

# 基于圈层人口变量的城市轨道交通车站客流预测

李俊芳, 杨冠华, 邹江源, 柴东

(同济大学 道路与交通工程教育部重点实验室, 上海 201804)

**摘要:** 考虑城市轨道交通车站客流(指进出站客流)吸引范围内不同距离的人口对车站客流贡献率不同的情况, 将人口按照距离车站的远近分为不同圈层, 以不同圈层人口作为变量进行车站客流预测。通过偏相关分析验证圈层人口作为变量的合理性, 同时获得影响车站客流的其他显著因素。针对线性多元回归预测模型的不合理性, 建立了可反映车站客流与自变量高度非线性关系的 BP(back propagation)神经网络预测模型。案例研究表明: 基于圈层人口变量和 BP 神经网络的车站客流预测模型在减小误差方面明显优于其他模型, 且具有很好的实时性。在上述模型的基础上, 构建了已知任意车站背景变量, 车站圈层人口对客流的贡献率模型。该模型验证的结果进一步说明基于圈层人口变量和 BP 神经网络的车站客流预测模型能够很好地反映圈层人口与其他影响车站客流的显著影响因素同车站客流之间的关系。

**关键词:** 城市轨道交通; 车站客流预测; 圈层人口变量; BP 神经网络; 客流吸引范围

**中图分类号:** U293

**文献标志码:** A

## Forecasting Method of Urban Rail Transit Ridership at Station-level Based on Population Variable in Circle Group

LI Junfang, YANG Guanhua, ZOU Jiangyuan, CHAI Dong

(Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China)

**Abstract:** Considering different contribution rates to station riderships (in this essay, it means exit and enter ridership) of population within different distance to the station, it is necessary to classify population into different circle groups according to their distance to the station as a variable of the forecasting model. Population variable in circle group has been certified by partial correlation analysis and at the same time, other significant factors influencing station ridership have been obtained. Because of the irrationality of linear multi-variable regression, the back propagation neural

networks forecasting model has been built to reflect the high non-linear relation between the independent variable and dependent variable. The case study indicates that the forecasting model based on population variable in circle group and BP neural networks significantly precedes other models and meanwhile, it is real-time. Based on the above model, the contribution rate model of population in different circle groups to station ridership has been built, where any background variable of the station has been known. The result of the contribution model also indicates the forecasting model based on population variable in circle groups and BP neural network can better reflect the relationship between station ridership and all the factors influencing the riderships.

**Key words:** urban railway transit; station-level forecasting ridership; population variable in circle group; back propagation neural networks; service area

城市轨道交通车站客流预测是城市轨道交通客流预测的重要部分, 是确定车站设计规模和车站接驳交通设施规模的主要依据。城市轨道交通车站的客流规模与其吸引范围内的土地利用关系密切。由于四阶段法<sup>[1-3]</sup>和出行链法<sup>[4-6]</sup>需要获取全市范围内的出行 OD 基础资料, 工作量大且费时, 一般难以根据车站周边土地利用的动态变化情况提供实时车站客流预测。

车站客流预测的多元回归预测方法以车站客流吸引范围内影响车站客流的因素(环境、社会、经济等)作为自变量, 车站客流为因变量进行回归。文献[7-10]具体分析了车站客流吸引范围内影响车站客流的因素, 居住人口和岗位人口均被认为是重要的车站客流影响因素, 其中文献[8-9]将车站客流吸引范围内的居住人口和岗位人口及其他相关变量作为自变量, 提出了基于线性回归的全日进出站客流预测方法。目前在基于多元回归的车站客流预测方法

研究中主要存在两方面问题:一是预测模型中人口变量采用车站客流吸引范围内的居住总人口和岗位总人口,认为不同距离的人口对车站客流的贡献是一样的,这与现实中客流随到车站距离的增加而减少的规律不符.因此,有必要将距离车站不同圈层的人口作为自变量来构建车站客流预测模型;另一个问题是采用线性回归预测方法的不合理性,线性回归预测方法对许多因素考虑不清楚,比如车站客流吸引范围内人口岗位的平衡度.而非线性表达式的关系有很多种,很难确定.寻找一种能够分析自变量和因变量之间的关系且不需要具体表达式的模型是很有必要的.

## 1 影响车站客流关键因素分析

基于文献[7-10],本文初步选择车站客流吸引范围内的圈层人口、道路网密度、接驳公交线路数、土地利用混合度、车站单向发车数、车站换乘线路数、车站停车换乘设施数、是否终点站、车站到市中心的距离作为车站客流的影响因素.由于我国数据的缺乏,日本数据比较细致且可以获得,因此本文采用日本首都圈的相关数据来说明以上各项因素的获取、计算方法以及下文的案例实证.

车站客流吸引范围内的圈层人口是按照道路网络距离划定 $>0\sim 1\text{ km}$ , $>1\sim 2\text{ km}$ , $>2\sim 3\text{ km}$ , $>3\sim 4\text{ km}$ , $>4\sim 5\text{ km}$ , $>5\sim 6\text{ km}$ , $>6\text{ km}$ 共7个圈层.根据日本2010年大都市圈交通调查<sup>[11]</sup>,车站客流吸引范围6 km以内的早高峰进站客流占总客流的97%以上,因此将6 km以上作为1个圈层.各圈层人口是将各圈层覆盖丁目町区域内的所有人口相加.

道路网密度用来衡量到达车站的便利程度.通过谷歌地球测量客流吸引范围内的道路总长度,然后除以其客流吸引范围面积获得;接驳公交线路数用来衡量车站公交接驳的便利性.通过谷歌地球读出车站周边200 m范围内的接驳公交线路数.

土地利用混合度可采用土地利用混合率来表示<sup>[12-13]</sup>,值越高,代表土地利用多样性越好,人口和岗位在该区域平衡的可能性越大.公式如下:

$$P_{\text{mix}} = |D_{\text{popu}} \lg D_{\text{popu}}| + \sum_{k=1}^n |(D_{\text{em}})_k \cdot \lg(D_{\text{em}})_k| \quad (1)$$

式中: $P_{\text{mix}}$ 为车站客流吸引范围内的土地利用混合率; $D_{\text{popu}}$ 为车站客流吸引范围内的人口密度; $(D_{\text{em}})_k$

为车站客流吸引范围内第 $k$ 类就业岗位密度; $n$ 为车站客流吸引范围内总的就业岗位类别数.土地利用混合度数据来源于文献[11,14-15],其中岗位类别数包括农林渔业、矿业、建设业等共16类<sup>[14]</sup>.

车站单向发车数是车站高峰小时单向发车数,如某车站有多条轨道交通线路,则为该站各条线路高峰小时单向发车数之和,它同车站换乘线路数一样,用来衡量车站的吸引力;车站停车换乘设施位数用来衡量接驳交通和城市轨道交通换乘衔接的便利性,车站换乘方便,将会吸引客流.车站是否终点站和车站到市中心的距离同样用来衡量车站的吸引力.车站是否终点站用哑元变量来表示.如果是终点站,该变量为1;如果不是,则为0.以上变量数据从日本地铁各大官方网站获取.

这里需要说明的是,实际数据证明,车站客流吸引范围内各圈层道路网密度差别不大,各圈层人口通过道路网到达车站的通畅程度相似,各圈层人口因其到车站距离的不同对车站客流的贡献率不同才是关键.不同车站客流吸引范围内道路网密度差异较大,所以将车站整个客流吸引范围内道路网密度作为变量衡量其对车站客流的影响,不对其分圈层计算;土地混合利用程度用来衡量整个车站客流吸引范围内的人口岗位平衡程度,不能分圈层.接驳公交线路同道路网密度的解释类似,不对其分圈层计算.车站单向发车数、车站换乘线路数、车站停车换乘设施位数、车站是否终点站、车站到市中心的距离属于车站自身属性,不需要分圈层.

经验表明:如果将以上因素都作为自变量来构建车站客流预测模型,则可能会因为变量过多而导致模型的预测精度降低.因此,有必要进一步分析车站客流的关键影响因素.由于偏相关分析可以避免两个变量之间的相关性分析受到第三个变量的影响,所以本文采用偏相关分析方法来分析车站客流的关键影响因素以及各因素之间的相关程度,以剔除非独立因素变量.偏相关分析的计算公式如下:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

$$r_{xy, z_1} = \frac{r_{xy} - r_{xz_1} r_{yz_1}}{\sqrt{(1 - r_{xz_1}^2)(1 - r_{yz_1}^2)}} \quad (3)$$

$$r_{xy, z_1 \dots z_n} = \frac{r_{xy, z_1 \dots z_{n-1}} - r_{xz_n, z_1 \dots z_{n-1}} r_{yz_n, z_1 \dots z_{n-1}}}{\sqrt{(1 - r_{xz_n, z_1 \dots z_{n-1}}^2)(1 - r_{yz_n, z_1 \dots z_{n-1}}^2)}} \quad (4)$$

式中: $r_{xy}$ 为变量 $x, y$ 的简单相关系数,这里变量包括车站客流、各圈层人口、车站客流吸引范围内道路网密度、接驳公交线路数、土地利用混合度、车站单向发车数、车站换乘线路数、车站停车换乘设施数、是否终点站、车站到市中心的距离; $r_{xy, z_1}$ 为控制变量 $z_1, x, y$ 的相关系数; $r_{xy, z_1 \dots z_n}$ 为控制变量 $z_1, \dots, z_n, x, y$ 的相关系数; $x_i, \bar{x}$ 为变量 $x$ 的第 $i$ 个值和变量 $x$ 的均值; $y_i, \bar{y}$ 的意义同 $x_i, \bar{x}$ 。

求得的偏相关系数需要进行显著性检验。假设待求偏相关系数与零无显著差异。采用 $t$ 统计量进行检验。公式如下:

$$t = \frac{r \sqrt{n-1}}{1-r^2} \quad (5)$$

式中: $t$ 为检验统计量; $r$ 为待求显著性程度的偏相关系数; $n$ 为样本量。

求取自由度为 $n-1$ 时,该检验统计量对应的概率值 $p$ ,比较其与0.01和0.05的大小。若 $p < 0.01$ ,认为假设不成立,该相关系数的两个总体相关关系非常显著; $0.01 \leq p \leq 0.05$ ,认为假设不成立,该相关

系数的两个总体相关关系一般显著; $p > 0.05$ ,认为假设成立,该相关系数的两个总体不具相关性。

以日本首都圈33个车站的相关数据作为案例研究对象,车站客流是居住人口产生的年平均日早高峰的进站客流,客流数据来源于文献[11]。该调查详细给出了首都圈的所有车站早高峰进站客流来源于哪些丁目町及具体数量。居住人口数据来源于文献[15]。表1和表2分别列出了通过公式(2)~(5)控制其他所有变量进行偏相关分析后得到各圈层居住人口变量同车站客流的相关系数及显著程度,非人口变量同车站客流的相关系数及显著程度。

从表1可以看出,随着人口圈层到车站距离的增大,圈层人口与车站客流的相关系数大致呈现越来越小的趋势, $>0 \sim 1$  km,  $>1 \sim 2$  km,  $>2 \sim 3$  km,  $>3 \sim 4$  km圈层人口与车站客流显著相关, $>4 \sim 5$  km,  $>5 \sim 6$  km圈层人口与车站客流一般相关, $>6$  km圈层人口和未分圈层的总人口与车站客流不具相关性。通过偏相关分析验证了采用圈层人口变量的合理性。

表1 各圈层人口变量与车站客流相关性分析

Tab.1 Correlation analysis of population variable of circle group and station ridership

| 人口变量 | 圈层人口           |                |                |                |                |                | 未分圈层人口  |
|------|----------------|----------------|----------------|----------------|----------------|----------------|---------|
|      | $>0 \sim 1$ km | $>1 \sim 2$ km | $>2 \sim 3$ km | $>3 \sim 4$ km | $>4 \sim 5$ km | $>5 \sim 6$ km | $>6$ km |
| 相关系数 | 0.365**        | 0.216**        | 0.356**        | 0.211**        | 0.202*         | 0.061*         | -0.186  |

注:\*表示显著程度一般,\*\*表示显著程度非常高。

表2 非人口变量与车站客流相关性分析

Tab.2 Correlation analysis of non-population variable and station ridership

| 非人口变量 | 道路网密度 | 接驳公交线路数 | 土地利用混合程度 | 单向发车数  | 轨道交通换乘线路数 | 停车换乘设施数 | 是否终点站  | 距市中心距离 |
|-------|-------|---------|----------|--------|-----------|---------|--------|--------|
| 相关系数  | 0.07* | 0.18*   | -0.164*  | 0.043* | 0.101     | -0.251  | 0.028* | 0.04   |

注:\*表示显著程度一般。

通过公式(2)~(5)同时可以得到,单向发车数与轨道交通换乘线路数相关系数是0.924,非常显著,不符合变量独立性原则,结合表2,剔除轨道交通换乘线路数变量;车站客流吸引范围内道路网密度和车站到市中心的距离相关系数是0.542,非常显著,结合表2,剔除车站到市中心的距离变量;同时结合表1和表2的结果,选择 $>0 \sim 1$  km,  $>1 \sim 2$  km,  $>2 \sim 3$  km,  $>3 \sim 4$  km,  $>4 \sim 5$  km,  $>5 \sim 6$  km圈层人口及车站客流吸引范围内道路网密度、接驳公交线路数、土地利用混合程度、车站单向发车数、是否终点站等11个变量来构建预测模型。

## 2 城市轨道交通车站客流预测模型

线性预测模型不能合理反映影响城市轨道交通

车站客流的关键因素同车站客流之间的关系。例如线性预测模型认为车站客流是和其客流吸引范围内的岗位和居住人口成正比,即岗位和居住人口越多,车站客流越大。但是如果车站客流吸引范围内土地利用混合程度高,岗位和居住人口平衡,上述线性关系就不成立了。如果考虑非线性表达式,其表达式有很多种,非常难确定。

BP(back propagation)神经网络<sup>[16]</sup>作为人工神经网络模型的代表,能学习和存贮大量的输入-输出模式映射关系,而无需事前揭示描述这种映射关系的数学方程,具有突出的大规模并行处理、信息分布存储及自学习功能,已广泛运用于函数逼近、模式识别、分类及数据压缩问题。BP网络的拓扑结构包括输入层、隐含层和输出层,学习算法由信息正向传播

(输入层—隐含层—输出层)和误差反向传播(输出层—隐含层—输入层)两个过程组成. 隐含层根据信息变化能力需求可以设计为单层和多层. 周而复始的信息正向传播和误差反向传播过程,是各层权值和阈值不断调整的过程,也是神经网络学习训练的过程,此过程一直进行到可以接受的程度为止.

本文将基于 BP 神经网络来构建城市轨道交通车站客流预测模型.

## 2.1 BP 神经网络数学模型

理论证明,3 层 BP 神经网络可以实现任意从输入到输出的非线性映射;因此本文采用单层隐含层神经网络来构建模型. 若以上述得到的 11 个关键影响因素作为输入变量(自变量),以车站客流作为输出变量,则输入、输出神经元个数分别为 11 和 1. 根据经验,BP 神经网络的隐含层神经元数是输入输出神经元数的积的平方根,这里为  $\sqrt{11}$ ,取为 4. 构造 BP 神经网络预测车站客流的示意图如图 1 所示. 图 1 中,  $x_1$  到  $x_6$  分别表示  $>0\sim 1$  km,  $>1\sim 2$  km,  $>2\sim 3$  km,  $>3\sim 4$  km,  $>4\sim 5$  km,  $>5\sim 6$  km 圈层人口变量.  $x_7$  到  $x_{11}$  分别表示车站客流吸引范围内道路网密度、接驳公交线路数、土地利用混合度、车站单向发车数以及车站是否终点站变量.

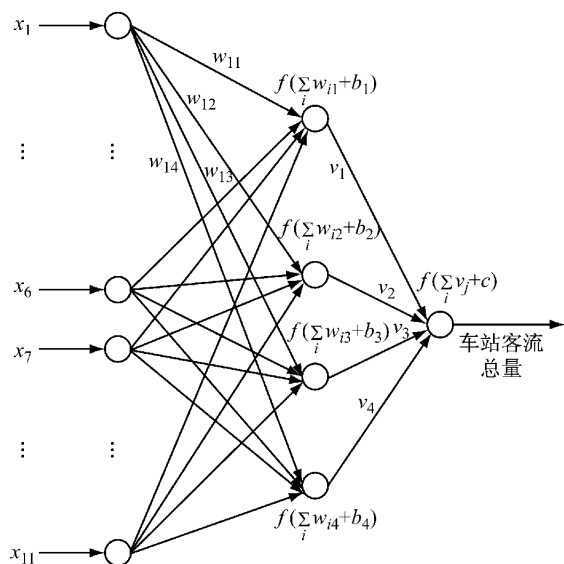


图 1 基于 BP 神经网络的城市轨道交通车站客流预测示意图

Fig. 1 Construction of BP neural networks model forecasting station ridership

从图 1 可以看出,这个预测模型中包含 48 个权重值,分别是输入层神经元对隐含层神经元的权重  $w_{ij}$ ,隐含层神经元对输出层神经元的权重  $v_{jk}$ . 模型中还包括 5 个阈值,分别是隐含层神经元阈值  $b_j$  和

输出层神经元阈值  $c$ . 其中  $i$  表示输入层第  $i$  个神经元( $i=1, \dots, 11$ ),  $j$  表示隐含层第  $j$  个神经元( $j=1, \dots, 4$ ).

以保证实际输出结果与期望输出结果的差值最小为目标,建立如下 BP 神经网络数学模型:

$$\min E(\mathbf{w}, \mathbf{b}, \mathbf{v}, \mathbf{c}) = \frac{1}{2q} \sum_{p=1}^q \sum_{k=1}^l (O_k(p) - \hat{O}_k(p))^2 \quad (6)$$

约束条件如下:

对于输出层

$$o_k = f\left(\sum_{j=1}^m v_{jk} y_j + c\right), \quad k = 1, 2, \dots, l, j = 1, 2, \dots, m \quad (7)$$

对于隐含层

$$y_j = f\left(\sum_{i=1}^n w_{ij} x_i + b_j\right), \quad j = 1, 2, \dots, m, i = 1, 2, \dots, n \quad (8)$$

式中:转换函数  $f(x) = 1/(1 + e^{-x})$ ;  $E(\mathbf{w}, \mathbf{b}, \mathbf{v}, \mathbf{c})$  为目标函数,以 BP 神经网络输出车站客流和实际车站客流差值平方和最小化为目标;  $\mathbf{w}, \mathbf{v}$  是待求权重值矩阵;  $\mathbf{b}, \mathbf{c}$  是待求阈值向量(这里需要说明的是,为了消除每个变量之间的量纲差异,代入模型的变量数据需要先量纲一化);  $O_k(p), \hat{O}_k(p)$  分别为第  $p$  个车站样本输出层第  $k$  个神经元的实际输出和期望输出;  $l$  为输出层神经元总数;本模型中,输出层仅有一个神经元,  $l=1$ ;  $v_{jk}$  为隐含层第  $j$  个神经元对输出层第  $k$  个神经元的权重;  $w_{ij}$  为输入层第  $i$  个神经元对隐含层第  $j$  个神经元的权重;  $y_j$  为隐含层第  $j$  个神经元输出;  $x_i$  为输入层第  $i$  个神经元的值;  $c_k$  为输出层第  $k$  个神经元的阈值,  $k=1$ ;  $b_j$  为隐含层第  $j$  个神经元的阈值;  $m$  为隐含层神经元个数,  $m=4$ ;  $n$  为输入层神经元数,  $n=11$ ;  $q$  为总的车站样本数.

## 2.2 BP 神经网络模型求解

BP 神经网络采取的权值更新方法为最速梯度下降法. 具体求解步骤如下:① 网络初始化. 给各连接权值分别赋一个区间  $(-1, 1)$  内的随机数,设定计算精度值和最大学习次数. ② 随机选择任意一个车站样本数据,用初始权值和阈值进行预测得到实际车站客流输出,同期望车站客流输出进行比较得到误差函数. 计算误差函数对输出层神经元的偏导数  $\delta_o = (o - \hat{o})o(1 - o)$ . ③ 利用隐含层神经元到输出层神经元的连接权值  $v_{jk}$ 、输出层神经元的偏导数  $\delta_o$  和隐含层神经元的输出  $y_j$  计算误差函数对隐含层各神经元的偏导数  $\delta_h(j) = \delta_o v_{jk} y_j (1 - y_j)$ . ④ 利用

输出层神经元的偏导数  $\delta_o$  和隐含层各神经元的输出来修正连接权值  $\Delta w_j = \eta \delta_o y_j$ 。利用隐含层各神经元的  $\delta_h(j)$  和输入层各神经元的输入  $x_i$  修正连接权  $\Delta w_{ij} = \eta \delta_h(j) x_i$ 。其中  $\eta$  是神经网络的学习率。⑤ 用步骤④获得的权值和阈值对所有样本数据进行预测,得到全局误差  $E$ 。当网络全局误差达到计算精度或者最大学习次数时,算法结束;否则,选取下一个学习样本及对应的期望输出,返回到步骤②,进入下一轮学习。

$$w = \begin{bmatrix} 9.0163 \times 10^{-6} & 6.0987 \times 10^{-6} & 4.4869 \times 10^{-6} & -5.1781 \times 10^{-6} & 6.9907 \times 10^{-6} & -2.9551 \times 10^{-6} & 0.0877 & 0.0996 & 2.3132 & 0.0132 & -0.2943 \\ -1.9465 \times 10^{-5} & 6.3686 \times 10^{-6} & -6.8716 \times 10^{-6} & -7.6285 \times 10^{-6} & -4.7524 \times 10^{-6} & -4.1400 \times 10^{-6} & 0.0819 & -0.2193 & 1.7022 & -0.0003 & 1.1325 \\ 1.5983 \times 10^{-5} & 6.5604 \times 10^{-6} & 4.7022 \times 10^{-6} & -7.3950 \times 10^{-6} & 1.2775 \times 10^{-5} & 1.0129 \times 10^{-5} & -0.0822 & -0.1318 & -1.3690 & 0.0069 & 0.8692 \\ 1.6976 \times 10^{-5} & -2.4846 \times 10^{-6} & -9.12185 \times 10^{-6} & 5.0729 \times 10^{-6} & -1.1286 \times 10^{-5} & 9.6219 \times 10^{-6} & 0.0778 & 0.0238 & 1.6076 & -0.0070 & 1.1679 \end{bmatrix}$$

$$v = \{-1.2524 \quad -0.0409 \quad -0.2176 \quad 0.5852\},$$

$$b = \{-7.1356 \quad -1.9314 \quad 0.1474 \quad -1.5717\}, \quad c = 0.$$

取其余 12 个车站的数据作为检验数据,图 2 显示了 12 个车站经模型预测的实际输出客流。为了验证基于圈层人口变量和 BP 神经网络的车站客流预测方法的预测精度,将其预测结果分别同采用总人口变量的线性预测模型、采用总人口变量的 BP 神经网络模型以及采用圈层人口变量的线性预测模型的预测结果进行比较。另外 3 种模型同样采用相同的 21 个车站的数据进行训练,12 个车站数据进行预测。其中线性预测模型基于 SPSS 软件进行回归。预测结果从最大相对误差、最小相对误差、平均相对误差、误差均方根几个方面进行比较。图 2 显示了 4 种模型的预测结果。表 3 列出了指标分析的具体结果。

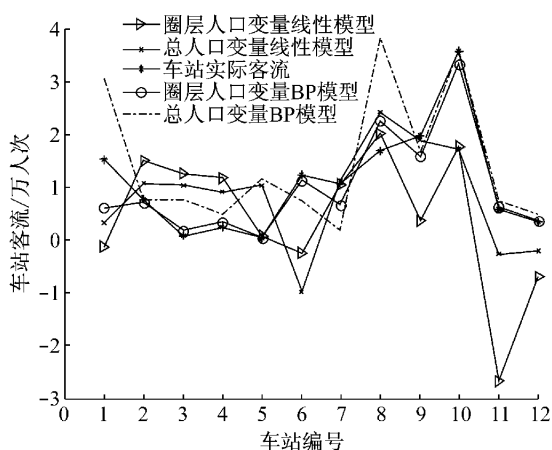


图2 4种模型预测结果比较图

Fig.2 Comparison of four models

从图 2 可以看出,基于圈层人口变量的 BP 神经网络模型预测的结果和实际客流最接近,另外 3 种模型的预测结果和实际客流相去甚远。具体分析表 3 中 4 种模型的误差指标,基于圈层人口变量的 BP 神

### 3 模型验证

根据上述算法,基于 MATLAB 7.7 编写程序,为避免量纲差异,各变量均进行量纲一化。取学习速率为  $\eta=0.8$ ,计算精度值 0.001,最大学习次数为 30 000 次。以 21 个车站的数据作为训练数据,当迭代次数达到 30 000 次,目标函数为 0.001 56。此时得到的权值和阈值如下:

表3 4种模型预测结果指标分析

| Tab.3 Index analysis of four models |        |        |        | %     |
|-------------------------------------|--------|--------|--------|-------|
| 模型                                  | 最大相对误差 | 最小相对误差 | 平均相对误差 | 误差均方根 |
| 总人口变量线性模型                           | 2 704  | 3.3    | 409    | 249   |
| 总人口变量 BP 模型                         | 3 014  | 0.7    | 369    | 262   |
| 圈层人口变量线性模型                          | 1 484  | 1.2    | 277    | 140   |
| 圈层人口变量 BP 模型                        | 125    | 0.8    | 31     | 13    |

经网络模型除了最小相对误差稍大于基于总人口变量 BP 模型外,其余指标均为最小。综合各项指标,说明基于圈层人口变量的 BP 神经网络模型用来预测车站客流具有稳定优势。基于总人口变量的线性模型的各项误差指标是基于圈层人口变量的线性模型的 2 倍,同时,基于总人口变量的 BP 模型除了最小相对误差外,其余误差项至少是基于圈层人口变量的 BP 模型的 10 倍以上,上述对比充分说明采用圈层人口作为变量的合理性。对比同是基于圈层人口变量的线性模型和 BP 神经网络模型,说明 BP 神经网络模型能更好反映圈层人口变量和其他变量同车站客流之间的内在关系,这种关系明确后,改变任意一个变量,如某个圈层的人口数发生变化,或者道路网络密度发生变化,或者增加就业岗位导致土地利用混合率发生变化等,BP 神经网络预测模型可以很快得到车站的客流结果,不需要对整个区域全方式交通量进行预测,因此基于圈层人口变量的 BP 神经网络车站客流预测模型具有很好的实时性。

鉴于 BP 神经网络模型能更好反映圈层人口变量和其他变量同车站客流之间的内在关系,下面探讨通过 BP 神经网络求解各个圈层人口对车站客流的具体贡献率。

#### 4 各圈层人口对车站客流贡献率模型

BP 神经网络能够很好反映圈层人口变量和其他变量同车站客流之间的内在联系,为了解具体某个圈层人口对车站客流的贡献率,可以置其他圈层人口变量为零,通过改变待求解圈层人口的值,观察车站客流的变化,从而求取贡献率。前提条件是道路网密度、接驳公交线路数、土地混合利用程度、单向发车数、车站是否终点站作为背景需要固定。即通过 BP 神经网络可以求解在背景变量已知的前提下任意圈层人口对车站客流的贡献率。图 3 是具体求解 0~1 km 圈层人口对车站客流贡献率的示意图,各变量的意义同图 1。

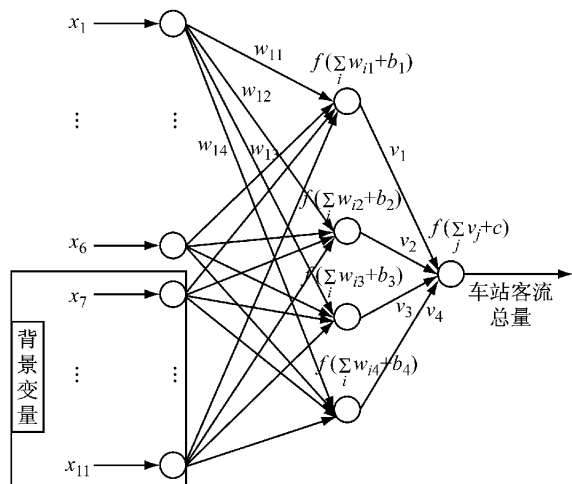


图 3 求解圈层人口对车站客流贡献率示意图

Fig.3 Construction of contribution model of population in circle groups to station ridership

建立求解具体圈层人口对车站客流贡献率的模型如下:

$$y = g(x), h = \frac{dy}{dx} \quad (9)$$

式中:  $h$  为具体圈层人口对车站客流的贡献率;  $x$  为具体圈层人口变量,  $x > 0$ ;  $y$  为通过 BP 网络计算的车站客流值;  $g(x)$  为 BP 神经网络模型。

模型求解采用数值分析法,改变  $x$  的值,观察  $y$  的变化,进而求取斜率近似解。

下面具体求解在背景变量位于 33 个车站数据均值,车站是否终点站变量取零的情况下,各圈层人口对车站客流的贡献率。采用数值分析法得到随各个圈层人口的变化,相应的车站客流变化曲线(图 4),其中圈层人口是从 1 万(符合实际圈层人口的量级)人次开始按增量 1 万人次递增至 32 万人次。从

图 4 可以看出,各圈层都是随着人口增加,产生的车站客流随之增加。表 4 列出了由公式(9)求得的各圈层人口对车站客流的贡献率和经过实际统计的 33 个车站的各个圈层人口对车站的贡献率平均值(33 个车站每个圈层的实际总进站客流除以 33 个车站每个圈层的实际总居住人口)。

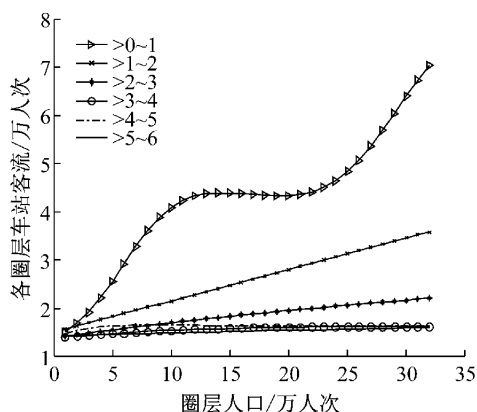


图 4 车站客流随各圈层人口变化曲线

Fig.4 Variation of station ridership by increasing population in some circle groups

表 4 模型求解和实际圈层人口对车站客流贡献率对比

Tab.4 Comparison of model values and value from reality

| 圈层人口    | 模型求解结果  | 实际值     | 相对误差/% |
|---------|---------|---------|--------|
| >0~1 km | 0.154 0 | 0.190 0 | 18.9   |
| >1~2 km | 0.065 0 | 0.070 0 | 7.1    |
| >2~3 km | 0.025 2 | 0.033 0 | 23.6   |
| >3~4 km | 0.009 5 | 0.012 0 | 20.8   |
| >4~5 km | 0.006 3 | 0.009 0 | 30.0   |
| >5~6 km | 0.005 5 | 0.006 6 | 16.7   |

从表 4 可以看出,模型求解圈层人口对车站客流贡献率值和实际值最大相对误差为 30%,最小相对误差为 7.1%。平均相对误差为 16.9%,在一定程度可以认为两者较为接近。分析产生误差的原因主要是代入 BP 神经网络预测模型进行拟合的 33 个车站数据,车站的圈层人口基数没有考虑车站与其相邻车站客流吸引范围重叠区域人口折减。例如,14201 居住区<sup>[11]</sup>是喜多見和狛江车站的客流吸引范围内重叠区域内的交通小区。其距离喜多見车站 1.7 km,距离狛江车站 1.8 km,人口基数为 14 775 人,理论上这些人中一部分属于喜多見车站的潜在客流人口,另一部分属于狛江车站的潜在客流人口。但是在预测模型拟合阶段,无论喜多見车站或狛江车站,代入模型的 >1~2 km 圈层人口均是按照 14 775 人进行计算,于是导致预测模型计算的 >1~2 km 圈层人口对车站的贡献率小。而贡献率模型是基于预

测模型分析的各圈层人口与车站客流的关系进行求解的. 实际 $>1\sim 2$  km 圈层人口对车站贡献率的计算是所有车站的 $>1\sim 2$  km 圈层到各车站的客流之和与该圈层的人口基数和(此时不考虑车站与相邻车站的客流吸引范围重叠区域)之比. 所以, 实际值相对于贡献率模型求解值偏大. 另外一个较为次要的原因是模型的背景变量中, 道路网密度、土地混合利用程度、接驳公交线路数、单向发车数采用的是33个车站的平均值, 车站是否终点站是0-1变量, 不能用33个车站的平均值, 取其为0, 即车站为非终点站. 但是实际统计的33个车站中, 有部分车站是终点站, 而终点站的圈层人口对车站客流贡献率要大于非终点站, 因此导致贡献率模型求解值整体小于实际值.

## 5 结论

本文在已有文献初步选择的城市轨道交通车站客流影响因素的基础上, 通过多元变量偏相关性分析, 验证了采用圈层人口变量的合理性并获得其他影响车站客流的显著影响因素. 针对现有城市轨道交通车站客流预测模型的缺陷, 构建基于圈层人口变量的BP神经网络车站客流预测模型. 采用日本首都圈21个车站的相关数据对模型进行拟合及验证, 说明基于圈层人口变量和BP神经网络的车站客流预测模型在减小误差方面具有稳定优势. 构建了已知除人口变量外任意车站背景变量下的车站圈层人口对客流的贡献率模型. 模型验证结果进一步说明了基于圈层人口变量和BP神经网络的车站客流预测模型能够很好地反映圈层人口与其他影响车站客流的显著影响因素同车站客流之间的关系.

## 参考文献:

- [1] McNally M G. The four step model[D]. Irvine: University of California, 2000.
- [2] 李明, 王海霞. 轨道交通车站客流预测模型研究[J]. 铁道工程学报, 2009(3): 67.  
LI Ming, WANG Haixia. Ridership prediction model research of transit railway station[J]. Rail Engineering Journal, 2009(3): 67.
- [3] 王文静. 福州市轨道交通客流预测研究[D]. 武汉: 武汉理工大学, 2009.  
WANG Wenjing. Ridership prediction research on transit railway of Fuzhou[D]. Wuhan: Wuhan Technology University, 2009.
- [4] 史晟, 杨超. 基于出行方式链的城市轨道交通客流分析方法[J]. 城市轨道交通研究, 2011, 14(4): 29.  
SHI Cheng, YANG Chao. Ridership analysis of urban rail transit based on trip-chain[J]. Research on Urban Rail Transit, 2011, 14(4): 29.
- [5] 董志国. 上海轨道交通出行方式链模型研究[J]. 城市轨道交通研究, 2012, 15(7): 15.  
DONG Zhiguo. Trip-chain model research of Shanghai railway transit[J]. Research on Urban Rail Transit, 2012, 15(7): 15.
- [6] 兰鹏. 基于出行目的链的轨道交通客流预测模型研究[D]. 北京: 北京交通大学, 2009.  
LAN Peng. Ridership prediction model research based on trip-chain[D]. Beijing: Beijing Jiaotong University, 2009.
- [7] Zhao F, Chow L F, Li M T, et al. Forecasting transit walk accessibility: A regression model alternative to the buffer method[J]. Transportation Research Record: Journal of the Transportation Research Board, 2003, 1835: 34.
- [8] Kuby M, Barranda A, Upchurch C. Factors influencing light-rail station boardings in the United States[J]. Transportation Research: Part A, 2004(38): 223.
- [9] Chu X. Ridership models at the stop level[M]. Tampa: University of South Florida, 2004.
- [10] Gutiérrez J, Cardozo D O, García-Palomares J C. Transit ridership forecasting at station level: An approach based on distance-decay weighted regression[J]. Journal of Transport Geography, 2011, 19(6): 1081.
- [11] 日本国土交通省. 日本三大都市圈第11次交通大调查报告[EB/OL]. (2012-03-20)[2013-10-08]. [http://www.mlit.go.jp/sogoseisaku/transport/sosei\\_transport\\_tk\\_000035.html](http://www.mlit.go.jp/sogoseisaku/transport/sosei_transport_tk_000035.html).
- [12] 林红, 李军. 出行空间分布与土地利用混合程度关系研究[J]. 规划研究, 2008, 32(9): 53.  
LIN Hong, LI Jun. Research of relationship between trip space distribution and mix of land use[J]. Program Research, 2008, 32(9): 53.
- [13] 钱林波. 城市土地利用混合程度与居民出行空间分布[J]. 城市研究, 2003(3): 7.  
QIAN Linbo. Trip space distribution of people and mix of land use[J]. City Research, 2003(3): 7.
- [14] 日本东京都统计局. 东京都按丁目町和产业分类的企业数和雇员数统计报告[EB/OL]. (2013-03-19)[2013-10-15]. <http://www.toukei.metro.tokyo.jp/jigyoku/2006/jg06v20100.html>.
- [15] 日本东京都统计局. 东京都按区市町村分类的昼夜人口统计报告[EB/OL]. (2013-08-14)[2013-10-20]. <http://www.toukei.metro.tokyo.jp/tyukanj/2010/tj-10index.html>.
- [16] 施彦, 韩力群, 廉小亲. 神经网络设计方法与实例分析[M]. 北京: 北京邮电大学出版社, 2009.  
SHI Yan, HAN Liqun, LIAN Xiaoqin. Design and case analysis of neural network[M]. Beijing: Beijing University of Posts and Telecommunications Press, 2009.