

基于 Prophet 算法的铁路客流量预测研究

王 晓, 揣锦华, 张立恒

(长安大学 信息工程学院, 陕西 西安 710064)

摘 要:客流量预测是铁路部门了解日常客运流量和冷热门线路具体情况的基础,是制定运输方案和列车开行计划的重要依据。针对铁路客流量变化受到季节周期、突发事件和节假日等因素影响的现象,提出一种基于 Prophet 时间序列算法的铁路客流量预测研究的新方法。实验选取了某铁路客运专线 2015 年至 2016 年的日客流量数据,经过预处理的客流量数据进行标准化得到客流量时间序列;构建节假日特征时间窗口列表和设置趋势的突变点,结合 Prophet 算法对节假日特征和突变点进行计算,并依此建立 Prophet 预测模型;利用可视化技术分析预测结果,调节参数进一步优化模型。重点结合节假日效应对未来两周客流量进行了预测分析。结果表明,Prophet 模型预测结果精确度高于 LSTM 模型,所建立的模型预测结果是合理和可靠的。

关键词:铁路客运专线;客流量;时间序列预测;Prophet 算法;节假日效应

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2020)06-0130-05

doi:10.3969/j.issn.1673-629X.2020.06.025

Research on Railway Passenger Flow Forecast Based on Prophet Time Series Algorithm

WANG Xiao, CHUAI Jin-hua, ZHANG Li-heng

(School of Information Engineering, Chang'an University, Xi'an 710064, China)

Abstract: The passenger flow forecast is the basis for railway departments to understand the daily passenger flow and the specific situation of cold and hot lines, and it is also an important basis for formulating transport plans and train running plans. A new method for railway passenger flow prediction based on Prophet time series algorithm is proposed to solve the problem that railway passenger flow is affected by seasonal cycles, emergencies and holidays. In the experiment, daily passenger flow data of a railway passenger dedicated line from 2015 to 2016 were selected. We construct the holiday characteristic time window list and set up the sudden change point of trend, calculate the holiday characteristic and sudden change point with Prophet algorithm, and build the Prophet prediction model based on it. Visualization technology is used to analyze the prediction results and adjust parameters to further optimize the model. The forecast and analysis of the passenger flow in the next two weeks are made based on the holiday effect. The results show that the Prophet model is more accurate than the LSTM model, and the predicted results are reasonable and reliable.

Key words: dedicated passenger railway line; traffic; time series prediction; Prophet algorithm; holiday effect

0 引言

近年来,随着人民生活水平的提高,人们对铁路客运服务的要求也逐渐提高。铁路部门为了提高运输服务能力,需要了解日常的铁路客运流量、淡旺季客流量波动指数和冷门线路情况等。其中,对客流量的充分了解和预测是提高铁路客运服务能力的首要前提,因此,对铁路客流量预测分析是目前一个重点的研究方向。

关于铁路客流量预测的分析,许多学者对此做过深入研究。常国珍等人认为铁路客流量变化具有复杂

性、随机性和周期性的特征^[1]。成强提出基于小波包分解与长短时记忆融合的深度神经网络用于铁路客流量的预测^[2]。Li L 等人提出了一种基于符号回归和 ARIMA 的混合模型预测地铁客流量^[3]。王伟炜提出了根据客流构成,将客流量分为趋势客流量、转移客流量和诱增客流量三部分,分别选用 BP 神经网络、灰色模型和重力模型建立高速铁路影响下的铁路客流量预测模型^[4]。路小娟等人利用 GA-BP 网络预测模型对铁路客流量进行了预测^[5]。许俊等人提出了通过若干组连续历史断面客流数据训练动态连续的

收稿日期:2019-08-05

修回日期:2019-12-05

基金项目:国家重点研发计划(2018YFB1600200);陕西省自然科学基金基础研究计划—重大基础研究项目(2017ZDJC-23)

作者简介:王 晓(1990-),男,硕士,研究方向为大数据处理的软件体系结构;揣锦华,教授,研究方向为大数据分析与管理。

神经网络预测未来客流^[6]。马彪采用灰色 GM(1,1) 预测模型对国内春运客流量进行了预测^[7]。Guo 等人提出了一种基于支持向量回归 (SVR) 和长短期记忆 (LSTM) 神经网络融合模型对城际轨道列车客流量进行了预测研究^[8]。Li B 等人针对影响客流的特殊因素建立组合模型进行了预测分析^[9]。Dou 等人提出了基于模糊时序逻辑的客流量预测模型对高速铁路短期客流量进行了较为精确的预测^[10]。江天河构建了 ARIMA 模型和 BP 神经网络模型对春运客流量进行了预测^[11]。冯冰玉等人提出了构建线性时间序列灰色 GM(1,1) 和非线性遗传算法优化 BP 神经网络模型用于铁路客流量的预测^[12]。黄召杰等人通过建立灰色预测模型对铁路客流量进行了预测^[13]。Xie 等人提出了基于神经网络和出发地矩阵估计的分而治之的方法预测高速铁路短期客流^[14]。Feng 等人建立了一个随机系数模型用于预测城市轨道交通车站的短期客流量^[15]。Ni 等人提出了一种基于参数和凸优化的方法并结合线性回归和自回归整合移动平均 (SARIMA) 模型对地铁客流量进行了预测^[16]。

为了分析季节性因素和节假日效应等对铁路客流量变化的影响,文中将基于 Facebook 的 Prophet 时间序列预测算法对某铁路客运专线过去 15 个月的日客流量进行拟合,对未来两周的客流量变化做出预测。

1 Prophet 时间序列预测算法原理及方法

1.1 时间序列预测算法

时间序列,也称时间数列或者动态数列,它是将某种现象的某一个统计指标在不同时间上的各个数值按时间的先后顺序排列而成的序列。时间序列预测算法是通过编制和分析时间序列,根据时间序列所反映出来的发展过程、方向和趋势,进行延伸,以此预测下一时间点或区间内可能达到的水平。时间序列预测算法实质上是一种回归预测方法,属于定量预测,其基本原理是承认事物发展的延续性,运用过去的时间序列数据统计分析,推测出事物的发展趋势,同时充分考虑到由于偶然因素影响而产生的随机性,为了消除这种随机波动产生的影响,利用历史数据进行统计分析,并对数据做预处理,对未来趋势的变化进行预测。目前,比较成熟的时间序列预测算法有移动平均法 (MA)、自回归模型 (AR)、自回归滑动平均模型 (ARMA)、指数平滑法、自回归求和滑动平均模型 (ARIMA) 和长短期记忆神经网络 (LSTM) 等。文中将基于 Prophet 时间序列预测算法对某铁路客运专线未来两周客流量进行预测分析。

1.2 Prophet 预测算法原理及方法

Prophet 算法是 Facebook 开源的时序预测框架,最

初是为了创建高质量的商业预测而研发,是一种新的基于时间序列的预测算法。Prophet 算法的原理是对各种时间序列特征分组件进行分析,如按时间序列的季节周期性、趋势性、节假日效应等组件分析。Prophet 算法通过拟合这几项,将它们累加起来得到时间序列的预测值。因此,Prophet 实质上是一个基于加法回归的模型。Prophet 模型的构成如下:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (1)$$

其中, $g(t)$ 是趋势变化函数,用来分析时间序列中非周期性的变化; $s(t)$ 是季节周期项,一般以年或以周为单位; $h(t)$ 属于节假日项,表示节假日因素造成的影响变化; ε_t 是误差项,表示突发事件的影响,通常建模为高斯噪声。

1.2.1 Prophet 算法的趋势项

(1) 基于逻辑回归的非线性饱和增长趋势 (Logistic)。

$$g(t) = \frac{C(t)}{1 + e^{-k(t)(t-m(t))}} \quad (2)$$

其中, $C(t)$ 是承载量, $k(t)$ 是增长率, $m(t)$ 是偏移量。在 Prophet 中 $C(t)$ 、 $k(t)$ 和 $m(t)$ 是随着时间 t 变化的函数。

(2) 分段线性趋势 (Linear)。

$$g(t) = (k(t) + a(t)^T \delta)t + (m(t) + a(t)^T \gamma) \quad (3)$$

其中, $k(t)$ 是随着时间 t 变化的增长率, δ 是增长率的变化量, $m(t)$ 是偏移量, $a(t)$ 表示在 t 时刻之前突变点发生变化的次数, $\gamma = -s \cdot \delta$, s 与突变点的选取有关。

1.2.2 季节周期性影响

时间序列通常会按每天、星期、月、年等季节性的影响而呈现周期性的变化,Prophet 中使用傅里叶级数模拟时间序列的周期性。其形式如下:

$$s(t) = \sum_{n=1}^N \left(a_n \cos \frac{2n\pi t}{P} + b_n \sin \frac{2n\pi t}{P} \right) \quad (4)$$

其中, P 是可以设置为想要的时间序列周期长度,如设 $P=365$ 可对年度数据进行的周期效应估计,同样 $P=7$ 是对以周为单位的周期效应估计。

设 Prophet 的季节项为:

$$s(t) = \beta X(t) \quad (5)$$

则有

$$\beta = (a_1, b_1, \dots, a_N, b_N)^T \quad (6)$$

$$X(t) = \left[\cos\left(\frac{2\pi(1)t}{P}\right), \sin\left(\frac{2\pi(1)t}{P}\right), \dots, \right.$$

$$\left. \cos\left(\frac{2\pi(N)t}{P}\right), \sin\left(\frac{2\pi(N)t}{P}\right) \right] \quad (7)$$

此时,参数向量 β 初始化为 $\beta \sim \text{Normal}(0, \sigma^2)$ 。

1.2.3 节假日效应

节假日因素对很多时间序列产生重要影响,这些影响效应通常是可以预测的。但是,因其周期性不固定,很难用一个模型去确定所有地区的节假日。为了克服不确定因素,Prophet 模型提供了一个灵活的拟合节假日的模型,即模型可以对节假日进行自定义,满足特定场景的预测需求。模型原理如下:

$$h(t) = \sum_{i=1}^L (k_i \cdot 1) \quad (t \in D_i) \quad (8)$$

设

$$Z(t) = [1(t \in D_1), 1(t \in D_2), \dots, 1(t \in D_L)] \quad (9)$$

$$h(t) = Z(t)k \quad (10)$$

其中, D_i 表示节假日窗口序列, $k \sim \text{Normal}(0, v^2)$ 。

2 Prophet 模型的构建与应用

2.1 原始数据及其预处理

2.1.1 原始数据

文中所采用的数据来自于铁路公司某客运专线过去连续 15 个月的日客流数据,这些数据中包含日期、车次、站点、客座率、上车和下车人数等信息。在对原始数据分析的过程中,发现数据中存在格式的不一致性、空缺值和重复记录等噪声数据,这些噪声数据将会不利于模型的训练,需要对原始数据进行预处理。

2.1.2 原始数据预处理

(1) 数据格式的转换。

原始数据的格式是半结构化形式,数据表中的信息没有按照字段对应。通过对原始数据的属性做相关性分析,找出与客流量相关的属性进行分类别建立子表。例如,分别按站点统计信息,建立客流量数据表或者按车次建立客流量数据表等。在保证原始数据完整性和准确性的同时,对原始数据格式进行转换,以满足算法模型对数据的要求。

(2) 数据空缺值的处理。

原始数据中存在着大量空缺值和重复值。例如,2015 年 7 月 26 日和 10 月 31 日等缺失了一整天的客流数据信息等。文中采用同类属性均值替代法,即用同类属性的其他已知对象的日客流数据计算并取均值来填补缺失的日客流数据。求均值的公式为:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (11)$$

经过对原始数据进行相应的预处理,得到了标准的、干净的和连续的实验数据。

2.2 基于 Prophet 模型的铁路客流量预测

以 2015 年 1 月至 2016 年 3 月的日客流量数据为基础构建模型,对铁路客运专线的某站点未来两周的

客流量进行预测。经过分析发现,铁路客流量是随时间发展变化的随机过程,而站点每天客流量也是不确定的,且受地理位置、季节周期、突发事件、节假日等多种复杂因素影响,而 Prophet 在构建模型时可以充分考虑到这些因素对客流量趋势变化的影响,确保对未来铁路客流量做出更精确的预测。

2.2.1 Prophet 预测算法的流程

原始数据经过格式转换、对空缺值和异常值等进行处理得到标准化后的客流量数据,将其代入构建的 Prophet 模型进行训练,按非周期性变化的趋势项、季节周期项和节假日效应对模型进行分解,进而对模型进行分组件训练,最后根据训练结果进一步优化模型,以提升模型的预测精确度。Prophet 预测算法的流程如图 1 所示。

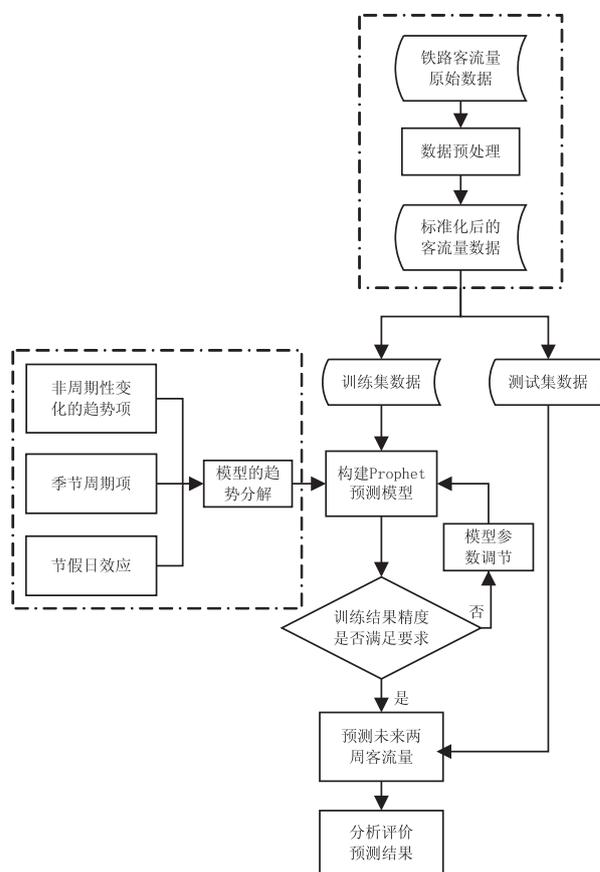


图 1 Prophet 预测算法的流程

2.2.2 Prophet 模型的参数说明

Prophet 本质上是一种可分解的加法回归模型,即按组件可将模型分解为非周期性变化的趋势项、周或年的季节周期项和节假日效应等。因此,Prophet 模型可以灵活地对各组件的参数进行设置,这些参数值的大小代表着各组件对模型预测结果的贡献度。通过分析模型的训练结果,可以调节模型各组件的参数,以便进一步提升预测精确度。文中所使用的 Prophet 模型的部分参数说明如表 1 所示。

表1 模型的参数

参数名称	类型	说明
growth	linear/logistic	模型趋势的增长方式
changepoints	double	指定潜在突变点
yearly_prior_scale	double	表示模型的年季节性灵活度
weekly_prior_scale	double	表示模型的周季节性灵活度
daily_prior_scale	double	表示模型的日季节性灵活度
holidays	Date	可设定节假日

2.2.3 节假日效应对预测的影响

在实际研究中,节假日或者一些重大的社会事件都会对时间序列造成很大影响,而这些因素往往不存在周期性。因此,对其单独建模分析是非常有必要的,如何有效地拟合历史数据中隐含的节假日效应是需要特别考虑的。

ARIMA 算法是目前时间序列预测算法中比较常用的,它可以很好地把握历史数据的季节周期性和未来趋势变化,但无法结合实际对节假日效应的影响进行有效分析,且模型预测的精确度也有待提高。Prophet 算法相比其他时间序列算法,模型可对节假日效应的影响进行有效分析。假设节假日的影响是独立的,通过设置特定的节假日窗口集,在特定的节假日窗口前后一段天内,将围绕窗口前后的每一天都视为节假日本身,例如“-1”表示节假日窗口下限往前延伸一天,同理,“1”表示节假日窗口上限往后延伸一天。表2为对节假日窗口日期的设置。

表2 节假日窗口日期

节假日名称	节假日窗口	节假日窗口下限(天)	节假日窗口上限(天)
元旦	2015/01/01---2015/01/03	-1	1
春节	2015/02/18---2015/02/24	-5	3
清明节	2015/04/04---2015/04/06	-1	1
劳动节	2015/05/01---2015/05/03	-1	1
端午节	2015/06/20---2015/06/22	-1	1
中秋节	2015/09/26---2015/09/28	-2	1
国庆节	2015/10/01---2015/10/07	-2	1

3 实验结果分析

3.1 Prophet 算法的实验结果分析

应用 Prophet 模型对过去一年的铁路客流量数据进行训练,对预测的总体变化进行组件化分解,得到了非周期性变化的趋势项、节假日效应、年或周的季节周期项等各组件对应的分析结果,最终得到铁路某站点未来两周的客流量预测结果。

(1) Prophet 模型各组件的预测结果及分析。

客流量的非周期性变化趋势如图2所示。

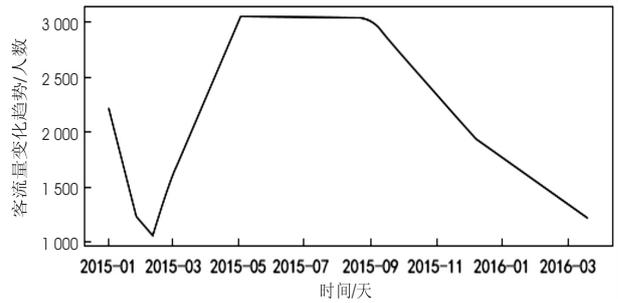


图2 客流量的非周期性变化趋势

从图2可以看出,客流变化趋势是比较明显的,并且没有明显的周期性。在2015年2月到5月有明显的上升趋势,5月至9月趋势变化相对稳定,而在9月以后出现明显的下降趋势。

节假日效应对客流量变化的影响如图3所示。

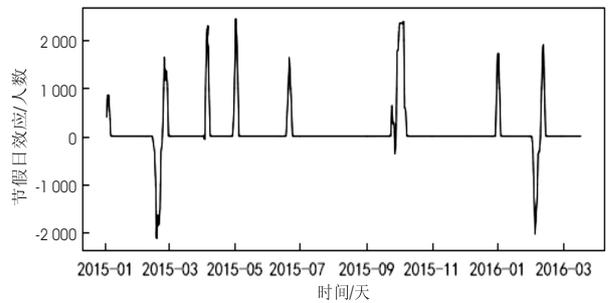


图3 节假日效应对客流量变化的影响

从图3很清晰地看到,在法定节假日因子影响下,客流量有着显著变化。例如,在2015年的2月18日至2月24日的春节期间,节假日因子对客流量变化有较强的削弱作用,分析其原因是因为春节期间流动人口较少,外出乘坐火车人数相比平时会少一些,而在“十一”这样的节假日窗口,由于大多数人会选择外出游玩而乘坐火车。因此,节假日因子对客流量变化具有明显的促进作用。

以年为单位的客流量季节性变化如图4所示。

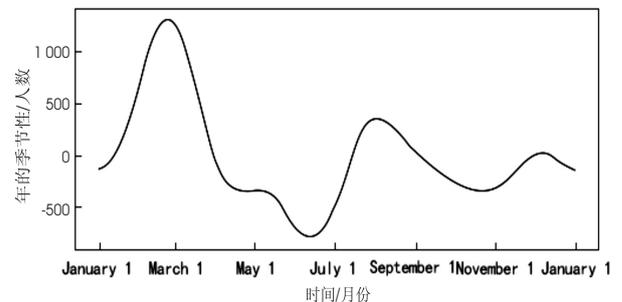


图4 以年为单位的客流量季节性变化

从图4可以看出,在1月至3月客流量变化有一个明显上升趋势,因为这一时间段有国家传统节日“春节”,节日窗口前后的客流量波动明显,如从3月以后出现急剧下降的趋势。6月到8月又急剧上升,之后趋势缓慢减少趋于稳定。模型为了更好地拟合数据,通过对已知的突变点检测,使得季节性因素的影响

作用显著增强。

(2) 客流量预测结果分析。

图 5 为铁路客运专线某站点客流量的预测结果。

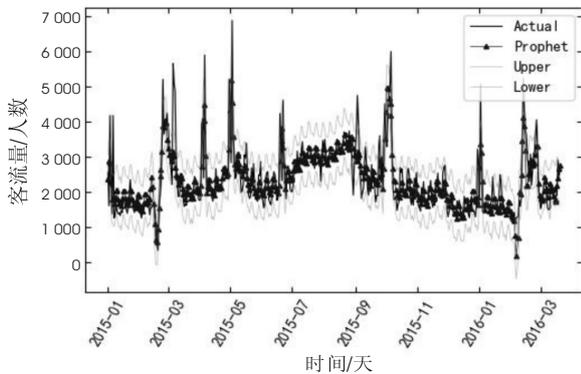


图 5 某站点的客流量预测结果

在图 5 中，“Actual”表示客流量真实数据，“Predict”表示预测值，“Upper”和“Lower”分别代表预测结果的置信区间上限和下限，灰色部分表示未来两周的客流量的预测部分。从图 5 的预测结果可以看出，模型对历史数据有较好的拟合，年的季节性效应和节假日效应有着明显的调节作用，预测结果的置信区间也包含了大部分的真实值，模型预测的精确度良好。

3.2 Prophet 和 LSTM 算法的预测结果对比分析及评价

为了验证 Prophet 算法的预测效果，将同样的数据代入 LSTM 算法得到相应的预测结果，并对这两种算法的预测结果进行对比分析和评价。

(1) Prophet 和 LSTM 算法的站点未来两周客流量预测结果对比分析。

Prophet 和 LSTM 算法的未来两周客流量预测结果对比如图 6 所示。文中所采用的 LSTM 算法选择 1 个客运日作为预测步长，即根据当天的铁路客流量数据对下一天客流量数据做出预测。从图 6 可以清楚地看到两种预测算法对未来两周的铁路站点客流量变化的预测结果对比，Prophet 相比于 LSTM 算法的预测结果更接近于真实值。

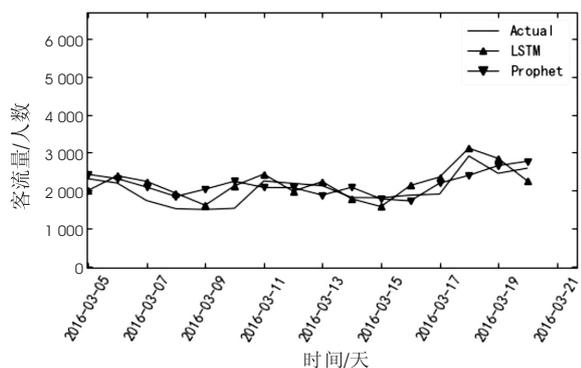


图 6 Prophet 和 LSTM 算法对未来两周铁路客流量预测结果对比

(2) Prophet 和 LSTM 算法的预测结果评价。

预测算法的评价指标主要有均方误差 (MSE)、均方根误差 (RMSE)、平均绝对误差 (MAE)、平均绝对百分比误差 (MAPE) 和对称平均绝对百分比误差 (SMAPE) 等。文中选取 RMSE、MAE 和 MAPE 作为评价指标分别对 Prophet 和 LSTM 算法的铁路客流量预测结果进行分析比较，如表 3 所示。从表 3 可以得到，Prophet 算法比 LSTM 算法的精确度略高一点。

表 3 Prophet 和 LSTM 算法的评价指标对比结果

预测评价指标	Prophet 算法	LSTM 算法
RMSE	324.587	321.131
MAE	271.441	283.633
MAPE	14.5114	15.7268

4 结束语

Prophet 算法是一种组件化的预测模型，即模型可分解为非周期性变化的趋势项、季节周期项和节假日效应等组件，从而可以具体分析各组件对预测结果变化的影响。相比于已有的时间序列预测算法，Prophet 最大的优势在于对历史数据中存在的节假日效应的影响具有很强的敏感性。文中结合 Prophet 算法这一优势，对铁路客运专线某站点的未来两周客流量进行了预测，实验结果表明，该模型中的节假日效应组件使模型对铁路客流量历史数据的拟合度增加，使其更加接近于真实情况，从而大大提高了模型最终的预测精确度。因此，将 Prophet 算法应用于铁路客流量预测分析是一种新的研究方法，对于铁路客流量的预测研究有一定应用价值。

参考文献：

[1] 常国珍,张前登.基于时间序列的我国铁路客流量预测[J].统计与咨询,2008(4):20-21.

[2] 成强.基于小波包与长短时记忆融合的铁路旅客流量预测模型[J].计算机系统应用,2018,27(7):121-126.

[3] LI L,WANG Y,ZHONG G, et al. Short-to-medium term passenger flow forecasting for metro stations using a hybrid model [J]. KSCE Journal of Civil Engineering, 2018, 22(1): 1937-1945.

[4] 王炜炜.高速铁路影响下铁路客流量预测研究[J].铁道运输与经济,2016,38(4):42-46.

[5] 路小娟,马宝峰,张武娟.铁路客流量预测分析与研究[J].兰州交通大学学报,2013,32(6):28-31.

[6] 许俊,王登,李晨毓,等.基于动态反馈神经网络的城市轨道交通短期客流预测[J].现代交通技术,2014,11(5):43-46.

[7] 马彪.GM(1,1)模型在春运铁路客流预测中的应用[J].

(下转第 150 页)