TRANSIT ASSIGNMENT MODEL FOR EXPRESS TRAINS BASED ON INVERSE OPTIMIZATION

KYUNG MIN KIM

Industrial Management and Engineering Department, Myongji University, South Korea

ABSTRACT

A loading balance between express and local trains is a crucial issue on the train operation plan of urban railway, which provides two services on the same line. This is because passenger's preference is affected by the train operation plan, such as headway and halting patterns. Recently, Seoul Metro Line 9, in Korea, has experienced an imbalance, as the preference for express train over local train services is significantly higher, so the load factor for the express train was twice as much as for the local train. As such, to analyse load balance, a schedule-based transit assignment to determine the ridership by section and by train is needed when a timetable for trains and passengers' trips is given. This paper develops the cost function considering the transfer between express and local trains as well as the train load factor in the time-expanded network via inverse optimization. We suggest a convex optimization model to ensure the path that the passenger has selected will cost less than the alternative path, and from this model, the parameters of the deterministic cost function were estimated. We employed the Frank-Wolfe-based transit assignment algorithm to estimate the equilibrium flow using our cost function. The algorithm was tested to Seoul Metro Line 9 in order to verify its accuracy. The results show that the difference in the actual ridership by train was less than 10%. Finally, we simulated reducing the load difference between the express and the local train through re-scheduling to increase the service frequency of express trains.

Keywords: inverse optimization, transit assignment, express trains.

1 INTRODUCTION

Train operation planning for an urban railway, which provides both express and local train service simultaneously, is necessary to maintain load balance between the two services [1]. This is because passenger preference is affected by the train operation plan, such as headway and halting patterns. Recently, Seoul Metro Line 9 in Korea has experienced an imbalance, as the preference for express train over local train service is significantly higher, so the load factor for express trains was twice as much as for local trains. As such, to analyse load balance, schedule-based transit assignment to determine the ridership by section and by train is needed when the timetable for trains and passengers' trips is given.

Schedule-based transit assignment in a time-expanded network allows for the expression of the events of arrival and departure train and the complete passenger's flow – boarding, alighting, and transfer. In a time-expanded network, we can analyse individual passenger choices. Although GPS or a Smart Card system can be used to collect massive and reliable trip data, difficulties in handling a large size network and obtaining massive data reveal that preferences at the microscopic level may have hampered the implementation of schedule-based transit assignment in real cases. However, Hong et al. [2] recently developed a precise algorithm that detects an actual train-path for metro passengers based on an observation that tag-out times of passengers from Smart Card data were clustered.

In this paper, we propose a deterministic path cost function with three attributes – in-vehicle time, wait time, and transfer time between express and local trains – to develop a schedule-based transit assignment model. Congestion effect measured by step function of relative weight on in-vehicle time increasing to predetermined crowding level. Stochastic assignment models which are based on random utility maximization (RUM) theory such as

logit estimate the coefficient of the cost function to maximize likelihood. However, this paper estimates parameters using inverse optimization due to assuming the deterministic path cost.

An optimization (forward optimization) addresses the problem of finding a point, an optimal solution, in the region delineated by a set of functions, or constraints, which minimizes a cost function called an objective. An *inverse optimization*, on the other hand, is the problem of calibrating a set of parameters in terms of objective and/or constraints, so that a prescribed solution becomes optimal [3]. The parameters should be what best explains user-optimality, underlying the path choice of passengers. This implies that an inverse optimization has relevance in doing so. We will show that the inference problem can be posed as a convex quadratic optimization that allows an efficient solution method.

Tested on a daily set of trains, with parameters calibrated by the inverse optimization, from Seoul Metro line 9, the results showed a mean absolute percent error of deviation between actual and predicted passenger load of trains within 10%. Our model seems precise enough for the prediction of passenger behaviour required by the city of Seoul.

This paper is organized as follows. In Section 2, we describe the time-expanded network and cost function in the proposed transit assignment model. Section 3 describes the methodology of inverse optimization to infer the coefficient of parameters of the cost function. Section 4 evaluates the path choice predictability of our model in comparison with actual choices. Finally, Section 5 provides some concluding remarks, and suggests some directions for further research.

2 TIME-EXPANDED NETWORK AND COST FUNCTION

Time-expanded network $G = (N, E)$ is defined as Set N of place-time-defined nodes and Set E of edges, which is defined as Subset of $N \times N$ [4]. As shown in Fig. 1, each node i $\in$ N has place, i.e., station and time value as attribute, and each edge represents an event in which a train departs from each station or arrives at the station.

On the time-expanded network, path P can precisely represent the movement of passengers who use local and express trains. As shown in Fig. 1, for example, a passenger who moves to Station E from Station A where local both trains and express trains stop can have *P1, P2* and *P3* as an *alternative path*. That is, the *alternative paths set* of this passenger is {*P1, P2, P3*}. *P1* means a path in which a passenger transfers to an express train at Station C after using a local train, while *P2* and *P3* are the paths only for express train and local train, respectively.

Deterministic path cost function with three attributes – wait time, in-vehicle time and transfer time between express and local trains for schedule-based transit assignment model. The cost of path P on the time-expanded network is assumed to be a linear function, as shown in eqn (1). In-vehicle time is divided into 3 categories depending on the degree of in-vehicle crowding

$$c(P) = \alpha W T_P + \beta_1 IVT_P^L + \beta_2 IVT_P^N + \beta_3 IVT_P^H + \gamma TT_P, \tag{1}$$

- WT_P : Waiting time of Path P
- IVT_P^L : In-vehicle time given less than 100% congestion-rate on Path P (low-level ridership)
- IVT_P^N : In-vehicle time given 100%~150% congestion-rate on Path P (normal-level ridership)
- IVT_P^H : In-vehicle time given more than 150% congestion-rate on Path P (high-level ridership)
- TT_P : Transfer time between local and express trains on Path P
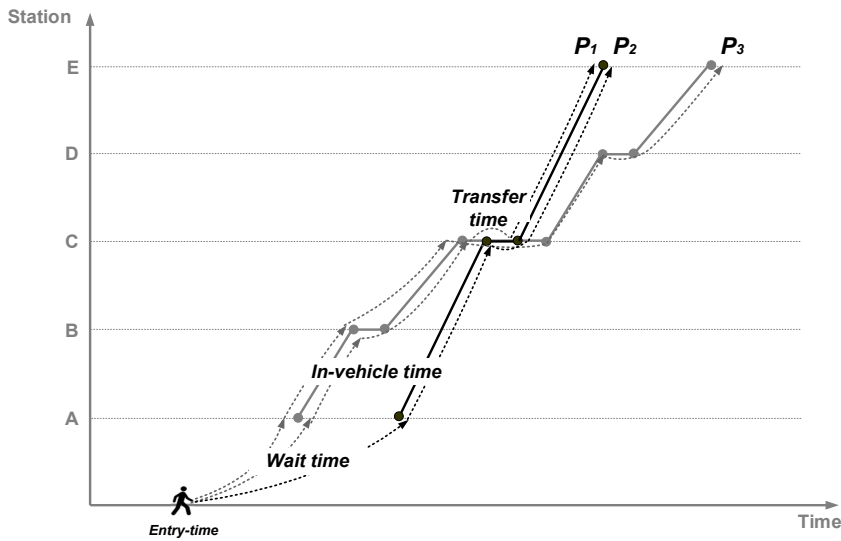
Figure 1: Time-expanded network and alternative paths set.

3 ESTIMATION OF COST FUNCTION VIA INVERSE OPTIMIZATION

In the stochastic model based on the Random Utility Maximization (RUM) theory including the logit model, the coefficient of cost function is estimated to be Maximum Likelihood Estimation (MLE). However, if the path cost is deterministic, as it is in this study, MLE is not available. As such, this study uses the inverse optimization model to estimate the coefficient of cost function.

Given that the path selection of passengers is selfish routing [5], $c(P) \leq c(Q)$ shall be fulfilled in the event that there are Path P selected by a passenger in equilibrium conditions, and alternative path Q. However, there are passengers who do not meet $c(P) \leq c(Q)$ due to an assumption of cost factor and function, which are not taken into account. As such, the convex quadratic optimization model was developed to minimize error related to those passengers

$$\min \quad \frac{1}{2}\sum_i e_i^2, \tag{2}$$

$$\text{s.t} \quad c(P^i) \leq c(Q^i) + e_i, \quad ,\forall i \in I, \tag{3}$$

$$\beta_1 = 1, \tag{4}$$

$$e_i \geq 0, \quad \forall i \in I, \tag{5}$$

$$\alpha, \beta_2, \beta_3, \gamma \geq 0. \tag{6}$$

Eqn (2) is an objective function to minimize error related to all passengers. Eqn (3) is a restriction in which the cost of the path that a passenger selects shall be less than that of an alternative path in equilibrium conditions. Eqn (4) means a restriction such that one of coefficients shall be fixed to ensure a feasible solution that makes all the coefficients 0 does not occur.

4 EXPERIMENT RESULTS

4.1 Data set and estimation result of inverse optimization

This section describes the implementation of the proposed model for the Seoul Metro 9. Seoul Metro 9 has operated both the local and express train since it launched commercial service in 2009. As shown in Fig. 2, there are 12 express stations and 18 express-local stations. Express and local trains are being serviced alternately every 3 min during the morning peak hour.

The RP-data of the real train choice of passengers over the interval was acquired through the path recovering method of Hong et al. [2] that was applied to two-day sets of Smart Card data: one on Monday, May 11 and another on Monday, May 18, 2015. The first data set is used for coefficient estimation, and the second data set for validation.

Table 1 shows coefficients of cost function per path estimated for 156,257 passengers during the peak hours (07:00 ~ 10:00) on Monday, May 11, 2015. The number of passengers with e_i more than 0 is 21, 148, which accounts for 13.5%. Coefficient for waiting time is 1.27, which appeared to be higher than in-vehicle time given less than 100% congestion. If the congestion-rate is 100%~150%, in-vehicle coefficient is 1.13, and if the congestion-rate is 150% or more, in-vehicle coefficient is 1.34. As such, the cost that the passengers feel is increased, even given the same in-vehicle time, if ridership increases. Finally, the coefficient of transfer between the local and express train is 4.86, and resistance to transfer appeared to be very high.
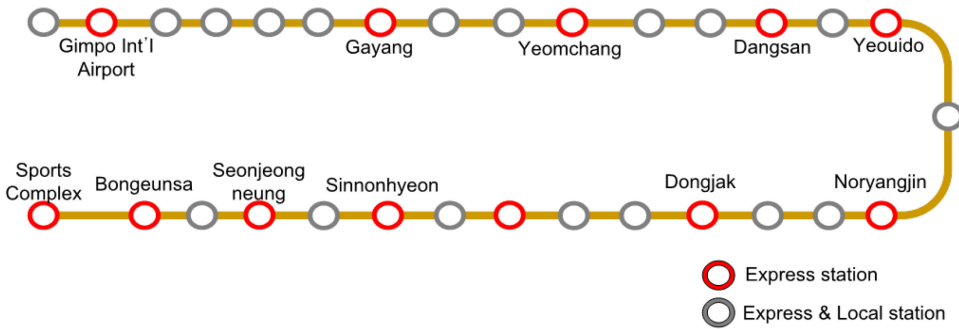

Figure 2: Map of Seoul metro line 9.

Table 1: Estimated parameters.

Parameter	α	β_1	β_2	β_3	γ
Coefficient	1.27	1	1.13	1.34	4.86

4.2 Performance evaluation

We conducted transit assignment using the cost function estimated in this study. The MSA-based Frank-Wolfe algorithm was applied, which is commonly used. Given that ridership of the observed train is f and flow of the estimated train is $\hat{f}$, the deviation of $\hat{f}$ from f is measured in both mean squared error (MSE) and mean absolute error (MAE), as in eqns (7) and (8)

$$\text{MSE} = \frac{1}{|E|} \sum_{e \in E} (\hat{f}_e - f_e)^2, \tag{7}$$

$$\text{MAE} = \frac{1}{|E|} \sum_{e \in E} \left| \frac{\hat{f}_e - f_e}{\hat{f}_e} \right|. \tag{8}$$

Table 2 shows the results of MSE and MAE calculated from the Estimation data set and Validation data set. For MSE, the difference in actual number of passengers is approximately 70 on average, and MAE is 0.09. Therefore, error in the transit assignment model presented in this study is approximately 10% or less.

Our transit assignment model implemented to Demand Responsive Metro Train Operation Planning Software (Dr.METRO) developed by KRRI [6]. Fig. 3 illustrates transit assignment results of Seoul metro line 9 using Dr.METRO. Dr.METRO display the ridership of each section and train by the color. Red, orange, yellow and green train-path indicate over 200%, 150%~200%, 150%~100% and less 100% congestion-rate, respectively. In time-space window 1 (from 07:20 to 08:40 and from Gayang station to Dongjak), we found almost every express and local train's congestion-rate are over 200%. Congestion-rate in express train is continued in time-space window 2, however, that of local train is decreased dramatically. So that, passengers prefer express trains.

Table 2: MSE and MAE of estimated link flows.

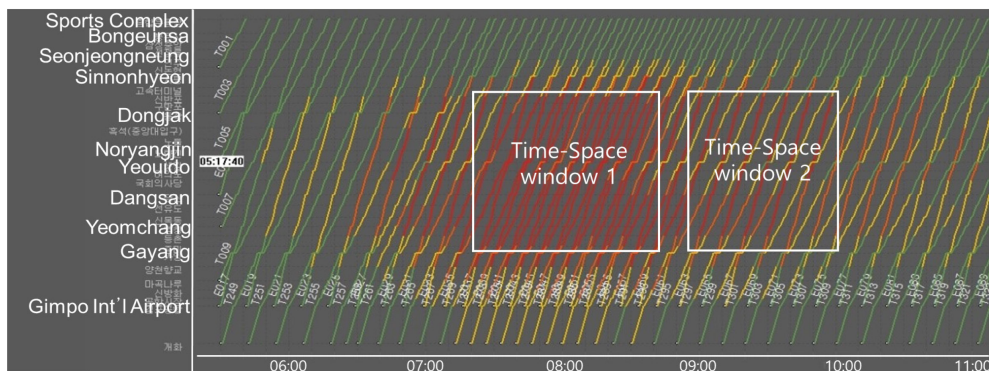|	MSE	MAE
Estimation set	66.2	0.08
Validation set	70.9	0.09

Figure 3: Transit assignment result of Seoul metro line 9 in a view of train-path diagram.

5 CONCLUSION

Express and local train service in urban railway provides an opportunity for passengers to meet the varying needs of passengers. After successfully launched on Seoul Metro Line 9, now the Korean government is considering an express-local train service on existing lines.

This study presents a methodology that estimates the coefficient of deterministic path cost for transit assignment by applying inverse optimization in the express-local train operation conditions. More specifically, our convex quadratic optimization model has estimated the coefficient of cost function to minimize passengers who violate the selfish routing condition. In addition, we validated the prediction of the model by applying the estimated cost function to the Seoul Metro 9. Through validation, it was shown that difference between actual and estimated ridership per train appeared to be less than 10%. Therefore, our model can be used to adjust train operation plan to reduce load imbalance between the local and express trains.

For future work, it would be interesting to develop the transit assignment model whose performance is better than the current ones and investigate the methodology that integrate the inverse optimization and stochastic model.

ACKNOWLEDGEMENT

REFERENCES

[1] Transit capacity and quality of service manual (3rd ed.), Technical Report TCRP Report 24 165, *Transportation Research Board of the National Academies*, 2013.

[2] Hong, S.-P., Min, Y.-H., Park, M.-J., Kim, K.M. & Oh, S.M., Precise estimation of connections of metro passengers from Smart Card data. *Transportation*, **43**(5), pp. 749–769, 2016.

[3] Ahuja, R.K. & Orlin, J.B., Inverse optimization. *Operations Research*, **49**(5), pp. 771–783, 2001.

[4] Hamdouch, Y. & Lawphongpanich, S., Schedule-based transit assignment model with travel strategies and capacity constraints. *Transportation Research Part B*, **42**, pp. 663–684. 2008.

[5] Correa, J.R., Schulz, A.S. & Stier-Moses, N.E., Selfish routing in capacitated networks. *Mathematics of Operations Research*, **29**(4), pp. 961–976, 2004.

[6] Oh, S.M. & Min, J.H., Dr. METRO: A demand-responsive metro-train operation planning program. *WIT Transactions on The Built Environment*, vol. 135, WIT Press: Southampton and Boston, pp. 595–601, 2014.