Correlation between characteristic of subway network and congestion in subway train

Yeonggyeong KANG
Graduate student, Department of Urban Engineering, University of Seoul, Seoul, Republic of Korea

Changhyo YI, Jooyeon GO
Research professor, Department of Urban Engineering, University of Seoul, Seoul, Republic of Korea

Seungil LEE
Professor, Department of Urban Engineering, University of Seoul, Seoul, Republic of Korea

ABSTRACT

Highest level of congestion in a subway train of the Seoul Metropolitan Area (the SMA) reaches 240% during rush hour in the morning (from the survey conducted by Ministry of Land, Infrastructure and Transport: MOLIT). Level of congestion is defined as percentage of ratio of capacity to the number of passengers. Transport capacity of most of subway trains run in Seoul is 160 persons per a subway train. High level of congestion in subway train causes unavoidable physical contact, risk of crime and accident. Then it lowers the level of service of subway. Meanwhile, the level of congestion in a certain link is very low in spite of at the same time. This imbalance brings inefficiency in subway investment. For even distribution of congestion, it is important to identify what correlate with congestion in subway train.

The goal of this paper is to find the factors of congestion in metro train. Hypothesis of this paper is that adjacent land use and network characteristic affect the number of passengers in metro train. Network characteristic used in this research is an index called betweenness centrality and the number of transfer stations. For empirical analysis, multiple linear regression model was used. The response variable is level of congestion in a subway train during rush hour in the morning on weekdays.

Findings from the study are as follows: Land-use factors have bigger influence on congestion than network factors do. Network characteristic affects the passengers in metro train and area of influence of land use factor has certain limit. Land-use factors have bigger influence on congestion than network factors do.

This result expected to be used as basic research to make subway system efficient through distributing the passengers of metro train.

Keywords: Subway, Congestion, Betweenness centrality, TOD (Transit Oriented Development), ROD (Rail Oriented Development)
1. Introduction

Subway captures 35.2% of total trips (Seoul Statistics-Population Trend in 2010/Seoul Statistics-Composition of Daily Passenger Transportation in 2009). It is appropriate for mass transit in a metropolitan city because it transports mass of people without being influenced by or influencing road congestion. However, its transport capacity has its limit like other transport modes. Highly congested train causes inconvenience such as unavoidable physical contact, higher risk of crimes and accidents. Level of congestion is defined as percentage of ratio of capacity to the number of passengers. Transport capacity of most of subway trains run in Seoul is 160 persons per a subway train. According to survey conducted by MOLIT, the level of congestion in metro train of the SMA(the Seoul metropolitan area) reaches up to 240% during rush hour in the morning. In recent years, highly congested train becomes an ever growing issue in New York, Toronto and Sao Paulo. (1)

Meanwhile, congestion level depends on the direction of subway and the local position even in the same time, and its imbalance is serious (Jang, Kang and Lee, 2012). The previous studies about TOD or compact city mainly focused on high density and mixed development. However, this approach results in imbalance of passengers of subway. Considering enormous cost of construction, the imbalance causes inefficiency in operating subway. To solve this problem, land use plan of the catchment area of the subway station should be established comprehensively in consideration of even distribution of congestion. For this, the extensive investigation and detailed analysis on patterns of subway use are required (Choi, Koo, Lee, Kim and Sung, 2012). Based on this background, this paper is aimed to find factors influencing congestion of the Seoul metropolitan a subway train.

2. Previous Studies

2-1. Transit Oriented Development

Calthorpe(1993) suggested Transit Oriented Development. It featured the concept that “moderate and high-density housing, along with complementary public uses, jobs, retail and services, are concentrated in mixed-use developments at strategic points along the regional transit system.”(Calthorpe, 1993) Further to this, Cervero and Kockelman (1997) examined the connection empirically between the 3Ds of the built environment and travel demand. 3Ds mean density, diversity, and design. (Cervero&Kockelman, 1997) From this empirical study, the concept of TOD developed rapidly. Follow-up studies focused on emphasis of public transport or demographic characteristics of individual and household or extend of application range of TOD and so on. (Choi & Koo, J., & Lee, S. & Kim, T., & Sung, 2012) In this regard, the influencing factors affecting demand of subway have been studied. Shin and Lee (2012) analyzed influencing factors TOD planning elements affecting demand of the Seoul subway. The study showed that influencing factors are mix of land use, mix of development density and accessibility to various transport.

However, in these previous studies, public transport merely was approached in terms of increasing of user. This fragmentary approach do not consider imbalance of public transport user. Imbalance of public transport user causes inefficiency of public transportation investment. This is eventual motivation of this paper.
2-2. Congestion in a subway train

Promoting subway is a valuable objective in terms of reduction of greenhouse gas emission and energy savings, but highly concentrated use in specific range lowers level of service of subway. However, it is difficult to evaluate congestion because congestion of subway does not have influence on punctuality unlike one of road. In response, Prud'homme, Koning and Lenormand (2012) equated missed convenience in congested subway with wasted time because of road congestion and evaluated congestion cost from the survey targeting traveler of Paris subway. Interviewers asked them the willingness to pay for non-congested travel to actual congestion levels. This study showed that what travelers would be ready to pay to avoid congestion in the Paris subway is on average about three times the amount of their out-of-pocket payments. Also, it showed an 8% increase in congestion densities experienced over the 2002–2007 period implies a welfare loss of at least 75 M€/year.

Thus, congested subway makes inefficiency. Jang et al.(2012) conducted an empirical study on influences of land use characteristics to congestion in a subway train. In the study, range of influence of land use was supposed as range of 10 subway stations. It is based on investigation that average commute distance is 9.4km in Seoul and average distance between stations is 1km (Public transportation using status survey 2010 by Seoul municipality). The study used multiple regression analysis, and its explanatory variables are total floor space by use in previous 10 stations and next 10 stations. It showed that non-residential floor space’s influence is significant statistically while residential floor space rarely influences congestion. Especially, Businesses floor space has high statistical influence.

3. Methodology
3-1. Multiple regression analysis

Multiple regression analysis is used to learn statistically relationship between more than two explanatory variables and a response variable. Multiple linear regression model assume that the each explanatory variable affects linearly a response variable. This can be represented as follows.

\[ y = \sum_{i=1}^{n} x_i \beta_i + \beta_0 + \varepsilon \]

Where \( y \) is a response variable and \( x_1, x_2, x_3 \cdots x_n \) are explanatory variables. Then, the expected value of \( y \) and the estimates of \( \beta_i \) and \( \beta_0 \) are

\[ \hat{y} = \sum_{i=1}^{n} x_i \hat{\beta}_i + \hat{\beta}_0 \]

\( \hat{\beta}_i \) represents average change amount of a response variable induced by changes in explanatory variables \( x_i \). It is estimated through ordinary least square method (OLS) in this study. However, regression analysis cannot identify causal relationship. For this, it is needed more qualitative study. The objective of this study is to identify statistically the explanatory variables relating with congestion in the subway train.
3-2. Assumptions & Variable selection

This study focuses on land-use within the range which Jang et al. (2012) analyzed. However, this study has some differences and improvements. First, instead of floor space of each station, the average floor space of previous 10 stations and that of next 10 stations were used to show difference between influences of previous range and next range. Second, instead of the floor space by urban land-use, the total floor space and the ration of the non-residential floor space to the total floor space were selected as explanatory variables. Third, this study focuses on not only land-use but also network. In aspect of network, betweenness centrality and the number of transfer within specified range were added.

Response variable

The unit of analysis is each link between every closest station on the same subway line, which are line number 1 to line number 8 in Seoul. Response variable is average congestion level in a subway trains passing through these links from 8 a.m. to 9 a.m. on weekdays. This assumed that, in this time period, destinations of subway users are similar to each other: from residential area to non-residential area, and the influence of land use is shown clearly. The average congestion level is lower than instant congestion level, which is practical sensory congestion level. However, to use instant congestion level as a response variable obscures analysis because it depends on local region and time to the minute. Hence, this paper selected average congestion level as a response variable.

Explanatory variables: Land use

The distribution of land-use determines the locations of human activities. The distribution of human activities in space requires trips in the transport system to overcome the distance between the locations of activities (Wegener & Fürst, 2004). On this basis, this study hypothesized that, for direction of movement of subway, land uses on the catchment area of the previous 10 stations and the next 10 stations by each link influence congestion of that link. The “link” means the section between each station. The each range, which is 10 stations, is based on the study of Jang et al. (2012). The catchment area of each station was defined within a 500-m radius from the center of a station. This is based on the previous study of the catchment area of subway station (An et al., 2012).

In the morning during weekdays, it is assumed that subway users move from residential areas to non-residential areas. Hence, this study selected the average of the total floor space density (total floor space density) and the average of non-residential floor space ratio to the total floor space (non-residential ratio) in the previous range (10 stations) as explanatory variables. The same things in the next range were selected. Density represents intensity of land use and is defined as follows:

\[
\text{density} = \frac{\text{total floor space}}{\text{the surface area of the catchment area (the circle with a radius 500m)}}
\]

These explanatory variables were selected to answer these research questions: Does
influence of land use depend on the direction of the movement of a subway? Which is a more major influencing factor - the intensity of land use or the mix of land use?

*Explanatory variable: Network*

Centrality determines the importance of a point within a network. It can be quantified by a few methods, which are degree centrality, betweenness, closedness and so on. Of these, betweenness centrality is defined as “a point in a network is central to the extent that it falls on the shortest path between pairs of other points” (Freeman, 1977). If, in traffic network, it is assumed that each individual is rational and selects the shortest path, betweenness centrality represents the necessity of a point in network. In this study, it was calculated upon each link within the subway network unlike “general” betweenness centrality calculated upon each node because the observation of this study is each link. Nonetheless, it still represents the necessity of the link.

Also, there are transfer stations in the subway network. Land use on the catchment area of these transfer stations may be a significant variable. To consider this influence, explanatory variables includes the numbers of transfer stations within the previous range and the next range influence congestion.

As mentioned so far, the variables of regression analysis in this study are in Table 1

<table>
<thead>
<tr>
<th>Response variable</th>
<th>$y$</th>
<th>The average level of congestion in a subway train for one hour from 8 a.m. to 9 a.m. (unit: %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explanatory variables</td>
<td>$x_1$</td>
<td>Average of total floor space of land use on the catchment area of the previous 10 stations</td>
</tr>
<tr>
<td></td>
<td>$x_2$</td>
<td>Average of total floor space of land use on the catchment area of the next 10 stations</td>
</tr>
<tr>
<td></td>
<td>$x_3$</td>
<td>Average of the total floor space and the average of non-residential floor space ratio to total floor space in the previous range</td>
</tr>
<tr>
<td></td>
<td>$x_4$</td>
<td>Average of the total floor space and the average of non-residential floor space ratio to the total floor space in the next range</td>
</tr>
<tr>
<td></td>
<td>$x_5$</td>
<td>Betweenness centrality</td>
</tr>
<tr>
<td></td>
<td>$x_6$</td>
<td>The number of transfer stations within the previous range</td>
</tr>
<tr>
<td></td>
<td>$x_7$</td>
<td>The number of transfer stations within the next range</td>
</tr>
</tbody>
</table>

3-3. Data

The congestion data for a subway train was obtained from each subway operating agency (Seoul Metro and Seoul Metropolitan Rapid Transit Corp). Data for land use of the catchment area of the subway station was calculated with GIS spatial analysis tool based on the account book for taxation in 2008. Betweenness centrality was also calculated with GIS spatial analysis tool.
4. Results

Table 2. Descriptive statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>The average level of congestion</td>
<td>488</td>
<td>2.00</td>
<td>176.00</td>
<td>69.9656</td>
<td>40.0783</td>
</tr>
<tr>
<td>total floor space density (prev.)</td>
<td>488</td>
<td>.016834570</td>
<td>2.655203300</td>
<td>1.63962901384</td>
<td>.414939927688</td>
</tr>
<tr>
<td>total floor space density (next)</td>
<td>488</td>
<td>.016834570</td>
<td>2.655203300</td>
<td>1.63011087516</td>
<td>.41203128626</td>
</tr>
<tr>
<td>non-residential ratio (prev.)</td>
<td>488</td>
<td>.094941100</td>
<td>1.000000000</td>
<td>.49053751572</td>
<td>.150004256538</td>
</tr>
<tr>
<td>non-residential ratio (next)</td>
<td>488</td>
<td>.094941100</td>
<td>1.000000000</td>
<td>.4895135041</td>
<td>.147794830666</td>
</tr>
<tr>
<td>Betweenness centrality</td>
<td>488</td>
<td>468</td>
<td>29715</td>
<td>7652.41</td>
<td>5547.672</td>
</tr>
<tr>
<td>The number of transfer station (prev.)</td>
<td>488</td>
<td>0</td>
<td>7</td>
<td>3.43</td>
<td>1.625</td>
</tr>
<tr>
<td>The number of transfer station (next)</td>
<td>488</td>
<td>0</td>
<td>7</td>
<td>3.43</td>
<td>1.575</td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>488</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The regression model of this study proved to be statistically significant within 1% of the significance level. Adjusted R square is 0.431. That is to say, this model explained 43% of variance of congestion in a subway train.

Table 3. Model summary

<table>
<thead>
<tr>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>.663a</td>
<td>.440</td>
<td>.431</td>
<td>30.21929</td>
<td>.000</td>
</tr>
</tbody>
</table>

The explanatory variables in aspect of land use proved to be statistically significant within 1% of the significance level. The influence of land use in the next range proved to be bigger than that of the previous range. The higher total floor space density or non-residential ratio in the previous range, the lower level of congestion is and in next range vice versa.

Of the explanatory variables in aspect of network, betweenness centrality proved to be statistically significant within 1% of the significance level. The number of transfer within the previous range proved to be statistically significant within 10% of significance level but that within the next range proved not to be statistically significant.
Table 4 Coefficient estimates

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t-value</th>
<th>Sig.</th>
<th>Collinearity Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
<td>Tolerance</td>
</tr>
<tr>
<td>(Constant)</td>
<td>15.099</td>
<td>8.408</td>
<td>-1.796</td>
<td>.073</td>
<td></td>
</tr>
<tr>
<td>total floor space density(prev.)</td>
<td>-16.729</td>
<td>4.132</td>
<td>-.173</td>
<td>-4.049</td>
<td>.000</td>
</tr>
<tr>
<td>total floor space density(next)</td>
<td>26.432</td>
<td>4.199</td>
<td>.272</td>
<td>6.294</td>
<td>.000</td>
</tr>
<tr>
<td>non-residential ratio(prev.)</td>
<td>-37.603</td>
<td>13.433</td>
<td>-.141</td>
<td>-2.799</td>
<td>.005</td>
</tr>
<tr>
<td>non-residential ratio(next)</td>
<td>110.502</td>
<td>13.282</td>
<td>.407</td>
<td>8.320</td>
<td>.000</td>
</tr>
<tr>
<td>Betweenness centrality</td>
<td>.001</td>
<td>.000</td>
<td>.130</td>
<td>3.232</td>
<td>.001</td>
</tr>
<tr>
<td>The number of transfer station(prev.)</td>
<td>-2.326</td>
<td>1.356</td>
<td>-.094</td>
<td>-1.716</td>
<td>.087</td>
</tr>
<tr>
<td>The number of transfer station (next)</td>
<td>1.272</td>
<td>1.358</td>
<td>.050</td>
<td>.937</td>
<td>.349</td>
</tr>
</tbody>
</table>

5. Discussion
5-1. Influence of land use

Standardized coefficients help to compare difference of influence between explanatory variables having quite different units. With standardized coefficients, the influence of land use - non-residential ratio, total floor space density - is bigger than that of network. High density of the total floor space is likely to mean commercial area since upper bound of floor area ratio in commercial area is higher than that in residential area and the supply of residential use land in commercial area has a limit. That is to say, the catchment area of the station with high total floor space density or non-residential ratio is likely to be a destination of subway users in the morning on weekdays. Hence, if this station exists in the next range, it is likely to make congestion in a subway train high.

On the other hand, if this station exists in the previous range, it is likely to make congestion in a subway train low. This result supports the previous study and the assumption of this study.

As an additional finding, non-residential ratio has a bigger influence on congestion than total floor density. High non-residential ratio means low job-housing balance. Low job-housing balance induce trip to commute to work or school in the morning on weekdays. Hence, if the catchment area of the station in next range has a high non-residential ratio, it is likely to make congestion in a subway train high. Mixed-use of non-residential land and residential is an important element of TOD. This result also supports the previous study about TOD.

5-2. Influence of Network

Betweenness centrality means the necessity of the link to move on the shortest path. Therefore, it is expected that betweenness centrality explain the demand of the link apart from land use. This assumption proved to be statistically significant. This means the importance of not only land use planning but also transportation planning in order to
evenly distribute congestion. Transportation planning should make high betweenness centrality in specific link low. However, a more detailed study about causal relationship is needed to use this result to make transportation planning: In which link this effect is bigger than land use? This remains a question for further research.

The number of transfer within the previous range is likely to explain that it has drain effect of subway user. However, the number of transfer within the next range is not statistically important. The regard of transfer station should be approached as land use. This remains a challenge for further research.

6. Conclusions

In promoting subway, even distribution of congestion is important in the aspect of efficiency of public transport investment. To consider this in planning, it is important to identify influencing factors of congestion. This is the reason that this study use regression analysis. This study selected explanatory variables based on in previous studies like total floor space density (Density), non-residential ratio (Diversity), which is similar to the 3Ds in TOD, and betweenness centrality and the number of transfer stations. The findings are as follows:

First, total floor space density and non-residential ratio (land use) have bigger influence on congestion in subway metro than betweenness centrality and the number of transfer stations (network).

Second, the most influencing factor is non-residential ratio in the next range and the land use in the next range has a bigger influence than the land use in the previous range.

Third, betweenness centrality has influence on congestion in subway metro. This study identified influencing factors on congestion in a subway train in light of both land use and network. This study is expected to be used as a basic study to make a transport and land use plan regarding congestion in a subway train. However, this study has limitations and further research questions. First, opportunity to transfer the subway line has less statistical influence. This study reflected opportunity to transfer as the number of transfer stations, but the approach regarding land use of the catchment area of a transfer station is more appropriate. Second, a more detailed study about betweenness centrality is needed: In which link this effect is bigger than land use? Third, this study did not consider an individual’s various trip behaviors. Some individuals may prefer to use nonstop subway in spite of longer trip. Further research should consider this trip behavior. These three points remain questions for further research.

Note
(1) Refer the following:
References


