

基于 PSO-FSVR 的城市轨道交通客流预测模型^{*}

孟 歌 郝晓培 张军锋^{**} 王洪业 李 永 毛子今

(中国铁道科学研究院集团有限公司电子计算技术研究所, 100081, 北京//第一作者, 助理研究员)

摘 要 目的:为了提高城市轨道交通(以下简称“城轨”)运营决策效率,控制运营成本,减少资源空置和浪费,有必要对城市轨道交通客流进行精准预测。方法:提出一种基于 PSO(粒子群优化)-FSVR(模糊支持向量回归机)的城轨客流预测模型。对城轨客流数据进行清洗和分析,根据获取的数据构建特征空间,利用 PSO 算法对 FSVR 模型中的参数进行优化以提高预测精度,引入模糊隶属度构建 PSO-FSVR 模型,增强客流预测方法的泛化能力。将该模型应用于北京地铁 2 号线某车站的进出站客流预测中,并引入 BP(反向传播)神经网络模型与 GRNN(广义回归神经网络)模型作为对比方法,采用均方根误差、显著性参数及相关系数等指标对预测结果进行评价。结果及结论:基于 PSO-FSVR 的城轨客流预测模型下,工作日、节假日客流预测绝对均方根误差分别为 7.010 4 和 8.778 5,相关系数分别为 0.993 0 和 0.955 8,相较其他两种模型,该模型客流预测性能最优。基于 PSO-FSVR 的城轨客流预测模型能够有效提高城轨客流预测的准确性,尤其在客流极值点处的表现比其他方法更加优秀,具有良好的预测性能。

关键词 城市轨道交通;客流预测;粒子群优化;支持向量回归机

中图分类号 U293.13

DOI:10.16037/j.1007-869x.2023.10.008

Urban Rail Transit Passenger Flow Prediction Model Based on PSO-FSVR

MENG Ge, HAO Xiaopei, ZHANG Junfeng, WANG Hongye, LI Yong, MAO Zijin

Abstract Objective: In order to enhance the efficiency of operational decision-making, control operating costs, and reduce resource wastage in URT (urban rail transit), it is essential to accurately predict passenger flow. Method: An URT passenger flow prediction model based on PSO (particle swarm optimization) and FSVR (fuzzy support vector regression) is proposed. The passenger flow data of URT are cleaned and analyzed, and a feature space is constructed based on the acquired

data. The PSO algorithm is utilized to optimize the parameters of the FSVR model for higher prediction accuracy. Fuzzy membership degrees are introduced to construct the PSO-FSVR model, thereby improving the model generalization ability. This model is applied to predict the inbound and outbound passenger flow of a station on Beijing Subway Line 2. BP (back-propagation) neural network and GRNN (general regression neural network) models are taken as comparative methods. The evaluation of prediction results is carried out using metrics such as RMSE (root mean square error), significance parameters, and correlation coefficients. Result & Conclusion: Under the PSO-FSVR-based URT passenger flow prediction model, the absolute RMSE for weekday and holiday passenger flow predictions are 7.0104 and 8.7785, respectively; the corresponding correlation coefficients are 0.9930 and 0.9558, respectively. Compared to the other two models, this model exhibits the best passenger flow prediction performance. The PSO-FSVR-based URT passenger flow prediction model effectively enhances the accuracy of passenger flow prediction, particularly excelling at predicting passenger flow extremes, demonstrating excellent predictive performance.

Key words urban rail transit; passenger flow prediction; particle swarm optimization; support vector regression machine

Author's address Institute of Computing Technology, China Academy of Railway Sciences Group Co., Ltd., 100081, Beijing, China

目前,城市轨道交通(以下简称“城轨”)正向智慧化、绿色化方向发展,其中重要的前提之一就是使轨道交通系统具备预测能力,以达到自动化、系统化、智慧化的目标。智慧轨道交通需要强大的感知能力作为支持,准确掌握城轨车站的客流量是关键基础。只有精准预测未来车站客流量,才能够做出准确决策并进行快速响应,以提高城轨系统的智

^{*} 中国铁道科学研究院集团有限公司科研项目(2021YJ191)

^{**} 通信作者

慧化水平。同时,在掌握客流变化趋势的基础上,可以合理安排资源配置,减少空置和浪费,达到绿色可持续发展的目标。

在交通预测领域,常用的方法主要包括专家意见法、回归分析法、支持向量机^[1]、神经网络^[2]及时间序列^[3]等。其中:专家意见法依靠人工判断,结果较为主观,不利于城轨信息系统的自动化建设发展;回归分析法适用于变量间的相关性为线性或简单非线性情况,不适用于复杂非线性情况;时间序列方法建立在线性平稳模型的基础上,但城轨客流并非正态分布的平稳序列,客流波动性强烈,会导致预测效果不佳;神经网络方法为非线性自回归模型,适用于大部分预测场景,但需要调整的参数较多,且网络的构建与构建者的经验有关,因此预测性能因人而异,泛化能力较弱;支持向量机具有坚实的理论基础和良好的泛化能力,尤其在解决小样本、非线性问题时具有许多优势,但由于其中参数选择的范围较大,难以获得最优参数,因此需要对其进行优化。基于此,本文提出了一种基于 PSO (粒子群优化)-FSVR (模糊支持向量回归机)的城轨客流预测模型(以下简称“基于 PSO-FSVR 的城轨客流预测模型”),并将其应用在北京轨道交通线路的实际客流数据预测中。与常用的 GRNN(广义回归神经网络)和 BP(反向传播)神经网络相比,本文方法具有更加良好的预测性能,可以有效对城轨车站客流数据进行精准预测。

1 FSVR 模型

SVR (支持向量回归机)是建立在有限样本 SLT(统计学习理论)基础上的通用学习方法^[4]。FSVR 是将模糊数学和 SVR 相结合的一种新型的 SVR,其通过引入模糊隶属度,来提高机器学习的鲁棒性和泛化能力。

FSVR 的基本原理是依靠预测误差最小化寻找某一目标函数,利用非线性映射函数 $\varphi(\cdot)$ 将输入空间的数据点 \mathbf{x}_i 映射至高维空间 H 中,进而在 H 中采取线性回归方法进行计算,以获得原来低维空间中非线性回归的效果。

由于在真实环境中,数据点 \mathbf{x}_i 在训练过程中的贡献度有所不同,因此 FSVR 将模糊参数引入模型中,以消除歧义点对训练过程的影响,进而解决由其导致的过学习问题,即存在 1 个模糊隶属度 s_i 与任意数据点 \mathbf{x}_i 有联系,则最终产生的训练集含有模

糊成员。

在 FSVR 构建时,设训练集 $S = \{(\mathbf{x}_1, y_1, s_1), (\mathbf{x}_2, y_2, s_2), \dots, (\mathbf{x}_i, y_i, s_i), \dots, (\mathbf{x}_N, y_N, s_N)\}$, 其中: $\mathbf{x}_i \in \mathbf{R}^n$ (\mathbf{R}^n 为 n 维实数集), $y_i \in \mathbf{R}$ (y_i 为真实值), $s_i \in [0, 1]$, $i = 1, 2, \dots, N$ 。在时间序列问题中, s_i 是时间序列 t_i ($i = 1, 2, \dots, N$) 的函数,模糊隶属度函数 $l(t_i)$ 是 t_i 的二次函数,即:

$$s_i = l(t_i) \quad (1)$$

$$s_i = (1 - \lambda) \left(\frac{t_i - t_1}{t_N - t_1} \right)^2 + \lambda \quad (2)$$

式中:

λ —— $(0, 1]$ 区间内的实数。

s_i 的边界条件为:

$$\begin{cases} s_1 = \lambda, & 0 < \lambda \leq 1 \\ s_N = l(t_N) = 1 \end{cases} \quad (3)$$

进而,求解二次规划问题:

$$\arg \min_{\omega, b, \xi, \xi^*} \phi_L(\omega, \xi_i, \xi_i^*) = \frac{1}{2} \omega^T \omega + C \sum_{i=1}^N s_i (\xi_i + \xi_i^*) \quad (4)$$

$$\text{s. t. } \begin{cases} y_i - \omega^T \varphi(\mathbf{x}_i) - b \leq \varepsilon + \xi_i \\ \omega^T \varphi(\mathbf{x}_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0 \end{cases} \quad (5)$$

式中:

ω ——回归超平面权值向量;

b ——偏差系数;

C ——惩罚系数;

ε ——回归超平面带宽;

$\phi_L(\cdot)$ ——拉格朗日目标函数;

ξ_i, ξ_i^* ——松弛变量。

式(4)的对偶形式为:

$$\arg \max_{\alpha, \alpha^*} \phi_L(\alpha_i, \alpha_i^*, \alpha_j, \alpha_j^*) = -\frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(\mathbf{x}_i, \mathbf{x}_j) - \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \quad (6)$$

$$\text{s. t. } \begin{cases} \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, s_i C] \end{cases} \quad (7)$$

式中:

$\alpha_i, \alpha_i^*, \alpha_j, \alpha_j^*$ ——拉格朗日乘子;

$K(\mathbf{x}_i, \mathbf{x}_j)$ ——核函数。

求解式(5),得到 FSVR 函数 $f(\mathbf{x})$:

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (8)$$

2 PSO-FSVR 模型概述

2.1 PSO 算法

启发式算法在工程应用中越来越流行,且由于其易于实现和不需要梯度信息的优点,经常被用于处理优化问题。PSO 算法具有快速、高效的特点^[5],因此在启发式算法中脱颖而出。该理论由 Eberhart 等人提出,它是一种模仿飞禽寻觅食物的 PSO 算法,可用于对预测模型的参数进行优化。根据该理论,每只飞禽都可以被看作 1 个粒子,则空间中的所有粒子就会组成 1 个飞禽群体,任意粒子都拥有自身的位置和速度元素,粒子群中每个粒子的信息都是不断迭代和更新的,迭代中获得的粒子局部最佳位置 $\mathbf{x}_{\text{best},i}$ 和全局最佳位置 \mathbf{x}_{best} ,就是潜在的理想解。

$$\begin{cases} \mathbf{x}_i(k+1) = \mathbf{x}_i(k) + \mathbf{v}_i(k+1) \\ \mathbf{v}_i(k+1) = \tau \mathbf{v}_i(k) + c_1 r_1 [\mathbf{x}_{\text{best},i}(k) - \mathbf{x}_i(k)] + c_2 r_2 [\mathbf{x}_{\text{best}}(k) - \mathbf{x}_i(k)] \\ r_1, r_2 \in [0, 1] \end{cases} \quad (9)$$

式中:

$\mathbf{v}_i(k)$ ——经过 k 次迭代后粒子 i 的速度;

$\mathbf{x}_i(k)$ ——经过 k 次迭代后粒子 i 的位置;

τ ——惯性权重;

r_1, r_2 —— $[0, 1]$ 区间内的随机数;

c_1, c_2 ——学习因子。

式(7)中, $\mathbf{v}_i(k)$ 和 $\mathbf{x}_i(k)$ 会不断进行改变。

2.2 PSO-FSVR 模型

选择不同的核函数,可以生成不同的 FSVR。

选取多项式和 Gauss 径向基两种核函数进行对比,就效果而言, Gauss 径向基核函数要优于多项式核函数。此外,在 FSVR 模型中需要确认 C 和 ε 。该模型参数的优化步骤如下:

1) 随机初始化粒子群,并对整个群体的 $\mathbf{x}_i(k)$ 和 $\mathbf{v}_i(k)$ 进行随机化。记录种群的全局最佳位置 \mathbf{x}_{best} 和第 i 个粒子的局部最佳位置 $\mathbf{x}_{\text{best},i}$, 其中 $i = 1, 2, \dots, N$ 。

2) 对于每次迭代,评估第 i 个粒子的适应度值。如果第 i 个粒子的适应度值大于记录值,则设置 $\mathbf{x}_{\text{best},i}$ 为迄今为止的个体最佳位置;如果群体适应度值大于群体记录值,则设定 \mathbf{x}_{best} 为迄今为止的群体最佳位置。

3) 计算惯性权重,并更新粒子的速度和位置。

4) 重复步骤 2)、3),直到达到迭代最大值。

5) 根据最佳粒子,计算 FSVR 模型所需的人参值。

3 基于 PSO-FSVR 的城轨客流预测模型数值分析

3.1 客流量预测值与原始值对比分析

以北京地铁 2 号线(以下简称“2 号线”)某车站 2023 年 6 月 7 日至 2023 年 6 月 13 日的进出站客流数据作为试验对象,以 15 min 为粒度对数据进行统计。由于地铁客流具有明显周期性,因此分别对工作日(周一—周五)及节假日(周六—周日)的进站和出站客流进行预测。经过数据清洗,筛选出 1 050 条数据进行数值试验,其中:训练数据 975 条,测试数据 75 条。2 号线某车站节假日及工作日进出站客流部分原始数据如图 1 所示。

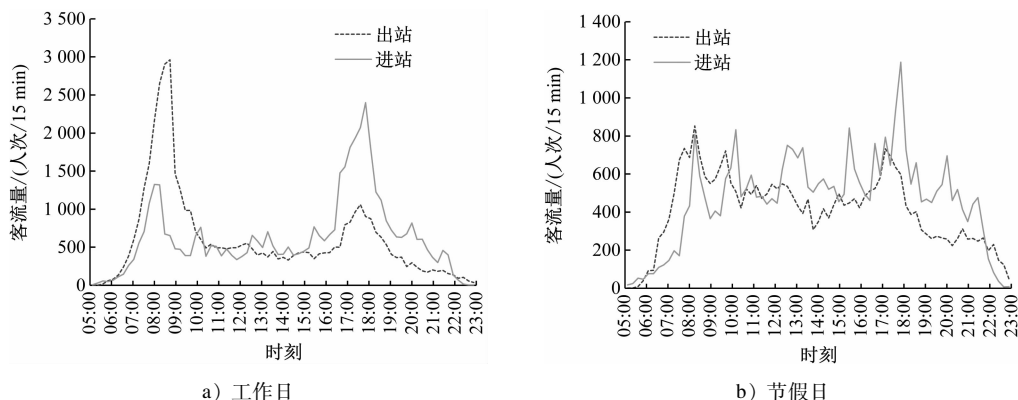


图 1 2 号线某车站进出站客流部分原始数据

Fig. 1 Partial original data of inbound and outbound passenger flows of a station on Line 2

由图 1 可以看出:地铁客流具有明显的周期特征,工作日客流量整体上大于节假日客流量;工作日高峰时段客流量可达 2 500 ~ 3 000 人次/15 min,节假日高峰时段客流量为 1 000 ~ 1 500 人次/15 min,且工作日客流具有明显的双峰值,分别为早、晚高峰时段;而节假日客流波峰不明显,基本呈现

平铺波动趋势。通过观察还可以发现,进站客流与出站客流呈近似对称波动。通过基于 PSO-FSVR 的城轨客流预测模型对 2 号线某车站工作日进出站客流量预测结果见图 2。节假日进出站客流量预测结果见图 3。

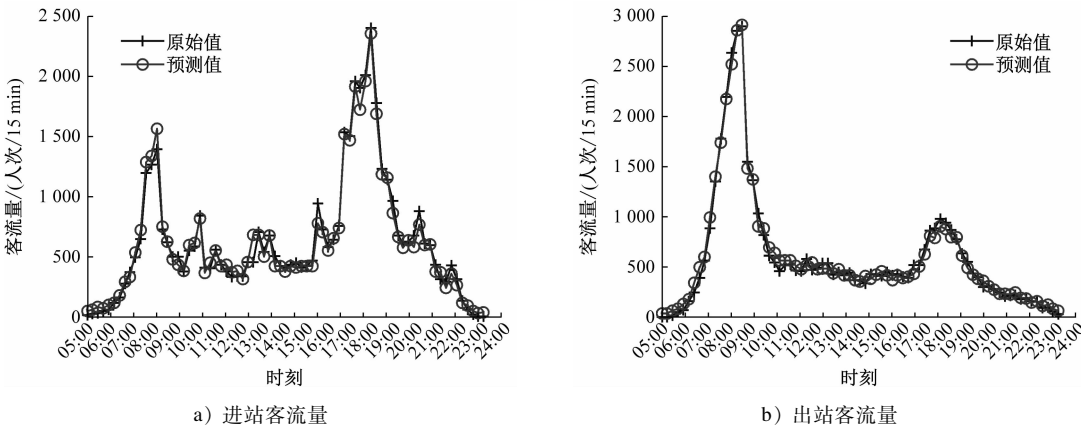


图 2 2 号线某车站工作日进出站客流量预测结果

Fig. 2 Prediction results of inbound and outbound passenger flow on weekdays of a station on Line 2

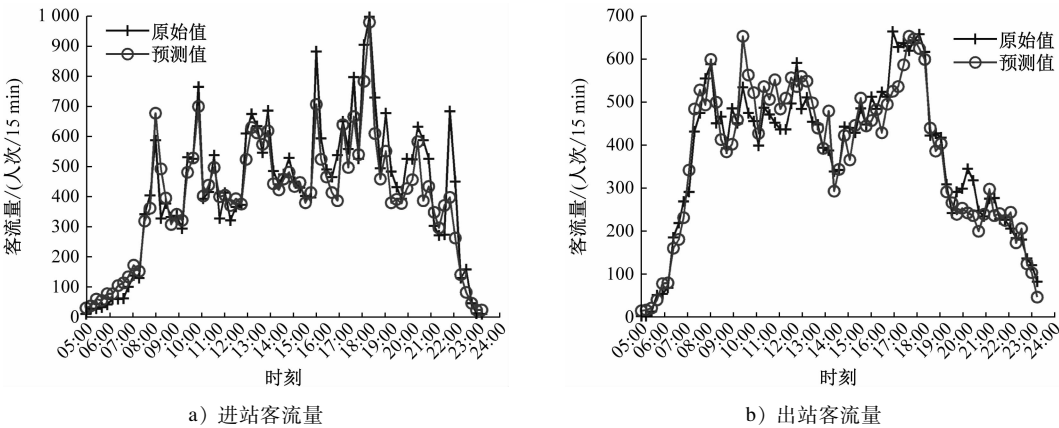


图 3 2 号线某车站节假日进出站客流量预测结果

Fig. 3 Prediction results of inbound and outbound passenger flows on holidays of a station on Line 2

由图 2 可以看出:2 号线某车站工作日进出站客流规律明显,尤其在早晚高峰时段客流突出,其余时段客流波动相对平稳;基于 PSO-FSVR 的城轨客流预测模型在工作日的客流量预测结果与原始值基本吻合,尤其在客流极值点处表现较好,拥有较高的预测准确性。

通过图 3 能够发现:2 号线某车站节假日客流量整体偏低,其峰值客流量明显小于工作日客流量;节假日各时段客流波动相对频繁,客流变化规律性较弱,预测难度相对较高。基于 PSO-FSVR 的城轨客流预测模型的客流量预测结果与原始值基

本一致,且在客流量波动频繁的拐点处依旧能够较好地刻画出其波动规律,这证明基于 PSO-FSVR 的城轨客流预测模型在工作日和节假日均能够准确地对进出站客流量进行预测。

3.2 模型预测性能对比分析

为了验证基于 PSO-FSVR 的城轨客流预测模型的性能,引入常用的 BP 神经网络模型和 GRNN 模型作为对照组进行数值分析,分别对工作日、节假日进站客流量进行预测。其中:BP 神经网络模型由正向和反向两种计算过程组成,该方法在网络理论和性能方面比较成熟,具有非线性映射能力强和

网络结构柔性好等优点,但学习训练速度较慢且容易陷入局部极值,推广能力稍弱。GRNN 是基于径向基函数网络的一种改进模型,学习收敛速度较快,非线性函数逼近能力强,同时克服了 BP 神经网络容易陷入局部极小值的问题,但解释性较差且参数确定困难。使用相同的训练集,利用 3 种方法分别对工作日、节假日进站客流进行预测。不同方法下 2 号线某车站工作日进站客流量预测结果如图 4—图 5 所示。

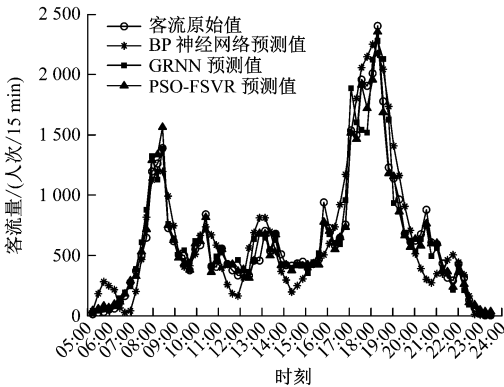


图 4 不同方法下 2 号线某车站工作日进站客流量预测结果
Fig. 4 Prediction results of inbound passenger flow on weekdays of a Line 2 station with different methods

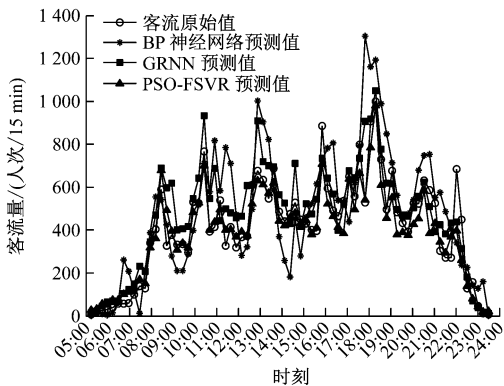


图 5 不同方法下 2 号线某车站节假日进站客流量预测结果
Fig. 5 Prediction results of inbound passenger flow on holidays of a Line 2 station with different methods

通过图 4 可以发现:在工作日进站客流预测中,基于 PSO-FSVR 的城轨客流预测模型的客流量预测效果优于改进的 GRNN 和模型 BP 神经网络模型,且该模型在客流高峰和低谷时段的客流量预测均表现较好;GRNN 模型在单调增长或下降区间内的客流量预测效果较好,但在客流高峰时段的极值点处表现不佳,BP 神经网络模型的客流量预测数据虽然与原始数据相似,但整体准确率较低,尤其在

客流低谷区域客流量预测结果偏小。
通过图 5 可以发现:GRNN 模型整体客流量预测结果偏大;BP 神经网络模型客流量预测趋势波动过大,且预测准确率较低;与两者相比,基于 PSO-FSVR 的城轨客流预测模型在节假日进站客流量预测中表现较为优秀,且该模型的客流量预测趋势更为准确,且预测精度更高。
为了更加直观地对比上述 3 种方法的预测性能,采用 E_{RMS} (均方根误差)、 P 值 (显著性参数) 及 k (相关系数) 作为指标对预测结果进行评价。其中: P 值表示出现比预测结果更极端结果的概率, P 值越小,预测结果越显著; k 表示预测值与原始值间相关性的紧密程度, k 越大,两者间的关系越紧密; E_{RMS} 表示预测值和原始值间的偏差与观测次数之比的平方根,用于体现两者间的误差, E_{RMS} 越小,预测精度越高。2 号线某车站工作日客流量预测模型性能对比结果如表 1 和表 2 所示。

表 1 2 号线某车站工作日客流量预测模型性能对比

Tab. 1 Comparison of passenger flow prediction model performance for weekdays of a Line 2 station

模型	E_{RMS}	P	k
PSO-FSVR 模型	7.010 4	0	0.993 0
GRNN 模型	14.768 1	0	0.968 3
BP 神经网络模型	23.206 8	0	0.934 2

表 2 2 号线某车站节假日客流量预测模型性能对比

Tab. 2 Comparison of passenger flow prediction model performance for holidays of a Line 2 station

模型	E_{RMS}	P	k
PSO-FSVR 模型	8.778 5	0	0.955 8
GRNN 模型	13.491 1	0	0.908 8
BP 神经网络模型	22.125 4	0	0.795 2

从表 1 和表 2 可以看出:基于 PSO-FSVR 的城轨客流预测模型下,2 号线某车站工作日、节假日进站客流量的 E_{RMS} 分别为 7.010 4 和 8.778 5,且在 3 种方法中最小,代表该方法预测精度最高;基于 PSO-FSVR 的城轨客流预测模型下,工作日、节假日进站客流量的相关系数分别达到 0.993 0 和 0.955 8,高于另外两种预测方法,表示 PSO 算法-FSVR 的客流量预测结果与原始数据的趋势更加一致;在显著性上,一般认为 $P < 0.05$ 即为显著,3 种方法下客流量预测结果的 P 均远小于 0.05,证明 3 种客流量预测模型的结果均显著。由此可见,在 2

号线某车站工作日、节假日的进站客流量预测中,基于 PSO-FSVR 的城轨客流预测模型的客流量预测性能均表现得更加优异。

4 结语

通过本文分析可知,基于 PSO-FSVR 的城轨客流预测模型能够对城轨进出站客流进行精准预测,且性能良好。常用的 GRNN 模型和 PB 神经网络模型虽然能够预测客流量,但精度和趋势一致性均有不足;而基于 PSO-FSVR 的城轨客流预测模型整体预测效果稳定,且在客流极值点处的预测结果更加准确。基于 PSO-FSVR 的城轨客流预测模型有效地处理了客流数据的非线性情况,为城轨客流预测提供了创新的方法,并且可在类似情况下推广应用。未来将对客流量预测方法的成本和复杂性做进一步深入分析,以期提高城轨客流量预测的效率,同时降低能耗,进一步向智能化、绿色化目标发展。

参考文献

[1] 张仪鹏,周玮腾,韩宝明. 基于支持向量机的地铁故障类型预测[J]. 北京交通大学学报, 2023, 47(1): 90.

ZHANG Yipeng, ZHOU Weiteng, HAN Baoming. Subway fault type prediction based on support vector machine[J]. Journal of Beijing Jiaotong University, 2023, 47(1): 90.

[2] 梁粤华,翟利华,刘旭,等. 基于 BP 神经网络的隧道下穿既有运营地铁线注浆施工参数预测:以广州地铁 18 号线隧道注浆工程为例[J]. 科学技术与工程, 2023, 23(7): 3045.

LIANG Yuehua, ZHAI Lihua, LIU Xu, et al. Prediction of grou-

ting construction parameters of tunnel undercrossing existing subway lines based on BP neural network[J]. Science Technology and Engineering, 2023, 23(7): 3045.

[3] 申雷霄,陆宇航,郭建华. 卡尔曼滤波短时交通流预测普通国道适应性研究[J]. 交通信息与安全, 2021, 39(5):117.

SHEN Leixiao, LU Yuhang, GUO Jianhua. Adaptability of Kalman filter for short-time traffic flow forecasting on national and provincial highways[J]. Journal of Transport Information and Safety, 2021, 39(5):117.

[4] 杨帮华,李昕,杨磊. 模式识别技术及其应用[M]. 北京: 科学出版社, 2016.

YANG Banghua, LI Xin, YANG Lei. Pattern recognition technology and its application[M]. Beijing: Science Press, 2016.

[5] 张爱军,方陆文,胡后征. 基于速度优化与粒子群优化算法的插电式燃料电池有轨电车能量管理策略[J]. 城市轨道交通研究, 2022, 25(10):249.

ZHANG Aijun, FANG Luwen, HU Houzheng. Energy management strategy of plug-in fuel cell tram based on speed optimization and PSO algorithm[J]. Urban Mass Transit, 2022, 25(10):249.

[6] 陈东洋,陈德旺,陈开河,等. 基于客流大数据分析和支持向量回归的地铁乘客出行时间预测研究[J]. 现代城市轨道交通, 2020(9): 70.

CHEN Dongyang, CHEN Dewang, CHEN Kaihe, et al. Study on subway passenger travel time prediction based on passenger flow big data analysis and support vector regression[J]. Modern Urban Transit, 2020(9): 70.

[7] 苗沁,刘慧婷. 城市轨道交通站点数据链客流预测方法研究[J]. 现代城市轨道交通, 2021(3): 64.

MIAO Qin, LIU Huiting. Research on passenger flow forecast method of urban rail transit station and node data link[J]. Modern Urban Transit, 2021(3): 64.

(收稿日期:2023-06-26)

城市轨道交通综合检测列车在北京地铁线网实现工程应用

近年来,中国铁道科学研究院集团有限公司(以下简称“铁科院”)聚焦城市轨道交通“绿色·融合”发展,积极发挥铁科院城市轨道交通中心的平台和窗口作用,不断推进铁路综合检测技术向城市轨道交通转移。

2016 年,经国家发展和改革委员会批复,铁科院承担了城市轨道交通综合检验试验项目的科研攻关,依托该项目建成世界首列 3B 型城市轨道交通综合检测列车。继完成北京地铁 19 号线综合检测服务之后,近期又完成了北京地铁 6 号线综合检测服务工作,标志着应用城市轨道交通综合检测列车开展更大范围的线网级应用已成为可能。对北京地铁 6 号线的检测内容涵盖:轨道动态几何、钢轨廓形、车辆动力学响应、钢轨波磨、隧道三维结构变形、接触网动态几何及弓网相互作用关系、轮轨和车内噪声检测,以及轨道状态、隧道衬砌状态、接触网悬挂状态巡检,并针对车辆异常晃动、车内噪声超标等多专业耦合的系统问题开展了详细分析。

铁科院 3B 型城市轨道交通综合检测列车在北京地铁 6 号线的工程示范应用,突破了既有专业界限,实现了“全线、全断面”健康状态的同步一体感知,显著提高了检测及数据分析效率与智能化水平,辅助运营公司形成了 6 号线运营期的高质量基础设施运行状态数据资产,为北京地铁高质量基础设施检测维护提供了全新的解决方案。

(供稿:铁科院(北京)工程咨询有限公司)