Hindawi Publishing Corporation
Discrete Dynamics in Nature and Society
Volume 2016, Article ID 9527584, 9 pages
http://dx.doi.org/10.1155/2016/9527584

Research Article

Forecasting Method for Urban Rail Transit Ridership at Station Level Using Back Propagation Neural Network

Junfang Li,[1] Minfeng Yao,[2] and Qian Fu[3]

[1]*College of Transportation Engineering, Tongji University, Shanghai 201804, China*
[2]*College of Architecture, Huaqiao University, Xiamen 361021, China*
[3]*College of Transportation Engineering, Leeds University, Leeds LS2 9JT, UK*

Correspondence should be addressed to Minfeng Yao; flyboot@hqu.edu.cn

Received 11 May 2016; Revised 11 August 2016; Accepted 22 August 2016

Academic Editor: Allan C. Peterson

Direct forecasting method for Urban Rail Transit (URT) ridership at the station level is not able to reflect nonlinear relationship between ridership and its predictors. Also, population is inappropriately expressed in this method since it is not uniformly distributed by area. In this paper, a new variable, population per distance band, is considered and a back propagation neural network (BPNN) model which can reflect nonlinear relationship between ridership and its predictors is proposed to forecast ridership. Key predictors are obtained through partial correlation analysis. The performance of the proposed model is compared with three other benchmark models, which are linear model with population per distance band, BPNN model with total population, and linear model with total population, using four measures of effectiveness (MOEs), maximum relative error (MRE), smallest relative error (SRE), average relative error (ARE), and mean square root of relative error (MSRRE). Also, another model for contribution rate of population per distance band to ridership is formulated based on the BPNN model with nonpopulation variables fixed. Case studies with Japanese data show that BPNN model with population per distance band outperforms other three models and the contribution rate of population within special distance band to ridership calculated through the contribution rate model is 70%~92.9% close to actual statistical value. The result confirms the effectiveness of models proposed in this paper.

1. Introduction

In transportation planning, ridership modeling and forecasting is the basis for analyzing project viability and sustainability in the long run. Urban Rail Transit (URT) ridership at station level is an important element of URT ridership, which is critical for determining scale of stations and access facilities. Four-step model has been a traditional method for transit modeling [1], a forecasting method mainly on a region scale, integrating four interrelated submodels, trip generation, trip distribution, mode split, and traffic assignment together [2, 3] which makes it inappropriate to forecast detailed transit ridership, such as at the station level. Furthermore, the four-step model needs all basic information involving trips on a regional scale, which requires large amount of resources. Despite its proven effectiveness, the high complexity and costs prevent it from providing quick and timely response to dynamic land-use change within service area of stations.

Forecasting method for URT ridership at station level with multivariate regression models, also known as direct-forecast method, can forecast ridership based on the changes in factors affecting ridership throughout service area of stations [4–8]. This method considers factors affecting ridership such as built environment, social and economic attributes within service area, and ownership of stations as independent variables and average daily or peak hour ridership as dependent variable for regression analysis.

Considerable research has been conducted on finding factors significantly affecting ridership. Cervero and Kockelman analyzed relationship between travel demand and 3Ds (density, diversity, and design) [9]. "Density" refers to population and employment density within the service area of the stations, which is considered as the most important factor [4–8]. "Diversity" means land-use type and land-use mix within the service area of the stations [10–12]. "Design" indicates whether or not the design of streets and roads

is convenient for people to access transit services [12, 13]. Additionally, station ownership can also explain the ridership changes well. The type of station (intermediate, terminal, interchange, and intermodal) influences the ridership as well [7, 8, 11, 12]. The distance from the station to central business district (CBD) was also found to be a significant factor [8, 12]. Parking and riding facilities and feeder services [7, 8, 12] also affect the ridership.

The tendency of using transit declines as the distance from stations/lines increases [7, 8, 14, 15]. Zhao et al. (2003) drew an exponential curve to depict the effect of walking distance to transit by surveying people living in station service areas [14]. Gutiérrez et al. drew a line to depict the ratio of inhabitants living in service areas of stations to daily boarding per distance band and an exponential curve to depict the ratio of employment population within service area of stations to daily alighting per distance band [8]. The previously mentioned line or curve is called distance-decay function curve. The literature recommends that the population variable in forecasting model needs to be weighted by the distance-decay function according to its distance to the station. However, it is problematic when the function is used to weight population living or working in the service area of stations, because distance-decay weight in real world may not change continually as the distance-decay function shows in special distance bands. It stays the same in per distance bands.

Forecasting models such as the ordinary least squares (OLS) regression model [12], Poisson regression [16], and other multivariate linear regression [8] models are fed by variables on the characteristic of stations and their catchment areas. However, linear regression models are not appropriate because relationship between some predictors and URT ridership at station level does not show linear trends. Taking population within service area as an example, if the degree of land-use mix is high, inhabitants and employment are balanced in this area. And it is likely that no extra transit demand is derived. For this, back propagation neutral network (BPNN) model which can reflect high nonlinear relationship between URT ridership and its predictors is put forward. BPNN model is trained using case study data, with predictors as input and ridership as output, to get parameters (weights and bias), and then it is used to forecast the URT ridership at station level.

The major contributions of this paper are as follows:

(i) Take population or employment per distance band as predictor directly.

(ii) Identify key factors affecting ridership through partial correlation analysis.

(iii) Formulate BPNN model to reflect exact relationship between URT ridership and its predictors.

(iv) Formulate a model of contribution rate of population per distance band to URT ridership using numerical analysis.

2. Methodology

2.1. Key Factors Affecting Ridership

2.1.1. Factors Affecting Ridership. Referring to previously mentioned researches, population per distance band, road density, number of shuttle bus lines, land-use mix within station service areas, train frequency in one direction during peak hour, number of lines through station (station on the line), number of park and ride facilities, station type (terminal or not), and distance from station to CBD are selected preliminarily as factors affecting URT ridership at station level. This paper uses data of Tokyo, Japan, to illustrate how values of the above factors are calculated or obtained and the case study is conducted using the same dataset.

The population distance bands include 0-1 km, 1-2 km, 2-3 km, 3-4 km, 4-5 km, 5-6 km, 6 km, and above, according to road network buffer by GIS (Geographic Information System) [17]. Data from "Total Rail Season Tickets Classified by Departure Stations and Inhabitant Zones of Tokyo, Japan" at 2010 are used [18]. The data show that, within 6 km of service areas, access ridership during morning peak hour is above 97% of the total access ridership. So, 6 km and above is taken as a unit distance band. Population of each distance band is sum of capita living/working in Chomes (smallest unit of street block in Japan) within this distance band.

Road density (unit: km/km^2) is used to measure the convenience to stations. Larger road density favors access to stations by walking and increase transit use [7, 8, 12]. This variable is calculated by total length of roads gauged by Google Earth divided by area of the service area.

Number of shuttle bus lines is used to measure convenience of shuttle bus access, which can be obtained from Google Earth within service area of 200 m radius from station. Both shuttle bus and URT belong to mass transit and they are strongly dependent on each other.

Land-use mix can be expressed by land-use mix ratio. The higher the value, the more diverse the land-use throughout this area and the more likely the fact that inhabitant and employment population are balanced here. The formula is as follows [19]:

$$P_{\text{mix}} = \left|D_{\text{popu}} \times \log_{10}\left(D_{\text{popu}}\right)\right| + \sum_{k=1}^{n} \left|\left(D_{\text{em}}\right)_k \times \log_{10}\left(D_{\text{em}}\right)_k\right|, \quad (1)$$

where P_{mix} is land-use mix ratio of service area; D_{popu} is population density of service area (unit: per/hectare); D_{em} is density of the kth type employment (unit: per/hectare) of service area; n is total number of employment types in service area.

Unidirectional peak-hour train frequency is sum of all trains in one direction from any line stopping at a station. Same as number of lines through station, station type (terminal or not) and distance from station to CBD are used to measure station attraction. The station type is a binary variable: "1" indicates terminal and "0" otherwise.

The number of parking and riding facilities is also an important variable. More facilities are likely to attract more

drivers, making them choose transit instead. Data for the last 5 variables can be obtained from official websites of railway operation industries in Japan.

2.1.2. Partial Correlation Analysis. Identifying key factors is critical. If all the above variables are used to formulate forecasting model, the performance of model will be unsatisfying because of the correlation between these variables.

Partial correlation analysis [20, 21] is used to find key factors, avoiding the influence of other confounding factors when analyzing relationship between two variables. The formulas are listed as follows:

$$\begin{aligned} r_{xy} &= \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(y_i-\overline{y})^2}} \\ r_{xy,z_1} &= \frac{r_{xy}-r_{xz_1}r_{yz_1}}{\sqrt{\left(1-r_{xz_1}^2\right)\left(1-r_{yz_1}^2\right)}} \\ r_{xy,z_1z_2} &= \frac{r_{xy,z_1}-r_{xz_2,z_1}r_{yz_2,z_1}}{\sqrt{\left(1-r_{xz_2,z_1}^2\right)\left(1-r_{yz_2,z_1}^2\right)}} \\ r_{xy,z_1,\ldots,z_n} &= \frac{r_{xy,z_1,\ldots,z_{n-1}}-r_{xz_n,z_1,\ldots,z_{n-1}}r_{yz_n,z_1,\ldots,z_{n-1}}}{\sqrt{\left(1-r_{xz_n,z_1,\ldots,z_{n-1}}^2\right)\left(1-r_{yz_n,z_1,\ldots,z_{n-1}}^2\right)}}, \end{aligned} \tag{2}$$

where r_{xy} is simple correlation coefficient of variable x and y; r_{xy,z_1} is correlation coefficient of variables x and y when controlling z_1; r_{xy,z_1z_2} is correlation coefficient of variables x and y when controlling z_1 and z_2; $r_{xy,z_1,\ldots,z_n}$ is correlation coefficient of variables x and y when controlling $z_1,\ldots,z_n$; x_i, $\overline{x}$ is the ith value of variable x and the mean value of x; the meaning of y_i, $\overline{y}$ is the same as x_i, $\overline{x}$.

The partial correlation coefficient needs to be tested for significance further. The null hypothesis is that the partial correlation coefficient to be tested is not significantly different from zero. The corresponding t-statistic for the test is shown in

$$t = \frac{r\sqrt{n-1}}{1-r^2}, \tag{3}$$

where t is statistic value of test; r is partial correlation coefficient to be tested; n is sample size.

When degree of freedom is $n-1$, the probability value p corresponding to t is obtained to be compared with 0.01 and 0.05. If $p < 0.01$, association between the two variables is strongly significant. If $0.01 < p < 0.05$, association between the two variables is generally significant. If $p > 0.05$, association between the two variables is not significant. Partial correlation coefficient is used to identify key factors in the case study.

2.1.3. Data Source and key Predictors. Data from 129 stations in Tokyo, Japan, are used for case study to get the key predictors. Tokyo is the capital of Japan, encompassing 23 special wards, 26 cities, 5 towns, and 8 villages, which is also called Tokyo Metropolitan Area. The whole area is 2188.67 square kilometers, and the population is 13.23 million. URT is the main commuter traffic mode of Tokyo, whose total line is approximately 1000 km long and amount of total station is approximately 800 excluding the line servicing just suburban area. The selected station for case study is mainly in 23 special wards where the economic and trade activity mainly develops. Figure 1 is the distribution of urban transit rail line and selected station in case study in Tokyo Metropolitan Area.

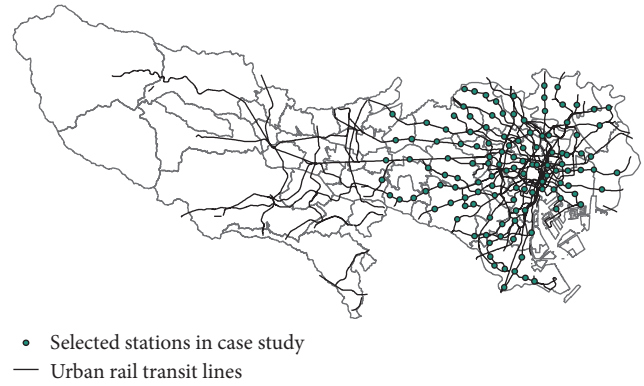

Figure 1: URT in Tokyo metropolitan area.

(1) Access Ridership (Annual Average Daily) in Morning Peak Hours. The data are obtained from "Total Rail Season Tickets Grouped by Departing Stations and Inhabitant Zones of Tokyo" at 2010 [18].

(2) Land-Use Mix within Service Area of the Station. The data is obtained from "The Number of Enterprises and Employment Grouped by Chomes and Industries" of Tokyo statistical information [22], where the 16 types of employment are agriculture-forestry-fisheries, mining, construction, manufacturing, electrical, gas heating and water supply Industry, information and communications, transport, wholesale and retail trade, finance and insurance, real estate, restaurant accommodation, medical welfare, education and learning supported industry, composite services, services not classified, and unclassified public services. Chomes are the smallest units for population, similar to blocks in America.

(3) Inhabitant Population within Service Area of the Station. The data is obtained from "Day and Night Population Grouped by Chomes" of Tokyo statistical information [23].

(4) Characteristic of Station. Variables like "unidirectional peak-hour train frequency," "number of lines through station," "number of parking and riding facilities, station type," and "distance from station to CBD" can be obtained from official websites of railway operation corporation in Japan [24–27].

Table 1 shows the correlation analysis of population per distance band and ridership using (2)–(3), when fixing all other variables. "Total" in Row 1 of Table 1 means population within all distance bands, which is taken as a single variable. Table 2 shows the correlation analysis of nonpopulation variables and ridership, when fixing all the other variables.

Table 1: Correlation analysis of population per distance band and ridership.

Correlation coefficient	0-1 km	1-2 km	2-3 km	3-4 km	4-5 km	5-6 km	Above 6 km	Total
Ridership	0.365**	0.216**	0.356**	0.211**	0.202*	0.061*	−0.186	−0.039

*: $0.01 < p < 0.05$, generally significant.
**: $p < 0.01$, strongly significant.

Table 2: Correlation analyses of nonpopulation variables and ridership.

Correlation coefficient	Road density	Number of shuttle bus lines	Land-use Mix	Peak-hour unidirectional train frequency	Number of lines through station	Number of park and ride facilities	Station type	Distance To CBD
Ridership	0.07*	0.18*	−0.164*	0.043*	0.101	−0.251	0.028*	0.04

*: $0.01 < p < 0.05$, generally significant.

Table 1 illustrates that correlation coefficient decreases as distance increases. Population in 0-1 km, 1-2 km, 2-3 km, and 3-4 km bands are strongly significantly correlated with ridership. Population in 4-5 km and 5-6 km bands are generally significantly correlated with ridership. Population above 6 km band and population within all bands are not correlated with ridership.

Table 2 shows that "road density," "number of shuttle bus lines," "land-use mix," "peak-hour unidirectional train frequency," and "station-type" are all generally significantly correlated with ridership. It needs to be noted that "land-use mix" is a negative predictor for ridership. That is because ridership in case study is the amount of annual average daily passengers entering station in the morning peak hour, which is mainly produced by inhabitants living in service area of the station. Imagine that there are two stations with same amount of inhabitants in their service area; if one station has no employment in its service area for inhabitants, inhabitants have no choice but find employment outside the region, so the URT railway may be used with high probability. Meanwhile, the area for the other station has enough employments for inhabitants, inhabitants would like to work near their home, and thus URT may be used with low probability. The result is ridership of station whose "land-use mix" is high is less than that of the station whose "land-use mix" is low. However, the case study only shows "land-use mix" is negative for ridership entering station in the morning peak hour, and how it influences the all-day ridership entering station needs further research.

Other than results in Table 2, correlation coefficient of "peak-hour unidirectional train frequency" and "number of lines through station" is 0.924 using (2)–(3), and they are strongly significantly correlated. Table 2 shows "number of lines through station" is not correlated with "ridership," while "peak-hour unidirectional train frequency" is generally significantly correlated with "ridership" even though the coefficient of the two is only 0.043. Since predictors should be independent of each other [28], the variable "number of lines through station" is abandoned. Correlation coefficient of "road density" and "distance to CBD" is 0.542 using (2)–(3), and they are strongly significantly correlated. Table 2 shows "distance to CBD" is not correlated with "ridership," while "road density" is generally significantly correlated with "ridership" and the coefficient of the two is 0.07. Thus, "distance to CBD" is abandoned.

Therefore, key predictors affecting ridership in this paper are obtained: population within distance bands (0-1 km, 1-2 km, 2-3 km, 3-4 km, 4-5 km, and 5-6 km), road density, number of shuttle bus lines, land-use mix, peak-hour unidirectional train frequency, and station type (terminal or not).

2.2. BPNN Model. BPNN model is selected as the forecasting model for URT ridership at station level [29–31]. BPNNs have hierarchical feed forward network architecture [32]. The processing procedures can be done with a minimum of three layers: one layer that receives and distributes the input pattern, middle or hidden layer that captures the nonlinearities of the input/output relationship, and one layer that produces the output pattern. BPNNs also may contain a bias node in the output and/or hidden layers that produce a constant output of 1 and is fully connected to the upper layer but receives no input. The procedures of the BPNNs are the error at output layer propagating backward to input layer through the hidden layer in the network to obtain the final expected outputs. BPNNs are trained by repeatedly presenting a series of input/output pattern sets to the network. The trained network is usually examined through a separate set of data called the test set to monitor its performance and validity. When the mean squared error (MSE) of the test set reaches a minimum, network training is considered to be complete and the weights are fixed.

Three-layer BPNN (input-hidden-output) is able to reflect any nonlinearity from input to output and thus is adopted in this paper. The number of key factors affecting ridership is 11, and it is also the number of nodes in input layer. In output layer, there is only 1 node: ridership. The number of nodes in hidden layer is $|\sqrt{11}| \approx 4$, which is square root of product of number of nodes in input layer and number of nodes in output layer.

The mathematical formulation of BPNN is as follows:

$$\text{Min} \quad E(w,b,v,c) = \frac{1}{2q}\sum_{p=1}^{q}\sum_{k=1}^{l}\left(O_k(p) - \widehat{O}_k(p)\right)^2, \quad (4)$$

subject to the following:

For output layer,

$$o_k = f\left(\sum_{j=1}^{m} v_{jk} y_j + c_k\right) \quad k = 1, 2, \ldots, l;\ j = 1, 2, \ldots, m. \tag{5}$$

For hidden layer,

$$y_j = f\left(\sum_{j=1}^{m} w_{ij} x_i + b_j\right) \quad j = 1, 2, \ldots, m;\ i = 1, 2, \ldots, n, \tag{6}$$

in which, transform function

$$f(x) = \frac{1}{1 + e^{-x}}, \tag{7}$$

where $E(w, b, v, c)$ is the objective function, minimizing the MSE of the actual and predicted ridership; w, v are matrix of weights variable; b, c are bias vector variable; $O_k(p), \widehat{O}_k(p)$ are the actual and predicted ridership of the kth neuron in output layer of the pth station sample. In our model, there is only one node in output layer and thus $l = 1$. v_{jk} is the weight of the jth node in hidden layer related to the kth node in output layer; w_{ij} is the weight of the ith node in input layer related to the jth node in hidden layer; y_j is the output of the jth node in hidden layer; x_i is the input of the ith node in output layer, that is, the value of key variables; c_k is the kth bias node in output layer; b_j is the jth bias node in hidden layer; m is the number of nodes in hidden layer, which is rounded off of $\sqrt{n}$. n is the number of nodes in input layer. q is the total number of station samples.

Steepest gradient descent method is adopted to update the weight and bias at each iteration. The detailed solution steps are described as follows:

Step 1 (initialization). Initialize weight and bias of BPNN. At the same time, for BPNN train, it is needed to set prediction accuracy and the maximum number of learning iterations.

Step 2. Select input (predictors)/output (actual ridership) of any station randomly. Use BPNN with initial weights and bias to forecast and obtain the predicted ridership. Then, MSE can be obtained by comparing the predicted ridership with actual ridership of the station. Calculate $\delta_o = (\widehat{o} - o)o(1 - o)$, where δ_o is the MSE derivation of output of output layer.

Step 3. Calculate $\delta_h(j) = \delta_o v_j y_j (1 - y_j)$, where $\delta_h(j)$ is MSE derivation of output of hidden layer.

Step 4. Adjust each weight v_{jk} using δ_o and y_j; that is, $\Delta v_j = \eta \delta_o y_j$. Adjust each weight w_{ij} using $\delta_h(j)$ and x_i; that is, $\Delta w_{ij} = \eta \delta_h(j) x_i$, where η is the learning rate of BPNN; Δv_j is the add value of v_j; Δw_{ij} is the add value of w_{ij}.

Step 5. Use BPNN and weights and bias obtained from Step 4 to forecast ridership of all stations using input (predictors) of these stations. Then, the objective function E can be obtained by comparing the predicted ridership with actual ridership of all stations and summing the differences up. If E satisfies the prediction accuracy requirement or the maximum number of learning iterations is satisfied, then the optimal value is obtained and the calculation procedures are terminated. Otherwise, select input (predictors)/output (actual ridership) of next station, go back to Step 2, and go to the next iteration.

2.3. Contribution Rate Model. Contribution rate model which predicts the contribution of population within specific distance band to ridership at station level is formulated in this section, by fixing the value of other variables. BPNN is able to reflect interrelationship between key predictors and ridership. To obtain contribution rate, population within other distance bands are set to zero. By changing population within specific distance band, we can observe corresponding changes of ridership. Thus, variables other than population (e.g., road density, number of shuttle bus lines, land-use mix, peak-hour unidirectional train frequency, and station type (terminal or not)) need to be known first. Figure 2 shows how to formulate BPNN model to obtain the contribution rate of population within 0-1 km band for ridership.

The detailed formulas are listed as follows:

$$y = g(x) \qquad h = \frac{dy}{dx}, \tag{8}$$

where h is contribution rate to ridership of population within specific distance band; x is population within specific distance band $x > 0$; y is ridership predicted by BPNN; $g()$ is the BPNN model.

The contribution rate model is solved by the following process. Plot dot (x,y) in the two-dimension coordinate which takes population within specific distance as x-axis and ridership as y-axis. Find a curve to comply with all the existing dots in the principle that variance is smallest. The curvature is the appropriate solution that is contribution rate to ridership of population within specific distance band.

3. Case Studies

3.1. BPNN Model. Data of previous 129 stations in Tokyo, Japan, are used as case study and implementing BPNN model to forecast ridership at station level. All variables are normalized due to dimension difference. This paper sets learning rate, prediction accuracy, and maximum number of learning iteration to be 0.8, 0.001, and 30000, respectively. BPNN is trained with data of 117 stations.

The optimal weights and bias are shown in Tables 3–5. Data of the other 12 stations are taken for model test. Weights and bias from the above training process are used to forecast the ridership of the 12 stations. Results are shown in Figure 3.

To verify forecasting accuracy of BPNN model in this paper, its results are compared with that of linear model with population per distance band, linear model with total population, and BPNN model with total population. The other three models are calibrated/trained using the same input/output pattern sets of 117 stations and are implemented

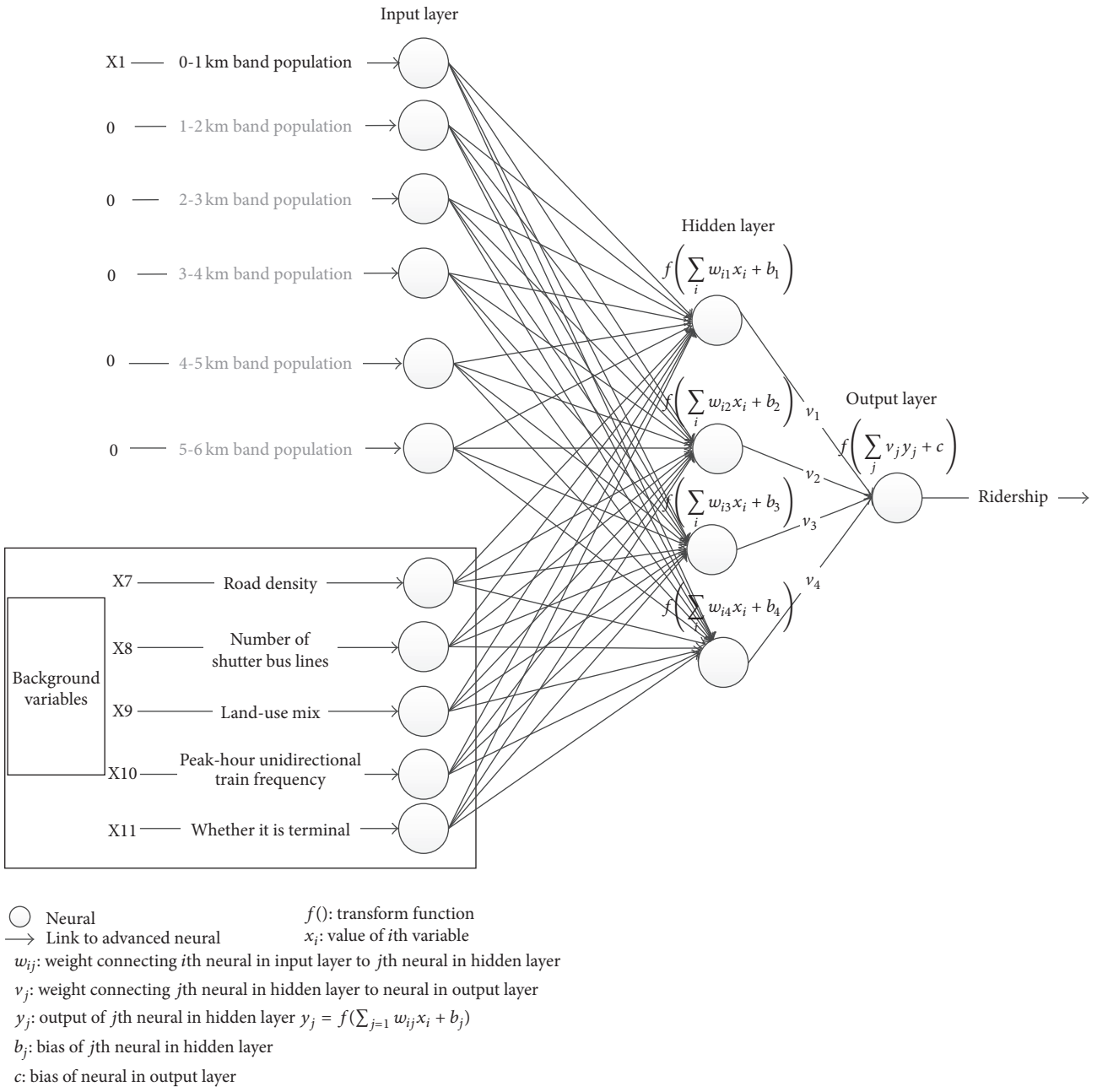

Figure 2: Contribution rate model for ridership of population within 0-1 km band.

for prediction with the same data of other 12 stations. Linear model is calibrated/trained by SPSS. The results are compared using multiple MOEs (unit: %), maximum relative error (MRE), smallest relative error (SRE), average relative error (ARE), and mean square root of relative error (MSRRE). Relative error is the difference between forecasting ridership of one model and actual ridership of each station divided by actual ridership. MRE, SRE, and ARE of a model indicate the maximum, the smallest, and mean values among relative errors of 12 stations, respectively. MSRRE of a model is the mean value of square root of relative errors of 12 stations. Figure 3 and Table 6 show results of four models graphically and numerically.

Figure 3 shows that, compared with other three models, result of BPNN model with population per distance band is the best and the difference of prediction of this model and the actual ridership is the smallest. In Table 6, except SRE, for other MOEs, BPNN model with population per distance band outperformed other three models. Values of each MOE of linear model with total population are nearly twice that of linear model with population per distance band. Meanwhile, values of each MOE of BPNN with total population are about

Table 3: Weights for nodes in input layer and hidden layer.

	Weights	Hidden layer			
		Node 1	Node 2	Node 3	Node 4
Input layer	Neural 1 (0-1 km band population)	$9.016E-06$	$-1.947E-05$	$1.598E-05$	$1.700E-05$
	Neural 2 (1-2 km band population)	$6.099E-06$	$6.369E-06$	$6.560E-06$	$-2.490E-06$
	Neural 3 (2-3 km band population)	$4.487E-06$	$-6.872E-06$	$4.702E-06$	$-9.129E-06$
	Neural 4 (3-4 km band population)	$-5.178E-06$	$-7.629E-06$	$-7.395E-06$	$5.073E-06$
	Neural 5 (4-5 km band population)	$6.990E-06$	$-4.752E-06$	$1.278E-05$	$-1.286E-05$
	Neural 6 (5-6 km band population)	$-2.955E-06$	$-4.138E-06$	$1.013E-05$	$9.622E-06$
	Neural 7 (road density)	0.0877	0.0819	−0.0822	0.0778
	Neural 8 (number of shuttle bus lines)	0.0996	−0.219	−0.132	0.0238
	Neural 9 (land use mix)	2.313	1.702	−1.369	1.608
	Neural 10 (peak-hour unidirectional train frequency)	0.0132	$-3.07E-04$	0.0069	−0.007
	Neural 11 (station type (terminal or not))	−0.294	1.1325	0.869	1.168

Table 4: Weights for nodes in output layer and hidden layer.

	Weights	Hidden layer			
		Node 1	Node 2	Node 3	Node 4
Output layer	Only one neural (ridership)	$9.016E-06$	$-1.947E-05$	$1.598E-05$	$1.700E-05$

Table 5: Bias for layers.

Hidden layer				Output layer
Neural 1	Neural 2	Neural 3	Neural 4	Only one neural
−7.1365	−1.9314	0.1474	−1.5717	0

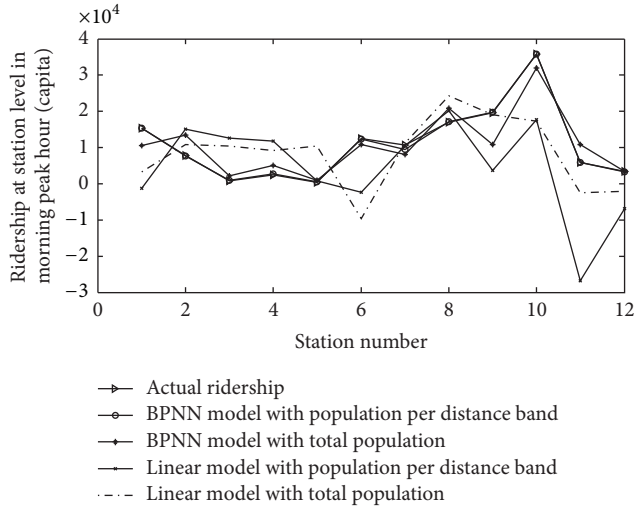

Figure 3: Comparison of four models.

10 times that of BPNN with population per distance band. The comparisons indicate that taking population per distance band as separate variable is necessary. The comparison of BPNN model and linear model both with population per distance band implies that BPNN model is able to better reflect nonlinearity relationship between ridership and its predictors than the linear model. After internal relationship has been obtained through training BPNN, ridership is predicted quickly and efficiently when changing value of any variable (e.g., when population in any distance band increases or decreases).

3.2. Contribution Rate Model. Mean values of every variable of 129 stations are used as background value. The station type is a binary variable, which makes us not able to take mean value of 129 stations as its background value. Here, it is set as 0. Under the above setting, increase population per distance band from 10000 (unit: persons) to 320000 by 10000 every time. Ridership is predicted by BPNN with weights and bias already obtained in Section 3.1 for per distance band every time. Taking 0-1 km distance band as an example, when the population is 10000, there is a ridership value predicted by previous BPNN model, and when the population is 20000, there is another ridership value predicted by previous BPNN model, and so on. We can get 129 ridership values corresponding to the changing population from 10000 to 320000 by 10000. We connect these ridership values to form a curve line time of population value. The slope ratio of this curve is contribution rate of population within 0-1 km distance band to ridership. And the procedures are repeated for contribution rate of population within 1-2 km, 2-3 km, 3-4 km, 4-5 km, and 5-6 km distance band to ridership.

The results are shown in Figure 4. Figure 4 shows the curve of contribution rate of population per distance band to ridership. By (8), we get contribution rate of population within 1-2 km, 2-3 km, 3-4 km, 4-5 km, and 5-6 km distance band to ridership. The results are compared with actual contribution rate of 129 stations by statistical analysis that total ridership from special distance band of 129 stations divided by total population within corresponding special distance band of 129 stations.

Table 6: MOEs of four models.

	MRE	SRE	ARE	MSRRE
Linear model with total population	2704	3.3	409	249
BPNN model with total population	3014	0.7	369	262
Linear model with population per distance band	1484	1.2	277	140
BPNN model with population per distance band	125	0.8	31	13

Table 7: Comparison of results from model and the real contribution rate.

		0-1 km	1-2 km	2-3 km	3-4 km	4-5 km	5-6 km
Contribution rate	Result from model	0.161	0.069	0.0255	0.0098	0.0067	0.0059
	Actual rate	0.19	0.07	0.033	0.012	0.009	0.0066
	Relative error (%)	18.9	7.1	23.6	20.8	30	16.7

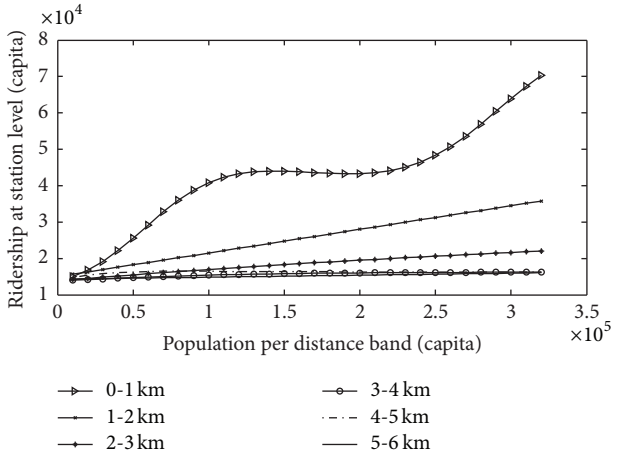

Figure 4: Curve of contribution rate.

Table 7 shows results from the BPNN model and the actual contribution rate are close. The relative error between the BPNN mode and actual data ranged from 7.1% to 30%. There are two possible explanations. The major one is that the value of population per distance band of one station may overlap with adjacent stations. For example, "14201 inhabitant zone" belongs to service area of both Kitami station and Komae station in Tokyo, Japan. It is 1.7 km from Kitami station and 1.8 km from Komae station. There are 14775 inhabitants in 14201-inhabitant zone. Potentially, some of the people may choose Kitami station while others may choose Komae station. However, in fact, population in 1-2 km band of the two stations are both set as 14775 in BPNN model, which leads to contribution rate of population within 1-2 km band to ridership decreasing. When calculating actual contribution rate of population within 1-2 km band to ridership, population within 1-2 km band of 129 stations is added up, and population in overlapping areas of two adjacent stations are just added once. Thus, actual rate is greater than the results from model. Another possible reason is that this paper sets the value of station type to 0 as the background value. In fact, some stations are not terminal. Population contribute less at nonterminal stations than at terminal station to ridership. So the rate from the model is smaller.

4. Conclusions

On the basis of previous researches, factors affecting URT ridership at station level are summarized and identified. Key factors are then obtained through partial correlation analysis, including population per distance band, road density, number of shuttle bus lines, land-use mix, peak-hour train frequency in one direction, and station type (terminal or not).

BPNN model is formulated to forecast ridership due to the nonlinear relationship between ridership and its predictors. Input (factors affecting ridership)/output (ridership) pattern sets of 117 stations in Tokyo, Japan, are adapted to train the model and data from other 12 stations are used to predict for test. The result obtained from BPNN with population per distance band is compared with that of BPNN with total population, linear model with population per distance band, and linear model with total population. The MOEs, for example, MRE, SRE, MRE, and MSRRE, are used to evaluate the model and results show BPNN model with population per distance band has the best performance. Since the model can reflect the internal relationship between ridership and its affecting factors, when one of factors varies, ridership can be quickly and efficiently predicted.

Based on BPNN model, contribution rate model of population per distance band to ridership is constructed, when setting other nonpopulation variables as background. Results of the case study show the effectiveness of the model. When population within special distance band changes, ridership from this population can be calculated quickly and timely by multiplying corresponding rate and population without performing BPNN model once more. Explanations of the relative errors are also presented.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The work is supported by the National Natural Science Foundation of China under Grant no. 51478198.

References

[1] M.-G. McNally, "The four step model," *Handbook of Transport Modelling*, no. 1, pp. 35–41, 2007.

[2] N. Marshall and B. Grady, "Sketch transit modeling based on 2000 census data," *Transportation Research Record*, vol. 1986, pp. 182–189, 2006.

[3] C. Yang and A. Chen, "Sensitivity analysis of the combined travel demand model with applications," *European Journal of Operational Research*, vol. 198, no. 3, pp. 909–921, 2009.

[4] R. Cervero, J. Murakami, and M. Miller, "Direct ridership model of bus rapid transit in Los Angeles County, California," *Transportation Research Record*, no. 2145, pp. 1–7, 2010.

[5] S. S. Pulugurtha and M. Agurla, "Assessment of models to estimate bus-stop level transit ridership using spatial modeling methods," *Journal of Public Transportation*, vol. 15, no. 1, pp. 33–52, 2012.

[6] N. Estupiñán and D. A. Rodríguez, "The relationship between urban form and station boardings for Bogotá's BRT," *Transportation Research Part A: Policy and Practice*, vol. 42, no. 2, pp. 296–306, 2008.

[7] J. Li, X. Ye, and J. Ma, "Forecasting method of urban rail transit ridership at station-level on the basis of back propagation neural network," in *Proceedings of the Transportation Research Board 94th Annual Meeting*, 2015.

[8] J. Gutiérrez, O.-D. Cardozo, and J. C. García-Palomares, "Transit ridership forecasting at station level: an approach based on distance-decay weighted regression," *Journal of Transport Geography*, vol. 19, no. 6, pp. 1081–1092, 2011.

[9] R. Cervero and K. Kockelman, "Travel demand and the 3Ds: density, diversity, and design," *Transportation Research Part D: Transport and Environment*, vol. 2, no. 3, pp. 199–219, 1997.

[10] K. Manaugh and T. Kreider, "What is mixed use? Presenting an interaction method for measuring land use mix," *Journal of Transport and Land Use*, vol. 6, no. 1, pp. 63–72, 2013.

[11] Y. Jiang, P.-C. Zegras, and S. Mehndiratta, "Walk the line: station context, corridor type and bus rapid transit walk access in Jinan, China," *Journal of Transport Geography*, vol. 20, no. 1, pp. 1–14, 2012.

[12] M. Kuby, A. Barranda, and C. Upchurch, "Factors influencing light-rail station boardings in the United States," *Transportation Research Part A: Policy and Practice*, vol. 38, no. 3, pp. 223–247, 2004.

[13] R. Cervero, *Transit Oriented Development in America: Contemporary Practices, Impacts, and Policy Directions*, National Center for Smart Growth Research and Education, 2004.

[14] F. Zhao, L.-F. Chow, M.-T. Li, I. Ubaka, and A. Gan, "Forecasting transit walk accessibility: regression model alternative to buffer method," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1835, pp. 34–41, 2003.

[15] O. D. Cardozo, J. C. García-Palomares, and J. Gutiérrez, "Application of geographically weighted regression to the direct forecasting of transit ridership at station-level," *Applied Geography*, vol. 34, no. 4, pp. 548–558, 2012.

[16] X. Chu, "Ridership models at the stop level," 2004.

[17] C. Upchurch, M. Kuby, M. Zoldak, and A. Barranda, "Using GIS to generate mutually exclusive service areas linking travel on and off a network," *Journal of Transport Geography*, vol. 12, no. 1, pp. 23–33, 2004.

[18] M. Japan, "Total rail season tickets classified by departure (arriving) stations and inhabitant (employment) zones of Tokyo," http://www.mlit.go.jp/en/.

[19] Y. Song, L. Merlin, and D. Rodriguez, "Comparing measures of urban land use mix," *Computers, Environment and Urban Systems*, vol. 42, pp. 1–13, 2013.

[20] J. Cohen, P. Cohen, S.-G. West, and L.-S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, Routledge, 2013.

[21] S.-B. Green and N.-J. Salkind, *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data*, Prentice Hall Press, 2010.

[22] S.-B. Tokyo, "The Number of Enterprises and Employment Classified by Chomes and Industries," http://www.toukei.metro.tokyo.jp.

[23] S.-B. Tokyo, "Day and Night Population Classified by Chomes," http://www.toukei.metro.tokyo.jp/.

[24] E.-J.-R.-C. Tokyo, "Timetable and Stations of Lines," http://www.jreast.co.jp/.

[25] K.-C. Tokyo, "Timetable and Stations of Lines," http://www.keio.co.jp/.

[26] T.-M. Tokyo, "Timetable and Stations of Lines," http://www.tokyometro.jp/index.html.

[27] T.-C. Tokyo, "Timetable and Stations of Lines," http://www.tokyu.co.jp/.

[28] T. Hochkirchen, "Modern multivariate statistical techniques: regression, classification, and manifold learning," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 173, no. 2, p. 467, 2010.

[29] Q. Zhuang and L. Chen, "Dynamic prediction of financial distress based on Kalman filtering," *Discrete Dynamics in Nature and Society*, vol. 2014, Article ID 370280, 10 pages, 2014.

[30] J. Ma and L. Liu, "Multivariate nonlinear analysis and prediction of Shanghai stock market," *Discrete Dynamics in Nature and Society*, vol. 2008, Article ID 526734, 9 pages, 2008.

[31] P. Sarvareddy, H. Al-Deek, J. Klodzinski, and G. Anagnostopoulos, "Evaluation of two modeling methods for generating heavy-truck trips at an intermodal facility by using vessel freight data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1906, pp. 113–120, 2005.

[32] J. L. McClelland and D. E. Rumelhart, "A simulation-based tutorial system for exploring parallel distributed processing," *Behavior Research Methods, Instruments, & Computers*, vol. 20, no. 2, pp. 263–275, 1988.