

基于月客流量残差的地铁客流季节指数预测方法

鲁文博 马超群 李国栋 曹蕊 徐金华

(长安大学交通运输工程学院, 710064, 西安//第一作者, 硕士研究生)

摘要 为精准反映城市轨道交通的季节性特征, 基于月客流残差提出了地铁客流季节区间分类方法。建立趋势方程并计算月客流残差, 以月客流残差曲线之间的相似度为基础对月份进行分类, 得出基于月客流残差曲线的季节分组区间(SIR)。分别以每月进行分类并计算传统月份分组区间(TMC)、以每季度进行分类并计算传统季度分组区间(TQC)以及SIR区间的季节指数并进行预测。预测结果表明, 提出的分类方法在相对误差、平均绝对百分比误差、均方根误差三方面都比其他分类方法更小。这说明所提出的分类方法在地铁客流预测中具有更好的适应性, 从而可以为地铁运力调整以及运营计划的制定提供参考依据。

关键词 地铁; 客流预测; 季节指数; 弗雷歇距离; 人工神经网络

中图分类号 U293.13

DOI:10.16037/j.1007-869x.2021.09.018

Seasonal Index Forecast Method of Metro Passenger Flow Based on Monthly Passenger Flow Residual

LU Wenbo, MA Chaoqun, LI Guodong, CAO Rui, XU Jinhua

Abstract To accurately reflect the seasonal characteristics of urban rail transit, the seasonal interval classification method of metro passenger flow is proposed based on the change of monthly passenger flow residual. The trend equation is established and the monthly passenger flow residual is calculated. The monthly classification is based on the similarity between the monthly residual curves, and the SIR (seasonal grouping interval) based on the residual curve is obtained. TMC (traditional month grouping interval) that is classified by month, TQC (traditional quarter grouping interval) and the seasonal index of SIR interval that are classified and calculated by quarter are predicted respectively. The prediction result shows that the proposed classification method is smaller in relative error, mean absolute percentage error, and root mean square error than other classification methods, indicating that the proposed classification method has better adaptability in the short-term forecast of metro passenger flow, providing reference for the

adjustment of metro transportation capacity and the formulation of operation plan.

Key words metro; passenger flow forecast; seasonal index; Frechet distance; artificial neural network

Author's address College of Transportation Engineering, Chang'an University, 710064, Xi'an China

城市轨道交通客流兼具长期的线性增长趋势、周期性季节变化及随机波动, 呈现复杂的非线性特点。传统的预测方法, 如线性回归模型、ARIMA模型等, 通过调整预测数据以达到对非线性客流量的预测^[1-2]。在季节指数应用方面, 目前主要是对其值的估计^[3-5]和预测模型的改进^[6-8], 缺少对月份分组和季度分组的比较以及更准确的季节分组的研究。

本文基于月客流残差, 以弗雷歇距离作为曲线相似性度量, 达到对季节分组的精确性把握。该季节分组即体现了同一分组内月份的特殊特征, 又突出了不同分组之间的特殊区别, 为确定最优的季节分组提供了新的思路。考虑到季节指数法能够有效地消除季节性因素影响, 而人工神经网络对非线性时间序列具有良好的适应性, 结合两种模型的优势, 对地铁月客流量进行预测, 以验证分类结果的准确性。

1 基于月客流残差的季节分组方法

1.1 总体流程

基于月客流残差的季节分组流程如图1所示。

假设历史数据的时间序列为 $\{Y_t\}$, $t=1, 2, 3, \dots, n$ 。

步骤1: 建立时间序列线性回归趋势方程:

$$Y_t' = \alpha_1 - \alpha_2 t \quad (1)$$

式中:

Y_t' ——因变量;

α_1, α_2 ——回归系数;

t ——时间。

步骤2: 根据趋势方程计算每月客流量的趋势值, 从而计算客流量残差 δ_t , 计算公式为:

$$\delta_t = Y_t - Y'_t \quad (2)$$

式中:

Y_t ——实际客流量,万人次。

步骤3:基于弗雷歇距离算法计算任意两个月客流残差曲线的弗雷歇距离及曲线相似度。

步骤4:根据相似度对月份进行区间分类。

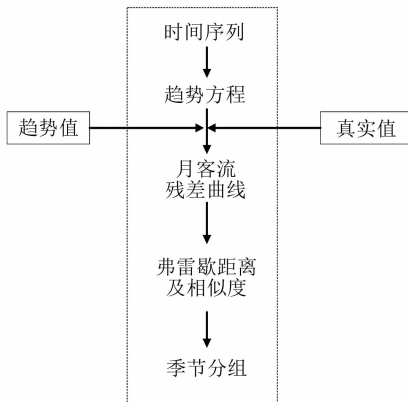


图1 季节分组流程

1.2 弗雷歇距离算法

距离空间的概念是法国数学家弗雷歇于1906年提出的,又称为弗雷歇距离。它将现实空间的距离概念推广到一般的集合上,为抽象空间之间的距离度量提供了理论基础。其着重将路径空间距离考虑进去,相比之下,其对有一定空间时序的曲线相似度评价效率更高。

设 A 和 B 是空间 S 上的两条连续曲线,即 $A:[0,1] \rightarrow S, B:[0,1] \rightarrow S$;又设 α 和 β 是单位区间的两个重新参数化函数,即 $\alpha:[0,1] \rightarrow [0,1], \beta:[0,1] \rightarrow [0,1]$;则曲线 A 与 B 的弗雷歇距离 $F(A,B)$ 定义为:

$$F(A,B) = \inf_{\alpha, \beta} \max_{t \in [0,1]} \{d(A(\alpha(t)), B(\beta(t)))\}$$

式中:

d —— S 上的度量函数。

弗雷歇距离越小,说明两条曲线之间的相似程度越高;弗雷歇距离越大,说明两条曲线之间的相似程度越低。因此对曲线 A 和 B 来说,相似度 $I_{A,B}$ 的定义如下:

$$I_{A,B} = \frac{1}{F(A,B)} \quad (3)$$

2 季节指数预测模型

2.1 季节指数

季节指数是基于统计学的一种简化时间序列

数据的方法,该方法对季节时间序列的季节性进行处理,同时使原始数据中的一些重要信息不会被丢失,一般采用直接平均季节指数法对原始客流数据进行预测前的调整。直接平均季节指数法用算数平均值直接计算各月或各季度的季节指数,将其扩展到季节区间的季节指数,其计算公式为:

$$S_j = \frac{\left(\frac{12}{k}\right) \cdot \sum_{i=1}^m \sum_{p=1}^k x_{ijp}}{\sum_{i=1}^m \sum_{j=1}^n \sum_{p=1}^k x_{ijp}} \quad (4)$$

式中:

S_j ——第 j 分组的季节指数;

x_{ijp} ——第 i 年第 j 分组中第 p 月的全线客流量,万人次。

利用季节指数对客流原始数据进行修正,修正值为 x'_{ijp} ,以降低季节对预测客流的影响,即:

$$x'_{ijp} = x_{ijp} / S_j \quad (5)$$

对原始客流数据处理之后,通过预测得到未来时期某月份的客流量 y'_{ijp} ,经逆向调整后的该月预测客流为:

$$y_{ijp} = y'_{ijp} S_j \quad (6)$$

式中:

y_{ijp} ——第 i 年第 j 分组中第 p 月的客流量,万人次。

2.2 Elman 神经网络

BP(误差反向传播)算法的多层感知器是至今为止应用最广泛的神经网络。在多层感知器的应用中,以单隐层网络的应用最为普遍。一般习惯将单隐层感知器称为三层感知器,所谓三层包括了输入层、隐含层和输出层。

训练的过程就是不断调整权值的过程,当误差小于设置值时停止训练。整个网络的信息都储存在权值矩阵中。当输入一个预测值时,网络根据权值矩阵进行计算,以达到记忆的目的。

Elman 神经网络是在 BP 神经网络基本结构的基础上,在隐含层增加一个承接层,作为一步延时算子,达到记忆的目的,从而使系统具有适应时变特性的能力,增强了网络的全局稳定性。

2.3 预测步骤

根据季节指数对原始客流数据进行调整,分别应用三种不同的季节指数调整后的数据进行预测,最后分别对预测后的数据进行逆向调整,以得到更精确的结果。

主要分为以下步骤:①对异常值进行修正;②根据时间序列数据建立趋势方程,并计算月客流趋势值、残差;③计算任意两月客流残差曲线的弗雷歇距离和相似度;④根据月客流残差曲线间的相似度进行季节区间分类;⑤根据式(4)计算不同分类方法中每个季节区间的季节指数,并对原始客流数据按式(5)进行调整,获得调整后的客流量;⑥由于Elman神经网络结构的特殊性,需要以原始客流数据对隐含层神经元数目进行寻优,采用优化后的网络预测得到的客流量按式(6)进行逆向调整,从而获得最终的预测客流量。

3 实例分析

选取西安地铁2号线全线2014年1月—2018年8月共56个月的客流量为基础数据,以此数据为原始客流数据,预测2018年9月—2019年2月的客流量。

3.1 修正异常值

由于西安地铁3号线在选取时间区间中开通,因此直接对2016年11月的客流造成影响。采用平均值法计算修正值,并在后续过程中采用修正值进行计算。

3.2 计算月客流残差

应用2014年1月—2017年12月共48个月的全线客流量建立趋势方程并计算月客流残差,结果如表1所示。

月份	不同年份的月客流残差/万人次			
	2014年	2015年	2016年	2017年
1月	-63.666	53.983	-103.383	-319.467
2月	-253.395	-329.226	-418.399	-312.647
3月	42.231	149.980	29.222	-8.728
4月	-31.862	89.981	-6.971	-8.637
5月	12.794	93.637	-69.777	-2.701
6月	16.035	-96.742	-152.933	-142.449
7月	273.466	42.894	-60.172	-35.393
8月	233.287	-37.805	-146.916	0.618
9月	177.408	-51.915	-96.741	-43.406
10月	283.030	16.101	6.823	79.431
11月	180.080	-98.078	-107.719	-15.869
12月	189.046	-54.442	5.951	128.112

注:“-”表示负数。

3.3 弗雷歇距离和相似度计算

根据表1中的数据绘制的不同年份的月客流残差曲线如图2所示。利用弗雷歇距离算法计算任意两条月客流残差曲线之间的相似度,结果如表2所示。

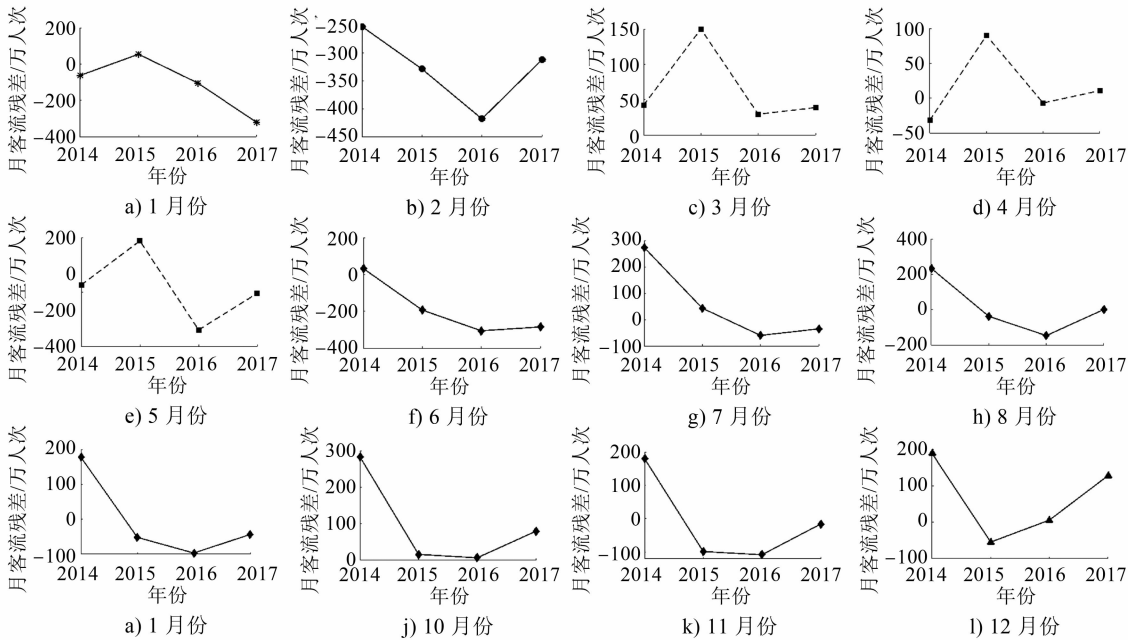


图2 西安地铁2号线2014年1月—2017年12月的月客流残差曲线

3.4 季节分组

对图2的月客流残差曲线进行定性分析,大致可以看到3~5月的客流残差曲线比较相似,6~11月的

客流残差曲线比较相似,12月、1月、2月的客流残差曲线各不相同。根据表2结果,绘制3~5月和6~11月的客流残差相似度曲线,分别如图3和图4所示。

表 2 利用弗雷歇距离算法计算的任意两个月客流残差曲线之间的相似度

月份	月客流残差曲线之间的相似度											
	1	2	3	4	5	6	7	8	9	10	11	12
1	Inf	0.973	1.218	1.170	1.366	1.064	1.044	0.957	0.924	0.834	0.825	0.752
2	0.973	Inf	0.869	0.864	1.402	1.626	1.634	1.793	1.712	1.582	1.665	0.654
3	1.218	0.869	Inf	2.814	1.931	0.897	0.886	0.845	0.814	0.753	0.748	0.658
4	1.170	0.864	2.814	Inf	1.823	0.798	0.792	0.803	0.753	0.710	0.711	0.680
5	1.366	1.402	1.931	1.823	Inf	1.110	1.094	1.152	1.004	0.913	0.918	0.839
6	1.064	1.626	0.897	0.798	1.110	Inf	7.578	3.038	4.656	2.777	2.536	1.245
7	1.044	1.634	0.886	0.792	1.094	7.578	Inf	3.179	5.299	2.996	2.713	1.277
8	0.957	1.793	0.845	0.803	1.152	3.038	3.179	Inf	4.357	3.538	3.807	1.908
9	0.924	1.712	0.814	0.753	1.004	4.656	5.299	4.357	Inf	6.812	5.553	1.588
10	0.834	1.582	0.753	0.710	0.913	2.777	2.996	3.538	6.812	Inf	7.771	1.827
11	0.825	1.665	0.748	0.711	0.918	2.536	2.713	3.807	5.553	7.771	Inf	2.008
12	0.752	0.654	0.658	0.680	0.839	1.245	1.277	1.908	1.588	1.827	2.008	Inf

由图 3 可见,3~5 月的月客流残差曲线相似度比较高,故认为 3~5 月为同一季节区间。

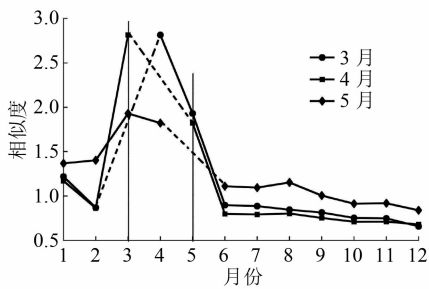


图 3 西安地铁 2 号线 2014 年 1 月—2017 年 12 月中 3~5 月的客流残差曲线相似度

由图 4 中可见,6~11 月客流残差相似度曲线的特征比较相似,故认为 6~11 月为同一季节区间。

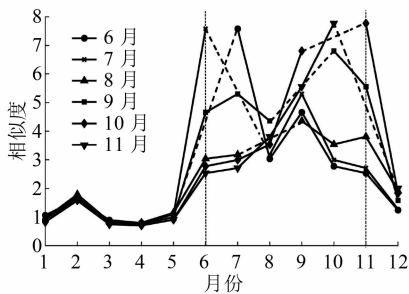


图 4 西安地铁 2 号线 2014 年 1 月—2017 年 12 月中 6~11 月的客流残差曲线相似度

图 5 为 12 月、1 月和 2 月的客流残差相似度曲线。由图 5 可见,12 月、1 月和 2 月的客流残差曲线相似度较小,且与其它月份的并不相同,因此认为 12 月、1 月和 2 月各为一个季节区间。根据上述分析,将地铁客流季节分为 5 个区间,即:3~5 月为第一季度,6~11 月为第二季度,12 月、1 月和 2 月各自为一个季度。

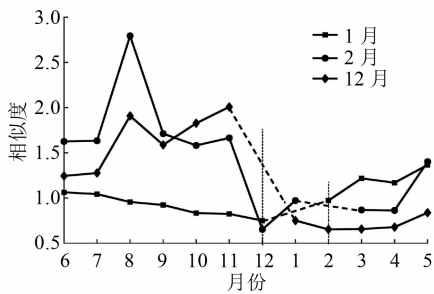


图 5 西安地铁 2 号线 2014 年 1 月—2017 年 12 月中 12 月、1 月和 2 月的客流残差曲线相似度

3.5 季节指数计算

以 2014 年 3 月—2018 年 2 月的客流量为基础数据,分别采用 SIR(季节分组区间)分类方法、TMC(传统月份分组区间)分类方法、TQC(传统季度分组区间)分类方法,按照直接平均法(式(4))进行计算,结果如表 3~5 所示。

表 3 基于 SIR 分类法计算的季节指数

季度	季节指数(s_j)	包含月份
1	0.981 856	3~5 月
2	1.048 788	6~11 月
3	1.130 046	12 月
4	0.865 879	1 月
5	0.765 779	2 月

3.6 优化网络结构

1) 以原始客流数据为基础,优化 Elman 神经网络的隐含层节点数。

(1) 选取 2014 年 3 月—2018 年 2 月基于 SIR、TQC、TMC 分类方法并计算调整后的数据作为训练集。

(2) 选取 2018 年 3 月—8 月基于 SIR、TQC、TMC 分类方法调整后的调整数据作为测试集。

表 4 基于 TMC 分类法计算的季节指数

季度	季节指数(s_j)	包含月份
1	0.865 879	1 月
2	0.765 779	2 月
3	0.981 581	3 月
4	0.974 798	4 月
5	0.989 191	5 月
6	0.951 178	6 月
7	1.044 986	7 月
8	1.038 398	8 月
9	1.046 014	9 月
10	1.113 877	10 月
11	1.098 274	11 月
12	1.130 046	12 月

表 5 基于 TQC 分类法计算的季节指数

季度	季节指数(s_j)	包含月份
1	0.981 856	3~5 月
2	1.011 521	6~8 月
3	1.086 055	9~11 月
4	0.920 568	12~2 月

(3) 先设置较少的隐含节点训练网络,然后逐渐增加隐含节点数,用同一样本集进行训练,从中确定网络误差最小时对应的隐含层节点数。

经过检验,最佳隐含层节点数取值公式为:

$$m = \sqrt{n + l} \pm \omega \tag{7}$$

式中:

- n ——输入层节点数(本文为 6);
- l ——输出层节点数(本文为 1);
- ω ——1~10 之间的常数;
- m ——隐含层节点数,取寻优范围为[1,16]。

(4) 确定适应度函数,以 2018 年 3 月—8 月预测客流量绝对误差作为适应度函数,即:

$$f = \sum_{j=3}^8 (|y_j' - x_j| + |y_j'' - x_j| + |y_j''' - x_j|) \tag{8}$$

式中:

x_j, y_j', y_j'', y_j''' ——为 2018 年第 j 月的实际客流量应用 SIR、TQC、TMC 分类法预测后的调整客流量。

由于寻优范围较小,采用遍历法对节点数进行寻优,寻优结果如图 6 所示。由图 6 可见,最优隐含层节点数为 3,即建立 6-3-1 的三层神经网络。

2) 以 2014 年 3 月—2018 年 8 月客流量为训练集,应用优化后的神经网络预测 2018 年 9 月—2019 年 2 月的客流量,经过逆向调整(式(6)),最终得到预测值,如表 6 所示。

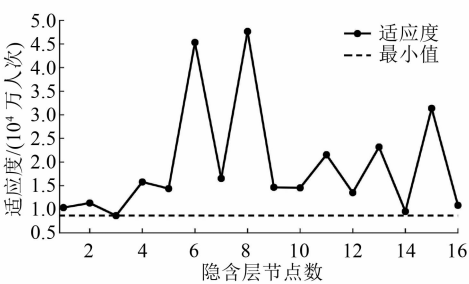


图 6 适应度函数曲线

表 6 西安地铁 2 号线 2018 年 9 月—2019 年 2 月的客流量预测结果

单位:万人次

月份	真实值	SIR 预测值	TQC 预测值	TMC 预测值
2018-09	2 837.012	3 099.517	3 152.149	3 155.275
2018-10	3 067.312	3 196.346	3 244.231	3 443.528
2018-11	2 899.607	3 220.592	3 300.129	3 445.611
2018-12	2 960.074	2 975.942	2 836.286	3 509.891
2019-01	2 840.498	2 793.543	2 946.825	2 796.631
2019-02	2 360.364	2 512.218	2 914.005	2 520.425

3.7 预测结果评价

为了消除随机性,分别训练 100 次并对预测客流取平均值,预测结果如表 6 所示。根据实际值计算相对误差(RE)、平均绝对百分比误差(MAPE)、和均方根误差(RMSE)来对预测结果进行分析,结果如表 7 所示。

表 7 西安地铁 2 号线 2018 年 9 月—2019 年 2 月的客流量预测误差对比分析

分类方法	指标	客流量预测误差/%					
		18-09	18-10	18-11	18-12	19-01	19-02
SIR 分类	RE	9.25	4.20	11.06	0.53	1.70	6.43
	RMSE	26.63	20.10	14.83	16.65	13.53	21.37
	MAPE	5.53					
TQC 分类	RE	11.10	5.76	13.81	4.18	3.74	23.45
	RMSE	28.50	21.73	20.00	18.93	14.71	21.45
	MAPE	10.34					
TMC 分类	RE	11.21	12.26	18.83	18.57	1.54	6.78
	RMSE	106.02	24.76	51.65	19.56	17.92	30.28
	MAPE	11.53					

进一步地,可由表 7 得到如图 7 所示的各分类方法预测误差对比图。

通过比较各分类方法的预测指标,SIR 分类预测所得到的相对误差、平均绝对百分比误差、均方根误差小于传统的季节分类方法的,相对误差控制在 10%以内(由于对 2016 年 11 月客流量的修正误差,其对应月份预测误差较大),个别月份达到 1%以内,预测精度相对较高。

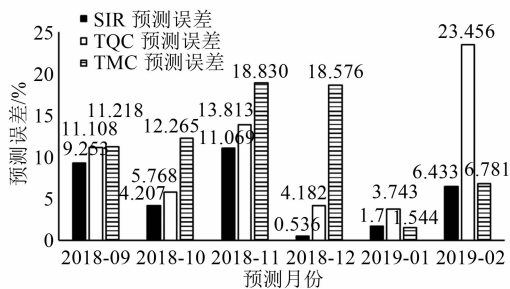


图7 西安地铁2号线2018年9月—2019年2月不同分类方法的客流量预测误差对比图

自然季节性和制度节性是影响季节性波动最主要的因素。TMC和TQC分类方法的季节分组只体现了自然季节性,而本文所采用的方法根据两者叠加的结果进行分组,更为准确。

综上所述,基于月客流残差的季节指数预测方法具有较高的预测精度和较好的稳定性。

4 结语

本文通过对月客流残差分析,提出了基于月客流残差的季节分类方法,以季节指数调整客流达到提高预测精度的目的。通过对三种分类方法进行对比分析,本文提出的分类方法在月度客流预测中精度较高,基本满足对客流量的预测需求。

(上接第76页)

反而会导致 ISCS 业务运行更加复杂,影响业务平台的可靠性、实效性,会起相反的作用。

3) 问题整改:测评认证是在技术与管理共用情况下的一个综合安全评定。问题整改是落实等级保护工作的关键,适合业务系统的整改策略,要充分利用各安全组件,严格按照网络安全等级保护标准要求、测评要求及时、规范地进行问题整改。

5 结语

根据等保2.0三级防护规范和实施指南要求,通过部署安全管理中心、加固计算环境安全、增强边界安全隔离、进行安全审计等措施,可以显著提升 ISCS 信息安全防护能力,满足等保三级要求。本文介绍的方案已在 ISCS 信息安全建设项目中进行实施,并通过了第三方机构的信息安全三级评测及 CNAS 的认证。当然,ISCS 的等保三级安全设计还是较新的课题,等级保护也只是现代信息系统安全应达到的底线,还需要根据后续的实际运行情况不断进行改进。

通过该分类方法计算出的季节指数用于预测时,能更好地适应城市轨道交通客流的季节性变化趋势,从而使运营公司及时调整相关运能运力,制定更加经济合理的月度或年度运营计划。

考虑到影响城市轨道交通客流季节性因素的多样性,后续需要对网络、不同线路和站点的季节区间进行对比分析。

参考文献

- [1] 王夏秋,张宁,王健.基于季节指数的城市轨道交通月度客流预测方法[J].城市轨道交通研究,2018(10):25.
- [2] 赵钰棠,杨信丰,杨珂.基于支持向量机的地铁客流量预测[J].都市快轨交通,2014(3):35.
- [3] 吉海兵,吕效国,周培,等.加权季节性指数法及其应用[J].数学的实践与认识,2010(12):80.
- [4] 何俊,刘会茹,张彦群.不变季节指数法的应用[J].黄冈师范学院学报,2008(3):25.
- [5] 邓明,张荷观.利用虚拟变量对季节指数的估计[J].统计与决策,2007(4):15.
- [6] 王跃军,刘万春,王正平,等.基于季节指数修正因子的交乘趋向预测方法[J].军民两用技术与产品,2006(4):43.
- [7] 郑薇,王灿强,李维德.基于季节指数调整与 HGWO-SVR 算法的农产品价格预测模型[J].统计与决策,2018(19):33.
- [8] 周华任,李浩然,孙学金,等.一种基于季节指数的灰色马尔科夫气温预测模型[J].数学的实践与认识,2016(4):167.

(收稿日期:2019-10-17)

参考文献

- [1] 中华人民共和国交通运输部.交通运输信息化“十三五”发展规划[R].北京:中华人民共和国交通运输部,2016.
- [2] 杨雪飞.浅谈等级保护2.0的新变化和等保工作的误区[J].IT经理世界,2019(1):51.
- [3] 中华人民共和国工业和信息化部电子工业标准化研究院.城市轨道交通综合监控系统工程技术标准:GB/T 50636—2018[S].北京:中国建筑工业出版社,2018.
- [4] 国家标准化管理委员会.信息安全技术网络安全等级保护基本要求:GB/T 22239—2019[S].北京:中国标准出版社,2019.
- [5] 国家标准化管理委员会.信息安全技术网络安全等级保护安全技术要求:GB/T 25070—2019[S].北京:中国标准出版社,2019.
- [6] 徐世亮,操屹.浅析构建国土资源网上交易系统安全防护体系[J].江西通信科技,2013(2):45.
- [7] 单联春.防火墙与入侵检测系统联动构建网络安全堡垒[J].有线电视技术,2005(10):80.
- [8] 高锐强,朱虹,贾立东,等.基于SSL安全协议实现工业控制通讯协议加密及认证的研究[J].化工设计通讯,2019(1):121.

(收稿日期:2020-12-10)