Automated Passengers Trip Phase Recognition and Public Transit Accessibility Level Analysis via Machine Learning Models Using GPS Data

SeyedHassan Hosseini
Department of Civil, Constructional and Environmental Engineering
Università di Roma La Sapienza, Roma, Italy
Email: seyedhassan.hosseini@uniroma1.it

Siavash Pourkhosro
Department of Civil, Constructional and Environmental Engineering
Università di Roma La Sapienza, Roma, Italy
Email: pourkhosro.1919784@studenti.uniroma1.it

Lory Michelle Bresciani Miristice
Department of Civil, Constructional and Environmental Engineering
Università di Roma La Sapienza, Roma, Italy
Email: lorymichelle.brescianimiristice@uniroma1.it

Francesco Viti
Faculty of Science, Technology and Medicine
University of Luxembourg, Mobilab Transport Research Group, L-4364 Esch-sur-Alzette, Luxembourg
Email: francesco.viti@uni.lu

Guido Gentile
Department of Civil, Constructional and Environmental Engineering
Università di Roma La Sapienza, Roma, Italy
Email: guido.gentile@uniroma1.it

Word Count: : 5436 words + 2 table (500 words per table) = 5,936 words

ABSTRACT

Determining true value of passengers waiting time, walking time and distance to and from public transit stops from passengers and public transit side is a key index to assessing potential demand, quality, and effectiveness of public transit services. Current studies heavily rely on household surveys, direct observation techniques, GIS, and telephone-based interviews. However, some investigations use GPS trajectories as a primary data source to detect trip phases, and the main drawback is saving a few trips. This study as a first investigation in this field tries to detect trip phases to and from public transit stops for both passengers and infrastructure side with an automated trip phase recognition algorithm. Passengers waiting time as a critical feature for public transit planning also infer from raw GPS data to analyze the performance of bus stations. Moreover, for bus stations in public transit-based trips the distribution of waiting time, access and egress time, and distance where there is GPS data have been computed. A random forest model, extract transit modes for each segment of a trip, and the results will be used to detect trip phases. There is a lack of labeled standing data in Geolife dataset to detect waiting time. Another novelty is to combine Sussex and Geolife as two large datasets to increase number of estimated modes, especially standing. Our results underscore the effectiveness of our automated approach in predicting different phases.

Keywords: Trip Phase Recognition, Transport Mode Detection, Machine Learning

INTRODUCTION

Urban designers now strive to make cities more walkable, and understanding access/egress part of passenger trips to public transport stops and waiting time at stations have become an essential feature for demand analysis. There are several methods to find access trips phases, such as passenger surveys and interviews (Besser & Dannenberg, 2005; Hess, 2012; Jiang et al., 2012; Psarros et al., 2011; Pueboobpaphan et al., 2022a), GIS-based approaches (García-Palomares et al., 2013a; Moniruzzaman & Páez, 2012), following passengers at stations (Kim, 2015a) and GPS and accelerometer data (Klinker et al., 2014; Southward et al., 2012). Cost and time consuming are the main limitations of the aforementioned methods.

In this study a trip phase recognition algorithm working with a random forest model in background will be present. All traditional methods suffer from human interaction to identify trip stages. We provide an innovative approach by feeding raw GPS inside our proposed framework, and the output will be detailed information about transit modes, passengers' trip stages, and distribution of access, egress, and waiting time for each single transit stop.

The rest of this paper is organized as follows: First, we review studies in transport mode recognition, access/egress distances, and waiting time at transit stops. The methodology section discusses aggregation of two primary datasets, preprocessing employed on GPS and linear accelerometer data, and an explanation of machine learning, and trip phase detection framework. In result and discussion section, the performance of random forest model trained with Geolife and Sussex data will be discussed, and the results of applying proposed framework on extracted trips to validate our framework will be explained. Finally, we conclude and make an outlook for this paper.

LITERATURE REVIEW

The initial step of trip phase recognition algorithm is recognizing passengers' transit mode. Therefore, this section discusses the application of machine and deep learning techniques in TMD and a review of prior studies about access, egress, and waiting time computation.

Transport Mode Detection

Capability of smartphones to capture GPS and inertial sensors has increased during previous years and this source of valuable data can be used to understand passengers behaviour in an urban context such as a commonly used mode of transport by commuters. Machine learning and deep learning models have shown their ability to predict transit modes and are among two frequently methods.

Machine Learning

For the purpose of recognizing transportation modes, several algorithms, including Bayesian belief network, SVM (support vector machine), decision trees, random forest, have been used in previous researches. Most studies followed specific criteria before training models including preprocessing and feature engineering. In this study (Shafique & Hato, 2015) different classifiers including support vector machines, adaptive boosting, decision trees, and random forests categorized segments of accelerometer data into different modes of transit. The primary data source for this study was accelerometer data collected from three cities in Japan. Different steps were taken, such as data pre-processing, generating tests and training subsets, and model evaluation.

GSM data has been used to detect commuter modes (Muller, 2006), and it has been observed that GSM signals are insufficient to differentiate between various transit modes. Multiple research studies (Duncan & Mummery, 2007; Murakami et al., 2004; Wolf et al., 1999) demonstrated that GPS loggers can boost prediction model accuracy compared to conventional techniques such as paper and telephone-based surveys. GPS loggers embedded in mobile phones record precise movements of individuals within urban areas and have become a valuable data source for transit modes.

Other investigations (Xiao et al., 2015; Zheng, Chen, et al., 2010; Zheng et al., 2008) experienced raw GPS data and extracting statistical features and machine learning techniques. The main limitation of these techniques is relying on manual feature extraction and selection, which are prone to human errors. Moreover, the effectiveness of machine learning models is heavily dependent on the extracted features.

Deep Learning

In recent years, deep learning models have been applied in transportation studies. Deep learning models possess the capability to automatically extract features, and due to the aforementioned benefit, they have been extensively employed to detect trip modes. In this study (Dabiri et al., 2019) an unsupervised method has been applied to categorize unlabelled data into different modes and a deep SEmi-Supervised Convolutional Autoencoder classified GPS data into five modes of transport (car, bus, train, bike, walk). In another paper (Liang & Wang, 2017) a deep convolutional neural network was employed to categorize accelerometer sensor data into different modes. The outcomes yielded a model with an accuracy of 94.48 percent in classifying data into seven distinct modes. Magnitudes of different axis were computed to reduce phone rotation noise.

Trip Phase Recognition (Access, Egress, and Waiting Time Detection)

There are different ways to determine how far commuters walk to and from public transport stops and the waiting time inside a bus or metro station. Self-reporting by users regarding walking distance and waiting time is a frequent method investigated in these studies (El-Geneidy et al., 2014; He et al., 2018; Pueboobpaphan et al., 2022b). These approaches are unable to capture the true value of trip phases (distance and duration), and the main difficulty is inaccurate data reporting. As a more accurate method, Google Maps' journey planner was used to extract passengers' trips phases. In this study (Lunke et al., 2021), travel characteristics for both car and public transit trips were obtained from Google journey planner. For each trip, travel time, number of transfers, distances that need to be walked to and between stations, and waiting time during transfers were computed. For a small number of trips, researchers can manually perform calculations. However, dealing with a large number of trips, an automated solution is the most optimal choice.

Another solution to compute passengers' walking distance is following pedestrians at stations up to their destinations. This method is relatively costly and more appropriate for scenarios in a few cases. In this paper (Kim, 2015b), a total of 139 valid records of pedestrians followed in the Glen Park was used and overall average walking distance was computed to 548 m. In another method (García-Palomares et al., 2013b) a GIS-based approach was used, and data was obtained from the Transport Authority of Madrid. All of the data was uploaded into a Geographic Information System. ArcGIS 9.3 was used to analyze trips, and SPSS19 was employed for statistical analysis. They computed statistical features, including mean, standard deviation, and percentiles, to explore the distances individuals walked from their residences to the Metro station.

In another article (Tennøy et al., 2022), the distance and time duration of walking as an access phase to public transport stops in four Norwegian cities were investigated. Participants were asked about the mode by which they travelled to their workplaces, and another survey was gathered at public transport stations. To understand passenger trips in more detail, respondents were interviewed about their actual time and distance of trips to public transport stops, and also duration and distances were computed based on the mean walking speed of 80 meters per minute(Bohannon, 1997).

Several research studies have attempted to calculate the access distance to public transit stations using GPS data. The utilization of GPS trajectories is an effective way to automatically understand passengers' behavior, and prediction accuracies can increase while using position data. In these studies, (Chaix et al., 2014; Hosseini & Gentile, 2022; Zuo et al., 2018), GPS and accelerometer data were used to compute access distance and time of passengers to public transport stops.

(Nygaard, 2016) proposed waiting time of 1145 passengers at 24 bus stations for 16 days. The passengers' arrival time at bus stop was precisely tracked, and waiting time of each passenger was computed as the time between arriving at the station and the departure time of boarded buses. The main limitation is cost and this method can be applied for a low number of trips.

Our proposed approach cover two main gaps. First, automatic trip phase recognition from passengers' trips, using GPS data rather than using handicraft methods. Second, analysing the quality of transit stations at stop level. Distribution of access time and distance, egress time and distance, and waiting time at station will be extracted from raw GPS data automatically. As shown in **(Figure 1)**, our main goal is to develop an automatic framework to find walking time and distance before boarding a bus or metro, waiting at stations, and the last walking time and distance known as egress part for every single trip. Moreover, the distribution data of access, egress, and waiting time for each single bus and metro station for all attracted trips.

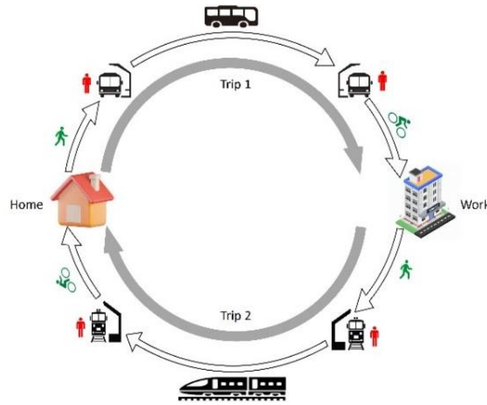

Figure 1 Illustrating trip legs

METHODOLOGY

There is a pre-trained transport mode classification model inside the core component of trip phase recognition algorithm. The initial stage involves training a prediction model using Geolife and Sussex datasets to identify transit modes based on extracted features of each segment and feeding predicted modes as an input to phase recognition algorithm. This section covers 1) Geolife and Sussex datasets, 2) Feature extraction and data preprocessing, 3) Machine learning classification 4) Trip phase recognition algorithms 5) Public transport location detection. In **(Figure 2)**, the main steps that need to be taken from data preprocessing, transport mode detection, and finally trip phase detection are presented.

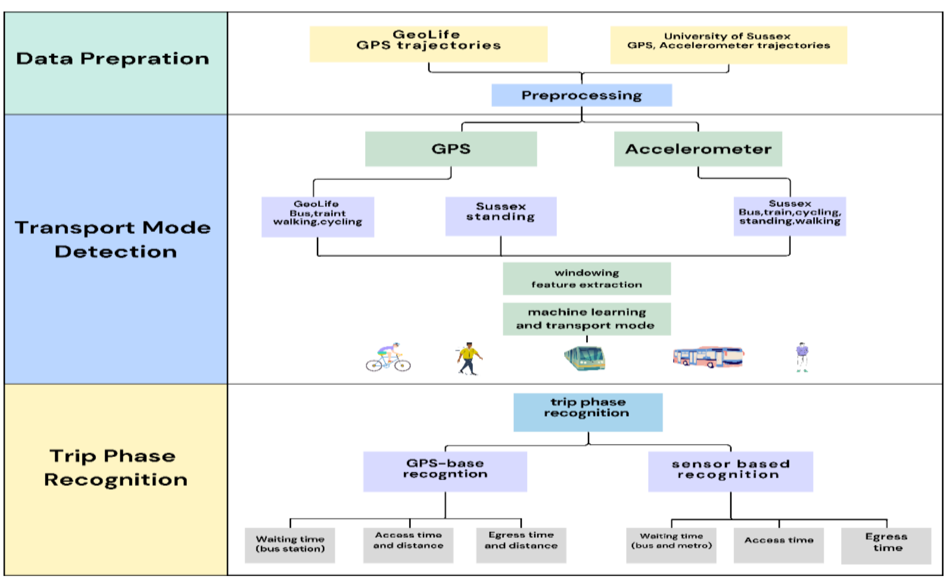

Figure 2 Our Proposed Approach

Geolife and Sussex Dataset:
In this study, we integrated two widely used and public datasets in the field of transport mode detection. There is no labelled data for standing in Geolife (saved from April 2007 to August 2012), and in order to predict waiting time at bus and metro stations, all standing labeled data were extracted from Sussex (2017) and used to train our classification model. Geolife contains (Zheng, Xie, et al., 2010), 189 users with 17,621 trips and 1.2 million kilometres distance and about 48,000+ hours data. More than 91 percent of GPS points were saved with a frequency of 1~5 . Sussex dataset (Gjoreski et al., 2018) benefits from the most common sensors embedded in mobile phones. Three participants collected real-life data with different labels (still,walk,running,car,bus,train,subway).
Data Cleaning:
Two datasets were cleaned up differently. First, all Geolife data outside April 2007–August 2012 were deleted. Moreover, for each mode of transport, certain restrictions on speed and acceleration have been applied following this study (Dabiri et al., 2019), as shown in **(Table 1),** for each data segment, specific thresholds for speed and acceleration were applied, due to the noise of GPS data in urban context. Another two assumptions are a minimum speed for cycling according to the mean value of the minimum speeds of all bicycles (Kassim et al., 2020) and a maximum 1 $\mathrm{m/s^2}$ speed for standing.

TABLE 1. Speed and Acceleration Limitation

Transit Mode	Maximum Speed (m/s)	Maximum Acceleration m/s^2
Bike	12	3
Walking	7	3
Bus	34	2
Train	34	3
Standing	1	-

Trip Identification, Segmentation and Windowing:
Geolife trips were not explicitly differentiated within the dataset, and they saved as a collection of consecutive timestamps. According to this study (Dabiri & Heaslip, 2018) a new trip starts when time interval between two consecutive GPS points exceeds 20 minutes. All trips from two datasets have been picked with following steps, first 212 trips were chosen to validate our proposed trip phase algorithm and not using in training process. From Geolife, 5617 trips and 165 trips from Sussex used to train our machine learning model. All trips were also windowed into equal segments (100 GPS points) to extract statistical features.

Motion Features
Different feature extraction processes have been used on Geolife dataset. First, speed, acceleration and the variations in acceleration between two GPS points as jerks were computed . Bearing rates (direction a vehicle or person when moving) are different in transportation mode headings and were selected as a valuable feature since pedestrians and cyclists can change their routes more easily than buses and trains as buses and trains are obligated to keep to pre-assigned routes. Meaningful extracted features are the mean velocity of each segment, calculated by dividing the cumulative distance of each 100 GPS over total time of each segment and also some other features included. Expectation speed as the mean of all observed speeds in a segment. Moreover, variance, median of speeds , and delta speed is calculated via the maximum speed minus the lowest speed for each section. The first three maximum in each segment and low speed ratio were also meaningful features to detect mode of transport. Low speed ratio is the number of GPS points less than 1.5 m/s^2 divided by total number of exist GPS points in each segment (100 points).This feature helps to detect bus and walking in traffic situation. Additionally, first three maximum acceleration, average acceleration, average jerk, and sum of all distance of each single segment were computed to use as a main features to train our random forest model.

Transport Mode Detection (Machine Learning Classification)
Random forest was selected as our main prediction model to categorized data into different mode of transport. In previous studies(Carpineti et al., 2018; Dabiri & Heaslip, 2018; Efthymiou et al., 2019;

Shafique & Hato, 2015) , random forest arrived to the highest results in transit mode classification problem. This model is an ensemble learning technique that consists of multiple tree classifiers. Each tree in the random forest is built using a subset of the training data and the related available attributes. The process of creating a tree involves recursively splitting the attributes at each node until reaching the final leaf node that contains the ultimate prediction. At each node, the best attribute is selected for decision tree induction based on a measure of impurity, such as entropy (**Equation 1**)

$$Entropy\ (S) = \sum_{i=0}^{c} -p_i \log_2 p_i \tag{1}$$

Where S represents attribute entropy, c is number of different values in features column, and pi is the proportion of S belonging to label i.

Trip Phase Recognition

The primary aims of this research study are twofold: First, we seek to develop an automated process for detecting different trip stages involved in urban trips. Secondly, it aims to extract novel data from raw GPS data, specifically focusing on waiting time experienced at public transport stations and analysing accessibility level of bus and metro stations from raw GPS trajectories saved by passengers . Three primary trip stages are typically access, motorized, and egress. The initial stage of urban trip, known as the access phase, start from origin to the nearest public transit stop. Motorized part includes taking bus or metro and ending trip with walking or cycling known as egress phase.

This study presents a novel framework for an automatic estimating the duration and distance of the access and egress phases, as well as the waiting time at bus and metro stations, using raw GPS data from urban trajectories. First, all data from a single journey are divided into equal segments. GPS points are chunked into 100 GPS points. After pre-processing , the data chunks have been fed into a random forest model, and the output involves predicted mode of transit for each time window. All predicted modes from each single trip are used as an input for trip phase recognition algorithm. In **(Figure 3)**, for each single trip, proposed algorithm start to find the index where there is two consecutive motorized prediction (bus or train) and all segments before motorized part are selected as the access phase of the trip.

Moving from destination towards the origin, two consecutive bus or train segments are labelled as start point of egress phase , all segments from corresponding index up to the end of trip are labelled as the egress section. Time durations and distance for each chuck of data (100 GPS point) and then the sum of all time and distance in each phase will be computed.

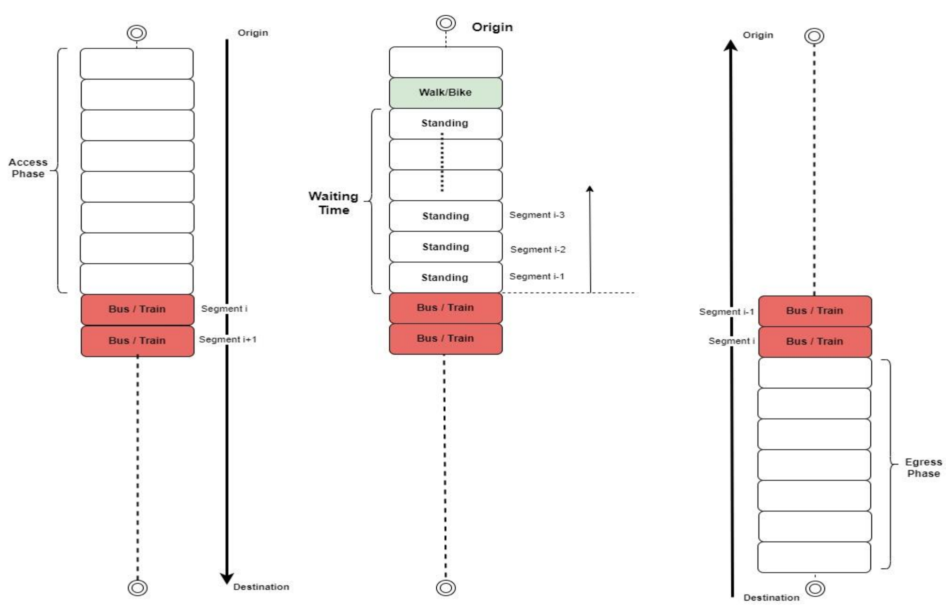

Figure 3 Trip Phase Recognition

$$\text{Access and Egress Time} = \sum_{i=1}^{n} t_i \quad (2)$$

$$\text{Access and Egress Distance} = \sum_{i=1}^{n} d_i \quad (3)$$

In (**Equation 2**) and (**Equation 3**), i is the index of first 100 GPS points in a trip and n is the index of segments before starting motorized. Moreover, in (**Equation 4**), i shows the index of ending walking segments and n demonstrates the index where motorize part start. For each equation, the sum of all segments are access, egress, and waiting time and distance.

$$\text{Waiting Time} = \sum_{i=1}^{n} w_i \quad (4)$$

Bus and Metro Station Detection

Raw GPS data can be used to determine the precise position of bus and train stations to analyze the quality of transit stops. More than 212 public transport-based journeys were captured from GeoLife dataset, with the condition to start with walking or cycling, continue with a bus or train, and then return to walking and cycling. The geographical location of each bus, metro, and train station in this 212 trips was extracted from trip phase algorithm results where motorized part of each trip start. Nearby Search Api supported by Google was used to find the closest public transit stops ranked by distance. Latitude and longitude of each starting point of motorized part were sent to the API and received a response for each single trip.

Geolife data was saved in 2007 and we called google API in 2023. Consequently, Google nearby search API failed to provide a response for all 212 trips and finally from 185 bus-based trips, google API founded the details address data of 81 bus station. Moreover, 27 train-based trips were sent to the API and receive the response as an address for station only for 17 train stations.

RESULTS AND DISCUSSION

In this section , we present the result of transport mode detection and trip phase recognition from passengers and transit station side.

Machine Learning Algorithm

In **(Figure 4),** a confusion matrix shows the result of random forest model to predict transit modes. The overall accuracy of model arrived at 93 percent. Standing as one of the prediction classes reported the highest prediction accuracy (100 percent). The main reason is that after feature extraction, the mean value of speed for all windows labeled as a standing observed to less than 1 m/s^2.Walking and bike arrived at 99 and 99 percent of accuracy respectively. The common problem among transport mode detection investigations is the detection of trains and buses. Traffic congestion and lack of GPS points in urban contexts are some difficulties to predict these modes of transport. Our final model predicted train segments with 90 percent accuracy whereas bus prediction arrived at 0.88 percent.

Figure 4 Confusion Matrix of Random Forest Model

Trip Phase Recognition Algorithm (Passenger Side)

In this study, the accuracy results of our proposed framework were evaluated with the percentage errors approach, which has the advantage of being scale independent. There are some widely used measures to evaluate the accuracy of forecasting models including, Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MdAPE), Root Mean Square Percentage Error (RMSPE), and Root Median Square Percentage Error (RMdSPE). In this paper, MAPE method was used to calculate the errors of the final results, which is defined as the mean of the absolute percentage errors between the forecast values and the actual values. Specifically, MAPE was calculated using this formula(Hyndman & Koehler, 2006)

$$MAPE = \left(\frac{1}{n}\right) \times \sum \left|\frac{actual - forcast}{actual}\right| \times 100 \quad (5)$$

In (**Equation 5**), n is the number of observations in the dataset, actual is the actual value of the variable being forecasted, and forecast is the prediction value of the variable. The MAPE is expressed as a percentage and represents the average magnitude of the percentage errors in the forecasting model. A lower MAPE indicates a more accurate model. Finally, 1 - MAPE was computed to obtain the final accuracy results. This study examines trip phase recognition for a subset of trips taken from the Geolife dataset. To be included, each trip had to begin with a walking or cycling phase, followed by a motorized phase, and end with another walking or cycling phase. From 220 selected trips, Our proposed approach is validated by trips with access time longer than 5 minutes. The access time, access distance, egress time, and egress distance for all selected trips were computed and the results are presented in **(Table 2).**

TABLE 2. Results Sussex

Phase	Prediction Accuracy	MAPE
Access Time	0.8567	0.1432
Access Distance	0.8608	0.1391
Egress Time	0.9469	0.0530
Egress Distance	0.9286	0.071

Trip Phase Recognition (Public Transit Side)

To understand, the accessibility level of each public transit stop and the distribution of passengers waiting time, a unique address ID and its corresponding address were found for each 97 bus and train journey from Google API. All 97 trips were divided into 81 trips labelled as bus-based trips and 17 trips selected as labelled train-based trips. For bus-based trips, a similar address ID was found for 71 trips and it means that all trips attracted by one single bus station. (Figure 5) illustrate the actual location of the bus station extracted from Google Maps.

Figure 5 Haidian District, China, 100190

(**Figure 5 and Figure 6**) shows actual access and egress time and distance and predicted access and egress time and distance for an aforementioned bus station. Trips were recorded by 15 different users where x-axis represent values of time and distance and y-axis shows trips frequency corresponded to x-axis.

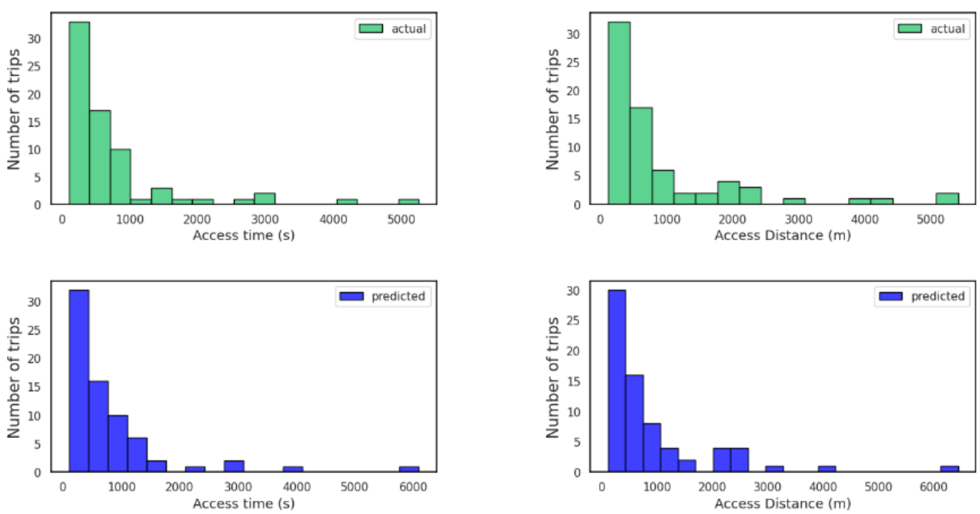

Figure 6 Actual and Predicted Access Time and Distance of 71 Trips Attracted by Single Bus Stop (Address = X8P8+9Q Haidian District, Beijing, China)

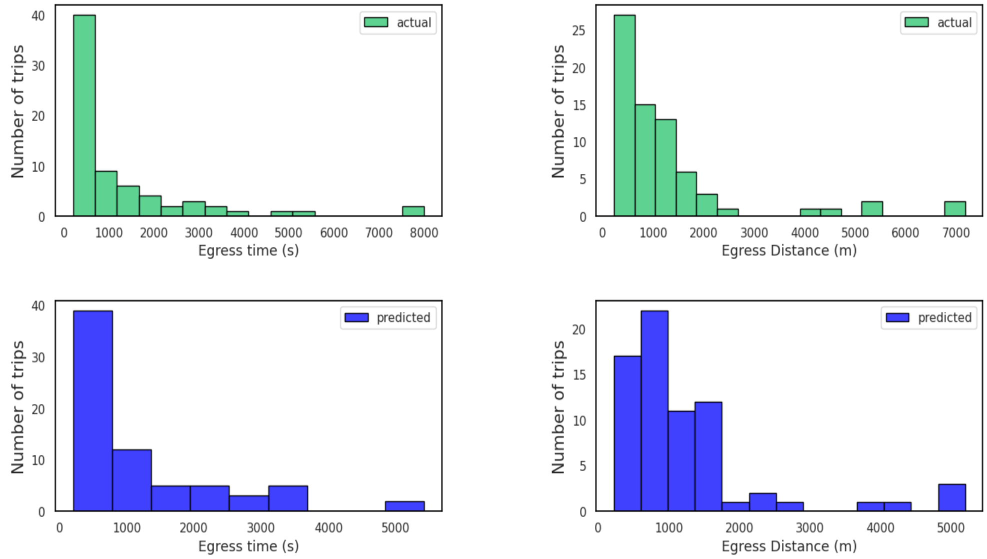

Figure 7 Actual and Predicted Egress Time and Distance of 71 Trips Attracted by Single Bus Stop (Address = X8P8+9Q Haidian District, Beijing, China)

(**Figure 8**) shows a better visualization of actual and predicted values for each phase where x-axis define the range of time and distance in data set and y-axis shows the probability density define as probability per unit value of variables on x-axis.

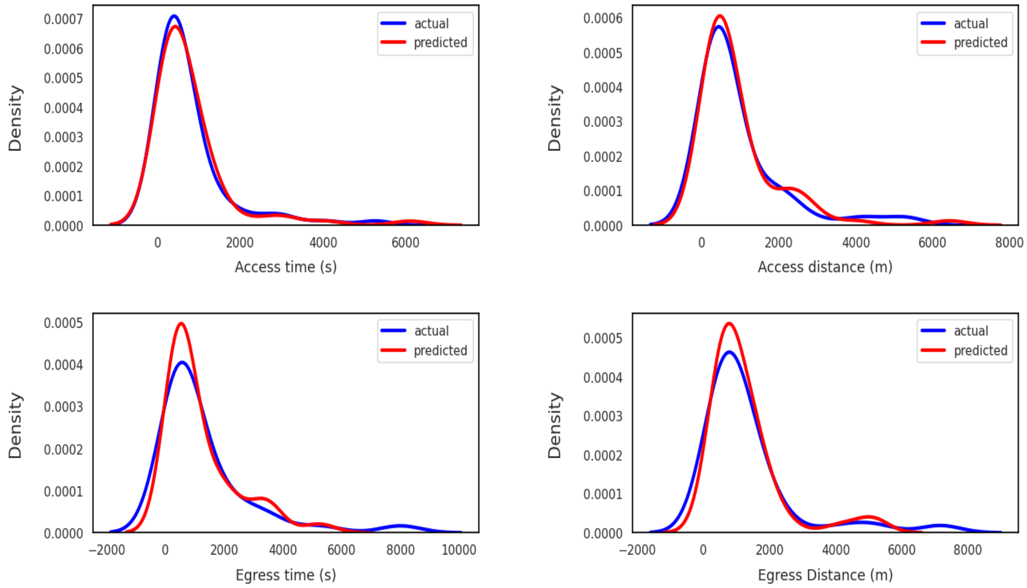

Figure 8 Kernel Distribution Estimation of Access and Egress Phases of 71 Trips Attracted by Single Bus Stop (X8P8+9Q Haidian District, Beijing, China)

Figure 9 Kernel Distribution Estimation of Waiting Time (19 Trips)

Finally, (**Figure 9**) demonstrates a kernel distribution of waiting time at stop level for 19 trips attracted by one single bus station. Our proposed framework is able to present the distribution of waiting time at each bus and train station where there is GPS data.

CONCLUSIONS

Urban planners require more accurate data regarding urban trips from both passengers and public transit sector in order to carry out a comprehensive analysis of the quality of public transport stops. At first, random forest algorithm demonstrate outstanding performance in predicting different transportation modes and the results have been used as an input for our trip phase detection algorithm. Various data collection methods, including surveys and interviews, GIS-based, and tracking passengers at stations are costly and require direct human interventions. Our findings indicate that the proposed method has attained a notable degree of accuracy in different stages and can be used as a viable alternative to traditional methods. The primary outcomes come up from an automated identification of access and egress phases within passengers' journeys and an understanding of the quality of public transit stops from access, egress, and waiting time for each single station. As a first investigation, our research tries to extract more valuable data from GPS data and use them to assess urban public transport stops.

REFERENCES

1. Besser, L. M., & Dannenberg, A. L. (2005). Walking to public transit: steps to help meet physical activity recommendations. *American Journal of Preventive Medicine*, *29*(4), 273–280.
2. Bohannon, R. W. (1997). Comfortable and maximum walking speed of adults aged 20—79 years: reference values and determinants. *Age and Ageing*, *26*(1), 15–19.
3. Carpineti, C., Lomonaco, V., Bedogni, L., Di Felice, M., & Bononi, L. (2018). Custom dual transportation mode detection by smartphone devices exploiting sensor diversity. *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 367–372.
4. Chaix, B., Kestens, Y., Duncan, S., Merrien, C., Thierry, B., Pannier, B., Brondeel, R., Lewin, A., Karusisi, N., & Perchoux, C. (2014). Active transportation and public transportation use to achieve physical activity recommendations? A combined GPS, accelerometer, and mobility survey study. *International Journal of Behavioral Nutrition and Physical Activity*, *11*(1), 1–11.
5. Dabiri, S., & Heaslip, K. (2018). Inferring transportation modes from GPS trajectories using a convolutional neural network. *Transportation Research Part C: Emerging Technologies*, *86*, 360–371.

6. Dabiri, S., Lu, C.-T., Heaslip, K., & Reddy, C. K. (2019). Semi-supervised deep learning approach for transportation mode identification using GPS trajectory data. *IEEE Transactions on Knowledge and Data Engineering*, *32*(5), 1010–1023.
7. Duncan, M. J., & Mummery, W. K. (2007). GIS or GPS? A comparison of two methods for assessing route taken during active transport. *American Journal of Preventive Medicine*, *33*(1), 51–53.
8. Efthymiou, A., Barmpounakis, E. N., Efthymiou, D., & Vlahogianni, E. I. (2019). Transportation mode detection from low-power smartphone sensors using tree-based ensembles. *Journal of Big Data Analytics in Transportation*, *1*, 57–69.
9. El-Geneidy, A., Grimsrud, M., Wasfi, R., Tétreault, P., & Surprenant-Legault, J. (2014). New evidence on walking distances to transit stops: Identifying redundancies and gaps using variable service areas. *Transportation*, *41*, 193–210.
10. García-Palomares, J. C., Gutiérrez, J., & Cardozo, O. D. (2013a). Walking accessibility to public transport: an analysis based on microdata and GIS. *Environment and Planning B: Planning and Design*, *40*(6), 1087–1102.
11. García-Palomares, J. C., Gutiérrez, J., & Cardozo, O. D. (2013b). Walking accessibility to public transport: an analysis based on microdata and GIS. *Environment and Planning B: Planning and Design*, *40*(6), 1087–1102.
12. Gjoreski, H., Ciliberto, M., Wang, L., Morales, F. J. O., Mekki, S., Valentin, S., & Roggen, D. (2018). The university of sussex-huawei locomotion and transportation dataset for multimodal analytics with mobile devices. *IEEE Access*, *6*, 42592–42604.
13. He, J., Zhang, R., Huang, X., & Xi, G. (2018). Walking access distance of metro passengers and relationship with demographic characteristics: A case study of Nanjing metro. *Chinese Geographical Science*, *28*, 612–623.
14. Hess, D. B. (2012). Walking to the bus: Perceived versus actual walking distance to bus stops for older adults. *Transportation*, *39*, 247–266.
15. Hosseini, S. H., & Gentile, G. (2022). Smartphone-Based Recognition of Access Trip Phase to Public Transport Stops Via Machine Learning Models. *Transport and Telecommunication*, *23*(4), 273–283.
16. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688.
17. Jiang, Y., Zegras, P. C., & Mehndiratta, S. (2012). Walk the line: station context, corridor type and bus rapid transit walk access in Jinan, China. *Journal of Transport Geography*, *20*(1), 1–14.
18. Kassim, A., Tayyeb, H., & Al-Falahi, M. (2020). Critical review of cyclist speed measuring techniques. *Journal of Traffic and Transportation Engineering (English Edition)*, *7*(1), 98–110.
19. Kim, H. (2015a). Walking distance, route choice, and activities while walking: A record of following pedestrians from transit stations in the San Francisco Bay area. *Urban Design International*, *20*, 144–157.
20. Kim, H. (2015b). Walking distance, route choice, and activities while walking: A record of following pedestrians from transit stations in the San Francisco Bay area. *Urban Design International*, *20*, 144–157.
21. Klinker, C. D., Schipperijn, J., Christian, H., Kerr, J., Ersbøll, A. K., & Troelsen, J. (2014). Using accelerometers and global positioning system devices to assess gender and age differences in children’s school, transport, leisure and home based physical activity. *International Journal of Behavioral Nutrition and Physical Activity*, *11*(1), 1–10.
22. Liang, X., & Wang, G. (2017). A convolutional neural network for transportation mode detection based on smartphone platform. *2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, 338–342.
23. Lunke, E. B., Fearnley, N., & Aarhaug, J. (2021). Public transport competitiveness vs. the car: Impact of relative journey time and service attributes. *Research in Transportation Economics*, *90*, 101098.

24. Moniruzzaman, M., & Páez, A. (2012). Accessibility to transit, by transit, and mode share: application of a logistic model with spatial filters. *Journal of Transport Geography*, *24*, 198–205.
25. Muller, I. A. H. (2006). Practical activity recognition using GSM data. *Proceedings of the 5th International Semantic Web Conference (ISWC). Athens*, *1*(8).
26. Murakami, E., Taylor, S., Wolf, J., Slavin, H., & Winick, B. (2004). GPS applications in transportation planning and modeling. *The Travel Model Improvement Program Connection Newsletter*, 1–3.
27. Nygaard, M. F. (2016). Waiting Time Strategy for Public Transport Passengers. *61*. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2405551
28. Psarros, I., Kepaptsoglou, K., & Karlaftis, M. G. (2011). An empirical investigation of passenger wait time perceptions using hazard-based duration models. *Journal of Public Transportation*, *14*(3), 109–122.
29. Pueboobpaphan, R., Pueboobpaphan, S., & Sukhotra, S. (2022a). Acceptable walking distance to transit stations in Bangkok, Thailand: Application of a stated preference technique. *Journal of Transport Geography*, *99*, 103296.
30. Pueboobpaphan, R., Pueboobpaphan, S., & Sukhotra, S. (2022b). Acceptable walking distance to transit stations in Bangkok, Thailand: Application of a stated preference technique. *Journal of Transport Geography*, *99*, 103296.
31. Shafique, M. A., & Hato, E. (2015). Use of acceleration data for transportation mode prediction. *Transportation*, *42*, 163–188.
32. Southward, E. F., Page, A. S., Wheeler, B. W., & Cooper, A. R. (2012). Contribution of the school journey to daily physical activity in children aged 11–12 years. *American Journal of Preventive Medicine*, *43*(2), 201–204.
33. Tennøy, A., Knapskog, M., & Wolday, F. (2022). Walking distances to public transport in smaller and larger Norwegian cities. *Transportation Research Part D: Transport and Environment*, *103*, 103169.
34. Wolf, J., Hallmark, S., Oliveira, M., Guensler, R., & Sarasua, W. (1999). Accuracy issues with route choice data collection by using global positioning system. *Transportation Research Record*, *1660*(1), 66–74.
35. Xiao, G., Juan, Z., & Zhang, C. (2015). Travel mode detection based on GPS track data and Bayesian networks. *Computers, Environment and Urban Systems*, *54*, 14–22.
36. Zheng, Y., Chen, Y., Li, Q., Xie, X., & Ma, W.-Y. (2010). Understanding transportation modes based on GPS data for web applications. *ACM Transactions on the Web (TWEB)*, *4*(1), 1–36.
37. Zheng, Y., Li, Q., Chen, Y., Xie, X., & Ma, W.-Y. (2008). Understanding mobility based on GPS data. *Proceedings of the 10th International Conference on Ubiquitous Computing*, 312–321.
38. Zheng, Y., Xie, X., & Ma, W.-Y. (2010). GeoLife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, *33*(2), 32–39.
39. Zuo, T., Wei, H., & Rohne, A. (2018). Determining transit service coverage by non-motorized accessibility to transit: Case study of applying GPS data in Cincinnati metropolitan area. *Journal of Transport Geography*, *67*, 1–11.