공학박사 학위논문

Design of Rapid Bus Routes Considering Transit Efficiency

대중교통 효율성을 고려한 급행 간선버스 노선 설계

2021년 2월

서울대학교 대학원

건설환경공학부

이 은 학

Design of Rapid Bus Routes Considering Transit Efficiency

지도교수 고 승 영

이 논문을 공학박사 학위논문으로 제출함
2021년 2월

서울대학교 대학원
건설환경공학부
이 은 학

이은학의 공학박사 학위논문을 인준함
2021년 2월

위 원 장 이 청 원 (인)

부위원장 고 승 영 (인)

위 원 김 동 규 (인)

위 원 홍 상 연 (인)

위 원 김 진 희 (인)

Abstract

As the use of public transport increases, more comfortable and faster service of public transportation is required to users. Rapid bus (RB) services have been regarded as one of the solutions that allow the public transportation system to operate efficiently. This study aims to optimize the RB routing problem, including both the selection of service areas and the optimization of the route. First, the efficiency of public transportation service for origin-destination (O-D) pairs was evaluated by data envelopment analysis. Second, the service areas such as inefficient O-D pairs were selected by their efficiency scores. With the selected origin-destination pairs, the number of vehicles and their service routes were optimized by a genetic algorithm. The proposed model aims to design the optimal routes of RBs by minimizing the total cost and maximizing the efficiency score. The decision variables were set to the number of vehicles and routes of each vehicle. The proposed model was applied to the transit system in Seoul, and the results showed that the RB service improved the efficiency score of O-D pairs significantly. Specifically, the efficiency score of 19 selected service areas was increased from 0.19 to 0.51 on average. Regarding the total cost and revenue, the routes from the proposed model were compared to the routes from other conventional models, e.g., high demand, long out of vehicle time, and long travel time-oriented models. As a result of the comparison analysis, the proposed model showed the highest value of a sum of the total cost and revenue, such as 15 (10,000/KRW) on average.

Keyword : Transit route network design problem, Vehicle routing problem, Rapid bus, data envelopment analysis, efficiency

Student Number : 2016-21263

Contents

Table List

Figure List

Chapter 1. Introduction

1.1. Background

As the use of public transport increases, more comfortable and faster service of public transportation is required for urban travelers. In several decades, the urban railway was regarded as a major mode to improve the public transportation system. One of the advantages of the rapid transit service is to transport users with high capacity and scheduled speed. Among public transportation modes, buses have advantages in operating costs and flexible route planning. With the advantages of buses, many cities are implementing preferential bus policies such as bus-only lane system, bus priority signals, upgraded buses, and bus intelligence systems. By combining these bus policies, the bus rapid transit (BRT) system was introduced considering as an effective way to solve urban traffic problems.

With low-cost and high-performance, rapid buses (RBs) are regarded as the most efficient mode to improve traffic congestion and public transportation services. The RBs transport the users faster than regular buses. RBs usually drive on main roads and stop at major bus stops. Specifically, the RBs are relatively advantageous for connecting urban and suburban areas. Figure 1 shows the comparison of two bus services, e.g., conventional bus and RB services. The conventional bus service stops at all the bus stops along the route. Conversely, the RB service only stops at the main stops along the route. The RB service provides faster mobility to users than conventional bus service by using the main roads and stopping a small number of stops.

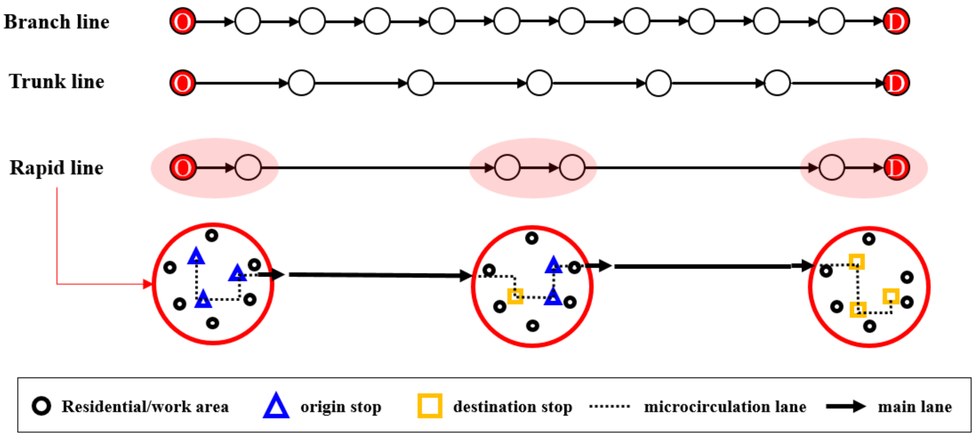
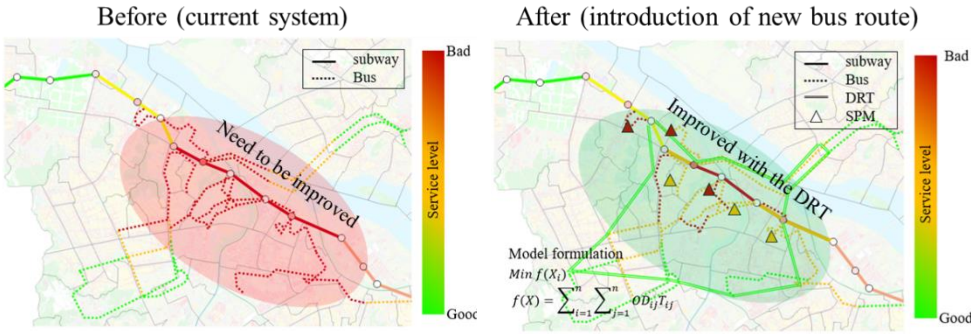


Figure 1 Comparison of bus services

There are two major issues, e.g., service area selection and routing problem, to introduce the new RB service. First, the service area selection is required to be performed to introduce the RBs. The operators and decision-makers usually explore areas where the transit service is vulnerable or the demand is excessive. With the selected areas, the operator designs a bus route considering maximum social benefits and a minimum travel cost. Service area and route planning usually has been carried out based on the survey and sampled revealed preference data. Figure 2 illustrates the example of the service quality before and after the introduction of new public transportation routes. The operator usually tried to find the service areas where improvement is required, as shown in Figure 2(a). Then, the quality of service is expected to be improved after the introduction of the new transit services, as shown in Figure2(b).



(a) (b)

Figure 2 The service improvement with the new transit route

In the academic field, the transit network design problem for introducing bus routes consists of two main topics, e.g., service area selection and bus line planning (vehicle routing problem). First, service area selection is a problem of selecting areas in need of improvement of transit service. The service areas are usually selected by the indicators related to traffic volume, socio−economics, and level of public transportation service. Second, line planning determines the optimal route with fleet size, stop station, and frequency. These two main topics, e.g., service area selection and bus line planning, are important in introducing new bus routes. Two main topics especially interact with each other and are required to be considered within the same framework. It is also necessary to design a process that integrates two topics with the same objective. Therefore, this study proposes RB routing model considering both service area selection and line planning based on the efficiency evaluation.

1.2. Purpose of the Study

This research aims to optimize the RB route considering transit efficiency. The process optimization consists of two stages, e.g., service area selection and line planning. The objective of all stages is to improve the efficiency of public transportation. The consistency of the objectives of each stage overcomes the limitation of integrating the objectives in the academic field and the field of demonstration. The two main problems to be solved in this study are shown in Figure 3 below.

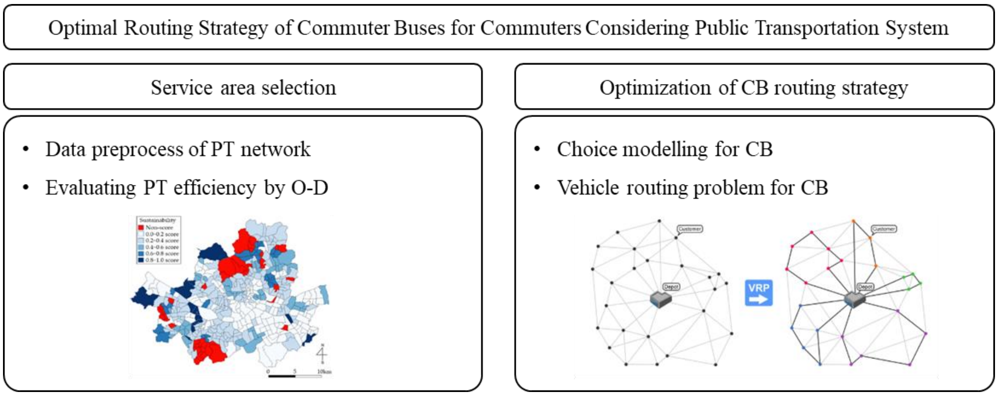


Figure 3 Main purposes of rapid bus routing problem

The contents of this study consist of six chapters. In chapter 2, related studies for the service area selection, optimization of bus routing problems, and evaluation of transit efficiency using data envelopment analysis (DEA) are reviewed. In chapter 3, the smart card data used in this study is illustrated with descriptive statistics. In chapter 4, two models, e.g., DEA model and optimization model, are developed to selecting the service area and optimizing routing problem, respectively. The DEA model is designed to measure the efficiency of public transportation of origin−destination by using the operator's travel distance and the user's travel time variables, e.g., waiting, in-vehicle, and transfer time. The bus route optimization model is designed to minimize the total cost, such as a sum of the operating cost and user travel cost. The decision variables are set to the number of vehicles, bus routes, and the stop station. The combinatorial optimization model is designed based on a genetic algorithm (GA) to optimize multiple RB routes concurrently. The objective function of the upper model is designed to minimize the total cost. In the lower level, the user's modal split for all modes (e.g., auto, transit, and RB) and assignment for RB are performed. A multinomial logit model is applied to perform a modal split and user assignment for RB.

In chapter 5, the proposed routing problem algorithm is applied to the transit network in Seoul. Firstly, service areas are selected by evaluating the efficiency of O-D based on the administrative unit of Seoul. Second, the optimization of RB routes is performed based on the bus stops located in the selected O-D. The evaluation of efficiency improvement is also performed to diagnose the optimized results. In chapter 6, the summary, contributions of the proposed model and future research are suggested.

Chapter 2. Literature Review

2.1. Selection of Service Area

Service area selection is including in the network sketch planning. The transit sketch planning process, such as service area selection, has been suggested to achieve sensitive transit planning procedures. The service area is defined as the areas where the transit service is required to be improved. The transit sketch planning has been carried out depending on the demand model. The demand models estimate demand by assigning passengers to alternative modes. However, these travel demand estimates are not sensitive to policy variables. The demand model process is also time-consuming, and it is difficult to use to test various transport policy alternatives.

To overcome this, some indicators based on simple variables have been used to select service areas. The travel time and demand are regarded as the main variables to evaluate transit service.

There are several previous studies to evaluate and select service areas using travel time and demand variables. For example, Viggiano et al. (2018) developed the indicator using Euclidean distance and travel time. The potential time savings were estimated by the developed indicator. The areas with the high value of travel time saving were mentioned as the service area for improving the transit service. Similarly, Matthew et al. (2018) developed the accessibility indicator using the travel time percentile. The areas with low accessibility were selected as the service area. The developed indicator was applied to the MRT line of Singapore. Vanderwaart et al. (2017) also planned and modified the bus services of Massachusetts Bay with the travel time variable. The travel pattern was estimated with the smart card data and the bus line was proposed to save the travel time. Jang (2010) and Park et al. (2008) evaluated the transit service using the travel time variable. The level of transit service was estimated by predominated traffic analysis zone (TAZ). Several previous studies have selected service areas using traffic volume variables (Cheranchert and Maitra, 2020; Bahbouh and Morency, 2014). The clustering analysis was performed to group the O-D pairs which have similar origin and destinations. These grouped O-D pairs were selected as the service area which needed to be improved in transit service. By introducing bus lines that connect directly to grouped O-D pairs, the passenger convenience and route efficiency are concurrently improved.

The indicators reviewed above have the advantage of being simple to calculate. However, some improvement is required on these indicators. First, objective measurement and relative evaluation between service areas are required. The result of the indicator is derived as a simple value, so it is difficult to know whether the corresponding value is a large value or a small value. Second, it is difficult to derive the degree of improvement. The indicator values simply provide the need for improvement, and no improvement criteria are provided. Third, it is required to select service areas considering the total travel time variable and other variables. For rigorous analysis, it is necessary to consider various variables that make up public transportation comprehensively. For example, the travel time variable consists of waiting time, in-vehicle time, and transfer time, and each effect of travel time variables required to be measured.

Table 1 Previous studies related to service area selection

No.	Author	Purpose	Method (Criteria: index)	Results
1	Viggiano et al. (2018)	Service area selection	Indicator: Euclidean distance ÷ travel time	Potential time savings
2	Conway et al. (2018)	Service area selection	Indicator: Accessibility (travel time percentile)	Application to MRT line (subway) in Singapore
3	Vanderwaart et al. (2017)	Service area selection and modification of transit line	Indicator: Travel time	Time savings and demand diversion
4	Huang and Levinson (2015)	Service area selection	Indicator: Circuity (Euclidean dis.÷ shortest dis.)	Improvement of circuity
5	Cheranchery and Maitra (2018)	Service area selection	C−means clustering	Measurement of the importance of variable
6	Bahbouh and Morency (2014)	Service area selection	flow	Selecting O−D pairs of Montreal
7	Qiu et al. (2018)	Estimating of travel patterns of transit users	Clustering the TAZ (DN algorithm)	Travel time distribution by TAZ
8	Jang (2010)	Evaluation of transit mobility	Indicator (travel time)	Travel time distribution by TAZ
9	Eom et al. (2015)	Evaluation of the level of service (LOS) of transit	Development of indicator (mobility and equity)	LOS by TAZ
10	Park et al. (2008)	Estimating of travel patterns of transit users	Development of indicator (travel time and flow)	Flow distribution by TAZ
This study		Service area selection	Modeling: distance÷travel time (wait, transfer, in−vehicle)	Rapid bus routes

2.2. Optimization of Bus Routing Problem

The vehicle routing problem (VRP) is an optimization and integer programming problem to optimize the vehicle routes for visiting a set of locations. Since VRP has a non-deterministic polynomial-time hard (NP-Hard) problem, optimization would be more difficult with an expansion of the service area.

The heuristic algorithm-based VRP has been widely used in route optimization problems. Since the VRP is the NP-hard problem, the metaheuristic algorithm is required to estimate the optimal solution. Representative metaheuristic methodologies include genetic algorithm(GA), simulated annealing algorithm and Tabu search algorithm. Simulated annealing algorithms are generally known to have a disadvantage in that it is difficult to consider various solutions compared to GA. It is also known that the Tabu search algorithm does not guarantee a better solution than a random search like the GA.

GA is widely used for VRP since it estimates a fast and accurate solution compared with other algorithms. Many previous studies have used GA to deal with complex issues, such as the bus routing problem. It was regarded as suitable for finding a global solution that goes beyond the local solution. The GA is applied to determine the optimal solution by checking each solution's convergence that evolves over generations. Many VRP related studies, therefore, conducted the routing problem of bus lines using GA.

GA is a method that provides the optimal solutions of both constrained and unconstrained optimization problems. The algorithm derives an optimal solution based on a natural selection process of biological evolution. The chromosome and the gene need to be defined to use GA. The chromosomes are defined as the stations composed of two types, i.e., express and local. The gene is the set of chromosomes which are the array of the stations. Since the objective function of this study is set to minimize the total travel time, the fitness of each gene evolves toward decreasing travel time.

The process of the GA consists of selection, crossover, mutation, and replacement steps. First, the selection is the most critical operation, a computational process for selecting a genetically good parent in a population. Parents that have evolved to meet the objective function and the constraint are selected. Second, the crossover is the operation of producing the offspring. The crossover uses the parents who are selected in the selection step. The crossover creates the offspring (a new array of the vehicles and visiting stations) by crossing the order of the visiting station. Third, the mutation is the step of modifying the generated gene of the offspring. In the mutation step, the visiting stations for each vehicle in the new array could be changed randomly. The mutation step prevents local minima problems. Finally, Replacement is to change the population to evolve into the next generation. Replacement constructs a population of new generations by replacing genes in the population with newly created genes. The population of a new generation could simply be substituted for all genes, or only the inferior genes could be substituted.

There are three typical types of GA, e.g., binary, permutation and real-value algorithm. Among these, the permutation algorithm is adopted to optimize the RB routes in this research. The general process of GA is shown in Figure 4.

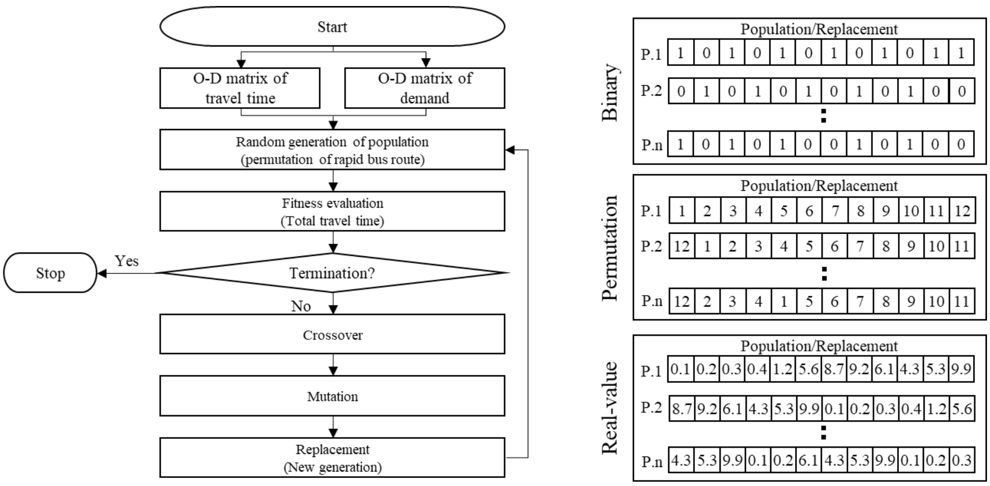


Figure 4 General process of the genetic algorithm

There are many previous studies to find optimal bus routes using VRP. For example, Guo et al. (2018) presented the modified VRP and minimized user and vehicle operation costs with the demand assumption. Similarly, Lyu et al. (2019) proposed the bus routing strategy considering the passenger's choice probability of bus routes. Zheng et al. (2019) optimized the vehicle route considering demand aggregation for each service area. Chen et al. (2018) applied a GA to the vehicle routing problem of bus lines using smart card and GPS data, and Qiu et al. (2018) selected a service area and optimizing the VRP of the bus routes based on passenger demand using the smart card data. Although previous studies have optimized bus routes with different ideas and objective functions, the routes were generated by heuristic algorithms.

The VRP concerns the service of passenger transportation. A passenger travels from the origin station to the destination station, and the bus runs the fastest routes to transport passengers. The travel cost and the travel time between each passenger and the station need to be identified to find the optimal bus route. The network is transformed into one where the vertices are stations, and the arcs are the links between the stations. The cost on each arc is the lowest cost between the two stations on the given network.

The formulation of fundamental VRP can be express as below. Let $\mathrm{G} = (V, A)$ be an undirected graph consisting of station node set $V = (0,1,2 \ldots, n)$, and edge $A = \{(i,j) | i,j \in V,\ i \neq j\}$. Node set V consists of the n stations, and $dist(i,j)$ refers to the distance between two nodes, e.g., i and j. The travel time from i to j is t_{ij}, which is calculated from $dist(i, j)$ and the speed of vehicle v. The term x_{ijk} implies whether the vehicle k traveled the link from s_i to s_j. If x_{ijk}=1, the vehicle was used; if x_{ijk}= 0, the vehicle was not used. It can be determined whether or not the k^{th} vehicle departed by checking $\sum_P x_{0jk}^h$. If the vehicle departed, there must be one station reached out of n, resulting in $\sum_S x_{0jk}^h$ =1, otherwise 0. Node set S is divided into boarding nodes $S^+ = \{s_i | s_i \in V,\ l_{jk} > 0\}$ and alighting nodes $S^- = \{s_i | s_i \in S,\ l_{jk} < 0\}$.

The objective function is to minimize the total travel cost.

$$\min \sum_{i \in V} \sum_{j \in V} c_{ij} x_{ij}$$

The constraint below is the in−degree, which denotes that exactly one arc leaves each node.

$$\sum_{i \in V} x_{ij} = 1, \forall j \in V$$

Likewise, the constraint below denotes the outdegree, which states that exactly one arc enters each node.

$$\sum_{j \in V} x_{ij} = 1, \forall j \in V$$

Two constraints below impose the requirements for the depot node.

$$\sum_{i \in V} x_{io} = K$$

$$\sum_{j \in V} x_{oj} = K$$

Capacity constraint ensures the vehicle capacity requirements and the connectivity of the solution are satisfied.

$$\sum_{i \notin S} \sum_{j \notin S} x_{ij} \geq r(S), \forall S \subseteq V$$

$$x_{ij} \in \{0,1\}, \forall j \in V$$

GA and VRP are the most useful mathematical programming to find the optimal bus routes. GA provides a flexible structure to design the routing problem in various forms. VRP also provides the optimal routes based on the objective function and constraints.

Table 2 Previous studies related to the bus routing problem

No.	Author	Service area selection (decision variables)		Routing problem			
		Criteria of service area selection	Quantification (ranking)	Objective	Decision variable		
					Line planning	Number of vehicles	Frequency
1	Wang et al. (2020)	High demand	X	Minimizing travel time and operation cost	–	1 by 1	–
2	Lyu et al. (2019)	High demand	X	Maximizing profit	–	1 by 1	–
3	Chen et al. (2018)	High demand	–	Maximizing demand/ minimizing travel time	–	1 by 1	X
4	Guo et al. (2018)	High demand	–	Minimizing user /operation cost	–	1 by 1	–
5	Zhang et al. (2018)	High demand	–	Minimizing operation cost	–	Multi	–
6	Li et al. (2018)	High demand	X	Minimizing user travel cost	–	1 by 1	X
7	Tong et al. (2017)	Predominant area	X	Maximizing profit	–	1 by 1	X
8	Zheng et al. (2015)	Predominant area	X	Minimizing environmental cost	–	1 by 1	–
This study		Efficiency score (DEA model)	–	Minimizing travel time and operation cost	–	Multi	–

2.3. Evaluation of Transit Efficiency

A major purpose of measuring efficiency using the DEA model was to determine the efficiency of firms, organizations, and industries. The concept of the DEA model was designed to evaluate the efficiency of the production functions. The DEA model was employed to evaluate the relative efficiency objectively with multiple input and output variables (Banker et al., 1984; Charnes et al., 1962; Farrell et al., 1962). Farrell et al. (1962) first suggested the concept of the DEA model to estimate efficient production functions. The productive frontier was estimated to deal with the non-convexity implicit in increasing returns and in estimating the efficiency of British farms. With the concept of the efficiency analysis, the DEA model was developed by Charnes, Cooper & Rhodes, and it had been called as CCR model, which assumes the constant returns to scale (CRS).

Rezaee et al. (2016) combined DEA and the Nash bargaining game as a cooperative game theory approach to evaluate the efficiency of transportation systems by a large scale of measures. Zhao et al. (2016) used the Charnes, Cooper, and Rhodes (CCR) model to evaluate the operational efficiency of transportation. The efficiency scores were estimated to compare the transportation efficiency of different periods. Some researchers have evaluated the performance of transportation systems with various fuzzy methods. Hanaoka and Kunadhamraks (2009) used a fuzzy analytical hierarchy process (AHP) to evaluate the performance of intermodal freight transportation systems. Celik et al. (2013) used a Fuzzy Multi-Criteria Decision Making (MCDM) to estimate customer satisfaction with the transit system. These studies used fuzzy logic for dealing with imprecise and qualitative data.

The DEA model seems to be more useful for evaluating public efficiency than other parametric methods, and previous researchers have made this point. Nishiuchi et al. (2015) sought to comprehend the use of public transportation systems based on smart card data in Kochi City, Japan. For example, Lee et al. (2019) evaluated the efficiency of the public transportation system of Seoul from smart card data using the data envelopment analysis (DEA) model. The results showed that the service areas with a low number of transit trips and long travel time were analyzed as inefficient areas. The results also mentioned that transit efficiency could be improved by adjusting public transportation routes. Similarly, Nishiuchi et al. (2015) also suggested that travel time and the number of trips are crucial factors to evaluate the efficiency of the public transportation system.

Hahn et al. (2013) and Hahn et al. (2017) developed a network DEA model for evaluating the efficiency of bus companies in Seoul, Korea. The model considered transportation services and reflected the non-storable nature of transit services. They evaluated the efficiency of 113 arterial bus routes in Seoul in 2009 using a modified Barker, Charnes, and Cooper (BCC) model that considered both desirable and undesirable outputs to improve the existing system used to evaluate bus services. Tobit regression analysis also was performed to identify the most effective variables for maximum efficiency. Hahn and colleagues also used a DEA model to evaluate the efficiency of the trucking industry in Korea (Hahn et al., 2015). Using smart card data from 10 transportation terminals in Beijing, Sun et al. (2010) combined individual performance measures into a single comprehensive measure based on the DEA model.

Network slacks-based measure (NSBM) DEA is used to DEA is a nonparametric method for estimating efficiency (production) frontiers. The DEA model identifies relative efficiencies using multiple input and output variables (Tone et al., 2010). The purpose of estimating efficiency is to control the strategy of a policy, enterprise, or organization.

Since the CRS condition assumes that the production unit is kept constant at the optimal scale, the output and input are scaled proportionally. The CCR model is the most important in the sense that it shows the most abbreviated methodological features. The CCR model estimates a ratio that can reduce the input as much as possible while keeping the output constant and vice-versa. As an example of the input-oriented CCR model, there are some considerations to estimate the efficiency score. The efficiency score is estimated by summing the weights of the output variables. The summed weights of output variables are not over the 1.0 value, and the weights of the output and input variables are over the 0.0 value. With the observed J DMUs (j = 1, ..., J), each DMU produces the M outputs using N inputs. The ratio of the input value versus output value is the efficiency score θ and the objective function is to minimize the θ^i which is the reduced ratio of the input variables of target DMU i. The input−oriented CCR model, therefore, measures the weights of input and output variables to minimize the θ^i, and the efficiency score is estimated by those of values. The maximum value of efficiency score is equal to or less than 1.0 value since the maximum value of the objective function is 1.0 with the constraints, i.e., $y, x > 0$ and $\lambda \geq 0$. The mathematical expression of the input−oriented CCR model is as follows:

$$\theta^{i^*} = \min_{\theta,\lambda} \theta^i$$

subject to:

$$\theta^i x_m^i \geq \sum_{j=1}^{J} x_m^j \lambda^j$$

$$y_n^i \leq \sum_{j=1}^{J} y_n^j \lambda^j$$

$$\lambda^j \geq 0$$

where θ^{i^*} is the efficiency score of target O−D pair i, y_n^j is the output variable n of an O−D pair j (n = 1, ..., N), x_m^j is the input number of each variable m of an O−D pair j (m = 1, ..., M), j is the number of observed O−D pairs (j = 1, ..., J), n is the output variable, m is the input variable, and λ^j is the weight of each O−D pair j.

Table 3 Previous studies related to evaluate transit efficiency

No.	Author	Purpose	Data	Method	Results
1	Lee et al. (2019)	Evaluation of efficiency of transit oriented development	Smart card data & socio−economic data	Network slacks−based data envelopment analysis	Efficiency score of station area and TAZ
2	Lee et al. (2019)	Evaluation of efficiency of bus and subway transfer	Smart card data & socio−economic data	Data envelopment analysis	Efficiency score of transfer station area
3	Jing et al. (2018)	Evaluation of efficiency of transit oriented development	Smart card data & socio−economic data	Data envelopment analysis	Efficiency score of transfer station area of Tokyu line
4	Han et al. (2015)	Evaluation of efficiency of bus route	Bus operation logs	Data envelopment analysis	Efficiency score of bus line
5	Nishiuchi et al. (2015)	Evaluation of efficiency of transfer stations	Smart card data	Data envelopment analysis	Efficiency score of transfer station area
6	Sun et al. (2010)	Evaluation of efficiency of transfer stations	Smart card data	Data envelopment analysis	Efficiency score of transfer station area
This study		Evaluation of efficiency of O−D pairs	Smart card data	Data envelopment analysis	Efficiency score of O−D pairs

2.4. Implication

With the literature review, implications for the three main topics, e.g., service area selection, bus line planning, and evaluation (objective) were derived. Service area selection is a problem of selecting areas in need of improvement. The service areas are usually selected by the indicators related to traffic volume, socio-economics, and level of public transportation service. Line planning determines the optimal route with fleet size, stop station, and frequency. Evaluation refers to the direction of service improvement and is related to the objective of the analysis. The evaluation (objective) is carried out from three perspectives (operator, user, and public transportation service). All three topics are important in introducing new bus routes. However, it is not easy to design an optimal bus route considering all three topics since it is a large-scale problem. Many previous studies usually perform optimal bus routing by considering only one or two major topics.

Previous research related to service area selection and line planning selects service areas using traffic volume and socio-economic indicators, and then optimizes routes for the selected areas. In research related to line planning, optimization is mainly performed from the operator's or the user's perspective.

This approach has a limitation in that it may not select areas where public transport services should be improved. Also, since this approach aims to maximize operating revenue, it leads to different results than the objective in the service area selection stage. In order to improve public transport services by introducing the new bus route, the objective of each stage must be consistent.

Chapter 3. Data Description

3.1. Description of Smart Card Data

The transit system of Seoul has been operating the automatic fare collection (AFC) system with the smart card since 2004. The transit fares are charged to the users based on their total traveled distance from the origin to the destination station. With the smart card, users can use any combination of public transit alternatives, e.g., urban railway, bus, and both. Since the smart card data in Seoul provides 99% of transit users' trip information, it is suitable for analyzing transit service. Table 4 shows the 38 indices of the smart card data in Seoul

.

Table 4 Description of smart card data (May 17, 2017)

No.	Data information	No.	Data information
1	Card ID	20	Boarding violation penalty
2	Transaction ID	21	Alighting violation penalty
3	Mode code	22	User code: general
4	Line ID	23	User code: student
5	Name of the transit line	24	User code: children
6	Vehicle ID	25	User code: others
7	Vehicle number	26	User type
8	Boarding station ID	27	User group
9	Alighting station ID	28	Company code
10	Name of boarding station	29	Company name
11	Name of alighting station	30	Time code
12	Boarding time	31	Starting run time
13	Alighting time	32	Ending run time
14	Number of transfers	33	Boarding date
15	Total travel distance	34	Alighting date
16	Total travel time	35	Year
17	Boarding fare	36	Zone code
18	Alighting fare	37	Transfer station ID
19	The number of users	38	Transfer time

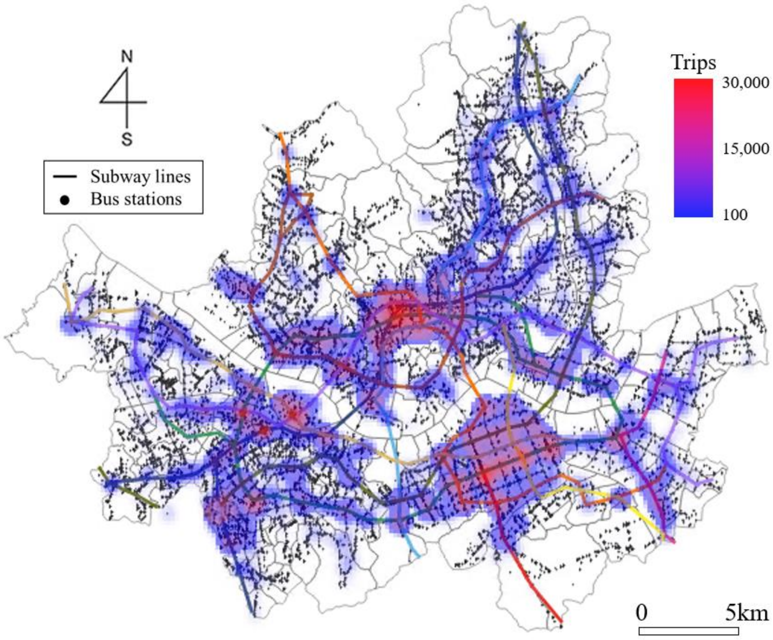


Figure 5 Transit network in Seoul (morning peak)

3.2. Descriptive Statistics of Smart Card Data

The proposed model was applied to optimize the routes of RBs at morning peak hours (7 to 9 AM). The application site is Seoul, which consists of 424 administrative districts. The public transportation system in Seoul includes 363 subway stations with 11 lines and 46,945 bus stations with 354 bus lines.

The smart card data provide about 99% of the public transportation trip information. In Seoul, the number of trips using the public transportation network in the morning peak hours, e.g., 7 to 9 AM, is 1,125,740 on May 17, 2017. The numbers of trips using the subway, bus, and transfer trips in the morning are 531,404, 277,730, and 316,606, respectively. In this study, smart card data of May 17, 2017, was used to select the RB service area and optimize the RB routes. Smart card data can be obtained from the Korea Transportation Safety Authority (KTSA).

With the pre−processing of the data, the number of chained trips is 15,330,903, i.e., 2,271,505 trips of morning peak hours, 13,059,398 trips of non−peak hours. The travel time of the users is 31.8 minutes per trip on average, i.e., 30.1 minutes of morning peak hours and 32.2 minutes of non−peak hours. The number of transfers is 0.32 per trip on average, i.e., 0.34 transfers of morning peak hours and 0.32 transfers of non−peak hours. The travel distance is 11.1km per trip on average, i.e., 10.8 km of morning peak hours and 11.2 km of non−peak hours. The descriptive statics of smart card data is shown in Table 5.

Table 5 The descriptive statics of the smart card data

	Number of trips (trips)	Average travel time (minutes)	The average number of transfers (transfers)	Average travel distance (km)
Total (0-24 hours)	15,330,903	31.8	0.32	11.1
Morning peak hours (7-9 hours)	2,271,505	30.1	0.34	10.8
Non-peak hours (0-7 & 9-24 hours)	13,059,398	32.2	0.32	11.2

With the smart card data, the visualization of descriptive statistics was performed to help understand the public transportation service in Seoul. Figure 6 shows the visualization of the number of trips and transfers per day. There are three major areas, e.g., Jongro, Yeouido, and Gangnam, in Seoul which has a high number of trips. These areas have both residential and commercial land−use features. The number of trips shows a pattern that occurs from the outskirt of the city. The number of transfers is also increased relatively at the outskirt of the city.

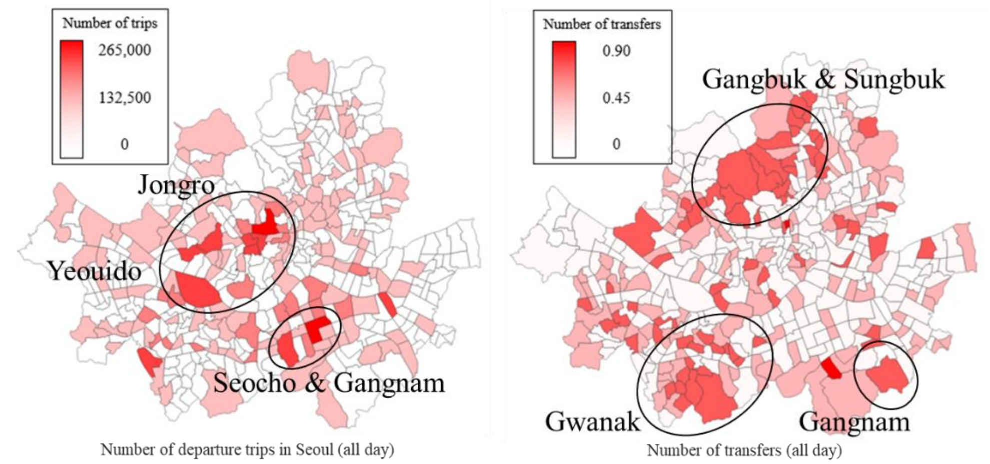


Figure 6 Visualization of trips and transfers per day

Figure 7 shows the number of trips and transfers during peak hours (7−9 AM). The number of trips and transfers is also visualized by the Dong unit. Conversely to the daily pattern, the numbers of trips and transfers during peak hours are high at the outskirt of the city. These areas have both residential land use features. The number of trips shows a pattern that occurs from the outskirt of the city. The number of transfers is also increased relatively at the outskirt of the city.

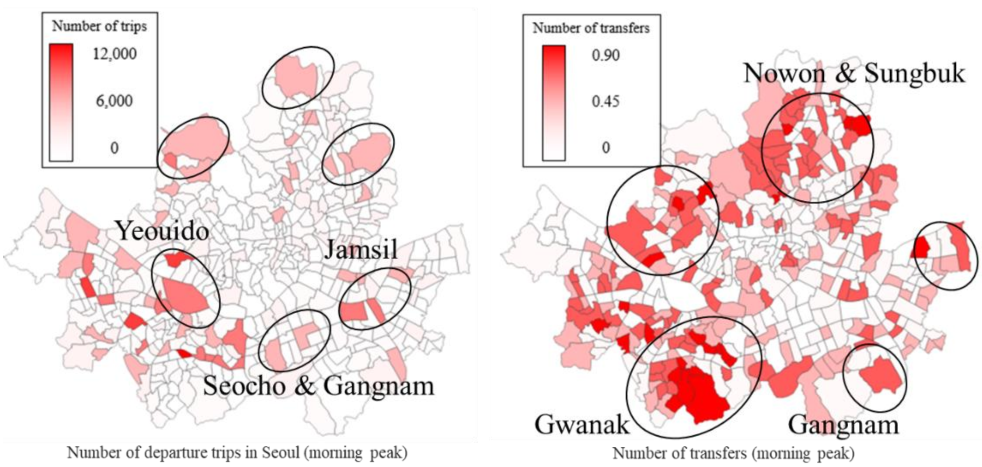


Figure 7 Visualization of trips and transfers during peak hours

Chapter 4. Methodology

4.1. Concept and Definition

This study aims to develop an optimization framework to optimize the RB routes. The framework is designed in two stages, e.g., service area selection and route optimization. First, the DEA model is developed to evaluate the efficiency score of O−D pairs and the inefficient O−D pairs are selected as the RB service areas. Second, the optimization model is developed to minimize the total cost of the RB service. The travel time related variables, e.g., waiting time, in-vehicle time, and transfer time, are used in both stages to synchronize each stage's objective. The efficiency score and the optimal routes are derived by using the travel time related variables. The framework of the proposed RB optimization model is shown in Figure 8.

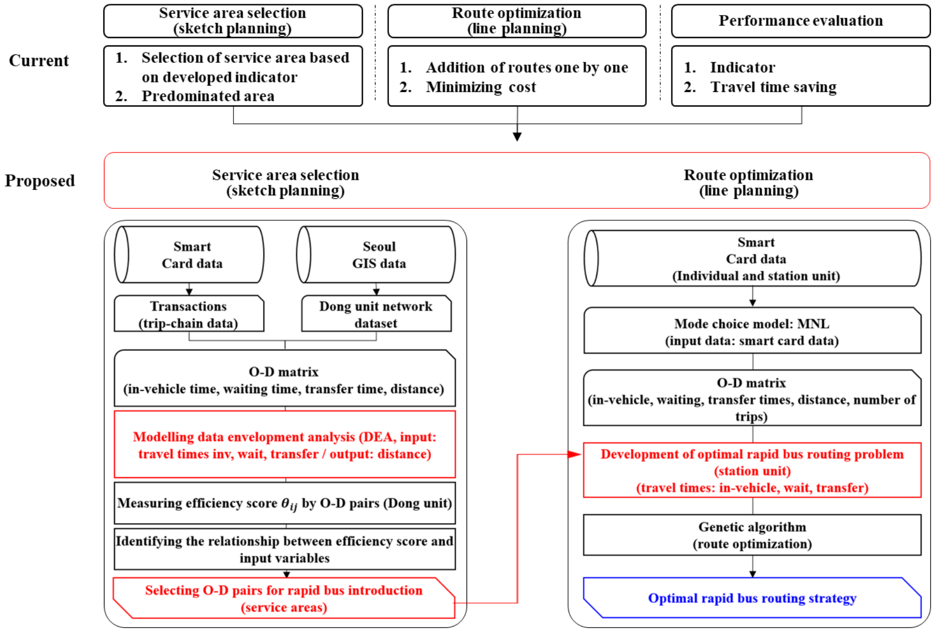


Figure 8 Framework of the proposed RB optimization model

4.2. Conditions and Assumptions

The RB is the new type of bus service that directly connects the major origin and destination areas. The attributes of RB are compared with those of the conventional bus services, e.g., feeder line and branch line. The feeder line buses run short distances to secure mobility within a specific area. The operation time and fare of feeder lines of Seoul are 5 to 24H and 1200 KRW, respectively. The trunk line buses run medium and long distances between areas. The operation time and fare of trunk lines of Seoul are all the same as those of feeder lines. The proposed RB line connects the selected O–D pairs directly during the morning peak hours. The RB provides a faster travel time to the passenger by stopping only designated stations of the O–D area. Thus, it is advantageous to connect relatively distant areas. The operation time and fare of RB lines are set to 7–9H and 2400KRW, respectively.

Table 6 Types of the bus line in the urban area

	Conventional		Proposed
	Feeder bus line	Trunk bus line	Rapid bus line
Route	fixed	fixed	fixed
Service area	Inner area	Inter area	Selected O–D area
Passenger	anyone	anyone	anyone
Station	Existing bus stop	Existing bus stop	Existing bus stop
Service type	All stop	All stop	Express
Operation time	All day (5–24H)	All day (5–24H)	Morning peak hours (7–9H)
Fare	1,200KRW	1,200KRW	2,400KRW

4.3. Service Area Selection with DEA

4.3.1. Conditions for Data Envelopment Analysis

The DEA model derives different results depending on the assumptions that are given to the production possibility set. Assumptions and measurement techniques should be applied to the model through a full review of the decision unit. There are four representative assumptions to develop to DEA model. First, the direction of efficiency evaluation is required to be set. The input oriented model measures efficiency based on the efficient DMU that uses the least inputs while maintaining output. The output oriented model measures efficiency based on the DMU that produces the most output while maintaining inputs.

Figure 9 shows the concepts of the input−oriented model and the output−oriented model. Figure 9(a) illustrates the input−oriented model and shows how much the input amount of inefficient DMU can be reduced. A point P is required to be reduced to P' to ensure efficiency. Figure 9(b) illustrates the output−oriented model and shows how much the output of an inefficient DMU can be reduced. A point U is required to be increased to U' to ensure efficiency.

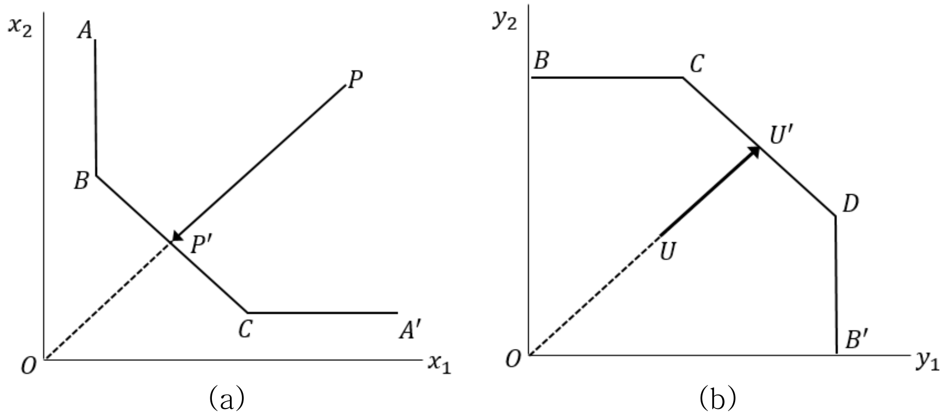


Figure 9 Concept of the input and output oriented model: (a) input oriented model; (b) output oriented model

The second is about the disposability assumption. All inputs and outputs are either strong or weak disposability. Strong (free) disposability means that inputs and outputs can be freely disposed of without external restrictions. Strong disposability assumes no additional costs associated with dealing with surplus inputs or producing undesirable outputs. Weak disposability refers to that inputs and outputs cannot be freely disposed of without external restrictions. Weak disposability assumes that there are additional costs associated with dealing with surplus inputs or producing undesirable outputs.

The efficient frontier varies depending on the type of disposability. Figure 10 shows the efficient frontier by disposability. The efficient frontier with the strong disposability, regards the negative marginal production as inefficient. Thus, the efficient frontier with the strong disposability assumption becomes S_s. On the other hand, the efficient frontier that assumes weak disposability includes the negative marginal production. Therefore, the efficient frontier becomes S_w. The efficiency score of the point P is OP_s/OP with strong disposability assumption, and OP_w/OP with weak disposability assumption.

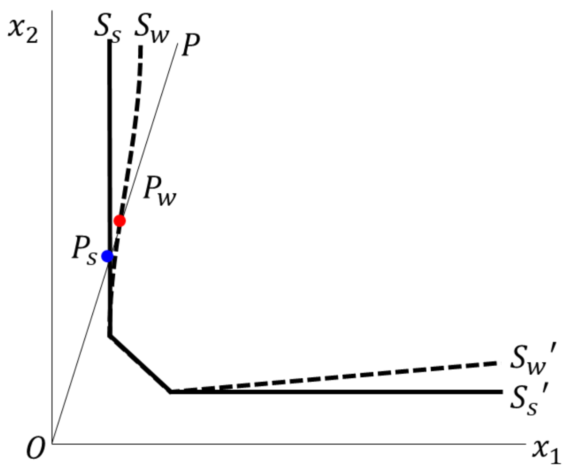


Figure 10 Concept of the efficient frontier by disposability

The third assumption is about the returns to scale. There are two types of returns to scale, e.g., constant returns to scale (CRS) and variable returns to scale (VRS). DEA model with CRS assumption measures the efficiency, assuming the current state is optimal. On the other hand, the DEA model with the VRS assumption measures the efficiency while excluding inefficiencies due to returns to scale. The efficiency score assuming CRS is estimated inefficiently when the decision unit is not optimal even though it is technically efficient.

Conversely, the efficiency score assuming VRS is estimated only considering technical efficiency. Therefore, the efficiency score with the VRS assumption is estimated to have a higher number of efficient DMUs and higher scores than the efficiency score with the CRS assumption. Figure 11 shows the two efficient frontiers according to CRS and VRS assumptions. The efficient frontier with CRS consists of a DMU that achieves the optimum scale. The efficient frontier with VRS is formed by reflecting both DMUs that reach the optimum scale in the long term and DMUs that do not. For point P, the efficient frontier with CRS is S_1 and the efficient frontier with VRS is S_2.

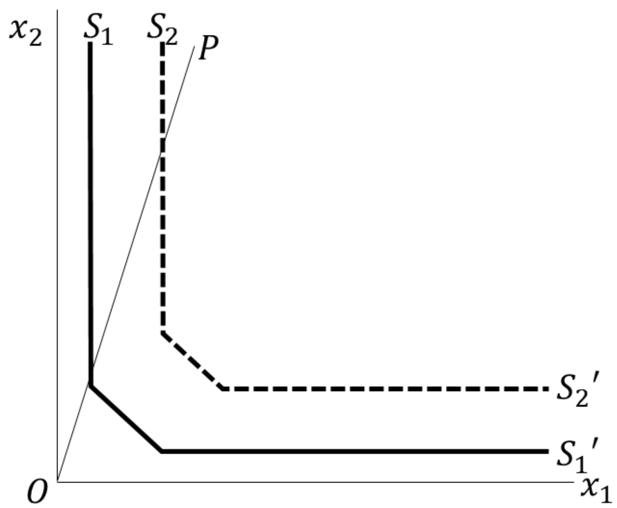


Figure 11 Concept of the efficient frontier by returns to scale

The fourth assumption is about radial form. The radial model measures the efficiency score based on the point where the target point and the efficient frontier meet. The radial model assumes that inputs and outputs can be proportionally increased or decreased. Figure 12 shows examples of radial and non−radial models. With the radial model, the efficiency score of the point P is measured based on the reference point P' on the efficient frontier Ss. With the non−radial model, the efficiency score of the point P is measured based on the point P''.

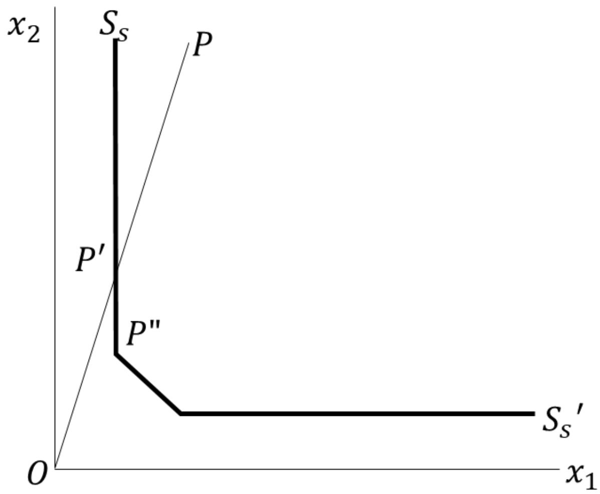
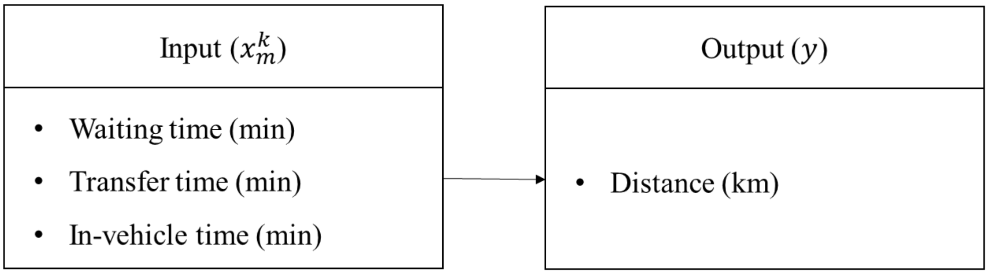


Figure 12 Concept of radial and non−radial model

4.3.2. Data Envelopment Analysis for Service Area Selection

In this study, the service areas were selected by evaluating the efficiency of public transportation service. The efficiency of public transportation can be measured by various factors, e.g., the number of trips, travel distance, travel time, and transfers (Lee et al., 2019). The DEA model is useful for evaluating the performance of public transportation, which consists of various variables. The DEA is a kind of nonparametric linear programming that can measure the efficiency of multiple decision-making units (DMUs) when the production process has a structure of multiple inputs and outputs (Charnes et al., 1962). Note that the DEA determines the relative efficiency of DMUs by estimating production frontiers, and it compares the performance of all DMUs in the dataset.

<Input and output variables for transit efficiency>



<Example of dataset>

DMU	Input variable			Output variable
	Waiting time (min)	Transfer time (min)	In-vehicle time (min)	Travel distance (km)
O1 - D2	5	15	20	10
O2 - O3	3	10	14	7
O3 - O5	4	0	10	8

Figure 13 Variable setting for evaluating the efficiency of the public transportation service

In this study, we used an input−oriented BCC model to evaluate the efficiency of the public transportation service by minimizing the inputs. The waiting time, in−vehicle time and the transfer time of O−D pairs on average were set as input variables, and the shortest road distance of O−D pairs was set as output variables. With the results of the DEA analysis, the operator can select the O−D pairs with low efficiency scores as the RB service areas. The input−oriented BCC model used in this study is expressed as Equation 1.

$$\theta^{k^*} = Min \;\; \theta^k \tag{1}$$

Subject to

$$\theta^k x_m^k \geq \sum_{j=1}^{J} x_m^j \lambda^j$$

$$y_n^k \leq \sum_{j=1}^{J} y_n^j \lambda^j$$

$$\sum_{j=1}^{J} \lambda^j = 1$$

$$\lambda^j \geq 0 \;\; (j = 1,2, \dots, J)$$

where θ^{k^*} is the efficiency score of DMU k; y_n^j is the nth output variable of a DMU j; x_m^j is the mth input variable of a DMU j; j is the number of DMUs (j = 1,2, .., 782); n is the output variable number (n = 1, y_1: travel distance); m is input variable number (m = 1, 2 and 3; x_1: waiting time, x_2: in−vehicle time, x_3: the number of transfers); and λ^j is the weight of each DMU, j.

4.4. Optimization of Bus Routing Problem

The demand estimation and vehicle routing were performed to optimize the RB routes of the service areas selected from the service area selection stage. For the demand estimation, the mode choice probability of RB was estimated by using the multinomial logit model. For the vehicle routing problem, minimizing total cost and several considerations were set as objective function and constraints, respectively.

4.4.1. Multinomial Logit Model for Demand Estimation of Rapid Bus

The demand is one of the most important factors to optimize the RB route. The mode choice model developed by Ministry of Land, Infrastructure and Transport (2017) is applied to estimate the demand of RB. The multinomial logit model (MNL) is the most typical method for discrete choice analysis in the transportation field. Each passenger has four mode alternatives of subway, bus, and subway+bus in this study. The choice probability of each alternative mode can be estimated with the MNL. The utility function used in this study is expressed in Equations 2 and 3.

$$pr_{ij}(r) = \frac{e^{U_{r,ij}}}{\sum_{q \in C_{ij}} e^{U_{q,ij}}}, \forall q \in C_{ij}, q \neq r \tag{2}$$

$$U_r = a + \beta_t t_r + \beta_{fare} fare_r + \beta_{tr} tr_r \tag{3}$$

$\forall r \in$ (sbw : subway, bus : bus, snb : subway+bus)

$U_{sbw} = \beta_t t_{sbw} + \beta_{fare} fare_{sbw} + \beta_{tr} tr_{sbw}$

$U_{bus} = a_{bus} + \beta_t t_{bus} + \beta_{fare} fare_{bus} + \beta_{tr} tr_{bus}$

$U_{snb} = a_{snb} + \beta_t t_{snb} + \beta_{fare} fare_{snb} + \beta_{tr} tr_{snb}$

where pr_{ij} the choice probability of the mode r from origin i to destination j; $U_{r,ij}$ is the utility function associated to alternative mode r from origin i to destination j (r = sbw: subway, bus: bus, sbs: subway + bus); C_{ij} is the feasible choice set from origin i to destination j; β_t is the coefficient of the travel time; β_{fare} is the coefficient of the fare; β_{tr} is the coefficient of the number of the transfer; t_r is the travel time (minutes) of the mode r; $fare_r$ is the fare of the mode r; and tr_r is the number of transfers of the mode r.

4.4.2. Optimal Strategy for Optimizing Rapid Bus Route

To optimize vehicle routes, GA was used in this study with the following considerations. First, minimizing the total cost (sum of operation and user travel cost) of RB is set as the objective function. Second, there is a capacity limit for the RB vehicle. Each number of RB vehicles increases when the required passenger load exceeds the passenger limit of the RB vehicle. Third, the travel time of the RB passenger is equal to or less than the travel time of the passenger using public transportation services. Fourth, when multiple RB vehicles are operated, the visiting orders of stations are re-estimated in order to equalize the travel time of each RB vehicle. This is the process of maximizing operational efficiency by adjusting the balance of travel time. The calculation of the phase is described in detail in the following section.

There are several assumptions for applying the proposed RB routing strategy as follows. (i) The route of RB uses the existing public transportation stations. (ii) The RB transports the passengers to the origin and destination stations recorded in the smart card data. (iii) The passenger's travel time of the RB is equal to or less than the travel time of public transportation. (iv) The numbers of passengers in RBs are non−negative and an integer with the maximum value of 45 (vehicle capacity). (v) There are four vehicles for two selected service areas, and all vehicles are available to use if necessary. (vi) RB does not visit the stations where passengers are not boarded or alighted. Note that the number of passengers for RB at each station is estimated by the mode choice model. (vii) Each passenger has four alternatives, e.g., subway, bus, subway + bus, and RB. (viii) The RB fare is set to 2400KRW.

4.4.2. Mathematical formulation for optimizing RB route

GA is a method that provides the optimal solutions of both constrained and unconstrained optimization problems. The algorithm derives an optimal solution based on a natural selection process of biological evolution. The chromosomes are defined as the vehicles, and each vehicle consists of the vectors of the visiting stations. The gene is the set of chromosomes which are the number of vehicles. Since the objective function of this study is set to minimize the total cost, the fitness of each gene evolves toward decreasing the total cost. The proposed GA consists of five stages, e.g., population generation, crossover, mutation, fitness evaluation, and termination. First, the chromosomes are randomly generated as parent generation. Each gene consists of a vector such as the set of the visiting stations. Second, new chromosomes such as spring generation are generated by crossing the gene of parent chromosomes. Third, the visiting stations of vehicles are mutated to prevent local minimum solution. Fourth, the fitness evaluation is performed by calculating the total cost of chromosomes. These four steps are iterated until satisfying the convergence condition.

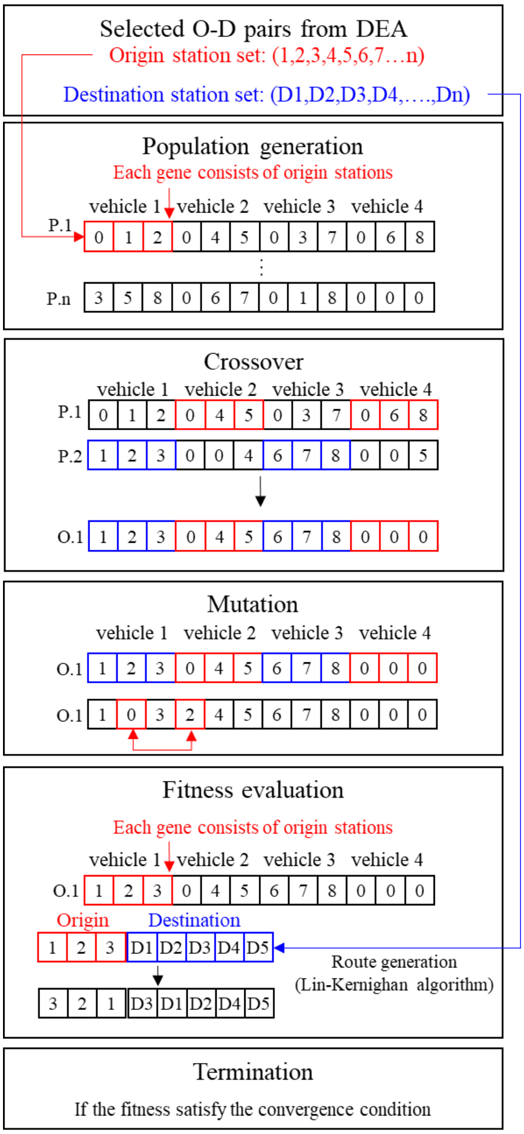


Figure 14 Genetic algorithm for optimal rapid bus routing.

The bus route is optimized by five steps of GA, as mentioned above. Figure 15 shows the example of the route generation with the GA process. First, the population step generates vehicles with specific routes. Second, the crossover step mixes vehicles and generates a new route. Third, the mutation step changes the route by switching only certain nodes in the route. The iteration of four steps is performed until the termination condition is satisfied.

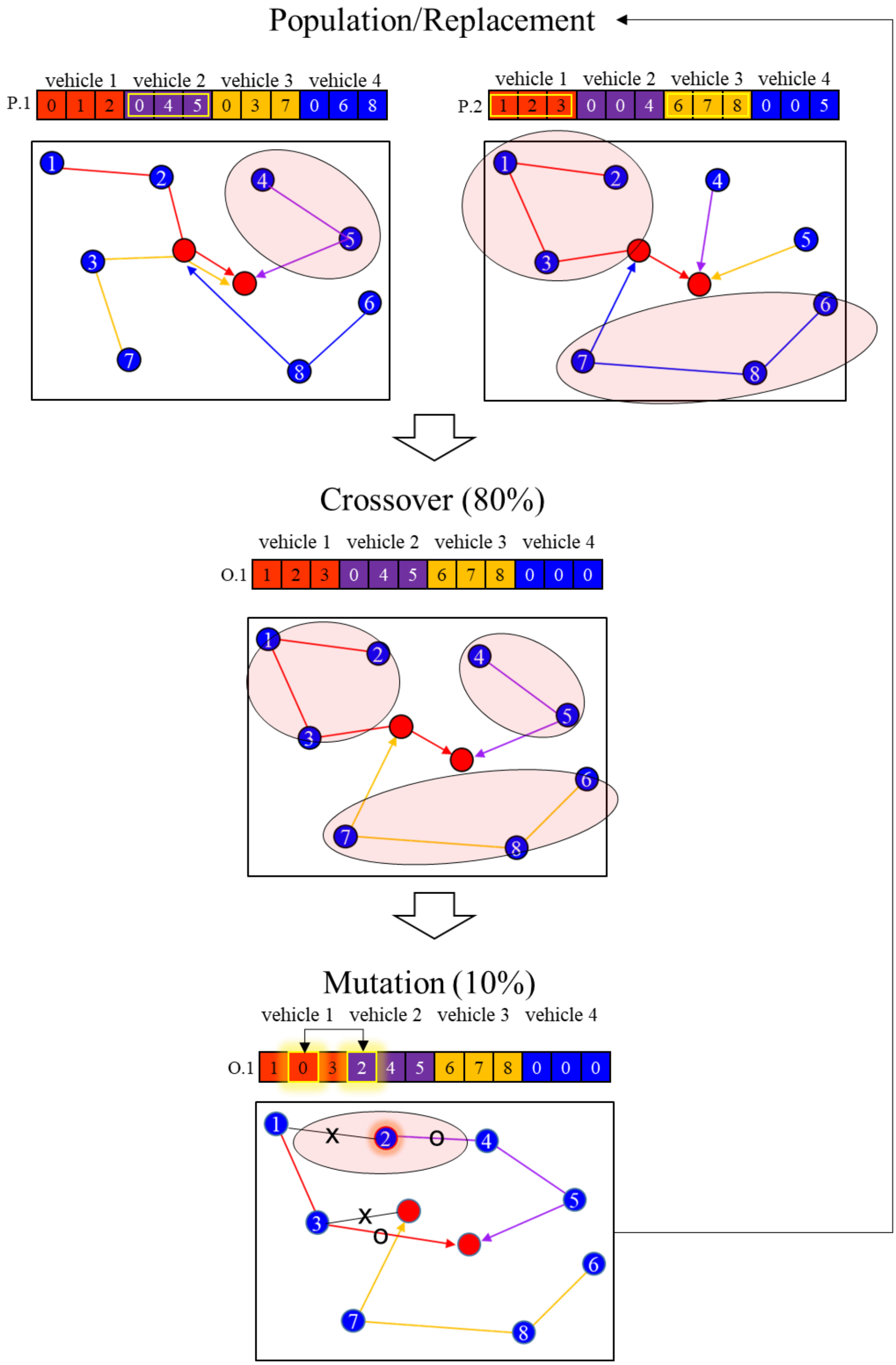


Figure 15 The route generation with genetic algorithm

The mathematical concepts of graphical nodes and edges are used to formulate the vehicle routing problem. Mathematical symbols are used to represent information of stations and links as follows. Let $\mathrm{G} = (S, E)$ be an undirected graph consisting of station node set $S = (s_1, \ldots, s_n)$, and edge $E = \{(s_i, s_j) | s_i, s_j \in S,\ s_i \neq s_j\}$. Node set S consists of the n stations, and $dist(s_i, s_j)$ refers to the distance between two nodes, e.g., s_i and s_j. The travel time from s_i to s_j is t_{ij}, which is calculated from $dist(s_i, s_j)$ and the speed of vehicle v. The term x_{ijk} implies whether the vehicle k traveled the link from s_i to s_j. If x_{ijk}=1, the vehicle was used; if x_{ijk}= 0, the vehicle was not used. It can be determined whether or not the k^{th} vehicle departed by checking $\sum_P x_{0jk}^h$. If the vehicle departed, there must be one station reached out of n, resulting in $\sum_S x_{0jk}^h$ =1, otherwise 0. Node set S is divided into boarding nodes $S^+ = \{s_i | s_i \in S,\ l_{jk} > 0\}$ and alighting nodes $S^- = \{s_i | s_i \in S,\ l_{jk} < 0\}$.

The objective function that minimizes the total cost is formulated in Equation 4. The total cost is composed of operation cost and users' travel cost. The users' travel cost is derived by multiplying passengers by the travel cost of the link. The constraints for the optimization are set as follows. Equation 5 implies that the numbers of inflows and outflows are the same for node i. Equation 6 and Equation 7 represent the conditions in which the passenger load of the vehicle is non-negative and does not exceed the capacity. Equation 8 implies how travel time for each link is obtained. Equation 9 implies that the passengers' travel time of the RB is less than the travel time of public transportation. Equation 10 is the objective function of the lower level. Equation 12 represents the calculated function of the choice probability of RB.

Upper level

Minimize $F = \sum_i \sum_j \left(f_{ij}^{cb} T_{ij}\right) + \sum_{k=1}^{m} \sum_i \sum_j (x_{ijk} t_{ij})$ (4)

Subject to

$\sum_i x_{ijk} = \sum_i x_{jik}$ (5)

$l_j + P_{jk} \geq 0$ (6)

$l_j + P_{jk} \leq Q$ (7)

$t_{ij} = dist(p_i,\ p_j)/vd$ (8)

$\sum_{i=i'}^{j'} \sum_{i=i'}^{j'} (x_{ijk} t_{ij} + t_{dwell}) < pt_{i'j'}$, $\forall i' \in S^+,\ j' \in S^-$ (9)

Lower level

Minimize $f = \sum_i \sum_j \left(f_{ij}^{cb} T_{ij}\right)$ (10)

Subject to

$T_{ij} = WT_i + INV_{ij} + TR_{i\prime}$ (11)

$f_{ij}^{rb} = pr_{ij}(rb) f_{ij}$ (12)

$pr_{ij}(rb) = \frac{e^{U_{cb,ij}}}{\sum_{q \in C_{ij}} e^{U_{q,ij}}}$ (13)

where x_{ijk} is the selected link (if k^{th} vehicle travel from s_i to s_j, x_{ijk} is 1 otherwise 0); t_{ij} is the travel time from s_i to s_j; t_{dwell} is the dwell time of the station; v is the travel speed for morning peak time (km/h); Q is the load capacity of the vehicle; P_{jk} is the number of passengers loaded on k^{th}vehicle when arrived s_j; l_j is the boarding and alighting number of passengers at s_j; $pt_{i'j'}$ is the travel time of the public transportation service from i' to j' (minutes); m is the number of vehicles departed $\{a | (a-1)Q < \sum_i \sum_j \left(f_{ij}^{cb}\right) \leq aQ\}$; f_{ij}^{cb} is the number of passengers who ride the RB from s_i to s_j; $pr_{ij}(cb)$ is the choice probability of RB; f_{ij} is the total number of passengers from s_i to s_j.

Chapter 5. Application

5.1. Application with Toy Network

The proposed model was applied to developed a toy network (35 O−D pairs). Firstly, the efficiency score of transit service was evaluated using the DEA model. The result of the efficiency score of 35 O−D pairs was estimated to be 0.66 on average. The average travel distance, in−vehicle time, transfer time, and waiting time were 13.0 kilometers, 32.2, 0.4, and 4.1 minutes, respectively. The slacks for the input variables, e.g., in−vehicle time, transfer time, and waiting time, were estimated to be 3.8, 0.4, and 1.4 minutes, respectively. Each slack value indicated that it is required to be reduced to achieve efficiency such as the 1.0 score.

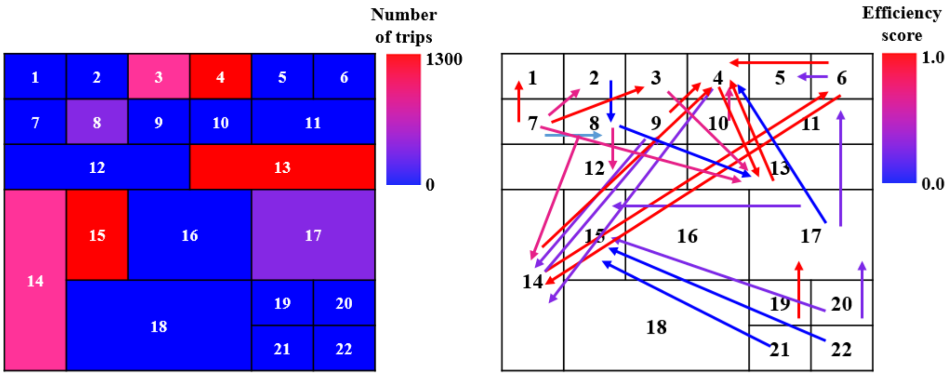
The results showed that the efficiency score of transit service is 0.66 on average (the efficiency score of the bottom 3 O−D pairs are 0.37)

Six O−D pairs were estimated to be efficient with 1.0 score. The travel distance, in−vehicle time, transfer time and waiting time of six efficient O−D pairs were 13.9 kilometers, 28.2, 0.3, and 3.1 minutes, respectively. For the bottom three O−D pairs, the travel distance, in−vehicle time, transfer time and waiting time were 14.3 kilometers, 57.1, 0.8, and 5.4 minutes, respectively. When comparing the bottom 3 and the top 6 O−D pairs, the average efficiency score of the bottom three O−D pairs was estimated to be low since the travel time was increased longer compared to the distance increased. Especially, it was analyzed that the influence of the in−vehicle time was the largest among the input variables. The slacks value of in−vehicle time was estimated to be 23.6 minutes.

Table 7 Results of the efficiency score of a toy network

		Efficiency score	Output variables	Input variables		
			Travel distance (km)	In-vehicle time (min)	Transfer time (min)	Waiting time (min)
Total (35 O-Ds)	Mean	0.66	13.0	32.2	0.4	4.1
	Std.	0.23	1.4	9.2	0.3	1.2
	Slacks	-	-	3.8	0.4	1.4
Top 6 O-Ds	Mean	1.00	13.9	28.2	0.3	3.1
	Std.	0.00	2.4	6.3	0.2	1.4
	Slacks	-	-	0	0	0
Bottom 3 O-Ds	Mean	0.30	14.3	57.1	0.8	5.4
	Std.	0.01	0.2	0.8	0.2	0.6
	Slacks	-	-	23.6	0.75	3.4
Std. means standard deviation						

Figure 16 showed the visualized results of the efficiency score of the toy network. Figure 16(a) illustrated the number of departure trips for each zone. The trips occurred intensively in zone number 4, 13, and 15. Figure 16(b) illustrated the efficiency score of O-D pairs, and many O-D pairs were directed to the zone number 4, 13, and 15. Among the 35 O-D pairs, the three O-D pairs, e.g., 17-4, 21-15, 22-15, were selected as service areas



(a) (b)

Figure 16 Visualized Results of the Efficiency score: (a) number of departure trips; (b) efficiency score of O-D pairs

The RB routes for three selected O−D pairs were optimized to improve the transit efficiency score. The results showed that three RBs were introduced by improving the efficiency score from 0.35 to 0.41. For RB only, the efficiency score, in−vehicle time, transfer time, and waiting time estimated to be 0.95, 27.1, 0.0, and 4.0 minutes, respectively. For transit service and RB, the efficiency score, in−vehicle time, transfer time, and waiting time estimated to be 0.41, 40.1, 0.5, and 4.5, respectively. The application results of the three O−D pairs are shown in Table 8 and figure 17.

Table 8 Results of the routing problem with toy network

O−D	Current transit system				Proposed							
					RB				Transit + RB			
	In−vehicle time (min.)	Trans. Time (min.)	Wait. Time (min.)	Eff. score	In−vehicle time (min.)	Trans. Time (min.)	Wait. Time (min.)	Eff. score	In−vehicle time (min.)	Trans. Time (min.)	Wait. Time (min.)	Eff. score
17 to 4 (5.9 km)	57.7	1.0	4.7	0.32	27.2	0.0	4.0	0.98	33.4	0.5	3.9	0.48
21 to 15 (6.7 km)	55.9	0.5	5.7	0.30	26.7	0.0	4.0	0.96	41.7	0.5	5.1	0.38
22 to 15 (7.7 km)	57.5	1.0	5.9	0.29	27.4	0.0	4.0	0.90	45.2	0.5	4.4	0.38
Average	57.0	0.8	5.4	0.30	27.1	0.0	4.0	0.95	40.1	0.5	4.5	0.41

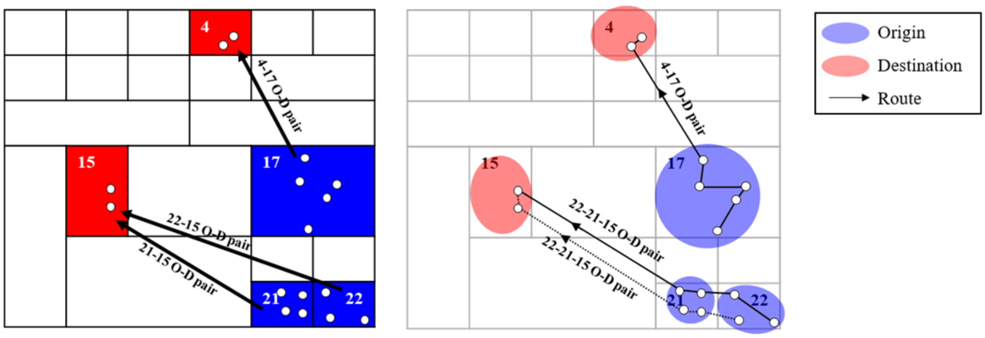


Figure 17 Visualized results of the routing problem

5.2. Application with Transit Network of Seoul

5.2.1. Assumptions

There are some assumptions for applying the proposed optimization model as follows. 1) The unit for service area selection is O–D pair based on Dong administrative unit, and the unit for optimization of the RB route is the bus station contained within the selected O–D pair. 2) The O–D pairs with more than 101 trips and 6.1 km distance are selected as the spatial range of analysis. Since RB service targets commuting demand during morning peak hours, the assumptions were set based on 99 percentile of O–D trips and the average commuting distance in Seoul. 3) For optimizing RB routes, the waiting time is set as the half of the headway. The in–vehicle time and transfer time are obtained from the smart card data. 4) The total travel time is a sum of the waiting time, in–vehicle time and transfer time. 5) The bus operation time is a sum of the link travel time and dwell time 6) The value of time for the passenger is 9011 KRW.

Table 9 Assumptions of the proposed model

No.	Category	Assumptions
1	Unit	Service area selection: Dong unit Route optimization: Station
2	O–D pair	Trips: >100 trips, Distance: >6.1 km
3	Travel time attributes	Waiting time: Half of headway In–vehicle time: alighting – boarding time Transfer time: boarding time of the second mode – alighting time of the first mode
4	Total travel time	Waiting time + In–vehicle time + Transfer time
5	Bus operation time	Link travel time + Dwell time (1minutes/stop)
6	Value of time	9011KRW per hour

5.2.2. Service area selection for the bus routing problem

As proposed earlier, the BCC DEA model was used to evaluate the efficiency of the public transportation service. The efficiency evaluation was performed on 782 O−D pairs with more than 104 transit trips during morning peak hours.

Table 10 shows the results of the efficiency score for the 782 O−D pairs in Seoul. The efficiency score was estimated to be 0.48, on average. 19 DMUs were estimated to be the most efficient O−D pairs with an efficiency score of 1.0. As the mean values of the input variables for the 19 efficient O−D pairs, the travel time, the transfer time, and travel distance were 33.3 minutes, 2.5 minutes, and 12.1 km, respectively. Three groups were selected according to the efficiency score in order to compare the DMUs with the high efficiency score and the low efficiency score. Each group consists of 19 DMUs. The efficiency score of the top 19 DMUs was estimated to be 1.0 as mentioned above. The efficiency score of the middle 19 DMUs and bottom 19 DMUs were estimated to be 0.48 and 0.19 on average, respectively. By comparing the top 19 DMUs to other DMUs, the efficiency score decreased as the travel time and the number of transfers increased. Also, the efficiency score decreased as the number of trips and travel distance decreased. In particular, the bottom 19 DMUs had a 4 minutes longer waiting time and 11 minutes longer transfer time than the tops 19 DMUs. The bottom 19 DMUs also had an in−vehicle time that was 10.1 minutes longer, despite the distance being 0.1 km shorter than the top 19 DMUs.

Overall, the bottom 19 DMUs showed a shorter travel distance (output) but longer travel time than other DMUs. Especially, the waiting time, in−vehicle time, and transfer times were about two times, 1.1 times, and four times longer than the top 19 DMUs.

Table 10 Results of efficiency evaluation of the transit service

		Efficiency score	Output variables	Input variables		
			Travel distance (km)	Waiting time (min)	In-vehicle time (min)	Transfer time (min)
Total (782 O-D pairs)	Mean	0.46	12.1	3.2	33.3	2.5
	Std.	0.13	3.7	1.0	9.5	3.1
Top 19 O-D pairs	Mean	1.00	10.5	2.6	26.9	0.5
	Std.	0.00	4.2	0.3	11.0	0.8
Middle 19 O-D pairs	Mean	0.48	13.4	2.8	33.8	1.5
	Std.	0.01	3.5	0.3	7.6	0.9
Bottom 19 O-D pairs	Mean	0.19	10.6	6.6	37.0	11.5
	Std.	0.02	1.4	1.5	4.5	4.2
Std.: standard deviation						

Figure 18 visually shows the results of the efficiency evaluation for the public transportation service in Seoul. The commute trips in Seoul are mainly connected to three inner-city areas, i.e., Gangnam, Jongro, and Yeouido. These areas are the main commercial and business districts in Seoul. Among the 102,252 trips of the 782 O-D pairs, 69,800 trips of 347 O-D pairs (68%) had these areas as destinations. The number of O-D pairs to Gangnam, Jongro, and

Yeongdeungpo are 147, 138, and 62, respectively. The O–D pairs directly connected by the public transportation service have a relatively high efficiency score, even if the travel distance between the origin and destination is long.

Conversely, O–D pairs without direct route service of public transportation have a relatively low efficiency score, even if the travel distance between the origin and destination is short. From the user's perspective, these results indicate that the commuters who travel along transfer lines of public transportation services experience relatively less convenience than the commuters who travel with the direct service. Based on these results, this study aims to introduce the optimal RB route that directly connects the origin and destination stations of O–D pairs that have a low efficiency score.

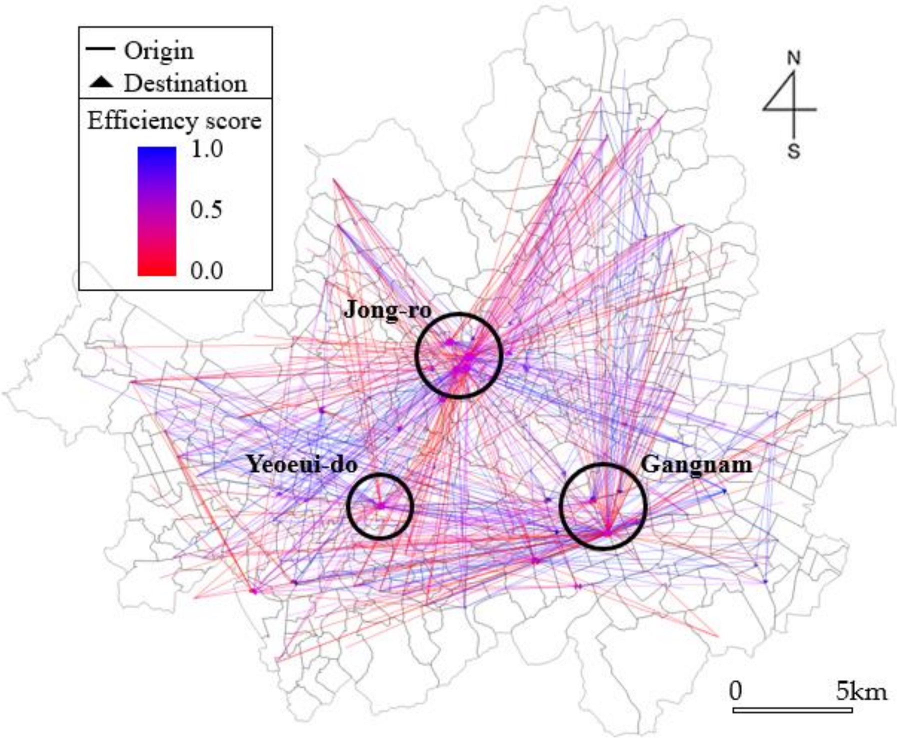


Figure 18 Result of the efficiency analysis for O–D pairs

To select the service area, the sensitivity analysis was performed using the estimated efficiency score. 782 O−D pairs were clustered into nine groups, since the nine points were estimated by the sensitivity analysis. The 19 O−D pairs of group 1 with the lowest efficiency scores were selected as the RB service areas. Since the number of efficient DMUs is 19, the bottom 19 DMUs are selected as the service area for equal comparison. The selected 19 service areas were analyzed to benchmark three efficient DMUs.

Table 11 Result of the sensitivity analysis

Group no.	Num. of DMUs	Range of efficiency score	Average efficiency score	Output variables	Input variables		
				Travel distance (km)	Waiting time (min)	In-vehicle time (min)	Transfer time (min)
1	19	0.00~0.21	0.19	10.6	6.6	37.0	11.5
2	66	0.21~0.27	0.24	10.6	4.3	34.6	6.8
3	52	0.27~0.31	0.29	10.8	3.9	33.3	3.7
4	76	0.31~0.35	0.33	11.0	3.9	33.1	2.7
5	116	0.35~0.41	0.38	11.1	3.6	30.9	2.2
6	129	0.41~0.48	0.44	11.9	3.9	31.7	1.5
7	219	0.48~0.64	0.56	12.9	3.7	33.6	1.2
8	89	0.64~0.99	0.72	14.8	4.0	36.4	0.6
9	16	0.99~1.00	1.00	12.8	3.8	30.6	0.1
Average			0.46	12.1	3.2	33.3	2.5

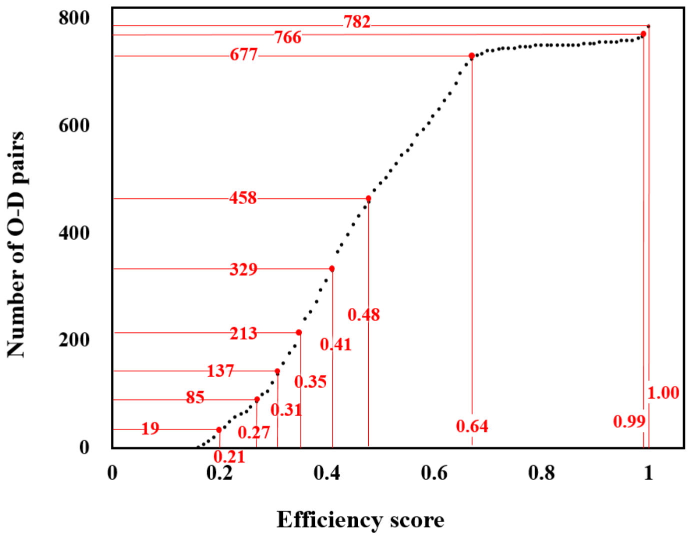


Figure 19 Sensitivity analysis for selecting service area

The results of the efficiency score of the selected O-D pairs were estimated to be 0.19. The slacks, which are the requirements to be efficiency score 1.0 for the in-vehicle, transfer, and waiting time, were estimated to be 26.0, 0.39, 2.5 minutes, respectively. The inefficient ratio for the in-vehicle, transfer, and waiting time were about 25%, 99%, and 62%. Table 12 shows the results of the efficiency score of selected O-D pairs

Table 12 Results of efficiency score of selected O-D pairs

O-D pair	Eff. score	Current			Requirements to be efficiency score 1.0		
		Inv-time	Transfer time	Waiting time	Inv-time	Transfer time	Waiting time
1	0.16	38.7	15.9	6.2	20.2	0.3	2.5
2	0.16	34.2	16.2	8.2	19.7	0.3	2.5
3	0.17	31.7	15.3	8.7	20.8	0.1	2.50
…	…	…	…	…	…	…	…
17	0.20	41.1	7.4	6.3	23.2	0.3	2.5
18	0.21	28.1	10.2	6.5	26.0	0.4	2.5
19	0.21	37.8	4.0	4.7	22.0	0.07	2.50
Average	0.19	37.0	11.5	6.6	20.8 (-25%)	0.2 (-99%)	2.5 (-62%)

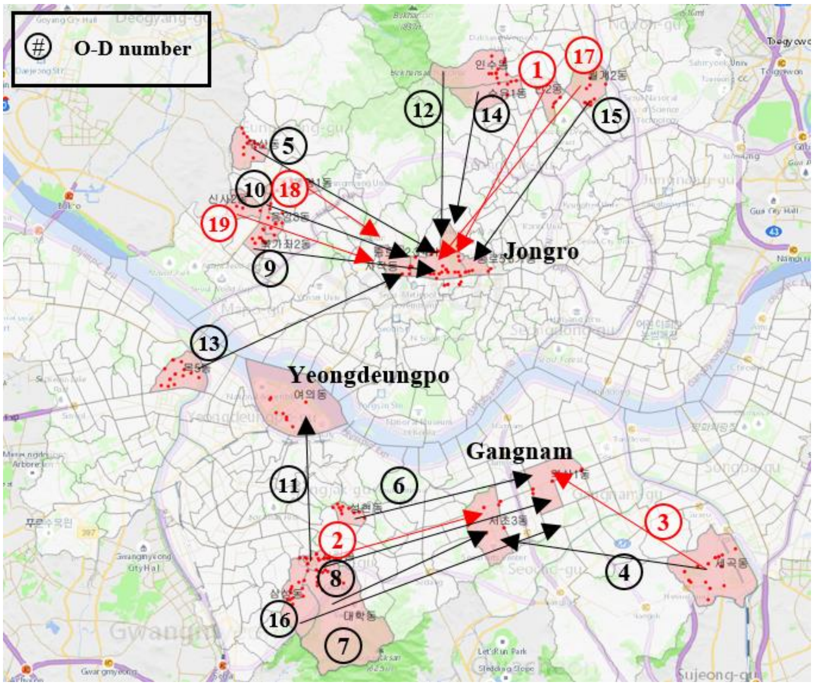


Figure 20 Results of efficiency score of selected O-D pairs

The selected O−D pairs could not be optimal O−D pairs. Therefore, the cost and revenue analysis was additionally performed to identify the profitable O−D pairs. As a result, the 112 O−D pairs were estimated to be profitable with introducing RBs. The efficiency score of 112 O−D pairs was improved from 0.29 to 0.34. in terms of the total system cost, the total system cost was saved up to 6 (/10000 KRW) with the RBs. The total expected benefit is 2 (/10000 KRW) considering operation cost and the revenue. The RC (revenue−cost ratio) was estimated to be 0.70 on average.

The top five and bottom five O−D pairs in order of system cost were compared to identify the relationship between efficiency score and total system cost. For the top 5 O−D pairs, the efficiency score and total system cost were estimated to be 0.37 and 571 (/10000 KRW), respectively. For the bottom 5 O−D pairs, the efficiency score and total system cost were estimated to be 0.17 and 114 (/10000

KRW), respectively. Although the system cost was higher in the top 5 O−D pairs, the improvement after the introduction of RB is greater in the bottom 5 O−D pairs. These results indicated that the efficiency score reasonably identified areas in need of reduced travel time.

Table 13 Results of the cost and revenue analysis

	Efficiency score			Total system cost (/10,000 KRW)			Op (D)	Rev (E)	Diff. (C'− D+E)	RC
	Cur. (A)	Pro. (B)	Diff. (C= B−A)	Cur. (A')	Pro. (B')	Diff. (C'= B'−A')				
Total: 111 O-D pairs (Mean)	0.29	0.34	0.05	225	219	6	15	11	2.0	0.70
Top 5 O-D pairs	0.35	0.38	0.029	645	635	10	15	11	6	0.72
	0.36	0.40	0.039	621	615	6	15	11	2	0.72
	0.38	0.41	0.030	557	549	8	15	11	4	0.72
	0.32	0.34	0.019	528	520	8	15	11	4	0.72
	0.41	0.52	0.117	503	495	7	15	11	3	0.72
(Mean)	0.37	0.41	0.047	571	563	8	15	11	4	0.72
Bottom 5 O-D pairs	0.17	0.24	0.075	132	116	16	15	11	0	0.72
	0.17	0.26	0.096	85	68	17	15	8	12	0.55
	0.16	0.39	0.235	91	75	16	15	9	10	0.63
	0.17	0.35	0.176	105	87	17	15	10	10	0.66
	0.18	0.24	0.060	157	138	19	15	11	12	0.72
(Mean)	0.17	0.31	0.14	114	97	17	15	10	15	0.66

Curr: current transit service
Pro.: proposed service (after the introduction of RB)
Diff: difference
Op.: operation cost (for RB) (/10,000 KRW)
Rev.: revenue (/10,000 KRW)
RC: (revenue ÷ cost) ratio

Figure 21 is a scatter plot of the efficiency score showing the relationship between the output and input variables. In Figure 21(a), a scatter plot of efficiency score illustrates the relationship between three input variables and efficiency score. The O-D pairs with low efficiency scores have the travel time in the range of 32 to 50 minutes. Overall, these inefficient O-D pairs have longer travel times, the fewer number of trips, and shorter distance compared to those of the other O-D pairs. In Figure 21(b), a scatterplot of efficiency score that represents the relationship between two output variables (travel distance, the number of trips) and one input variable (the number of transfers). Based on the scatter plot, we can observe that the efficiency score is inversely proportional to the number of transfers. The number of transfers of O-D pairs with low efficiency score is 0.34, on average, which is relatively smaller than those of the other O-D pairs. Therefore, the RB service area can be selected by comparing the efficiency scores estimated by outputs and inputs of O-D pairs.

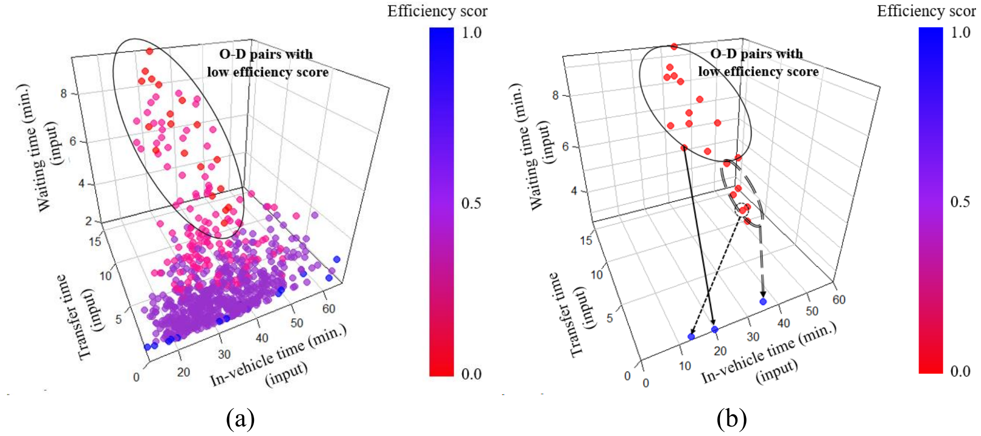


Figure 21 Illustration of the relationship between input and the efficiency score: (a) results of 782 O-D pairs; (b) results of bottom 19 O-D pairs with benchmarking O-D pairs

5.3. Design of Rapid Bus Routes

5.3.1. Estimating RB demand using MNL model

When RBs were introduced, the mode choice ratio could be varied according to the travel time, fare, and the number of transfers from origin to destination stations within the service areas. Utility parameters from the guidelines for evaluating the investment in transportation facilities (Ministry of Land, Infrastructure and Transport, 2017) were used to estimate the demand for RBs. The constant parameter of RB was assumed to be the same as that of the bus. Table 14 provides the parameters of the utility functions. The parameters were used to estimate the mode choice probability of each O−D pair in Seoul. As a result of the utility function, alternative specific constants for subway, bus, and subway + bus were estimated to be 0, −0.051, and −0.433, respectively. The parameter of travel time (β_t), fare (β_{fare}), and the number of transfers (β_{tr}) was estimated to be −0.076, −0.141, and −0.417, respectively. Since the choice utility increases as the travel time, fare, and the number of transfers decrease, the minus sign (−) of the estimated parameter of these variables shown to be reasonable. The p−value is estimated to be less than 0.01, which is statistically significant at the 99% confidence level. The pseudo R^2 was estimated to be 0.785, which is about 78.5% explanatory power. With the estimated parameters, the utility function for four modes, e.g., subway, bus, subway+bus, and RB, were developed. As proposed earlier, the fare parameter is estimated using transit and taxi fare data. The utility of RBs is calculated by bus constant, coefficients of travel time, fare, and the number of transfers.

Table 14. Utility Parameters of alternative modes

Variable	Constant	Coefficient		
		Travel time	Travel cost	Transfer time
Auto (a_{auto})	1.22179	−0.03051	−0.14217	−
Taxi (a_{taxi})	−2.08676	−0.03051	−0.14217	−
Bus (a_{bus})	0.892104	−0.03051	−0.03053	−0.20831
Subway (a_{sub})	2.34424	−0.03051	−0.03053	−
Subway + bus (a_{snb})	−	−0.03051	−0.03053	−0.20831

Pseudo R^2: 0.4874
Source: Ministry of Land, Infrastructure and Transport, 2017

5.3.2. Results of rapid bus routes

To identify the fitness convergence of the proposed model, the optimization process of GA was tested by iteration. Figure 22 illustrated the fitness convergence by iteration. By the 50th generation, the four RB routes were analyzed to have a high degree of redundancy. However, 500th generation showed the separation of the four routes. The results showed that the routes of RBs for 19 O−D pairs were optimized by minimizing fitness function as generations evolve.

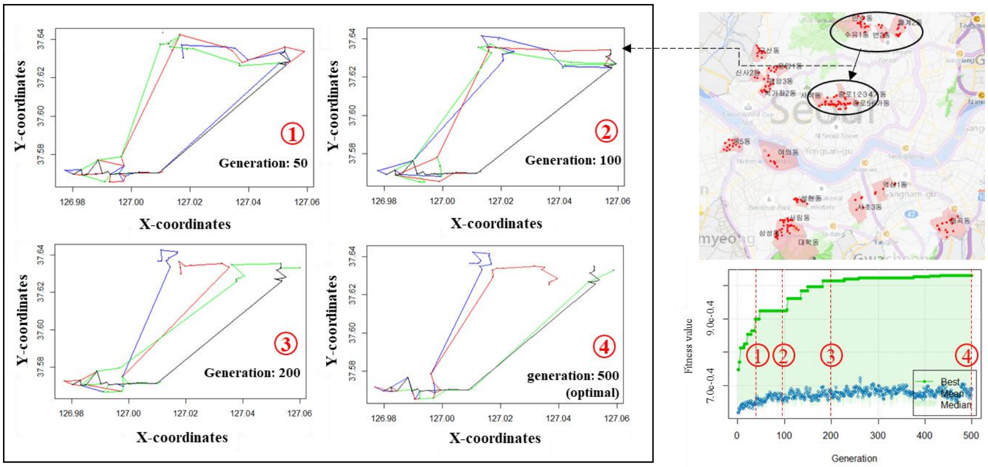


Figure 22 Fitness convergence by iterations

5.3.3. Evaluation of the public transportation service with rapid bus

With the 19 selected service areas, routes of RBs were optimized to improve the efficiency of the transit service. The results of the optimization derived the 21 vehicles and improved the efficiency score of 19 O−D pairs from 0.19 to 0.32 on average. The total demand diversion was estimated to be 825. Specifically, the demand diversions were estimated at auto and transit to RB were estimated to be 171 and 654, respectively. If only RB services were considered, the efficiency score was improved from 0.19 to 0.51 on average. Especially, the O−D pair from Segok to Yeoksam was estimated to have improved most among the 19 O−D pairs from 0.17 to 0.52. In the O−D pair from Seorim to Yeouido, the efficiency score was estimated to be the least improved among the 19 O−D pairs from 0.21 to 0.50. Although the efficiency of all O−D pairs was increased, the degree of efficiency improvement was estimated differently according to road conditions. the results of the efficiency of the RB service was shown in Figure 23 and Table 15.

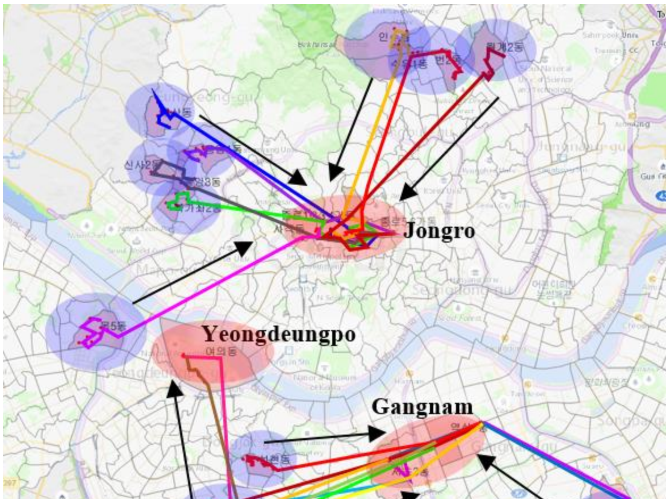
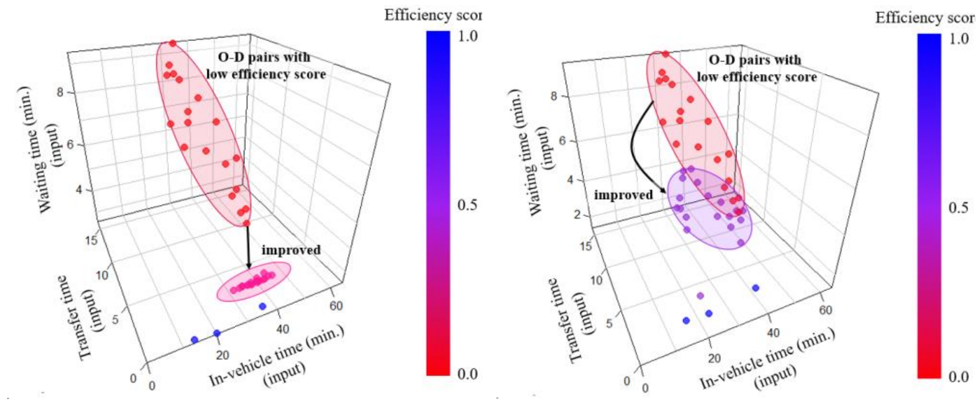


Figure 23 Results of efficiency evaluation of the RB service

Table 15 Results of efficiency evaluation of the RB service

No. O−D	Origin	Destination	Demand diversion			Efficiency score		
			Auto	Transit	Total trips	Current	proposed	
							Rapid only	Transit+RB
1	Segok	Yeoksam	8	36	44	0.17	0.52	0.24
2	Segok	Seocho	5	29	34	0.17	0.47	0.27
3	Seorim	Seocho	9	23	32	0.17	0.49	0.28
…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…
17	Bukgajwa	Jongro	8	24	35	0.21	0.48	0.30
18	Eungam	Jongro	6	22	32	0.21	0.52	0.32
19	Seorim	Yeoeuido	16	44	60	0.21	0.50	0.28
Average						0.19	0.51	0.32
Total			171	654	825	−	−	−

Figure 24 was a scatter plot of the efficiency score showing the change in the relationship between the output and input variables. In Figure 24(a), a scatter plot of the efficiency score of RB service illustrated the relationship between the efficiency score and three input variables. The O−D pairs with low efficiency scores had the travel time in the range of 30 to 45 minutes. The distance of selected 19 O−D pairs with low efficiency score, on average, were 10.6 km. Overall, these inefficient O−D pairs had longer travel times, fewer trips, and a shorter distance than those of the other O−D pairs. In Figure 24(b), a scatterplot of the efficiency score of transit service with RBs represented the relationship between efficiency score and three input variables. Based on the scatter plot, the efficiency score was inversely proportional to the number of transfers. On average, the number of transfers of O−D pairs with low efficiency score was 11.5 minutes, which was relatively smaller than those of the other O−D pairs. Therefore, the RB service area could be selected by comparing the efficiency scores estimated by outputs and inputs of O−D pairs.



(a) (b)

Figure 24 Illustration of the relationship between inputs and the efficiency score: (a) efficiency score of RB service; (b) efficiency score of RB+ transit service

To illustrate the evaluation results, 19 O-D pairs were clustered into seven spatial groups. The efficiency score was improved from 0.19 to 0.32 by introducing the RBs on average. The in-vehicle, transfer and waiting time were saved to be about 0.6, 2.0, and 0.3 minutes, respectively. Among the seven grouped O-D pairs, groups 3, 5, and 6 were estimated to be the most improved groups with an improvement in efficiency score of 0.13. Specifically, the O-D pair from Gangnam to Seocho, the efficiency score was improved from 0.19 to 0.32. The in-vehicle time, transfer time, and waiting time were saved to be about 0.9, 1.6, and 0.4 minutes, respectively. With the descriptive statics, the transfer time was the most saved compared to the other input variables, e.g., in-vehicle time and waiting time. Since the public transport system provides a transfer service with no additional charge, many transit users use the transfer service. Therefore, it was analyzed that the effect of transfer time was the greatest on the efficiency improvement.

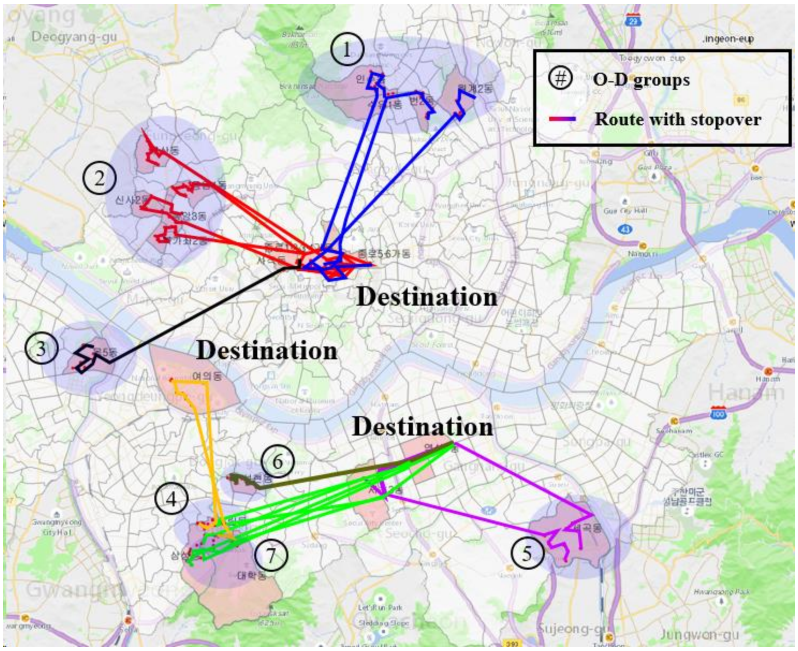


Figure 25 Results of optimal RB routes for 19 O-D pairs

Table 16 Results of efficiency evaluation of the RB service

No. O-D	Origin	Destination	Current transit				Rapid bus				Transit+rapid bus			
			Eff.	Inv	Tr	Wait	Eff.	Inv	Tr	Wait	Eff.	Inv	Tr	Wait
1	Gangbuk, Nowon	Jongro	0.18	34.0	12.8	7.7	0.67	31.8	0.0	4.0	0.29	33.9	10.4	7.0
2	Eunpyeong, Seodaemoon	Jongro	0.19	40.8	10.5	6.2	0.47	35.5	0.1	4.0	0.31	38.8	8.3	5.8
3	Yangcheon	Jongro	0.19	35.0	6.4	6.5	0.73	32.8	0.0	4.0	0.32	34.7	5.1	5.9
4	Gwanak	Yeongdeungpo	0.18	38.0	15.3	6.8	0.48	34.8	0.1	4.0	0.30	37.5	12.7	6.3
5	Gangnam	Gangnam, Seocho	0.19	33.4	8.5	5.8	0.38	28.8	0.1	4.0	0.32	32.5	6.9	5.4
6	Gwanak	Gangnam	0.19	31.7	15.3	8.7	0.39	25.9	0.2	4.0	0.26	32.5	12.6	7.8
7	Gwanak	Gangnam, Seocho	0.17	36.7	13.2	6.4	0.48	30.8	0.1	4.0	0.29	35.4	10.8	6.0
Average			0.19	37.0	11.5	6.6	0.51	30.6	0.1	4.0	0.32	36.4	9.5	6.3
Eff.: efficiency score; Inv.: in-vehicle time; Tr.: transfer time; Wait: waiting time														

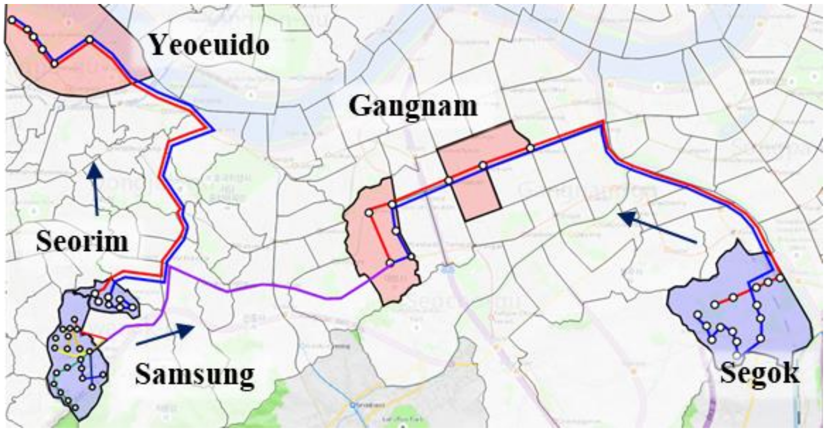
5.3.4. Comparison analysis for validating proposed routing problem

Table 17 shows the results of the RB routes results of the proposed model. Three cases were selected as the comparison areas, e.g., Segok–Gangnam, Seorim–Yeoeui, and Samsung–Yeoksam.

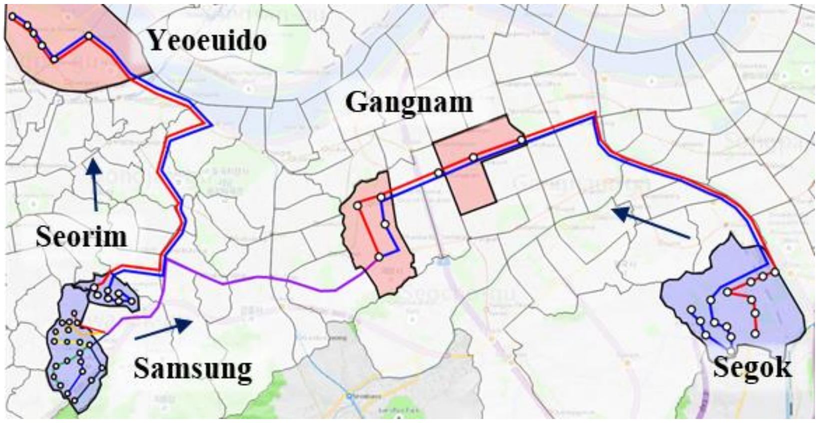
In the case of Segok–Gangnam, the number of trips was 78, and the number of vehicles was estimated to be 2. As a result of the conventional model (one by one), the average operation distance was estimated to be 13.2 km, and the average user travel time was estimated to be about 37.0 minutes on average. As a result of the proposed model (multiple vehicles concurrently), the average operation distance was estimated to be 12.7 km, and the average user travel time was estimated to be about 35.1 minutes on average. By comparing two models, the proposed model showed the less value on 0.5 km of the operation distance and 1.9 minutes of user travel time, respectively. In the case of Serim–Gangnam, the number of trips was 75, and the number of vehicles was estimated to be 2. As a result of the conventional model (one by one), the average operation distance was estimated to be 16.5 km, and the average user travel time was estimated to be about 45.1 minutes on average. As s result of the proposed model (multiple vehicles concurrently), the average operation distance was estimated to be 16.1 km, and the average user travel time was estimated to be about 43.5 minutes on average. By comparing two models, the proposed model showed the less value on 0.4 km of operation distance and 1.6 minutes of user travel time, respectively. In the case of Samsung–Yeoksam, the number of trips was 170, and the number of vehicles was 4. As a result of the conventional model (one by one), the average operation distance was estimated to be 13.1 km, and the average user travel time was estimated to be about 53.0 minutes on average. As a result of the proposed model (multiple vehicles concurrently), the average operation distance was estimated to be 12.9 km, and the average user travel time was estimated to be about 50.7 minutes on average. By comparing two models, the proposed model showed the less value on 0.2 km of operation distance and 2.3 minutes of user travel time, respectively. Overall, the proposed model showed less value on 0.3 km of operation distance and 2.1 minutes of user travel time, respectively.

Table 17 Comparison results of the proposed model (bus routing problem)

No	Origin	Desti-nation	Number of trips	Number of vehicles	Conventional (one by one)		Proposed (multiple vehicles concurrently)	
					Average operation distance (km)	Average user travel time (min)	Average operation distance (km)	Average user travel time (min)
1	Segok	Gangnam	78	2	13.2	37.0	12.7	35.1
2	Seorim	Yeoeui	75	2	16.5	45.1	16.1	43.5
3	Samsung	Yeoksam	170	4	13.1	53.0	12.9	50.7
Total			323	8	111.8	15,279	109.2	14,619
Average			-	-	14.0	47.3	13.7	45.2



(a)



(b)

Figure 26 Comparison results of the proposed model (bus routing problem)

5.3.5. Comparison analysis for validating proposed service area selection and optimal routing problem

To validate the proposed model, the comparison analysis was performed with models based on other selection criteria. In previous studies, there were several criteria of service area selection, e.g., high demand oriented, long out of vehicle oriented, and long travel time oriented models. Specifically, the high demand oriented model selected the service area where demand is high. The high demand oriented model aims to maximize the profit or benefit when introducing the new transit lines. The long out of vehicle time oriented model selected the service area where the transfer and waiting time was long. The long out of vehicle time oriented model aims to save the waiting time and transfer time by connecting the origin and destination directly without transferring between modes. The long in−vehicle time model selected the service area with many detours or stops. The long in−vehicle time model aims to save the in−vehicle time by simplifying the routes and skipping stops. In this study, the proposed model was compared to the other three models mentioned above.

Table 18 shows the results of the comparison analysis with high demand oriented model. As a result of the high demand oriented model, the in−vehicle and waiting time were increased by 0.35 and 2.53 minutes, respectively. The transfer time only was decreased from 0.41 to 0.39. The efficiency score was decreased from 0.65 to 0.63. The results of the proposed model, the in−vehicle transfer and waiting time were decreased as much as 1.12, 0.32, and 0.93 minutes, respectively. The efficiency score was decreased from 0.16 to 0.44.

O−D pairs for each model were selected to understand the results of the model. O−D pair from Seorim to Seocho was selected for the proposed model, and O−D pair from Sillim to Seocho was selected for the high demand oriented model.

In Figure 27, the blue line and red line illustrated the selected O−D pairs of the high demand oriented and proposed model, respectively. As a result of the high demand oriented model, the efficiency score of the O−D pair from Sillim to Seocho was decreased. Since the subway line connects Sillim−Gangnam O−D, it was analyzed that the introduction of RB rather impeded the transit efficiency. The efficiency score decreased as demand diverted to RB, which has a slower travel time than the current public transport service. On the other hand, the red line connects the O−D pair selected by the proposed model. Since users are required to experience transfers and detours, the efficiency score was estimated to low. This result implied that service areas with high demand might already be well−equipped with public transport infrastructure. In terms of excessive investment, there is no need to introduce an additional RB route.

Table 18 Results of the comparison analysis (high demand oriented model)

Selection criteria	Analysis results				
	Mode	In−vehicle time (minutes)	Transfer time (minutes)	Waiting time (minutes)	Efficiency score
High demand oriented	Current transit	24.82	0.41	2.53	0.65
	Transit + RB	25.17	0.39	2.95	0.63
	Difference	0.35	-0.02	0.42	-0.02
Low efficiency score (proposed)	Current transit	31.73	15.34	8.66	0.16
	Transit + RB	30.78	12.77	7.87	0.44
	Difference	-1.12	-3.02	-0.93	0.28

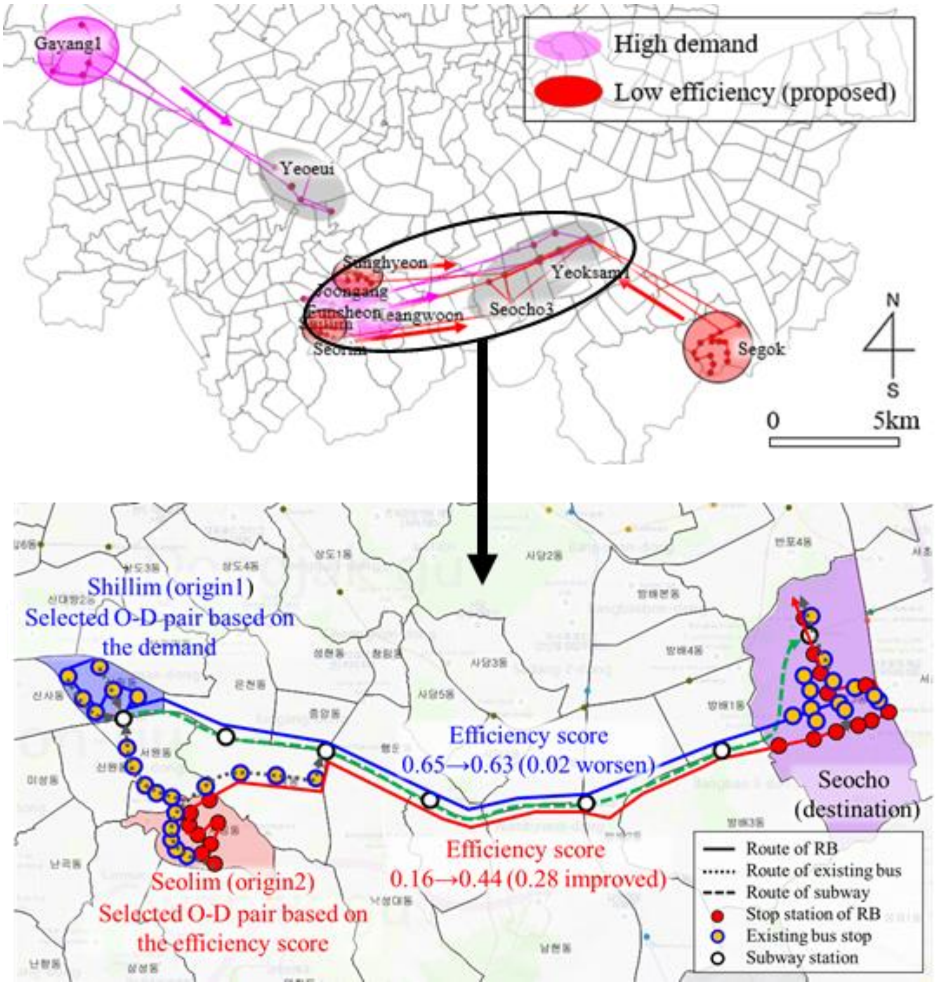


Figure 27 Results of the comparison analysis (high demand oriented model)

Table 19 shows the results of the comparison analysis with long out of vehicle time oriented model. The in−vehicle, transfer, and waiting time of the out of vehicle time oriented model, were saved as 0.93, 3.22, and 0.23 minutes, respectively. The efficiency score was improved from 0.27 to 0.32. The results of the proposed model, the in−vehicle, transfer, and waiting time were decreased as 1.12, 3.01, and 0.93 minutes, respectively. The efficiency score was improved from 0.16 to 0.44. The difference in efficiency score of the long out of vehicle time oriented and proposed models was estimated to be 0.04 and 0.28, respectively.

In Figure 28, the blue line and red line illustrated the selected O−D pairs of the long out of vehicle time oriented and proposed model, respectively. As a result of the long out of vehicle time oriented model, the efficiency score of the O−D pair from Doksan to Seocho was increased. Since the O−D pair from Doksan to Seocho was selected based on the long out of vehicle time, it can be seen that the reduction ratio of the waiting time is large. The efficiency score of the O−D pair from Seorim to Seocho, which was selected by the proposed model, was also increased.

Overall, the efficiency score of the long out of vehicle time oriented model improved. However, the improvement of efficiency score was less than that of the proposed model since only out of vehicle time variable was considered.

Table 19 Results of the comparison analysis (out of vehicle time oriented model)

<table>
<tr><th rowspan="3">Selection criteria</th><th colspan="5">Analysis results</th></tr>
<tr><th rowspan="2">Mode</th><th rowspan="2">In−vehicle time (min.)</th><th colspan="2">Out of vehicle time (min.)</th><th rowspan="2">Efficiency score</th></tr>
<tr><th>Transfer time</th><th>Waiting time</th></tr>
<tr><td rowspan="3">Long out of vehicle time oriented</td><td>Current transit (A)</td><td>43.75</td><td>15.74</td><td>5.45</td><td>0.27</td></tr>
<tr><td>Transit + RB (B)</td><td>42.82</td><td>12.52</td><td>5.22</td><td>0.32</td></tr>
<tr><td>Difference</td><td>−0.93</td><td>−3.22</td><td>−0.23</td><td>0.04</td></tr>
<tr><td rowspan="3">Low efficiency score (proposed)</td><td>Current transit</td><td>31.73</td><td>15.34</td><td>8.66</td><td>0.16</td></tr>
<tr><td>Transit + RB</td><td>30.78</td><td>12.77</td><td>7.87</td><td>0.44</td></tr>
<tr><td>Difference</td><td>−1.12</td><td>−3.01</td><td>−0.93</td><td>0.28</td></tr>
</table>

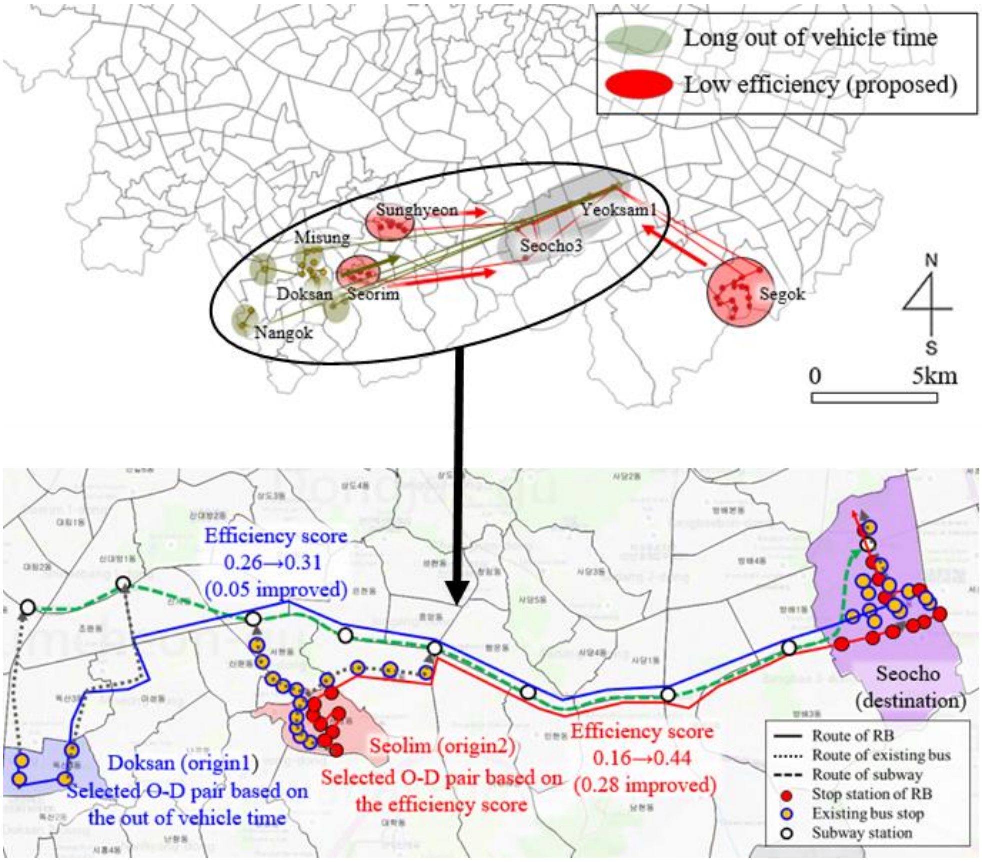


Figure 28 Results of the comparison analysis (out of vehicle time oriented model)

Table 20 shows the results of the comparison analysis with the travel time oriented model. The in−vehicle, transfer, and waiting time of the travel time oriented model, were saved as 1.62, 1.36, and 0.05 minutes, respectively. The efficiency score was improved from 0.26 to 0.31. The results of the proposed model, the in−vehicle, transfer, and waiting time were decreased as 1.35, 2.84, and 0.40 minutes, respectively. The efficiency score was improved from 0.16 to 0.39.

In Figure 29, the blue line and red line illustrated the selected O−D pairs of the travel time oriented and proposed model, respectively. As a result of the travel time oriented model, the efficiency score of the O−D pair from Jeongleung to Jongro was increased. Since the O−D pair from Jeongleung to Jongro was selected based on the travel time, it could be seen that the reduction ratio of the in−vehicle time was large. The efficiency score of the O−D pair from Bun2 to Jongro, which was selected by the proposed model, was also increased. Overall, the efficiency score of the travel time oriented model improved. However, the improvement of efficiency score was less than that of the proposed model since only the travel time variable was considered.

Table 20 Results of the comparison analysis (travel time oriented model)

Selection criteria	Analysis results				
	Mode	In−vehicle time (min.)	Transfer time (min.)	Waiting time (min.)	Efficiency score
Long travel time oriented	Current transit (A)	41.10	6.81	4.44	0.26
	Transit + RB (B)	39.48	5.45	4.35	0.31
	Difference	−1.62	−1.36	−0.09	0.05
Low efficiency score (proposed)	Current transit (A)	39.73	15.86	6.22	0.16
	Transit + RB (B)	38.38	13.02	5.82	0.39
	Difference	−1.35	−2.84	−0.40	0.23

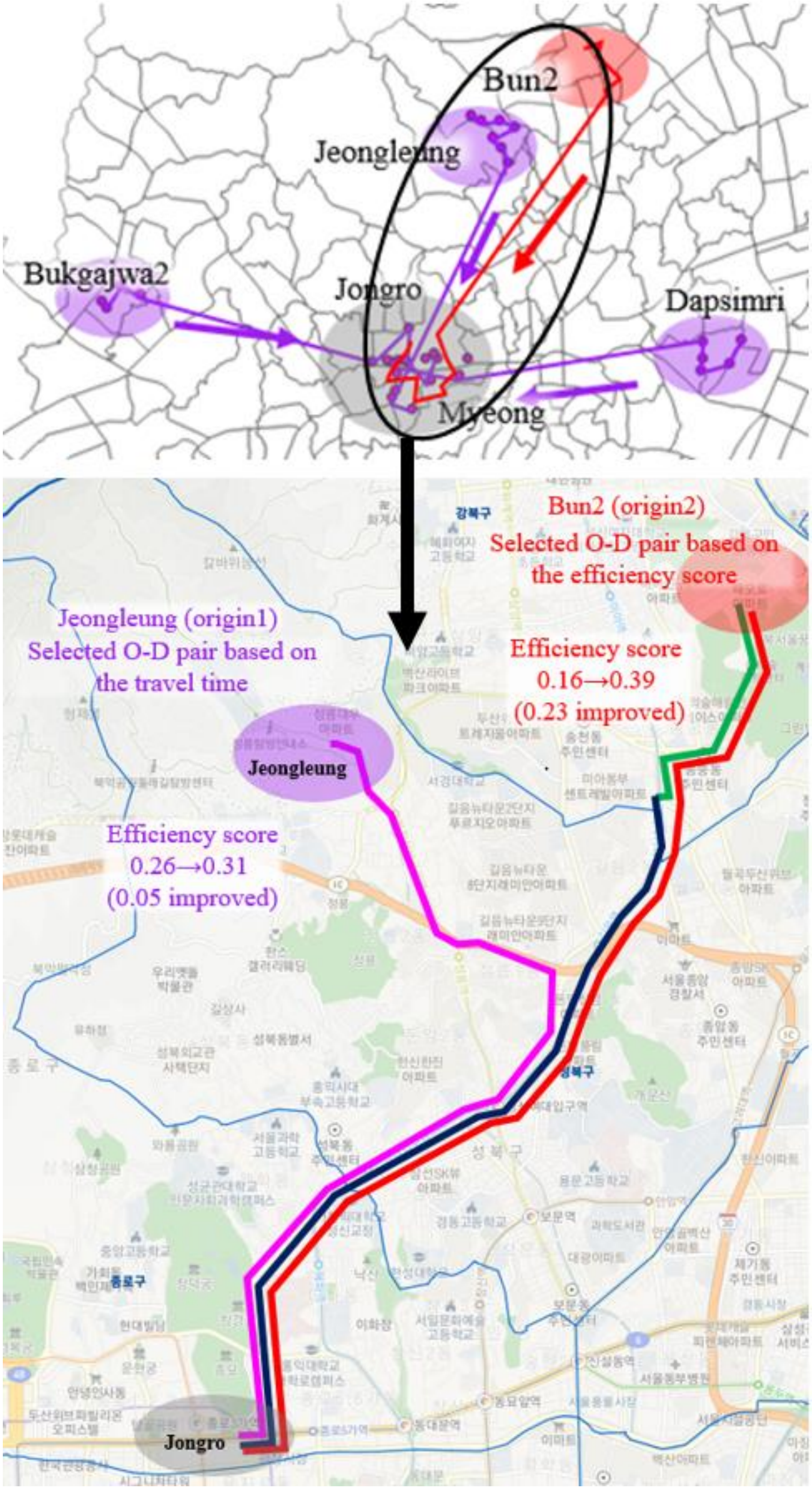


Figure 29 Results of the comparison analysis (travel time oriented model)

To identify the performance in terms of the cost and revenue of the RB routes, the total system cost, operation cost and revenue were calculated. First, in the high demand oriented model, the efficiency score decreased by 0.02 points on average. The total system cost of the high demand oriented model was calculated to be 1,188 (/10,000 KRW). After introducing the RB, the total system cost was increased to be 1,201(/10,000 KRW). The benefit was estimated to be −32 (/10,000 KRW) regarding the total system cost, operation cost, and revenue. Since the high demand oriented model selected the O−D pairs where the public transport infrastructures were well− equipped, both efficiency and benefit were decreased.

The total system cost of the long out of vehicle time oriented model was calculated to be 66 (/10,000 KRW). After introducing the RB, the total system cost was decreased by 10 and was estimated to be 56 (/10,000 KRW). The benefit was estimated to be 2 (/10,000 KRW) regarding the total system cost, operation cost, and revenue. The total system cost of the long travel time oriented model was calculated to be 91 (/10,000 KRW). After introducing the RB, the total system cost was decreased by 6 and was estimated to be 85 (/10,000 KRW). Regarding the total system cost, operation cost and revenue, the benefit was estimated to be −3 (/10,000 KRW).

The total system cost of the proposed efficiency oriented model was calculated to be 114 (/10,000 KRW). After introducing the RB, the total system cost was decreased by 17 and was estimated to be 97 (/10,000 KRW). Regarding the total system cost, operation cost and revenue, the benefit was estimated to be 15 (/10,000 KRW). Since the efficiency oriented model considers all factors such as waiting time, in−vehicle time, and transfer time, it showed the best performance among the comparison models.

Table 21 Results of the cost and revenue analysis

	Efficiency score			Total system cost (/10,000 KRW)			Op (D)	Rev (E)	Diff. (C'−D+E)	RC
	Cur. (A)	Pro. (B)	Diff. (C=B−A)	Cur. (A')	Pro. (B')	Diff. (C'=B'−A')				
High demand oriented model	0.65	0.63	−0.026	1,696	1,710	−14	89	57	−46	0.64
	0.63	0.60	−0.027	1,459	1,475	−16	89	56	−46	0.63
	0.61	0.59	−0.015	1,262	1,275	−13	75	48	−49	0.65
	0.54	0.54	−0.007	942	950	−8	45	32	−40	0.71
	0.57	0.56	−0.015	583	597	−14	45	28	−21	0.62
(Mean)	0.60	0.58	−0.02	1,188	1,201	−13	69	44	−32	0.65
Long out of vehicle time oriented model	0.27	0.34	0.070	55	46	9	15	6	−37	0.43
	0.27	0.32	0.049	65	56	9	15	6	1	0.38
	0.26	0.31	0.052	91	79	12	15	9	0	0.60
	0.26	0.30	0.043	47	37	11	15	5	6	0.35
	0.32	0.54	0.218	71	60	11	15	6	1	0.39
(Mean)	0.28	0.36	0.08	66	56	10	15	6	2	0.43
Long travel time oriented model	0.34	0.35	0.007	126	118	8	15	10	2	0.68
	0.38	0.39	0.007	83	77	6	15	7	3	0.48
	0.32	0.35	0.038	85	78	7	15	9	−2	0.58
	0.40	0.41	0.015	61	54	7	15	7	1	0.47
	0.27	0.30	0.034	101	97	4	15	8	−1	0.55
(Mean)	0.34	0.36	0.02	91	85	6	15	8	−3	0.55
Efficiency oriented model (Proposed)	0.17	0.24	0.075	132	116	16	15	11	0	0.72
	0.17	0.26	0.096	85	68	17	15	8	12	0.55
	0.16	0.39	0.235	91	75	16	15	9	10	0.63
	0.17	0.35	0.176	105	87	17	15	10	10	0.66
	0.18	0.24	0.060	157	138	19	15	11	12	0.72
(Mean)	0.17	0.31	0.14	114	97	17	15	10	15	0.66

Curr: current transit service
Pro.: proposed service (after the introduction of RB)
Diff: difference
Op.: operation cost (for RB) (/10,000 KRW)
Rev.: revenue (/10,000 KRW)
RC: (revenue ÷ cost) ratio

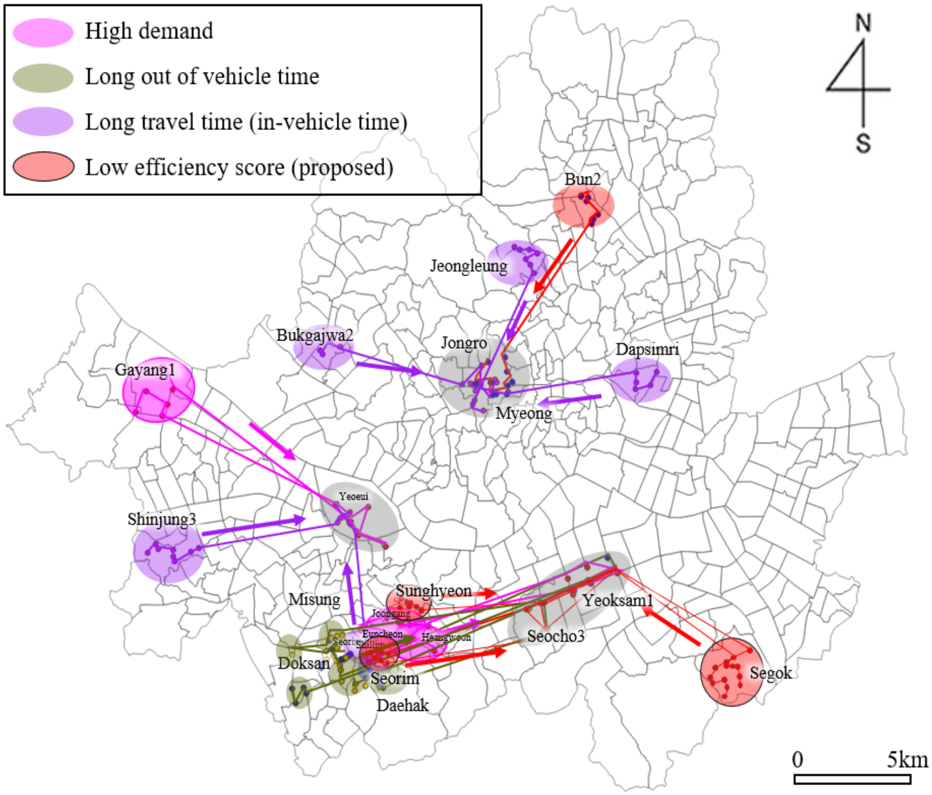


Figure 30 Results of the comparison analysis (overall)

Chapter 6. Conclusion

This study proposed RB routing strategy for commuters, including both service area selection and vehicle route optimization. Specifically, the service area, such as the O−D pair, was selected based on the efficiency score from the DEA model. With the selected O−D pairs, the number of vehicles and the operating routes was optimized using the GA. The strategy was developed to minimize the total travel time with a minimum number of vehicles.

The proposed strategy was applied to real−world cases, i.e., the public transportation system in Seoul, South Korea, and it was evaluated quantitatively. First, the efficiency evaluation was performed on 782 O−D pairs, and 19 O−D pairs were selected as a service area with a low efficiency score. Second, the route optimization was performed to selected O−D pairs. As a result, the efficiency score was improved from 0.19 to 0.32 by introducing the RBs on average. To validate the proposed model, the comparative analysis was also performed with models based on other selection criteria. As results, the proposed model showed the best performance among the comparison models. Since the efficiency oriented model considers all factors such as waiting time, in−vehicle time, and transfer time,

This study proposes a framework to develop RB routing strategies considering both service area selection and vehicle route optimization with efficiency evaluation. There are several issues that merit future investigations. This study used the DEA model to select the RB service area. With the results of the DEA, we selected two O−D pairs as the service area with the lowest score. Although the two service areas were selected based only on the efficiency scores in this study, they could have been selected differently according to the policy direction, mobility environment, and other O-D characteristics. For example, if RB aims to reduce the number of transfers, the service area can be selected as an O-D pair with a high number of transfers and a low efficiency score. Similarly, various additional and detailed data would be required, such as sociodemographic factors, socioeconomic factors, and other mobility modes to select the service area using the DEA model. Further research is on-going to consider these aspects in the process of optimal routing strategy by the authors.

Reference

1. Barnum, D. T., McNeil, S., & Hart, J. (2007). Comparing the efficiency of public transportation subunits using data envelopment analysis. Journal of Public Transportation, 10(2), 1.
2. Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management science, 30(9), 1078-1092.
3. Calthorpe, P. (1993). The next American metropolis: Ecology, community, and the American dream. Princeton architectural press.
4. Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European journal of operational research, 2(6), 429-444.
5. Chen, X., Wang, Y., & Ma, X. (2018, July). Customized bus line design model based on multi-source data. In International Conference on Transportation and Development 2018: Traffic and Freight Operations and Rail and Public Transit (pp. 218-228). Reston, VA: American Society of Civil Engineers.
6. Choi, J., Lee, Y. J., Kim, T., & Sohn, K. (2012). An analysis of Metro ridership at the station-to-station level in Seoul. Transportation, 39(3), 705-722.
7. Eom, J. K., Song, J. Y., & Moon, D. S. (2015). Analysis of public transit service performance using transit smart card data in Seoul. KSCE Journal of civil engineering, 19(5), 1530-1537.
8. Farrel, M. J. (1957). The Measurement of Productive Efficiency," Journal of the Royal Statistical Society, Series A, Part 3.
9. Farrell, M. J., & Fieldhouse, M. (1962). Estimating efficient production functions under increasing returns to scale. Journal of

the Royal Statistical Society: Series A (General), 125(2), 252-267.

10. Florida, D. O. T. (2000). The Florida Reliability Method in Florida's Mobility Performance Measures Pr
11. ogram. Florida Department of Transportation.
12. Golani, H. (2007). Use of archived bus location, dispatch, and ridership data for transit analysis. Transportation research record, 1992(1), 101-112.
13. Guo, R., Guan, W., & Zhang, W. (2018). Route design problem of customized buses: Mixed integer programming model and case study. Journal of Transportation Engineering, Part A: Systems, 144(11), 04018069.
14. Guo, Z., & Wilson, N. H. (2007). Modeling effects of transit system transfers on travel behavior: case of commuter rail and subway in Downtown Boston, Massachusetts. Transportation Research Record, 2006(1), 11-20.
15. Hahn, J. S., Kho, S. Y., Choi, K., & Kim, D. K. (2017). Sustainability evaluation of rapid routes for buses with a network DEA model. International Journal of Sustainable Transportation, 11(9), 659-669.
16. Hahn, J. S., Kim, D. K., Kim, H. C., & Lee, C. (2013). Efficiency analysis on bus companies in Seoul city using a network DEA model. KSCE Journal of Civil Engineering, 17(6), 1480-1488.
17. Hahn, J. S., Sung, H. M., Park, M. C., Kho, S. Y., & Kim, D. K. (2015). Empirical evaluation on the efficiency of the trucking industry in Korea. KSCE Journal of Civil Engineering, 19(4), 1088-1096.
18. Khoshnevis, P., & Teirlinck, P. (2018). Performance evaluation of R&D active firms. Socio-Economic Planning Sciences, 61, 16-

28.
19. Kittelson & Associates, United States. Federal Transit Administration, Transit Cooperative Research Program, & Transit Development Corporation. (2003). Transit capacity and quality of service manual (Vol. 42). Transportation Research Board.
20. Laporte, G. (2009). Fifty years of vehicle routing. Transportation science, 43(4), 408-416.
21. Lee, E. H., Lee, I., Cho, S. H., Kho, S. Y., & Kim, D. K. (2019). A travel behavior-based skip-stop strategy considering train choice behaviors based on smartcard data. Sustainability, 11(10), 2791.
22. Lee, E. H., Lee, H., Kho, S. Y., & Kim, D. K. (2019). Evaluation of transfer efficiency between bus and subway based on data envelopment analysis using smart card data. KSCE Journal of Civil Engineering, 23(2), 788-799.
23. Lee, E. H., Shin, H., Cho, S. H., Kho, S. Y., & Kim, D. K. (2019). Evaluating the Efficiency of Transit-Oriented Development Using Network Slacks-Based Data Envelopment Analysis. Energies, 12(19), 3609.
24. Lee, H., Park, H. C., Kho, S. Y., & Kim, D. K. (2019). Assessing transit competitiveness in Seoul considering actual transit travel times based on smart card data. Journal of Transport Geography, 80, 102546.
25. Lyu, Y., Chow, C. Y., Lee, V. C., Ng, J. K., Li, Y., & Zeng, J. (2019). CB-Planner: A bus line planning framework for customized bus systems. Transportation Research Part C: Emerging Technologies, 101, 233-253.
26. Meng, W., Hu, Z., & Liu, W. (2006). Efficiency evaluation of basic

research in China. Scientometrics, 69(1), 85-101.

27. Metropolitan Transportation Authority. (2015). Joint business of current O/D in the metropolitan area in 2014, http://www.mta.go.kr/2015mta_renew/condition/passResearch_business.jsp

28. Nishiuchi, H., Todoroki, T., & Kishi, Y. (2015). A fundamental study on evaluation of public transport transfer nodes by data envelop analysis approach using smart card data. Transportation Research Procedia, 6, 391-401.

29. Pattnaik, S. B., Mohan, S., & Tom, V. M. (1998). Urban bus transit route network design using genetic algorithm. Journal of transportation engineering, 124(4), 368-375.

30. Pillac, V., Gendreau, M., Guéret, C., & Medaglia, A. L. (2013). A review of dynamic vehicle routing problems. European Journal of Operational Research, 225(1), 1-11.

31. Qiu, G., Song, R., He, S., Xu, W., & Jiang, M. (2018). Clustering passenger trip data for the potential passenger investigation and line design of customized commuter bus. IEEE Transactions on Intelligent Transportation Systems, 20(9), 3351-3360.

32. Renne, J. L. (2005). Transit-oriented development: Developing a strategy to measure success (No. 294). Transportation Research Board.

33. Rezaee, M. J., Izadbakhsh, H., & Yousefi, S. (2016). An improvement approach based on DEA-game theory for comparison of operational and spatial efficiencies in urban transportation systems. KSCE Journal of Civil Engineering, 20(4), 1526-1531.

34. Sheffi, Y. (1985). Urban Transportation Networks Prentice-Hall. Englewood Cliffs, NJ.

35. Singh, Y. J., Fard, P., Zuidgeest, M., Brussel, M., & van Maarseveen, M. (2014). Measuring transit oriented development: a spatial multi criteria assessment approach for the City Region Arnhem and Nijmegen. Journal of Transport Geography, 35, 130-143.
36. Sohn, K., & Shim, H. (2010). Factors generating boardings at metro stations in the Seoul metropolitan area. Cities, 27(5), 358-368.
37. Song, J. Y., Eom, J. K., Lee, K. S., Min, J. H., & Yang, K. Y. (2015). Public transportation service evaluations utilizing seoul transportation card data. Procedia Computer Science, 52, 178-185.
38. Spiess, H., & Florian, M. (1989). Optimal strategies: a new assignment model for transit networks. Transportation Research Part B: Methodological, 23(2), 83-102.
39. Sung, H., & Oh, J. T. (2011). Transit-oriented development in a high-density city: Identifying its association with transit ridership in Seoul, Korea. Cities, 28(1), 70-82.
40. Sun, L., Rong, J., & Yao, L. (2010). Measuring transfer efficiency of urban public transportation terminals by data envelopment analysis. Journal of urban planning and development, 136(4), 314-319.
41. Utsunomiya, M., Attanucci, J., & Wilson, N. (2006). Potential uses of transit smart card registration and transaction data to improve transit planning. Transportation research record, 1971(1), 118-126.
42. Wang, J., Yamamoto, T., & Liu, K. (2019). Role of customized bus services in the transportation system: insight from actual performance. Journal of Advanced Transportation, 2019.

43. Yu, X., Lang, M., Gao, Y., Wang, K., Su, C. H., Tsai, S. B., ... & Li, S. (2018). An empirical study on the design of China high-speed rail express train operation plan—From a sustainable transport perspective. Sustainability, 10(7), 2478.
44. Zhao, J., Jiang, Y., & Zhang, X. (2015). Research on transportation efficiency evaluation based on DEA model. In CICTP 2015 (pp. 2364-2375).
45. Zheng, H., Chen, J., Zhang, X., & Yang, Z. (2019). Designing a New Shuttle Service to Meet Large-Scale Instantaneous Peak Demands for Passenger Transportation in a Metropolitan Context: A Green, Low-Cost Mass Transport Option. Sustainability, 11(18), 5025.

국문 초록

대중교통 효율성을 고려한 신규 급행버스 노선 설계

서울대학교 대학원
공과대학 건설환경공학부
이 은 학

대중교통 이용이 증가하면서 보다 편리하고 신속한 대중교통 서비스가 요구되고 있다. 급행버스는 특정 기종점을 직결 운행하는 버스 노선으로 운영 효율과 이용자 편의성 제고 측면에서 각광 받는 수단이다. 본 연구는 대중교통 효율성을 고려한 신규 급행버스 노선을 설계하는 것을 목표로 한다. 신규 급행버스 도입을 위해서 서비스 지역 선정과 노선 최적화 단계로 모델 프레임워크를 개발했다.

먼저, 서비스 지역 선정 단계는 대중교통 효율성의 개선이 필요한 지역을 선정하는 단계이다. 대중교통 효율성은 기종점의 최단 도로거리 대비 대중교통 통행시간(대기시간, 차내시간, 환승시간)으로 설정했다. 대중교통 효율성 평가를 위해서 자료포락분석(DEA: data envelopment) 모델을 개발했으며, 투입지향 VRS(variable returns to scale) 가정을 채택했다. DEA 모형의 산출변수는 도로거리, 투입변수는 대기시간, 차내시간, 환승시간으로 설정했다. DEA 모형은 기종점간의 상대적인 효율성 평가가 가능하고 투입변수에 대한 개선 방향과 크기를 파악할 수 있다는 장점을 지닌다.

두번째 단계는 노선 최적화 단계로 유전자 알고리즘(GA: genetic algorithm) 기반의 차량경로문제(VRP: vehicle routing problem) 모델을 개발했다. 최적화 모형의 결정변수는 버스 차량대수, 운행노선으로 설정했으며, 여러 차량의 노선 조합들을 고려한 최적해를 도출하기 위해 GA 구조를 수정했다. 목적함수는 총비용 최소화로

설정했으며, 대기시간, 차내시간, 환승시간의 값이 총비용 연산에 포함되어 효율성이 개선될 수 있도록 모델을 설계했다.

서울시 대중교통망 네트워크를 대상으로 모형을 적용했다. 서비스 지역 선정 단계에서는 서울시 행정동 단위 기종점, 노선 최적화 단계에서는 기존 대중교통 정류장 단위로 분석을 수행했다. 분석자료는 2017년 스마트카드 및 지리정보시스템 데이터를 이용했다. 첫번째 단계인 효율성 분석 결과, 782개의 행정동 단위 기종점의 평균 효율성 점수는 0.46로 분석되었다. 효율적인 기종점은 19개로 효율성 점수는 1.00으로 분석되었다. 비효율적인 기종점 19개의 효율성 점수는 0.19로 분석되었으며, 통행거리 10.6km, 대기시간 6.6분, 차내시간 37.0분, 환승시간 11.5분 나타났다. 비효율적인 19개 기종점은 효율적인 기종점과 비교하여 평균적으로 짧은 거리를 더욱 오래 통행하는 것으로 분석되었다. 두번째 단계에서는 서비스 지역 선정 단계에서 선정된 19개 기종점을 대상으로 급행버스 노선 최적화를 수행했다. 최적화 결과, 21개의 버스 노선이 생성되었으며 825명의 수요를 수송하는 것으로 분석되었다. 신규 급행버스 도입 결과, 급행버스 이용자의 효율성 점수는 기존 0.19에서 0.51로 증가했으며, 기존 대중교통 이용자까지 고려하는 경우, 효율성 점수는 0.32로 증가했다. 개발된 모형의 성능을 평가하기 위해서 3개의 기존 모형(통행량, 차내시간, 차외시간 모형)의 결과와 비교 분석을 수행했다. 제안한 효율성 기반 모형은 효율성 및 총비용 측면에서 가장 크게 개선될 수 있는 노선들을 선정하였다. 특히, 3개의 기존 모형은 통행량 또는 특정 통행시간 요소를 기반으로 급행버스를 도입한 반면, 제안한 효율성 기반 모형은 대기시간, 차내시간, 환승시간을 모두 고려하기 때문에 전체적인 통행시간 절감효과 측면에서 가장 우수한 효과를 나타냈다.

주요어 : 대중교통 노선망 설계, 차량경로문제, 급행버스, 자료포락분석, 효율성

학번 : 2016-21263