TRIP-CHAIN PATTERN ANALYSIS OF COMMERCIAL VEHICLES USING PROBE DATA

Ryosuke MAEDA [a], Hiroe ANDO [b], Takuya MARUYAMA [c]

[a] *Graduate School of Science and Technology, Kumamoto University, Kumamoto, 860-8555, Japan*
[a] *E-mail: 226d8367@st.kumamoto-u.ac.jp*
[b] *Center for Water Cycle, Marine Environment and Disaster Management, Kumamoto University, Kumamoto, 860-8555, Japan*
[b] *E-mail: hiroeando@kumamoto-u.ac.jp*
[c] *Faculty of Advanced Science and Technology, Kumamoto University, Kumamoto, 860-8555, Japan*
[c] *E-mail: takumaru@kumamoto-u.ac.jp*

Abstract: This study explored the behavioral patterns of commercial vehicles by using probe data from April 2019 to March 2020 in Kyusyu, Japan. The Jaccard and the trip pattern sequence index were proposed to evaluate the similarity among the behavioral patterns. A non-hierarchical cluster analysis was used to classify vehicles with behavioral patterns. The results revealed differences in the behavioral patterns by day of the week and time of the day, trends in trip duration, and distance for each trip chain. Commercial vehicles departing between 3:00 and 6:00 revealed less similar behavioral patterns, whereas vehicles departing between 17:00 and 21:00 revealed similar patterns. Vehicles with long travel distances have similar patterns and stay longer at their destination. Vehicles with less similar behavioral patterns tend to increase their travel from January to March. These findings have potential applicability in future logistics studies.

Keywords: probe data, freight vehicle, activity pattern

1. INTRODUCTION

In recent years, the characteristics and behavior of freight transportation have changed dramatically owing to the globalized logistics networks and the increase in online shopping. It is important to plan and operate the transportation systems corresponding to these rapid changes in freight demand. The origin-destination (OD) flows and trip patterns of commercial vehicles are important data for planning and operation.

The availability of data, such as daily driving reports written by drivers at the end of a trip and data from analog loggers installed in vehicles, used to understand the behavior of commercial vehicles was limited. However, in recent years, probe vehicle data have become widespread, thus making it easy to obtain trajectory data for many commercial vehicles. Additionally, the vehicle IDs included in this data enabled us to continuously analyze road usage at an individual vehicle level. This study aims to analyze the behavioral patterns of commercial vehicles traveling within Kyushu, Japan, using probe vehicle data.

Trip-chain-based analyses have been used to investigate the movements of commercial vehicles. Seyama et al. (2019) analyzed the behavior of rail container pickups and delivery trucks using digital operation record data. They identified three main types of movements: (1) piston-type with unloading, (2) piston-type with loading, and (3) triangular. Furthermore, Siripirote et al. (2020) proposed an approach to estimate the truck activity time, activity location, and trip chain from the truck GPS data. They identified the trends in the distribution of trip distances and rest periods, trip-chain patterns by the type of major cargo transported, and differences in the loading and unloading operation times. Khan et al. (2017) employed data from Austin, Texas, to develop a behavioral model of tour and stop choice. The findings indicated that various factors influenced tour choice, including freight characteristics, trip characteristics, base location, and intermediate stop location. Similarly, Lin et al. (2013) utilized data from a logistics survey conducted across five counties in Texas. They employed a binary logit model to identify the factors that impact the choice of tour type for freight vehicles in metropolitan areas. Their study demonstrated that freight type, purpose, travel time, length of stay, and destination characteristics significantly affected the tour type choice of freight vehicles. Additionally, Morgul et al. (2013) demonstrated the usefulness of taxi GPS data in estimating travel times for commercial vehicle trips.

In recent studies, GPS data have been used to understand the changes in commercial vehicle movements. Kawasaki et al. (2019) proposed a method using machine learning to detect detour routes for commercial vehicles during disasters efficiently. The analysis results revealed that the risk of flooding contributed to inefficient detour behavior during heavy rainfall in western Japan in 2018. Furukawa et al. (2023) proposed a detection method for vehicles whose behavior differs from that in daily life by calculating the similarity of converted strings of commercial vehicle data. The difference in behavior refers to speed changes and detouring behavior. Alho et al. (2019) proposed and evaluated a tour-type identification method using GPS data surveyed in Singapore in 2017-2019. The homogeneity of daily behavior patterns of freight vehicles was investigated using entropy, and the differences in operation were revealed according to the vehicle type and industry. Using GPS data, Akter and Hernandez (2023) explored the truck activity patterns. Wu et al. (2016) employed cluster analysis of GPS data to categorize commercial vehicle driving behavior into acceleration, deceleration, and speeding-prone profiles, which can help identify driving risks and enhance safety. Maeda and Maruyama (2022) analyzed truck OD flow fluctuations using probe vehicle data. However, they did not reveal the differences in the daily trip patterns of commercial vehicles at a vehicular level.

Several studies have explored the similarities in behavioral patterns. Hosoe et al. (2021) proposed a method to identify similar behavioral patterns among public transportation users using smart card data. They used the Jaccard index to calculate the similarity of behaviors on the day of the week, time of the day, boarding station, and alighting station among the users. They proposed a graph-polishing method to extract 141 characteristic behavioral patterns.

The contributions of this study are twofold. First, this study entailed the application of novel methods—evaluating the similarity of behavioral patterns by using the Jaccard and novel indices to understand the behavior of commercial vehicles. Additionally, a non-hierarchical cluster analysis using variables representing the features of the trip chain is performed. Second, this study entailed an analysis that considered a wide spatial-temporal range. Although many studies have explored the behavioral patterns of users or vehicles using transportation-related big data, no prior study has analyzed such large-scale data over a wide area over a long period.

The remainder of this paper is organized as follows. Section 2 outlines the data and methods used in this study. Section 3 describes the basic analysis. Section 4 presents the cluster analysis results and discusses the behavioral trends in clusters. Finally, Section 5 concludes the study.

2. DATA AND METHOD

2.1 Data

The data used in this study are probe vehicle data provided by Transtron. This probe data is based on data collected from a digital tachograph. Real-time operation data, such as speed, travel time, distance traveled, and location, were accurately recorded while driving. In Japan, with an aim to reduce the number of traffic accidents, it has been made mandatory to install such tachographs in commercial freight vehicles with a maximum loading capacity of four tons or more and weight of seven tons or more, as well as vehicles that are newly purchased.

The OD data of commercial vehicles arriving in and departing from Kyushu, Japan, were used. A mesh of 1 km was the smallest size of the traffic analysis zone. The data consists of departure/arrival time, trip distance, and loading information (occupied, empty, or undefined), in addition to the origin and destination. The period covered was from April 1, 2019, to March 31, 2020. Six days of data were excluded owing to missing data, from July 5 to 8, 2019, and from February 18 to 19, 2020. This data processing excluded 253,502 trips. Table 1 summarizes the data, and Figure 1 shows the target area and distribution of the trip generation.

Table 1. Summary of data

Item	Data before exclusion	Data after exclusion*
Number of days	366 (days)	360 (days)
Total number of trips	21,208,613 (trips)	8,011,177 (trips)
Total number of vehicles	26,236 (vehicles)	19,556 (vehicles)
Total number of meshes	22,469 (meshed)	17,377 (meshes)
Total number of OD pairs	1,595,853	725,660

Note: * Irregular data, such as extremely long travel time and irregular trip-chain sequences, were excluded. See 2.2.

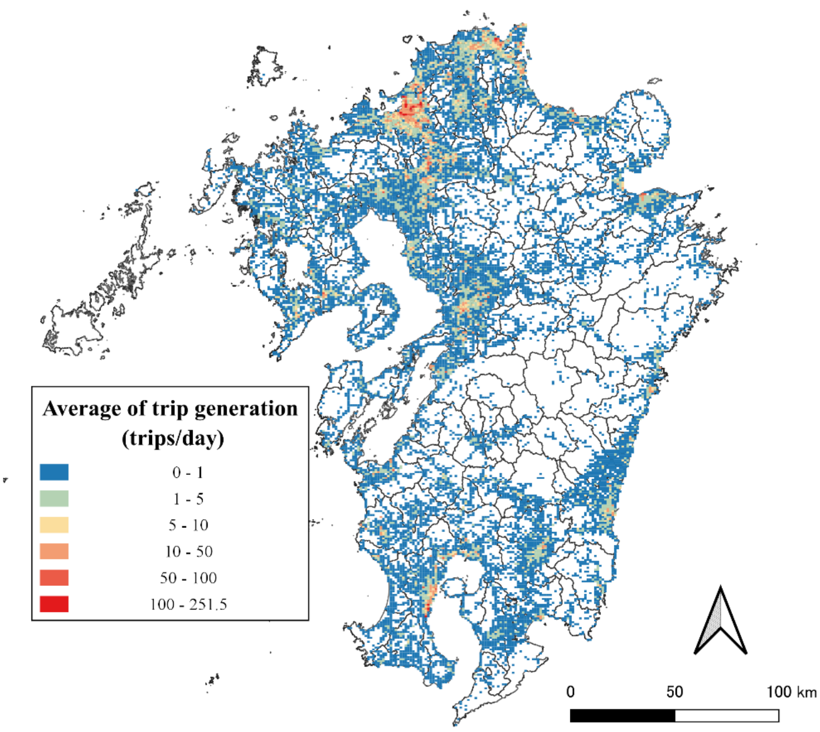

Figure 1. Target area and daily average of trip generation (trip/day)

2.2 Detecting base mesh and trip chains

First, we identified the base mesh for each vehicle to identify the trip chains. This study defines the mesh with the most frequent departure and arrival locations as the base mesh for the vehicle. Subsequently, a trip chain is defined as the sequence of trips departing from the base mesh and arriving at the base mesh.

Irregular data were excluded. The arrival mesh should be the departure mesh for the next trip. Data that violated this sequence were excluded. This data processing excluded 12,931,432 trips. Additionally, trip chains with a travel time of > 24 h were excluded. This data processing excluded 253,502 trips. Table 1 summarizes the data before and after the data exclusion.

2.3 Assessing the similarity of behavior patterns

Two indices—the Jaccard and trip pattern sequence (TPS) indices—are proposed to explore the similarity of the behavior patterns. The notion behind the indices is that two similar trip chains should have common trip ends (or mesh in our data).

The Jaccard index is expressed as:

$$Jaccard(a, b) = \frac{a \cap b}{a \cup b} \tag{1}$$

where $Jaccard(a, b)$ is the Jaccard index of trip-chain a and b, $a \cap b$ is the number of meshes included in both trip-chain a and b, $a \cup b$ is the number of meshes included in either trip chain a or b. The index takes a value of 0 to 1, and a higher value indicates that the two sets are more similar. The indices were calculated for every combination of trip chains in a vehicle, and the averages were used.

To evaluate the order of visiting places in a trip chain, the TPS index was proposed to overcome the limitation of the Jaccard index. The TPS index is defined as the percentage of frequency of trip chains that appear most frequently over all trip chains for the target vehicle in the target period. In other words, by enumerating all the trip chains in a vehicle, the percentage of the most often observed trip chains was set as the index.

The two indices take a value between 0 to 1, and a higher value indicates that the two trip chains are more similar. Figure 2 illustrates the examples of these two indices. While the Jaccard index focuses only on the location similarity of the mesh distribution in a trip-chain, the TPS index additionally considers the mesh sequence in a trip-chain.

The Jaccard index enabled us to evaluate the similarity between the trip chains with different numbers of trips. Hosoe et al. (2021) used the index to examine behavioral patterns among public transportation users. However, this study focused on the similarity of the daily behavior patterns of freight vehicles over a year. This index was used to determine whether a vehicle travels through similar destinations.

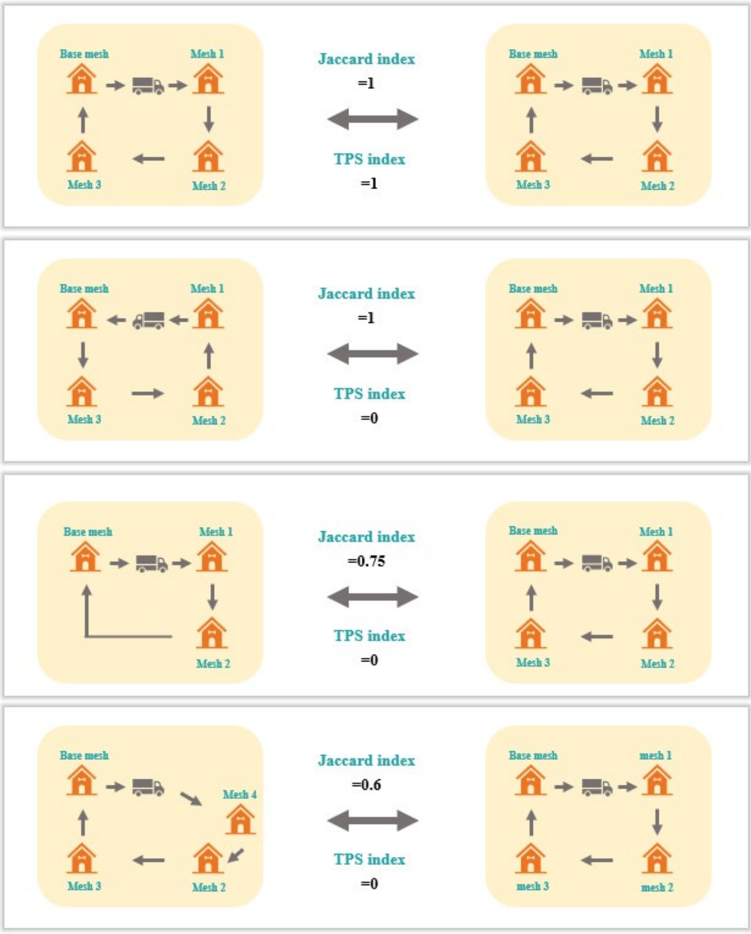

Figure 2. Illustration of the Jaccard index and trip pattern sequence (TPS) index

3. RESULTS AND DISCUSSION

3.1 Distribution of base mesh and trip-chain

Fukuoka Prefecture contained most base meshes (Table 2); thus, Figure 2 shows the distribution of base meshes in Fukuoka Prefecture. The mesh with the most base meshes is located in Higashi Ward, Fukuoka City, where many logistics facilities, including the Fukuoka Cargo Terminal, are located. For many vehicles, other meshes with many logistics facilities were found to be the base meshes. Table 3 presents the number of logistics facilities in the base mesh. The bonded area has the most logistics bases, followed by the truck terminals. Therefore, it can be inferred that many vehicles are transporting foreign cargo in Kyushu.

The distribution of trips per trip chain is shown in Figure 4. The trip chain with two trips (piston-type) was the most frequent, and the observation of trip chains with more than two trips decreased with the increase in the number of trips.

Table 2. Number of base meshes by prefecture

Prefecture	Number of meshes
Fukuoka	821
Saga	183
Nagasaki	130
Kumamoto	280
Oita	155
Miyazaki	203
Kagoshima	280

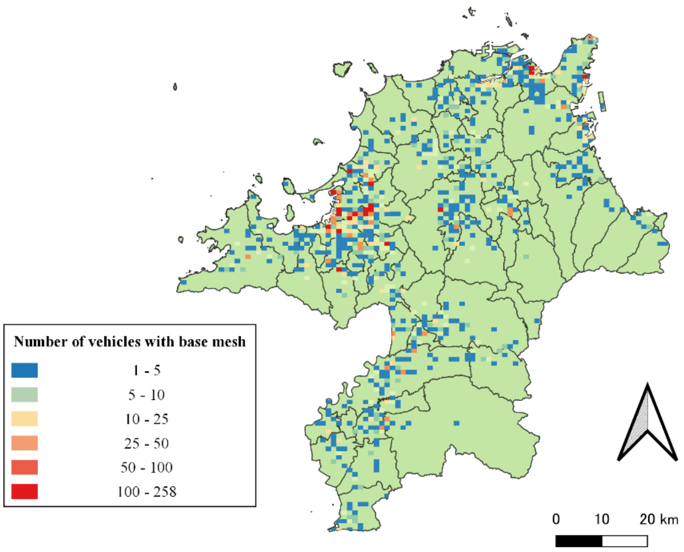

Figure 3. Distribution of base mesh in Fukuoka Prefecture

Table 3. Type of logistics facility of base mesh

Type	Count
Countainer terminal	63
Air cargo terminal	3
Railroad freight station	12
Bonded area	388
Truck terminal	90
Wholesale market	26

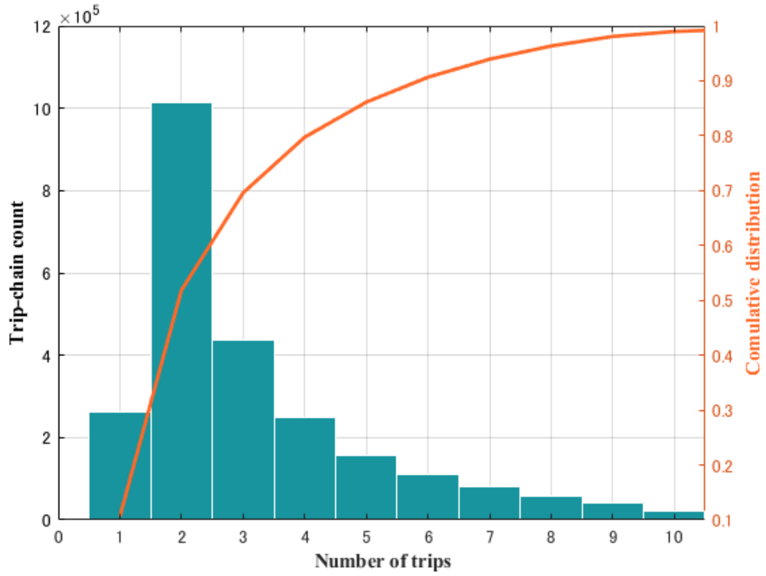

Figure 4. Distribution of the number of trips

3.2. Distribution of travel time and distance

Figure 5 shows the distribution of distances per trip chain. The average distance was 87.89 km, and the observation tended to decrease with the increase in the trip distance. Some trip chain distances exceeded 300 km, suggesting that such vehicles may use expressways.

Figure 6 shows the distribution of the travel times per trip chain. The average travel time was 145.18 min. Most trip chains with short travel times were one-trip or two-trip chains, while most trip chains with three or more trips had longer travel times. Additionally, the length trends of trip distances and trip times varied among the trip chains.

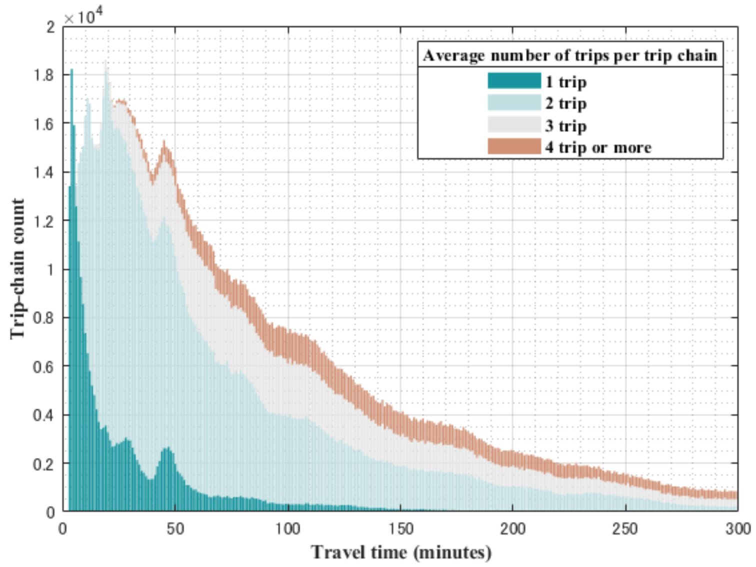

Figure 5. Trip-chain travel time distribution

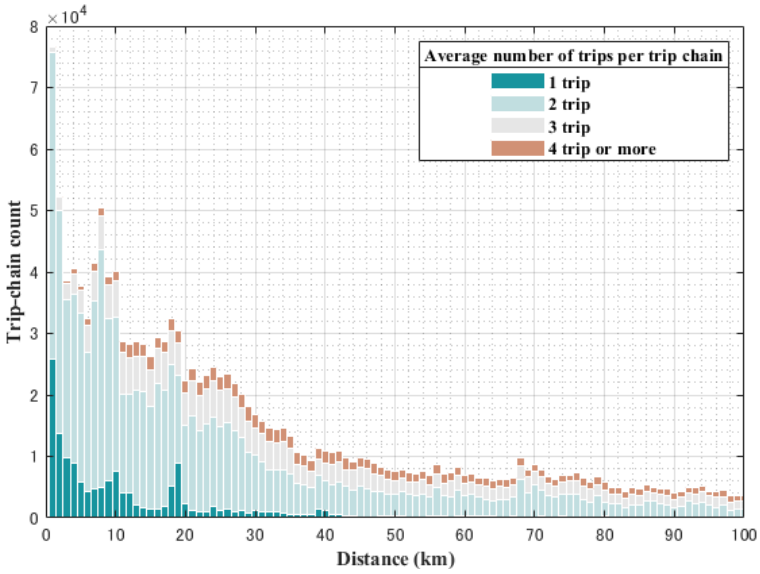

Figure 6. Trip chain distance distribution

3.3 Distribution of the similarity index

Figures 7 and 8 show the distribution of the Jaccard and TPS indices. The peak values of the Jaccard and TPS indices lie in the range 0.2–0.3 and 0.1–0.2, respectively. Many vehicles had a Jaccard index of 0.6 or higher, indicating similar behavioral patterns during the target period. Vehicles with many trip chains, higher Jaccard, and TPS indices may be engaged in route delivery operations.

Figure 9 shows the Jaccard and TPS indices based on the departure time from the base mesh. Vehicles departing from the base between 3:00–6:00 tended to be less similar in their daily activity patterns. In contrast, vehicles departing between 17:00–21:00 tended to be more similar in their daily activity patterns.

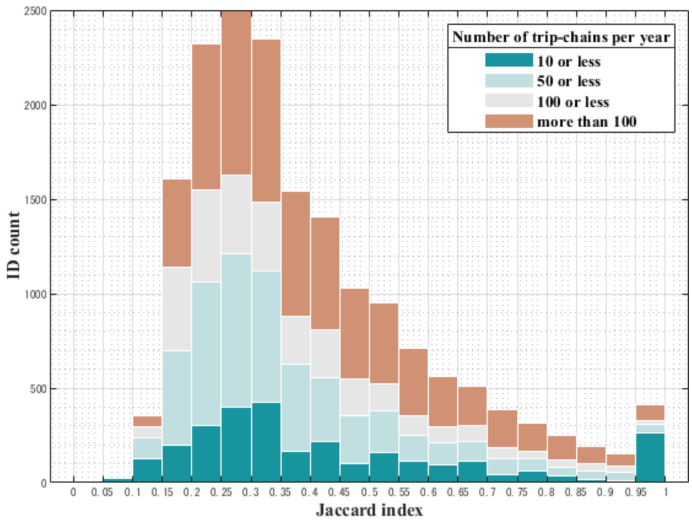

Figure 7. Jaccard index distribution

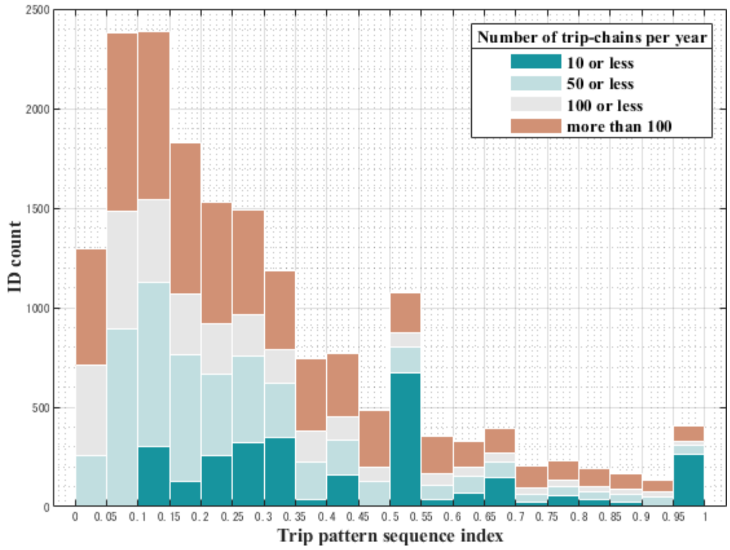

Figure 8. Trip pattern sequence index distribution

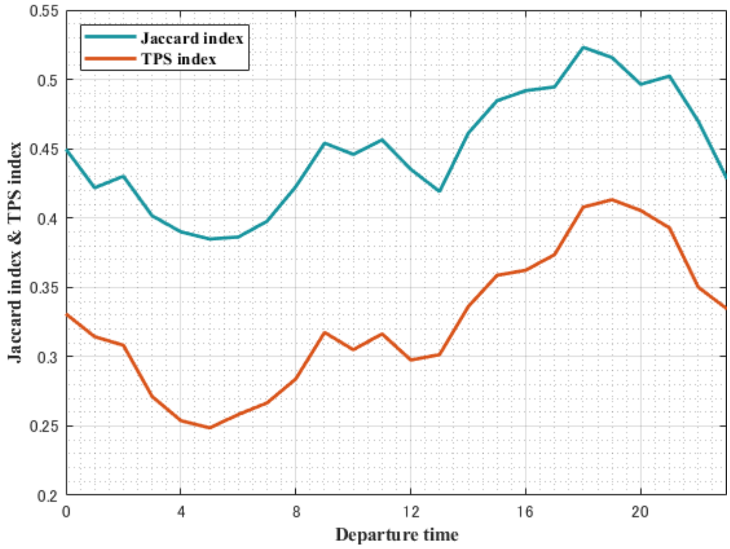

Figure 9. Average Jaccard and TPS index by departure time

3.4. Non-hierarchical cluster analysis

Vehicles were classified through non-hierarchical cluster analysis using the k-means method. This analysis used five variables: (1) number of trip chains, (2) average number of trips per trip chain, (3) average distance traveled, (4) Jaccard index, and (5) TPS index. Because each variable has a different scale, standardized values with a mean of 0 and a standard deviation of 1 were used. The results are presented in Table 4 and Figure 10. Cluster 1 had a larger average number of trips per trip chain and less similar behavioral patterns. Cluster 2 had a longer average distance per trip chain and a smaller number of trip chains per year. For each variable, cluster 3 exhibited a smaller value. Cluster 4 revealed larger trip chains per year, fewer trips per trip chain, and shorter distances per trip chain. Cluster 5 exhibited high Jaccard and TPS indices. Additionally, vehicles with a high Jaccard index had a high TPS index, and they were included in similar clusters.

Figure 11 shows the box-and-whisker plot of the number of trip chains for each cluster. Vehicles in Cluster 4 tended to have more trip chains than those in the other clusters. Figure 12 shows the box-and-whisker plot of the average number of trips per trip chain. Vehicles in

Cluster 1 have a higher number of trips, indicating that they tend to pass through multiple destinations per trip chain. Figure 13 shows the box-and-whisker plot of the average distance traveled per trip chain. Vehicles in cluster 2 traveled longer distances per trip chain.

Table 4. Result of cluster analysis: The standardized average value for each cluster

	Cluster type				
	1	2	3	4	5
Number of the trip-chain per year	-0.144	-0.196	-0.163	2.8	-0.127
The average number of trips per trip chain	1.711	0.314	-0.252	-0.412	-0.597
Average distance per trip-chain	0.285	2.059	-0.204	-0.471	-0.528
Jaccard index	-0.746	-0.091	-0.352	0.396	1.591
Trip pattern sequence index	-0.638	-0.373	-0.478	0.01	1.105

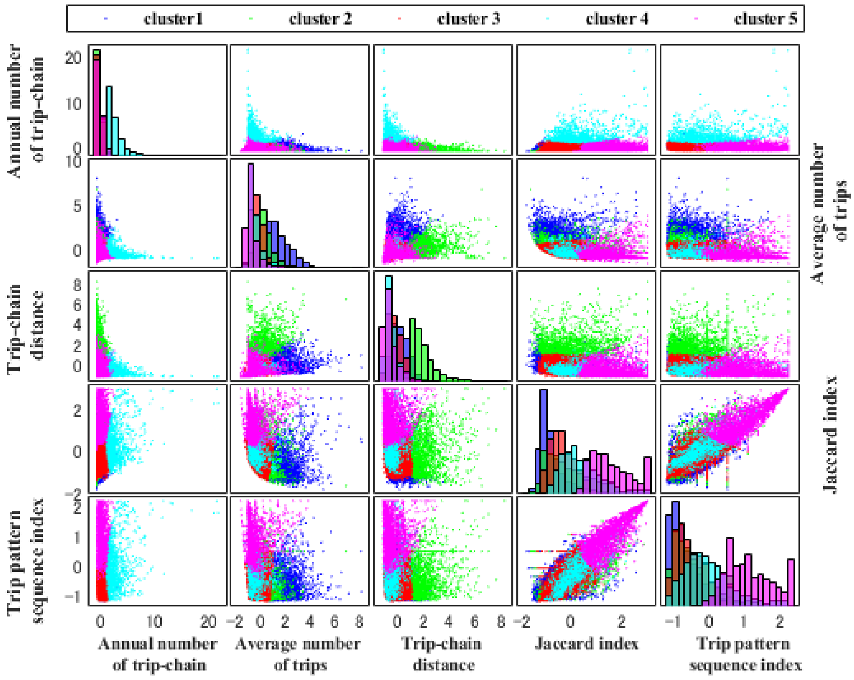

Figure 10. Matrix scatter plot

Figures 14 and 15 show box-and-whisker plots of the Jaccard and TPS indices. The vehicles in Cluster 5 have higher indices, suggesting that they engage in route deliveries. The vehicles in Cluster 3 had smaller indices. In this study, we named cluster 1 as the “multiple-destination type,” cluster 2 as the “long-distance type,” cluster 3 as the “variety type,” cluster 4 as “multiple departures from bases type,” and cluster 5 as the “similarity type.” (Table 5)

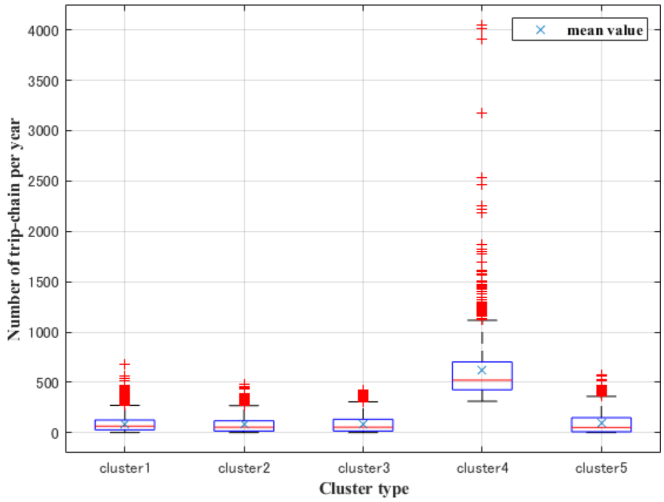

Figure 11. Number of trip chains per vehicle per year

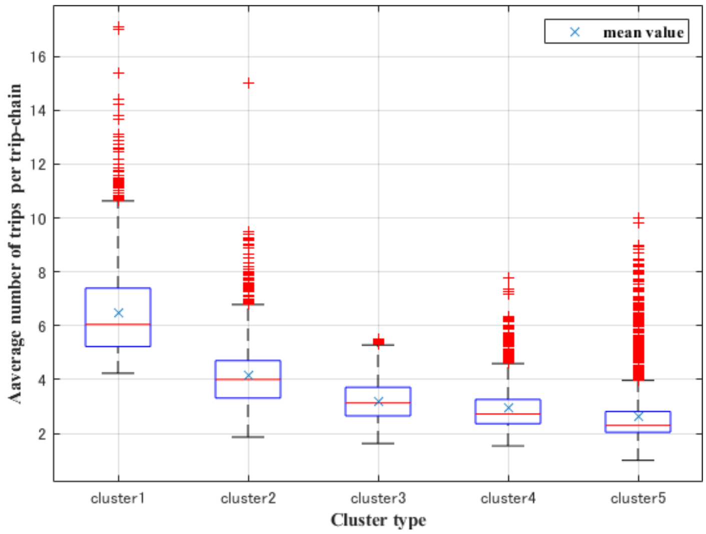

Figure 12. Average number of trips per trip chain per vehicle

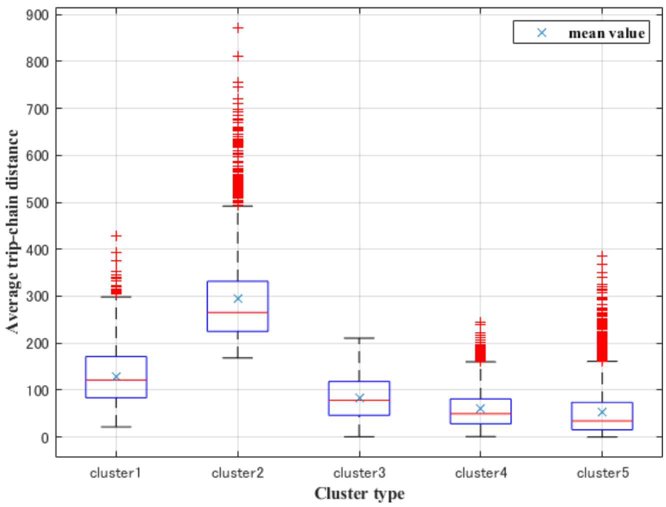

Figure 13. Average trip chain distance per vehicle

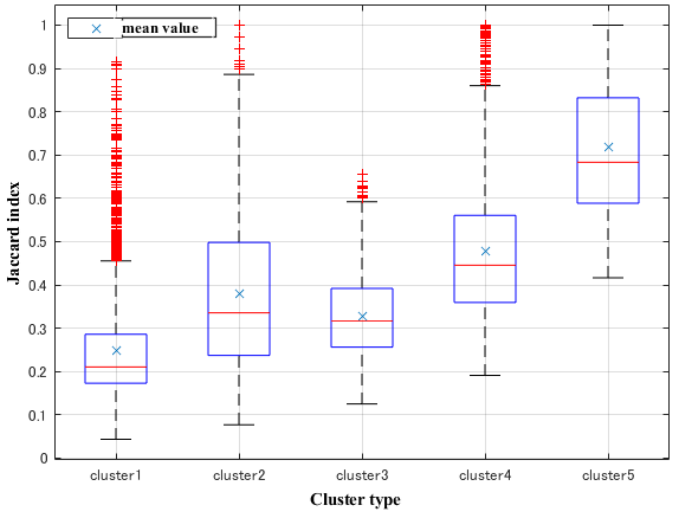

Figure 14. Jaccard index for each cluster

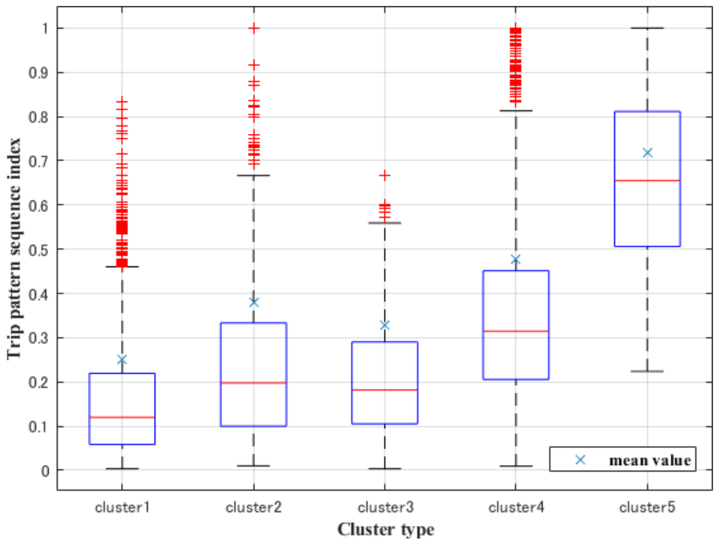

Figure 15. Trip pattern sequence index for each cluster

Table 5. Trend by cluster

Cluster	Trend	Name
1	Vehicles tend to pass the multiple destinations	Multiple-destination type
2	Vehicles take long-distance travel	Long-distance type
3	Vehicles go to fewer destinations and have various chain patterns	Variety type
4	Vehicles reveal many trip chains per year	Many-trip-chain type
5	Vehicles reveal similar behavioral patterns	Similarity type

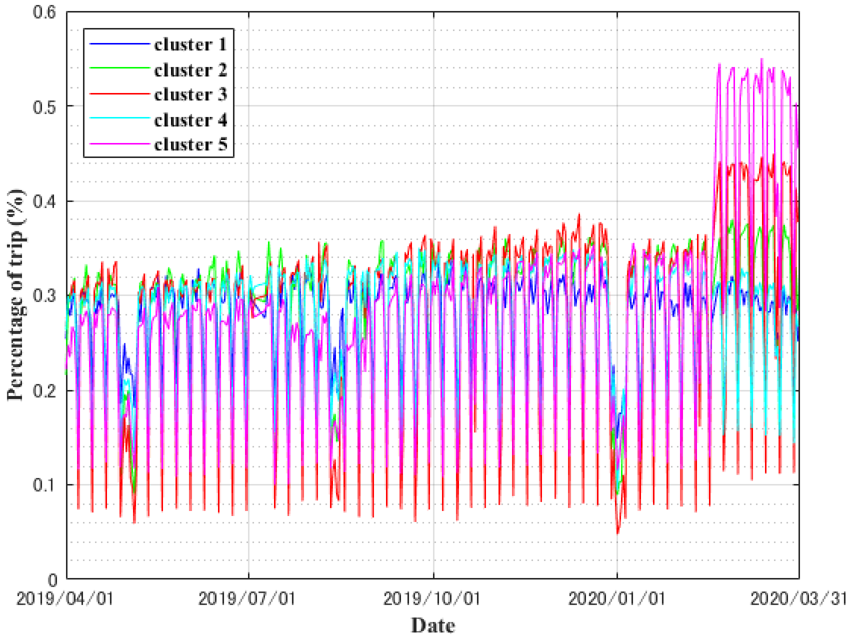

Figure 16. Daily trip variation for each cluster: The percentage of trips on the day / the summarized trips on the year

Figure 16 shows the daily trend of the total number of trips per cluster. In all the clusters, the number of trips tended to decrease during the weekends and holidays. These results are naturally due to the impact of work holidays. Clusters 3 and 5 were observed to have more trips in March. The results can be attributed to the effect of the moving season in Japan, which may have increased the demand for truck transportation.

Figure 17 shows the trip departure distributions for each cluster. Cluster 1 revealed a large number of vehicles departing at 6:00 and 13:00. The departure times of vehicles that travel through multiple destinations are concentrated at the same time of day. The result suggests that the working patterns among the vehicles with multiple-destination are similar. Cluster 2, the long-distance travel type, had many vehicles departing at 5:00 a.m. Because these vehicles are long-distance travelers, they may tend to depart early when traffic congestion is low. Cluster 3 had a large number of vehicles departing during the daytime. Cluster 4 showed no significant decrease in the number of departures, indicating that the departure time differed among vehicles. Cluster 5 showed increased departures in the evening, with a peak at 15:00. Thus, departure times tended to differ among the clusters.

Figure 18 shows the box-and-whisker diagram of the average number of trips per day for the clusters. In this figure, we focus on meshes with 10 trips/day or more. Cluster 4 had a large quartile range, indicating that the mesh tended to have many vehicles with a large amount of generated traffic. Cluster 4 tends to have many trip chains, and the Jaccard and TPS indices are also high, suggesting that these factors are responsible for a large amount of generated traffic.

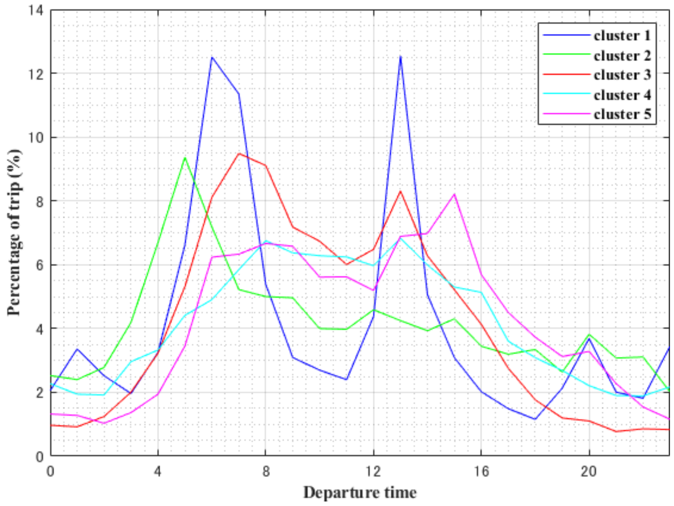

Figure 17. Departures time distribution by clusters

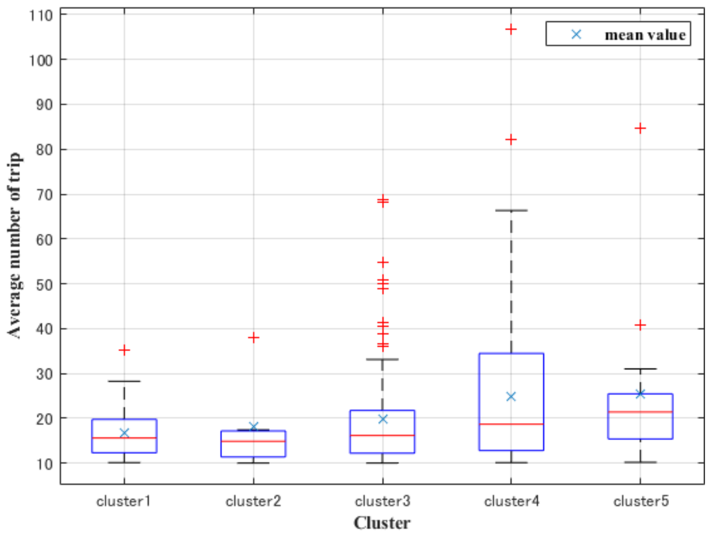

Figure 18. Average number of trips per day

Figure 19 shows a box-and-whisker diagram of the monthly OD flow with five trips/day or more for Cluster 3. The OD flow increased in December, January, and March compared to other months. However, the other clusters showed little variation in the monthly OD flow, indicating that the difference in monthly OD flow is larger for vehicles in the variety-type cluster.

Figure 20 shows the daily trend of the percentage of vehicles unloaded by the cluster. The unloaded vehicle rate in Clusters 1 and 4 tended to be low throughout the year. Vehicles with multiple destinations (Cluster 1) and many trip chains (Cluster 4) appeared to provide efficient transport. The unloaded vehicle rates for Clusters 2, 3, and 5 are similar.

Figure 21 shows the duration of stay per stop for each cluster. The durations tended to be longer in Clusters 2 and 5. Vehicles with long distances (Cluster 2) and similar daily patterns (Cluster 5) appeared to have longer loading/unloading times.

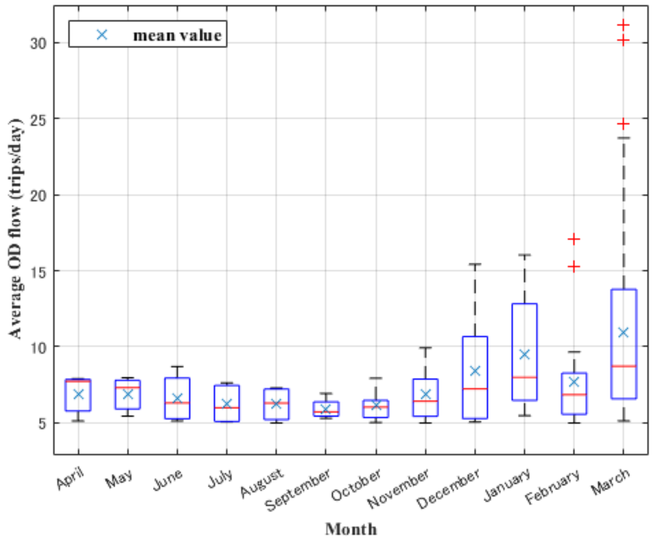

Figure 19. Average OD flow for Cluster 3

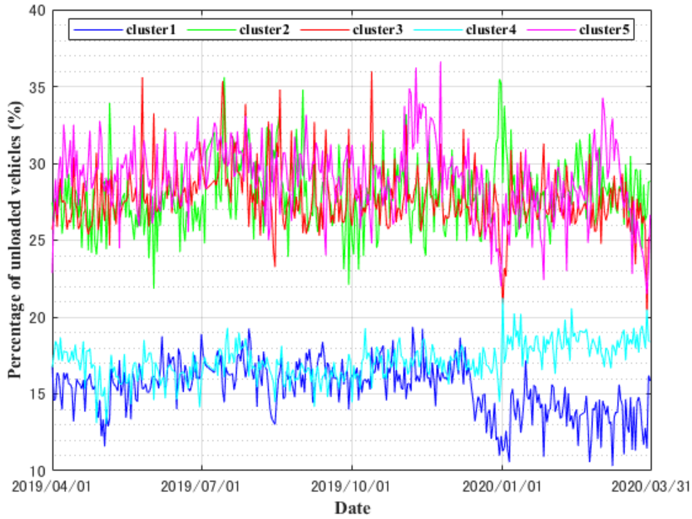

Figure 20. Change in the percentage of unloaded vehicles

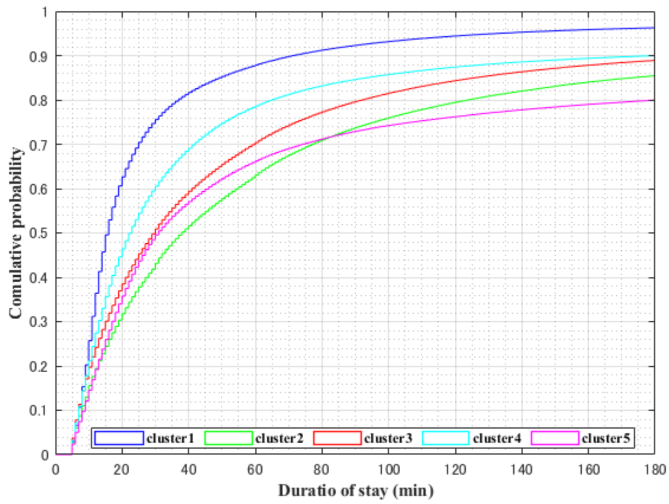

Figure 21. Duration of stay per stop

4. CONCLUSIONS

In this study, we analyzed the behavioral patterns of commercial vehicles in Kyushu, Japan, using digital tachograph data. Our exploratory analysis yielded the following findings:

1) Some vehicles revealed a higher similarity in terms of behavioral patterns throughout the year.
2) Commercial vehicles departing between 3:00 and 6:00 tend to be less similar in terms of their daily activity patterns. In contrast, commercial vehicles departing between 17:00 and 21:00 tend to have similar daily activity patterns.
3) The number of trips tends to decrease on weekends and holidays.
4) The total number of trips tends to increase in March for commercial vehicles with short trip distances, number of trips, and number of trip chains, and for commercial vehicles with low similarity and similar behavioral patterns during the year.
5) Commercial vehicles with long-distance trip chains tend to depart at 5:00.
6) Commercial vehicles that pass through multiple destinations exhibit a similar trend in the departure time from the base, with the most frequent departure times at 6:00. and 13:00.
7) Commercial vehicles that are less similar in terms of their behavioral patterns during the year tend to increase their OD flow from January to March.
8) Unloaded vehicle rate tends to be low throughout the year for vehicles with a higher number of destination stop and trip chains.
9) Commercial vehicles that travel long distances and have similar daily activity patterns tend to stay longer at their destination.

The findings were based on a novel methodology of measuring the trip chain similarity on the basis of the Jaccard and TPS indices. However, several limitations should be carefully addressed in future studies. First, the present study is an exploratory analysis that lacks a clear hypothesis. A hypothesis testing approach for the data will be useful in obtaining clear policy implications. Second, our analysis is vehicle-based, and the difference in drivers is neglected. Third, certain vehicles cannot be accurately determined under the current base-mesh determination for the trip chain. This method is required to be improved by considering the stop duration. Further analysis of the stop duration and unloaded rate using probe data will be useful in developing logistic policies in the future.

REFERENCES

Akter, T., Hernandez, S. (2023). Representative truck activity patterns from anonymous mobile sensor data. *International Journal of Transportation Science and Technology*, 12(2), pp. 492-504.

Alho, A, R., Sakai, T., Chua, M., Jeong, K., Jing, P., Ben-Akiva, M. (2019). Exploring algorithms for revealing freight vehicle tours, tour-types, and tour-chain-types from GPS vehicle traces and stop activity data. *Journal of Big Data Analytics in Transportation,* 1,175-190.

Furukawa, T., Ando, H., Kakimoto, R. (2023). The detection method of abnormal vehicle behavior using point sequence vehicle trajectory data. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, 78(5), I_385-I_394.

Hosoe, M., Kuwano, M., Moriyama, T., Miyazaki, K., Ito, M. (2021). Extraction of

similar behavior from smart card data. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, 76(5), I_957-I_966.
Joubert, J. W., Meintjes, S. (2016). Freight activity chain generation using complex networks of connectivity. *Transportation Research Procedia*, 12, 425-435.
Kawasaki, Y., Umeda, S., Kuwahara, M. (2019). Detection and analysis of detours of commercial vehicles during heavy rain in western Japan using machine learning technology. *Journal of Japan Society of Civil Engineers, Ser. B1(Hydraulic Engineering),* 75(1), 143-154.
Khan, M., Machemehl, R (2017). Analyzing tour chaining patterns of urban commercial vehicles. *Transportation Research Part A: Policy and Practice,* 102, 84-97.
Lin, JJ., Zhou, W (2013). How urban commercial vehicle trips are chained: Case studies in Texas and Idaho. *Transportation Research Board 92nd Annual Meeting* No.13-3121.
Maeda, R., Maruyama,T. (2022). Variation analysis of truck OD flow using commercial vehicle probe data. *Proceedings of Infrastructure Planning Conference, JSCE,* 65.
Morgul, E.F., Ozbay, K., Iyer, S., Holguin-Veras, J. (2013). Commercial vehicle travel time estimation in urban networks using gps data from multiple sources. *Transportation Research Board 92nd Annual Meeting*, No. 13-4439.
Nishiuchi, H., Todoroki, T. (2012). Behavior analysis of public transport passengers for discussing transport marketing schemes using smart card data. *Journal of Japan Society of Civil Engineers, Ser. F3 (Civil Engineering Informatics)*, 68(2), II_8-II_17.
Seya, H., Zhang, J., Chikaraishi, M., Jiang, Y. (2020). Decisions on truck parking place and time on expressways: an analysis using digital tachograph data. *Transportation*, 47, 555-583.
Seyama, S., Akira, N., Odani, M. (2019). Study on truck movement for collecting delivering railway container cargo using digital tachograph. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, 75(5), I_659-I_668
Siripirote, T., Sumalee, A., and Ho, HW (2020). Statistical estimation of freight activity analytics from Global Positioning System data of trucks. *Transportation Research Part E*, 140, 101986.
Tamagawa, D., Yokota, T., Maekawa, K., Komono, I. (2011). Analysis of behaviour of scheduled delivery/pick up trucks in urban area using probe data and ETC data. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, 67(5), I_715- I_726.
Torii, Y., Kobayakawa, S., Kikuchi, H. (2020). An analysis of concentrated districts of freight vehicles by using traveling data of freight vehicles and on-street parking data. *Journal of Japan Society of Civil Engineers, Ser. D3 (Infrastructure Planning and Management)*, 75(6), I_575-I_582.
Wu, C., Sun, C., Chu, D., Huang, Z., Ma, J., Li, H. (2016). Clustering of several typical behavioral characteristics of commercial vehicle drivers based on GPS data mining: Case study of highways in China. *Transportation Research Record*, 2581(1), pp.154-163.