SHORT CONTRIBUTION

GIS-Based Analysis of Railway’s Origin/Destination Path-Selecting Behavior

Zhongzhen Yang*
Department of Civil Engineering, Dalian University of Technology, Dalian 116023, China

&

Yoshitsugu Hayashi
Graduate School of Environmental Studies, Nagoya University, Nagoya 464-8603, Japan

Abstract: This study aims to analyze a railway’s OD path-selecting behavior with microsimulation analysis in big cities. In big cities, railway networks often consist of lines belonging to several different companies, making railway OD path-selecting behavior very complex. Usually, the second or third nearest stations are used instead of the nearest station. This study takes buildings rather than zones as origins or destinations for simulating railway path-selecting behavior. First, in the GIS database, we consider whether the destinations are within walking distance of their origins. Then we modify the map-based railway network with GIS to make it represent the real urban railway network more clearly. Then, from all the available stations for each building, and all the paths between the origin and destination stations, the one involving the shortest distance is the one used.

1 BACKGROUND

Land use and transportation models are essential tools in deciding on land use or transport policies. In both the four-step traffic demand model and activity-based travel analysis, we must calculate travel time/cost for each travel mode when estimating traffic assignment or modal split. In fact, this is a simulation of travellers’ rational behavior using mathematical models; the main theoretical basis is that travellers will select the shortest time/cost path when they travel between origin and destination. Therefore, the least time/cost path-searching method or the largest entropy method is usually applied to transportation networks (Bell et al., 1997). Since an urban railway network is less dense than a road network, simple methods are often used to estimate the path selection and its required time/cost. For example, railway company timetables are often used to find the time/fare between a pair of stations, and travellers are expected to use the nearest stations travelling from one zone to another. This method is useful whenever analysis is based on zone and there are only a few railway lines owned by just one company. However, there are often several companies operating many lines in large urban developments, making railway path-selection behavior much more complex. Travellers are then willing to compare the time/cost between several available OD paths to find the best one.

Since urban railways are less dense than road networks, there are often several stations with similar distances between origin and destination. In such cases, it is reasonable to suppose that travellers will choose stations on the basis of the direction they are travelling in. So even for the same traveller, the stations used will change accordingly. Very often, the second or third nearest station is chosen instead of the nearest one. Also, where railway networks consist of several companies’ lines, train connections and fare systems are complicated where movement

*To whom correspondence should be addressed. E-mail: yangzhongzhen@hotmail.com.
between different companies' lines is involved. Sometimes, even though changing from one company's line to another's saves time, and although travellers have to pay the lowest fare to the last company, this fare might be several times the additional fare for continuing to use the original line. As a result, travellers will continue to use the initial lines to the full or keep to the additional line and walk the rest of the way to their destination if it is then within walking distance. We can say that when urban railways have lines belonging to several companies, estimating OD travel time using timetables is inadequate. Moreover, traditional land use and transportation models used zone to assimilate travellers' attributes, and travellers in the same zone were considered to be with the same attribute. The main problem with this approach is the use of nonlinked, static trip-ends, generated and spatially distributed via aggregated households and zonal-level travel model (McNally, 1997). Therefore, simulated trip tables using this approach are not the real results of railway OD path-selecting behavior.

This study aims to simulate travellers' path-selecting behavior for getting accurate railway OD travel time, in the context of analyzing the traffic flow of railways in conurbations. To achieve this, we will use residential buildings and shops as trip origin or destination, and then analyze the OD path-selecting behavior in the context of shopping traffic. From the study of shopping traffic flow in each railway link, the effects of train frequency and fare on travellers' path selection can be accurately analyzed.

### 2 STUDY METHOD

Conventional transportation models use ordinary database rather than GIS database for input and output. The disadvantage of the ordinary database is that it cannot deal with data's spatial relationship. Spatial analyses between network and zone must be done manually. So while the database is being built up, available railway stations are selected for zones. According to the urban map and zone, traffic should flow to each station on a determined ratio (Asami et al., 1998; Ieda et al., 1998). In this study, instead of zone we use residences and shops as spatial analysis units. It means that we have to collect, manage, and produce large amounts of data and find spatial relationships between the subject matter of the data. Since an ordinary database is inadequate for the job, GIS (MapInfo & MapBasic) is used to manage the database.

Unlike road travel, railway links do not have an obvious speed-volume curve. However, railway fare attributes differ from road in both time and road cost attributes. Both of these are cumulative, while railway fares are stepped. For example, when a traveller changes from one company's line to another one, the required cost will increase with a big jump. Even within the same company's lines the fare is also a stepped one. For example, the same fare will be charged for travelling between 1–3 stations, while for travelling 4–5 stations, a higher fare will be charged but applied again to more than one station. This study will integrate the stepped fare system into the shortest path-searching algorithm. In the algorithm, as a node is added to the path tree we total the path distance for each different company's lines, and then calculate the fares paid to each company, according to each company's fare charging system. Finally, we convert fare to time with time value to get the generalized time index for the path in each search step. Here time value index from an existing study (Miwa and Morikawa, 1998) will be used.

We will use pair-wise comparison to solve the problem of using the second or third nearest station. In fact, with GIS we find all of the available stations for the origin and destination, and then compare paths' generalized time between each pair of OD stations to find the shortest path. This process is illustrated in Figure 1. Taking origin and destination as the center and walking distance as the radius, we found three and four stations available for origin and destination, respectively. Therefore, there are 12 paths for the OD trip, and then we can use the Logit model to calculate the probability of selection for each path, or just take the path with the least generalized time as the path actually used. We can also use GIS to assess whether there is an available station, or whether travellers use the railway or not. If there is no station in one of the circles, it is reasonable to suppose that no railway is available between the OD. If the destination and origin are located within the same circle, it is reasonable to suppose that for trips between the OD the railway will not be used. (Bus transportation is omitted.)

### 3 BUILDING GIS DATABASE

Trip origin and destination are the first data needed. To represent properly the true situation, we need to understand the origins and destinations in detail. We also need to analyze shopping trip information, such as a household's location, its membership structure, and collective purchasing ability. Information about shops would include their location, the size of the business, goods category, and price. It is difficult to collect such detailed data completely, but even some of it
is worthwhile. Nagoya City, in central Japan, will be taken as our study area. There are about 196 km of urban railway, with lines belonging to four companies, and the city's population is about 2.15 million. First, we input spatial and attribute data of almost 0.4 million residential buildings into the MapInfo table. In the table, spatial data are represented by polygons in the map window and attributes are stored in the browser window. This GIS layer will represent shopping trip origins, while trip destinations are assumed to be the relevant large stores, namely the shopping center (business space is over 1000 m²). Using the Nagoya urban commercial statistical map, we obtained another GIS layer representing shopping trip destinations (Economic Bureau of Nagoya City, 1999). These two layers are illustrated in Figure 2.

We can also add population and income data into each residential building and business scale, and goods category data into stores according to census and commercial statistics. However, since they are used to analyze shopping behavior rather than railway path selection behavior, they are beyond this study and omitted here. Another datum is the urban railway network, which consists of lines of four companies and each company has its own fare system. We divided the network into two GIS layers, namely node layer and link layer, which consist of point and polylines, respectively. In the layers other than spatial data, we also add names of line and station, travel time between stations, stopping time and diagram at station, and owning company into the attribute tables.

4 MODIFYING RAILWAY NETWORK FOR MICRO-SIMULATION

The spatial data of the network described above is just a digitized result of a map-based urban railway. When applying the shortest path-search algorithm to it, some difficult problems emerge. For example, there are as many stations as lines where the lines intersect. On the map, however, there is usually only one node to represent them. This creates difficulty for computer simulation of transfer time and cost. Also, there are sometimes several stations within walking distance of each other, but as there are no links between these stations, potential transfer is omitted from the simulation process. To deal with these problems, we must modify the original digitized network.

First, we divide the node, where several stations are integrated, into as many nodes as the number of lines, then connect each pair of stations with dummy links. These links represent transfer routes and have no length attribute but do have time and direction attributes. Time attribute represents transfer time and can be obtained through filed survey. Figure 3 is a part of the modified railway network, where an integrated station is divided into several stations. Secondly, we search for other companies' stations or those on other lines' stations for all stations within walking distance. Those stations are treated as the transferable stations. This study defines walking distance as 800 m. As shown in Figure 4, we found two stations within 800 m of West Jingu station. Travellers can change from subway 4 to Tokaido or Meitetsu lines, even though there is no link between them, and vice versa. Therefore, we create dummy links between them and

Fig. 2. Residence, shopping center, and railway.

Fig. 3. Separation of integrated station.

Fig. 4. Stations on other lines in walking distance.
Different methods.

Figure 2: Comparison of some shortest paths using

Time of one iteration (seconds)

[Graph showing comparison of different methods]

We found that about 78% million entries of OD pairs were found to support the shortest paths results. These showed that 105 million paths (number of agents) (C) are found. In our studies, there are about 0.4 million entries and 159 stations. Therefore, we suggested the use of OD pairs for the shortest paths.

(1) If $a = \min\{\frac{\theta_f}{\alpha_f}, \frac{\theta_t}{\alpha_t}\} + \frac{\theta_f}{\alpha_f} + \frac{\theta_t}{\alpha_t}$

The algorithm is divided into two stages: the projection and the decision stage. The projection stage determines the stations, and the decision stage selects the optimal path.

Figure 4: Part of network and traffic in our algorithm

The algorithm is divided into two stages: the projection and the decision stage. The projection stage determines the stations, and the decision stage selects the optimal path. The projection stage is divided into an inner (inside) and an outer (outside) part. The outer part determines the optimal path, while the inner part selects the optimal path.

Table 1

<table>
<thead>
<tr>
<th>Distance (km)</th>
<th>Time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>320</td>
<td>15</td>
</tr>
<tr>
<td>290</td>
<td>4</td>
</tr>
<tr>
<td>260</td>
<td>7</td>
</tr>
<tr>
<td>230</td>
<td>11</td>
</tr>
<tr>
<td>200</td>
<td>13</td>
</tr>
<tr>
<td>Zone</td>
<td></td>
</tr>
</tbody>
</table>

6. COMPARING RESULTS OF DIFFERENT ALGORITHMS

(1) $O = f (\text{Dense}) + (\text{Sparse}) f (\text{Sparse}) + (\text{Sparse}) f (\text{Sparse})$

The comparison of results is calculated with (1):

![Graph showing comparison of different algorithms]

2. RAILWAY SHORTEST PATH-SEARCH

The algorithm is divided into two stages: the projection and the decision stage. The projection stage determines the stations, and the decision stage selects the optimal path. The projection stage is divided into an inner (inside) and an outer (outside) part. The outer part determines the optimal path, while the inner part selects the optimal path.

Based on the above algorithm, we build up our path search.
that households in about 0.2 million residential buildings can access 133 shopping centers by rail with a MapBasic program, so about 28 million shortest railway paths were obtained by calculation. The vertical axis in Figure 7 shows the generalized time of the shortest OD path calculated using the Moore algorithm and access/egress of nearest stations, and the horizontal axis shows the same index obtained by our method. Among the 23,000 OD paths, 6,700 are different, and the rest are the same.

Tables 2, 3, and 4 are the results obtained from random selection. From Tables 2 and 3 we can see that the nearest stations to residences 43,386 and 43,884 and shops 53 and 62 are stations 65, 107, 36, and 27, respectively. However, Table 4 shows that from 43,386 to 62, stations 64 and 154 are used; from 43,884 to 62, stations 142 and 27 are used; from 43,884 to 53, stations 62 and 150 are used. If travellers use the nearest stations, generalized times from 43,386 to 62, from 43,884 to 62, and from 43,884 to 53 are 42.1, 46.8, and 37.5 min, respectively. All of these times are greater than the results obtained by our method. The average generalized time for the 28 million shortest paths obtained by our algorithm is 48.5 min, while the value obtained by the ordinary method is about 54.6 min.

7 CONCLUSIONS

We integrated the railway fare system and diagram at each station into the Moore algorithm and used GIS to link traffic origin and destination with network dynamically, so that the algorithm could simulate Mega City’s railway OD path-selecting behavior with the microanalysis method. The method considered walking time, waiting time, riding time, and transfer time and cost between different lines. Also, as residential and department store buildings are used as spatial analyzing units, this method reproduced the real railway’s OD path selection more clearly. Twenty-eight million calculated generalized times reflect travellers’ travel time and cost very accurately. Trip distribution simulated using this method will really illustrate the market share of each railway company. This method will be very useful for analyzing the impact of shortening times, fare adjustments, and train frequency, and for assessing the effect of railway construction on road transport.

Our method of calculation requires much more time than the conventional method. Using a Windows computer with 400 MHz CPU, it took us about a week to get the 28 million shortest paths already mentioned out of the 78 million potential paths. However, in view of the difference between the results obtained from the two methods, it is worthwhile to use our method rather than the conventional method. We believe that as computer hardware advances, the computing time will be very much shortened.

Basically we should assign railway OD traffic to the railway network, based on our path search method, and then compare the assigned traffic with the observed one to verify our method. However, as the result of omitting bus lines, link traffic fed by bus could not be obtained, so no comparison was made in this area. Because the bus network in Nagoya City is very complex and completely separate from the railway network, we could not integrate it into the railway network in this study. As Nagoya City will create a digital bus network by the end of this year, we plan to combine it with the railway network in our future study.

REFERENCES


An activity-based micro-simulation model for travel demand forecasting, Activity-Based Approaches to Travel Demand Forecasting, New York, 31-54.

M. F. P. (1997), An activity-based micro-simulation model for travel demand forecasting, Activity-Based Approaches to Travel Demand Forecasting, New York, 31-54.


McNally, M. G. (1997), An activity-based micro-simulation model for travel demand forecasting, Activity-Based Approaches to Travel Demand Forecasting, New York, 31-54.