Transdisciplinary Engineering Methods for Social Innovation of Industry 4.0
M. Peruzzini et al. (Eds.)

doi:10.3233/978-1-61499-898-3-857

Estimation of Delivery Time Considering with Stochastic Service Times

Kenji TANAKA[1], Kazuho KOZAWA and Kenji ARAOKA
Department of systems innovation, School of Engineering, The University of Tokyo

Abstract. This paper talks about how to plan efficient delivery route by analyzing drivers' behavior. Today, as E-commerce services represented by Amazon are growing, demand of delivery service is getting bigger. It is a problem that not only the amount of delivered baggage is expanding, but also the level of delivery service is rising. On the other hand, the number of drivers is reducing because of a low salary and poor working conditions. From these background, it is required to improve efficiency of delivery to keep delivery service. Especially in urban areas, the most time-consuming work in last mile delivery is not moving between customers by a truck but work occurring after parking (including parking). In spite of this, many researches focus on how to minimize the total time of moving between customers, like solving traveling salesman problem (TSP). So this paper has tried analyzing drivers' "parking behavior", consisting of a successions of behavior such as searching for parking place, parking a truck, taking out baggage from a truck and visiting customers to entrances.On this paper, I have tried analyzing by following steps. First, I have detected each parking behavior from the GPS data of the trucks by 2-stage clustering method, which has resistance to noise. Next, I have matched each parking behavior with each customers. Then, I can analyze how drivers decide parking behavior by customers, and we can plan efficient delivery routes.

Keywords. last mile delivery, stochastic service time, GPS, clustering

Introduction

Today, as E-commerce services represented by Amazon are growing, demand of delivery service is getting bigger. It is a problem that not only the amount of delivered baggage is expanding, but also the level of required service such as time specified delivery is rising. On the other hand, the number of delivery personnel who drive a truck is reducing and the delivery infrastructure is collapsing. From this background, various studies have been done with the aim of improving the efficiency of delivery. However, these studies almost only focuses on distance cost in delivery. In many areas where customers are concentrated, such as Japan, the work except moving between customers, such as parking trucks and delivery from parking space to customers, accounts for most of the delivery time. Thus, not only improving the efficiency of delivery routes but also considering the time required for these work is important. In this study, the time required for these work is referred to as service time. It is required to quantify the service time, and to estimate service time as considering its uncertainty for supporting to develop a delivery plan.

[1] Corresponding Author, Mail: tanaka@tmi.t.u-tokyo.ac.jp.

For quantifying the service time, it is realistic to use the GPS data obtained from the mobile devices owned by delivery personnel. In regard to the method of estimating the service time from the GPS data, there are some related researches on determination of staying points using GPS data. One of the methods of determining the staying point is the spatio-temporal Mean-Shift method of Nishida et al. [1]. However, Kitazawa et al. [2] have pointed out a defect that the determined staying point may be separated due to continuous noise. As to the estimation of the total delivery time, research has focused mainly on the Vehicle Routing Problem (VRP), but the service time has been fixed value in many studies. There are some studies, Liu [3] and Binart et al. [4] for instance, focusing on the Stochastic Vehicle Routing Problem (SVRP), which treats the required time as uncertainty. However, although stochastically assumption of the service time, since that assumption is not based on the analysis of actual data, those studies have a low credibility.

In the conventional research on improving efficiency of the last mile delivery, the service time wasn't considered sufficiently. In contrast to this, in this study, it is the purpose that the service time is quantified and the time required for delivery is estimated, to support delivery. And to achieve this, the following two approaches are listed.

- Apply determination of staying points by the spatio-temporal Mean-Shift to GPS data to quantify the service time and estimate the distribution of the service time for each customer
- Estimate delivery time on the basis of the distribution of the service time and judge availability of order for delivery request on the day

1. Definition of delivery status and analysis of data accompanying delivery

1.1. Definition of delivery status

In this study, we divide the delivery status into four categories: "traveling between a depot and a delivery area", "traveling in a delivery area", "correspondence to customers" and "waiting". "Traveling between a depot and a delivery area" is defined as the state from departing a depot until getting close to an initial customer, or from departing near a last customer until reaching a depot. "Traveling in a delivery area " is defined as the state from departing near a customer until getting close to a next customer. "Correspondence to customers" is defined as the state of being parking a truck, and unloading and delivering baggage(s) to customers. This is the main target of analysis and defined as "service time" in this study. "Waiting" is defined as the state of waiting in the track because of a next customer requiring time specified delivery. The conceptual diagram of the definition is shown in Figure 1.

1.2. Analysis of data accompanying delivery

In this section, we analyze the time required for each delivery status by using the data which has been measured when the author had accompanied the delivery. The routes of the delivery are shown in Figure 2. The breakdown of the delivery time is shown in Figure 3. In the delivery of 4 hours 15 minutes from 8:24 to 12:39, the service time was

2 hours 22 minutes, accounting for 53.5% of the total, which also makes it important to estimate the service time.

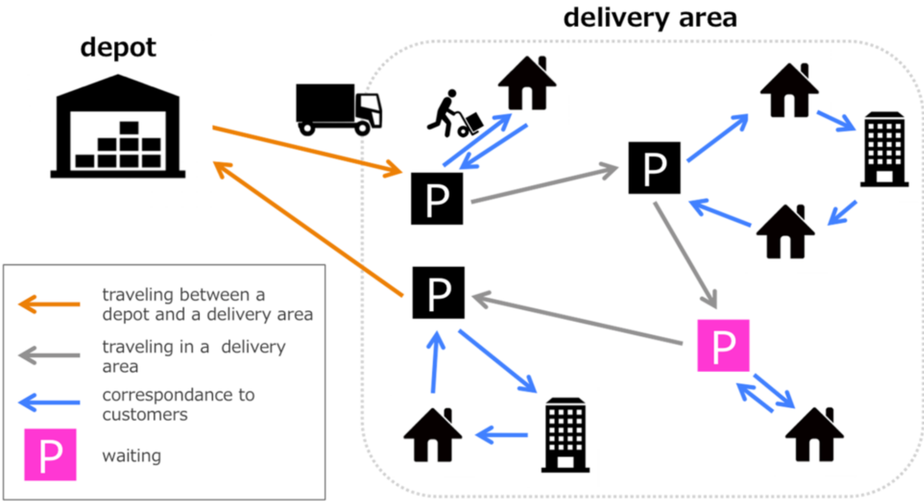

Figure 1. Conceptual diagram of the definition of delivery status.

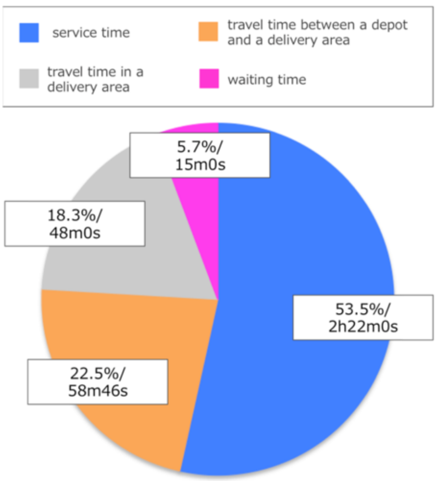

Figure 2. Routes of the delivery. **Figure 3.** Breakdown of the delivery time.

2. Estimation of service time distribution of each customer

2.1. Overall of the method

The overall flowchart of the method of estimating the service time of each customer in this study is shown in Figure 4. For input data, we use two types of data, GPS data and delivery result data. The method mainly consists of 4 steps. First, the staying point is determined from the one day's GPS data by the two-stage clustering method which is proposed in this research. Next, estimate parking and waiting behaviors by linking these staying points with delivery result data. Then, we can calculate the service time of each customer. As this repeated for all data sets, we estimate the service time distribution based on the actual data.

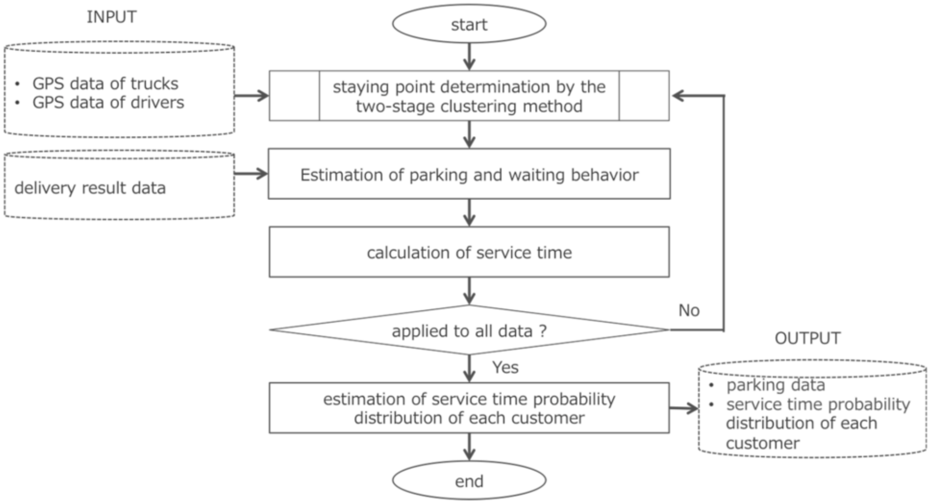

Figure 4. Flowchart of the method of estimating the service time of each customer.

2.2. Staying point determination by two-stage clustering

In this study, service time is estimated by applying the method which is named two-stage clustering method to GPS data. The two-stage clustering method is application of the Nishida et al.'s spatio-temporal Mean-Shift method. The flowchart of two-stage clustering method is shown in Figure 5. First, GPS data p_i is initialized as cluster C_i by converting position and time information of GPS data to symmetrical three-dimensional data. Next, Mean-Shift clustering is performed to this data by using the bandwidth b determined based on the distance kernel width ϕ_{dist}[m] and the time kernel width ϕ_{time}[s], thereby generating many clusters having a short duration. Then, Mean-Shift clustering by using only distance as a feature value is first performed, where the bandwidth is obtained by multiplying the bandwidth b used in the first stage clustering by the distance coupling coefficient λ_{dist}. After second clustering, the all pairs of clusters made in first clustering whose time interval is less than the combination time width λ_{time}[s] is merged as one cluster, and determine the clusters which duration is longer than the minimum staying time ψ_{time}[s] as the staying point. The feature of this method is that by setting an arbitrary combination time width λ_{time}[s] in the second stage clustering, it is possible to make the staying points determination more robust against continuous noise.

2.3. Estimation of parking and waiting behavior

We determine the parking and waiting behavior by linking the staying points generated in 3.2 with each customer using the delivery result data, which is consist of the each customer’s address and the recorded delivery time.

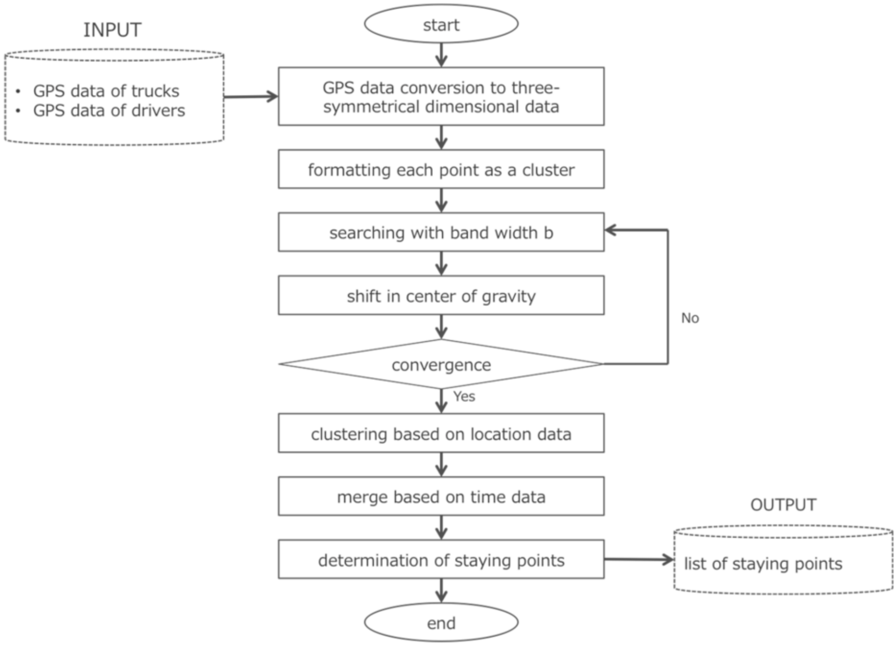

Figure 5. Flowchart of two-stage clustering method.

In this study, among the staying points extracted from the GPS data of truck, those which distance d_{trk}[m] to a customer is within the parking behavior linking distance d_{link_trk} [m] and which staying time includes the recorded delivery time of the customer are considered to parking behavior. Also, among the staying points extracted from the GPS data of delivery personnel, those which distance to the parking place of a parking behavior d_{dvr} [m] is less than the waiting behavior linking distance d_{link_dvr}[m] and which staying time is included in parking time of the parking behavior are considered to waiting behavior.

The linkage rate r_{fit} is introduced as an evaluation indicator for measuring the accuracy of two-stage clustering. r_{fit} is the value which evaluate that more parking behaviors can be estimated by appropriate staying time, and defined as:

$$r_{fit} = \frac{n_d}{n_{all}} \tag{1}$$

r_{fit} : Linkage rate

n_d : The number of linked baggage

n_{all} : The number of all baggage

2.4. Calculation of service time

The waiting behavior is regarded as unnecessary time for the original delivery, so the service time t^d is calculated according to:

$$t^d = t^p - t^w \tag{2}$$

t^d : The service time of a customer

t^p : Duration of parking[s]

t^w : Duration of wating[s]

2.5. Calculation of service time distribution by each customer

In this study, we estimate the service time distribution by creating a histogram showing the distribution of service time for each customer, based on the result of service time calculation, and approximating it by probability density function. As the probability density function, a normal distribution, a lognormal distribution, a gamma distribution, and a beta distribution are used.

3. Estimation of delivery time

The flowchart of estimation of delivery time is shown in Figure 6. The required time for each delivery status is estimated according to the definition in 2.1. First, to calculate travel time between a depot and a delivery area, we reproduce a delivery order based on a delivery result data. Then, we identify first and last customers and calculate the distance between these customers and the depot using the Google Maps Directions API. With this distance and v_{round}[m / s], travel time between a depot and a delivery area is given. Second, to calculate travel time in a delivery area, the customer centroid distance method of Fukuhara[5] is used. Eq. (3) gives the formulation to calculate the distance in a delivery area in the method. Here, we can calculate the rough estimate of travel distance in a delivery area without depending on a delivery order. Third, to calculate total service time, we obtain the convolution integration of service time probability distributions of each customer calculated in Section 3. Finally, by summing up these three times, we can get delivery time.

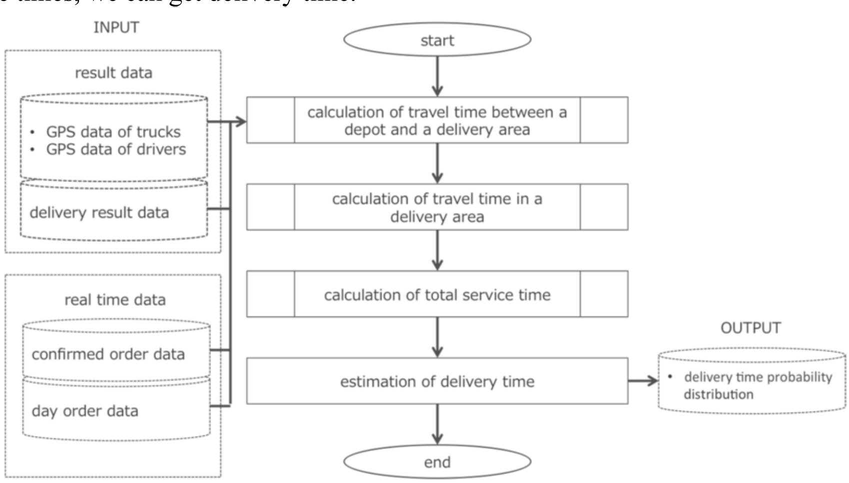

Figure 6. Flowchart of estimation of delivery time.

$$d_{area} = \alpha\sqrt{n} + \beta \cdot d_{ave\_g} \cdot Cmp_{area} + \gamma \tag{3}$$

d_{area}	:	Travel time in a delivery area[s]
n	:	The number of customers
d_{ave_g}	:	The customer centroid distance [m]
Cmp_{area}	:	The complexity of a delivery area
α, β, γ	:	Regression coefficients

4. Case study

4.1. Analyzing data and setting parameters

In this study, we conducted a case study using actual data provided by Company X. The software we used is Python 3. GPS data we used had been collected by Android smartphones on trucks or with delivery personnel for 4 months. First, we analyzed the data and set each parameter used in the two-stage clustering method. Here, we used the grid search method and evaluate each parameter on a linkage rate. As a result, the distance kernel width ϕ_{dist} = 80[m], the time kernel width ϕ_{time} = 120[m], the coupling distance coefficient λ_{dist} = 0.5, the coupling time width λ_{time} = 180[s], and the minimum residence time ψ_{time} = 60[s] are obtained.

4.2. Application results of the two-stage clustering method

In this case study, we compared application results of the two-stage clustering method and the spatio-temporal Mean-Shift clustering method. An example of the application result is shown in Figure 7. Here, staying point is divided into two by a noise of GPS data with using the coventional method. However, only one staying point is extracted with using the proposal method. As a result, the linkage rate of the conventional method is 0.564, and that of the proposal method is 0.675. This shows the superiority of the proposal method.

An example of judgment of waiting based on the two-step clustering method is shown in Figure 8. In this case, the driver who came from the upper right of the map parked near the customer's building and delivered, then went to the lower left of the map. As a result, 7 minutes 37 seconds is spent on waiting within parking time 22 minutes 18 seconds, and the service time is 14 minutes 41 seconds.

4.3. Result of estimation of service time probability distribution of each customer

In this section, based on the method described in Section 3, we estimate service time probability distribution of each customer. 95.9% of customers in the data live in apartment houses, so we estimate service time probability distribution of each building. In this study, in addition to this, the customer is classified into three types, "high-rise apartment group", "large apartment group" and "small apartment group", and estimate service time probability distribution of each group. This enables us to estimate new customer's service time probability distribution.

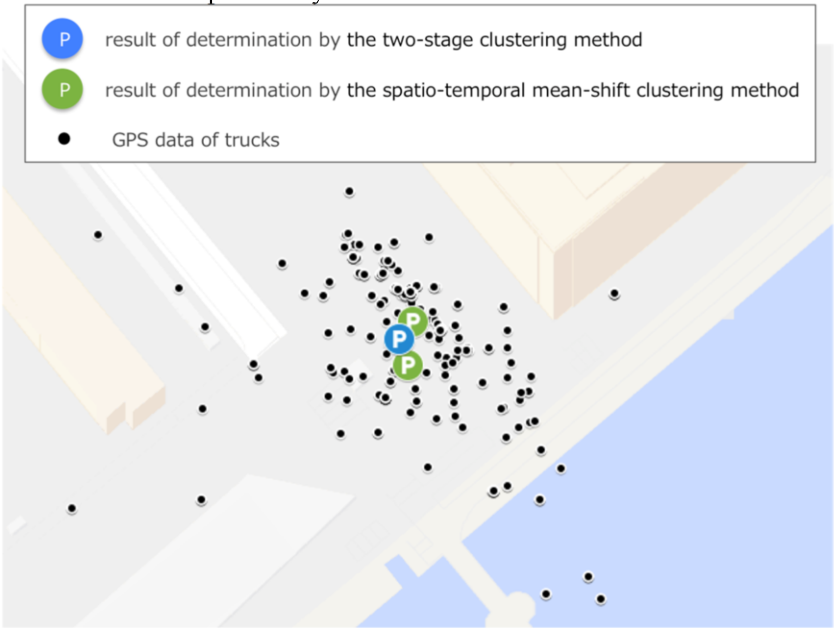

Figure 7. Example of the application result.

Figure 9 shows the service time probability distribution of a certain building belonging to a large apartment group. The relative frequency was the maximum at 480 to 600 seconds, 0.237, and the class of 360 to 600 seconds contained 44.0% of the total, and the service time was distributed around this class. For the approximation to the probability distribution function, it is approximated by the gamma function, and the median value was 525 seconds.

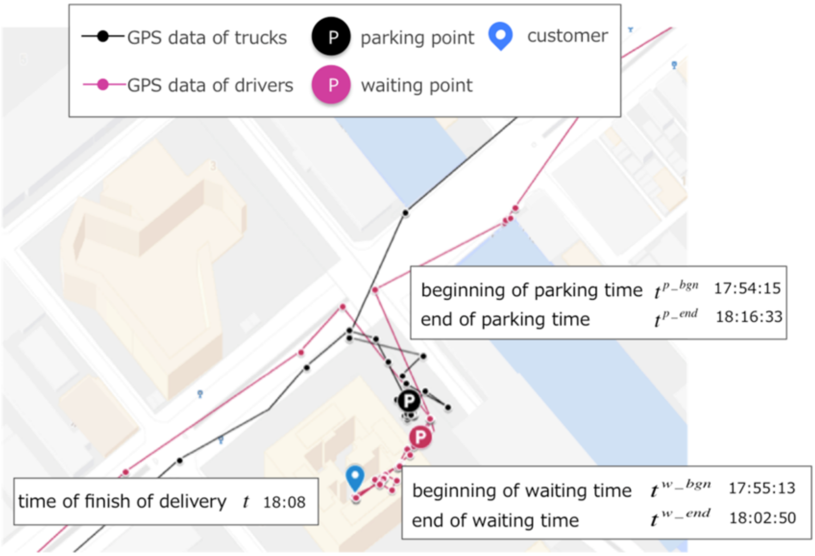

Figure 8. Example of judgment of waiting.

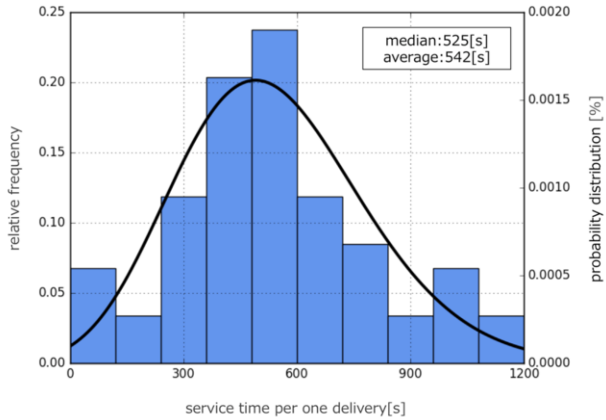

Figure 9. Service time probability distribution of a certain building belonging to a large apartment group.

4.4. Result of estimation of delivery time

In this section, validity is verified by estimating delivery time based on estimation result of service time distribution of each building conducted in 5.3 and comparing the result with existing method. The data used here consists of 15 customers in 235 minutes from 8:31 to 12:26. The results of the comparison are shown in Figure 10. Here, as an existing study, we use the model in which the service time is fixed to 5 minutes 33, which is used in Fukuhara's study. The median value of the result of the proposed method was 214.7 minutes and the error from the actual data was 8.6%. On the other hand, despite that the fixed service time was designed from the result of Fukuhara accompanying delivery, the result of the existing method has estimated the fixed value so small that the total required time was 160.0 minutes and the error rate was 25.5%. It can be seen that from the human factors measuring in accompanying delivery is not always accurate.

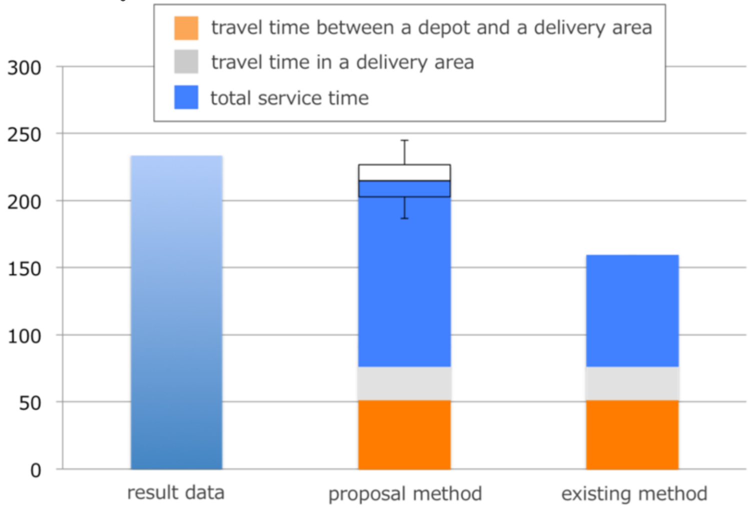

Figure 10. Result of estimation of delivery times.

5. Conclusion

The conclusion of this study is described below.

- We proposed a two-stage clustering method in the staying point determination and showed that this method is more robust than the existing method
- Based on the result of the staying point determination, the service time was calculated and the service time probability distribution of each building was estimated
- We estimated the delivery time and compare the results with the existing method to show its superiority

The method of estimating the service time probability distribution proposed in this research and its result can be incorporated into existing SVRP and it has great impact. In addition, estimation of delivery time can lead to improvement of driver's work environment and maintenance of delivery infrastructure.

References

[1] K. Nishida, H. Toda, T. Kurashima and T. Uchiyama, Probablistic Visited-POI Analysis : User Modeling from Spatio-Temporal Trajectories, In:*DICOMO*, Tokachigawa, 2013, pp. 334-345.

[2] T. Kitazawa, D. Sakamoto, T. Azumi, M. Mochiduki, N. Nishio, Analysis method of noise and loss in GPS logs, In: *DICOMO*, Tsukioka, 2014, pp. 1595-1602.

[3] X. Li, P. Tian and S.C.H. Leung, Vehicle routing problems with time windows and stochastic travel and service times: Models and algorithm, *International Journal of Production Economics*, Vol. 125, 2010, Issue 1, pp.137-145.

[4] S. Binart, P. Dejax, M. Gendreau, F. Semet, A 2-stage method for a field service routing ploblem with stochastic travel and service times, *Computers & Operations Research*, Vol.65, 2016, pp. 64-75.

[5] K. Fukuhara, *Mode choice and accommodation between areas in home delivery*, Thesis, The University of Tokyo, 2017.