공학석사학위논문

승객의 선호를 고려한 호기반 확률적 통행배정에 관한 연구

A study on link-based stochastic algorithm for transit assignment considering passengers' preferences

2015년 8월

서울대학교 대학원
산업공학과
황 준 하

승객의 선호를 고려한 호기반 확률적 통행배정에 관한 연구

A study on link-based stochastic algorithm for transit assignment considering passengers' preferences

지도교수 홍 성 필

위 논문을 공학석사 학위논문으로 제출함.

2015년 6월

서울대학교 대학원

산업공학과

황 준 하

황준하의 공학석사 학위논문을 인준함.

2015년 6월

위원장	박 진 우	(인)
부위원장	홍 성 필	(인)
위 원	이 경 식	(인)

ABSTRACT

Junha Hwang
Department of Industrial Engineering and Naval Architecture
The Graduate School
Seoul National University

In transit assignment problem, it is assumed that passengers choose path which is, they consider, the best among the alternative paths. Thus, path choice of passengers reveals their taste heterogeneity, and generating path choice set is an important part of solving transit assignment problem. In this paper, we discuss how to make path choice set efficiently. The network of Seoul Metro consists of a large number of nodes and links; in some area, it has a grid structure. Due to this, a method which does not require path enumeration should be used; Dial's algorithm is the one. In order to generate path choice set more accurately, it is essential to use a precise link cost function. To develop this, we should take crowding and taste heterogeneity into account. We show that the new link cost estimates path choice set more accurately than the previously used link cost.

Keywords : Dial's algorithm, link cost function, taste heterogeneity, crowding

Student Number : 2013 – 23212

Contents

영문 초록 **i**

1 Introduction **1**

2 Literature review **3**

2.1 Travel cost function 3

2.2 Transit assignment algorithms 4

2.2.1 Outline of Transit assignment algorithms 4

2.2.2 Link based stochastic algorithm for transit assignment: Dial's algorithm 7

2.3 Drawbacks of Dial's algorithm 13

2.3.1 Drawbacks of reasonable link 13

2.3.2 Pros and cons of previous studies which tried to resolve drawbacks of Dial's algorithm 16

2.4 Path choice set generation methods 18

2.5 A method which finds out passengers' preferences 20

3 Development of link cost functions **24**

3.1 Difficulty of developing link cost function 24

3.2 A method which finds out a substitution for $|P|$ 27

3.3 Three link cost functions which are used in experiment 29

3.4 Development of parameters, which minimize the errors in logit model 30

4 Experiment **31**

4.1 Calibrating result of parameters 31

4.2 Results of path set generation 33

5 Conclusion and Further research **36**

References **37**

국문 초록 **40**

List of Tables

2.1 Classification of transit assignment algorithms 6

2.2 9 cases that can be possible according to the value of r_i and s_i 8

2.3 Studies on the link-based algorithm for transit assignment . . 23

3.1 Comparison of congestion cost of path 28

4.1 Result of path set generation using first link cost function . . 33

4.2 Result of path set generation using second link cost function . 34

4.3 Result of path set generation using third link cost function . 35

List of Figures

2.1 The condition which link (i, j) is reasonable link 9
2.2 5×5 grid network 10
2.3 Computation result of link likelihood of link $(7, 12)$ 11
2.4 Computation result of link weight of link $(7, 12)$ 12
2.5 Computation result of traffic of link $(7, 12)$ 13
2.6 Reasonable path in 5×5 grid network 13
2.7 Drawback of reasonable link 14
2.8 Preliminaries step of Dial's algorithm : Apgujeong-Suseo station 15
2.9 A cyclic flow between Gangnam-Euljiro 3(sam)-ga station (($H \geqslant 0.6$) 18
2.10 Proportion of passengers who chose a 500-less-passenger-load path, at the expense of transit time during peak hours 20
2.11 Proportion of passengers who chose an alternative path to avoid a transfer, at the expense of transit time during peak hours 21

3.1 Comparison of congestion cost of path A and B 25

4.1 Estimated joint and marginal distribution of α, β and $1 - \alpha - \beta$ 32

1. Introduction

Since the opening of Seoul-Cheongnyangni station, Seoul Metro has evolved to meet increasing demand for public transport in accordance with population increase in metropolitan area. As a result, in 2015, about 8 million people a day use Seoul Metro. In traffic engineering, the problem of estimating the path choice of passengers is called *Transit assignment problem*, and this problem is important because metro operators must know travel behavior of passengers to plan operating strategies well.

To solve transit assignment problem, it is important to generate path choice set efficiently. For this, we should understand network structure and select appropriate method based on it. As you know, the metro network of Seoul metropolitan area consists of a large number of nodes and links; in some area, it has a grid structure. This results in enormous alternative paths between an O-D pair, and enumerating them all requires an inordinate cost. Dial's algorithm, a link-based algorithm, does not require path enumeration. In addition, it does not generate path choice set which passengers may not travel. Therefore, Dial's algorithm is an appropriate method for generating path choice set in Seoul Metro.

In transit assignment problem, the assumption that passengers choose the cheapest path is rational and appears as a basic assumption in many studies. Therefore, it is essential to use a precise link cost function for generating path choice set. In previous studies about cost function in metro, in-vehicle time, waiting time, and transfer are mainly used cost factors. However, in Seoul Metro, a portion of passengers chooses a path which is less crowded but takes more time to avoid crowding especially during peak hours. We should, therefore, consider crowding as a cost factor. In the past, the development of link cost function in metro which reflects crowding was difficult, because there was no way to find out passengers' real flow exactly. However, recently, a method which can precisely estimate exact flow of passengers in Seoul Metro using the Smart Card Automated Fare Collection System, or *Smart Card* in short, was developed by Hong et al. [12]. Based on this method, we can develop link cost function which reflects crowding.

In addition to crowding, there is one more thing that we should take into account in the development of a link cost function. It is passengers' preferences. According to Hong et al. [11], taste heterogeneity affects path choice of passengers significantly larger than perception error. They derived a joint probability of the taste parameters, which determine the way a passenger relatively weighes cost factors, by using inverse optimization to explain the path choice diversity. Therefore, to develop precise link cost function, we should take passengers' preferences into account.

The goal of this paper is as follows. First, we develop a link cost function which reflects both crowding and passengers' preferences. We explain some difficulties of development process, which is caused by congestion cost of path, and propose a method that overcomes these difficulties. Second, we show that the new link cost estimates path choice set more accurately than the previously used link cost.

2. Literature review

In this chapter, we discuss the theoretical background of this paper. In section 2.1, we discuss the concept of travel cost function. In section 2.2, we discuss classification of transit assignment algorithms and Dial's algorithm, which is used in experiment. In section 2.3, we discuss drawbacks of reasonable link defined by Dial. In section 2.4, we discuss path choice set generation methods. At last, in section 2.5, we discuss the method which finds out diversity in the path choice of metro passengers.

2.1 Travel cost function

In general, travel cost function is defined as a function of link's traffic, as follows.

$$c_a = c_a(x)$$

At this time, $x = \{\cdots, x_a, \cdots\}$ means traffic of all the links including relevant link, and if travel cost of link is only related to its own traffic, it is called *separated cost* and expressed as $c_a = c_a(x_a)$. Ortuzar & Willumsen [17] suggested desirable conditions which traffic cost function should satisfy, as follows.

1. The cost function should conform to the reality.

2. The cost function should be non-decreasing and monotonic with respect to traffic.

3. The cost function should be continuous and differentiable.

4. The cost function should be able to express the situation when traffic demand exceeds the capacity.

5. The cost function should be easily manipulated according to the realistic demands.

BPR (Bureau of Public Road) cost function [15], which is a representative link cost function, is easy to manipulate and monotonically decreasing function with respect to traffic. So it guarantees unique solution. The basic form of the BPR function is given by

$$c_a(x_a) = t_0\{1 + \alpha(\frac{x_a}{Q_a})^\beta\}$$

where t_0 means travel time when no traffic is existed on link a, x_a means traffic of link a, Q_a means capacity of link a, and α, β are parameters. However, BPR function is criticized by Spiess [19] because of its unreality.

2.2 Transit assignment algorithms

2.2.1 Outline of Transit assignment algorithms

According to Lim [8], transit assignment algorithm is a technique which loads traffic on transportation network. Transit assignment algorithm can be classified in two ways, whether to reflect traffic and whether to consider passenger's perceived error. First, according to the reflection of traffic, transit assignment algorithm can be classified into network equilibrium algorithm or network loading algorithm. In network equilibrium algorithm, link cost function is defined as a function of link traffic and it finds out equilibrium where cost of a path which passengers choose, is less than equal to cost of other alternative paths. This can be explained by game theory in management science [3]. For example, consider the case which a passenger travels from origin station s, to destination station t in Seoul Metro, and there are multiple paths between $s-t$. In this case, if passenger is rational, he/she will select the minimum cost path for his/her travel. This is called *Selfish Rouging* and traffic of whole metro network is decided by individual passengers' choices. Unlike other cost factors such as time, distance and fare, crowding is unique because it mutually affects the path choice of passengers. That is, if we assume crowding as a factor which affects travel cost, the path choice of passengers becomes a game. In general, this is called *Routing game* and traffic of all paths becomes equilibrium of routing game. On the contrary, in network loading algorithm, which does not reflect traffic, link cost function is constant regardless of the traffic.

Transit assignment algorithm can be classified into deterministic or stochastic according to consideration of passengers' perception error. In deterministic transit assignment algorithm, it is assumed that passengers know whole information about network, so there are no perception errors. On the other hand, in stochastic transit assignment algorithm, travel cost is divided into observable cost such as time, distance, etc., and unobservable cost such as passengers' perception error. In stochastic transit assignment algorithm, it is assumed that unobservable cost follows specific probability distribution. When network becomes larger and more complex, the passengers' degree of understanding network, becomes lower. So, it is impossible for passengers in large network such as Seoul Metro, to know whole information about network. Therefore, perception error is occurred when passengers choose path. Due to this, even if passengers choose same path, their perceived cost is different by passengers, and even if their origin and destination are same, they choose different paths. Therefore, stochastic transit assignment algorithm describes reality more accurately than the deterministic transit assignment algorithm. Table 2.1 represents a classification of transit assignment algorithm according to above two criteria.

First, deterministic network loading is a technique which does not reflect traffic and does not consider passengers' perception error. All-or-nothing assignment, which is one of deterministic network loading technique, is mainly

Table 2.1: Classification of transit assignment algorithms

		Consideration of perceived error	
		not consider	consider
Consideration of traffic	not consider	Deterministic network loading	Stochastic network loading
	consider	User equilibrium	Stochastic user equilibrium

used technique in the early studies of traffic engineering. It finds out shortest path between an $O-D$ pair, and assigns all traffic demand to shortest path. All-or-nothing assignment is time-efficient, but small change of link cost can cause a great change in transit assignment result. Moreover, its basic assumption that passengers know whole information about network is unrealistic. Secondly, stochastic network loading is a technique which does not reflect traffic, but considers passengers' perception error. Stochastic network loading is a special case of binary selection model. In stochastic network loading, if passengers' perception error follows *Gumbel* distribution, it becomes *logit* model, and if passengers' perception error follows *normal* distribution, it becomes *probit* model. In stochastic network loading, traffic is allocated to network based on these two models. Thirdly, user equilibrium is a technique which reflects traffic, but does not consider passengers' perception error. Lastly, stochastic user equilibrium is a technique which reflects traffic and considers passengers' perception error. In stochastic user equilibrium, cost function is developed by reflecting traffic, and by using recursive technique, it reflects congestion effects. That is, stochastic network loading is a sub model of stochastic user equilibrium.

2.2.2 Link based stochastic algorithm for transit assignment: Dial's algorithm

Transit assignment algorithm can be divided into path-based algorithm and link-based algorithm according to its unit of analysis. Unit of analysis of path-based algorithm is path, so it needs all the alternative paths between Origin*(O)* and Destination*(D)*. Seoul Metro consists of a great number of nodes and links, and in some area, it has grid structures. So, there are enormous paths between $O-D$ pair, and enumerating all the paths requires a lot of storage and cost. Therefore, it is inappropriate to use path-based algorithm to solve transit assignment problem in Seoul Metro.

Unlike path-based algorithm, unit of analysis of link-based algorithm is link, so it does not need path enumeration. Therefore, it can be easily applied to a network where enormous paths between $O-D$ pair are exist, and an appropriate method to generate path choice set. Dial's algorithm [9], which is one of the link-based algorithms, is based on logit model, and it has some features as follows.

1. Dial's algorithm does not allocate traffic on links which passengers may not travel, if they are reasonable.

2. Dial's algorithm does not need path enumeration between $O-D$ pair, so it can be easily applied to large network.

3. If alternative path sets of logit model and path sets which Dial's algorithm assign traffic, are equal, transit assignment results in both model are same.

Reasonable link(reasonable path) is a link that consists of path which Dial's algorithm assigns traffic. In Dial's algorithm, a path is reasonable if every link in it has its initial node closer to the origin than is its final node, and has its final node closer to the destination than is its initial node. Notations used in Dial's algorithm are as follows.

1. r_i : the shortest path distance from origin s to node i.
2. s_i : the shortest path distance from note i to destination t.
3. d_{st} : the traffic demand between origin-destination $s - t$ pair.
4. δ_i^+ : the set of all links whose initial node is node i.
5. δ_i^- : the set of all links whose final node is node i.

At this time, the definition of reasonable link is like definition 2.2.1.

Definition 2.2.1 If link (i, j) satisfies following condition, define link (i, j) as a reasonable link.

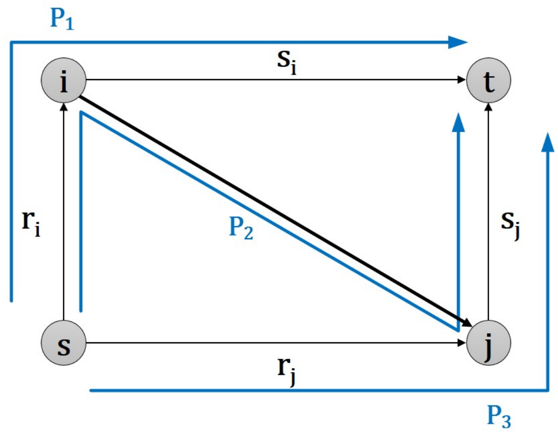
$$r_i < r_j, s_i > s_j \tag{2.1}$$

Let's check the reason why link (i, j) is reasonable if it satisfies definition 2.2.1. Table 2.2 indicates 9 cases that can be possible according to the value of r_i and s_i. As you can see in figure 2.1 and table 2.2, if link (i, j) satisfies definition 2.2.1, there is a chance of allocating traffic on link (i, j). For example, in the case of $r_i = 5, r_j = 3$, $s_i = 1, s_j = 3$, the cost of path P_1, P_2, and P_3 is $6, 8 + c_{ij}, 6$, respectively. Therefore, if passengers are reasonable, they may choose P_1 or P_3 instead of P_2.

Table 2.2: 9 cases that can be possible according to the value of r_i and s_i

	$s_i < s_j$	$s_i = s_j$	$s_i > s_j$
$r_i < r_j$	P_1	P_1	P_1 or $\mathbf{P}_2$ or P_3
$r_i = r_j$	P_1	P_1 or P_3	P_3
$r_i > r_j$	P_1 or P_3	P_3	P_3

Figure 2.1: The condition which link (i, j) is reasonable link

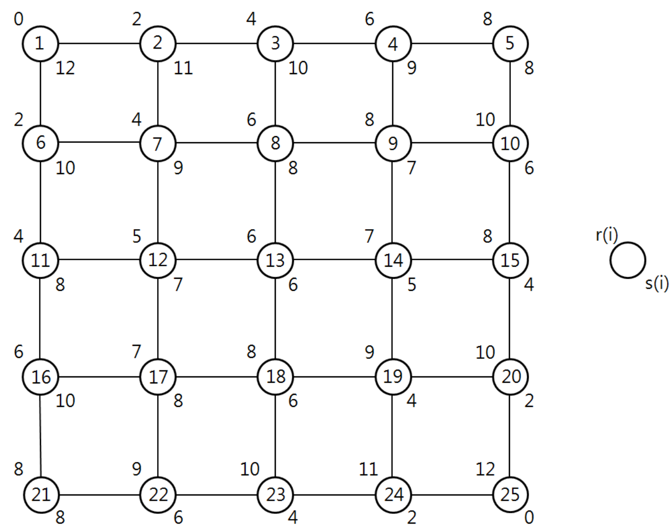


Dial's algorithm assigns traffic through preliminaries, forward pass, backward pass, while avoiding path enumeration. Let's see how the Dial's algorithm proceeds in 5×5 grid network, figure 2.2. In figure 2.2, origin node is 1 and destination node is 25, traffic demand is 700, and link costs are 2, except on the links forming the bisecting horizontal path between node 11 and node 15, where link costs are 1. And above and to the left of node i is r_i, below and to the right of node i is s_i. During preliminaries, first compute r_i and s_i of all nodes, and assign positive *link likelihood* on the link which satisfies definition 2.2.1. Link likelihood of link (i, j) is calculated by equation 2.2.

$$l_{ij} = \begin{cases} \exp[\theta\,(r_j - r_i - c_{ij})] & \text{if } r_i < r_j,\ s_i > s_j \\ 0 & \text{Otherwise.} \end{cases} \tag{2.2}$$

In this case, θ is a parameter that reflects passengers' perception error. If the value of θ is zero, same traffic volumes are assigned to all paths which consist of reasonable links, and if the value of θ becomes larger, more traffic volumes are assigned to the cheaper path. After preliminaries, *link weights*

Figure 2.2: 5×5 grid network



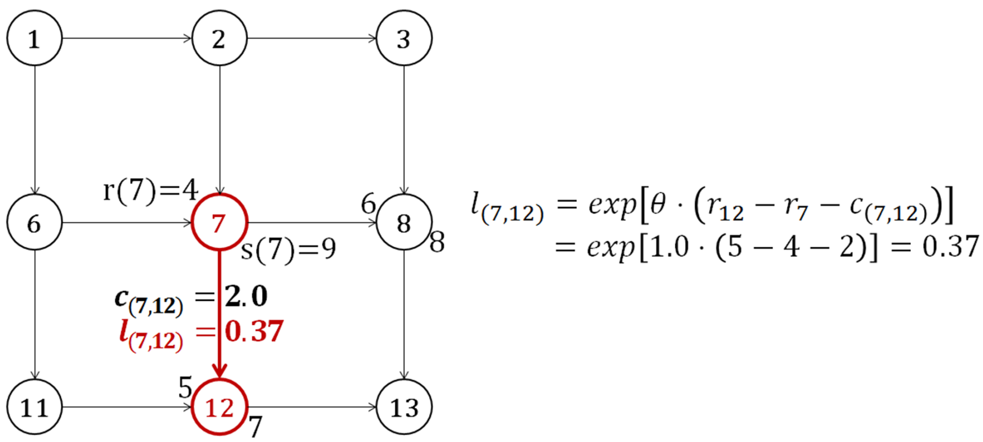
are calculated in forward pass. In forward pass, by examining all nodes i in ascending sequence with respect to r_i, their distance from the origin, link weight of (i, j) is calculated, until the destination node t is reached. For node i, link weight of (i, j), $j \in \delta_i^+$ is calculated by equation 2.3.

$$w_{ij} = \begin{cases} l_{ij} & \text{if } i = s \\ l_{ij} \sum_{m \in \delta_i^-} w_{mi} & \text{Otherwise.} \end{cases} \tag{2.3}$$

After forward pass, traffic is assigned to link which satisfies definition 2.2.1 in backward pass. In backward pass, starting with the destination node t, examine all nodes j in descending sequence with respect to s_j, and assign traffic to link, until the origin node s is reached. For node j, traffic of link (i, j), $i \in \delta_j^-$, x_{ij} is calculated by equation 2.4.

$$x_{ij} = \begin{cases} d_{st} \frac{w_{ij}}{\sum_{m \in \delta_j^-} w_{mj}} & \text{if } j = t \\ \left(\sum_{m \in \delta_j^+} x_{jm} \right) \frac{w_{ij}}{\sum_{m \in \delta_j^-} w_{mj}} & \text{Otherwise.} \end{cases} \tag{2.4}$$

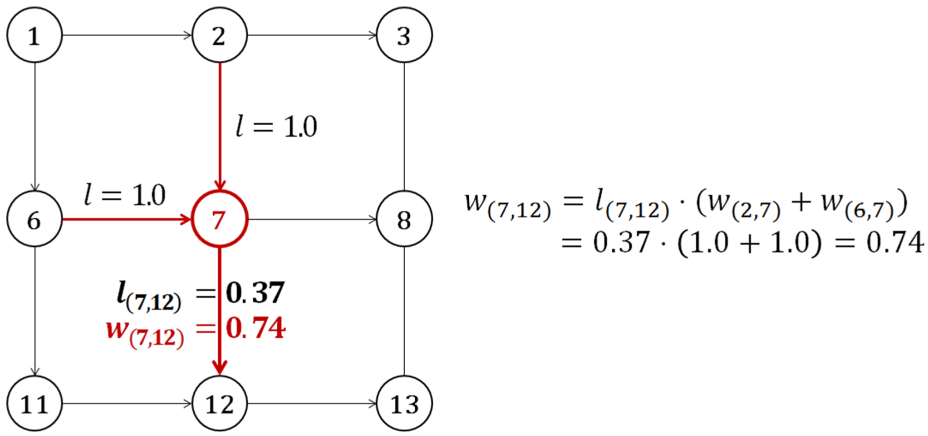
Figure 2.3: Computation result of link likelihood of link $(7, 12)$



After preliminaries, forward pass, and backward pass, Dial's algorithm assigns traffic demand between $O-D$ pair, by using the definition of reasonable link. At the same time, Dial's algorithm excludes paths which passengers may not travel. In figure 2.2, the number of possible alternative paths is $\frac{8!}{4!4!} = 70$, but Dial's algorithm considers only 9 paths, like figure 2.6.

Focusing on these advantages, previous studies which apply Dial's algorithm to public transport network is as follows. First, Jeong [2] defined reasonable link as its initial node closer to the origin than is its final node, and applied Dial's algorithm to Seoul Metro. In his study, distance is only cost factor of link, and he defined travel link cost as moving distance between stations, and transfer link cost as 3.4km. By adjusting parameter θ for each origin-destination, he got assignment result that substantially coincides with actual passengers' movement. But using same θ for every origin-destination pair, assignment result was quite different from actual passengers' movement. Lee et al. [1] used time and fare as a cost factor of link, and applied Dial's

Figure 2.4: Computation result of link weight of link $(7, 12)$



algorithm to the Sioux Falls network. They showed that Dial's algorithm was better than TRANPLAN and EMME/2, which are commercial program, in theoretical background and analysis of flexibility.

Figure 2.5: Computation result of traffic of link $(7, 12)$

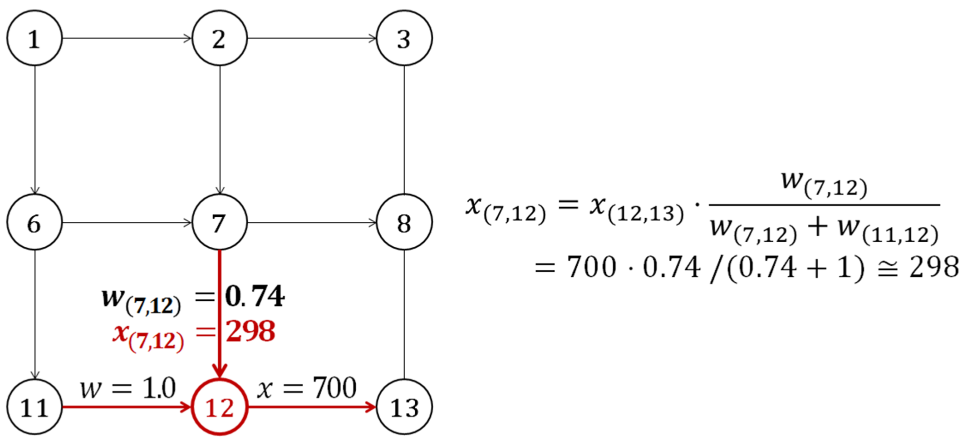
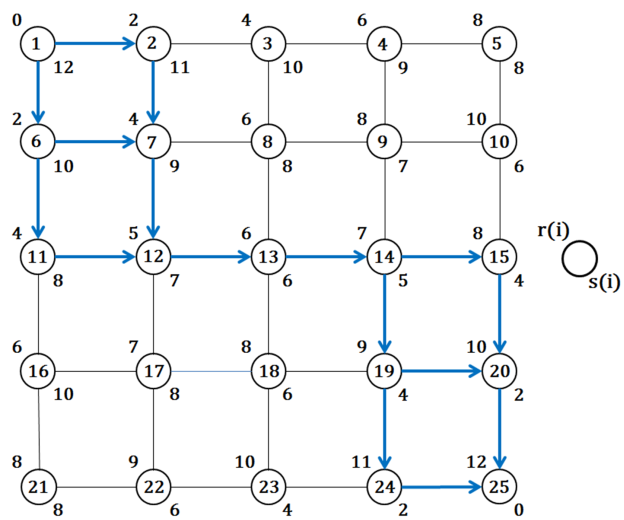


Figure 2.6: Reasonable path in 5×5 grid network

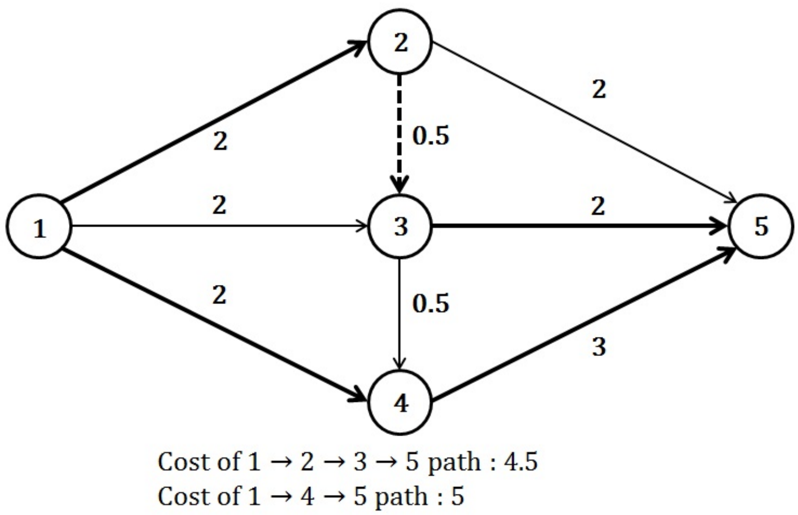


2.3 Drawbacks of Dial's algorithm

2.3.1 Drawbacks of reasonable link

Dial's algorithm assigns traffic demand between $O - D$ pair, by using definition of reasonable link. At the same time, Dial's algorithm excludes paths which passengers may not travel. However, according to Akamatsu [4], Dial's algorithm excludes some links, because they does not satisfy the condition of reasonable link, even if passengers may travel those links. This is because although cost of path is cheap, some links that consist of path do not satisfy the condition of reasonable link.

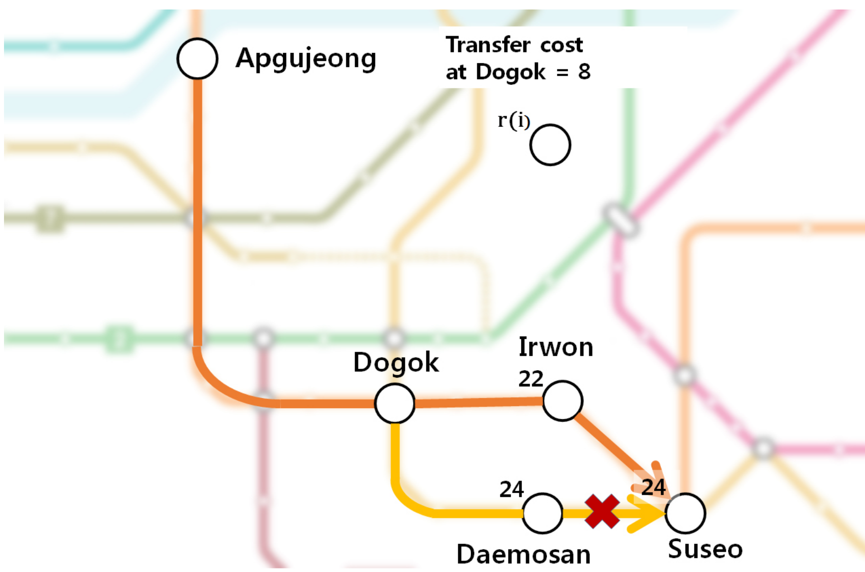
Figure 2.7: Drawback of reasonable link



In figure 2.7, define path A and B as : $1 \rightarrow 2 \rightarrow 3 \rightarrow 5$, $1 \rightarrow 4 \rightarrow 5$, respectively. The cost of path A is 4.5 and B is 5. However, link $(2, 3)$ which consists of path A, does not satisfy condition of reasonable link. Therefore, no traffic is allocated to path A, although its cost is cheaper than path B. This problem can be happened when we apply Dial's algorithm to Seoul Metro. When we apply it to Seoul Metro, due to condition of reasonable link, no traffic is allocated to paths which passengers actually travelled. For example, there are two paths, which passengers actually travelled, between Apgujeong station and Suseo station. One path is without transfer, and the other path has one transfer at Dogok station. Figure 2.8 represents preliminaries step of

Dial's algorithm when we apply it to Apgujeong station-Suseo station pair. In figure 2.8, above and to the left of node i is r_i, below and to the right of node i is s_i and measure is minute. If we look at the Daemosan station and Suseo station, value of r_i is same. Therefore, link (Daemosan station, Suseo station) does not satisfy condition of reasonable link, so no traffic is allocated to path with one transfer at Dogok station. This phenomenon mainly occurs at $O-D$ station pairs where destination station is transfer station, or line of origin station and destination station is different, so there are multiple paths in $O-D$ station pair. This means that Dial's algorithm finds out less number of paths than passengers actually travelled by excluding paths which are unreasonable.

Figure 2.8: Preliminaries step of Dial's algorithm : Apgujeong-Suseo station



2.3.2 Pros and cons of previous studies which tried to resolve drawbacks of Dial's algorithm

Reasonable link, which Dial defined, has some advantages., it excludes unreasonable paths and prevents cyclic flows. However, number of paths that Dial's algorithm assigns, is smaller than number of paths that passengers actually travelled. To resolve this problem, Bell [5] and Akamatsu [4] proposed transit assignment methods that consider all possible paths in network. First, Bell proposed logit based transit assignment method, which does not need path enumeration. Bell's method is based on Floyd-Warshall algorithm, and his method computes link weight matrix recursively, and sum these computed link weight matrixes all. And then, compute probability of using alternative paths based on inverse matrix of link weight. However, if network size is large, Bell's method takes lots of time in computing inverse matrix of link weight. And Bell assumed that sum of link weight matrix converges, but the convergence was only proved in acyclic network, not in general network, by Wong [22]. Akamatsu proposed two logit based transit assignment methods. First one, which is same as Bell's method, uses Markov chain. And second one uses decomposition of entropy function. Second method is based on the equivalence of maximum entropy principle and logit model. But second method has drawback., it is not efficient in large network.

Tobin [21] proposed a method that modified definition of reasonable link by introducing coefficient of discrimination, δ, to resolve drawbacks of reasonable link. Tobin focused on the fact that degree of recognition of time difference between shortest path and alternative path, differs according to travel time. So he suggested that if time difference between shortest path and alternative path is smaller than function of path times coefficient of discrimination, then passengers recognize that travel time of alternative path and shortest path are same. Let $t(P)$ be travel time for path P, $t(P^{'})$ be travel time for shortest path, $g(P)$ be a function of path P. If path P satisfy equation 2.5, then passengers recognize that travel time of alternative path and shortest path are same.

$$t(P) - t(P^{'}) \leqslant \delta g(P) \tag{2.5}$$

Based on this, Tobin modified definition of reasonable link like equation 2.6.

$$r(i) + t(i,j) - r(j) \leqslant \delta g, \text{ and } r(i) \leqslant r(j) \tag{2.6}$$

But the problem of Tobin's method is that it is difficult to infer coefficient of discrimination and function of path.

To resolve drawbacks of reasonable link, Bing-Feng Si [18] proposed that if cost of alternative path does not exceed $(1+H)$ times cost of shortest path, alternative path is a reasonable path. Let $r-s$ be origin-destination node, k be the kth alternative path between origin-destination node, c^{rs}_{min} be cost of shortest path between $r-s$. Then, definition of reasonable path is as follows.

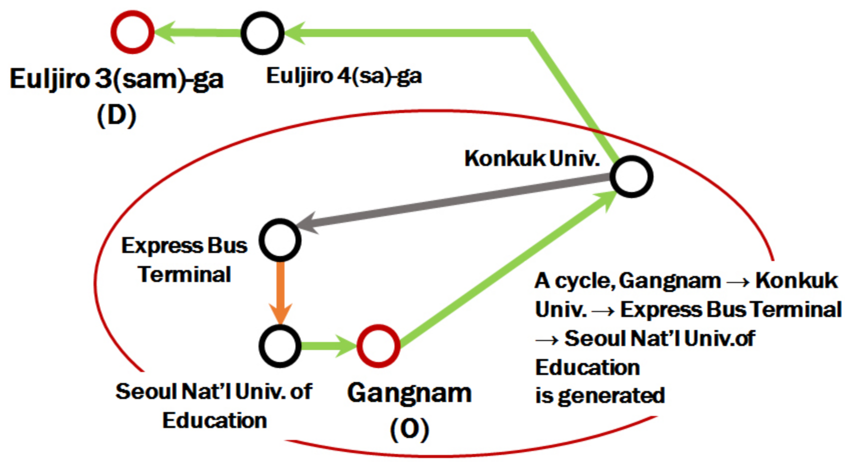
$$c^{rs}_{k} \leqslant (1+H)c^{rs}_{\text{min}} \tag{2.7}$$

Based on this, Bing-Feng Si modified definition of reasonable link like equation 2.8.

$$r(i) + t_{ij} + s(j) \leqslant (1+H)c^{rs}_{\text{min}} \tag{2.8}$$

Bing-Feng Si showed that by using his modified definition of reasonable link, transit assignment results in Beijing Subway network, described reality more accurately than using Dial's original definition of reasonable link. However, the problem in Bing-Feng Si's method is that it can be applied only in acyclic network. Basically, cyclic flow can be possible in Seoul Metro. So as you can see in figure 2.9, cyclic flow is generated when the value of H is above a certain value, if we apply Bing-Feng Si's method to Seoul Metro.

Figure 2.9: A cyclic flow between Gangnam-Euljiro 3(sam)-ga station (($H \geqslant 0.6$)



2.4 Path choice set generation methods

There are some path choice set generation methods in traffic engineering. First one is *k-th shortest paths*(KSP). This method is proposed by Lawler [14]. In KSP, it generates path set which consists of the k-shortest paths according to a given criteria. And path choice set is generated by excluding some links and repetitive applications of shortest path algorithm. The termination criteria is based on the given maximum number k of paths to be generated. KSP has an advantage that it matches only few requirements for generating path choice set, and one of the requirements is acyclic criterion. However, in KSP, it is hard to decide appropriate k, and the size of path choice set may be much larger than the desired size in some cases. Moreover, generated paths tend to be very similar in spatial sense and also largely overlapping.

Second one is *Link elimination* method. This method is proposed by Bellman and Kabala [6], Martin, E.Q.V. [16]. In link elimination method, path choice set contains paths that are variation of the shortest path by deleting some or all links of the shortest path. This method stops if no more paths are generated. However, link elimination method has a disadvantage that it may remove essential links in shortest path. As a result, network becomes disconnected, and no more paths is generated.

Third one is *Labelling approach*. This method is proposed by M Ben-Akiva et.al [7]. In labelling approach, path choice set consists of all labelled shortest paths that each are optimal for a specific criteria, or label, from a given set of labels. In labelling approach, minimum time, minimum distance, minimum fare, minimum traffic, etc. can be a label criteria. It generates multiple paths that reflecting the fact that passengers may have different preferences, so their objective functions are different. This method stops when all labels are considered. However, labelling approach has a disadvantage that same path may be generated even if labels are different. Also, spatial variability criteria is not considered.

Previous studies which utilize these techniques is as follows. Tan et.al [20] applied path choice set generation methods to Singapore transportation network. In [20], passengers' actual path were obtained by using smart card data. They presented the results of path choice set generation methods and evaluate this results in coverage. They said that using combination of path choice set generation methods, 93.5 % of actual path is covered. However, efficient coverage, which is the percentage of generated paths being a actual path, is only 1.0%, which is low. S Fiorenzo-Catalano et.al [10] applied a combination of Labelling method and simulation method to transportation network which connects the cities of Dordrecht and Rotterdam in the Netherlands. They generated path choice set in three ways., first one randomizes link attributes, second one randomizes passengers' preferences and third one randomizes both at the same time. They observed that path choice set generate by third one covered passengers' actual path set with a percentage of

78%.

2.5 A method which finds out passengers' preferences

Hong et al. [11] found that in Seoul Metro, there were some passengers' movement, which was observed by the method in [12], that can not be explained by passengers' perception error. Figure 2.10 indicates proportion of passengers who chosen an alternative path, which is less than 500 passenger load at the expense of transit time during peak hours. For example, in figure 2.10, to avoid congestion, proportion of passengers who chosen an alternative path, which is 200 seconds longer, is 45%.

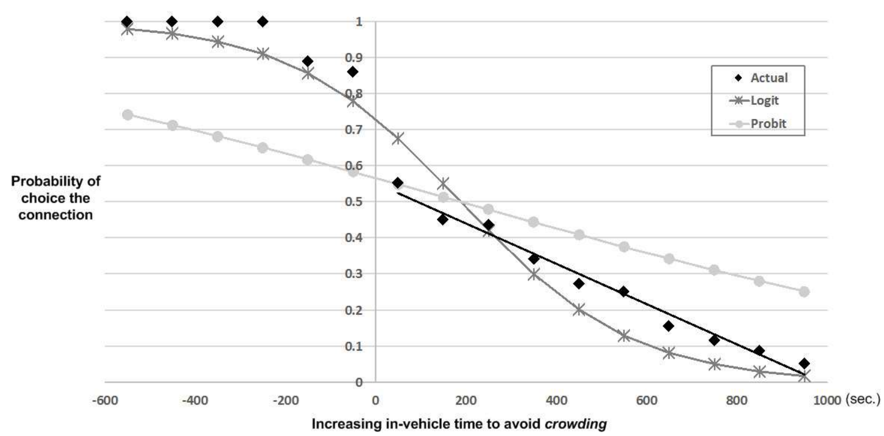


Figure 2.10: Proportion of passengers who chose a 500-less-passenger-load path, at the expense of transit time during peak hours

In figure 2.10, we can observe that proportion of passengers who chosen an alternative path to avoid congestion, decreases linearly as transit time increases. This phenomenon can not be explained by the logit model or probit model, which assumes passengers' perception error as Gumbel distribution, or Normal distribution, respectively. Figure 2.11 indicates proportion of passengers who chose an alternative path to avoid a transfer at the expense of transit time during peak hours. For example, in figure 2.11, to avoid transfer, proportion of passengers who chose an alternative path which is 240 seconds longer, is 50%.

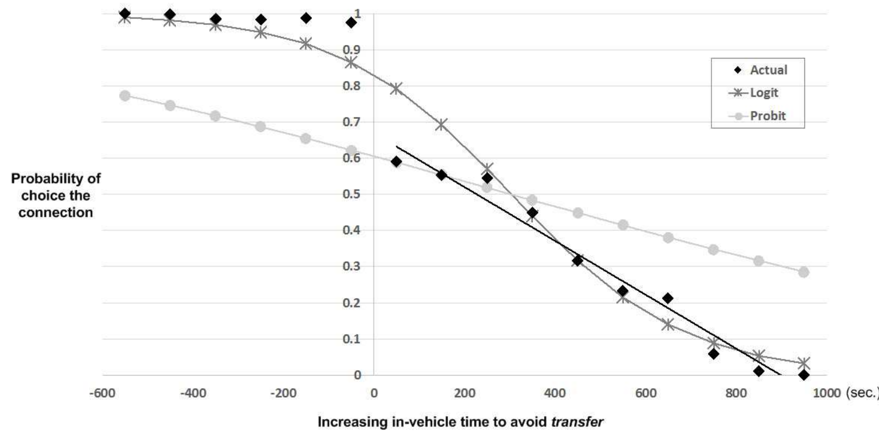


Figure 2.11: Proportion of passengers who chose an alternative path to avoid a transfer, at the expense of transit time during peak hours

As you can see in figure 2.11, proportion of passengers, who chose an alternative path to avoid a transfer, can not be explained by logit or probit model.

Based on these two observations, the reason why passengers choose different path, is not because they make perception error, but because their preferences are various. To find out passengers' preferences, Hong et al. [11] assumed form of disutility function of path P as Cobb-Douglas form, like equation 2.9, if current passengers' flow is f and parameters α, β follow probability distribution $g(\alpha, \beta)$. In equation 2.9, $s_P, l_P(f), t_P$ means in-vehicle time of P, average passenger load of P, transfer time of P, respectively. They derived estimated parameter distribution for g, $\widehat{g}$, by solving inverse optimization model in equation 2.10.

$$u_p(f,\alpha,\beta) = \alpha \ln s_P + \beta \ln l_p(f) + (1-\alpha-\beta)\ln t_p \tag{2.9}$$

$$\begin{gathered}\min \textstyle\sum_{P\in\mathcal{P}} \delta_P^2,\\ \int_{\Gamma_P^k(\hat{f})} g = \hat{\pi}_P^k(\hat{f}) + \delta_P,\ \forall P \in P_k, \forall k\\ \int_\Gamma g = 1,\ g \geqslant 0.\end{gathered} \tag{2.10}$$

They compared transit assignment results using $\widehat{g}$ with transit assignment results using logit model which has same cost factors. The results turned out that error of transit assignment results using $\widehat{g}$ with respect to actual passengers' flow, is 9% lower than transit assignment results using logit model which has same cost factors.

So far observed, to be a precise link cost function in Seoul Metro, passengers' preferences should be taken into account. Table 2.3 summarizes previous studies in domestic and foreign, and model that is proposed in this paper.

Table 2.3: Studies on the link-based algorithm for transit assignment

	Cost factors					Features	Applied network
	Time	Transfer	Distance	Fare	crowding		
[9]	O					Propose Dial's algorithm	Sample network
[18]	O	O				Modify condition of reasonable link	Beijing Subway Network
[2]		O	O			Adjusting θ	Seoul Metro
[1]	O	O		O		Compare Dial's algorithm with commercial program	Sioux Falls network
[21]	O					Introducing coefficient of discrimination	Sample network
Ours	O	O			O	Inverse optimization, passengers' preferences	Seoul Metro

3. Development of link cost functions

In generally, it is assumed that cost of path is equal to sum of cost of links. At this time, commonly used cost factors are time, distance, fare, crowding, etc. Unlike other cost factors, crowding is unique because congestion cost of path can not be expressed as sum of crowding of links. In this chapter, we deal with the structure of link cost function and how to develop link cost function considering passengers' preferences.

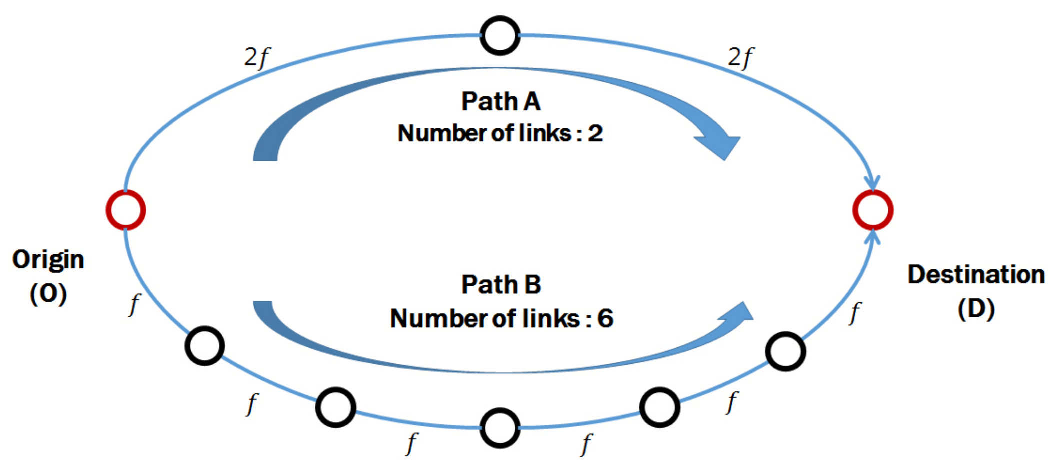
3.1 Difficulty of developing link cost function

Path cost functions, which reflects crowding in Seoul Metro, so far, use average passenger load of path as a factor that determines congestion cost of path [11], [13]. In [11], path cost function has Cobb-Douglas from, like equation 2.9. Therefore, we need to convert it into link cost function. However, logarithmic function in it makes conversion difficult. For example, if path cost function has Cobb-Douglas form, in-vehicle time cost of path is equal to $\ln(\sum_{a\in P} IVT_a)$. Here IVT_a means in-vehicle time of link a. If we convert this into link cost function, in-vehicle time cost of link is equal to $\ln(IVT_a)$. As stated before, cost of path is equal to sum of cost of links. So, in-vehicle time cost of path is equal to $\ln(\prod_{a\in P} IVT_a)$. As a result, in-vehicle time cost of path, which is expressed as sum of cost of links, is greater than in-vehicle time cost of path which has Cobb-Douglas form. However, according to [11], improvement of correctness of transit assignment result of path cost function which has Cobb-Douglas form, compared to that of path cost function which has linear from, is only 1.3%. Therefore, we assume the structure of link cost function as linear form like equation 3.1, not Cobb-Douglas form. In equation 3.1, TT_a means transfer time of link a, x_a means traffic of link a, and traffic of link is defined as average passenger load of trains which pass link a.

$$c_a = \begin{cases} \alpha\ IVT_a + \beta x_a & \text{if link } a \text{ is travel link} \\ (1-\alpha-\beta)\ TT_a & \text{if link } a \text{ is transfer link} \end{cases} \tag{3.1}$$

But, if we assume link cost function like equation 3.1, unlike in-vehicle time and transfer time, congestion cost of path is exaggerated. This is because congestion cost of path becomes bigger when number of links in it becomes bigger, although degree of congestion is low. For example, in figure 3.1, path A is more crowded than path B, so passengers who use path A, feel more congestion cost than passengers who use path B. However, if we assume link cost function like equation 3.1, because number of links which consists of path A is smaller than that of path B, congestion cost of path B is larger than that of path A.(Number of passengers of path A and B is 4000, 6000 people, respectively.)

Figure 3.1: Comparison of congestion cost of path A and B



Not to exaggerate congestion cost of path, we take logarithm function on traffic of a link, and use it as a congestion cost of link. By doing this, congestion cost of path becomes lower than congestion cost of path is calculated as a sum of the numerical value of traffic of links. This can be easily proved. Assume that there is a path which consists of 3 links, and traffic of link is x_a, x_b, x_c in the order. Then, congestion cost of path is equal to $x_a + x_b + x_c$ when cost of link is like equation 3.1, and $\ln(x_a \times x_b \times x_c)$ when we take logarithm function on traffic of link. As you know, $x_a + x_b + x_c$ is equal to $\ln(e^{x_a \times x_b \times x_c})$, and there are at least tens of passengers at a link in Seoul Metro. So, $e^{x_a}, e^{x_b}, e^{x_c}$ is greater than x_a, x_b, x_c respectively. Therefore, when we take logarithm function on traffic of link, congestion cost of path becomes lower. However, this method still can exaggerate congestion cost of path. In figure 3.1, congestion cost of path A is $\ln(2000^2)$ and that of path B is $\ln(1000^6)$, so congestion cost of path B is still larger.

To overcome this problem, we compare congestion cost of path taking logarithm function on traffic of link, with congestion cost of path using average passenger load of path. The congestion cost of path P, which takes logarithm function on traffic of link, is equal to equation 3.2 and congestion cost of path, which uses average passenger load of paths, is equal to equation 3.3. In those two equations, $|P|$ means number of links that consist of path P.

$$\text{Congestion cost of } P \text{ taking logarithm function on traffic} = \ln(\prod_{a \in P} x_a) \tag{3.2}$$

$$\text{Congestion cost of } P \text{ using average passenger load} = \ln(\frac{\sum_{a \in P} x_a}{|P|}) \tag{3.3}$$

When we take $(1/|P|)$ root of $\prod_{a \in P} x_a$ in equation 3.2, this and equation 3.3 becomes arithmetic geometric relation like equation 3.4.

$$\ln\left(\frac{\sum_{a \in P} x_a}{|P|}\right) \geqslant \ln\left(\prod_{a \in P} x_a^{\frac{1}{|P|}}\right) \tag{3.4}$$

Therefore, if we assume congestion cost of link like equation 3.5, we can reflect average crowding of path and can not exaggerate congestion cost of path.

$$\text{Congestion cost of link } a = \ln(x_a^{\frac{1}{|P|}}) \tag{3.5}$$

However, Dial's algorithm, which is used in experiment, is link-based transit assignment algorithm. So, unlike path based transit assignment algorithm, we can not know the number of links that consist of path. Therefore, calculation of $|P|$ is impossible.

3.2 A method which finds out a substitution for $|P|$

As stated in section 3.1, the fact that we can not calculate $|P|$ during process of Dial's algorithm, makes development of link cost function, which reflects crowding, difficult. However, we can find out a substitution for $|P|$, $\widehat{|P^k|}$, depending on origin and destination station. According to Hong et al. [12], they assumed that if an alternative path has more stations than shortest one, which has same number of transfer with alternative path, by 10 sections or more, passengers might not travel alternative path. Moreover, among 137,150 origin-destination pairs, which passengers travelled in November 21, 2011, in Seoul Metro, only 0.13% of origin and destination pair had a path that was longer than a shortest path by 10 sections or more, regardless of number of transfer. Therefore, if we denote minimum number of stations that is needed to travel between origin and destination station pair k as

$|P_{sp}^k|$, then maximum number of stations that passengers may travel is equal to $|P_{sp}^k| + 10$.

To find out a substitution for $|P|$, we compare congestion costs of path. First congestion cost is obtained by assuming $\widehat{|P^k|}$ is equal to $|P_{sp}|$, second one is obtained by assuming $\widehat{|P^k|}$ is equal to $|P_{sp}| + 10$, and third one is obtained by using average passenger load of path.

Paths that passengers travelled between selected 1312 origin-destination station pairs in Seoul Metro, during peak hours (7:00 AM $\sim$ 10:00 AM, 6:00 PM $\sim$ 9:00 PM) in November 21, 2011, are used for comparison. Table 3.1 indicates comparison result. In table 3.1, it is turned out that the difference between first congestion cost and third congestion cost, is significant. However, the difference between second congestion cost and third congestion cost is a small value, 0.1532. Therefore, we can define a substitution for $|P^k|$, $\widehat{|P^k|}$, as $|P_{sp}^k| + 10$ for origin-destination pair k, and congestion cost of link is calculated like equation 3.6. So, if we calculate congestion cost of link differently depending on the origin-destination station pair, without the information about path, congestion cost of path becomes similar with the cost which reflects average passenger load of path.

$$\text{Congestion cost of link } a = \ln(x_a^{\frac{1}{\widehat{|P^k|}}}), \widehat{|P^k|} = |P_{sp}^k| + 10 \tag{3.6}$$

Table 3.1: Comparison of congestion cost of path

	Avg. passenger load	$\|P_{sp}\|$	$\widehat{\|P\|}$	Avg. passenger load - $\widehat{\|P\|}$	Avg. passenger load - $\|P_{sp}\|$
Avg. congestion cost of paths	6.7396	8.6846	6.5864	0.1532	-1.9450

3.3 Three link cost functions which are used in experiment

There are three kind of link cost function which is used in experiment. In first link cost function, cost factor is only time. In-vehicle time(IVT) determines travel link cost, and transfer time(TT) determines transfer link cost. First link cost function is like equation 3.7.

$$c_a = \begin{cases} \alpha \ IVT_a & \text{if link } a \text{ is travel link} \\ \beta \ TT_a & \text{if link } a \text{ is transfer link} \end{cases} \tag{3.7}$$

In second link cost function, cost factors are time and crowding. In-vehicle time and traffic determines travel link cost, transfer time determines transfer link cost. Second link cost function is like equation 3.8.

$$c_a = \begin{cases} \alpha \ IVT_a + \gamma \ln(x_a^{\frac{1}{|P|}}) & \text{if link } a \text{ is travel link} \\ \beta \ TT_a & \text{if link } a \text{ is transfer link} \end{cases} \tag{3.8}$$

In third link cost function, cost factors are time and crowding. But Unlike other two link cost functions, third link cost function reflects passengers' preferences. In third link cost function, parameters α, β follow probability distribution $g(\alpha, \beta)$ and estimated parameter distribution is obtained by solving inverse optimization model in equation 2.10. Third link cost function is like equation 3.9.

$$c_a = \begin{cases} \alpha \ IVT_a + \beta \ln(x_a^{\frac{1}{|P|}}) & \text{if link } a \text{ is travel link} \\ (1 - \alpha - \beta) \ TT_a & \text{if link } a \text{ is transfer link} \end{cases} \tag{3.9}$$

3.4 Development of parameters, which minimize the errors in logit model

First link cost function and second link cost function are developed by minimizing the errors in logit model. For origin-destination station k, arbitrarily choose two paths P, Q , which passengers actually travel. Let $\Pr_P, \Pr_Q$ be a proportion of passengers who use path P or Q, respectively. Then, equation 3.10 is satisfied.

$$\frac{\exp(-c(P))}{\Pr_P} = \frac{\exp(-c(Q))}{\Pr_Q} \tag{3.10}$$

However, transit assignment result obtained by logit model, does not match passengers' actual path choice. If we denote the errors by e_{pq}, the mathematical model which finds out parameters α, β, γ that minimize the errors, is like equation 3.11.

$$\begin{gathered} \min \tfrac{1}{2} \sum_{P,Q \in \mathcal{P}_i, i=1,\ldots,k} e_{PQ}^2, \\ \ln(\Pr_Q) - c(P) = \ln(\Pr_P) - c(Q) + e_{PQ}, \quad P, Q \in \mathcal{P}_i, \quad i = 1, \ldots, k \\ \alpha, \beta, \gamma \geqslant 0 \end{gathered} \tag{3.11}$$

4. Experiment

Passengers who used Seoul Metro during peak hours (7:00 AM ∼ 10:00 AM, 6:00 PM ∼ 9:00 PM) in November 21, 2011, are used in experiment. The Number of passengers used in experiment is 1,940,730 people, and the number of paths that passengers actually travelled is 76,284.

4.1 Calibrating result of parameters

The calibrating result of parameters in first and second link cost function, developed by minimizing the errors in logit model, are in equation 4.1, 4.2, respectively.

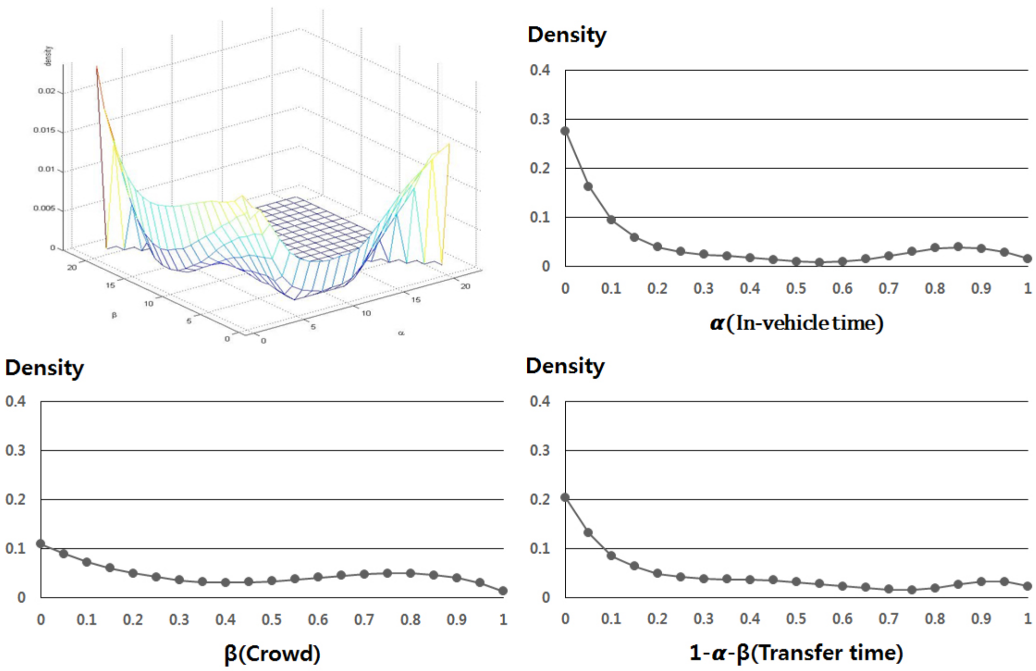
$$c_a = \begin{cases} 0.0017\ IVT_a & \text{if link } a \text{ is travel link} \\ 0.0039\ TT_a & \text{if link } a \text{ is transfer link} \end{cases} \tag{4.1}$$

$$c_a = \begin{cases} 0.0040\ IVT_a + 0.0019 \ln(x_a^{\frac{1}{|P|}}) & \text{if link } a \text{ is travel link} \\ 0.0058\ TT_a & \text{if link } a \text{ is transfer link} \end{cases} \tag{4.2}$$

The results show that if we only use time as a cost factor, passengers perceive transfer time as 2.3 times in-vehicle time. And If we use time and crowding as cost factors, passengers perceive transfer time as 1.5 times in-vehicle time.

The estimated probability distribution of third link cost function, which has time and crowding as cost factors and reflects passengers' preferences, is illustrated in figure 4.1.

Figure 4.1: Estimated joint and marginal distribution of α, β and $1-\alpha-\beta$



The parameter α, which is related to in-vehicle time, and $1-\alpha-\beta$, which is related to transfer time, has similar probability over the entire region while having relatively large mess around 0. The parameter β, which is related to congestion cost, has similar probability over the entire region.

4.2 Results of path set generation

Before the analysis of results, let's define some terminologies. Let $\mathcal{P}_R$ be the set of paths that passengers actually travelled, $\mathcal{P}_i$ be the set of paths that Dial's algorithm generates using i-th link cost function. Then, $\mathcal{P}_R \cap \mathcal{P}_i$ becomes the set of paths that passengers actually travelled and Dial's algorithm generates, $\mathcal{P}_i \setminus \mathcal{P}_R$ becomes the set of paths that Dial's algorithm generates but passengers actually did not travel. And $\frac{\mathcal{P}_R \cap \mathcal{P}_i}{\mathcal{P}_R}$ means coincidence rate, the percentage of paths which passengers actually travelled being available in the generated path choice set. The results are summarized in table 4.1, 4.2, 4.3.

Table 4.1: Result of path set generation using first link cost function

Num. transfer	$\mathcal{P}_R \cap \mathcal{P}_i$	$\mathcal{P}_i \setminus \mathcal{P}_R$	$\frac{\mathcal{P}_R \cap \mathcal{P}_i}{\mathcal{P}_R}$(%)	0 transfer (%)	1 transfer (%)	2 transfers (%)
0 (23.4 %)	14,242	1,591	98.6	100	-	-
1 (53.5 %)	28,982	6,490	75.2	-	100	-
2 (13.3 %)	4,780	3,402	49.7	-	-	100
1,2 (7.8 %)	4,860	1,027	44.1	-	50.5	49.5
0,2 (1.0 %)	635	146	45.9	49.1	-	50.9
0,1 (0.9 %)	542	111	47.0	48.6	51.4	-
0,1,2 (0.1 %)	33	3	30.6	32.4	32.4	35.2
Total	54,074	12,770	70.9	20.6	58.6	20.7

Table 4.2: Result of path set generation using second link cost function

Num. transfer	$\mathcal{P}_R \cap \mathcal{P}_i$	$\mathcal{P}_i \setminus \mathcal{P}_R$	$\frac{\mathcal{P}_R \cap \mathcal{P}_i}{\mathcal{P}_R}$(%)	0 transfer (%)	1 transfer (%)	2 transfers (%)
0 (23.4 %)	14,245	1,076	98.6	100	-	-
1 (53.5 %)	29,005	5,090	75.2	-	100	-
2 (13.3 %)	5,103	2,675	53.0	-	-	100
1,2 (7.8 %)	4,919	753	44.6	-	50.5	49.5
0,2 (1.0 %)	632	70	45.7	49.1	-	50.9
0,1 (0.9 %)	540	53	46.8	48.6	51.4	-
0,1,2 (0.1 %)	34	3	31.5	32.4	32.4	35.2
Total	54,478	9,720	71.4	20.6	58.6	20.7

Table 4.3: Result of path set generation using third link cost function

Num. transfer	$\mathcal{P}_R \cap \mathcal{P}_i$	$\mathcal{P}_i \setminus \mathcal{P}_R$	$\frac{\mathcal{P}_R \cap \mathcal{P}_i}{\mathcal{P}_R}$(%)	0 transfer (%)	1 transfer (%)	2 transfers (%)
0 (23.4 %)	14,319	2,243	99.1	100	-	-
1 (53.5 %)	33,631	10,077	87.2	-	100	-
2 (13.3 %)	7,476	3,655	77.7	-	-	100
1,2 (7.8 %)	8,709	1,270	79.0	-	50.5	49.5
0,2 (1.0 %)	1,018	171	73.6	49.1	-	50.9
0,1 (0.9 %)	904	220	78.4	48.6	51.4	-
0,1,2 (0.1 %)	67	7	62.0	32.4	32.4	35.2
Total	66,124	17,643	86.7	20.6	58.6	20.7

In table 4.1, 4.2, 4.3, the first column indicates portion of $O-D$ pair according to number of transfer. For example, in 53.5% of Origin-Destination pairs, paths which passengers actually travelled have only one transfer. From the tables, if time is the only cost factor, coincidence rate of paths is only 70.9 %. And if we use time and crowding as cost factors, coincidence rate only improves 0.5 %. However, if we use link cost function reflecting both crowding and passengers' preferences, coincidence rate improves about 15 %. This coincidence rate is higher than S Fiorenzo-Catalano's result and lower than Tan's result. However, in our result, the percentage of generated paths being a actual path is 78.9 %, which is much higher than Tan's results.

5. Conclusion and Further research

The contents which are dealt in this study, are as follows.

After observing passengers' real flow in Seoul Metro, we found out that we should take crowding and taste heterogeneity of passengers into account to develop a precise link cost function. We observed some difficulties of development process which is caused by congestion cost of path. We resolved these difficulties by calculating congestion cost of link differently depending on the origin-destination pair. By using inverse optimization, we made estimated joint and marginal distribution of taste parameters. We generated path choice set in Seoul Metro based on the new link cost function. The results turned out that coincidence rate is about 86 %, which is better than using previous link cost function.

In this paper, passengers who used Seoul Metro during peak hour(7:00 AM $\sim$ 10:00 AM, 6:00 PM $\sim$ 9:00 PM) in November 21, 2011, are used in experiment. Travel pattern of passengers may be different depending on day of the week or a time zone. Therefore, it is required to compare the changes in parameters of a link cost function, according to day of the week or a time zone. And in our result, except for the case where an $O-D$ pair has 0 or 1 transfer, coincidence rate is below 80%. So a method which increases coincidence rate in other cases, is required.

References

[1] 김익기 이재섭. 대중교통 통행배정을 위한 링크기반 확률적 알고리즘 연구: 도시철도 통행배정을 대상으로. *대한교통학회지*, 19(2):53–60, 2001.

[2] 정동재. 대중교통 통행배정을 위한 링크기반 확률적 알고리즘 연구: 도시철도 통행배정을 대상으로. *서울대학교 석사 졸업 논문*, 2012.

[3] 홍성필. 경영과학 제 2판. *율곡출판사*, 2014.

[4] Takashi Akamatsu. Cyclic flows, markov process and stochastic traffic assignment. *Transportation Research Part B: Methodological*, 30(5):369–386, 1996.

[5] Michael GH Bell. Alternatives to dial's logit assignment algorithm. *Transportation Research Part B: Methodological*, 29(4):287–295, 1995.

[6] Richard Bellman and Robert Kalaba. On kth best policies. *Journal of the Society for Industrial and Applied Mathematics*, pages 582–588, 1960.

[7] M Ben-Akiva, MJ Bergman, Andrew J Daly, and Rohit Ramaswamy. Modeling inter-urban route choice behaviour. In *Proceedings of the 9th International Symposium on Transportation and Traffic Theory, VNU Press, Utrecht*, pages 299–330, 1984.

[8] 임강원 · 임용택. 교통망분석론(개정판). *서울대학교출판부*, 2008.

[9] Robert B Dial. A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation research*, 5(2):83–111, 1971.

[10] Stella Fiorenzo-Catalano, Rob Van Nes, and Piet HL Bovy. Choice set generation for multi-modal travel analysis. *European journal of transport and infrastructure research EJTIR, 4 (2)*, 2004.

[11] Sung-Pil Hong, Kyung Min, and Geunyeong Byeon. Transit assignment embracing diversity as well as rationality. *submitted*.

[12] Sung-Pil Hong, Yun-Hong Min, Myoung-Ju Park, Kyung Min Kim, and Suk Mun Oh. Precise estimation of connections of metro passengers from smart card data. *Transportation*, pages 1–21, 2015.

[13] Kyung Min Kim, Sung-Pil Hong, Suk-Joon Ko, and Dowon Kim. Does crowding affect the path choice of metro passengers? *Transportation Research Part A: Policy and Practice*, 77:292–304, 2015.

[14] Eugene L Lawler. *Combinatorial optimization: networks and matroids*. Courier Corporation, 1976.

[15] Traffic Assignment Manual. Bureau of public roads. *US Department of Commerce*, 1964.

[16] Ernesto de Queiros Vieira Martins. An algorithm for ranking paths that may contain cycles. *European Journal of Operational Research*, 18(1):123–130, 1984.

[17] J.D. Ortuzar and L.G. Willumsen. Modelling transport. *Wiley*, 1994.

[18] Bing-Feng Si, Ming Zhong, Hao-Zhi Zhang, and Wen-Long Jin. An improved dial's algorithm for logit-based traffic assignment within a directed acyclic network. *Transportation planning and technology*, 33(2):123–137, 2010.

[19] Heinz Spiess. Technical note-conical volume-delay functions. *Transportation Science*, 24(2):153–158, 1990.

[20] Rui Tan, Steve Robinson, Der-Horng Lee, and Moshe E Ben-Akiva. Evaluation of choice set generation methods with smart card data in large-scale public transport network.

[21] Roger L Tobin. An extension of dial's algorithm utilizing a model of tripmakers' perceptions. *Transportation Research*, 11(5):337–342, 1977.

[22] SC Wong. On the convergence of bell's logit assignment formulation. *Transportation Research Part B: Methodological*, 33(8):609–616, 1999.

초 록

서울대학교 대학원
산업·조선공학부
황 준 하

지하철을 이용하는 승객들은 대안경로 중에 본인이 판단하기에 가장 좋은 경로를 선택하여 이동한다. 따라서 승객이 어떤 경로를 선택하여 이동하는지 추정하는 것은 단순히 승객의 이동경로를 파악하는 데에 그치지 않고 그 안에 숨어 있는 승객의 합리성과 다양성을 밝힐 수 있는 중요한 자료가 된다. 또한 배차 간격 조정, 급행열차 도입, 역내 시설 개선과 같이 지하철 시스템의 정책결정 및 운영을 위해서는 승객이 어떻게 지하철을 이용하고 있는지 알아야한다. 최근 스마트 카드 데이터를 활용하여 서울 지하철 네트워크를 이용하는 승객의 경로를 정확히 추정할 수 있는 방법과 경로 선택에 있어 승객의 다양성과 합리성을 반영할 수 있는 방법론이 개발되었다. 이는 통행량과 승객의 다양성을 반영한 호 비용함수의 개발이 가능하다는 것을 의미한다. 본 연구에서는 통행량과 승객의 다양성을 고려한 호 비용함수를 개발한다. 그리고 Dial 알고리듬을 활용하여 통행량과 승객의 다양성을 반영한 호 비용함수가 그렇지 않은 호 비용함수들보다 승객의 실제 이동 행태를 더 잘 반영함을 확인한다.

주요어 :
Dial 알고리듬, 호 비용 함수, 승객들의 선호, 통행량

학번 : 2013 – 23212