Journal of Traffic and Transportation Management
Volume 2, No. 2 (2020) pp. 25-33

Predicting Travel Behavior by Analyzing Mobility Transactions

J. Slik[a]*, S. Bhulai[b]

[a]*Pon, Amsterdam, Netherlands, 1076 AM*
[b]*Vrije Universiteit Amsterdam, Amsterdam, Netherlands, 1081 HV*

Abstract

Urban planning can benefit tremendously from a better understanding of where, when, why, and how people travel. Through advances in technology, detailed data on the travel behavior of individuals has become available. This data can be leveraged to understand why one prefers one mode of transportation over another one. In this paper, we analyze a unique dataset through which we can address this question. We show that the travel behavior in our dataset is highly predictable, with an accuracy of 97%. The main predictors are reachability features, more so than specific travel times. Moreover, the travel type (commute or personal) has a considerable influence on travel mode choice.

Keywords: *Mobility Analysis, Travel Mode Choices, Trip Data, Machine Learning, Logit Model*

1. Introduction

The analysis of mobility is of key importance to tackle major urban planning challenges [1]. It is projected that by 2030, today's 1.2 billion global car fleet could double [2]. This has a big impact on the dynamics in urban areas: traffic delays, unhealthy smog levels, noise, routine irritations of urban lives, and others. The introduction of other modes of transportation can help to alleviate these mobility challenges. One can think of public transport, bikes, and trains, as well as shared services or combinations thereof. In addition, [3] introduces many other (sub)urban transformations and transit-oriented developments to improve urban lives. At the same time, it is argued that in order to decide where to invest in requires a good understanding of the travel behavior of individuals and their underlying reasons.

Much research has been devoted to mobility patterns within cities. The gravity model is the prevailing framework for discovering and modeling these patterns [4]. This model is rather data-intense, in the sense that it requires specific parameters fitted from a continuous collection of traffic data. When these measurements are not available or not complete, the gravity model cannot be applied. In response to that, trip distribution models have been introduced [5, 6]. However, these models also rely on context-specific parameters. In [7], the authors explain how radiation models [8] can model mobility patterns based on only the population's spatial distribution as input.

Through advances in technology, trip information per individual can be recorded more precisely. E.g., telematics modules in cars can record the time and location when a car starts or turns off the engine. Travel cards issued by companies store information on the usage of public transport, bikes at the start of the trip, and the end of the trip. Such detailed information per individual allows one to perform mobility analysis on a much more personalized level. The previously mentioned mobility models cannot handle such a granularity of information and cannot be applied. Models have been introduced to analyze trip data from one mode of transport to reveal travel patterns [9, 10]. However, such analyses do not reveal the underlying reasons why an individual chooses one mode of transport over another one. In the literature, there is a gap in analyzing these mobility choices at such granularity due to the lack of high-quality data.

In this paper, we analyze the travel mode choices based on a unique dataset consisting of mobility transactions on an individual level. This allows us to follow individuals throughout the day and the year. We propose a measure to quantify the speed of any transport type for any neighborhood and show how to combine relevant external data sources. By doing so, we can accurately predict travel mode choices. Our model shows the opportunity to influence travel mode choices and gives the potential to simulate the impact of infrastructure changes.

This paper is organized as follows. Section 2 describes the dataset and the data preparation. Section 3 illustrates the impact of the data preparation and introduces the models for understanding mobility. The results of the analysis are discussed in Section 4. Section 5 draws a conclusion and discusses on the results. Lastly, Section 6 describes research opportunities.

2. Data

The data used in this research is gathered from multiple sources by a company that provides mobility to customers through a mobility card. Individuals can use different travel modes using this card. The card enables one to use a car, all forms of public transport, a taxi, car-sharing, and bike-sharing. The customers using this service are all employed by one specific company in the Netherlands, which has multiple offices spread throughout

* Corresponding author. E-mail: jesper.slik@pon.com

DOI: 10.5383/JTTM.02.02.004

the country. Travel usage is registered through automated systems and stored as transactions.

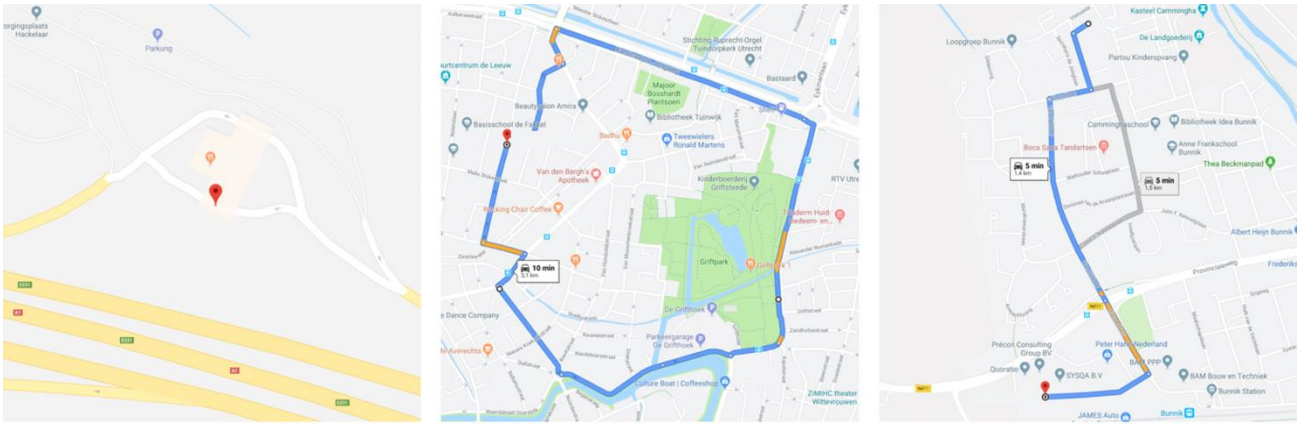

Fig. 1. Three categories of trips that are filtered from the data: (a) gas stations; (b) round trips; (c) relatively short trips.

Car transactions are registered automatically by a built-in telematics module in the cars, which have trip registration as their sole purpose. All collected transactions are considered private; however, under strict conditions, analysis of this data is allowed. Consequently, due to these conditions, we cannot directly determine the identities behind the person identifiers in the data.

The full dataset contains over half a million mobility transactions from over a thousand employees. This concerns a period of one entire year, 2018. We filter the data by individuals having access to both public transport and a telematics-enabled car. For each mobility transaction, we know the transport type, start and end date and time, start and end location, and costs. We have aggregated statistics for each individual, such as the city of residence, lease category, and commute mileage. Other individual-specific attributes such as age, gender, and fuel compensation are not taken into account for privacy reasons.

Analyzing such a dataset imposes various challenges. In the next section, we discuss the data cleaning. Then we explain how to estimate statistics on the alternative travel mode. Repeating choices will be discussed in the subsequent section, followed by a method to compute the start and end locations of public transport transactions more accurately. We conclude with an examination of relevant external data sources.

2.1. Data Cleaning

The dataset consists of all transactions of all individuals for the year 2018. However, we do not consider all transactions in this research for various reasons. Specifically, three categories of transactions are filtered: transactions to gas stations, transactions with a similar start and end location, and relatively short transactions. Figure 1 visualizes the three categories. We filter out these transactions for reasons that we will explain next.

Transactions departing or ending at gas stations are filtered as they are not considered the `true' start or destination of a transaction. Typically, it is a compulsory stop en route to a different destination. As all locations of gas stations in the Netherlands are publicly available, these transactions can be filtered easily.

Transactions with a similar start and end location (within one transaction) are difficult to analyze as we do not know what happened during the trip. It is impossible to calculate accurate statistics on alternative travel modes. Therefore, we filter transactions of which the start and end location are within a distance of 200 meters.

Relatively short transactions are not considered either, since the availability of alternative modes can be questioned. Additionally, our focus is not on these short trips. For example, if a transaction by car exists with a length of 1 kilometer, we could compute statistics on public transport on this transaction. However, the resulting statistics could be similar to those of walking. Therefore, we decided to filter all trips with a distance shorter than 4 kilometers.

2.2. Estimating Statistics on the Alternative

The transactions contained within the dataset show statistics on the chosen mobility type. For scenario analysis, we are interested in the statistics on the alternative. Specifically, we are interested in the travel time, distance, and CO2 emissions. We compute these by using external APIs. A wide variety of them is available, including the Open Routing Service, Google Maps, Bing Maps, City Mapper, and Tripgo. We chose to use the HERE API [11] for estimating statistics on car alternatives and the TravelTime Platform API [12] for public transport. This choice is made based on the availability and cost of the services. An estimate of CO2 emissions is made for cars by analyzing all historical transactions of the individuals (including liters tanked), and for public transport by using statistics from research on emission factors in the Netherlands [13].

The performance of these APIs can be measured by requesting statistics on known transactions and comparing those to the observed statistics. Figure 2 shows a comparison of observed and estimated travel times for (a) cars and (b) public transport. The red line is used as a reference and has slope 1 in both graphs. Interestingly, estimated car travel times are slightly underestimated. This can be partially explained by congestion. Concerning public transport, we can see that there is a significant variance in the observed travel time for similar estimates. This can be explained by irregularities in schedules and varying arrival times of individuals at stations.

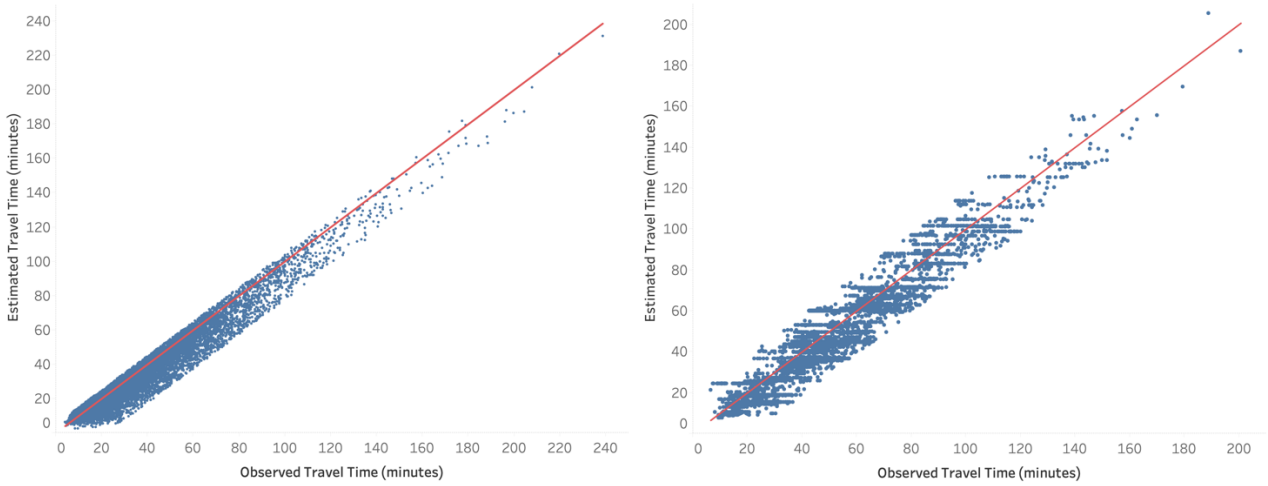

Fig. 2. Observed vs estimated travel time: (a) car; (b) public transport.

To improve the accuracy of the estimates, we create a linear model on top of the API estimation. For cars, this model is based on the API estimation, start hour of the transaction, and a weekend indicator. For public transport, this model is based on the API estimate and start hour of the transaction. The choice for these features is based upon the significance of their results. The linear model results in an increase of the coefficient of determination R^2 from 0.835 to 0.873 for cars and from 0.759 to 0.814 for public transport.

Figure 3 (a) shows the model coefficients for re-estimating car trips. Interestingly, there is a clear relationship between the start hour of a trip and the fitted model coefficient on the data. During rush hours, in the morning and in the afternoon, the model coefficients are positive; outside rush hours the coefficients are negative. The highest model coefficient corresponds to the hour that is often considered as the hour with the highest congestion level: 16:00. To simplify the model, we chose to group all hours during the night in a category labeled as '0'.

2.3. Start and End Locations

The start and end location of the transactions are difficult to interpret, as they only provide an estimate of the `true' start and end location. For cars, these locations are generally close by, as parking spots are widely available. However, for public transport, this can be assumed to be less accurate as the transactions provide us with the check-in and check-out locations. These locations are always at stations.

To improve this estimation, we re-estimate the locations of public transport transactions. Using the TravelTime Platforms Time Map feature, we calculate the area that can be reached from each station in the Netherlands by 10 minutes of cycling. Combining this with a data source containing coordinates of all addresses in the Netherlands (BAG), we compute all reachable addresses for all stations in the Netherlands. Next, we sample one address for each transaction concerning public transport from all reachable addresses from the corresponding station. We use this address instead of the address that is shown in the raw data.

This process is visualized in Figure 4. On the left (a), all addresses within 10 minutes of cycling from a station are visualized. The station is located at the orange cross, all addresses reachable within the 10-minute threshold are colored orange and the others blue. The orange area resembles a circle, however, this is not necessarily true. In some areas, there might be natural obstacles or little infrastructure. This will influence the travel time towards these areas, and hereby the shape of the area. On the right (b), the sample density is shown. In some neighborhoods, the addresses are more densely packed and, therefore, should have a higher probability of being selected. If we sample at random from all known addresses, we automatically correct for the population density.

2.4. Repeating Choices

A potential challenge with fitting models on the transactions is that the model becomes biased towards choices that are often repeated. For example, a person might decide once on his or her commuting transport mode and hereafter execute it hundreds of times. In contrast, an occasional trip to a specific destination might only appear once. We want to present a dataset with variety to the models in order to prevent the models from being biased towards choices that are repeated often. We cannot filter based on the trip type, as individuals can commute to multiple offices, can have business trips to similar locations, or change behavior over time.

To increase the variety of the data, we remove rows that are highly similar to others. For this, we define the distance $d(r_i,r_j)$ between record r_i and r_j as in Equation (1):

$$d(r_i, r_j) = \sum \frac{f * |r_i - r_j|}{s} \quad (1)$$

In this equation, f is the vector with the feature importance acting as a weight, and s is the feature standard deviations. The vector f is based on the features of the model developed in Experiment 2 in Section 3.2. We sort the data by date and time, partition by person id, and remove all rows that have a distance lower than a certain threshold, for index j > i.

Determining the threshold implies balancing the variety and availability of the data. Having a dataset with a large variety implies having low volume; having a dataset with high volume implies having little variety. Figure 3 (b) visualized this trade-off by showing the relation between the threshold and dataset size. A high similarity score as threshold implies more choices are seen as similar, which results in lower data volume. We empirically set the threshold to 0.04, which keeps roughly 30% of the data.

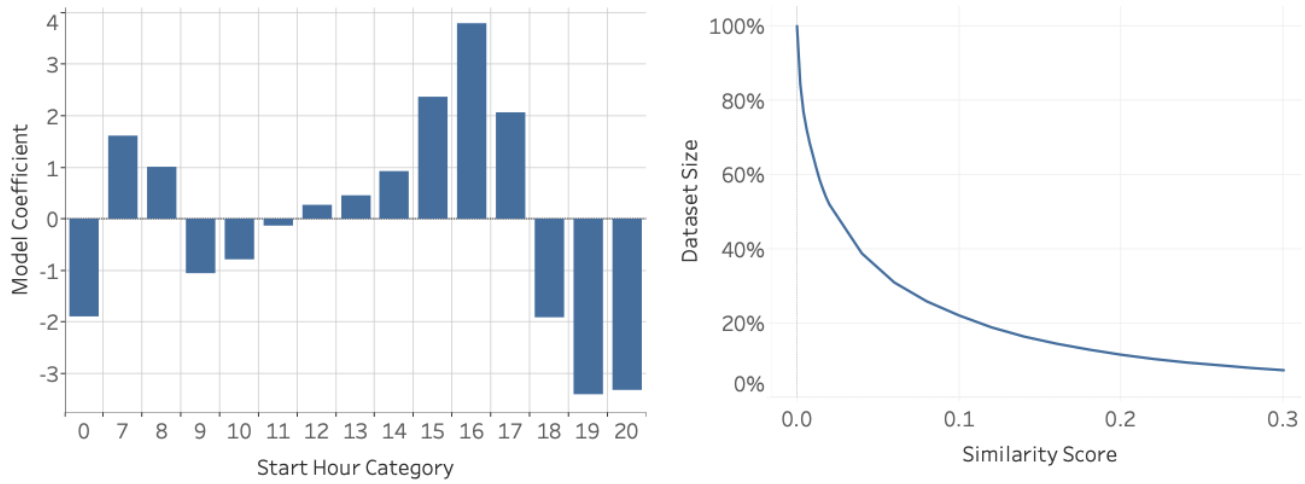

Fig. 3. Data processing: (a) model coefficients for re-estimating car trips; (b) trade-off between variety and availability of data.

Table 1 shows an example of how the repeating choices are identified. Taking the first row as a reference, the distance between the following rows is computed according to Equation 1. Taking the threshold of 0.04, rows 2, 3, and 4 will be filtered. This does not guarantee the other rows will not be filtered, as the process will be repeated from row 5 onward.

2.5. External Sources

Besides the transactions and personal statistics, we use external data sources to calculate features. This concerns data on congestion, reachability of neighborhoods, and weather conditions.

Congestion is measured using data from the Dutch Nationale Databank Wegverkeersgegevens (NDW). The NDW continuously measures the speed and volume of cars driving over their sensors on federal roads. This concerns 37 thousand sensors across the Netherlands, which report statistics by the minute. Figure 5 visualizes this data. On the left (a), it shows the measurement locations on a highway around Amsterdam. The congestion level is indicated by color, red meaning high congestion and green meaning low congestion levels. Clearly, the highway is congested in a single direction. On the right (b), it shows the normalized speed on all measurement sites between two Dutch cities (Amsterdam and Almere) in opposite directions. The graph shows that congestion in the morning is heavy in one direction, whereas the opposite direction is hardly congested. This data allows us to quantify congestion on the road at a specific time, between the start and end locations, and in the corresponding direction for all transactions in our dataset.

A second feature we compute is the so-called `reachability' of neighborhoods. This captures the general speed of a particular transport type in a certain area. To compute these, we start by taking the definition of a neighborhood from the Dutch CBS. These neighborhoods are similar to postal code definitions. However, it provides a higher detail level and is still feasible for this analysis. For each neighborhood in the Netherlands, we compute the speed (distance over time) at which the 4,000 surrounding neighborhoods can be reached by both a car and public transport. To compute the travel time, we use the APIs selected in Section 2.2. To compute distance, we take the celestial distance. Both measures are calculated between the building lying closest to the center point of the neighborhoods. Next, we average these speed values to gain one numeric value per neighborhood. The resulting measure is visualized in Figure 6. It shows the reachability of the neighborhoods in the Randstad region in the Netherlands by (a) public transport and (b) car.

Finally, we add statistics on weather conditions. This includes wind, rain, temperature, sunshine, wind speed, and rain duration. These statistics are historically made available by the Dutch KNMI. They are measured on 50 locations spread throughout the Netherlands. For each transaction, we take the measurement values from the nearest station to the middle coordinate of the transaction.

Table 1. Quantifying repeating choices

Time PT	Cost PT	CO2 PT	Time car	Cost car	CO2 car	Similarity
77.42	22.3	7.82	42.42	4.9	7.83	n/a
77.36	22.3	7.82	42.29	4.91	7.85	0.001
77.78	22.	7.82	42.24	4.91	7.84	0.002
78.59	22.3	7.82	42.1	4.9	7.83	0.006
42.26	9.61	3.37	28.56	2.97	4.74	0.788
92.86	15.27	0.41	63.37	8.75	13.99	0.922
138.57	32.63	0.88	109.28	19.83	31.69	2.036
197.97	49.01	7.77	120.82	23.05	36.84	2.421

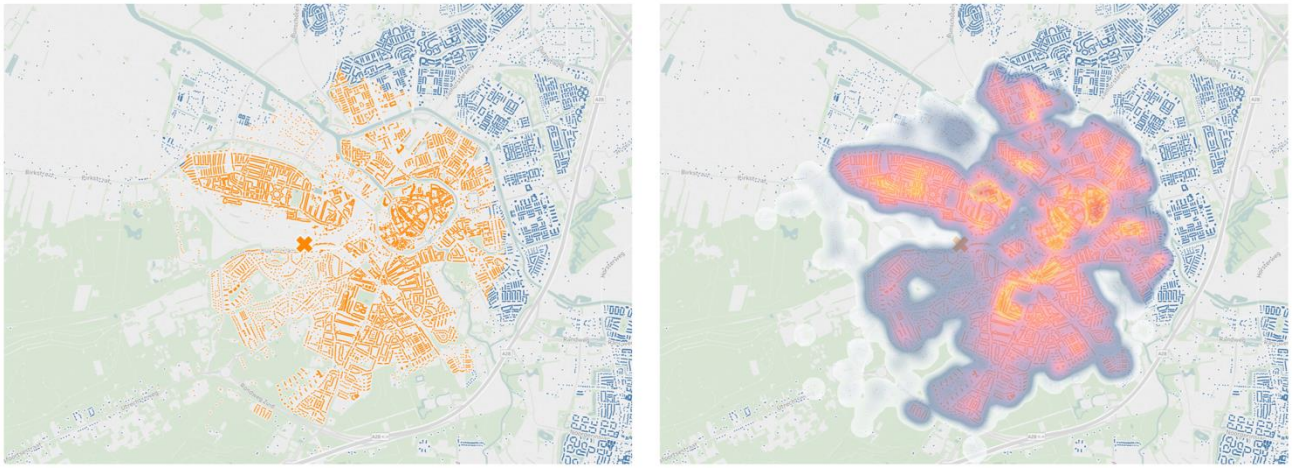

Fig. 4. Estimating start and end locations: (a) addresses inside (orange) and outside (blue) 10 minutes cycling from a station (orange cross); (b) sample density of households within 10 minutes cycling range.

3. Numerical Experiments

The data processing steps are pre-requisites to understand better and to predict travel behavior. Therefore, we create multiple models for describing the mobility transactions, define multiple experiments to assess the impact, and evaluate them accordingly. These three steps are described in this section.

3.1. Models

We fit five different models on the mobility transactions. The first model is a model familiar to the transport science field, a logit choice model. The other four models are commonly used in the machine learning field: logistic regression, feedforward neural network, gradient boosted decision trees, and random forest. The alternatives for all models are using the car or using public transport, making it a binary problem. The models are evaluated using 5-fold cross-validation and are implemented in Python. The logit choice model using the PandasBiogeme [14] package, the logistic regression, neural network, and random forest using the scikit-learn [15] package, and the gradient boosted trees using the xgboost [16] package. The logit model is highly similar to the logistic regression model, however, they are implemented through different libraries. The model parameters are determined by a grid search procedure, the feedforward neural network performs best with a single hidden layer containing 10 neurons.

3.2. Experiments

To highlight the impact of the data processing, we define four experiments. We start by fitting models on relatively raw data and step-by-step work through the processing steps to highlight their impact. The final experiment can be considered the most realistic and important.

- In Experiment 1, we take the raw data, filter it (Section 2.1), calculate features on the alternative (Section 2.2), and fit the models. The features used by the models are the travel time, costs, and CO2 emissions of both alternatives (car and public transport).
- In Experiment 2, we take the data from Experiment 1, change the start and end locations of public transport trips (Section 2.3), re-calculate the features on the alternatives, and re-fit the models of Experiment 1.
- In Experiment 3, we take the data from Experiment 2, filter by removing repeating choices (Section 2.4), and re-fit the models of Experiment 2.
- In Experiment 4, we take the data from Experiment 3, add features, remove correlated features, and re-define the models of the previous experiments. Added features are as described in Section 2.5, combined with the aggregated personal statistics, and a classification of the transaction as indicated by the individuals (private, commute, or business).

3.3. Evaluation

For each experiment, we fit the models on the data. We evaluate the performance by measuring the accuracy and the AUC [17]. Additionally, we use SHAP [18] to measure the impact of all features for the machine learning models. The SHAP value represents the impact on the model output. For each feature, it holds that the larger its absolute SHAP value, the larger its importance.

4. Results

Table 2 shows the accuracy and the AUC of all four experiments. As a benchmark, always predicting the car will result in an accuracy of 81% and an AUC of 0.50. The random forest model shows to have the highest performance, accurately predicting the mobility choice of 97% of the transactions in the last experiment. Generally, the accuracy and the AUC of all models in all experiments is relatively high, with minima of 88% and 0.91, respectively. The random forest outperforms the other models in most experiments, closely followed by the gradient boosted trees.

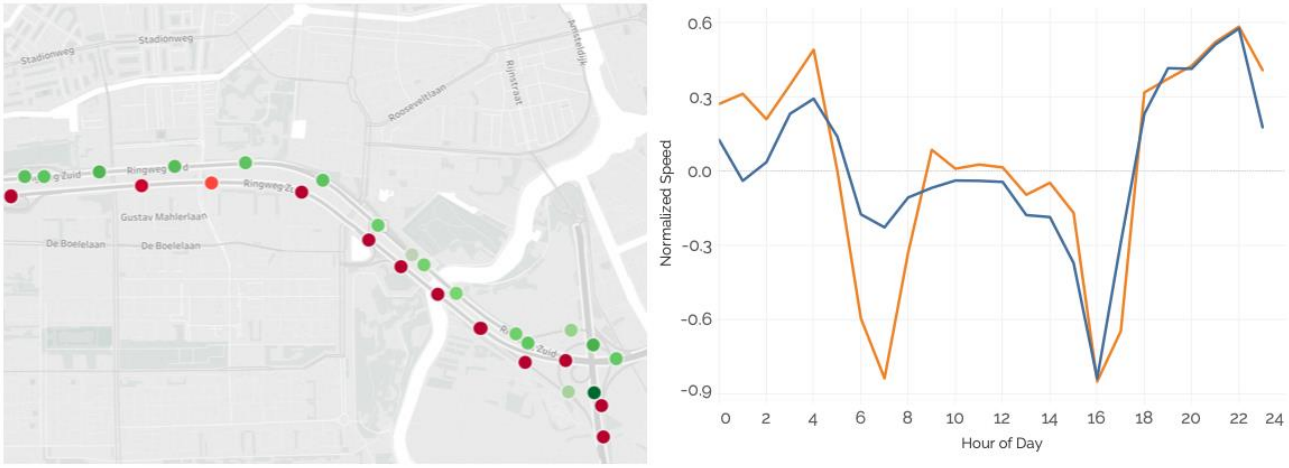

Fig. 5. Measuring congestion: (a) measurement locations: having high (red) and low (green) congestion levels; (b) congestion levels: from Amsterdam to Almere (blue) vs from Almere to Amsterdam (orange).

As expected, the performance of the models is high in Experiment 1 and decreases in Experiments 2 and 3. This can be explained as in the first experiment, unrealistic start and end locations and repeating choices help the models boost their performance. In Experiment 4, the performance increases, showing the relevance of the external data sources. Especially the random forest and the gradient boosted trees benefit. Surprisingly, the performance of the neural network decreases in Experiment 4 and achieves lower performance than the logistic regression. This might be attributed to overfitting and can possibly be prevented by a more advanced parameter selection procedure.

Figure 7 shows the feature importance for the random forest model of Experiment 4. On the left (a), the mean absolute importance is shown. Interestingly, all features related to the reachability of the neighborhood are important. Hereafter, the indication for commute has a large impact. The specific travel times (time PT, time car) show to be relatively unimportant in comparison. Congestion also seems to have relatively little effect, as well as features corresponding to weather conditions.

On the right (b), the impact of all feature values on the model outcome is shown. A negative SHAP value (negative impact on the model outcome) implies that the prediction tends to go towards public transport, a positive value towards the car. The more extreme the SHAP value is, the higher the impact of the feature on the model output. The results confirm our intuition, a low ratio (car over public transport) of reachability results in a large negative impact. A low reachability by public transport or a low travel time by car result in a large positive impact. Interestingly, personal trips are preferred by car and commute trips by public transport. Also, a low commuting distance implies a preference for public transport. Lastly, congestion seems to have a positive model impact, implying a slight preference for cars when roads are congested.

5. Conclusion and Discussion

Our results show that the travel behavior in our dataset is highly predictable, as we can predict the individual transactions with an accuracy of 97%. Compared to the benchmark of 81%, this is a significant increase. Additionally, by quantifying the importance of all features, we show insights into why travel behavior is predictable.

The main features contributing to the models are our proposed reachability features. This conforms to our intuition; however, surprisingly, their importance is much higher than the specific travel times of the specific transactions. The general reachability of an area is more important for modal choice than the specific reachability. General reachability refers to the reachability of the neighborhoods, specific reachability to that of the specific trip the person is planning. People might not take the effort to check the travel times for the alternative and decide based on their general knowledge of the destination (neighborhood) reachability. This insight can be seen as an opportunity to inform individuals to stimulate behavioral change proactively.

Table 2. Experimental results: accuracy and (AUC)

Model	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Binary Logit	98% (0.99)	91% (0.91)	88% (0.92)	89% (0.93)
Logistic Regression	98% (0.99)	91% (0.91)	91% (0.93)	91% (0.94)
Neural Network	98% (0.99)	95% (0.97)	93% (0.96)	88% (0.92)
Gradient Boosted Trees	98% (0.99)	94% (0.95)	94% (0.91)	97% (0.99)
Random Forest	99% (0.99)	95% (0.97)	92% (0.95)	97% (0.99)

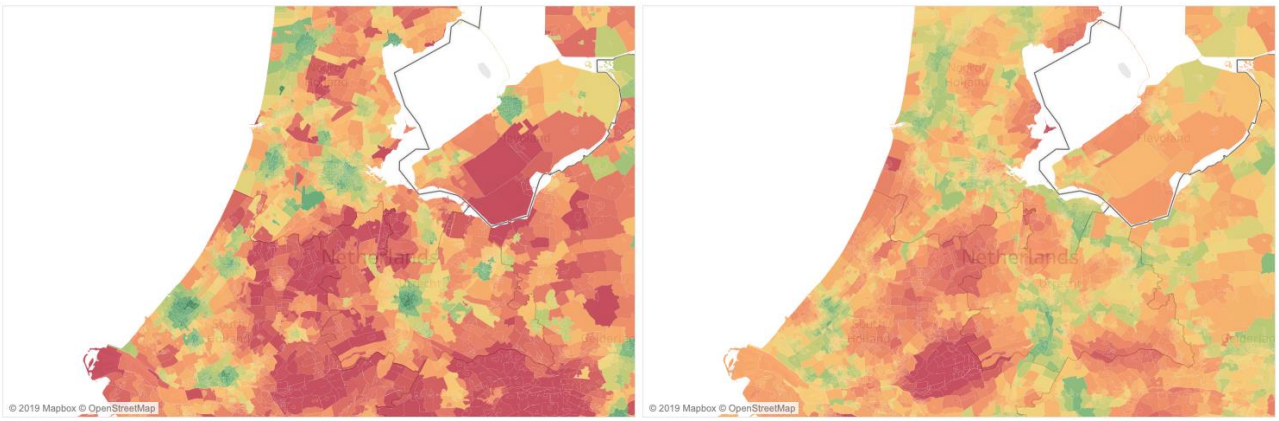

Fig. 6. Measuring reachability of neighborhoods: (a) public transport; (b) car. Green indicates a high reachability, red indicates a low reachability.

Besides the reachability, the travel type has a considerable influence on travel mode choice. Commute trips are favored by public transport but private trips by car. This might be influenced by the travel policy of the company or the ability to work during public transport. The CO2 emissions of public transport also have relatively high importance. However, this might be interpreted by the model as an indication of whether the transaction contains bus trips. In the Netherlands, the train, metro, and trams are relatively low on CO2 emissions. The emissions are high only when transactions would involve the bus.

The features concerning congestion and weather are of little influence. The congestion might be explained as roads are typically congested at the start and end of a working day, and all persons in our dataset are employed. The SHAP values even indicate that high congestion corresponds to a positive impact on model output, meaning a higher probability of taking the car.

Our experimental setup shows that data processing is critical for the evaluation of the models. If we would simply only execute the first experiment, we could present models with even higher accuracy. However, they would explain modal choices in a limited fashion. For example, public transport transactions can be predicted easily as their start and end location are at stations and travel times between stations are relatively fast by public transport. Additionally, the experimental setup highlights differences between the models. In the first experiment, all models perform similarly, however, in the final experiment differences are clearly visible.

6. Research Opportunities

The models developed in this research show promising results and give insights into the mobility behavior of the individuals. Still, they can be improved and used for further purposes.

Firstly, the models can be used to predict the impact of changes in infrastructure. The mobility behavior of the individuals is incorporated into the models. When the infrastructure changes, the features in the data change, and the mobility choices of the individuals might as well. Our model can be used to quantify to what extent investments in infrastructure lead to different mobility behavior.

Next, we can introduce more specialized models to gain more accurate predictions. Specifically, the availability of alternatives and repeating choices can be incorporated explicitly in a model. The advantage is that we can use more data to fit the model, as currently, we filter the data on these conditions.

Additionally, we can introduce trip chaining. This incorporates the fact that trips of individuals are linked throughout time and possibly influence each other. For example, if an individual first needs to take their children to school and after that directly go to work, the modal choice for the trip to work is influenced by the trip to the school. In the data presented in this research, we can follow the choices of an individual throughout the day, hereby combining the trips that need to be executed throughout the day. For example, if one of the destinations throughout this day has historically only been executed by car, we need to take this preference into account for the other trips during that day.

Furthermore, the estimation of start and end locations can be further improved. At the moment, we take a constant travel time of 10 minutes by bike and consider the buildings in the surrounding area weighted by population density. However, the willingness to bike might vary per station and region. For example, the density of stations in cities varies, which possibly has an impact on the willingness to bike. Also, it might be more relevant for business trips to weight the buildings by the number of employees.

Also, this research can be extended towards influencing the travel mode choice of individuals. We can use the predictions of the model to compare individuals amongst each other, and amongst the expected behavior. If the choices of an individual differ from a cluster or the expectation of the model, this might be an indication that the behavior can be changed. This potential change can be communicated easily on an individual level through a mobile application.

Finally, we can investigate whether we can influence the travel time of individuals. We know the expected choices of individuals and we know the expected choices of our whole population. We can combine these to spread the traffic flows on multiple travel modes. This in order to minimize congestion or the stress on a system. Especially during crisis situations, such as the COVID-19 virus, such an extension could be relevant.

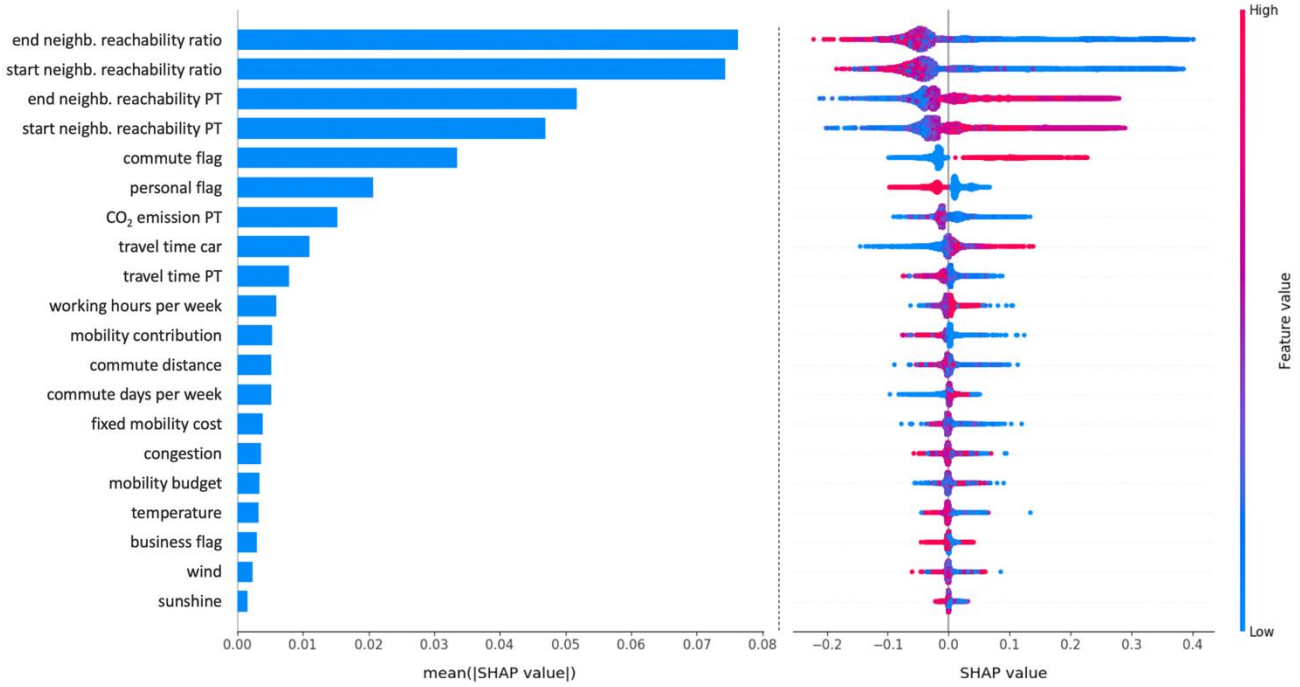

Fig. 7. Feature importance: (a) mean absolute importance; (b) all feature values.

References

[1] J. Slik and S.Bhulai. Transaction-driven mobility analysis for travel mode choices. Proceedings of the 11th International Conference on Ambient Systems, Networks and Technologies (ANT), pages 169–177, 2019. https://doi.org/10.1016/j.procs.2020.03.022

[2] S. Bouton, S.M. Knupfer, I. Mihov, and S. Swartz. Urban mobility at a tipping point. Technical report, 2017.

[3] R. Cervero, E. Guerra, and S. Al. Beyond mobility: planning cities for people and places. Island Press, 2017. https://doi.org/10.5822/978-1-61091-835-0

[4] George Kingsley Zipf. The p1 p2/d hypothesis: On the intercity movement of persons. American Sociological Review, 11 (6): 677–686, 1946. https://doi.org/10.2307/2087063

[5] T.A. Domencich and D. McFadden. Urban travel demand - a behavioral analysis. North-Holland Publishing, 1975.

[6] Samuel A. Stouffer. Intervening opportunities: A theory relating mobility and distance. American Sociological Review, 5 (6): 845–867, 1940. https://doi.org/10.2307/2084520

[7] Universal predictability of mobility patterns in cities. Journal of the Royal Society Interface, 11, 2014. https://doi.org/10.1098/rsif.2014.0834

[8] Filippo Simini, Marta C. González, Amos Maritan, and Albert-Lászloó Barabási. A universal model for mobility and migration patterns. Nature, 484 (7392): 96–100, 2012. https://doi.org/10.1038/nature10856

[9] Xi Liu, Li Gong, Yongxi Gong, and Yu Liu. Revealing travel patterns and city structure with taxi trip data. Journal of Transport Geography, 43:78 – 90, 2015. https://doi.org/10.1016/j.jtrangeo.2015.01.016

[10] Lei Gong, Takayuki Morikawa, Toshiyuki Yamamoto, and Hitomi Sato. Deriving personal trip data from GPS data: A literature review on the existing methodologies. Procedia - Social and Behavioral Sciences, 138:557 – 565, 2014. https://doi.org/10.1016/j.sbspro.2014.07.239

[11] Here technologies, 2020. Retrieved from https://developer.here.com/.

[12] Traveltime platform, 2020. Retrieved from https://www.traveltimeplatform.com/.

[13] Lijst emissiefactoren, totale lijst, 2020. Retrieved from https://www.co2emissiefactoren.nl/lijst-emissiefactoren

[14] M. Bierlaire. Pandas Biogeme: a short introduction. Technical Report TRANSP-OR 181219, 2018.

[15] Pedregosa et al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.

[16] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM. https://doi.org/10.1145/2939672.2939785

[17] J.A. Hanley and B.J. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143(1): 29–36, 1982. https://doi.org/10.1148/radiology.143.1.7063747

[18] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in

Neural Information Processing Systems 30, pages 4765–4774. Curran Associates, Inc., 2017.