Analysis of Features of the Routes Using Probe Trajectory Data

Yosuke Kawasaki [a], Shogo Umeda [b], Masao Kuwahara [c]

[a] *Graduate School of Information Sciences, Tohoku University, Japan*
Email: kawasaki-y@plan.civil.tohoku.ac.jp
[b] *Graduate School of Information Sciences, Tohoku University, Japan*
Email: umeda@plan.civil.tohoku.ac.jp
[c] *Graduate School of Information Sciences, Tohoku University, Japan*
Email: kuwahara@plan.civil.tohoku.ac.jp

Keywords: probe vehicle trajectory, rat-run, clustering, dynamic time wrapping, non-negative matrix factorization

1. Introduction

This study proposes a method to detect rat runs using probe trajectory data and analyzing their feature quantities. In this study, rat run is defined as a route that includes a community road. However, it is assumed that there is a main route (a route that does not include a community road) in addition to a rat run. A community road is defined as a road that residents use to move throughout an area or access the main road. In urban areas, rat-run traffic has become an issue due to the frequent use of these community roads to avoiding traffic congestion. However, a comprehensive understanding of rat runs is lacking, and the characteristics of rat runs differ depending on the location, making them difficult to evaluate. In general, rat runs are routes that are shorter and require less travel time than main routes. However, we speculate that there are routes used as rat runs even if the time and distance of the commute are not shortened. Therefore, we try to clarify which feature quantities cause residents to choose rat runs by clustering probe trajectories based on route characteristics.

Previous studies have proposed using time, distance, and space (TSD) to evaluate the difference between trajectories using the position of probe trajectories of two vehicles as feature quantities (Yoshida et al., 2018). Using vehicle trajectory, they attempt to determine which detours residents in the area are likely to use in the event of a disaster. This study differs in that it deals with rat runs and analyzes the feature quantities, including vehicle behavior data, in addition to information regarding the position of the probe vehicle.

2. Methodology

This study used ETC2.0 probe data. ETC2.0 records a vehicle's GPS position data at every fixed distance. Therefore, the number of measurement points differs for each piece of trajectory data. ETC2.0 also records data regarding the vehicle's acceleration and deceleration.

First, the feature quantities of the route (i.e. vehicle trajectory) were defined in order to detect rat runs. Table 1 shows a list of feature quantities. As shown in Table 1, there are 13 feature quantity indicators divided into four categories: basic information, vehicle speed, sudden behavior, and route shape. Basic items and vehicle speed mainly include indicators related to traffic congestion, such as required time and speed. In addition, data regarding sudden behavior was collected. Sudden behavior allows us to evaluate the driving comfort and safety of the route. The differences in the shapes of the routes were evaluated using dynamic time wrapping (DTW), which is a technique that calculates the path with the shortest distance between series by comparing the distance between each point of two time series trajectories in a brute force manner (Donald and James, 1994). DTW can be applied even when the observation points of two trajectories are different, as they are in the present study.

Table 1. Feature quantities of routes

Item	Feature quantity	
Basic information	Departure time [t = 0, 1, …, 23]	
	Travel time [min]	
	Distance [km]	
Vehicle speed	Maximum speed (Max speed) [km/h]	
	Average speed (Ave speed) [km/h]	
	Minimum speed (Min speed)[km/h]	
Sudden behavior	Maximum longitudinal acceleration (Max long acc) [G]	
	Maximum lateral acceleration (Max lat acc) [G]	
	Number of sudden behaviors	
Shape of route	Maximum angle [degrees]	
	Dynamic time wrapping (DTW):	Average value (Ave DTW)
	DTW:	DTW with the shortest path trajectory (Shortest DTW)

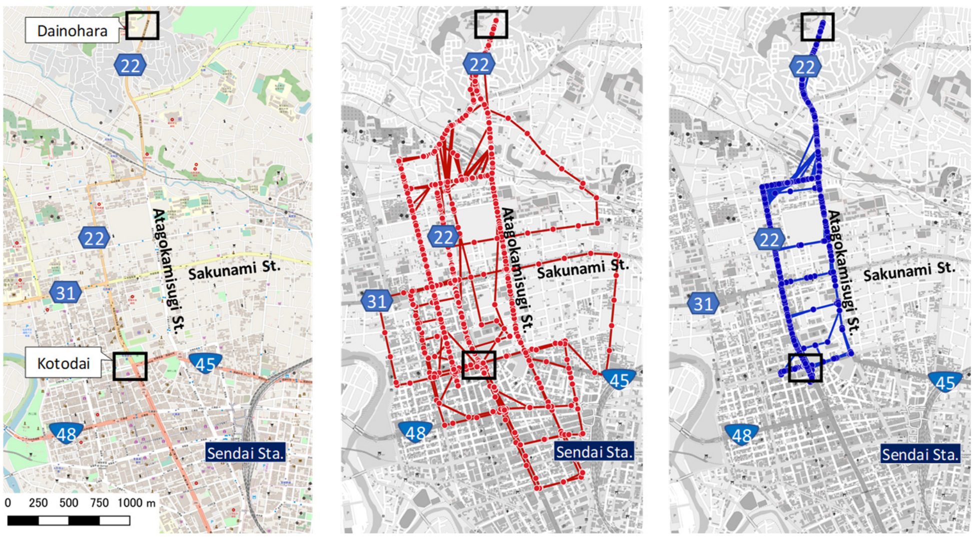
This section describes the vehicle trajectory clustering method used in this study. For clustering, non-negative matrix factorization (NMF) was used (Lee and Seung, 1999). This method is a type of matrix decomposition and can be applied to clustering. Until now, it has been used for various types of data, such as sound, image, text, and genetic data. It has been pointed out that results obtained using NMF depend on the initial value. Therefore, non-negative double singular value decomposition, proposed by Boutsidis and Gallopoulos (1999), was used to set the basic matrix and determine the initial value. This method is unique in that the initial value is determined systematically. Details of the proposed method are described in the full paper.

3. Results

The route we analyzed lies between Kotodai mesh (250m×250m) and Dainohara mesh (250m×250m) in Sendai City, Miyagi Prefecture, Japan, as shown in Fig. 1(a). The date was collected over the course of three months, from

May to July 2019. The total number of vehicle trajectories during that period is 2,762. The vehicle trajectory data is clustering by NMF using feature quantities. Here, for clustering, it is necessary to give the number of clusters K. In this study, the number of clusters K (K = 10) when the coefficient was maximized was set by trial calculation of the cophenetic correlation coefficient. As a result of clustering, all vehicle trajectories were classified into 10 groups. The set route of each group was checked; we named that the group with the highest usage of the community road was the "rat-runs group ," while the group with the lowest usage of the community road was the "main-route group." In the following, we will analyze these two groups. Figures 1(b) and 1(c) show the rat-runs group and main-route group, respectively. As is evident by the figure, the rat-runs group has a larger variety of route patterns compared with the main-route group. And looking at the rat-run group, we can see a route that passes through the area south of the origin. Probably, it is a pick-up actions (e.g. delivery, drop-in). The full paper will try to eliminate pick-up actions.

The importance of the feature quantities of the rat-runs and main-route groups was analyzed using the random forest method (Breiman, 2001). Random forest is used to evaluate the importance of features that contribute to data classification. Figure 2 shows a parallel coordinate plot of feature quantities. They are arranged from left to right in order of importance. As shown in the figure, the average speed of vehicles that use rat runs is low, and the travel times of these routes tend to be long. As mentioned above, in general, rat runs indicate routes that are shorter and require less travel time than the main route, but they have different feature quantities. In addition, maximum angle of rat--runs tends to be smaller than the main route. Also, no sudden behavior has occurred in rat-runs. Based on these facts, it can be inferred that rat runs are used by drivers who dislike the shape(curve) and sudden behavior of the main route.



(a)Analysis target (b) Rat-runs group (n = 82) (c) Main-route group (n = 476)

Figure 1. Result of clustering vehicle trajectories

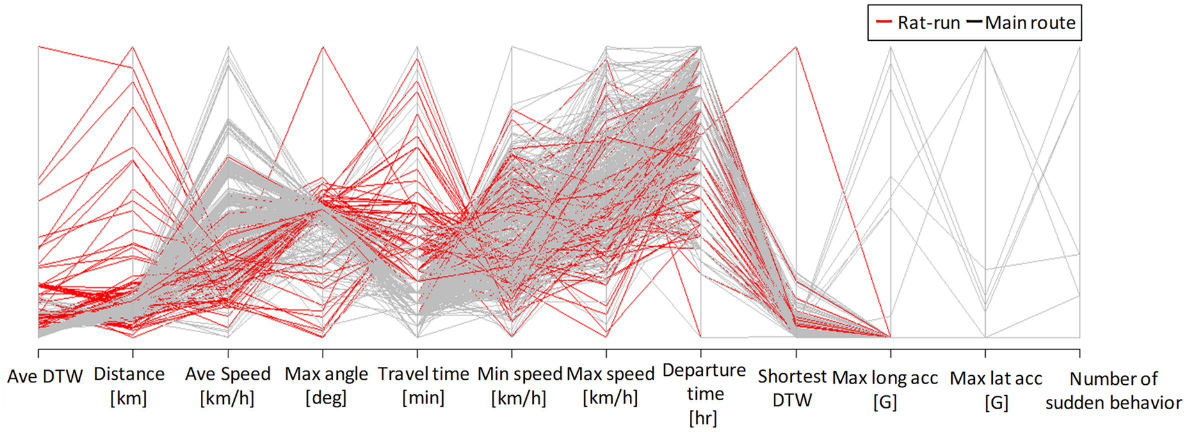


Figure 2. Parallel coordinate plot of routes' feature quantities

4. Conclusion

This study proposed a method to identify rat runs using probe trajectory data and analyzed the feature quantities of rat runs. In this paper, only the rat-runs and main-route groups were analyzed; thus, in future research, other groups must be analyzed. Moreover, it is necessary to expand the analysis and confirm the robustness of this method.

Acknowledgments

This work is supported by the Ministry of Land, Infrastructure, Transport and Tourism's "Technology Research and Development Contributing to Improving the Quality of Road Policies," "Technology Research and Development on Location and Traffic Management Based on Multi-Scale Traffic Cooperation" and JSPS KAKENHI grant number JP19K15107.

References

Breiman, L.: Random Forests, Machine Learning, Vol. 45(1), pp. 5–32, 2001.

Donald, J. B. and James C.: Using dynamic time warping to find patterns in time series. In KDD workshop, Vol. 10, pp. 359–370, 1994.

Lee, D. D. and Seung, H. S.: Learning the parts of objects with nonnegative matrix factorization, Nature, 401, 788–791, 1999.

Yoshida, M., Umeda, S., Kawasaki, Y. and Kuwahara, M.: Incident alert by an anomaly indicator of probe trajectories, Transportation Research Procedia, Vol. 34, pp. 179–186, 2018.