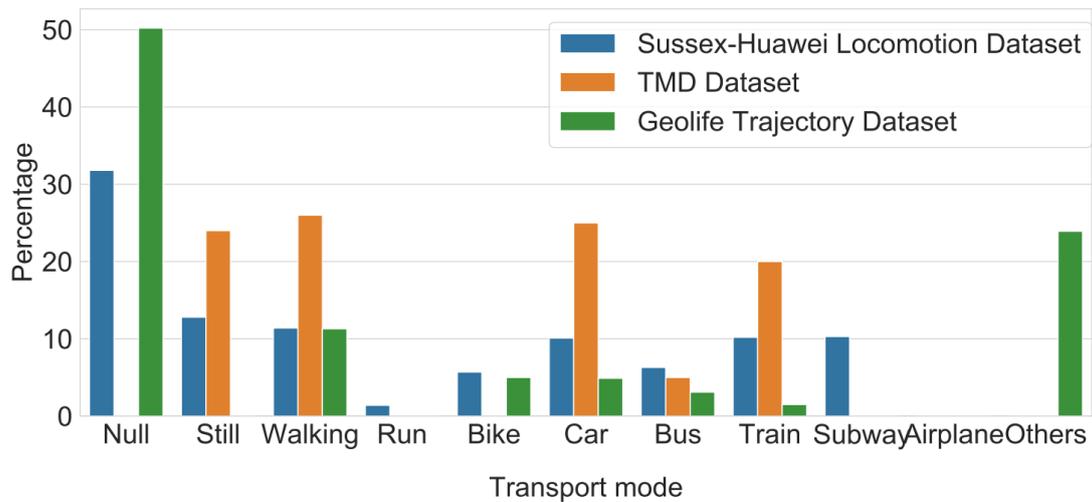


Figure 2: Percentages of labels of each means of transport and activity.

detect the carrying mode or algorithms that consider how the different smartphone carrying modes effect other applications, e.g. the classification of the means of transport or multimodal positioning, i.e. positioning in different means of transport.

The Sussex-Huawei Locomotion Database and the Geolife GPS Trajectory Dataset include positioning information based on global navigation satellite system (GNSS) measurements. However, GNSS measurements cannot be considered ground truth in urban environments due to the presence of multipath or absence of satellite view in urban canyons, indoors or underground. Thus, these measurements cannot be used as reference to evaluate the performance of smartphone-based positioning systems. Nonetheless, GNSS measurements can be used to obtain a coarse estimate of the trajectory followed by the users.

Figure 2 presents the total percentage of each label in each dataset. The first highlight is that the TMD Dataset does not provide unlabelled data. This fact, however, is not an indication that there was no unlabelled data but rather that the authors of the database only published the data for which they had a label. The second highlight is that the Sussex-Huawei Locomotion Dataset has at least three times more unlabelled data, i.e. a *null* label in Figure 2, than labelled data of any other means of transport within their database. The Geolife Trajectory Dataset has approximately 50% of unlabelled data, yet we need to highlight that it is the largest dataset among the three here reviewed.

The fact that the Sussex-Huawei Locomotion Dataset and the Geolife Trajectory Dataset contain so much unlabelled data is yet another example of the challenge that data labelling poses in the creation of public databases for urban mobility. For instance, a significant effort is required to appropriately label all the activities or means of transport within the dataset. Otherwise, it is possible to skip batches of data where no label can be provided, and thus making the unlabelled batch unsuitable to be used in any training.

Figure 3 shows an example of the acceleration norm of the smartphone introduced in the

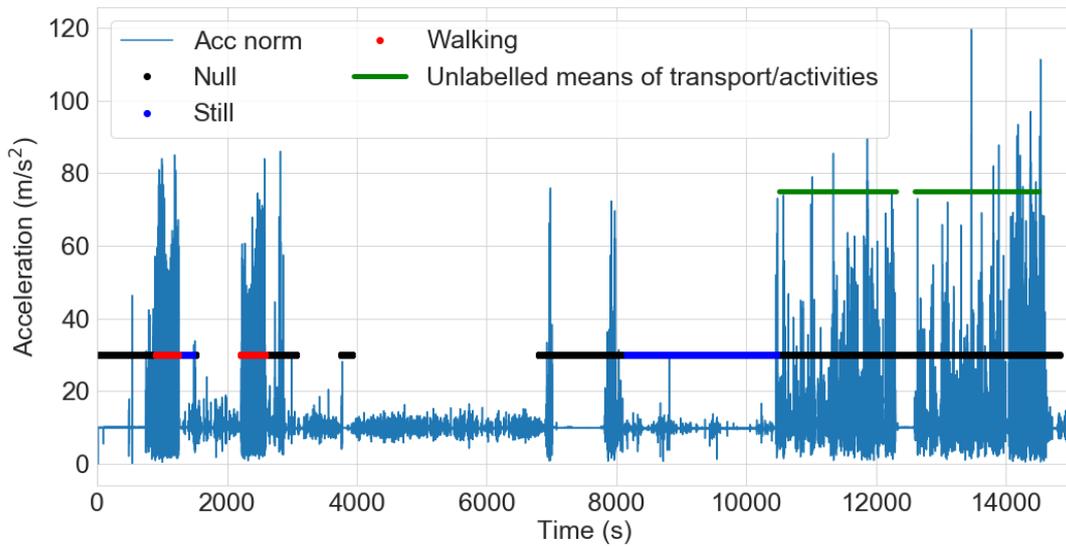


Figure 3: Example of data labelling of the Sussex-Huawei Locomotion Dataset. Only three labels are shown: *null*, *still* and *walking*. The horizontal green lines indicate unlabelled time frames that could contain relevant information of mean of transport or activity.

front pocket of the trousers of the Sussex-Huawei Locomotion Dataset. We plot also three types of labels provided in this dataset: *null*, for no information, *still* for a passenger standing still and *walking*. The horizontal green lines in Figure 3 show clearly that during these time frames the passenger was probably walking or in some mean of transport. However, this period of time is labelled as *null* by the authors. The reason for the lack of a label may have been a mistake during the data recording which caused the authors to discard this batch of data rather than label it.

3.2. Criterion 2: Size of the Database

Table 2 presents the size of the databases regarding different metrics, namely the amount of hours of data, the number of users and the average number of hours per user. The latter has been introduced to ease the comparison of the size of the databases.

We can see significant differences among the sizes of the databases, e.g. the number of hours of data in the Geolife GPS Trajectory Dataset is two orders of magnitude greater than the number of hours of data in the TMD dataset. A similar fact occurs between the number of users in the Geolife GPS Trajectory Dataset and the Sussex-Huawei Locomotion Dataset. The TMD dataset follows a strategy of a few number of users, i.e. three, and a large number of hours per user, i.e. 27.8 hours per user, whereas the latter follows a strategy of a great number of users, e.g. 13, and a few hours per user, i.e. 2.4 hours per user.

Considering the number of hours per user, we can see that the Geolife GPS Trajectory Dataset collected in average 270.8 hours per user whereas the Sussex-Huawei Locomotion Dataset or the TMD Dataset collected one or two orders of magnitude less of data per user. In fact, if we

Table 2

Size of the databases in terms of hours of data, number of users and average hours per user. The values are given per device and not the total number of devices used during the test.

Database	Hours of data	No. of users	Hours/ user
Geolife GPS Trajectory Dataset	48203	178	270.8
Sussex-Huawei Locomotion Dataset	83	3	27.8
TMD Dataset	31	13	2.4

consider an average of 9 hours of data collection per day, the Geolife GPS Trajectory Dataset collected data for approximately one month for each user. In contrast, the Sussex-Huawei Locomotion Dataset collects data for 3 days and the TMD Dataset for less than a day.

An important aspect to highlight is the inconsistency in the amount of hours published by the Sussex-Huawei Locomotion Dataset. The authors claim that they publish approximately 700 hours of data, i.e. taking into account all the data from the four smartphones that each user carries during each day of measurements. However, according to our calculations, the downloadable version of the Sussex-Huawei Locomotion Dataset contains approximately 332 hours of data, i.e. 83 hours of data per day times 4 smartphones carried by each user. Therefore, we believe there is an inconsistency between the amount of data that the authors claimed to have published and the actual amount of data published.

3.3. Criterion 3: Carrying Mode, Means of Transport and Activities

Table 3

Details regarding criterion 3: the carrying mode, the means of transport used and the activities done.

Title	Device	Carrying mode	Means of transport	Activity
Geolife GPS Trajectory Dataset	GNSS logger or smartphone	-	Bike, bus, car, taxi, train, airplane, others	Walking, sports, shopping
Sussex-Huawei Locomotion Dataset	Smartphone	Bag, chest, hand, hip	Bike, bus, car, subway, train	Still, walking, running
TMD Dataset	Smartphone	-	Bus, car, train	Still, walking

Table 3 details the carrying modes of the smartphone, the means of transport and the passenger activities considered in each public database. In the Sussex-Huawei Locomotion Dataset the users carry simultaneously four different smartphones in four carrying modes, namely bag,



Figure 4: Percentages of data normalized to the total amount of data of the means of transport and activities chosen, i.e. *still*, *walking*, *car*, *bus* and *train*.

chest, hand and hip. In the Geolife GPS Trajectory Dataset and the TMD Dataset the users carry only one smartphone and the carrying mode is not specified.

Regarding the means of transport, all databases consider the bus, the car (or taxi) and the train. The Geolife GPS trajectory dataset considers bicycles and airplanes as well and the Sussex-Huawei Locomotion Dataset considers bicycles and other means of transports like subway. None of the selected databases include scooters or trams.

Provided that the focus is urban mobility, it seems natural that the two main activities considered are standing still, e.g. while waiting for a bus or train to arrive, and walking. This is the case for both, the Sussex-Huawei locomotion dataset and the TMD Dataset. The Geolife GPS Trajectory Dataset only considers walking and other interesting activities like shopping or sports, however they do not provide labels of the passenger activity activity, as Table 1 indicates.

Figure 4 shows the percentage of time of each mean of transport common to both the Sussex-Huawei Locomotion Dataset and the TMD Dataset. For a fair comparison, we have normalized the percentages of the Sussex-Huawei Locomotion Dataset and the Geolife Trajectory Dataset to the duration of the common means of transport, namely *car*, *bus* and *train*, and activities, namely *still* and *walking*. In addition, we have considered only the labeled data.

According to Figure 4, both databases are approximately balanced regarding the duration of each means of transport. The largest difference is the amount of *bus* label, where the Sussex-Huawei Locomotion Dataset contains approximately twice the amount of labelled data than the TMD Dataset.

3.4. Criterion 4: Sensors and Sampling Frequency

Table 4 lists the sensors used to record each database, which are a key aspect to determine what application can be implemented with each database. The Geolife GPS Trajectory Dataset includes only GPS data collected from either a smartphone or a dedicated GPS receiver. In

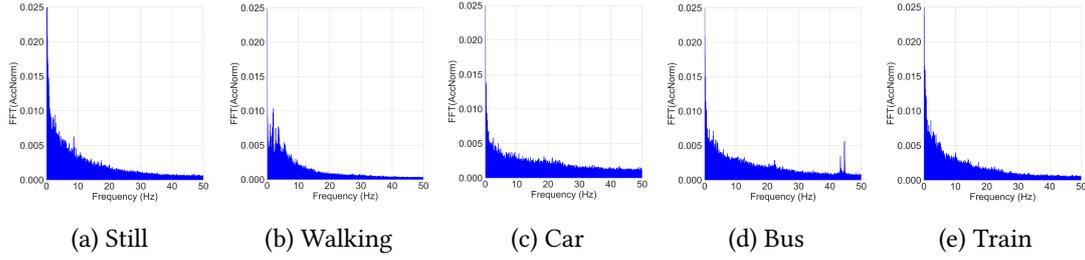


Figure 5: Normalized frequency distribution of the norm of the acceleration.

contrast, the Sussex-Huawei Locomotion Dataset includes the data from all available smartphone sensors. The TMD Dataset includes also all the sensor in a smarthpone, with the exception of the pressure, temperature, altitude and GNSS sensors.

Table 4

Features regarding the sensor data included in each public database.

Title	F_s	Sensor data
Geolife GPS Trajectory Dataset	-	GPS
Sussex-Huawei Locomotion Dataset	100 Hz	Inertial, magnetic, orientation, gravity, linear acceleration, pressure, altitute, temperature, GNSS
TMD Dataset	20 Hz	Inertial, magnetic, orientation, linear acceleration, pressure

One of the main features when recording data is the sampling frequency, since it determines which dynamics are observable through the data. The two dabases we are comparing have a fundamental difference: the sampling frequency of the Sussex-Huawei Locomotion Dataset is one order of magnitude greater than the sampling frequency of the TMD Dataset. This makes the TMD Dataset unsuitable for a frequency analysis of the accelerometer, gyroscope or magnetometer data in the means of trasport that entail motion.

In Figure 5, we show the normalized frequency distribution of the five motion modes for the Sussex-Huawei Locomotion dataset. The most variable distribution is the one of *walking*, where we observe the main frequency component around 1-1.5 Hz, which corresponds approximately to a walking speed of one step per second. The remaining motion modes, including *still*, have a similar frequency profile. This indicates that while the user is inside a moving vehicle, there is no recognisable pattern in how the vehicle accelerates with the exception of the *bus*. In the case of the *bus*, Figure 5 shows two frequency peaks around 45 Hz. The frequency distribution of the norm of the gyroscope shows similar results to ones of Figure 5.

Figure 6 shows an example of the sampling time of two trajectories. The Sussex-Huawei Locomotion Dataset presents a stable sampling time of 10 ms, i.e. a sampling frequency of 100 Hz. In contrast, the TMD Dataset presents an unstable sampling time, which does not correspond to the 20 Hz claimed by the authors. This variability in the sampling time hinders

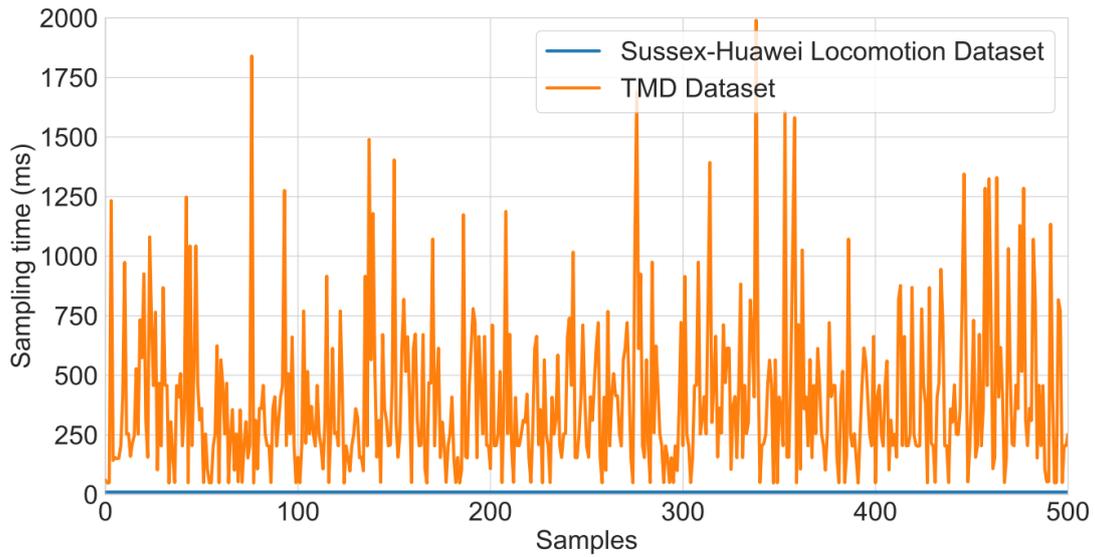


Figure 6: Sampling time of the first 500 samples of two example trajectories of the Sussex-Huawei Locomotion Dataset and the TMD Dataset. The sampling Sussex-Huawei Locomotion Dataset presents a constant sampling frequency of 10 ms.

the use of the TMD Dataset for applications that require a stable sampling time. An alternative is to downsample the TMD Dataset to, e.g. 10 Hz, which may imply the impossibility of observing certain dynamics of some means of transport, which are already unobservable at 20 Hz, see Figure 5.

It is crucial to point out that not all the data within the databases is usable. That is, there is missing data usually due to the sampling frequency of the smartphones. As Figure 7 shows, the TMD Dataset contains a large percentage of missing data in each sensor axis when compared to the missing data in the Sussex-Huawei Locomotion Dataset. In fact, in order to generate this plot we downsampled the data in the TMD Dataset to 10 Hz since the original sampling frequency of 20 Hz was not consistent.

4. Conclusion

In this article, we have defined our concept of urban mobility, identified its main elements and key applications. We have reviewed three public databases regarding four main criteria: the ground truth of each database, the size of the database, the carrying modes, means of transport and activities considered in each database and the sensors recorded. One of the advantage of our work is that the analysis is transferable to other types of databases.

This article allows us to draw four main conclusions. Firstly, it is crucial to develop a robust concept for data collection. In doing so, we avoid basic flaws like missing the recording of labels or ensure that the data collected is consistent regarding sampling frequency. In our work, we have observed that one of the databases actually missed the publications of its labels and that

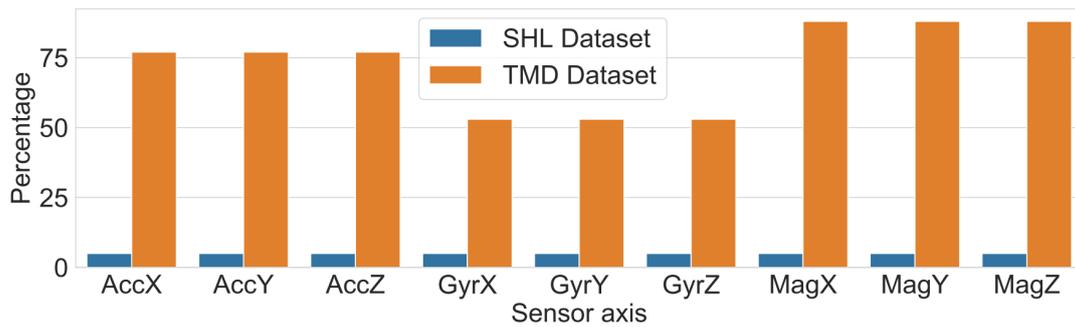


Figure 7: Percentages of NaN values in data in the Sussex-Huawei Locomotion Dataset and the TMD Dataset.

another one claims to have published more data than it actually has.

Second, the ground truth remains one of the biggest challenges in public databases. The community needs to dedicate more effort into developing systematic concepts to collect ground truth, thus increasing the usability of the database. Third, authors should make sure to clean the data before making it public. This effort would facilitate the use of the database by the community.

Finally, the successful development of machine learning and artificial intelligence methods for urban mobility is tightly coupled to the availability of public databases. At this point in time, we believe that urban mobility has barely made its way into big data.

References

- [1] T. Klinger, Moving from monomodality to multimodality? changes in mode choice of new residents, *Transportation Research Part A: Policy and Practice* 104 (2017) 221–237. doi:10.1016/j.tra.2017.01.008.
- [2] European Commission, A concept for sustainable urban mobility plans., https://eur-lex.europa.eu/resource.html?uri=cellar:82155e82-67ca-11e3-a7e4-01aa75ed71a1.0011.02/DOC_4&format=PDF, 2013. Accessed on 28th May 2021.
- [3] R. Abduljabbar, H. Dia, S. Liyanage, S. A. Bagloee, Applications of artificial intelligence in transport: An overview, *Sustainability* 11 (2019) 189. doi:10.3390/su11010189.
- [4] S. Richoz, L. Wang, P. Birch, D. Roggen, Transportation mode recognition fusing wearable motion, sound and vision sensors, *IEEE Sensors Journal* (2020) 1–1. doi:10.1109/jsen.2020.2987306.
- [5] Y. Qin, H. Luo, F. Zhao, C. Wang, J. Wang, Y. Zhang, Toward transportation mode recognition using deep convolutional and long short-term memory recurrent neural networks, *IEEE Access* 7 (2019) 142353–142367. doi:10.1109/access.2019.2944686.
- [6] K. Murphy, *Machine learning : A probabilistic perspective*, MIT Press, Cambridge, Mass, 2012.
- [7] I. H. Witten, *Data mining : Practical machine learning tools and techniques*, Morgan Kaufmann, Burlington, MA, 2011.

- [8] D. Bousdar Ahmed, L. E. Diez, E. Munoz Diaz, J. J. Garcia Dominguez, A survey on test and evaluation methodologies of pedestrian localization systems, *IEEE Sensors Journal* 20 (2020) 479–491. doi:10.1109/jsen.2019.2939592.
- [9] Y. Zheng, X. Xie, W.-Y. Ma, GeoLife: A collaborative social networking service among user, location and trajectory, *IEEE Data(base) Engineering Bulletin* (2010). URL: <https://www.microsoft.com/en-us/research/publication/geolife-a-collaborative-social-networking-service-among-user-location-and-trajectory/>.
- [10] H. Gjoreski, M. Ciliberto, L. Wang, F. J. O. Morales, S. Mekki, S. Valentin, D. Roggen, The University of Sussex-Huawei locomotion and transportation dataset for multimodal analytics with mobile devices, *IEEE Access* 6 (2018) 42592–42604. doi:10.1109/access.2018.2858933.
- [11] C. Carpineti, V. Lomonaco, L. Bedogni, M. D. Felice, L. Bononi, Custom dual transportation mode detection by smartphone devices exploiting sensor diversity, in: *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, IEEE, 2018. doi:10.1109/percomw.2018.8480119.
- [12] M.-C. Yu, T. Yu, S.-C. Wang, C.-J. Lin, E. Y. Chang, Big data small footprint: The design of a low-power classifier for detecting transportation modes, *PVLDB* 7 (2014) 1429–1440.
- [13] S. Hemminki, P. Nurmi, S. Tarkoma, Accelerometer-based transportation mode detection on smartphones, in: *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems - SenSys '13*, ACM Press, 2013. doi:10.1145/2517351.2517367.
- [14] A. Bayev, I. Chistyakov, A. Derevyankin, I. Gartsev, A. Nikulin, M. Pikhletsy, RuDaCoP: The dataset for smartphone-based intellectual pedestrian navigation, in: *2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2019. doi:10.1109/ipin.2019.8911823.
- [15] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, A. Chaintreau, CRAWDAD dataset cambridge/haggle (v. 2009-05-29), Downloaded from <https://crawdad.org/cambridge/haggle/20090529>, 2009. doi:10.15783/C70011.
- [16] N. Akki, New York City bike share dataset, <https://www.kaggle.com/akkithetechie/new-york-city-bike-share-dataset>, 2017.
- [17] Y. Zheng, H. Fu, X. Xie, W.-Y. Ma, Q. Li, Geolife GPS trajectory dataset - User Guide, geolife gps trajectories 1.1 ed., 2011. URL: <https://www.microsoft.com/en-us/research/publication/geolife-gps-trajectory-dataset-user-guide/>, geolife GPS trajectories 1.1, <https://www.microsoft.com/en-us/research/publication/geolife-gps-trajectory-dataset-user-guide/>.
- [18] L. Wang, H. Gjoreski, M. Ciliberto, S. Mekki, S. Valentin, D. Roggen, Enabling reproducible research in sensor-based transportation mode recognition with the Sussex-Huawei dataset, *IEEE Access* 7 (2019) 10870–10891. doi:10.1109/access.2019.2890793.
- [19] C. Carpineti, V. Lomonaco, L. Bedogni, M. D. Felice, L. Bononi, TMD dataset, <http://cs.unibo.it/projects/us-tm2017/download.html>, 2017.