

doi: 10.7690/bgzd.2025.10.010

基于改进 MD 模型的高速铁路短期客运量预测方法

张 怡, 史 歌

(西安铁路职业技术学院交通运输学院, 西安 710026)

摘要: 针对铁路的客运量数据规模较大, 具有较大的随机性和不稳定性, 难以保证短期客运量预测精度的问题, 提出一种基于 GIS 地理信息技术的城市土地利用动态数据整合方法。选取影响高铁出行的个性化需求因素, 以修正因子对历史数据扩展, 选择时间序列法提取有效结构成分。在原有 MD 模型中融入舒适度指标, 设置改进 MD 模型的预测流程, 通过影响出行牺牲量的因素, 设计改进后的出行牺牲量函数, 实现高速铁路短期客运量预测。实验结果表明: 对于不同城市的高铁短期客运量, 该方法的预测精度可达 99.5%, 具有一定应用价值。

关键词: 预测; 短期客运量; 高速铁路; 改进 MD 模型

中图分类号: TP301.6 **文献标志码:** A

Short-term Passenger Volume Forecasting Method of High-speed Railway Based on Improved MD Model

Zhang Yi, Shi Ge

(School of Transportation, Xi'an Railway Vocational and Technical Institute, Xi'an 710026, China)

Abstract: In view of the large scale, randomness and instability of railway passenger volume data, and the difficulty to ensure the accuracy of short-term passenger volume forecast, a dynamic data integration method of urban land use based on GIS geographic information technology was proposed. The individual demand factors affecting high-speed rail travel are selected, the historical data are expanded by the correction factor, and the effective structural components are extracted by the time series method. The comfort index is integrated into the original MD model, and the prediction process of the improved MD model is set up. Through the factors affecting the travel sacrifice, the improved travel sacrifice function is designed to realize the short-term passenger volume prediction of high-speed railway. The experimental results show that the prediction accuracy of the method can reach 99.5% for the short-term passenger volume of high-speed rail in different cities, which has certain application value.

Keywords: forecast; short-term passenger volume; high-speed railway; improved MD model

0 引言

近年来, 随着我国科学技术与经济飞速发展, 各客运方式不断提高服务水平, 引导居民选择适宜的方式进行出行。科学客运量预测对不同区域内的客运方式规划产生巨大影响。随着高速铁路客运的成熟发展, 客运行业的发展重心发生转移, 对航空和公路客运都造成一定影响, 因此对高速铁路的短期客运量进行有效预测非常重要。杨飞等^[1]提出基于 Newton 插值与超松弛技术对铁路客运量进行预测, 以解决因节假日造成的数据异常问题。该方法将 Newton 插值法与神经网络相结合, 提出一个组合预测模型, 并以北京市的铁路客运作为测试案例, 具有一定寻优能力, 可较快完成客运量预测。曹武军等^[2]选择 SSA-GM-GPR 技术对铁路的客运量进行分析, 为解决数据的非稳定性问题, 对大量数据

的噪声进行剔除, 并在高斯回归模型中对数据进行整理, 最终达到预测的效果。该方法对数据的整理具有较高水平, 但由于回归预测和集成分析需要大量测试准备, 难以在短时间内获取精准结果, 虽然具有一定精度, 但不能保证时效性。综合上述传统方法的应用效果, 其主要是应用在铁路和航空客运中, 对于高速铁路的客运量预测没有统计。笔者以此为基础, 为帮助高铁的组织性管理提供有力依据, 选择改进 MD 模型作为技术支持, 针对于高速铁路的客运量进行预测。

1 选取高速铁路出行的个性化需求因素

对高速铁路短期客运量的预测, 需要充分考虑出行者的需求, 筛选出影响出行者选择高速铁路的魅力因素, 其中包括正向因素和反向因素 2 种形式。此次以竞争策略为研究方式, 分别对高铁负责人、

收稿日期: 2024-10-26; 修回日期: 2024-11-28

基金项目: 陕西省“十四五”教育科学规划 2022 年度课题(SGH22Y1646)

第一作者: 张 怡(1981—), 女, 河北人, 硕士。

出行人以及管理人员进行因素标定, 针对具体的因素进行选择, 具体情况如表 1 所示^[3]。

表 1 高速铁路出行需求

类别	因素
必须需求	舒适
	安全
期望需求	快捷
	卫生好
	乘车环境好
	购票方便
	天气影响小
	乘车班次多
魅力需求	容易换乘
	出行时间方便
	乘客少
无差异需求	个人偏好
	票价便宜

根据表 1 内容, 依据筛选原则, 此次只保留期望需求和魅力需求 2 个因素, 剔除无差异需求和必须需求 2 个因素。将剩余的影响因素进行规范化处理, 建立决策矩阵, 对出行者的个性化需求进行规范化处理:

$$x_{cv} = z_{cv} / \sum_{c=1}^b z_{cv}; \quad (1)$$

$$m = (x_{cv})_{b \times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1v} \\ x_{21} & x_{22} & \cdots & x_{2v} \\ \vdots & \vdots & & \vdots \\ x_{c1} & x_{c2} & \cdots & x_{cv} \end{bmatrix}_{b \times n}。 \quad (2)$$

式中: z_{cv} 为第 c 个出行者对第 v 个需求项的评价分数; $c=1, 2, \dots, b$; b 为选取的出行者总数量; $v=1, 2, \dots, n$; n 为选择的个性化需求项总数, 此次 $n=11$;

x_{cv} 为规范化后的分数^[4]; m 为规范化的决策矩阵。将规范化的个性化需求项作为出行者选择高铁的依据, 选择历史客运量数据进行分析, 提取数据中的有效结构成分, 以实现预测。

2 高铁历史客运量数据结构成分提取

此次高铁历史数据是在特定日期以及特定时间段内进行选取的。为保证数据具有普遍性, 需要对其进行扩样分析, 在修正完毕后提出有效结构成分, 修正公式如下:

$$W_{ERT} = Y_{ERT} \times (W_T / Y_T)。 \quad (3)$$

式中: T 为高铁车次^[5]; W_{ERT} 为 E 区域到 R 区域的年平均日分布客运量; W_{ERT} 为 E 区域到 R 区域实际调查的客运量; W_T 为年平均日客运量; Y_T 为样本客运量。根据式(1), 车次的样本修正系数计算公

式为:

$$\chi_T = W_T / Y_T。 \quad (4)$$

在对样本数据修正后, 直接对未来数据的预测是具有较高难度的, 但未来数据中的随机分量可能会与过去某个时间段的随机分量相似, 以此采用时间序列方式对样本数据进行分解, 可增加预测准确的可能性^[6]。

时间序列要求具有平稳性, 考虑到高铁客运量历史数据的随机性, 从样本数据中选择大小相同的多个数据进行拼接, 形成新的序列组合形式。通过确定数据的序列的长度, 对结构成分进行提取, 矩阵为:

$$\begin{bmatrix} I_1 \\ \vdots \\ I_A \\ \vdots \\ I_{P-O+1} \end{bmatrix} = \begin{bmatrix} S_1 & \cdots & S_O \\ S_2 & \cdots & S_{O+1} \\ \vdots & & \vdots \\ S_A & \cdots & S_{P+(O-1)} \\ \vdots & & \vdots \\ S_{P-(O-1)} & \cdots & S_P \end{bmatrix}_{(P-O+1) \times O}。 \quad (5)$$

式中: $S = \{S_1, S_2, \dots, S_P\}$ 为时间序列长度为 P 的历史客运量数据集^[7]; I 为数据的结构, 生成 $P-O+1$ 个数据结构。从样本集合中随机选择数据进行连接, 能够筛选出所需的结构成分 I_{P-O+1} 。通过对数据的修正以及结构的提取, 以改进 MD 模型对高铁客运量进行预测。

3 舒适性指标改进 MD 模型预测短期客运量

MD 模型最先应用于轨道客流预测, 其特征是对出行者的行为心理进行分析, 以此评估出行者的潜在出行率。该模型的研究基础为经济学效应理论, 整体上以出行者个体作为研究载体, 在进行分析时只考虑了时间和费用 2 个因素, 在现阶段的研究中较为单一。笔者为提高预测精度, 融合舒适性指标对 MD 进行改进^[8]。整个模型结构不变, 仍以出行者的行为心理出发, 立足于供给与需求的关系, 综合考虑高速铁路的运输方式服务特性对出行者的影响, 则改进的 MD 模型关系如图 1 所示。

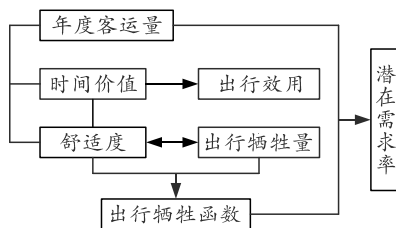


图 1 融合舒适性指标的 MD 模型

根据图 1 内容, 将出行者的时间价值对应出行效应, 将舒适性指标对应出行牺牲量函数中, 以此为基础对影响出行牺牲量的因素进行分析, 以此来完成出行者的潜在客运需求的求取。本次研究默认为, 只有出行者的潜在需求大于出行牺牲量时, 出行者才会出行, 并且出行者在该模型中被认定只会选择牺牲量最小的交通方式^[9]。基于此, 对涵盖的多个指标因素进行计算:

$$p_y = \ln(a_y + s \times d_y); \quad (6)$$

$$t_y = (u_y \cdot p_y) / i_y. \quad (7)$$

式中: y 为高速铁路出行方式; p_y 为 y 的出行牺牲量; a_y 为出行费用; s 为时间价值, 表示以单位时间换算的单位货币系数; d_y 为出行时间; u_y 为高速铁路的客运需求量, 表示从出发点到目的地的实际社会个体总人数; i_y 为潜在的客运需求量; t_y 为潜在需求量显化率。

在出行过程中, 出行者的心理行为会受到多种因素影响, 其中身体疲劳为出行必须付出的代价之一, 因此在预测模型中需要增加该指标, 衡量影响出行牺牲量的重要因素。将舒适度指标作为评价基础, 其中包括服务质量、运行平稳性等多个因素^[10]。但由于上述参数不容易被量化, 将舒适度转化为出行疲劳度进行分析。以此为基础, 将原有 MD 模型中的出行时间进行量化, 表示为:

$$f_y = h_y(g) \times s_{f_y}. \quad (8)$$

式中: g 为乘车时间^[11]; s_{f_y} 为恢复疲劳的时间价值; f_y 为高铁出行的疲劳度; $h_y(g)$ 为疲劳恢复的时间, 与出行时间的长短存在一定关系, 因此可将疲劳恢复时间归属与出行时间, 具体计算方式如下:

$$h_y(g) = \frac{j d_{\max}}{(\beta_y \times \exp(-\alpha_y \times g) + 1)}. \quad (9)$$

式中: $j d_{\max}$ 为疲劳恢复需要的最长时间^[12]; β_y 为选择的客运方式出行的应用系数, 此次为高速铁路, 取值为 $\beta_y = 50$ ^[13]; α_y 为单位出行时间 $h_y(g)$ 的强度系数, 单位为 h^{-1} , 当 $\alpha_y > 0$ 时表示 $h_y(g)$ 越长。基于此, 建立改进后的 MD 模型预测函数:

$$\ln p_y = \ln(a_y + s \times d_y + k_y + l_y). \quad (10)$$

式中: k_y 为关于时间价值 s 的表达式; l_y 为关于时间 g 的表达式。综上, 通过对出行牺牲量的分析和计算, 能够直接对出行者的心理进行分析, 直接推导出出行者的客运方式选择结果。而直接代入高速铁路的出现方式, 则可直接获取出行者选择该方式意愿值, 实现客运量的预测。

4 实验测试分析

笔者通过改进 MD 模型设计了新的方法, 为验证该方法能够实现高速铁路短期客运量的有效预测, 以对比测试的方式完成论证。采用 Newton 插值法的预测方法和 SSA-GM-GPR 技术的预测方法作为对照测试组, 分别与笔者所提方法进行对比, 验证不同方法的应用价值。

4.1 选取测试数据

为验证不同方法对高速铁路短期客运量的预测具有应用价值, 本次选择精准度作为评价指标, 直接以未来“春节”假后前 4 天的客运量作为测试对象, 分别选择不同地区的历史数据作为测试集合, 以此分析各地区高速铁路的客运量的差异性。从数据库中随机调取 5 组城市的历史数据, 具体如表 2 所示。

表 2 不同地区“春节”假期客运量情况 万人

年份	天数	城市				
		A	Q	Z	W	S
2019	假后第 1 天	2.1	0.5	0.8	1.2	1.4
	假后第 2 天	2.8	0.4	0.7	1.2	1.3
	假后第 3 天	2.6	0.4	0.6	1.2	1.2
	假后第 4 天	2.6	0.4	0.6	1.2	1.2
2020	假后第 1 天	2.2	0.4	1.5	1.6	1.8
	假后第 2 天	2.8	0.4	1.2	1.6	1.6
	假后第 3 天	2.6	0.5	1.2	1.5	1.4
	假后第 4 天	2.6	0.5	1.2	1.5	1.4
2021	假后第 1 天	2.2	0.4	1.4	1.8	1.8
	假后第 2 天	2.0	0.4	1.4	1.8	1.6
	假后第 3 天	2.8	0.5	1.2	1.6	1.5
	假后第 4 天	2.8	0.3	1.2	1.6	1.6
2022	假后第 1 天	2.2	0.3	1.6	1.9	1.8
	假后第 2 天	2.0	0.5	1.6	1.8	1.8
	假后第 3 天	2.0	0.4	1.6	1.8	1.9
	假后第 4 天	2.8	0.4	1.6	1.8	1.9

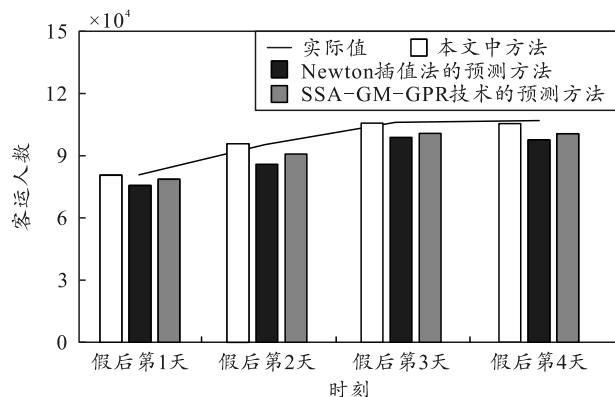
根据表 2 可知, 此次选择近 4 年的数据作为测试集合, 分别对不同城市的“春节”假期的高铁客运量进行统计, 其中 A 城市的客运量较大, 一直保持在最高水平; Q 城市的客运量相对较少; Z 城市的客运量居中, 但变动幅度较大; W 城市与 S 城市的客运量变动较小, 客运量相比较 A 城市较小, 但高于 Z 城市。不同城市的客运量具有差异性, 且变动情况各不相同, 符合此次测试要求。

由于此次是对短期客运量的预测, 为保证测试的准确性和真实性, 把近 4 年的数据统计为 4 组数据, 将统计后的结果作为此次预测样本, 将统计后的数据结果作为各城市“春节”假期可能达到的客运量, 其中 A 城市的客运量仍为最大值。将上述样

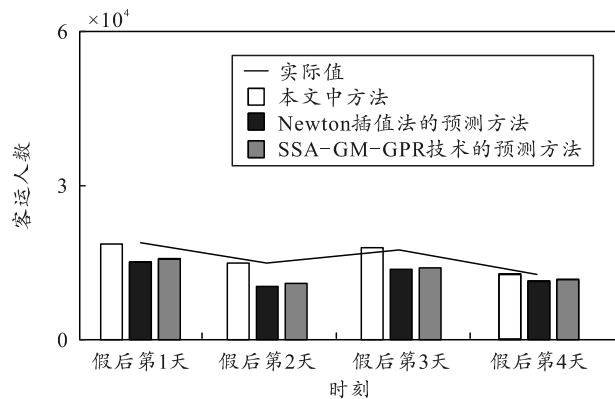
本上传至 Matlab 测试平台,分别连接 4 组预测方法,分别依据上述数据对 2023 年春运期间前 4 天的客运量进行预测,并对结果进行分析。

4.2 预测精度测试结果及分析

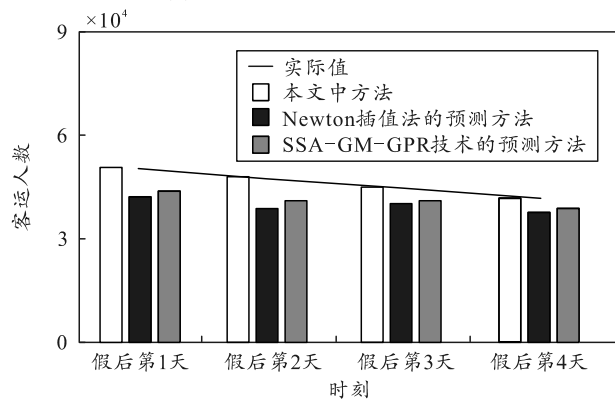
将预测精度作为测试的评价指标,只有具有较高的预测精度,才能对高速铁路的车次进行及时调整和有效安排。除此以外,对客运量的预测也能够为所在区域的城市发展规划提供依据,不同城市的营销政策会影响高铁短期客运量。为减少多种因素对客运量的影响,此次不考虑城市的营销政策,直接以设计的样本数据作为预测基础,完成各组方法的预测精度测试,具体结果如图 2 所示。



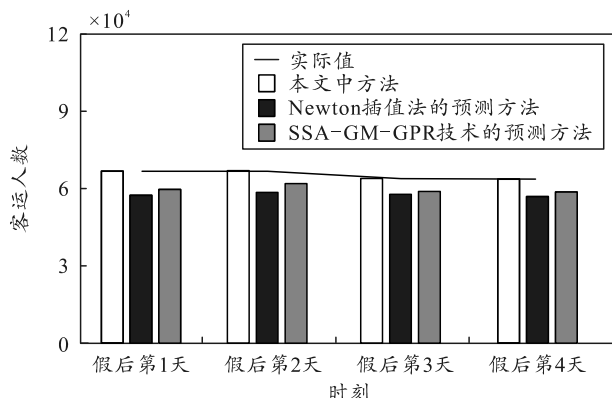
(a) A 城市客运量预测结果



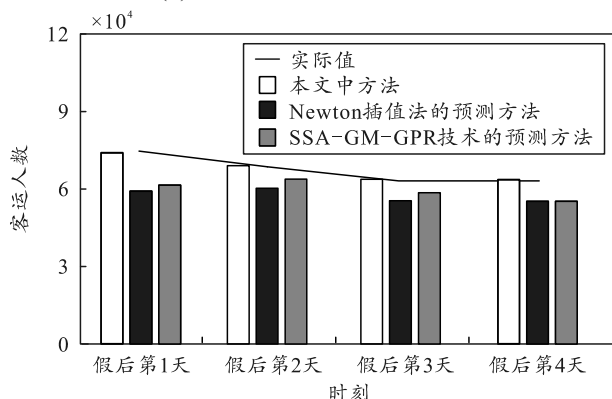
(b) Q 城市客运量预测结果



(c) Z 城市客运量预测结果



(d) W 城市客运量预测结果



(e) S 城市客运量预测结果

图 2 不同方法的预测结果

根据图 2 中内容所示,在笔者所提方法应用下可对样本数据进行精准预测,与实际数据基本一致,而 3 组传统方法等预测结果各不相同,其预测结果与实际数据相差较大。对其精度进行具体计算,计算公式为:

$$q = \frac{w}{e} \times 100\% \quad (11)$$

式中: q 为预测精度; w 为预测的样本数据结果; e 为实际的样本数据量。将每组方法预测数据代入式 (11),本文中方法精度最大为 99.5%,最小为 87%。综合结果说明:本文中方法的预测精度更高,能够实现对高速铁路短期的客运量预测,具有更高的应用价值。

5 结束语

实验结果表明:与传统方法相比,本文中方法具有更高的预测精度,可以投入现实中进行应用,为不同城市的规划提供理论依据。由于预测过程受到多种因素影响,在设计过程中存在一些不足之处,如未考虑不同地区的政策变动,没有进行全面地分析,具有一定局限性。后续研究中,会对上述问题进行分析,结合不同因素对客运量的影响,提

出更加全面地预测，为其在实际中的应用提供理论支持。

参考文献：

- [1] 杨飞, 贾俊芳, 刘岩岩, 等. 基于 Newton 插值与超松弛技术的铁路客运量预测研究[J]. 铁道运输与经济, 2023, 45(3): 44-52.
- [2] 曹武军, 李中强. 基于 SSA-GM-GPR 的铁路客运量预测研究[J]. 现代电子技术, 2022, 45(14): 159-163.
- [3] 陈聪聪, 李程. 基于 PSO-ELM 算法的民航客运量预测[J]. 控制工程, 2022, 29(11): 2042-2047.
- [4] 杨飞, 王洋. 基于改进自然选择粒子群算法的铁路客运量组合预测研究[J]. 高速铁路技术, 2022, 13(6): 63-68, 73.
- [5] 徐学明, 王仕杰, 陈艳艳, 等. 基于 MD 模型公路建设项目诱增交通量方法研究[J]. 公路, 2019, 64(5): 200-203.
- [6] 刘尚东, 孙东永, 许晶晶, 等. 基于 ESMD-NNBR 耦合的月径流预测模型[J]. 水文, 2023, 43(1): 52-56.
- *****
- (上接第 15 页)
- [5] 单雨丝, 陈波, 程朋飞. 基于快速特征点提取和描述算法与色调、饱和度和明度的图像特征点匹配算法[J]. 激光与光电子学进展, 2022, 59(12): 153-159.
- [6] 刘自强, 周铁军, 傅冬和, 等. 基于颜色和形状的鲜茶叶图像特征提取及在茶树品种识别中的应用[J]. 江苏农业科学, 2021, 49(12): 168-172.
- [7] 杨燕, 武旭栋, 杜康. 基于小波变换及注意力机制的 T 型图像去雾网络[J]. 湖南大学学报(自然科学版), 2022, 49(10): 61-68.
- [8] 江泽涛, 覃露露, 秦嘉奇, 等. 一种基于 MDARNet 的低照度图像增强方法[J]. 软件学报, 2021, 32(12): 3977-3991.
- [9] 刘云, 董守杰. 基于 CUDA 核函数的多路视频图像拼接加速算法[J]. 计算机科学, 2022, 49(S1): 441-446.
- [10] 杨棚, 朱大明, 杨润书, 等. 匹配点云结合 HSI 色彩分量的无人机 RGB 影像信息提取方法[J]. 农业工程学报, 2021, 37(22): 295-301.
- [11] 支俊俊, 董娅, 鲁李灿, 等. 基于无人机 RGB 影像的玉米种植信息高精度提取方法[J]. 农业工程学报, 2021, 37(18): 48-54.
- [12] 纪玉华, 洪婉敏, 谢明晏, 等. 基于 UPLC 特征图谱及色彩图像技术的当归与酒当归差异比较研究[J]. 中药材, 2022, 45(2): 321-326.
- [13] 贾静, 曹竞文, 徐平华, 等. 基于京剧脸谱意象色彩的服饰纹样自动配色[J]. 纺织学报, 2022, 43(12): 160-166.
- [14] 王金星, 刘雪梅, 刘双喜, 等. 基于组合色彩特征的苹果树叶片各生长期氮含量预测[J]. 农业机械学报, 2021, 52(10): 272-281, 376.
- [15] 李世镇, 钱俊, 余映, 等. 基于凸包计算和颜色特征的显著性检测算法[J]. 云南大学学报(自然科学版), 2021, 43(2): 254-262.