HAL
open science

Floating Car Data mining: Identifying vehicle types on the basis of daily usage patterns

Danyang Sun, Fabien Leurent, Xiaoyan Xie

To cite this version:

Danyang Sun, Fabien Leurent, Xiaoyan Xie. Floating Car Data mining: Identifying vehicle types on the basis of daily usage patterns. Transportation Research Procedia, 2020, 22nd EURO Working Group on Transportation Meeting, EWGT 2019, 18th – 20th September 2019, Barcelona, Spain, 47, pp.147-154. 10.1016/j.trpro.2020.03.087 . hal-02955682

HAL Id: hal-02955682

https://hal.science/hal-02955682v1

Submitted on 31 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 47 (2020) 147–154

Transportation Research Procedia

www.elsevier.com/locate/procedia

22nd EURO Working Group on Transportation Meeting, EWGT 2019, 18-20 September 2019, Barcelona, Spain

Floating Car Data mining: Identifying vehicle types on the basis of daily usage patterns

Danyang Sun[a*], Fabien Leurent[a], Xiaoyan Xie[a]

[a]*Laboratoire Ville Mobilité Transport, UMR T 94 03, ENPC – IFSTTAR – UPEM.*
Ecole des Ponts ParisTech, 6-8 Avenue Blaise Pascal – 77455 Marne-la-Vallée cedex 2, France

Abstract

This paper presents a novel approach for exploring vehicle daily mobility patterns using Floating Car Data in the Paris region. The objective is to reveal vehicle types and analyze usage patterns on different roadway sub-classes. Firstly, mobility representative features are recovered to recognize vehicle activity contexts, which includes modeling single trip pattern in terms of time-window, traveling-distance, and speed; and building vehicle mobility profile of trip combinations. Based on that, a two-step clustering algorithm is developed to explore the constitution of trip patterns and cluster vehicle types. Characterization analysis is conducted to find out outstanding features of clustered groups, thus helping to categorize the vehicles behavioral types. As a result, 4 major vehicle types were identified over the Paris region with the 2 comparative leading groups as those mainly composed by morning-activity trips and long-distance trips. Then, statistical association assessment by Configural Frequency Analysis is employed to examine the usage intensity on different roadway classes of identified types. The association analysis reveals that the identified types have statistically significant differences in the usage of different roadway classes. Furthermore, this approach can be expected to provide more representative results with more generalized sampling by the development of connected vehicles.

© 2020 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the scientific committee of the 22nd Euro Working Group on Transportation Meeting

Keywords: Vehicle typology; Road usage assessment; Machine learning; Floating Car Data; Large scale network;

1. Introduction

Over the years, more and more territories have been facing the challenge of mobility issues. One of the key causes is the inconsistency between mobility patterns of people and the spatial structure of the region (Liu et al., 2015), which therefore prompts the need to understand people's spatial-temporal mobility patterns to reveal the

* Corresponding author.
E-mail address: danyang.sun@enpc.fr

2352-1465 © 2020 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the scientific committee of the 22nd Euro Working Group on Transportation Meeting.
10.1016/j.trpro.2020.03.087

underlying issues. In past studies, the availability of data has been a major concern for mobility analysis. Traditional data collection methods, such as household mobility surveys, loop detectors, vehicle diaries, and video cameras are inherently limited as concerns both spatial coverage and temporal resolution (Sun et al., 2014). With the growing diffusion of GPS devices, Floating Car Data (FCD) has emerged as a new data source to address mobility analysis in a systematic and cost-effective way. Since no special equipment is required to set up, this technology can be widely deployed and has the potential of providing data for large scale network and up-to-date mobility demand.

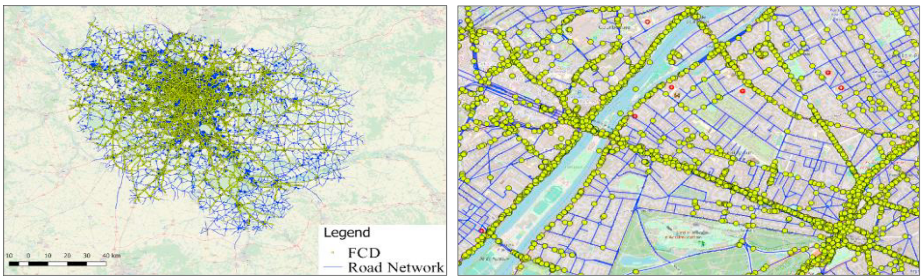
FCD-related studies in existing literature can be generally grouped according to two ranges of issues: "physical issues" for recovering the state of road network performance and "behavioral issues" for revealing the microeconomic mechanism behind trips. More specifically, the first part can be subdivided in three major issues: 1) estimation of traffic speed (Fabritiis et al., 2008; Fusco et al., 2016; Rempe et al., 2017); 2) estimation of travel times (Hunter et al., 2009; Jenelius and Koutsopoulos, 2013; Rahmani et al., 2017); 3) determination of traffic conditions (Brockfeld et al., 2007; Li and Graham, 2016). As for behavioral analysis, several studies have been targeted to route choice or origin-destination distribution problems (Ciscal-Terry et al., 2016; Sun et al., 2014; Zhu and Levinson, 2015). These studies addressed many traffic issues, however, so far, limited work has made use of FCD to investigate mobility issues in terms of vehicle usage typology. While, FCD actually tracks the traces for spatial-temporal patterns, which can be reviewed to better understand vehicle usages. In fact, the issue of vehicle usage type has been commonly investigated using conventional vehicle diaries, especially for freight vehicles, results of which could further contribute to business monitoring, activity configuration and multiple other sector analyses ("Survey of the use of road freight vehicles (TRM)," n.d.). However, such analysis is limited due to the time-intensive nature and in-complete path capturing of conventional vehicle diaries. Therefore, new kinds of digital data such as FCD are expected to eliminate the restrictions and improve such studies (Nguyen et al., 2017). While in the literature, only limited reference was found to classify the vehicle types using GPS records collected in a more specialized way than anonymized FCD (Simoncini et al., 2018, 2016; Sun and Ban, 2013). Sun and Ban (2013) developed a method using supporting vector machine (SVM) to distinguish delivery trucks from passenger cars, using field collected GPS data from traffic mobile sensors with a frequency of every 1s and 3s for passenger cars and trucks respectively. However, this study only classified the vehicles into two classes and only limited data on arterial streets were involved. Besides, the data used in this study were recorded with a rather high frequency, while a more common practice is to transfer less data with less frequency, e.g. every minute or even longer. In Simoncini et al. (2016), the authors claimed that they were the first effort in tackling the problem of vehicle classification using low-frequency GPS data, collected from the installed devices in commercial fleets. A binary SVM classifier was developed to distinguish light-duty vehicles from large size ones, based on a combination of features that were identified to be most predictive using a recursive feature elimination procedure. A more recent study by Simoncini et al. (2018) was conducted to classify vehicles from a lower frequency GPS data collected from connected vehicles by every minute. The authors employed recurrent neural networks to categorize the vehicles into the types of small-duty, medium-duty, and heavy-duty vehicles. An approach based on Long Short-Term Memory (LSTM) recurrent neural networks was proposed to learn effective hierarchical and stateful representations for temporal sequences. However, one research gap found in these two studies was that the employed data mining approaches, SVM and LSTM were all based on supervised learning, which requires the vehicle type labels. While, due to the increasing need of privacy and regulation of data protection, the trend of the data available to wide research use would be fully anonymous, which prompts the need of unsupervised data-driven techniques in exploring such kind of data.

Acknowledging the needs to address the above-mentioned problems, the overall objective of this paper is to develop a methodology to explore vehicle mobility patterns from Floating Car Data. To better understand the roadway traffic, one important issue is that of vehicle typology with respect to its mobility usage: each type should be associated with a specific behavior of trip-making, such as quantitative patterns including trip frequency, distance traveled, driving speed, and qualitative patterns reflected by the "semantics" of spatial-temporal trip makings. The research aim of this study is thus to build such a vehicle typology in terms of their diverse activity characteristics. Considering the data are recorded anonymously, extrapolations can be made using data mining approaches based on the traces. By identifying the vehicle typology, corresponding performance on the usage of different roadway classes can be assessed for potential future regulation. The structure of the paper is as follows. The data preparation is presented in section 2. Section 3 describes the methodology for vehicle type clustering and corresponding results. The roadway usage assessment is described in section 5, followed by a concluding discussion in Section 6.

2. Data preparation and initial analysis

2.1. Data

FCD explored in this study is collected by onboard GPS devices from Coyote, which is one of the major roadway navigation service providers in France. Such data contain the localization coordinates of vehicles (longitudes and latitudes) together with moving speed, heading and the timestamp periodically. The raw dataset is organized by a sequence of records, each of which represents an instantaneous state of a vehicle. The recording frequency is noted around 30s to 60s, which is subject to the device configuration of different models and the transmission speed under different circumstances. In the following analysis, one-day data on February 14th, 2018 over the Paris region were adopted for carrying the analysis due to the limitation of data availability. Besides, a shapefile containing the road network was extracted from OpenStreetMap, which contains properties information of each road segment. All recording logs vehicles were geocoded and overlaid with the road network layer as illustrated in Fig 1. Those recording logs were plotted as individual points on the GIS map. By using geo-processing functions in QGIS, these observations were bounded with the nearest road segment to join the attribute of corresponding roadway classes.



(a) (b)

Fig. 1 FCD geo-matched with the road network in QGIS: (a) the Paris region, and (b) zoomed view

2.2. Trip generation

To explore the mobility information conveyed by FCD, it is important to recover trajectories for each vehicle and segment them into meaningful trips. This is because the prevailing FCD for the general public users does not have the absolute indicator in the data to indicate whether a trip stops or not. Therefore, a trip generation approach was proposed based on the original FCD to recover the trip information of each vehicle. As a consequence, a new dataset of trips was generated along with a series of describing features computed from raw FCD. The layout of the data constituted by single trips is illustrated in Fig. 2, which includes trip frequency (*Trip_No:* the sequence order of the trip), origin-destination location (*Orig. and Dest.* latitudes and longitudes in WGS 84), time window (*OrigTs* and *DestTs,* in seconds with the origin 0s at midnight), trip distance (*Dist* in km) and driving speed (*Speed* in km/h).

	ID	Trip_No.	OrigLat	OrigLong	OrigTs	DestLat	DestLong	DestTs	Speed	Dist
0	00822DEFA18D49245FA420390DDEACA8	1	48.8800	2.6891	22320	48.8602	2.2904	27204	30.200000	37.320184
1	00822DEFA18D49245FA420390DDEACA8	2	48.8624	2.3029	54581	48.8329	2.5591	58935	18.548611	22.989519
2	00822DEFA18D49245FA420390DDEACA8	3	48.8329	2.5589	64845	48.8791	2.6882	66900	28.780822	15.997341
3	008D380B47E10A599FB11A2879D2C8C2	1	48.8967	2.1576	48193	48.8168	2.1280	49541	29.782609	10.030994
4	00A35D54400EDC60DB87E89AD42355BB	1	49.0089	1.4136	57400	48.9586	1.8215	58900	74.714286	30.973641
5	00A35D54400EDC60DB87E89AD42355BB	2	48.9651	1.7652	61368	49.0202	1.4893	65030	59.545455	33.118691

Fig. 2. Sample of reconstituted single trips

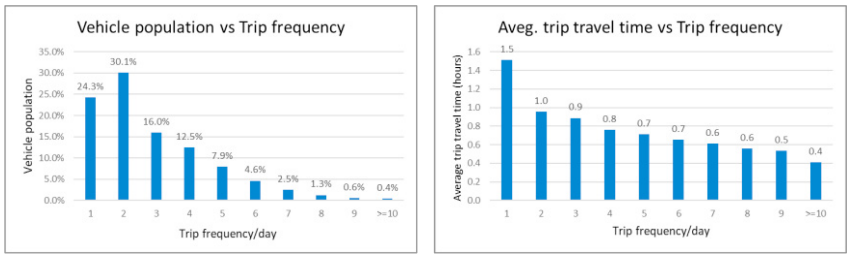
In the processing, vehicle ID was used to match the belonging records for each vehicle. As for segmenting the logs in to trips, it still remains as an unclear issue in existing literature, as the time interval threshold between two trips can vary a lot in different cases. A recent study by Dabiri and Heaslip (2018) indicated that 20 minutes could be used as the interval threshold to segment meaningful trips at urban settings. To further check, an experimental analysis was performed, which inspected the time intervals between two succeeding logs of the same vehicles by plotting the distribution. It can be assumed that the distribution frequency at interval values corresponding to device

recording should be apparently higher than those corresponding to trip stops. In our case, an elbow point in the interval distribution was found around 1200s to 1300s, which indicates there could be a distinction around such values. Considering the logical meaning of 20 minutes interval in Dabiri and Heaslip (2018), 20 minutes (1200s) was finally selected to segment trips in this study. Travel distance was calculated as the overall traveling distance of the trip, which was the sum of distances for the shortest path between every two succeeding points. Driving speed was calculated as the average speed of all corresponding recordings of the trip. Besides, the departure time of a trip and its location, as well as the arrival time and its location, were also extracted for each single trip.

By reconstituting the data, a total of 68,613 unique vehicles with trip information was recovered from FCD over the Paris region, with a total number of 19,654 single trips identified. The overall statistics of trips can be summarized as in Table 1. In addition, Fig. 3(a) shows the vehicle population in terms of different trip frequency. Vehicle making 2 trips in the day was found most prevailing with a percentage of 30.1% of the total. Fig. 3(b) shows the average trip travel time of vehicles in terms of trip frequency. It was found that the more trip the vehicle made in the day, the less average travel time on a trip it spent.

Table 1. Data description of recovered trips

	Average	Median	1st Quartile	3nd Quartile
# Trips per vehicle	2.9	2.0	2.0	4.0
Speed (km/h)	32	28	14	45
Distance (km)	27	14	4	37
Trip Duration (h)	0.8	0.5	0.2	1.0



(a) (b)

Fig. 3. (a) Vehicle population vs different trip frequencies; (b) Average travel time vs trip frequency

3. Type identification

3.1. Method: a two-step clustering

To identify the vehicle types, a two-step clustering method is proposed and described as follows.

1) Step 1: Considering the trip is a more describable unit that can be recovered with explicit features describing its characteristic, the first step was to conduct a clustering analysis in an attempt to identify and characterize the type of trips. Each trip can then be labeled with a specific type number.
2) Step 2: The second step was to cluster the vehicle types according to the trip profile of the vehicle, which was generated by matching the trips with identified types back to corresponding vehicles.

The k-means clustering algorithm was employed to partition both of trips and vehicles into different k homogeneous groups respectively according to their relative variation of corresponding features. To begin with, feature selection and standardization were required. To describe trips, the independent features in the recovered trip dataset: departure time, trip distance, and trip average speed, were selected for the trip clustering analysis. As for the vehicle clustering in the second step, the trip profile was firstly generated for each vehicle by counting the number falling in each trip type. These trip numbers of different types were then used as the features to cluster vehicles. Since the range of values may vary widely due to different dimensional quantity, all features were rescaled to weight each dimension equally. Another key issue of clustering analysis is to determine the number of clusters. It is commonly acknowledged that there is no unambiguous answer to this question. The optimal number k of clusters is relatively subjective and depends on the methods and the data used for partitioning. In this study, two widely used

methods, the elbow method (Thorndike, 1953) and average silhouette method (Rousseeuw, 1987) were employed to determine the optimum cluster number for the k-means clustering. Combine the results of good candidates from the two analyses, a final judge could be made to determine the number of k.

To further understand the characteristic of the clusters and check whether the partitioning result was logical, a further characterization analysis was conducted to explore significant variations between groups. By finding out prominent characteristic differences among each group, such a process could help define the types (Ren et al., 2018). Different approaches can be employed to make the comparison between groups. Considering the feature dimensionality, partitioned trip clusters were visualized for comparison into a 3D scatter plot to in terms of the 3 trip attributes. While the vehicle clusters were compared by the statistics of the trip profiles through box plots. Detailed results are described in the section below. Due to the processing speed of limited computation power, only 1000 random selected vehicles, corresponding to 2904 trips, were finally analyzed in type identification. However, it should be noted that this won't affect the methodological findings in a meaningful way.

3.2. Results of trip type identification

The recover trips were clustered into 3 clusters by the proposed approach. To understand the characteristic of the clusters, the identified clusters are visualized in Fig. 4 in different colors using a 3D scatter plot. By further looking at the statistics of each type, the 3 trip types can be characterized as follows.

- Trip type 1 (in red): Long distance trips with relatively medium to high speed. No specific characteristic of departure time.
- Trip type 2 (in green): Evening short trips with relatively low speed.
- Trip type 3 (in blue): Morning short trips with relatively low speed.

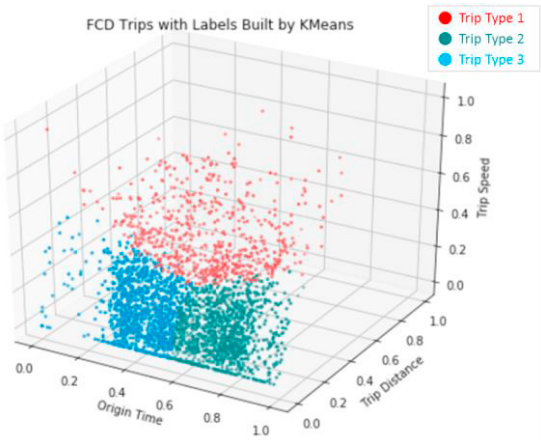


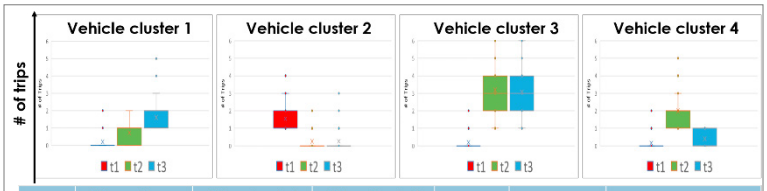
Fig. 4. Trip type clustering result by k-means

Although there were no significant gaps between each group, such clustering result is consistent with the practical scenarios of mobility activities, where there is no definite distinction among various trips. Moreover, such clustering results also provide a way to segment the trips from a quantitative perspective based on trip features. These identified trip types were further used for the clustering analysis of vehicle types in the next step.

3.3. Results of vehicle type identification

Based on the generated trip profiles, the optimal number k of clusters for vehicle clustering were determined as 4 by employing the elbow method and average silhouette method. Therefore, the vehicles were clustered into four groups (also referred to as vehicle types). To further understand the features of each group, we calculated the statistics of trip frequency for each trip type and visualized that of the four vehicle groups using box plots, which is shown in Fig. 5. By comparing the statistics, the four vehicle types can be characterized as the table shown in Fig.5. According to these trip counts, four types of vehicle were thus identified, with the defined descriptions as follows.

- Vehicle type 1: Morning activity based vehicles, whose trip profile consists of a medium amount of morning short trips, low amount of evening short trips and almost none of long distance trips.
- Vehicle type 2: Long distance traveling vehicles, whose trip profile mainly consists of long distance trips.
- Vehicle type 3: Frequent activity based vehicles, which have a high frequency of both morning and evening short trips.
- Vehicle type 4: Evening activity based vehicles, whose trip profile consists of a medium amount of evening short trips, low amount of morning short trips and almost none of long distance trips.



Vehicle Type	T1 (long distance trip) (frequency)	T2 (evening short trip) (frequency)	T3 (morning short trip) (frequency)	Vehicle population	Distance Travelled	Inference
1	Rare	Low	Medium	32.8%	25%	**Morning Activity Based**
2	Medium	Rare	Rare	31.1%	48%	**Long Distance Traveling**
3	Rare	High	High	13.3%	15%	**Frequent Activity Based**
4	Rare	Medium	Low	22.7%	12%	**Evening Activity Based**

Fig. 5. Vehicle type clustering and characterization

The vehicle population (of all analyzed vehicles) and network share of the traveled distance of each identified types were also calculated in the table. It was found that morning activity based vehicles and long distance traveling vehicles were the two leading groups, accounting for 32.8% and 31.1% of the population and 25% and 48% of the total distance respectively. Evening activity based vehicles rank the 3rd, taking up 22.7% of the total population and 12% of the distance. Frequent activity based vehicles were found to be the least with the population percentage of 13.3% and a slightly higher share of the distance of 15%.

4. Statistical association assessment on roadway usage

4.1. Method: Configural Frequency Analysis

This section proposes an approach to assess the association between the identified vehicle types (clusters) and their usage intensity on different classes of roadways. A Configural Frequency Analysis (CFA) was performed to examine the association. By bounding FCD with the GIS road networks, a cross-tabulation table was assembled for the statistical analysis, which was constituted by two categories (vehicle types and roadways classes). Each cell in the cross-tabulation table represents a certain pattern combined by a particular vehicle type and a class of the roadway, which was referred to as a configuration. The value of a configuration was counted by the observations corresponding to it. As the recording frequency is relatively constant, around 30s, it is reasonable to assume the more observations of a configuration, the higher usage the vehicle has on that class of roadways. CFA is a widely adopted non-parametric statistical method for multivariate data analysis, whose goal is to detect types and antitypes in cross classification of categorical variables. Types are defined as configurations that occur significantly more than expected, while antitypes are those that occur significantly less, both of which are identified by comparing observed and expected frequencies. To perform a CFA, two main routine steps were carried out: 1) Calculating the expected configuration frequencies under the assumption that no category relationship exists (Eye et al., 1996). The widely used First Order CFA model was employed as the base model in this study; 2) Comparing the observed and expected frequencies using a statistical testing. The Z test was employed to detect the types and antitypes.

4.2. Results of roadway usage assessment

Based on the roadway matched vehicle observations, the final assembled cross-tabulation table of vehicle usage observations along different roadway classes is summarized using percentage, as shown in Table 2.

Table 2. FCD observations over different vehicle clusters and roadway classes

Vehicle Clusters	Roadway Classes Usage						Observations
	Motorway	Primary	Secondary	Tertiary	Trunk	Residential	Total records
Morning Activity Based	14%	21%	18%	11%	8%	28%	64609
Long Distance Traveling	34%	19%	14%	10%	13%	9%	54526
Frequent Activity Based	14%	21%	23%	15%	6%	22%	44327
Evening Activity Based	13%	22%	21%	14%	8%	22%	31508

Table 3 summarizes the results of the analysis, where all configurations were tested at a significance level of 99.9%. Recall that in CFA types are defined as patterns of variable categories that occur significantly more than expected by chance, while antitypes are "significantly less". As the antitype configuration may shows as a reverse side effect of the increasing number in type configuration and the primary concern is to determine particularly "more" roadway usages, only type results are discussed.

Table 3. CFA results of vehicle roadways usage on different roadway classes

Vehicle Clusters	Roadway Classes Usage					
	Motorway	Primary	Secondary	Tertiary	Trunk	Residential
Morning Activity Based	A	-	-	A	A	T
Long Distance Traveling	T	A	A	A	T	A
Frequent Activity Based	A	-	T	T	A	T
Evening Activity Based	A	T	T	T	A	T

Where “T” represents “type”; “A” represents “antitype”; “–” indicates as expected.

It was found that Long Distance Traveling vehicles had a significantly higher usage on Motorways and Trunk roadways, while the other 3 vehicle types having significantly higher usage on Residential roadways. This well corresponds with the fact short trips are normally not distributed on higher class of roadways. Besides, Frequent Activity Based vehicles and Evening Activity Based vehicles were found significantly more presented on secondary and tertiary roadways. Only Evening Activity Based vehicles were found significantly related to the usage of primary roads. As for Morning Activity Based vehicles, beside the "type" pattern on residential roads, no significant more usage on other roadways was detected. These results indicate there is significant usage difference between vehicles by its temporal tendency in different activity patterns.

5. Conclusions and discussions

This research proposed an explorative approach on Floating Car Data to analyze the mobility patterns of vehicles. The objective was to identify the vehicle types and assess their usage behavior on different sub-categories of roadways based on low frequency Floating Car Data over the Paris region. A set of mobility representative features was generated to recognize main vehicle activity context. On top of that, a two-step clustering-based algorithm was developed to explore the constitution of different trip patterns and cluster vehicle types respectively. Statistical characterization analysis was conducted to find out the outstanding features based on clustered groups, thus helping to categorize the vehicles behavioral types. A configural frequency analysis was then employed to conduct the association analysis to examine the main phenomena of interest with respect to identified typologies. Consequentially, a total of 68,613 unique vehicles with trip information was recovered with an average daily trip frequency of 2.9. Four major vehicle types were identified based on their mobility features, which were defined as 1) Morning Activity Based vehicles; 2) Evening Activity Based vehicles; 3) Long Distance Traveling vehicles and 4) Frequent Activity Based vehicles. The population of the four types was found accounting for 32.8%, 22.7%, 31.1% and 13.3% of the total respectively. The association analysis revealed that the identified types had statistically significant differences in the usage of different roadway classes. This explorative approach could be used for future roadway planning in terms of analyzing its particularly targeted users. Although sampling bias is a common problem for the current FCD, such method can be expected to provide more representative results with the development of connected vehicles in the future.

By linking vehicle type identification with post behavior performance assessment through statistical analysis, this research provides a methodological proposal for detecting significant relations of vehicles to specific phenomena of interest, which can be transferrable on other roadway issues, such as accidents and emission. Moreover, this study offers a big data-driven instance to recover and better understand mobility patterns, the knowledge of which can be useful in improving network regulation and control. However, due to the availability of data of this study, future work can be done to extend the analysis by either considering patterns over days or involving a larger regional scale for inter-city patterns. Besides, by integrating with other data sources, more features can be recovered to describe the motility patterns, such as crossing with land use and special patterns of geographical areas. Other alternative machine learning methods can also be explored and compared to improve the identification of vehicle typology.

Acknowledgments

The research was supported by Chair Eco-design Vinci at Ecole des Ponts ParisTech. The authors would like to thank Geolytics project, with partners IT4PME and Coyote, for the data and assistances, and Chair ENPC-IdFM in facilitating this research.

References

Brockfeld, E., Lorkowski, S., Mieth, P., Wagner, P., 2007. Benefits and limits of recent floating car data technology–an evaluation study, in: 11th WCTR Conference, Berkeley, USA.

Ciscal-Terry, W., Dell'Amico, M., Hadjidimitriou, N.S., Iori, M., 2016. An analysis of drivers route choice behaviour using GPS data and optimal alternatives. J. Transp. Geogr. 51, 119–129.

Dabiri, S., Heaslip, K., 2018. Inferring transportation modes from GPS trajectories using a convolutional neural network. Transp. Res. Part C Emerg. Technol. 86, 360–371.

Eye, A., Spiel, C., Wood, P.K., 1996. Configural frequency analysis in applied psychological research. Appl. Psychol. 45, 301–327.

Fabritiis, C. de, Ragona, R., Valenti, G., 2008. Traffic Estimation And Prediction Based On Real Time Floating Car Data, in: 2008 11th International IEEE Conference on Intelligent Transportation Systems. Presented at the 2008 11th International IEEE Conference on Intelligent Transportation Systems, pp. 197–203. https://doi.org/10.1109/ITSC.2008.4732534

Fusco, G., Colombaroni, C., Isaenko, N., 2016. Short-term speed predictions exploiting big data on large urban road networks. Transp. Res. Part C Emerg. Technol. 73, 183–201.

Hunter, T., Herring, R., Abbeel, P., Bayen, A., 2009. Path and travel time inference from GPS probe vehicle data. NIPS Anal. Netw. Learn. Graphs 12, 1–8.

Jenelius, E., Koutsopoulos, H.N., 2013. Travel time estimation for urban road networks using low frequency probe vehicle data. Transp. Res. Part B Methodol. 53, 64–81.

Li, H., Graham, D.J., 2016. Quantifying the causal effects of 20 mph zones on road casualties in London via doubly robust estimation. Accid. Anal. Prev. 93, 65–74.

Liu, X., Gong, L., Gong, Y., Liu, Y., 2015. Revealing travel patterns and city structure with taxi trip data. J. Transp. Geogr. 43, 78–90.

Nguyen, T.T., Armoogum, J., Madre, J.L., Pham, T.H.T., 2017. GPS and travel diary: Two recordings of the same mobility, in: ISCTSC, 11th International Conference on Transport Survey Methods. p. 13p.

Rahmani, M., Koutsopoulos, H.N., Jenelius, E., 2017. Travel time estimation from sparse floating car data with consistent path inference: A fixed point approach. Transp. Res. Part C Emerg. Technol. 85, 628–643. https://doi.org/10.1016/j.trc.2017.10.012

Rempe, F., Franeck, P., Fastenrath, U., Bogenberger, K., 2017. A phase-based smoothing method for accurate traffic speed estimation with floating car data. Transp. Res. Part C Emerg. Technol. 85, 644–663.

Ren, K., Kim, A.M., Kuhn, K., 2018. Exploration of the Evolution of Airport Ground Delay Programs. Transp. Res. Rec. 2672, 71–81.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65.

Simoncini, M., Sambo, F., Taccari, L., Bravi, L., Salti, S., Lori, A., 2016. Vehicle Classification from Low Frequency GPS Data, in: 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW). Presented at the 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), pp. 1159–1166. https://doi.org/10.1109/ICDMW.2016.0167

Simoncini, M., Taccari, L., Sambo, F., Bravi, L., Salti, S., Lori, A., 2018. Vehicle classification from low-frequency GPS data with recurrent neural networks. Transp. Res. Part C Emerg. Technol. 91, 176–191. https://doi.org/10.1016/j.trc.2018.03.024

Sun, D., Zhang, C., Zhang, L., Chen, F., Peng, Z.-R., 2014. Urban travel behavior analyses and route prediction based on floating car data. Transp. Lett. 6, 118–125.

Sun, Z., Ban, X. (Jeff), 2013. Vehicle classification using GPS data. Transp. Res. Part C Emerg. Technol. 37, 102–117. https://doi.org/10.1016/j.trc.2013.09.015

Survey of the use of road freight vehicles (TRM) [WWW Document], n.d. URL https://www.statistiques.developpement-durable.gouv.fr/enquete-sur-lutilisation-des-vehicules-de-transport-routier-de-marchandises-trm (accessed 4.16.19).

Thorndike, R.L., 1953. Who belongs in the family? Psychometrika 18, 267–276.

Zhu, S., Levinson, D., 2015. Do people use the shortest path? An empirical test of Wardrop's first principle. PloS One 10, e0134322.