

文章编号: 2095-2163(2022)07-0080-05

中图分类号: U469.79

文献标志码: A

基于 ARIMA-BP 组合模型的城市公交客运量预测

刘 昶, 吴君华, 张凤娇, 于多友

(南京林业大学 汽车与交通工程学院, 南京 210037)

摘要: 随着社会经济的快速发展,城市居民日常公共交通出行供需矛盾日益突出,城市公共交通客运量需求预测成为其发展规划中的重要研究课题之一。针对于此,本文利用时间序列(ARIMA)模型与神经网络(BP)模型相结合的组合模型,以南京市地铁客运量为例,预测城市公共交通客运量的变化情况。结果证明,本文提出的 ARIMA-BP 组合模型预测精度优于单一模型。

关键词: 客运量预测; ARIMA 模型; BP 模型; 组合模型预测

Prediction of urban subway passenger volume based on ARIMA-BP combined model

LIU Chang, WU Junhua, ZHANG Fengjiao, YU Duoyou

(School of Automotive and Traffic Engineering, Nanjing Forestry University, Nanjing 210037, China)

[Abstract] At present, with the rapid development of social economy, the contradiction between supply and demand of daily public transport trips of urban residents has become increasingly prominent, which is a severe problem faced by the development planning of public transport in China's big cities. The prediction of passenger volume demand of urban public transport has become an important research topic in its development planning. Therefore, it is of great significance to accurately forecast the passenger volume of urban public transportation. Based on the combined model of ARIMA and BP, this paper takes Nanjing metro passenger volume as an example to predict the change of urban public transport passenger volume. The results show that the ARIMA-BP combined model used in this paper has better prediction accuracy than the two single models effectively.

[Key words] passenger volume forecast; ARIMA model; BP model; combined model prediction

0 引言

现如今,城市公共交通是城市基础设施的重要组成部分。在中国的经济发展和城市建设和社会生活中发挥着重要作用,直接关系到城市的经济发展和居民的生活,对城市经济发展有主导性的影响^[1]。但是,随着城市化进程的推进,机动车数量不断增加,出现了道路拥堵、堵塞等城市交通问题^[2]。优先发展城市公共交通,调整优化城市交通结构,对解决城市交通问题具有重要意义^[3]。由于公共交通系统的复杂性和可变性,客运量预测通常是不科学的,因此有必要更好地预测公交客流量,为城市规划提供更好的参考^[4]。

然而,城市公交客运量数据的收集存在大量未知、不确定因素和干扰。例如,人口密度、人口流动等能够真实反映客运量^[5]、特征和分布规律的历史

数据很难获得,也没有形成有效的历史数据^[6]。为此,许多专家学者对公交客流量预测方法进行了大量研究,并采用了不同的研究方法进行预测,每种预测方法都各有优缺点^[7-10]。因此,研究城市常规公交客运量预测方法,选择最合适的城市可持续发展方式,具有重要的理论和实践价值。

本文在 ARIMA 预测模型和 BP 预测模型的基础上,引入了基于方差倒数法的组合模型算法。以南京地铁客运量为例,首先利用训练样本数据对两种单一模型进行预测训练,随后基于两种单一模型的预测结果,运用方差倒数法,确定两种单一预测结果各自的权重,得到加权组合模型,最后进行预测试验,并给出了相关结果及分析。

1 单一模型预测

1.1 ARIMA 模型预测

ARIMA 模型是一种时间序列预测方法,其将非

基金项目: 南京林业大学高层次人才创新基金(G2014026)。

作者简介: 刘 昶(1998-),男,硕士研究生,主要研究方向:交通运输规划与管理;吴君华(1974-),男,博士,教授,主要研究方向:交通运输规划与管理。

通讯作者: 吴君华 Email: wj_h@163.com

收稿日期: 2022-01-13

平稳时间序列转化为平稳时间序列,然后将因变量的滞后值以及随机误差项的现值进行回归所建立的模型。

ARIMA (p, d, q) 中,AR 是“自回归”;I 为差分;MA 为“滑动平均”; p 为自回归项数; d 为使之成为平稳序列所做的差分次数(阶数); q 为滑动平均项数。 ACF 自相关系数可决定 p 的取值, $PACF$ 偏自相关系数能够决定 q 的取值。

自回归模型(AR)描述历史值与当前值的关系,用变量自身的历史时间数据进行预测,必须满足平稳性的要求,并具有自相关性,只在自相关系数小于 0.5 的情况适用。 p 阶自回归定义如式(1):

$$y_t = \mu + \sum_{i=1}^p \gamma_i y_{t-i} + \varepsilon_t \quad (1)$$

其中, y_t 是当前值; μ 是常数项; p 为阶数; y_i 是自相关系数; ε_t 是误差; y_{t-i} 为 y_t 前的值。

移动平均模型(MA)关注的是自回归模型中误差项的累加,能够有效地消除预测中的随机波动。 q 阶回归的定义如式(2):

$$y_t = \mu + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (2)$$

自相关函数(ACF)反映了同义序列在不同时期取值之间的相关性,如式(3):

$$ACF(k) = \rho_k = \frac{Cov(y_t, y_{t-k})}{Var(y_t)} \quad (3)$$

ARIMA (p, d, q) 模型的一般形式如式(4):

$$\begin{cases} \Phi(B) \tilde{N}^d x_t = \mu + \Theta(B) \varepsilon_t \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E x_s \varepsilon_t = 0, \forall s < t \end{cases} \quad (4)$$

其中, $\nabla^d = (1 - B)^d$ 为高阶差分; $\Phi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ 为平稳可逆模型的自回归多项式系数; $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ 为平稳可逆模型的移动平滑系数多项式。式(4)可简化为:

$$\tilde{N}^d x_t = \mu + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t \quad (5)$$

其中, ε_t 是白噪声序列, μ 是时间序列 x_t 的均值。

本文以 2006~2014 年南京客运量为训练数据,对 2015~2018 年的客运量进行预测,与真实值进行比较,并对模型预测结果进行比较分析。南京地铁客流量年度数据见表 1。

表 1 南京地铁年度客流量

Tab. 1 Annual passenger volume of Nanjing Metro

年份	客流量	单位
2006	5 789	万人
2007	8 016	万人
2008	10 379	万人
2009	11 353	万人
2010	21 459	万人
2011	34 370	万人
2012	40 060	万人
2013	45 215	万人
2014	50 300	万人
2015	71 700	万人
2016	83 067	万人
2017	97 711	万人
2018	111 250	万人

数据来源:南京统计发布

由训练数据进行预测训练,选用 ARIMA $(3, 0, 3)$ 模型,如图 1 所示。

ARIMA(3,0,3) Model:

Conditional Probability Distribution: Gaussian

Parameter	Value	Standard Error	t-Statistic
Constant	0	Fixed	Fixed
AR(1)	-0.308207	12.0575	0.0255614
AR(2)	0.308207	-1.99217	0.15471
AR(3)	-1	3.53855	-0.282602
MA(1)	1.19593	2.43314	0.454188
MA(2)	1.35975	1.50651	0.899924
MA(3)	0.555318	0.131396	4.22638
Variance	3.55975e+07	2.694e-06	1.32136e+13

图 1 ARIMA $(3, 0, 3)$ 模型

Fig. 1 ARIMA $(3, 0, 3)$ model

预测结果如图 2 所示。

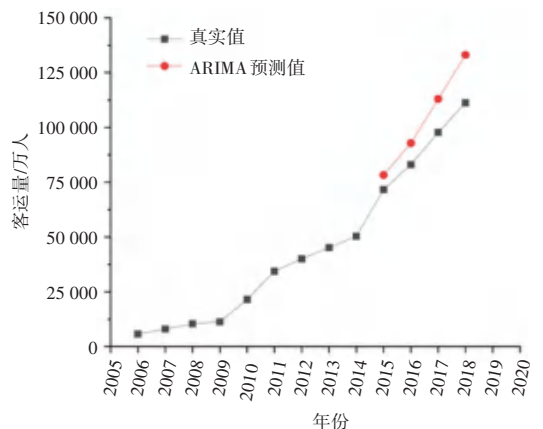


图 2 ARIMA 预测结果

Fig. 2 ARIMA model prediction results

1.2 BP 模型预测

BP (Back Propagation) 网络是一种按照误差逆向传播算法训练的多层前馈神经网络,是目前使用较为广泛的一种参数学习算法。

BP 神经网络分为两个过程:工作信号正向传递过程和误差信号反向传递子过程。在 BP 神经网络运算中,单个样本有 m 个输入、 n 个输出。其输出结果采用前向传播,误差采用反向传播方式进行,在输入层和输出层之间通常还有若干个隐含层。

Robert Hecht-Nielsen 在 20 世纪 90 年代就已证明:对于任何闭区间内的一个连续函数,都可以用一个隐含层的 BP 网络来逼近,这就是万能逼近定理。所以一个 3 层的 BP 神经网络就可以完成任意的 m 维到 n 维的映射,即这 3 层分别是输入层 (Input)、隐含层 (Hidden) 和输出层 (Output)。输入层接收数据,输出层输出数据,前一层神经元连接到下一层神经元,收集上一层神经元传递来的信息,经过“激活”把值传递给下一层。其工作模式如图 3 所示。

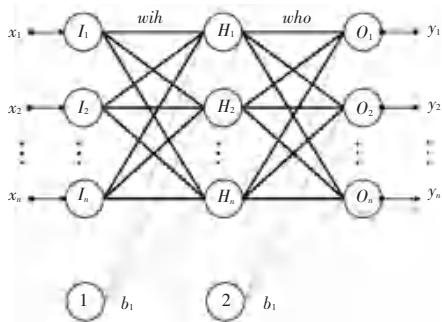


图 3 BP 神经网络

Fig. 3 BP neural network

图 3 中, I 是输入层; H 是隐含层; O 是输出层; x, y 分别表示输入和输出; w 为权重; b 为偏置; 每个圆圈表示一个神经元。

通常用来估量模型预测值与真实值不一致程度的损失函数有两种: Sigmoid 函数和 relu 函数。

由于 relu 函数是一次函数,并且随着自变量的增大而增大,同时方便在求梯度时求导,没有值的大小限制,且求导较为容易,常用在回归中使用。因此,本文选用 relu 函数作为激活函数。

同样以 2006~2014 年的客运量为训练数据,对 2015~2018 年的客运量进行预测,与真实值进行比较,对模型预测结果进行比较分析。

由训练数据进行预测训练,得到的预测结果如图 4 所示。

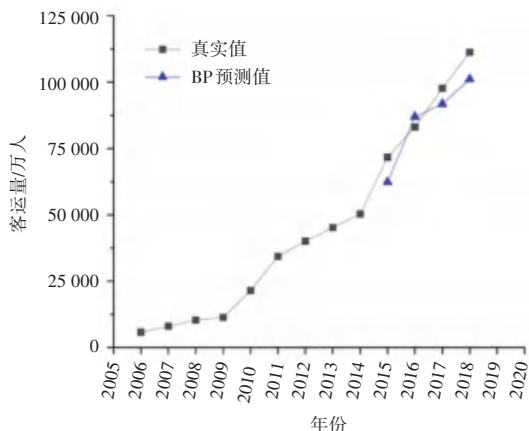


图 4 BP 预测结果

Fig. 4 BP neural network prediction results

2 组合模型预测

本章将在两个单模型预测所得结果的基础上,根据两种结果与真实值的误差,设计一种组合模型,对客运量进行预测。

2.1 组合模型构建

本文提出的基于 ARIMA 模型和 BP 模型的组合模型的设计思路如图 5 所示。

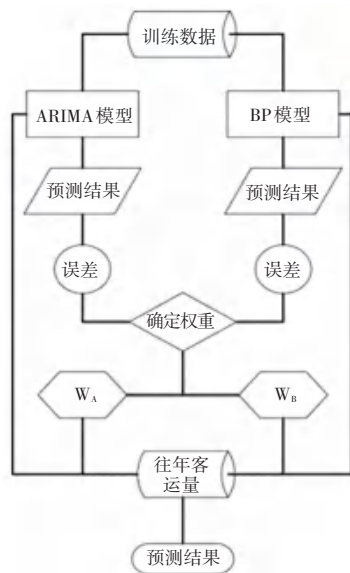


图 5 组合模型

Fig. 5 Combined model

求解两种单个模型各自权重的方法如式(6):

$$\begin{aligned} \begin{cases} D_{Ai} = |R_{Ai} - S_i| \\ D_{Bi} = |R_{Bi} - S_i| \\ \vdots \\ w_A = \frac{\overline{D_{Bi}^2}}{D_{Ai}^2 + \overline{D_{Bi}^2}}, i = 2015, 2016, 2017, 2018 \\ \vdots \\ w_B = \frac{\overline{D_{Ai}^2}}{D_{Bi}^2 + \overline{D_{Ai}^2}} \end{cases} \quad (6) \end{aligned}$$

其中, S_i 为第 i 年的客运量真实值; R_{Ai} 为 ARIMA 模型第 i 年的预测值; D_{Ai} 为第 i 年 ARIMA 模型的预测误差; R_{Bi} 为 ARIMA 模型第 i 年的预测值; D_{Bi} 为第 i 年 ARIMA 模型的预测误差; w_A 为 ARIMA 模型的权重; w_B 为 BP 模型的权重。

本文利用误差取平方的方式, 将两种单一模型的预测误差程度放大, 再将几年的误差平方取均值。以该平均值作为确定权重的依据, 根据公式求得两种模型各自在组合模型中的权重, 从而使组合模型的预测结果更加准确。

2.2 组合模型预测结果

使用上述组合模型, 同样以 2006~2014 的客运量为训练数据, 对 2015~2018 年的客运量进行预测, 与真实值进行比较, 并对模型预测结果进行比较分析。

由训练数据进行预测训练, 得到的预测结果如图 6 所示。

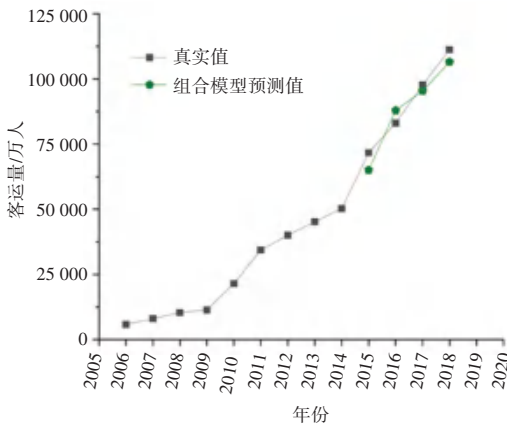


图 6 组合模型预测结果

Fig. 6 Combined model prediction results

3 预测结果对比分析

本节对 ARIMA 模型、BP 模型以及组合模型的预测结果进行对比分析, 其结果如图 7 所示。

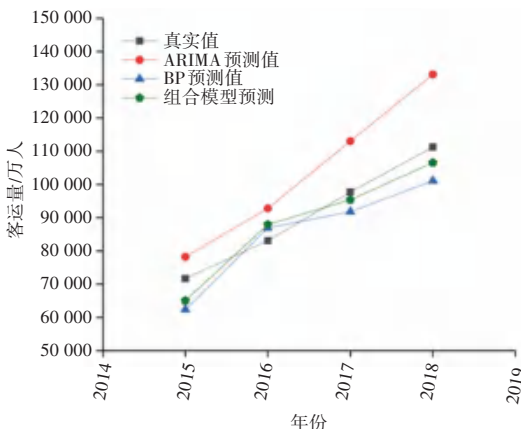


图 7 预测结果对比

Fig. 7 Comparison of prediction results

本文以年均误差率 \bar{T} 作为预测准确度的判断标准, 其计算方法如式(7):

$$\bar{T} = \frac{\sum_{i=2015}^{2018} (R_i - S_i)^2}{\sum_{i=2015}^{2018} S_i^2}, \quad i = 2015, 2016, 2017, 2018 \quad (7)$$

得到 3 种模型预测结果的年均误差率见表 2。

表 2 预测结果年均误差率

Tab. 2 Annual average error rate of forecast results

年份	实际客运量/万人	ARIMA 预测值/万人	BP 预测值/万人	组合模型预测值/万人
2015	71 700	78 253	62 354	65 057
2016	83 069	92 814	86 934	87 934
2017	97 711	113 005	91 785	95 392
2018	111 250	133 105	101 095	106 537
年均误差		0.14	0.08	0.05

由表 2 中结果可以看出, 组合模型预测结果的年均误差率明显低于两种单一模型。

4 结束语

本文基于 ARIMA 模型和 BP 模型, 设计了一种组合模型预测城市地铁客运量的变化趋势, 利用误差取平方的方式, 将两种单一模型的预测误差程度放大, 再将几年的误差平方取均值, 以该平均值作为确定权重的依据, 以进一步使误差相对较大的模型获得一个相对较小的权重, 而误差相对较小的模型获得一个相对较大的权重。本文以年均误差率为标准, 通过实验结果的对比分析表明: 组合模型相对于两种单一模型具有更高的预测准确度。

参考文献

- [1] 张鹏, 唐琴琴. 基于小波神经网络对铁路客运量的预测研究[J]. 山西大同大学学报(自然科学版), 2021, 37(3): 29-32, 113.
- [2] 曾华燕, 郑琼芳, 蔡铭洲. 利用简化时间序列法预测公共交通客运量[J]. 城市公共交通, 2021(3): 37-40.
- [3] CHENG Maolin, SHI Guojun, HAN Yun. An Extended Grey Model GM(1, 1, exp(bk)) and Its Application in Chinese Civil Air Passenger Volume Prediction[J]. Journal of Systems Science and Information, 2019, 7(5): 486-496.
- [4] XU Yuping, WEI Xu, CHEN Siwei. Research on Railway Passenger Volume Prediction Based on LSTM Neural Network[J]. IOP Conference Series: Materials Science and Engineering, 2019, 688(4): 004017.
- [5] 吴宇, 孙宏宇, 孙明辰, 等. PDTR 模型对城市流动人口的预测[J]. 智能计算机与应用, 2021, 11(9): 156-160.
- [6] DENG Wu, LI Wen, YANG Xinhua. A novel hybrid optimization algorithm of computational intelligence techniques for highway passenger volume prediction[J]. Expert Systems With Applications, 2010, 38(4): 4198-4205.