Application of Simulation Method and Regression Analysis to Optimize Car Operations in Carsharing Services: A Case Study in South Korea

Jongtae Rhee, Ganjar Alfian, Byungun Yoon*
Dongguk University-Seoul

Abstract

A carsharing service is a form of public transportation that enables a group of people to share vehicles based at certain stations by making reservations in advance. One of the common problems of carsharing is that companies can have difficulty optimizing the number of vehicles in operation. This paper reports on investigations of the relationship between the number of cars and the number of reservations per day with either the acceptance ratio or utilization ratio based on the commercially-operational dataset of a carsharing company in Korea. A discrete event simulation is run to analyze a round-trip service for every possible number of cars and number of reservations with the output acceptance ratio and utilization ratio. The simulation data revealed that increasing the number of reservations with respect to a certain number of cars will decrease the acceptance ratio, thus increasing the percentage of the utilization ratio. Based on the simulation data results, a rational regression model can achieve high precision when predicting the acceptance ratio or the utilization ratio compared to other prediction algorithms such as the Multi-Layer Perceptron (MLP) and the Radial Basis Function (RBF) models. K-means clustering was
used to understand the pattern and provide additional policies for carsharing companies. Consequently, opening a carsharing business is very promising in terms of profit, escalating the level of customer satisfaction. In addition, a small reduction in the utilization ratio by operators will create a large increase in the acceptance ratio.

Introduction

As the world population grows, private vehicles are becoming more attractive, leading to high energy consumption and high vehicle emission levels. Carsharing is one of the transportation strategies that can reduce personal transportation usage and its negative impacts. Because of the worldwide environmental benefits involved, carsharing evolved out of the economic motivations of individuals who could not afford to purchase a vehicle into a mainstream, worldwide transportation system. In recent carsharing systems, customers can access the portal of a carsharing company and easily make a reservation via an Internet connection or by phone. The information, including traveled distances and rent duration, is recorded and charged as to the customer’s bill. An intelligent transportation system can play an important role in making a carsharing system user-friendly, easy to manage, and efficient.

Because of these benefits, carsharing as an alternative transportation paradigm has become increasingly popular in many countries (Barth and Todd 1999). Previous research has demonstrated that the benefits of carsharing include reducing costs and the negative impacts of private vehicle ownership and the environmental impacts of auto usage (e.g., congestion, energy consumption, vehicle emissions, and inefficient land use). In North America, the impact of carsharing includes the reduction of emissions as a result of less driving and a 27 percent reduction in the average number of observed vehicle kilometers traveled per year (Martin and Shaheen 2011). According to another review, an additional benefit is cost savings, which was reported to be the main motivation for new memberships from 2006 to 2010. In addition, there has been a change in carsharing activity, as can be seen from the number of worldwide carsharing memberships. In 2006, Europe was the epicenter, but it shifted to North America in late 2010. Stabilized growth in neighborhood residential carsharing and rapid growth in the business and university markets in North America from 2006–2011 was the key trigger. Importantly, there was a worldwide increase in the number of carsharing memberships and in total vehicles and member-vehicle ratios from 2006–2010. As carsharing increasingly becomes a mainstream transportation mode, it is expected that it will be further
integrated into metropolitan transportation, land use strategies, and multimodal nodes (Shaheen and Cohen 2013).

Up-to-date carsharing systems enable a car to be driven among multiple stations (one-way service), whereas traditional service (round-trip/two-way) allows users to use a car and return it to the same station only. Although one-way service can provide convenience for customers, the cars from each station become disproportionally distributed. Thus, a strategy of vehicle relocation is necessary to elevate the satisfactory level of users. A carsharing system must be efficient, user-friendly, easy to manage, and advantageous to both companies and customers (Barth et al. 2001).

Studies concerning data mining have been intensively conducted in carsharing-related research areas. In particular, the forecasting technique is used to predict the net flow of vehicles in a three-hour period by using neural networks and support vector machines (SVM) (Cheu et al. 2006), and the results show that multilayer perceptron has slightly better accuracy compared to SVM. In another case, such as the one-way type, it is difficult to maintain the distribution balance of parked vehicles among stations. A method for the optimization of vehicle assignment is used according to the distribution balance of parked vehicles; thus, it is possible to maintain distribution balance of parked vehicles and keep the convenience of the carsharing system (Uesugi et al. 2007).

In regard to car optimization, one study shows an international comparison regarding carsharing services (Shaheen and Cohen 2007). The paper shows that the member-vehicle ratio is an important key factor that characterizes worldwide carsharing operations. The comparison demonstrates that the member-vehicle ratio based on the survey of each country is different; Asia, Australia, Europe, and North America are 26:1, 17:1, 28:1 and 40:1, respectively. The estimation for the average national ratios are approximately 20:1 and are lower in new markets where carsharing companies must first position their vehicles to gain membership. However, in other research (Morency et al. 2007; Habib et al. 2012; Costain et al. 2012), studies about user behavior in carsharing transaction data sets show interesting results. The data are from Communauto, Inc., a carsharing company in Montreal from January –December 2004. The result reveals that there is variability in the number of transactions and distance traveled by each customer. Another study (Costain et al. 2012) found that increasing the home-to-parking-lot distance reduces trip duration. Thus, it is important to evaluate the member-vehicle ratio with respect
to other parameters such as variability of the number of transactions, traveled distance, and traveled time by the customer.

Advanced simulations in carsharing have focused on developing a relocation model to evaluate one-way car availability (Kek et al. 2009). In addition, a forecasting model for relocation has been suggested to optimize the results of relocation and predict efficient routes (Cheu et al. 2006; Wang et al. 2010; Karbassi and Barth 2003; Correia and Antunes 2012). However, to implement those models, it is important for carsharing companies to decide first on the initial vehicles before focusing on relocation models. Because it is difficult to predict the initial number of cars needed without losing customer interest and company profits, this paper aims to demonstrate that a simulation model must be developed first to evaluate the acceptance ratio and utilization ratio for traditional, round-trip services based on traveling frequency, number of vehicles, and Vehicle Hours Traveled (VHT) and Vehicle Kilometers Traveled (VKT) patterns.

Two output parameters were used in this paper. The first was the acceptance ratio, which can be simply explained as successful reservations over total reservations made by customers; this parameter can be expected to reveal general customer satisfaction. The second parameter is the utilization ratio, which is the percentage of total actual driving hours of rented cars over the total possible driving hours of cars, which elucidates company profits. Later, the simulation data results are analyzed using regression and other forecasting techniques to generate a prediction model. This paper aims to focus on how to develop a model that can be used to optimize the number of cars needed with respect to a certain number reservations per day, time patterns, and thresholds of either the acceptance ratio or the utilization ratio.

Section 2 of this paper provides an overview of the results of the literature review. Section 3 describes the methodology of the simulation and algorithm analyses. Results and a discussion of the proposed model in are presented in section 4, and limitations and future research of this paper are discussed in section 5.

**Background**

**Carsharing Service**

Carsharing services can be placed under shared-use vehicle system models based on the similarities in types and models of service. A shared-use vehicle system consists of a vehicle that is used by several groups of people throughout the day. To create a formal structure, previous research developed a classification system
for evaluating various models. Generally, the classification of shared-use vehicle systems consists of neighborhood carsharing, station cars, multi-nodal shared-use, and hybrid models. Carsharing—or what is traditionally referred to as neighborhood carsharing—began in Europe and placed a network of vehicles in strategic parking areas (mostly in residential neighborhoods) located throughout denser cities. The second type is the station cars model, in which typical car stations are placed at major rail stations along a commuting corridor, thus enhancing transit connectivity and providing a convenient way to access a user’s home or work from the public transit station. Another model is the multi-nodal shared-use model, which allows customers travel from one center to another, as in, at resorts, recreational areas, and corporate university campuses. The trips are more likely to be one-way service. The hybrid model or the future of the shared-use vehicle system has the characteristics of many of these systems. In the hybrid system, the vehicles used may be linked to transit (referred to as station cars) and left at transit stations and could also be used for several other purposes such as daily-use trips of both a business and a residential nature (Barth et al. 2002).

The history of successful experiences of carsharing began in Europe in the mid-1980s, and carsharing organizations in Europe are now firmly established and on steep growth trajectories. Meanwhile, the North American experience with carsharing is far more limited. One of the formal carsharing demonstrations in the United States was Mobility Enterprise, operated as a Purdue University research program from 1983 to 1986. As carsharing emerges, researchers have concluded that operators are more likely to be economically successful when they provide a dense network and a variety of vehicles; serve a diverse mix of users; create joint-marketing partnerships; design a simple, flexible rate system; and provide easy emergency access to taxis and long-term car rentals (Shaheen et al. 1998).

Carsharing services represent an intermediate service that bridges public transportation and private vehicle ownership to reduce the number of cars, provide cost savings, and reduce parking demand, among other benefits. To clarify, carsharing was first implemented in Europe but has gained popularity in North American cities (Cervero and Tsai 2004; Zhou and Kockelman 2011) and Asia, including Singapore and Japan. Basically, members subscribe to a carsharing company and are able to use cars by making reservations in advance. The vehicle is picked up at the start of the trip and returned to the original station at the end of the trip (two-way or round-trip). Members pay a fee each time they use a vehicle, which covers the cost of vehicle use, insurance, maintenance, and fuel. An example, a carsharing
study in the U.S. was a pilot program called CarLink, which categorized users as home-based users, work-based commuters, and work-based day users. During the field test, each group paid a different fee according to the duration of usage. All user fees included fuel, insurance, and maintenance costs (Shaheen and Wright 2001). A carsharing company generally offers different service options based on these categories. The service options generate different benefits and satisfy each member’s requirements.

South Korea is a densely populated country in East Asia with about 48 million inhabitants. The process of rapid industrialization over the last few decades has transformed South Korea into an economic hub of Asia. One of the factors that has always played an important role in influencing the formation of urban societies is transportation. Advances in transportation have made possible changes in our way of living and the way in which societies are organized, and they, therefore, have a great influence in the development of civilizations. The big challenge for the implementation of carsharing services in South Korea, especially in Seoul, is public transportation, because most Koreans use public transportation. Information released by Seoul Metro about the transport mode share in Seoul reveals that the subway, city buses, and passenger cars have market shares of 34.7, 27.6, and 26.3 percent, respectively, and the rest comprises taxis, with a total number of daily passengers of about 4.04 million people (Seoul Metro 2011). To address real situations, one research project and paper has been published about carsharing as one of the product service systems that defined a service blueprint for carsharing in Korea (Yoon et al. 2012). The research revealed that a new carsharing service model is applicable to South Korea because it would foster sustainable development while reducing traffic problems and air pollution. The Korean carsharing service model interfaces with a public transport system for increasing mobility. It serves people who are not sufficiently mobile. Therefore, car-sharing stations need to be installed at transport interchanges and in areas with low access to public transportation.

The first pilot program of carsharing in South Korea began by offering round-trip service from November 2011 to June 2012 from and to the campus of Dongguk University. The pilot program was implemented to minimize the negative impact of the first carsharing market in South Korea. In addition, similar to the CarLink pilot program in the U.S., it was important to run a pilot program first before progressing to a larger market area. Previous research has revealed that the success of pilot programs will lead to the success of continuous programs, and this is main reason that pilot programs must be developed first in South Korea. In the campus
pilot program, the operator offered off-road parking in the general campus area and parking lots in residential areas nearby. The program attracted approximately 500 total customers, which consisted of staff members, students, and residents near campus.

Upon first implementing the pilot program, the proportion of residential users was small because of the limited number of parking lots near the campus and the limitation of service promotions, while the biggest users were staff members and students, respectively. The type of car that was offered was a small, domestic type, with a total of 50 cars. The member-vehicle ratio upon start-up was about 10:1 in order to gain membership. During the field test, each member paid a fee based on the duration of car use (a combination of distance and time), and all user fees included fuel and maintenance costs. The Dongguk campus program combined short-term rental vehicles with communication and reservation technologies (i.e., an automated reservation system by website and phone, GPS for vehicle tracking, and smartcards for vehicle access) to facilitate easy access. In addition to vehicle support services, staff supported the program with cleaning and maintenance and by maintaining the customer service via phone.

Korea Carsharing is the first carsharing company in South Korea that successfully transferred a pilot program to a larger area once the initial pilot program was completed. The program was successful in upgrading the quality of service by identifying the need for increasing public area parking lots and identifying hardware and software problems during the pilot program. Currently, the number of stations is increasing as an improvement in carsharing services in South Korea, and the primary focus of services is on residential, business, and public venues, as seen in the increase in residential customers, with a total number of 1,000 members. As the number of customer has increased, the user type has changed from mostly on-campus staff and students to business users and residential users. The emergence of carsharing services in South Korea involves the government and automobile manufacturers who are quite active in helping to sponsor programs. The increase in memberships required the development of integrated carsharing technologies, such as upgrading the system for coordinate vehicle tracking, and a reservation system (WeShareCar 2013). Reservations by smartphone now enable members to make reservations, and, thus, technology is able to enhance service capabilities.
Methodology

Simulation Model

A simulation approach is a process to design and conduct experiments for the purpose of understanding system behavior or evaluating various strategies for the operation of the system. A good solution from the results of the simulation is recommended for implementing a new system. In a discrete-event simulation, the operation of a system is represented as a chronological sequence of events. Each event occurs at an instance in time and marks a change of state in the system. The structural components of a discrete event simulation include entities, activities and events, global variable, random number generators, and calendar. The idea of a discrete event simulation is that the clock jumps to the next event as the simulation proceeds (Ingalls 2001).

A simulation approach is used for testing the relocation techniques, namely shortest time and inventory balancing (Kek et al. 2006). Shortest time relocation involves a process to move a car from a neighboring station in the shortest possible time. Inventory balancing relocation is an approach to moving a car to a station that has a shortage of cars from another station that has an oversupply of cars. Another simulation study proposed a static relocation to move a car immediately after a customer requests one (Barth and Todd 1999). In particular, a forecasting model has been implemented to predict the net flow of vehicles in a three-hour period by using neural networks and support vector machines (Cheu et al. 2006). The results of the simulation experiment demonstrate that all of the aforementioned techniques have the potential to improve carsharing services in a realistic situation.

In general, the simulation implementation will greatly assist a carsharing company in evaluating its policies before implementing a service in a realistic situation.

In this paper, a simulation model that reflects a reservation algorithm is presented to evaluate round-trip service only, which allows customers to use a car and return it to the same station. A comparison could not be presented in this paper of the acceptance and utilization ratios for other services such as one-way and open-ended services, which offer flexibility to users without identifying the ending time for a reservation (Schwieger and Wagner 2003). The simulation tool for a carsharing reservation system has been designed to be as realistic as possible. The reservation acceptance and car utilization ratios are presented to evaluate round trips for every possible number of cars and number of reservations. The car utilization ratio is very important for a carsharing company to optimize operation time, which can improve profits and reduce operational car costs. The reservation acceptance ratio
is important to customers, and it can provide a benchmark for revealing customer satisfaction. Thus, the number of initial cars must be chosen carefully with respect to the thresholds of the acceptance and utilization ratios.

First, artificial data are generated, using a trip generator based on customer travel demand distribution (VKT, VHT, time of day, day of week). Second, the artificial data are simulated with a simulation tool to evaluate round-trip service for every combination of the number of cars and the number of reservations. Finally, the simulation results are presented and are analyzed with prediction techniques to define the proposed model.

**Regression Analysis**

**Multiple Linear Regression**

Multiple linear regression attempts to model the relationship between the dependent variable and one or more independent variables, by fitting a linear equation to the observed data. The goal of regression analysis is to model the expected value of a dependent variable ŷ in terms of the value of an independent variable (or vector of independent variables) \( x \). In simple linear regression, the model the dependent variable (ŷ) is given by:

\[
\hat{y} = \beta_0 + \sum_{i=1}^{l} \beta_i x_i + \epsilon
\]

where \( x_i \) \( (i = 1, \ldots, l) \) are the explanatory independent variables, \( \beta_i \) \( (i = 1, \ldots, l) \) are the regression coefficients, and \( \epsilon \) is the error associated with the regression and assumed to be normally distributed with both the expectation value of zero and constant variance (Agirre-Basurko et al. 2006). Multiple regression has been implemented in many areas, such as building areas (Catalina et al. 2008) and brain research areas (Klein et al. 2005), and has shown good prediction models.

**Polynomial Regression Analysis**

Polynomial regression is nonlinear, which describes the relationship between any set of independent and dependent variables. The polynomial regression model, which contains more than two predictor variables, is called MPR (Multiple Polynomial Regression) (Zaw and Thinn 2009). Polynomial regression models are usually fit using the method of least squares. The least-squares method minimizes the variance of the unbiased estimators regarding the coefficients, under the conditions of the Gauss–Markov theorem. In general, we can model the expected value of \( y \) as an \( n \)-th order polynomial, yielding the general polynomial regression model.
A rational function is a function $f$ that is a quotient of two polynomials, that is, $f(x) = \frac{p(x)}{q(x)}$ where $p(x)$ and $q(x)$ are polynomials and where $q(x)$ is not the zero polynomial. The domain of $f$ consists of all inputs $x$ for which $q(x) \neq 0$. Typically, the rational model is a class of model description, which is nonlinear in the parameters. The following is a brief review of the work in the identification of nonlinear rational models.

\[
\hat{y} = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0
\]  

(2)

**Rational Function**

Multilayer Perceptron Neural Network

Multilayer perceptron (MLP) has to be configured such that the application of a set of inputs produces (either “direct” or via a relaxation process) the desired set of outputs. The ANN learning algorithm used here is back propagation. Various methods to set the strengths of the connections exist. One way is to set the weights explicitly, using a priori knowledge. Another way is to “train” the neural network by feeding it teaching patterns and letting it change its weights according to some learning rule. During this process, inputs are fed forward from the input layer and through the hidden layers, and, ultimately, the network provides its output, which for an untrained network is different from the known target output. The training process consists of estimating weights, which minimize deviations between network outputs and actual data. The deviations are then propagated backwards through the network and weights are adjusted to reduce error. Here, three layers were used in the ANN: input, hidden, and output layers. The detail explanations about MLP are described elsewhere and are not repeated here (Larose 2005; Krose and Van Der Smagt 1996).

\[
\hat{y} = \frac{a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0}{b_m x^m + b_{m-1} x^{m-1} + \cdots + b_2 x^2 + b_1 x + b_0}
\]  

(3)

with $n$ denoting a non-negative integer that defines the degree of the numerator and $m$ is a non-negative integer that defines the degree of the denominator. For fitting rational function models, the constant term in the denominator is usually set to 1. Rational functions are typically identified by the degrees of the numerator and denominator. For example, a quadratic for the numerator and a cubic for the denominator is identified as a quadratic/cubic rational function (Dette et al. 1999; Zhu 2005).
Radial Basis Function Network
The Radial Basis Function (RBF) network emerged as a variant of the artificial neural network in the late 1980s. However, their roots are entrenched in much older pattern recognition techniques as, for example, potential functions, clustering, functional approximation, spline interpolation, and mixture models. The construction of an RBF network in its most basic form involves three entirely different layers. The input layer is made up of source nodes (sensory units). The second layer is a hidden layer realizing the radial basis function with high enough dimensions, which serves a different purpose from that in a multilayer perceptron. The output layer supplies the response of the network to the activation patterns applied to the input layer. The transformation from the input space to the hidden unit space is nonlinear, whereas the transformation from the hidden unit space to the output space is linear. The detail explanations about the RBF network are described elsewhere and are not repeated here (Zhao et al. 2010; Li et al. 2010).

Data Collection and Analyses
The best way to investigate the impact of carsharing in detail is through targeted sample data collection. Therefore, this paper presents the results of an investigation of carsharing user behavior through the examination of the dataset from a carsharing service in South Korea. Although the choice of information used in this study is limited by data availability, sufficient information is available to investigate key issues of interest. The aforementioned Dongguk pilot program was successfully implemented in the campus area and has now become one of the stations for Korea Carsharing. For the simulation, the data distribution must be set to obtain good results, and, thus, the input parameters were collected based on the carsharing pilot program that operated from November 2011 to June 2012. More details on the input parameters are shown in Table 1.

Based on the Korean Carsharing pilot program dataset, the traveling time of customers is between 30 minutes and 6 hours. Customers traveling for less than 30 minutes prefer to use a taxi service; for 6 hours or more hours of travel, customers prefer to rent a car. VHT and its distribution can be seen in Figure 1(a). The dataset reveals that the average VHT by customers is 2–3 hours. The dataset provides detail about the trip behavior of the carsharing member, and it is interesting to note that trips are made by carsharing members throughout the whole day. The information in the dataset indicates that time of day distribution is grouped into three clusters: morning, afternoon, and night. It reveals that the majority of trips are made at night, beginning immediately after the end of Korean work time at around 6 PM.
and lasting until midnight, with the average time being 9 PM, as seen in Figure 1(c).

VKT illustrates that the majority of trips made by carsharing members are short-distance trips of less than 100 km, and the average is 20–30 km. Details are shown in Figure 1(b).

**Table 1. Input Parameters**

<table>
<thead>
<tr>
<th>Input Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total operated cars</td>
<td>Start with 5, increase until 100</td>
</tr>
<tr>
<td>Total stations</td>
<td>Automatically generated, depends on number of cars (1 station has about 5 cars)</td>
</tr>
<tr>
<td>Operation time</td>
<td>1 week, 24 hours per day</td>
</tr>
<tr>
<td>Service</td>
<td>Round-trip</td>
</tr>
<tr>
<td>Reservation per day</td>
<td>Start with 5, increase until 2,000</td>
</tr>
<tr>
<td>VHT (Vehicle Hours Traveled)</td>
<td>Between 30 minutes to 6 hours, with distribution</td>
</tr>
<tr>
<td>VKT (Vehicle Kilometers Traveled)</td>
<td>Between 5–120 km, with distribution</td>
</tr>
<tr>
<td>Time of day</td>
<td>Distribution of customer reservations in 24 hours</td>
</tr>
</tbody>
</table>

**Figure 1a. Customer travel patterns**
In addition, to understand the trip behavior of carsharing members, Figure 1d presents the week distribution of the trips. The result shows that the major peak occurs on the weekends, which mean customers prefer to travel during the weekend, starting from Friday night around 6 PM until Sunday midnight. In terms of the
day of the week, the percentage of trips is lowest at the beginning of the week and increases as the week progresses.

A trip generator was developed to transform the time of day, VHT, VKT and the week-long distribution into artificial reservation data based on these distributions (see Figure 1) for every number of reservations in a week. In this paper, we focused only on one dataset for investigating carsharing behavior; therefore, the use of similar datasets with different distributions from other companies would contribute to an increase in understanding but would generate different simulation data results. However, there is a similarity in the distribution of trips between a Toronto case study (Costain et al. 2012) and our dataset regarding trip length distribution and day of the week. In the Toronto case study, more than 60 percent of trips were less than 40 km, whereas in our dataset, trip length was mostly 10–40 km. This indicates that carsharing contributes to an increase in short-distance urban auto trips in Seoul and Toronto, and this is also true in other cities around the world (Morency et al. 2007; Zhou and Kockelman 2011). In addition, there is a similarity in day of the week distributions shown in the Toronto study. The percentage of trips is lowest at the beginning of the week and increases as the week continues, which is similar to our dataset in which major peaks occur on the weekends. Meanwhile, the time of the day shows a different pattern: in our dataset, peak travel occurred at night, whereas in the Toronto case study, the majority of trips were made between 9–11 AM, which is immediately after the morning peak period. These similarities illustrate that there are general patterns of customer usage among carsharing

Figure 1d. Customer travel patterns
operators in different parts of the world. Therefore, the simulation model that has been developed for this project can be used by other operators as a general benchmark of the relationship between utilization and acceptance ratios of operators.

In addition, the artificial reservation data are designed to be similar to a reservation table in a real carsharing system. Each record in a carsharing system database consists of a single reservation with a member identification number, vehicle identification number, transaction number, and the time and date (beginning and end of the reservation). This transaction table can be linked to other tables such as a member table, a car table (year, model, parking ID), and a parking lot table (capacity, location) (Morency et al. 2007). The artificial reservation table in this study consists of several columns (transaction number, member identification number, service identification number, vehicle identification number, beginning reservation, end of reservation, station to station, date of reservation, and calling time), which can be seen in Figure 2.
<table>
<thead>
<tr>
<th>id_reservation</th>
<th>id_customer</th>
<th>id_service</th>
<th>numb_car</th>
<th>start_time</th>
<th>end_time</th>
<th>from_location</th>
<th>to_location</th>
<th>id_day</th>
<th>call_time</th>
<th>id_service_detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13192</td>
<td>15</td>
<td></td>
<td>04:39:48</td>
<td>06:39:48</td>
<td>140</td>
<td></td>
<td>140</td>
<td>01:04:47</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>13239</td>
<td>15</td>
<td></td>
<td>18:26:45</td>
<td>20:26:45</td>
<td>140</td>
<td></td>
<td>140</td>
<td>18:06:19</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>12517</td>
<td>15</td>
<td></td>
<td>14:35:59</td>
<td>16:35:59</td>
<td>140</td>
<td></td>
<td>140</td>
<td>04:49:59</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>12941</td>
<td>15</td>
<td></td>
<td>17:57:01</td>
<td>19:57:01</td>
<td>140</td>
<td></td>
<td>140</td>
<td>05:48:41</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>12652</td>
<td>15</td>
<td></td>
<td>19:52:37</td>
<td>21:52:37</td>
<td>140</td>
<td></td>
<td>140</td>
<td>09:47:18</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>12727</td>
<td>15</td>
<td></td>
<td>07:59:51</td>
<td>09:59:51</td>
<td>140</td>
<td></td>
<td>140</td>
<td>05:34:34</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>12256</td>
<td>15</td>
<td></td>
<td>15:59:37</td>
<td>17:59:37</td>
<td>140</td>
<td></td>
<td>140</td>
<td>11:59:05</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>11551</td>
<td>15</td>
<td></td>
<td>19:25:45</td>
<td>21:25:45</td>
<td>140</td>
<td></td>
<td>140</td>
<td>03:18:25</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>12638</td>
<td>15</td>
<td></td>
<td>18:38:40</td>
<td>20:38:40</td>
<td>140</td>
<td></td>
<td>140</td>
<td>05:16:13</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>12350</td>
<td>15</td>
<td></td>
<td>02:41:31</td>
<td>04:41:31</td>
<td>140</td>
<td></td>
<td>140</td>
<td>01:28:30</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>12959</td>
<td>15</td>
<td></td>
<td>08:59:54</td>
<td>10:59:54</td>
<td>140</td>
<td></td>
<td>140</td>
<td>06:54:53</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>12140</td>
<td>15</td>
<td></td>
<td>11:51:50</td>
<td>13:51:50</td>
<td>140</td>
<td></td>
<td>140</td>
<td>09:45:20</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>12406</td>
<td>15</td>
<td></td>
<td>17:45:24</td>
<td>19:45:24</td>
<td>140</td>
<td></td>
<td>140</td>
<td>15:33:07</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>12882</td>
<td>15</td>
<td></td>
<td>17:50:59</td>
<td>19:50:59</td>
<td>140</td>
<td></td>
<td>140</td>
<td>04:50:58</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. Artificial reservation table
Selection of Output Variables for Simulation Model

Two outputs (car utilization ratio and reservation acceptance ratio) are defined in the simulation to evaluate the performance of round-trip service.

Company Car Utilization Ratio

The car utilization ratio is the percentage of total actual driving hours of rented cars over the total possible driving hours of cars per day. In addition, in this discrete-event simulation, the data are generated and simulated for one week. Since a car-sharing company normally wants to optimize the number of operation cars, the company needs to ensure that all cars can be rented (fully operated) to increase the profit and reduce operational car cost. Thus, the formula for calculating the utilization ratio in this simulation tool is:

\[
\text{Car Utilization Ratio} = \frac{\text{vehicle-hours of cars used}}{\text{available vehicle-hours of entire fleet}}
\]

Reservation Acceptance Ratio

The Car Acceptance Ratio is information on how many reservations are accepted over the total number of reservations. Accepted reservations mean that when a customer makes a reservation, the carsharing reservation system will check whether the customer can acquire an available car or not. If they receive an available car, and there is an empty space at a destination station, the reservation is accepted or, otherwise, rejected. Since the system does not suggest a customer to delay his/her reservation to get the other car, the customer is expected to find another reservation that has no conflict with others. In this paper, all reservations are assumed to be done by customers in a problem-free scenario, such as there is no conflict of destination stations when the cars are parked. This reservation acceptance ratio can provide the ideal situation to reveal customer satisfaction. The formula for calculating the car acceptance ratio in this simulation tool is:

\[
\text{Acceptance ratio} = \frac{\text{complete reservations}}{\text{total reservations}}
\]

Experimental Scenarios

A reservation algorithm that can handle reservations was implemented in the simulation tool and its data stored in the database. In this paper, the discrete-event simulation is implemented on the basis of structural components as shown in Table 2.
Table 2. Structural Component of Discrete Event Simulation

<table>
<thead>
<tr>
<th>Component</th>
<th>Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entities</td>
<td>List of customers who intend to use car by making reservation in advance.</td>
</tr>
<tr>
<td>Events</td>
<td>Customer makes phone call for reservation, customer picks up car, drives car,</td>
</tr>
<tr>
<td></td>
<td>then returns car to destination station.</td>
</tr>
<tr>
<td>Random number generator</td>
<td>Generates number of reservations to reservation table based on input distribution. Random number generator generates arbitrary names of customers who will make reservation and decide starting time, VKT, VHT by its distribution.</td>
</tr>
<tr>
<td>Queue (wait for an unspecified period)</td>
<td>Time for customer to pick up car and return it must be explicitly decided for queue activity.</td>
</tr>
<tr>
<td>Logic activity</td>
<td>Decision whether customer gets free car or not, depending on availability of car.</td>
</tr>
<tr>
<td>Global variable</td>
<td>Available to entire model for all times, e.g., station characteristics, operation time.</td>
</tr>
<tr>
<td>Calendar (list of events)</td>
<td>Assigned from calling time, starting time, and ending time from artificial reservation table.</td>
</tr>
</tbody>
</table>

This simulation tool will check the event from the calendar sequentially from the earliest event until the last event, and the simulation tool will implement the task based on the calendar. For instance, if the CurrentTime is 07.00 and that time is the actual calling time, the simulation tool will check if there is a car available at that time and, if at least one car is available, will assign a car to the reservation and change the status of that car from “Parked” to “Booked.”

In addition, if the CurrentEvent is at the starting time of a reservation, the simulation tool will change the status of the car from “Booked” to “On Road.” Moreover, if the CurrentEvent is at the ending time, the status of the car on the road is changed to “Parked” again. The reservation system in this simulation tool is basically the same idea as the common reservation system in carsharing services. The system checks customer reservations sequentially, and if there is a car available at a departure station, then it will assign the car to that reservation or it will be rejected (see Figure 3). The simulation tool is used for all 24 hours in a week for round-trip service regarding every step number of cars and reservations. At the end of the week, the simulation will show the average car utilization ratio and reservation acceptance ratio for a certain number of cars and reservations. All simulation results are collected and ready to be analyzed by the proposed prediction techniques.
Figure 3a. Reservation simulation
Training and Testing Procedure

In this paper, the accuracy of the proposed prediction techniques was tested and compared with each other. There were two possibilities when developing the model. First, the model is too simple and not able to learn the specificities of the data (underfitting) and second, it is too complex and will learn irrelevant details of the data and eventually its noise (overfitting). Thus, a solution to solve this problem was to rate the different complexity models with their cross-validation error estimator and to choose the superior one. This is a good solution to find which model is adapted to a certain data set. In this paper, to prevent overfitting and underfitting when predicting the data, a tenfold cross-validation was used to select the optimal model.

The difference comparison between the predicted and actual value was assessed by the correlation coefficient $R$, root mean square error (RMSE), average absolute error (AAE), maximum absolute error (MAE), and residual, as defined in Table 3. The RMSE gives an indication of the overall accuracy of the approximation, whereas MAE indicates the presence of a range that exhibits poor approximation capabilities (Al-Anazi and Gates 2010). The correlation coefficient is widely used as a measure of the strength of linear dependence between two variables (actual value and predicted value); the residual is the difference between the actual value and the estimated function value. The error measurement above becomes the benchmark to reveal the accuracy of the models. The experiment is run with a 95% confident interval with subjects $N=100$ for the simulation data.
Table 3. Error Measures for Accuracy Assessment

<table>
<thead>
<tr>
<th>Accuracy Measure</th>
<th>Mathematical Expression</th>
</tr>
</thead>
</table>
| Correlation coefficient (R)      | \[
\frac{\sum_{i=1}^{l}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{l}(y_i - \bar{y})^2 \sum_{i=1}^{l}(\hat{y}_i - \bar{\hat{y}})^2}}
\] |
| Root mean square error (RMSE)     | \[
\sqrt{\frac{1}{l} \sum_{i=1}^{l} (y_i - \hat{y}_i)^2}
\] |
| Average absolute error (AAE)      | \[
\frac{1}{l} \sum_{i=1}^{l} |y_i - \hat{y}_i|
\] |
| Maximum absolute error (MAE)      | \[
\max \ |y_i - \hat{y}_i|, i = 1, ..., l
\] |
| Residual                         | \[
y_i - \hat{y}_i, i = 1, ..., l
\] |
| Max residual                     | \[
\max (y_i - \hat{y}_i), i = 1, ..., l
\] |
| Min residual                     | \[
\min (y_i - \hat{y}_i), i = 1, ..., l
\] |

Results and Discussion

**Relationship between Number of Reservations and Number of Cars over Acceptance Ratio**

The simulation data results (number of cars, number of reservations, and acceptance ratio) were plotted in three-dimensional data with its models, as can be seen in Figure 4. The objective of a multiple regression analysis is to predict the single dependent variable (acceptance ratio) using a set of independent variables (number of cars, number of reservations). The purpose of this model for operators is to use it to predict their acceptance ratio based on their recent information on total operational cars and number of reservations. The simulation data revealed that if the number of reservations increases with respect to a certain number of cars, then the acceptance ratio will decrease (most customers will not receive a free car), but if the number of cars increases with respect to a certain number of reservations, then the acceptance ratio will increase (customers have a greater chance of receiving a free car). The maximum point for increasing the acceptance ratio up to 100 percent means that all customer reservations are absolutely accepted, and the lowest percentage is close to 1 percent (which means that only 1% of the total reservations will receive a free car).
Figure 4a. Prediction models for evaluating acceptance ratio

Figure 4b. Prediction models for evaluating acceptance ratio
As can be seen in Figure 4, this paper attempts to find the appropriate model to fit the data, so that the model can be used for prediction. Table 4 shows the percentages by the average error measures (RMSE, AAE, MAE, R, Residual) and correlation coefficient (R) of the models when predicting the acceptance ratio.
Table 4. Comparison of Models to Predict Acceptance Ratio

<table>
<thead>
<tr>
<th>Model</th>
<th>Detail</th>
<th>RMSE</th>
<th>AAE</th>
<th>MAE</th>
<th>R</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>( \hat{y} = a + bx_1 + cx_2 )</td>
<td>21.376</td>
<td>17.984</td>
<td>46.829</td>
<td>0.820</td>
<td>-46.829 to 33.261</td>
</tr>
<tr>
<td>Quadratic regression</td>
<td>( \hat{y} = a + bx_1 + cx_2 + dx_12 + ex_22 + fx_1x_2 )</td>
<td>14.675</td>
<td>11.391</td>
<td>39.462</td>
<td>0.920</td>
<td>-39.462 to 30.317</td>
</tr>
<tr>
<td>Cubic regression</td>
<td>( \hat{y} = a + bx_1 + cx_2 + dx_12 + ex_22 + fx_1x_12 + fx_1x_2 + jx_1x_2 )</td>
<td>11.696</td>
<td>8.750</td>
<td>34.796</td>
<td>0.950</td>
<td>-34.796 to 25.096</td>
</tr>
<tr>
<td>Rational regression</td>
<td>( \hat{y} = (a + bx_1 + cx_2)/(1 + dx_1 + ex_2) )</td>
<td>5.512</td>
<td>4.057</td>
<td>15.774</td>
<td>0.989</td>
<td>-13.874 to 15.774</td>
</tr>
<tr>
<td>Multi-layer perceptron</td>
<td>2 input node, 1 hidden layer, learning rate 0.3, momentum 0.2</td>
<td>13.835</td>
<td>10.551</td>
<td>37.772</td>
<td>0.930</td>
<td>-35.078 to 37.772</td>
</tr>
<tr>
<td>Radial basis function</td>
<td></td>
<td>32.738</td>
<td>28.044</td>
<td>72.086</td>
<td>0.484</td>
<td>-72.086 to 53.514</td>
</tr>
</tbody>
</table>

Multiple linear regression (Figure 4a) is applied first and results in an inaccurate prediction with RMSE 21.38; thus, the quadratic regression (non-linear regression) is applied to get the better model from the data. The quadratic regression model improves the prediction with an RMSE of about 14.67, and cubic regression is expected to smooth the prediction by showing a slight improvement of accuracy with an RMSE of about 11.7.

Rational regression (Figure 4b) was also used to predict the data, and among the proposed models, it generated the best results, with an RMSE of 5.51. The strength of dependence between the two variables (actual value and predicted value) for the rational regression is 0.989, the highest of all the results. The equation for the rational regression for predicting the acceptance ratio with the independent variables number of reservations and number of cars is described as follows:

\[ \hat{y} = (a + bx_1 + cx_2)/(1 + dx_1 + ex_2) \]  

where \( \hat{y} \) is the dependent variable acceptance ratio percentage while \( x_1 \) is the number of reservations and \( x_2 \) is the number of cars. Moreover, for other prediction algorithms, MLP achieves an RMSE of only 13.84, and RBF achieves an RMSE of 32.74. Both residuals can be seen in Figures 4c and 4d, respectively.
**Relationship between Number of Reservations and Number of Cars over Utilization Ratio**

Prediction techniques were also used to analyze the utilization ratio based on a certain number of cars and reservations. Since the real information on the utilization ratio can be extracted from a real transactional dataset, the difference when compared to the prediction result can be used to measure the maximum error of our prediction model. In addition, this model can be used to predict future utilization ratios as the operator predicts the increase of customers in the future or predicts the effect of new policies on increasing the capacity of car operations. For each operator, the model does not predict exactly or perfectly because of the variation in datasets, but it can be used to understand the pattern of the acceptance ratio or the utilization ratio given total reservations and the total number of cars operated. The similarities in the trip patterns of operators in many parts of the world (Costain et al. 2012; Morency et al. 2007; Zhou and Kockelman 2011) to our dataset can serve as one of the measurements that this simulation model is able to use to interpret general information regarding acceptance ratios or utilization ratios. More details on the results of the simulation data and its models can be seen in Figure 5.

![Figure 5a. Prediction models for evaluating utilization ratio](image-url)
Figure 5b. Prediction models for evaluating utilization ratio

Figure 5c. Prediction models for evaluating utilization ratio
Investigating the simulation data reveals that by increasing the number of reservations with respect to a certain number of cars, the percentage of the utilization ratio will be increased, which means that the company will gain more profit (more cars will be in operation) and operational car costs will go down. But, if the number of cars increases with respect to a certain number of reservations, then the utilization ratio will decrease. This would mean that a lot of cars are not being operated, which creates costs for the company. The lowest percentage of the utilization ratio is close to 0.7 percent (which means that the minimum average car can be optimized only 0.7% of the time during any given day), whereas the utilization ratio can increase to no more than 70 percent, which means the maximum average car can be optimized nearly 70 percent of the time during any given day. In this paper, the maintenance and cleaning time variables (and other variables that may reduce the utilization ratio) are not used as input, but rather it is assumed that every free car is ready to be used for a reservation (all cars are working perfectly without any problems). But in a real situation, the maintenance variable (the time required for operators to perform maintenance for each car) and the cleaning time variable (the time required for the operator to clean the car) will definitely affect the utilization ratio, and thus the maximum utilization could be predicted to be less than 70 percent. As the maintenance time for each vehicle increases, the error of prediction in this model could increase as well. If the operator uses electric vehicles, the error
prediction is not as high compared to the use of conventional vehicles, such as a gasoline or diesel cars that require more maintenance time. Because of the unused maintenance and cleaning parameters in the simulation, this result is quite surprising, because the expectation of the company could reach 100 percent (cars could operate for 24 hours nonstop), which can provide a big advantage to the company. In other words, even though the company increases the number of reservations to the maximum level or decreases the number of cars into the minimum threshold level, the operation of cars cannot be fully optimized because there will be time conflict during reservations made by the customers. The idea behind this simulation is to generate artificial reservation data based on the distribution of VHT, VKT, day of the week, and time of day (see Figure 1), and thus the conflict time during the reservations are absolutely possible.

Figure 5a reveals that the prediction by linear regression achieves low accuracy with RMSE 14.52, and it shows improvement by quadratic regression of about RMSE 9.72, whereas the cubic regression can increase into RMSE 7.44. Rational regression was also applied in this simulation data, which achieved the best accuracy, up to RMSE 2.22, while the strength of dependence between two variables (actual value and predicted value) for rational regression 0.995 showed the highest result compared to others (see Figure 5b for the model). The equation for rational regression to predict utilization ratio with the independent variable number of reservations and number of cars described as follow:

\[
\hat{y} = \frac{(a + bx1 + cx2)}{(1 + dx1 + ex2)}
\]  

(5)

where \(\hat{y}\) is the dependent variable utilization ratio percentage, while \(x_1\) is the number of reservations, and \(x_2\) is number of cars. In addition, the prediction algorithm, MLP, shows RMSE of 6.62, whereas RBF did not show good model prediction, with only RMSE 20.11 (both residual Figure 5c and 5d). The detailed results show the comparison between those models to predict the utilization ratio, as can be seen in Table 5.


Table 5. Comparison of Models to Predict Utilization Ratio

<table>
<thead>
<tr>
<th>Model</th>
<th>Detail</th>
<th>RMSE</th>
<th>AAE</th>
<th>MAE</th>
<th>R</th>
<th>Residual Min</th>
<th>Residual Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>$\hat{y} = a + bx_1 + cx_2$</td>
<td>14.516</td>
<td>12.327</td>
<td>35.226</td>
<td>0.752</td>
<td>-20.201</td>
<td>35.226</td>
</tr>
<tr>
<td>Quadratic regression</td>
<td>$\hat{y} = a + bx_1 + cx_2 + dx_{12} + ex_{22} + fx_{1x2}$</td>
<td>9.716</td>
<td>7.691</td>
<td>27.463</td>
<td>0.897</td>
<td>-18.578</td>
<td>27.463</td>
</tr>
<tr>
<td>Cubic regression</td>
<td>$\hat{y} = a + bx_1 + cx_2 + dx_{12} + ex_{22} + fx_{13} + gx_{23} + hx_{1x2} + ix_{12x2} + jx_{1x22}$</td>
<td>7.437</td>
<td>5.594</td>
<td>23.022</td>
<td>0.941</td>
<td>-16.223</td>
<td>23.022</td>
</tr>
<tr>
<td>Rational regression</td>
<td>$\hat{y} = (a + bx_1 + cx_2)/(1 + dx_1 + ex_2)$</td>
<td>2.218</td>
<td>1.792</td>
<td>6.2</td>
<td>0.995</td>
<td>-4.744</td>
<td>6.2</td>
</tr>
<tr>
<td>Multi-layer perceptron</td>
<td>2 input node, 1 hidden layer, learning rate 0.3, momentum 0.2</td>
<td>6.623</td>
<td>4.848</td>
<td>20.802</td>
<td>0.954</td>
<td>-17.125</td>
<td>20.802</td>
</tr>
<tr>
<td>Radial basis function</td>
<td></td>
<td>20.107</td>
<td>16.785</td>
<td>49.802</td>
<td>0.411</td>
<td>-27.111</td>
<td>49.802</td>
</tr>
</tbody>
</table>

**Relationship between Acceptance Ratio and Utilization Ratio**

The acceptance ratio is an important parameter, as it can be one parameter to reveal customer satisfaction, but it is difficult to acquire this information in the real carsharing system, because this simulation idea and real carsharing implementation is totally different. Information communication technology is widely used in many areas, especially in carsharing. Internet access can be easily used by the customer to make the reservation, and it can also be seen in the carsharing system always having a portal website, which allows the customer to easily make a reservation. Customers can avoid conflicting times by choosing different reservation times if they have a flexible time schedule. Otherwise, they will find another car from a different company or alternative transportation, which means the previous company is losing money. Thus, it is difficult to trace acceptance ratio information (searching history of customer is not stored in carsharing database). In this simulation, the reservation data are generated from the distribution of VHT, VKT, and time of the day, which is similar to phone reservations (not by website). Thus, this simulation focuses only on the assumption that every customer makes an appointment/reservation by phone (or website, with a condition that every customer’s searching history can be traced). Afterwards, the system will store all request reservations and evaluate whether the reservation is accepted or not.
In another case, the utilization ratio is one parameter to measure the profit of a company that can be obtained easily from the real carsharing system database. The information about cars operated is standard information in a carsharing database; thus, this parameter can be easily implemented in the real carsharing system. Based on this problem, the relationship between the utilization ratio and the acceptance ratio is an important issue. The value of the acceptance ratio can be predicted (dependent variable) if the value of the utilization ratio (independent variable) is obtained first. This result, shown in Figure 6, shows the relationship between the acceptance ratio and the utilization ratio by simulation. The investigations reveal that the acceptance ratio is an inverse negative logistic in regard to the utilization ratio with respect to a certain number of cars and reservations.

![Figure 6a. Prediction models for evaluating acceptance ratio (utilization ratio as input)](image-url)
Figure 6b. Prediction models for evaluating acceptance ratio (utilization ratio as input)

Figure 6c. Prediction models for evaluating acceptance ratio (utilization ratio as input)
As can be seen in Figure 6, the relationship of both parameters can be mapped into the proposed prediction models. Again, regression analysis is the best way to obtain the model from the information above. The investigations reveal that the logistic curve (Figure 6b), which obtained RMSE 7.71, is the best model compared to other regression models, whereas the prediction algorithms MLP and RBF achieve only RMSE 8.55 and 16.41, respectively (both residuals of the model can be seen in Figure 6c and 6d). As explained before, the acceptance ratio can approach 100 percent, whereas the utilization ratio is only 70 percent. Thus, the linear, quadratic, and cubic regression models are not good for predicting the highest point acceptance ratio and the sigmoidal regression model (Figure 6a) is not good for predicting the highest utilization ratio (the predicted value increases to 100%). The logistic curve model (Figure 6b) is the best model, with a maximum point nearest 100 percent for the acceptance ratio and reaches to about 70 percent for the maximum utilization ratio. The equation for logistic curve to predict acceptance ratio with independent variable utilization ratio described as follows:

$$
\hat{y} = a + \frac{(b - a)}{1 + (x1/c)(-d)}
$$

(6)
where $\hat{y}$ is the dependent variable acceptance ratio percentage and $x_1$ is the utilization ratio. However, if there is a finding in the real data where the utilization ratio can be optimized until 100 percent, not 70 percent as the predicted by simulation, others propose that models such as sigmoidal regression (Figure 6a) can be predicted as the better model. The detailed model can be seen in Table 6.

### Table 6. Comparison of Models to Predict Acceptance Ratio (utilization ratio as input)

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>RMSE</th>
<th>AAE</th>
<th>MAE</th>
<th>R</th>
<th>Residual Min</th>
<th>Residual Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression</td>
<td>$\hat{y} = a + bx_1$</td>
<td>12.683</td>
<td>10.373</td>
<td>29.747</td>
<td>0.941</td>
<td>-29.747</td>
<td>21.606</td>
</tr>
<tr>
<td>Quadratic regression</td>
<td>$\hat{y} = a + bx_1 + cx_1x_2$</td>
<td>8.039</td>
<td>5.921</td>
<td>37.298</td>
<td>0.977</td>
<td>-15.173</td>
<td>37.298</td>
</tr>
<tr>
<td>Cubic regression</td>
<td>$\hat{y} = a + bx_1 + cx_1x_2 + dx_1x_3$</td>
<td>7.803</td>
<td>5.842</td>
<td>37.228</td>
<td>0.9779</td>
<td>-15.647</td>
<td>37.228</td>
</tr>
<tr>
<td>Sigmoidal regression</td>
<td>$\hat{y} = a \exp(-\exp(-(x_1 - b)/c))$</td>
<td>7.791</td>
<td>5.482</td>
<td>40.620</td>
<td>0.9780</td>
<td>-15.359</td>
<td>40.620</td>
</tr>
<tr>
<td>Logistic curve</td>
<td>$\hat{y} = a + (b - a)/(1 + (x_1/c)(-d))$</td>
<td>7.710</td>
<td>5.386</td>
<td>39.553</td>
<td>0.9785</td>
<td>-16.027</td>
<td>39.553</td>
</tr>
<tr>
<td>Multi layer perceptron</td>
<td>1 input node, 1 hidden layer, learning rate 0.3, momentum 0.2</td>
<td>8.559</td>
<td>6.129</td>
<td>42.92</td>
<td>0.974</td>
<td>-16.553</td>
<td>42.92</td>
</tr>
<tr>
<td>Radial basis function</td>
<td></td>
<td>16.410</td>
<td>12.561</td>
<td>43.45</td>
<td>0.898</td>
<td>-43.45</td>
<td>22.737</td>
</tr>
</tbody>
</table>

In addition, as can be seen from Figure 6b, the slope of the logistic function is highly negative; in other words, increasing a little input will create a high decrease of output. Thus, it is appropriate for the operator to be more careful when deciding to propose utilization ratio.

**Clustering Acceptance and Utilization Ratio**

It is important for the company to decide the appropriate ratio before implementing its policy of carsharing service in real situations. Since the threshold of the proposed acceptance ratio and utilization ratio can be standard to determine the number of cars with respect to the information of a certain number of reservations, the company is faced with three big policy choices: increase profit (utilization ratio), which will reduce customer satisfaction (acceptance ratio); decrease profit, which will increase customer satisfaction; or choose the appropriate profit and provide satisfaction to the customer. However, to understand whether the accep-
tance ratio or utilization ratio is high, medium, or low is difficult for managers; thus, it is appropriate to cluster the simulation data into three clusters, which represent the three issues above. The K-Mean clustering was used to cluster the data by using the Euclidian distance technique. The details can be seen in Figure 7.

As can be seen from Table 7, the three clusters are found with its centroid, or, in other words, the policy of the company can be divided into three options (see Table 8). Every company has a different assessment to measure the level of profit and customer satisfaction. Thus, this information can not only be the one measure in regards to the level of profit and customer satisfaction but it also can be a benchmark to simply understand the grouping of simulation data. Based on Table 8, the company can consider its policy to refer to the three clusters. Companies can either increase their profit and lose customer satisfaction (Cluster 1) or vice versa (Cluster 2) or take the safe route, increasing profit without losing customer satisfaction (Cluster 3).

In addition, as can be seen from Table 7, the probability of the simulation result becoming cluster 1, 2, or 3 is about 53, 25, and 22 percent, respectively (seen from the total data in each cluster). It means that there is a 53 percent chance of the company starting its standard mode carsharing business with Cluster 1. This result demonstrates that opening a carsharing business is very promising in terms of
profit, but it is appropriate to encourage the level of customer satisfaction. Furthermore, as can be seen in Figure 6f, the slope of the logistic function is highly negative, which means that reducing only a little input value of the utilization ratio will create a big increase of the acceptance ratio. In other words, without too much profit or loss, a company can substantially increase the acceptance ratio.

### Table 7. Cluster of Acceptance and Utilization Ratio

<table>
<thead>
<tr>
<th>Attribute</th>
<th>K-Mean Cluster</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utilization Ratio</td>
<td>Centroid</td>
<td>65.9515</td>
<td>15.3648</td>
<td>50.4805</td>
</tr>
<tr>
<td>Acceptance Ratio</td>
<td>Centroid</td>
<td>12.602</td>
<td>99.147</td>
<td>53.344</td>
</tr>
<tr>
<td>Total data</td>
<td></td>
<td>53</td>
<td>25</td>
<td>22</td>
</tr>
</tbody>
</table>

### Table 8. Details of Clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Company Profit</th>
<th>Customer Satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>2</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>3</td>
<td>Medium</td>
<td>Medium</td>
</tr>
</tbody>
</table>

### Conclusions and Future Work

Because it is difficult to predict the initial number of operation cars needed in carsharing without losing customer interest and company profit, this paper demonstrates that a simulation model must be developed first to evaluate the acceptance and utilization ratios for traditional round-trip service that is based on traveling frequency, number of vehicles, and VHT and VKT patterns. The two evaluation parameters proposed in this paper are the acceptance ratio, which is the parameter that reveals customer satisfaction, and the utilization ratio, which reveals operator profit. In this paper, the Korea Carsharing (WeShareCar) dataset was collected and converted into artificial reservation data according to its distribution. The discrete event simulation was developed and run to analyze the acceptance and the utilization ratios for every combination of the number of cars and the number of reservations in a week. The simulation data revealed that increasing the number of reservations with respect to a certain number of cars will decrease the acceptance ratio (most customers will not receive a free car), but it will increase the percentage
of the utilization ratio (more cars will be rented by customers). Based on this result, regression analysis is the best model for predicting the percentage of the acceptance and the utilization ratios compared with other prediction algorithms such as MLP and RBF. Later, both can be used as a threshold for carsharing companies to optimize the number of operating cars with respect to their recent number of reservations. In addition, in this paper, a prediction model is proposed to investigate the relationship between the acceptance and the utilization ratios. Thus, through using this model and the real utilization ratios that the company collects from its operational database, it can predict the general acceptance ratio of customers. Our investigations have revealed that if the percentage of the acceptance ratio is increased, the utilization ratio will decrease and vice versa.

Put simply, the simulation data are clustered into three groups that can be considered as additional options for company policy before starting their business. Companies can either increase their profit and lose customer satisfaction, or vice versa, or take the safe route and increase profit without losing customer satisfaction. In addition, the cluster results of simulation data show that half of the companies that start carsharing businesses will make a profit, but they need to maintain and increase customer satisfaction levels. This result can be used as additional evidence to strengthen the case for the benefits of carsharing that have been demonstrated by previous research and that have concluded that operators are more likely to be economically successful. Furthermore, as an effect of the need to maintain customer satisfaction, the relationship model of the acceptance and the utilization ratios reveal that a small reduction in the input value of the utilization ratio will create a large increase in the acceptance ratio. The implication for the company is that without too much loss in profits, a company can substantially increase the acceptance ratio (customer satisfaction).

Finally, there were evident limitations to this project. First, only the operational dataset was used for basic round-trip service in carsharing because of the necessity for the preliminary step of implementing carsharing in South Korea. In the future, increasing the size of the dataset, increasing the sample of the subject experiment, and introducing and upgrading the simulation model for additional services such as one-way and open-ended service in this simulation might be considered as future projects. The evaluation of other parameters in the future might also be considered, such as the option of relocating with its costs, the pricing of services, the number of customers (its relation with the number of reservations), fuel costs,
profit, the distance between home and parking lot, cleaning costs, and maintenance costs.

**Acknowledgments**

This research was supported by the Industrial Technology Innovation Programs-International Collaborative Research and Development Program (N0000701) of the Ministry of Trade, Industry & Energy, Korea.

**References**


**About the Authors**

**Jongtae Rhee** ([jitrhee@dgu.edu](mailto:jitrhee@dgu.edu)) is a Professor in the Industrial & Systems Engineering Department of Dongguk University. His focus is on RFID, supply chain management, carsharing service, and data mining.

**Ganjar Alfian** ([ganjar@dongguk.edu](mailto:ganjar@dongguk.edu)) is a Ph.D. candidate in the Industrial & Systems Engineering Department of Dongguk University. His focus is on RFID, simulation, data mining, and service development.

**Byungun Yoon** ([postman3@dongguk.edu](mailto:postman3@dongguk.edu)) is an Associate Professor in the Industrial & Systems Engineering Department of Dongguk University. His work experience includes being an IT consultant for LG and a visiting scholar in the Centre for Technology Management (CTM) of the University of Cambridge. His focus is on patent analysis, new technology development methodology, and visualization algorithms. His current interest is in enhancing technology roadmapping and product design with data mining techniques. He has authored articles published in *R&D Management, Technological Forecasting and Social Change*, and *Technology in Society*, among others.