

马超群,李培坤,朱才华,等.基于不同时间粒度的城市轨道交通短时客流预测[J].长安大学学报:自然科学版,2020,40(3):75-83.
MA Chao-qun,LI Pei-kun,ZHU Cai-hua,et al.Short-term passenger flow forecast of urban rail transit based on different time granularities
[J].Journal of Chang'an University:Natural Science Edition,2020,40(3):75-83.

DOI:10.19721/j.cnki.1671-8879.2020.03.008

基于不同时间粒度的城市轨道交通短时客流预测

马超群,李培坤,朱才华,鲁文博,田 甜

(长安大学 运输工程学院,陕西 西安 710064)

摘 要:为探究城市轨道交通进站客流量预测精度与时间粒度之间的关系,以西安地铁自动售检票(AFC)系统连续 50 d 进站客流量数据为研究依据,将地铁运营有效时间划分为 5、15、30 min、1、2 h 及 1 d 等不同时间粒度,并对不同时间粒度下客流量时间序列采用 Pearson 系数法进行相似性度量,然后利用差分整合移动平均自回归(ARIMA)模型对不同时间粒度下的全网进站量进行拟合。以 Pearson 系数等于 0.95 作为短时客流预测时间粒度的选取阈值,最终选取在 30、60 min 及 1 d 三种时间粒度下用 ARIMA 模型进行短时客流预测,并与自回归滑动平均(AR)模型、支持向量回归(SVR)模型和 BP 神经网络预测模型的预测结果进行比较分析。研究结果表明:时间粒度相关性系数变化呈现单波峰形态,在 30、60 min 及 1 d 时间粒度下,ARIMA 模型预测结果平均相对误差分别为 4.12%、3.54%、4.97%;在这 4 种模型中,ARIMA 模型平均预测精度最高,在不同时间粒度下这 4 种模型的预测误差呈现相同的变化趋势,即平均预测误差由大到小的 3 种时间粒度分别为 1 d、30、60 min。因此,时间粒度大小选取的极端化并不会带来短时客流预测效果的直接提升,优化后的时间序列模型在西安地铁全网进站客流量短期预测方面具有较高的精度,研究成果可为城市轨道交通行车组织优化提供技术支持。

关键词:交通工程;城市轨道交通;短时客流预测;ARIMA 模型;时间粒度;相似性度量

中图分类号:U492.413

文献标志码:A

文章编号:1671-8879(2020)03-0075-09

Short-term passenger flow forecast of urban rail transit based on different time granularities

MA Chao-qun, LI Pei-kun, ZHU Cai-hua, LU Wen-bo, TIAN Tian

(College of Transportation Engineering, Chang'an University, Xi'an 710064, Shaanxi, China)

Abstract: In order to explore the relationships between forecast accuracy and time granularity of passenger flow in urban rail transit, based on the inbound passenger flow data of automatic fare collection (AFC) system for 50 consecutive days in Xi'an Metro, the effective time of metro operation was divided into different time granularities, such as 5 minutes, 15 minutes, 30 minutes, 1 hour, 2 hours and 1 day. The similarity of passenger flow time series under different time granularities was measured by Pearson coefficient method. The autoregressive integrated moving average (ARIMA) model was used to fit and predict the total network entry under

收稿日期:2019-12-19

基金项目:国家自然科学基金项目(71871027)

作者简介:马超群(1978-),男,浙江东阳人,副教授,工学博士,E-mail:machaoqun314@163.com。

different time granularities. Taking Pearson coefficient equal to 0.95 as the threshold of time granularity selection for short-term passenger flow forecast, ARIMA model was finally used for short-term passenger flow forecast under three time granularities of 30 minutes, 60 minutes and 1 day. The forecast results of ARIMA model were compared with those of autoregressive (AR) model, support vector regression (SVR) and BP neural network model. The results show that the time granularity correlation coefficient change presents a single peak form, and the average relative error of ARIMA model for 30 minutes, 60 minutes and 1 day time granularities are 4.12%, 3.54% and 4.97% respectively. The forecast results of the four models show that the average forecast accuracy of the ARIMA model is the highest. The four methods have the same change tendency for the forecast error at different time granularities, that is, the average forecast error decreases according to the three time granularities of 1 day, 30 minutes and 60 minutes. Therefore, the extreme selection of time granularity will not directly improve the effect of short-term passenger flow forecast, the optimized time series model has high accuracy in the analysis and short-term prediction of passenger flow data of Xi'an Metro stations, the research results can provide technical support for the optimization of the operation organization of urban rail transit. 7 tabs, 10 figs, 24 refs.

Key words: traffic engineering; urban rail transit; short-term passenger flow forecast; ARIMA model; time granularity; similarity measure

0 引言

随着中国城市轨道交通的快速发展,大客流现象频繁发生,城市轨道交通客流预测的精细化发展成为一种趋势^[1]。这种精细化的发展可以为城市轨道交通运营部门的管控措施、行车调度等提供技术支持。Li 等提出了一种动态径向基函数(RBF)神经网络,所提出的方法可以提前 30 min 确定关键的车站和时段,这对考虑采取主动客流控制以缓解地铁网络高峰时段的拥堵很有帮助^[2]。曹双等给出了区间客流来源和高峰小时内超高峰时刻的预测方法,为限流车站和限流时间的确定提供了帮助^[3]。

在客流预测方法上,城市轨道客流预测以中长期的客流预测为主,且多以传统四阶段法为基础,或在此基础上进行改进^[4]。徐瑞华等使用基于乘客多路径选择概率模型预测客流分布,为轨道交通网络化运营下的客流分布实时计算提供了理论基础^[5]。吴祥云等基于均衡分配原理,并采用 Frank-Wolf 算法进行客流分配,但计算结果缺少实践的检验^[6]。

在短时交通需求预测方法应用上,较为成熟的预测模型分为两类。一类是基于数理统计及数值分析的理论模型,主要有时间序列预测模型、卡曼尔滤波预测模型及灰色模型等。Liu 等提出一种灰色神经网络模型用于提升网络流量预测精度,可有效降

低网络流量预测误差,对流量管控起到重要作用^[7]。Williams 基于 ARIMA 模型原理,提出了一种包含多个传感器的多元 ARIMAX 预测模型,并对法国高速公路数据进行了预测,结果表明 ARIMAX 模型在数据缺失下的鲁棒性分析表现较好^[8]。Guo 等改进了传统卡尔曼滤波在预测过程中方差不变的特点,并使用改进后的自适应卡尔曼滤波方法对实际交通流量进行了预测,研究表明,改进后的模型对交通流量较大的情况具有较好的适应性^[9]。另一类是基于计算机仿真技术的预测模型,如神经网络模型、遗传算法模型等。Yin 等开发了一种模糊神经网络模型(FNM)来预测城市道路的交通流量,这种方法的优点是可以通过在线循环训练程序实时调整模型参数来增强 FNM 的预测能力^[10]。张纪升等认为路网关键道路识别是交通需求预测的基础,因此,提出一种交通分配理论与 K-短路径相结合的算法来提升道路关键路段识别能力,并在特定路网条件下证明了模型的有效性^[11]。伍拾煤等针对城镇客流分布预测的需求,利用最大熵原理构建市区、市域 2 个层级的最大熵预测模型,但模型的提出受不确定因素影响较多^[12]。Dimitriou 等在遗传算法框架基础上采用混合模糊系统(FRBS)方法,对城市主干道车流量进行了短时预测,其应用结果优于传统的遗传算法预测结果^[13]。

国内外学者在短时交通需求预测的工作中,缺

乏对预测对象系统性的分析,而先验数据时间粒度的选取是短时交通需求预测工作的基础,直接影响预测结果的精确度。目前学者们在短时交通需求预测时间粒度的选取上尚缺乏标准。在时间粒度的选取上,李得伟等使用 0.5、3、15 min 3 个时间粒度分别建立对应预测模型,对北京市某车站短时客流进行预测,并将此方法应用于北京市地铁安全防范实践中,取得较好的效果^[14]。Ma 等使用交互式多模型混合方法,对 7、1 d 及 1 h 为单位的短期乘客量进行了预测,并证明混合模型的有效性,为更短时间粒度的预测提供了基础^[15]。Huang 等提出了一种基于人类记忆特性的自旋网络预测方法 (SPN),预测了 15、30 min 时间粒度内的道路车辆数,其预测精度在可接受范围内^[16]。刘美琪等分别以北京市地铁 13 号线早高峰、平峰、晚高峰及全天的客流量为研究对象,采用 K 临近非参数回归模型进行客流预测,结果表明平峰时段预测效果较好^[17]。张晚笛等通过北京市地铁刷卡数据来分析不同时间粒度下的乘客出行规律,结果表明工作日早高峰与平峰客流比晚高峰历史同期客流相似度高,这为客流预测时间粒度的选取提供了依据^[18]。Zhong 等通过对伦敦、新加坡、北京 3 座城市基于智能交通卡的乘客出行流动模式研究,指出了不同时间粒度变化下客流数据的多样性和规律性^[19]。目前,学者对先验数据的时间粒度大小选择处于探索阶段,因此,本文的主要目的在于探寻先验数据时间粒度大小与预测精度之间的关系。

经典回归模型主要通过不同变量之间的因果关系对事物发展做出预测,但实际问题中某些影响因素难以采集以及变量的多样性,难以建立回归模型去研究其发展规律,而时间序列模型不需要建立因果关系,只根据数据自相关动态建模。因此,ARIMA 模型作为一种时间序列模型,被广泛应用于交通领域,此模型是在 ARMA 模型基础上对数据进行差分减噪优化得到的^[20]。史其信等使用历史平均、ARIMA、非参数回归、RBF 神经网络等模型对路网交通流量进行了预测,结果表明,ARIMA 模型在先验数据无缺失的情况下具有自适应调节能力且预测精度满足要求^[21]。孙湘海等在 ARIMA 模型的基础上,提出一种季节自回归求和移动平均模型与广义回归神经网络模型相结合的组合 (SARIMA-GRNN) 模型,此模型在城市道路短期交通流预测上表现优异^[22]。

本文以连续时间段内西安地铁全网进站客流量

为基础,首先对前期数据进行处理与分析^[12],选取数据相关性较高的时间粒度下客流进站量建立相应 ARIMA 模型,用此模型进行预测分析,并与其他客流预测方法预测精度进行比较。

1 时间粒度比选

从国内外研究现状分析可知,同比短时客流预测是利用历史同期数据作为先验条件预测当前时段客流量;环比预测法是利用预测时段前数小时客流数据作为先验条件,对所选时段客流量进行预测。对于相对稳定的客流大数据,有研究表明,同比预测法预测精度较高^[18],本文选取 5、15、30 min、1、2 h 及 1 d 时间粒度进行比选,即: $\Delta t \in \{5, 15, 30, 60, 120, 840\}$ min。为保证客流数据无较大波动,选取西安地铁全网客流相对稳定的 7:00~21:00 时段(计 840 min 时长)连续 50 d 进站客流量数据进行分析。将全天进站量数据分割为 6 个间隔段,即: $n \in \{168, 56, 28, 14, 7, 1\}$ 。用改进的时间序列描述车站进站量,即

$$\mathbf{X}_N = \{x_1, x_2, \dots, x_t, \dots, x_n\} \quad (1)$$

式中: \mathbf{X}_N 为第 N 期所选取时间内进站客流量形成的矩阵, N 为连续天数序列, $N \in \{1, 2, \dots, 50\}$; n 为在时间粒度 Δt 下每天地铁的有效运营时间 840 min 被平均分割的间隔段数, $n = 840 / \Delta t$; x_t 为第 t 个间隔段内的进站客流量, $t \in \{1, 2, \dots, n\}$ 。

对于时间上具有连续性的数据, Pearson 相关系数可以较好地判别数据之间的关联度^[23]。因此,本文对原始数据进行提取后,利用 Pearson 系数法进行数据相似性度量。假设全网第 i 天和第 j 天时间粒度为 Δt 的 Pearson 系数为 $r_{\Delta t}(\mathbf{X}_i, \mathbf{X}_j)$, 则其表示为第 i 天 \mathbf{X}_i 与第 j 天 \mathbf{X}_j 两矩阵之间 Pearson 系数乘积,具体计算方法如下

$$r_{\Delta t}(\mathbf{X}_i, \mathbf{X}_j) = \frac{\sum_{t=1}^n (x_{it} - \bar{x}_i)(x_{jt} - \bar{x}_j)}{\sqrt{\sum_{t=1}^n (x_{it} - \bar{x}_i)^2} \sqrt{\sum_{t=1}^n (x_{jt} - \bar{x}_j)^2}} \quad (2)$$

式中: x_{it} 为第 i 天第 t 间隔段进站客流量; \bar{x}_i 为第 i 天 Δt 时间粒度下的进站客流量均值。

为提高相似性度量精度,选取一周中相同星期天数进行计算,且令 $j = i + 1$, 因为要满足 Pearson 系数计算中样本个数要求,因此以 1 d 为时间粒度的客流预测单独进行计算,若 $r_{\Delta t}(\mathbf{X}_i, \mathbf{X}_j)$ 绝对值越大,则相关性越强;越接近 0,则相关性越弱。

从表 1 可知,不同时间粒度下的 Pearson 系数都接近 1,具有很强的相关性。总体而言,历史同期

表 1 不同时间粒度下 Pearson 相关系数

Tab. 1 Pearson correlation coefficients of different time granularities

$\Delta t/\text{min}$	5	15	30	60	120	240
范围	0.821~ 0.864	0.898~ 0.946	0.944~ 0.981	0.934~ 0.975	0.944~ 0.955	0.922~ 0.938
样本个数	560	560	280	140	70	70
加权均值	0.841	0.923	0.964	0.951	0.947	0.933

客流相似度越高,其短时进站客流预测精度越好,根据 Pearson 系数定义,以 0.95 作为客流规律性较强阈值,选取 30、60 min 及未进行客流相似性分析的 1 d 作为实例分析的时间粒度。

2 ARIMA 模型基本原理及应用

时间序列分析方法主要是从数据序列自相关角度去归纳揭示时间序列规律的一种分析方法^[24]。若随时间变化的数据序列随机过程的特征随着时间变化,则此过程是非平稳的;相反,如果随机过程的特征不随时间而变化,就称此过程是平稳的。自回归-积分-滑动平均 (ARIMA) 模型是一种常见时间序列。若导致非平稳的原因是随机的,则可以使用 ARIMA 及自回归条件异方差模型等。首先将预测对象按照时间排列顺序作为一随机序列,然后用数学模型来近似识别序列,最后可以从过去及现在时间序列值来预测未来值。

ARIMA(p, d, q)模型包含 3 个部分:即 AR 表示自回归模型;I 为单整阶数;MA 为移动平均模型。参数 p 代表预测模型中采用的时序数据本身的滞后数, d 代表时序数据在稳定之前需要进行几阶差分化, q 代表预测模型中采用的预测误差滞后数。

ARIMA(p, d, q)模型的一般方程可以表示为
$$\bar{y}_T = c + \varphi_1 x_{T-1} + \varphi_2 x_{T-2}, \dots, + \varphi_p x_{T-p} + \omega_1 \alpha_{T-1} + \omega_2 \alpha_{T-2}, \dots, + \omega_q \alpha_{T-q} \quad (3)$$
式中: \bar{y}_T 为对时间序列 X_N 经 d 阶差分后的第 T 期数值; $\varphi_1, \varphi_2, \dots, \varphi_p$ 为自回归模型系数; $\omega_1, \omega_2, \dots, \omega_q$ 为 MA 模型参数; c 为常数项; x_{T-1}, \dots, x_{T-p} 为先验数据值; $\alpha_{T-1}, \dots, \alpha_{T-q}$ 为预测误差值。

对于城市轨道交通进站客流量数据,可按以下步骤来建立 ARIMA 模型。

- (1)对车站时间排序进站客流量进行剔除及插值,得到初步优化时间序列数据。
- (2)根据序列图确定数据平稳趋势性差分阶数,计算差分后的时间序列数据自相关性 (ACF),并根据模型识别准则选择多组参数,建立 ARIMA(p, d, q)

- 模型。
- (3)对所建模型进行参数估计,并根据残差分析确定最终模型。
- (4)对模型进行残差白噪音检验,并结合起始数据进行分析及预测。

3 数据分析

3.1 数据预处理

本文选取西安地铁全网自动售检票系统 (AFC)统计获得的进站客流数据,数据起始日期为 2017 年 6 月 1 日,结束日期为 7 月 20 日。规定每日进站量为 1 期,共 50 期。

针对 30、60 min 时间粒度,选取连续 30 d 相同时间晚高峰进站客流量作为模型校验数据,后 20 期作为模型精度检验数据;对于以 1 d 为时间粒度的客流预测选取连续 40 d 西安地铁全网进站客流量作为模型校验数据,后 10 期作为模型精度检验数据。短时时间粒度可选取范围较广,为保证数据分析研究的实际利用价值,选取进站客流量较高的晚高峰 17:30~18:30 为研究时间段。

3.2 差分阶数的确定

分析不同时间粒度的进站客流量可知,进站客流量呈现以 7 d 为周期的上下浮动变化规律,并不围绕一个值波动,而利用 ARIMA 模型的前提条件为时间序列需围绕某一数值上下波动,反映在图形上表现为所有样本皆在某一水平线上下随机波动。因此,需首先对 3 种时间粒度下的进站客流数据进行一阶差分,其结果如图 1 所示。

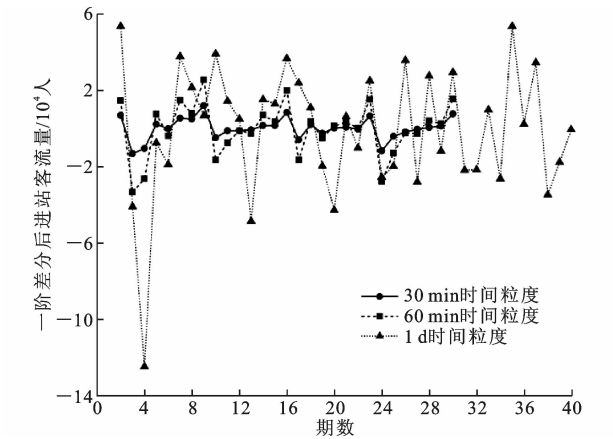


图 1 不同时间粒度一阶差分

Fig. 1 First order difference with different time granularities

利用 EVIEWS 软件对一阶差分后的地铁进站客流量进行 ADF 检验,其结果如表 2 所示。由表 2 可以看出,一阶差分后的 P 检验值小于 0.01,说明

在 1% 的显著水平下无单位根的存在,进一步说明一阶差分后的时间序列是平稳的,可对数据进行 ARIMA 建模分析。

表 2 ADF 平稳性检验
Tab. 2 ADF stability test

方法	时间粒度	t 值	P 值
扩展 Dickey-Fuller 检查 统计量	30 min	-4.490 322	0.001 2
	60 min	-3.245 154	0.003 1
	1 d	-3.154 901	0.002 2
ADF 检验(1%水平)	30 min	-3.661 661	
	60 min	-2.960 411	
	1 d	-2.619 160	

4 模型比选与参数确定

4.1 1 d 时间粒度模型参数确定

经过一阶差分处理后的时间序列在构建 ARIMA 模型时,参数的确定可以结合此时间序列的自相关与偏自相关图来确定,可用 Q 统计量反映差分数据中没有被主元模型所解释的变化范围,其对应的 P 值大小可以说明这种变化幅度如表 3 所示。由表 3 可知:偏自相关系数具有截尾趋势,自相关系数具有拖尾趋势,因此满足 ARIMA 模型的要求;且 PAC 在二阶截尾,因此子模型 AR 参数取值为 2,子模型 MA 取值在二阶及三阶差分处截距消失,则其取值为 2 或者 3。此时 ARIMA 模型基本确定。在 1 d 时间粒度预测模型选取时需对 2 种备选模型进行比选,使用 EVIEWS 对 ARIMA(2,1,2) 及 ARIMA(2,1,3) 模型进行最优分析,挑选最优预测模型,对模型进行拟合建模,检验结果如表 4 所示。可知,当 q 值选取为 3 时, P 值为 0,能较好拟合数据,因此 ARIMA(2,1,3) 优先作为数据预测模型。

4.2 30 min 时间粒度模型参数确定

对 30 min 时间粒度下差分后的时间序列数据做自相关及偏自相关,如图 2 和图 3 所示。可见,自相关系数具有二阶拖尾现象,偏自相关系数从二阶开始具有截尾趋势,因此,30 min 时间粒度时应选择 ARIMA(2,1,2) 模型。

4.3 60 min 时间粒度模型参数确定

同样对 60 min 时间粒度一阶差分后的时间序列做自相关图及偏自相关图,如图 4 和图 5 所示。可见,自相关系数具有二阶拖尾现象,偏自相关系数从二阶开始具有截尾趋势,因此,60 min 时间粒度时同样选择 ARIMA(2,1,2) 模型。

表 3 1 d 时间粒度自相关及偏自相关
Tab. 3 Autocorrelation and partial autocorrelation of
1 day time granularity

编号	自相关系数	偏自相关系数	Q 统计量	P 值
1	-0.314	-0.314	4.145 3	0.042
2	-0.397	-0.550	10.952	0.004
3	0.179	-0.296	12.381	0.006
4	0.216	-0.062	14.516	0.006
5	-0.252	-0.224	17.493	0.004
6	-0.239	-0.550	20.252	0.002
7	0.582	0.084	37.158	0.000
8	-0.243	-0.362	40.198	0.000
9	-0.166	-0.082	41.660	0.000
10	0.157	-0.110	43.022	0.000
11	0.148	0.040	44.268	0.000
12	-0.269	-0.158	48.557	0.000
13	-0.109	-0.064	49.290	0.000
14	0.418	-0.084	60.450	0.000
15	-0.216	-0.082	63.565	0.000

表 4 ARIMA(2,1,3) 模型检验结果
Tab. 4 ARIMA(2,1,3) model test results

变量	系数	标准差	t 值	P 值
c	2 236.846	537.845 0	4.158 905	0.000 2
AR(1)	-0.430 188	0.051 352	-8.377 043	0.000 1
AR(2)	-0.965 628	0.042 501	-22.720 110	0.000 0
MA(1)	-0.378 250	0.086 744	-4.360 543	0.000 1
MA(2)	0.386 788	0.091 959	4.206 111	0.000 2
MA(3)	-0.928 482	0.039 431	-23.547 240	0.000 0
AR 特征根倒数	-22+96i	-22-96i		
MA 特征根倒数	97	-30-93i	-30+93i	

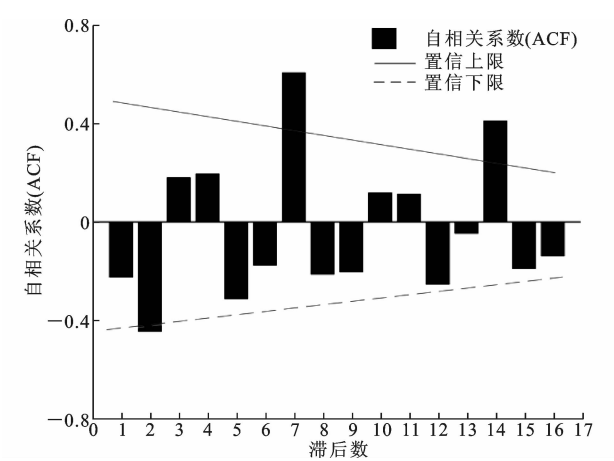


图 2 30 min 时间粒度自相关
Fig. 2 Autocorrelation of 30 min time granularity

4.4 残差分析

对于所建模型进行协整检验,利用 EVIEWS 软

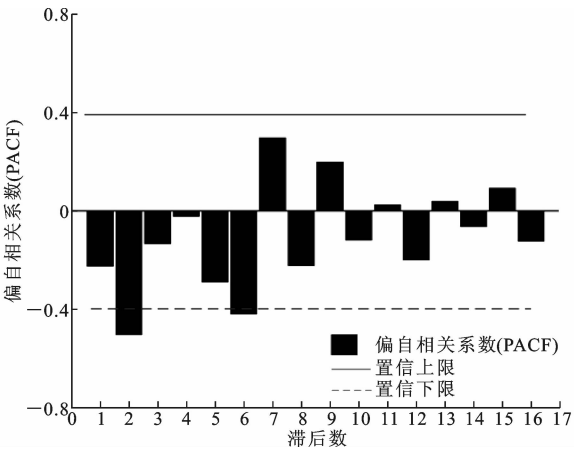


图3 30 min时间粒度偏自相关

Fig. 3 Partial autocorrelation of 30 min time granularity

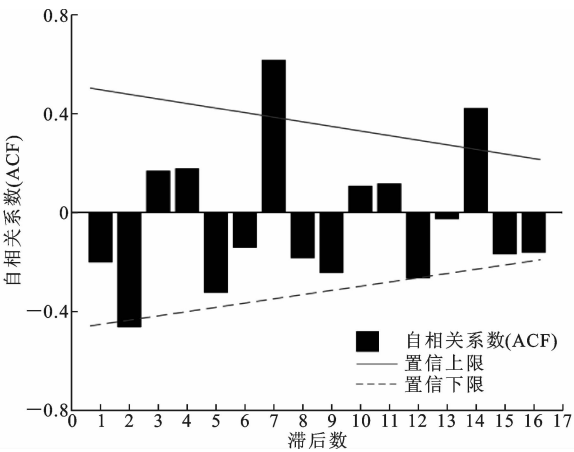


图4 60 min时间粒度自相关

Fig. 4 Autocorrelation of 60 min time granularity

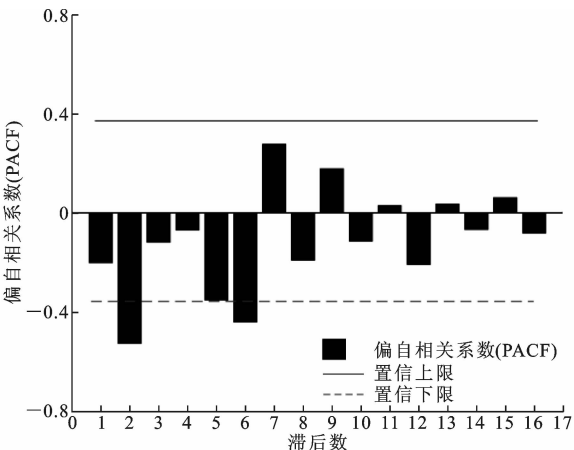


图5 60 min时间粒度偏自相关

Fig. 5 Partial autocorrelation of 60 min time granularity

件得到残差项时间序列,对1 d时间粒度时间序列进行 ADF 检验,见表 5。由表 5 可知, t 值为 $-5.663\ 119$,小于临界值,因此残差项平稳,所选模型稳定。

表 5 残差 ADF 平稳性检验

Tab. 5 ADF test of the residual data stability

方法	t 值	P 值
扩展 Dickey-Fuller 检查	$-5.663\ 119$	0.0012
不同显著水平(%)下的 ADF 检验	1	$-3.670\ 170$
	5	$-2.963\ 972$
	10	$-2.621\ 007$

对 30、60 min 时间粒度所选模型作残差相关函数关系,如图 6 和图 7 所示。可见,绝大部分残差相关函数值均在 95%置信区间内,即可认为残差为白噪声序列,模型通过检验。

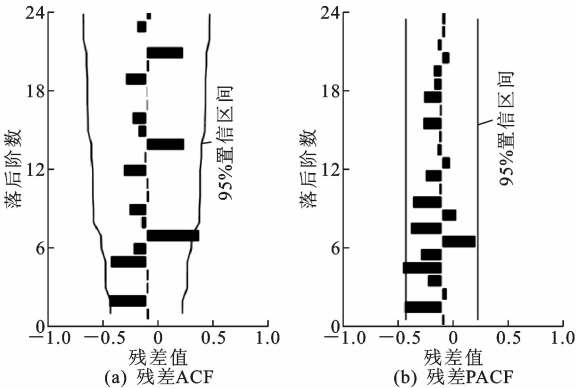


图6 30 min时间粒度残差相关函数

Fig. 6 Residual correlations function of 30 min time granularity

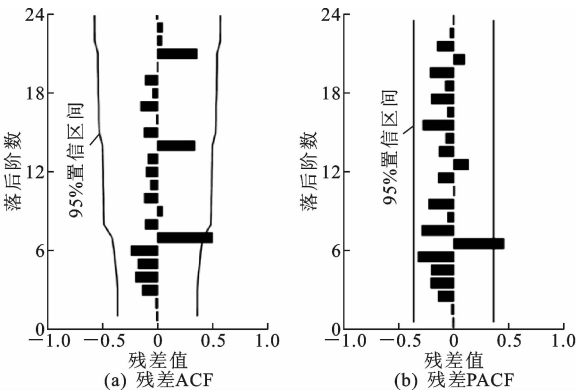


图7 60 min时间粒度残差相关函数

Fig. 7 Residual correlations function of 60 min time granularity

5 模型拟合与预测应用

5.1 1 d 时间粒度模型数据检验

由上述模型参数分析可知,1 d 时间粒度时 ARIMA(2,1,3)模型方程为

$$dx_T = 2\ 236.846 - 0.430x_{T-1} - 0.965x_{T-2} - 0.378\alpha_{T-1} + 0.387\alpha_{T-2} - 0.928\alpha_{T-3} \quad (4)$$

式中: dx_T 为一阶差分后第 T 期进站客流量数据; x_{T-1} 为差分数据的第 $T-1$ 期数据; x_{T-2} 为差分数据的第 $T-2$ 期数据; α_{T-1} 为差分数据第 $T-1$ 期误

差值; α_{T-2} 为差分数据第 $T-2$ 期误差值; α_{T-3} 为差分数据第 $T-3$ 期误差值。

利用已建立的 ARIMA(2,1,3)模型对前 40 期数据进行动态拟合,并将一阶差分所得数据进行初始化处理,所得拟合数据如图 8 所示。

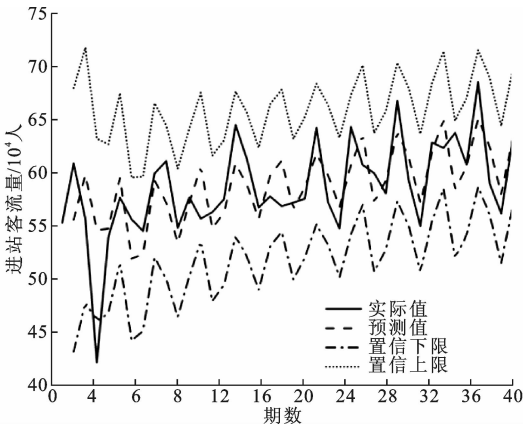


图 8 1 d 时间粒度拟合预测

Fig. 8 Fitting-forecast of 1 day time granularity

5.2 30 min 时间粒度模型数据检验

30 min 时间粒度 ARIMA(2,1,2)模型方程为
$$dx_T = 187.196 + 0.069x_{T-1} - 0.409x_{T-2} - 0.823\alpha_{T-1} - 0.118\alpha_{T-2} \quad (5)$$

以此模型方程对前 30 期数据进行拟合,并将拟合值与实际值进行比较,拟合结果如图 9 所示,可知实际进站量全部在模型拟合的置信区间内。

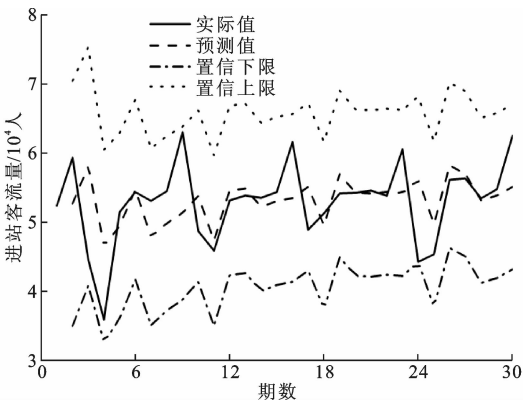


图 9 30 min 时间粒度拟合预测

Fig. 9 Fitting-forecast of 30 min time granularity

5.3 60 min 时间粒度模型数据检验

60 min 时间粒度 ARIMA(2,1,2)模型方程为
$$dx_T = 454.990 + 0.112x_{T-1} - 0.432x_{T-2} - 0.859\alpha_{T-1} - 0.086\alpha_{T-2} \quad (6)$$

以此模型方程对前 30 期数据进行拟合,将拟合值与实际值进行比较,拟合结果如图 10 所示,可知实际进站量也全部在模型拟合的置信区间内。

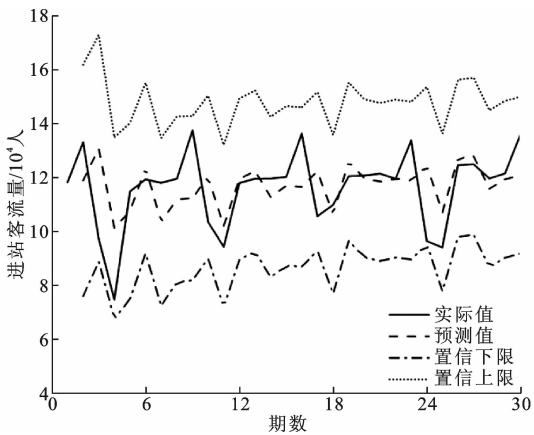


图 10 60 min 时间粒度拟合预测

Fig. 10 Fitting-forecast of 60 min time granularity

5.4 ARIMA 模型预测分析

将拟合值与实际数据相比,取置信区间为 95%,所建模型拟合数据与实际数据相差较小,拟合结果整体较为理想。利用所建模型对后 10 期或 20 期数据进行预测,所得数据分析如表 6 所示。

表 6 不同时间粒度下的预测误差

Tab. 6 Forecast error of different time granularities

粒度选取	期数	平均相对误差/%	最大相对误差/%	P 值
30 min	20	4.12	7.45	0.802
60 min	20	3.54	7.23	0.833
1 d	10	4.97	6.21	0.842

由表 6 可知:平均相对误差均控制在 5% 以内,预测结果显著性都较高,基于 ARIMA 模型所得结果可以动态预测时间序列的变化;满足置信区间 95% 的要求,且符合实际客流量变化;60 min 时间粒度对应的预测误差最小,因此在实际短时客流预测中选取 60 min 时间粒度较为精准。

5.5 预测模型比选

为了进行预测性能比较,使用其他 3 种预测模型对选取时间粒度下的进站客流量进行预测,并将预测结果与 ARIMA 模型进行比较。

自回归滑动 (AR) 模型在较为平稳的数据序列预测过程中具有广泛的应用;支持向量回归 (SVR) 模型是一种用于分类和线性回归分析的机器学习方法,此研究方法可剔除大量冗余数据,具有很好的鲁棒性;BP 神经网络模型是一种经典神经网络预测模型,其输出结果误差反向传播,随着训练迭代次数的增加,可减少预测误差的产生。

本文中 BP 神经网络参数设置为:输入层设置 7 个节点、输出层 1 个节点、隐藏层设置 5 个节点、学习速率为 0.005、最大迭代次数为 500。3 种模型与

ARIMA 模型训练数据数相同,即对于 30、60 min 时间粒度下的客流进站量,训练集样本数选取为 30 期,1 d 时间粒度训练集样本数为 40 期。预测效果如表 7 所示。

表 7 不同预测方法的预测误差

Tab. 7 Forecast error of different forecast methods

预测模型	时间粒度	平均相对误差/%	最大相对误差/%	P 值
AR 模型	30 min	8.61	10.45	0.712
	60 min	6.54	8.23	0.766
	1 d	6.97	6.21	0.826
SVR 模型	30 min	7.74	9.45	0.756
	60 min	6.12	8.23	0.798
	1 d	6.43	8.97	0.843
BP 神经网络模型	30 min	6.73	10.87	0.826
	60 min	5.33	8.34	0.862
	1 d	6.27	8.02	0.912

从表 6、表 7 可知:ARIMA 模型相较于其他 3 类预测方法,在 3 种时间粒度下预测平均相对误差最小,SVR 及 BP 神经网络模型预测效果次之,AR 模型预测效果最差;BP 神经网络模型 *P* 值较大表明预测波动较小,ARIMA 模型次之,但优于 SVR 及 AR 模型;且 4 种预测模型对 3 种时间粒度预测误差均呈现相同趋势。

在选定的时间粒度下,由于地铁进站客流量的波动较小,且差分后的进站客流数据更加稳定,因此采用 ARIMA 模型进行地铁进站量预测的精确度更高。同时,由于 ARIMA 模型训练复杂程度相对较小且无需大量先验数据,对客流的快速预测更加有效。

6 结 语

(1)本文首先对不同时间粒度下连续进站客流量使用改进 Pearson 系数法,进行了关联度比选,其中时间粒度为 30、60 min 时数据关联度最高。

(2)ARIMA 模型可以较好描述地铁进站客流量波动情况,预测精度均能保持在较高的水准,基本能满足地铁运营对客流量的管控要求。

(3)通过与 AR、SVR 及 BP 神经网络模型预测结果对比发现,由于先验数据波动较小,ARIMA 模型以全网进站量为先验数据的平均预测误差最小。且 4 类预测模型对不同时间粒度下的进站量预测误差呈现相同的规律,即 60 min 时间粒度下的进站客流量平均预测误差最小。

(4)短时客流预测时间粒度的选取并非越小或越大就好。首先时间粒度较大对实际客流变化的掌控

不具有很好的现实意义;其次,当时间粒度较小时,某些数据可能因突发恶劣天气或大型演出等引起的客流激增导致数据大幅波动,会影响预测的精确性。

参考文献:

References:

[1] 王玉萍,马超群.城市轨道交通客运量影响因素与成长规律[J].长安大学学报:自然科学版,2013,33(3):69-75.
WANG Yu-ping, MA Chao-qun. Influencing factors and development trends of urban rail transit passenger flow[J]. Journal of Chang'an University: Natural Science Edition, 2013, 33(3): 69-75.

[2] LI H, WANG Y, XU X, et al. Short-term passenger flow prediction under passenger flow control using a dynamic radial basis function network [J]. Applied Soft Computing, 2019, 83: 105620.

[3] 曹 双,马 骊.基于区间客流推算的城市轨道交通多车站协同限流研究[J].交通运输工程与信息学报, 2018, 16(4): 142-151.
CAO Shuang, MA Si. Collaborative multi-station flow restraints of urban rail transit based on the calculation of section passenger flow[J]. Journal of Transportation Engineering and Information, 2018, 16 (4): 142-151.

[4] 马超群,陈宽民,王玉萍.城市轨道交通客流预测方法[J].长安大学学报:自然科学版,2010,30(5):69-74.
MA Chao-qun, CHEN Kuan-min, WANG Yu-ping. Forecasting model of urban rail transit volume[J]. Journal of Chang'an University: Natural Science Edition, 2010, 30(5): 69-74.

[5] 徐瑞华,徐永实.城市轨道交通线路客流分布的实时预测方法[J].同济大学学报:自然科学版,2011,39(6):857-861.
XU Rui-hua, XU Yong-shi. Real-time forecast of passenger flow distribution on urban rail transit line[J]. Journal of Tongji University: Natural Science, 2011, 39(6): 857-861.

[6] 吴祥云,刘灿齐.轨道交通客流量均衡分配模型与算法[J].同济大学学报:自然科学版,2004,32(9):1158-1162.
WU Xiang-yun, LIU Can-qi. Traffic equilibrium assignment model specially for urban railway network [J]. Journal of Tongji University: Natural Science, 2004, 32(9): 1158-1162.

[7] LIU Y, CAO J. Network traffic prediction based on grey neural network integrated model [C]//WANG F, XIA H G. Proceedings of the International Confer-

- ence on Computer Science and Software Engineering. New York:IEEE,2008:915-918.
- [8] WILLIAMS B M. Multivariate vehicular traffic flow prediction; Evaluation of ARIMAX modeling [J]. Transportation Research Record, 2001 (1776): 194-200.
- [9] GUO J, HUANG W, WILLIAMS B M. Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification [J]. Transportation Research Part C, 2014, 43: 50-64.
- [10] YIN H, WONG S C, XU J, et al. Urban traffic flow prediction using a fuzzy-neural approach [J]. Transportation Research Part C, 2002, 10(2): 85-98.
- [11] 张纪升,贾利民,牛树云,等. 基于 K-短路径的路网关键路段集合的辨识与分析[J]. 长安大学学报:自然科学版, 2015, 35(3): 122-129.
- ZHANG Ji-sheng, JIA Li-min, NIU Shu-yun, et al. Identification and analysis of road network key segments set based on K-shortest path [J]. Journal of Chang'an University: Natural Science Edition, 2015, 35(3): 122-129.
- [12] 伍拾煤,裴玉龙,程国柱. 密集城镇群客流分布双层最大熵模型[J]. 中国公路学报, 2014, 27(5): 164-170.
- WU Shi-mei, PEI Yu-long, CHENG Guo-zhu. Double maximum entropy distribution model for passenger flow of dense urban group [J]. China Journal of Highway and Transport, 2014, 27(5): 164-170.
- [13] DIMITRIOU L, TSEKERIS T, STATHOPOULOS A. Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow [J]. Transportation Research Part C, 2008, 16(5): 554-573.
- [14] 李得伟,尹浩东. 基于物联网设备检测数据的城市轨道交通车站站台多维度实时客流密集度指数预测方法研究[J]. 铁道学报, 2014, 36(3): 9-13.
- LI De-wei, YIN Hao-dong. Real-time forecast of passenger crowd index on urban rail transit station platform based on test data of internet of things [J]. Journal of the China Railway Society, 2014, 36(3): 9-13.
- [15] MA Z, XING J, MESBAH M, et al. Predicting short-term bus passenger demand using a pattern hybrid approach [J]. Transportation Research Part C, 2014, 39: 148-163.
- [16] HUANG S, SADEK A W. A novel forecasting approach inspired by human memory; The example of short-term traffic volume forecasting [J]. Transportation Research Part C, 2009, 17(5): 510-525.
- [17] 刘美琪,焦朋朋,孙 拓. 城市轨道交通进站客流量短时预测模型研究 [J]. 城市轨道交通研究, 2015, 18(11): 13-17, 29.
- LIU Mei-qi, JIAO Peng-peng, SUN Tuo. On short-term forecasting model of passenger flow in urban rail transit [J]. Urban Mass Transit, 2015, 18(11): 13-17, 29.
- [18] 张晚笛,陈 峰,王子甲,等. 基于多时间粒度的地铁出行规律相似性度量 [J]. 铁道学报, 2018, 40(4): 9-17.
- ZHANG Wan-di, CHEN Feng, WANG Zi-jia, et al. Similarity measurement of metro travel rules based on multi-time granularities [J]. Journal of the China Railway Society, 2018, 40(4): 9-17.
- [19] ZHONG C, BATTY M, MANLEY E, et al. Variability in regularity: Mining temporal mobility patterns in London, Singapore and Beijing using smart-card data [J]. PLOS ONE, 2016, 11(2): e0149222.
- [20] CHEN L L, HE P H, CAO L, et al. Passenger flow prediction of exhibition based on ARMA [J]. Applied Mechanics & Materials, 2014, 667: 11-15.
- [21] 史其信,郑为中. 道路网短期交通流预测方法比较 [J]. 交通运输工程学报, 2004, 4(4): 68-71, 83.
- SHI Qi-xin, ZHENG Wei-zhong. Short-term traffic flow prediction methods comparison of road networks [J]. Journal of Traffic and Transportation Engineering, 2004, 4(4): 68-71, 83.
- [22] 孙湘海,刘潭秋. 基于神经网络和 SARIMA 组合模型的短期交通流预测 [J]. 交通运输系统工程与信息, 2008, 8(5): 32-37.
- SUN Xiang-hai, LIU Tan-qiu. Short-term traffic flow forecasting based on a hybrid neural network model and SARIMA model [J]. Journal of Transportation Systems Engineering and Information Technology, 2008, 8(5): 32-37.
- [23] ZIEGEL E R. Standard probability and statistics tables and formulae [J]. Technometrics, 2001, 43(2): 249.
- [24] 王 莹,韩宝明,张 琦,等. 基于 SARIMA 模型的北京地铁进站客流量预测 [J]. 交通运输系统工程与信息, 2015, 15(6): 205-211.
- WANG Ying, HAN Bao-ming, ZHANG Qi, et al. Forecasting of entering passenger flow volume in Beijing subway based on SARIMA model [J]. Journal of Transportation Systems Engineering and Information Technology, 2015, 15(6): 205-211.