Managing Parking Fees Based on Massive Parking Accounting Data

Yuichi Enoki[1], Ryo Kanamori[2], and Takayuki Ito[3]

[1] Department of Computer Science and Engineering,
Graduate School of Engineering, Nagoya Institute of Technology, Japan
enoki.yuichi@itolab.nitech.ac.jp
[2] Institute of Innovation for Future Society, Nagoya University, Japan
kanamori.ryo@nagoya-u.jp
[3] School of Techno-Business Administration,
Graduate School of Engineering, Nagoya Institute of Technology, Japan
ito.takayuki@nitech.ac.jp

Abstract. As parking accounting data of automatic payment system is accumulated, a managing parking fees in accordance with characteristics of parking utilization is expected. The purpose of this paper is to analyze the characteristics of parking utilization from a big data and to propose a procedure of parking fee management by developing of a simple simulator from a history of parking utilization. In concrete terms, we classify 1,050 parking lots by cluster analysis and analyze influence of a charge revision on parking time by survival analysis from 22.5 million parking accounting data in the past year. Further, we consider the appropriateness of modified fee by estimating parking time with a hazard-based duration model.

Keywords: characteristics of parking utilization, cluster analysis, survival analysis.

1 Introduction

Pay-by-the-hour parking spaces are actively used around downtown areas, moreover, large quantities of parking accounting data are accumulated because parking accounting systems are online. Considering characteristics of parking utilization is important for appropriate pricings of pay-by-the-hour parking spaces. The appropriate pricings according to users' demands are necessary for increasing users' convenience and profitability of parking lots. Currently, a parking charge is set by investigating utilization characteristics for each parking lot. However, there are problems in that variation of revenue becomes large because of the difference in ability for setting an appropriate parking charge, or the load of investigating all parking lots is large.

The purpose of this paper is to propose a procedure of pricing based on data analysis by pay-by-the-hour parking data. We consider utilization characteristics of each parking lot with parking accounting data. Moreover, we calculate revenue by simulating changes of parking time in accordance with changing pricing

D.-N. Pham and S.-B. Park (Eds.): PRICAI 2014, LNAI 8862, pp. 1050–1062, 2014.

based on characteristics of parking utilization. In the estimation of revenue by the simulation, it becomes possible to set parking charges by trial and error. Therefore, the difference in pricing ability can be reduced. In addition, we can grasp utilization characteristics based on data analysis to reduce the load of investigating the parking lots.

The structure of this paper is explained below. In section 2, we analyze parking data. In detail, we analyze characteristics of parking utilization by cluster analysis, influence of a charge revision on parking time by survival analysis, and change point of utilization characteristics by cluster analysis. In section 3, we propose a more appropriate pricing from the analysis results in section 2, and consider appropriateness of the pricing by simulation. In section 4, we give an overview of the related work. In section 5, we conclude this paper.

2 Analysis of Large Scale Parking Data

2.1 Parking Data

Pay-by-the-hour parking data provided by Meitetsu Kyosho Corp. is used for analysis in this study. Such information as parking location, parking capacity, parking fee, user's payment and parking time can be retrieved from this data. Data for analysis is parking accounting data on the 1,050 parking places around Nagoya City. This data set contains 22.5 million accounting data from October 1, 2011 to October 3, 2012.

2.2 Analysis of Characteristics of Parking Utilization

The parking lots are divided into eight clusters by k-means clustering [1,2] for classifying the parking into each utilization characteristic. Variables include utilization rate of parking, parking time, availability of commuter pass, utilization rate of point card, discount rate of coupon, and discount rate by maximum charge.

In the spatial distribution feature of the clusters, the clusters are divided into a cluster that is located in a city center (urban types), a cluster that has a spatial wide distribution (suburban types), and a cluster that has a distribution feature between urban types and suburban types (middle types). Distributions of cluster 3, cluster 7, and cluster 8 are illustrated in Figure 1. Then, average and variance of distance of the nearest station are calculated for each cluster. As a result, difference in average distance of the nearest station is almost nothing. However, there are many parking lots around stations in cluster 1, cluster 4, and cluster 7 because these variances are low.

Characteristics of cluster 3, cluster 7, and cluster 8 are explained according to the above results.

Cluster 3. This cluster's distribution is a middle type. The utilization rate of parking and the parking time have middle Characteristics; therefore, this cluster is a cluster to which general parking spots used for various applications belong.

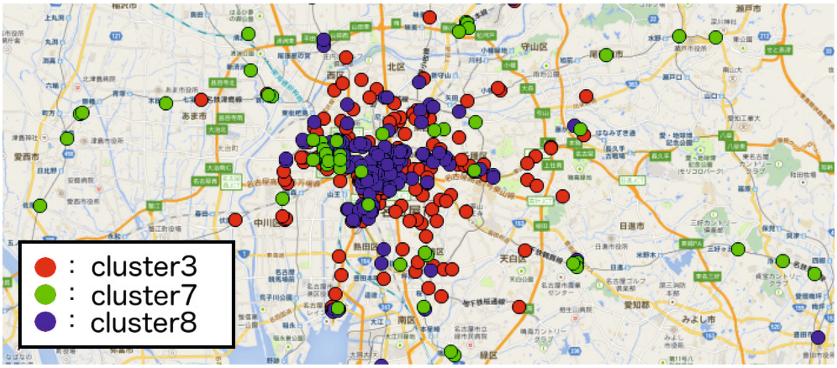

Fig. 1. Cluster distribution

Cluster 7. This cluster's distribution is a suburb type. Most of parking spots that belong to cluster 7 are located around train stations. The parking time between 1 o'clock and 7 o'clock is long. This cluster is used for a Park and Ride system. Motorists can park their cars near the station, and use trains to reach their destinations.

Cluster 8. This cluster's distribution is an urban type. The utilization rate of parking on holidays is high. The parking time is short. The discount rate of maximum charge is low. This cluster is one to which parking areas used for a short term belong.

2.3 Parking Time Model

Survival analysis [3] is a method that analyzes a relation between the time remaining until an event occurs and the event itself. Survival analysis is used for analyzing time until death by disease or until breakdown of a mechanical system in the fields of medical service, engineering, and so on. Time until an event occurs is called survival time. An analysis of relation between each covariate becomes possible by regarding parking time as survival time. A probability distribution is a weibull distribution [4] because parking time is drawn from weibull distribution [5]. From the above assumption, a probability of survival can be represented by the following formulation as in (1)

$$S(x|Z) = [exp(-\lambda x^{\alpha})]^{exp(\beta^T Z)} \tag{1}$$

where $S(x|Z)$ is the probability of survival given Z at survival time x, Z is a column vector of covariates, β^T is a row vector of coefficients, λ and α are parameters of weibull distribution. Influence of charge per hour, whether to set maximum charge, time zones in parking, holiday, and payment method as covariates on parking time is examined for each cluster by survival analysis.

Charge Per Hour. Value of charge per unit time is a variable of charge per hour.

Whether to Set Maximum Charge. If the maximum charge is set at the time a car enters a parking lot, the value of the variable is 1; otherwise, the value is 0.

Time Zones in Parking Time zones in parking are presented with three variables. Three variables are all 0, or one of the three is 1. Spans divided into four time zones in a day are assigned each variable.

Holiday. If a day when a car enters a parking lot is a Saturday, a Sunday or a festival day, the value of the variable is 1; otherwise, the value is 0.

Payment Method. Three variables indicate whether to use a coupon ticket, a credit card, or a point card in charge payment. If a coupon ticket, a credit card, or a point card are used, the value of the variables is 1; otherwise, the value is 0.

If the value of each coefficient is positive, the coefficients have the probability of exiting a parking lot decreasing; otherwise, the coefficients have the probability increasing. The higher the absolute value of a coefficient is, the bigger the influence of a coefficient is. The value of each covariate is considered in cluster 3, cluster 7, and cluster 8.

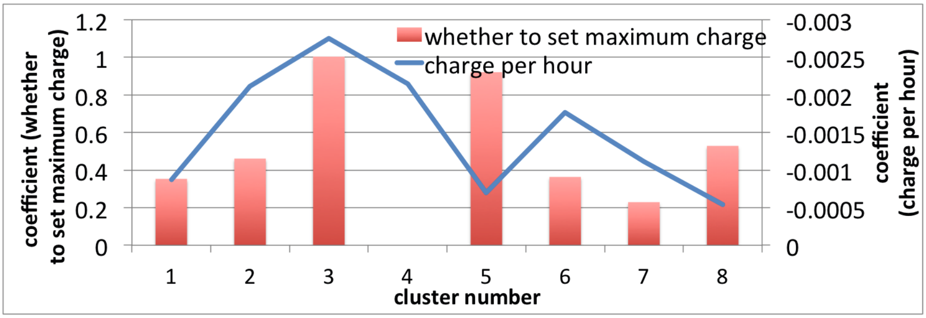

Fig. 2. Values of charge per hour and whether to set maximum charge for each cluster

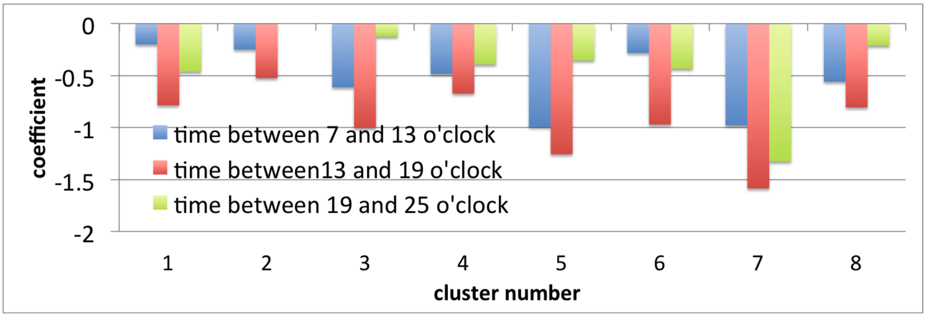

Fig. 3. Value of time zones in parking for each cluster

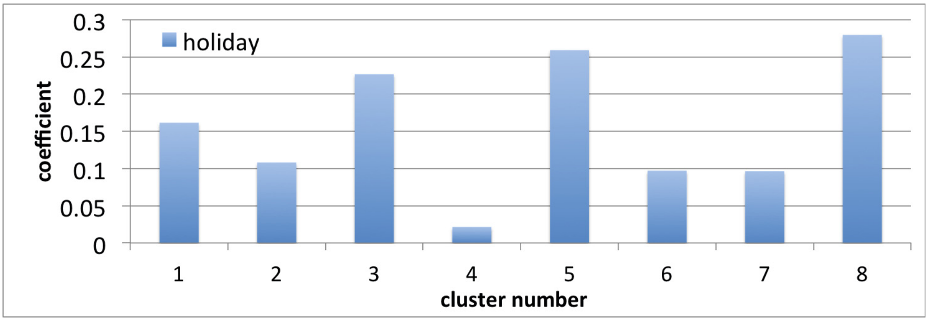

Fig. 4. Value of holiday for each cluster

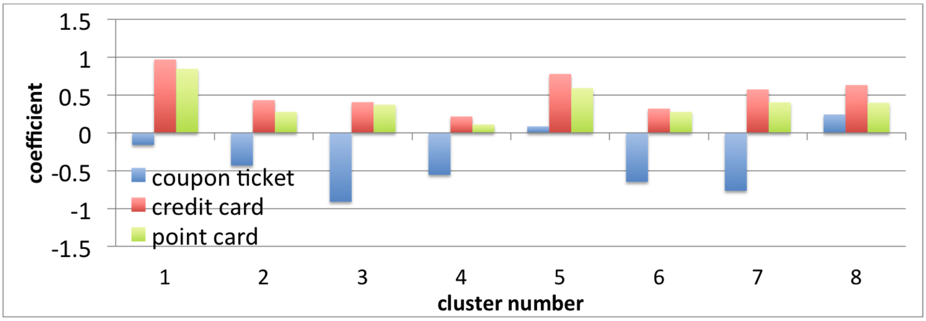

Fig. 5. Value of payment method for each cluster

Setting Price. The coefficient of setting price means the coefficient of charge per hour, and the coefficient of whether to set maximum charge is considered. The coefficient of setting price is presented in Figure 2. As for the coefficient of whether to set maximum charge, the influence is significant in cluster 3, and is not significant in cluster 7. As for the coefficient of charge per hour, the influence is significant in cluster 3, and is insignificant in cluster 7 and cluster 8. Cluster 3 is a cluster where change in parking time depending on charge revision is the most drastic because the absolute values of charge per hour and whether to a set maximum charge are the biggest. Cluster 7 is a cluster where charge in parking time depending on change of charge is small because the absolute values of charge per hour and whether to a set maximum charge are small.

Time Zones in Parking. The coefficients indicating time zones in parking are presented in Figure 3. The coefficients are negative in all clusters. Therefore, users who park between 1 and 7 o'clock are users who park most probably for a long time. Conversely, users who park between 13 and 19 o'clock are users who park most probably for a short time because the absolute values of coefficients are the biggest. In cluster 7, parking time changes drastically depending on time zones in parking because the absolute values of the coefficients are big.

Holiday. The coefficients about holidays are presented in Figure 4. Parking time becomes long on holidays because the coefficients are positive in all clusters. The influence of these coefficients is significant in cluster 3 and cluster 8.

Payment Method. The coefficients about payment method are presented in Figure 5. The values of the coefficients regarding credit card and point card are all positive. On the other hand, parking time regarding users using coupon tickets tends to become short except in cluster 8. Parking time tends to become long in cluster 8. Cluster 3 is a cluster where parking time becomes short by using a coupon ticket, and it is the most significant of all the clusters. The parking time of cluster 3 is average length. Influence of charge per hour and whether to set maximum charge in cluster 3 is the most significant of all the clusters. Thus, incentive to minimize parking charge is the highest, and parking time of users using a coupon ticket becomes short.

The influence of each covariate for cluster 3, cluster 7, and cluster 8 on the parking time is summarized from the above examination.

Cluster 3. The influence of setting price, holiday, and using coupon ticket is significant. Prudence is necessary for setting price because cluster 3 is a cluster where willingness to minimize parking charge is high.

Cluster 7. The influence of setting price is insignificant. The influence of time zones in parking is significant. Changing the utilization status is possible by setting different prices depending on time zones, for example, setting price lower except between 1 and 7 o'clock.

Cluster 8. The influence of charge per hour is not so significant. The influence of holidays is significant. Increasing revenue is possible by increasing charge per hour on holidays because parking time becomes longer on holidays.

2.4 Analysis of Change Points of Utilization Characteristics

The characteristics of parking utilization are changed by season or surrounding environment. Therefore, detecting change points of utilization characteristics is important for setting parking fee. In this study, we try detecting change points by assigning utilization rate and parking time per day to each typical one with cluster analysis, and considering changes in these as changes in clusters. The utilization rate and the parking time are calculated in each time zone dividing a day into 8 zones of 3 hours each from 0 o'clock, and cluster analysis is executed regarding changes in each utilization rate and parking time as 8 variables by data for 365 days on the 1,050 parking lots.

The cluster number of the utilization rate is 9. The cluster number of the parking time is 7. Cluster numbers are assigned in an ascending order of summing each value in cluster centers. In the next section, we consider whether influence of charge revision causes a change leading to increased revenue by changes in clusters, and the examination is used for setting price (e.g., Figure 8 in Section 3.2).

3 Setting Price Method and Examining Appropriateness Based on Data

3.1 Summary of Setting Price Method and Simulation Based on Data

In the section, we propose a pricing procedure. First, we present the results of cluster analysis in 2.2 and detect a parking lot in which it is necessary to adjust the charge. The parking clusters existing in great number are adapted for users' demands because users' demand for nearby parking is similar. A parking belonging to a parking cluster differing from parking clusters existing in great number are not adapted for users' demands. Thus, if a parking lot belonging to a parking cluster differing from a parking cluster of great number exists in some regions, the parking lot is regarded as needing to adjust the charge. Second, better pricing is considered by examining the difference between the parking lot needing to adjust its charge and surrounding parking lots. As for a parking cluster to which parking lots needing to adjust their charge belong, a parking lot increasing revenue by charge revision is detected by checking change of clusters regarding the utilization rate and parking time made in 2.4, and the results are used for setting price in the parking lot needing to adjust its charge. Finally, whether improved pricing is appropriate is considered by simulating how long the parking time changes by setting price method with survival analysis result in 2.3, and considering influence on revenue. A flowchart of the above pricing procedure is presented in Figure 6.

3.2 Executing Setting Price Based on Data

The parking lots around Hisaya-odori station are presented in Figure 7. The pink pins mean the lots belonging to parking cluster 2. The orange pins mean the lots belonging to parking cluster 6. The white pin means the lot belonging

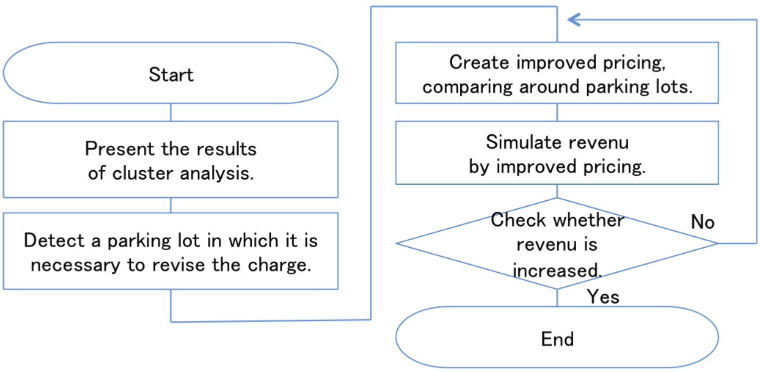

Fig. 6. Flowchart of pricing procedure

Fig. 7. Parking lots around Hisaya-odori station

Table 1. Pricing in parking lot 1 and parking lot 2

	charge type	day type	time zones	unit time	unit charge
1	unit time	all day	0 - 8	60 min	100 yen
	unit time	all day	8 - 24	30 min	200 yen
	maximum charge	all day	0 - 8	-	500 yen
2	unit time	all day	0 - 24	30 min	200 yen
	maximum charge	all day	22 - 8	-	500 yen
	maximum charge	weekday	8 - 22	-	1500 yen
	maximum charge	sat, sun, holidays	8 - 22	-	1000 yen

to parking cluster 8. The parking lot belonging to parking cluster 8 is indicated by a red arrow in Figure 7. Therefore, parking lot 1 needs to revise its charge. The improved pricing is considered by comparing the pricings or the changes of utilization rate clusters in the nearest parking lot 2 along the same road.

The pricings in parking lot 1 and parking lot 2 are presented in Table 1. Comparing pricing in parking lot 1 and parking lot 2, the charges per unit time are mostly the same; however, the settings of maximum charge are very different. Thus, setting no maximum charge in the parking lot 1 between 8 and 24 o'clock causes a large decrease in the utilization rate. Figure 8 shows the changes of the utilization rate clusters in the parking lot belonging to the same parking cluster 8 as lot 1. In Figure 8, there are many days assigned the utilization rate cluster 1 and 2 because the utilization rate is low before the 310th day; however, there are some days assigned the utilization rate cluster 7 and 8 because the utilization rate is increasing by setting a maximum charge near the 310th day. Setting maximum charge is effective for increasing parking time because influence of maximum charge is more significant than charge per hour in parking cluster 8. We propose pricing that sets 1,500 yen in maximum charge between 8 and 24 o'clock from the above consideration as better pricing in parking lot 1. This proposed pricing is called improved pricing 1. We consider appropriateness of improved pricing 1 in the next section.

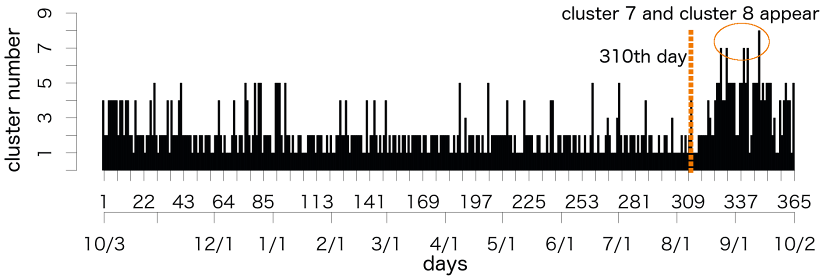

Fig. 8. Changes of utilization rate clusters in lot belonging to parking cluster 8

3.3 Evaluating Appropriateness of Pricing by Simulation

Simulation Method. The appropriateness of the pricing is evaluated by estimating parking time with survival analysis, and calculating revenue.

1. An accounting data in a target parking lot is retrieved.
2. The covariates except charge per hour and whether to set maximum charge from the retrieved data are decided.
3. The values of charge per hour and whether to set maximum charge is decided based on considering pricing, and a survival function $S(x|Z)$ is constructed.
4. A value between 0 and 1 is generated randomly, and the value is defined as p.
5. x is incremented by 10 from 10 until $S(x|Z) < p$ is corrected, then parking charge is calculated and stored regarding x satisfying the conditions as parking time.
6. Procedures 1 to 5 are repeated until all target accounting data are retrieved.

The value of x meaning the parking time is incremented by 10 from 10 because a minimum value of unit time in charge per unit time is 10 minutes. However, maximum parking time is 4,320 minutes, equal to 3 days, in order not to generate too long a parking time.

Simulation Result. Figure 9 shows a distribution of the parking time estimated by not changing pricing and a distribution of the parking time estimated by the improved pricing 1. The parking time is estimated every 10 minutes in the simulation; however, the users are counted every 30 minutes from 0 minutes in Figure 9 because the minimum value of unit time in charge per unit time is 30 minutes. The horizontal axis indicates the parking time, and is shown until 900 minutes when the number of users approaches 0. Table 2 and Table 3 show average parking time, the number of overflowing users, and total revenue in actual parking time, the parking time estimated by not changing pricing, and the parking time estimated by changing pricing. The number of overflowing users indicates the number of users who can't park because the parking lot is full. The longer the parking time becomes, the more the number of overflowing users becomes. The number of overflowing users needs to be considered because these become the factor decreasing the revenue in parking lots where maximum charge is set. Thus, two values of whether to consider the number of overflowing users are presented regarding the total revenue in Table 3. The value of not considering the number of overflowing users is indicated by A; in contrast, B indicates the value of considering the number of overflowing users. In the case of not considering the number of overflowing users, a total revenue is calculated by all users even if the parking lot is full. In the case of considering the number of overflowing users, a total revenue is calculated by the users who can park.

The appropriateness of the improved pricing 1 is considered. Comparing the distributions of parking time in Figure 9, the parking time becomes long by setting the maximum charge as the improved pricing. An increase in the utilization rate is expected by lengthening the parking time. However, if pricing becomes excessively inexpensive in order to increase the utilization rate, parking spaces tend to fill up by lengthening the parking time, and then the revenue decreases because the number of overflowing users increases. Therefore, pricing that doesn't make the turnover rate too low is needed. Comparing the data estimated by not changing pricing and by changing pricing, the average parking

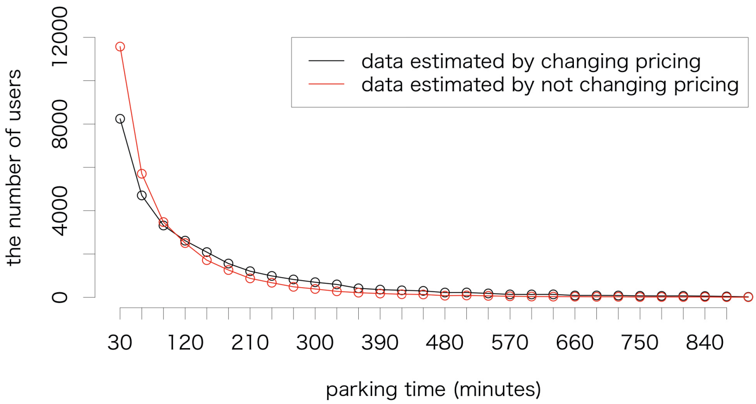

Fig. 9. Distribution of parking time estimated by not changing pricing and by changing pricing

Table 2. Comparison 1 between actual parking time and estimated parking time

	the average parking time (minutes)	the number of overflowing users
the actual data	93.35	-
the data estimated by not changing pricing	85.33	2,437
the data estimated by changing pricing	150.20	6,957

Table 3. Comparison 2 between actual parking time and estimated parking time

	the total revenue (yen)	
	A	B
the actual data	17,246,400	-
the data estimated by not changing pricing	17,224,500	15,886,500
the data estimated by changing pricing	23,938,400	18,705,300

Table 4. Simulation result in case of setting excessively inexpensive maximum charge

the total revenue (yen)	
A	B
20,031,600	15,661,900

time became longer for approximately 65 minutes in the data estimated by not changing pricing. The total revenue not considering the number of overflowing users (the total revenue A) increases approximately 6.7 million yen, and the total revenue considering the number of overflowing users (the total revenue B) also increases approximately 2.8 million yen.

Table 4 shows the simulation result by pricing that sets 1,000 yen in maximum charge of parking lot 1 between 8 and 24 o'clock in Table 1. The total revenue A increases comparing the data estimated by not changing pricing in Table 3; however, the total revenue B decreases by approximately 0.2 million yen. In addition, the total revenue A and the total revenue B decrease comparing the data estimated by changing pricing by changing pricing. Therefore, 1,000 yen is excessively inexpensive regarding maximum charge. The improved pricing setting 1,500 yen as maximum charge is an appropriate pricing for increasing revenue.

4 Related Work

A lot of data is generated and accumulated from various sources as information technology is developed. These data are called Big Data. Big Data are high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making [6].

Revenue Management has been used as a important technique in some fields of airline [7], financial service [8], etc., and Big Data enhance a revenue management system to manage in more detail. For example, Hashimoto et al. [10] suggest an effective parking reservation system, which assigns parking spaces and decides parking charge by auction system. As an evaluation of the proposed reservation system, a simulation considering price elasticity is executed. For this simulation, clusters are composed by cluster analysis from actual data for parking utilization, and parking time models are created for each composed cluster by survival analysis.

Furthermore, a system providing better service is available by taking advantage of not only system data such as just availability of flight but also individual customer behavior data [9]. For example, Ishigaki et al. [11] suggest that customer behavior can be understood by constructing a Bayesian network based on categories after extracting potential customers and product categories from questionnaire data and ID-POS data.

In this study, parking charges adapted for the utilization characteristics are set by extracting the utilization characteristic from large-scale data about parking accounting, and constructing a hazard-based duration model. As a concrete method, we classify characteristics of parking utilization into some clusters by cluster analysis, then we create parking time models for each cluster by survival analysis such as [10]. Unlike [10], we execute analysis in more detail by actual pricing change data in addition to parking accounting data. We propose a more appropriate pricing from the result of cluster analysis and survival analysis, then we consider appropriateness of the parking fee by simulating total revenue.

5 Conclusion

In this study, a pricing procedure based on data analysis was proposed based on parking accounting data on 1,050 parking places around Nagoya City over recent one year. First, the parking lots were classified according to utilization characteristics by cluster analysis. Second, Influence on parking time, which occurred by a difference of covariates such as the pricing and the time zones when users parked, was analyzed for each cluster by a hazard-based duration model. Third, the utilization characteristics and the parking time per day were classified into clusters by another cluster analysis. Finally, the appropriateness of pricing is evaluated by simulating parking time. The proposed pricing procedure enables a non-expert to set changes adapted for the characteristics of the parking utilization.

Acknowledgement. The parking accounting data was provided by Meitetsu Kyosho Corp. This work is partially supported by the Funding Program for Next Generation World-Leading Researchers (NEXT Program) of the Japan Cabinet Office and JSPS KAKENHI Grant Number 25870320. We would like to express our gratitude.

References

1. Xu, R., Wunsch, D.C.: Clustering. Wiley (2009)
2. Hartigan, J.A., Wong, M.A.: A K-means Clustering Algorithm. Applied Statistics 28, 100–108 (1979)
3. Klein, J.P., Melvin, L.M.: Survival Analysis: Techniques for Censored and Truncated Data. Springer (2005)
4. Weibull, W.: A Statistical Distribution Function of Wide Applicability. Journal of Applied Mechanics 18, 293–297 (1951)

5. Kawaura, K.: The Study on Actual State of Parking in Service Area of Expressway (Part 2): Distribution of Parking Durations. Seisan Kenkyu 20(7), 362–364 (1968)
6. Iafrate, F.: A Journey from Big Data to Smart Data. Digital Enterprise Design & Management 261, 25–33 (2014)
7. Christopher, P.W., Harry, G., Robert, A.S.: Dynamic Revenue Management in Airline Alliances 44(1), 15–37 (2010)
8. Emmanuel, D.(M.)H., Suresh, K.N., Michael, P.: Production and Operations Management 19(6), 633–664 (2010)
9. Cenk, K., Kivilcim, D.: An empirical investigation of consumers' willingness-to-pay and the demand function: The cumulative effect of individual differences in anchored willingness-to-pay responses 25(2), 139–152 (2014)
10. Hashimoto, S., Kanamori, R., Ito, T.: Auction-based Parking Reservation System with Electricity Trading. In: IEEE International Conference on Business Informatics (CBI), pp. 33–40 (2013)
11. Ishigaki, T., Takenaka, T., Motomura, Y.: Customer Behavior Prediction System by Large Scale Data Fusion in a Retail Service. The Japanese Society for Artificial Intelligence Journal 26(6), 670–681 (2011)