

引用本文:史小丽,王晓伟,邱晨阳,等.基于机器学习法的高速公路桥涵小修工程费用模型预测[J].建筑科学与工程学报,2023,40(4):125-134.

SHI Xiaoli, WANG Xiaowei, QIU Chenyang, et al. Cost model prediction of minor repair project of bridge and culvert based on machine learning method[J]. Journal of Architecture and Civil Engineering, 2023, 40(4): 125-134.

DOI:10.19815/j.jace.2021.12106

# 基于机器学习法的高速公路桥涵小修工程费用模型预测

史小丽<sup>1</sup>, 王晓伟<sup>2</sup>, 邱晨阳<sup>3</sup>, 高楠<sup>4</sup>

(1. 长安大学 公路学院, 陕西 西安 710064; 2. 中国路桥工程有限责任公司, 北京 100011;  
3. 苏交科集团股份有限公司, 江苏 南京 210019; 4. 陕西交通控股集团有限公司, 陕西 西安 710065)

**摘要:**基于陕西省 11 条高速公路 2008~2015 年桥涵小修工程量清单历史数据,采用机器学习算法,以组成桥涵各构件的小修费用作为因变量,研究桥涵小修费用预测模型。通过灰色关联度模型分析桥涵小修费用的影响因素,采用皮尔逊相关系数检验法对各影响因素进行多重共线性检验,筛选出解释变量,使用岭回归和 Lasso 回归对桥涵各构件小修费用模型进行回归分析,得到桥涵小修总费用预测模型。结果表明:桥涵小修工程费用的影响因素主要有通车年限、桥梁(涵洞)长度、年平均当量轴次、桥涵所处地区的年均降雨量和温度及车道数;基于模型预测所依托高速公路 2016~2017 年桥涵小修费用,并与该年度实际费用进行 Wilcoxon 符号秩检验,检验结果均大于 0.05,验证了机器学习法预测桥涵小修工程费用模型的有效性,预测结果能为分配养护费用、提高养护决策水平提供合理建议。

**关键词:**桥涵资产;小修费用;灰色关联度;岭回归;Lasso 回归

中图分类号:U445.2

文献标志码:A

文章编号:1673-2049(2023)04-0125-10

## Cost model prediction of minor repair project of bridge and culvert based on machine learning method

SHI Xiaoli<sup>1</sup>, WANG Xiaowei<sup>2</sup>, QIU Chenyang<sup>3</sup>, GAO Nan<sup>4</sup>

(1. School of Highway, Chang'an University, Xi'an 710064, Shaanxi, China; 2. China Road & Bridge Corporation, Beijing 100011, China; 3. JSTI Group Co., Ltd., Nanjing 210019, Jiangsu, China; 4. Shaanxi Transportation Holding Group Co., Ltd., Xi'an 710065, Shaanxi, China)

**Abstract:** Based on the historical data of bill of quantities of bridge and culvert minor repair project on 11 expressways in Shaanxi province from 2008 to 2015, the machine learning algorithm was used to study the prediction model of bridge and culvert minor repair cost with the minor repair cost of each component of bridge and culvert as the dependent variable. Through the grey correlation degree model, the influencing factors of the minor repair cost of bridge and culvert were analyzed. The Pearson correlation coefficient test method was used to test the multicollinearity of each influencing factor, and the explanatory variables were selected. Ridge

收稿日期:2021-12-05

基金项目:中国路桥工程有限责任公司科研项目(2020-zlkj-03);陕西省自然科学基金基础研究计划项目(2022JM-307)

作者简介:史小丽(1979-),女,工学博士,副教授,硕士生导师,E-mail:glxl@gl.chd.edu.cn.

regression and Lasso regression were used to analyze the minor repair cost model of each component of bridge and culvert, and the total cost prediction model of minor repair of bridge and culvert was obtained. The results show that the main influencing factors of the cost of bridge and culvert minor repair projects are years of operation, length of bridge (culvert), annual average equivalent axles, annual average rainfall and temperature in the area where the bridges and culverts are located, and number of lanes. Based on the model prediction, the bridge and culvert minor repair cost of the expressway from 2016 to 2017 is carried out, and the Wilcoxon signed rank test is carried out with the actual cost of the year. The test results are all greater than 0.05, which verifies the effectiveness of the machine learning method to predict the bridge and culvert minor repair project cost model. The prediction results can provide reasonable suggestions for allocating maintenance costs and improving maintenance decision-making level.

**Key words:** bridge and culvert asset; minor repair cost; grey correlation degree; Ridge regression; Lasso regression

## 0 引言

小修作为修复性养护中发生频率最高的养护方式,其养护成效和费用投入直接影响了大中修的养护时机和养护费用。本文通过对陕西省陕北、陕南以及关中 3 个地区部分高速公路的小修工程费用进行分析,发现桥涵资产小修工程费用占高速公路小修工程总费用的比重随着通车年限呈总体上升趋势,且在通车 5 年后所占比重超越各项资产小修工程费用比重均值。因此,在养护资金受限的情况下,对桥涵养护费用进行分析并建立有效预测模型,对提高养护费用预算决策水平具有十分重要的意义<sup>[1]</sup>。

费用预测方法主要分为单位成本法和模型回归法。单位成本法通过单位长度桥涵养护费用与桥涵总长度的乘积得到桥涵养护费用<sup>[2-3]</sup>;模型回归法是建立有效模型对桥涵养护费用进行预测,这种分析方法应用较为广泛。部分学者通过制定养护目标,建立养护效益模型预测桥涵养护费用<sup>[4-7]</sup>,部分学者通过分析桥涵养护费用影响因素,并将其作为解释变量进行建模来预测桥涵养护费用,常用的费用预测模型有固定效应模型、随机效应模型、最小二乘法<sup>[8-12]</sup>。上述研究均将桥涵养护费用作为一个整体来分析,但组成桥涵的不同构件在使用过程中发挥的功能不同,其使用寿命及相应的养护费用支出影响因素也必然不同,所以对不同构件养护费用影响因素进行分析,并建立有效预测模型能提高预测的准确性。Python 能快速读取并处理大量数据,且岭回归模型和 Lasso 回归模型较上述模型而言,在处理非平稳状态下的数据时具有一定优势,所以本文通过分析陕西省部分高速公路桥涵资产历史养护数

据,采用 Python 对桥涵构件小修费用影响因素进行灰色关联度分析,分别建立基于岭回归和 Lasso 回归的桥涵各构件小修费用模型,并引入小修概率的思想,建立桥涵小修总费用模型。

## 1 基于灰色关联度分析法的桥涵小修工程费用影响因素

### 1.1 影响因素指标的量纲一化处理

灰色关联度模型分析法是从多角度考虑,寻求两事物或两系统之间关联的一种因素分析法,所以建立可靠的桥涵小修工程费用预测模型,必须对影响桥涵小修工程费用的因素进行分析。影响桥涵小修工程费用的因素是多种多样的,本文选取以下 6 个影响因素作为分析对象:通车年限( $x_1$ )、桥梁(涵洞)长度( $x_2$ )、年平均日当量轴次( $x_3$ )、桥涵所处地区的年均降雨量( $x_4$ )、年均温度( $x_5$ )、车道数( $x_6$ )。

建立灰色关联度模型时,首先应确定参考数列和比较数列。反映系统行为特征的数据序列称为参考数列。影响系统行为的因素组成的数据序列称为比较数列。

根据表 1 的桥涵小修工程量清单,在桥涵小修工程影响因素分析中,设桥涵各构件小修工程费用  $y_i = \{y_i(k) | k = 1, 2, \dots, n\}$  为参考数列(又称母序列),其中: $y_1$  为混凝土护栏修补费用; $y_2$  为防抛网小修费用; $y_3$  为伸缩装置小修费用; $y_4$  为排水设施小修费用; $y_5$  为金属护栏小修费用; $y_6$  为涵洞渗漏修堵费用; $y_7$  为涵洞、通道修补费用; $y_8$  为钢扶手小修费用。设通车年限  $x_1 = \{x_1(k) | k = 1, 2, \dots, n\}$ 、桥梁(涵洞)长度  $x_2 = \{x_2(k) | k = 1, 2, \dots, n\}$ 、交通量  $x_3 = \{x_3(k) | k = 1, 2, \dots, n\}$ 、桥涵所处地区的年

表 1 桥涵小修工程工程量清单

Table 1 Bill of quantities of bridge and culvert minor repair project		
对象	序号	名称
桥梁混凝土护栏 维修	1	桥梁混凝土护栏修补
	2	桥梁混凝土护栏刷涂料
	3	桥梁混凝土护栏伸缩缝铝扣板更换
桥梁钢扶手维修	1	更换钢管
	2	更换底座
	3	钢扶手除锈刷漆
伸缩装置维修	1	伸缩缝混凝土维修
	2	伸缩缝更换
	3	橡胶条更换、维修
	4	伸缩缝挡板更换
涵洞渗漏修堵	1	裂缝封闭、灌浆
	2	裂缝粘贴碳纤维布
桥梁金属护栏维修	1	金属护栏更换
	2	金属护栏刷漆
桥梁防抛网维修	1	更换防抛网
	2	更换防抛网立柱(含底座基础)
	3	防抛网刷漆(含底座)
排水设施维修	1	桥梁集中排水管(PVC)更换
	2	更换泄水孔
	3	泄水孔盖板补缺
	4	更换、修补铁算子
	5	打排水孔
涵洞、通道修补	1	混凝土结构物修补
	2	砌体勾缝抹面(厚度大于 2 cm)

平均降雨量  $x_4=\{x_4(k)|k=1,2,\cdots,n\}$ 、年平均温度  $x_5=\{x_5(k)|k=1,2,\cdots,n\}$ 、车道数  $x_6=\{x_6(k)|k=1,2,\cdots,n\}$  为比较数列(又称子序列)。由参考数列和比较数列构成的矩阵为

$$[y_i\ x_1\ x_2\ \cdots\ x_6]=\begin{bmatrix}y_i(1) & x_1(1) & \cdots & x_6(1) \\ y_i(2) & x_1(2) & \cdots & x_6(2) \\ \vdots & \vdots & & \vdots \\ y_i(n) & x_1(n) & \cdots & x_6(n)\end{bmatrix}$$

由于 6 个影响因素指标具有不同的量纲,所以

表 2 桥涵各构件小修费用与各影响因素指标的关联度

Table 2 Correlation between minor repair cost of bridge and culvert components and indicators of influencing factors								
指标	混凝土护栏 修补费用	防抛网小 修费用	伸缩装置 小修费用	排水设施 小修费用	金属护栏 小修费用	涵洞渗漏 修堵费用	涵洞、通道 修补费用	钢扶手小修 费用
通车年限( $x_1$ )	0.845	0.740	0.924	0.955	0.859	0.873	0.747	0.952
桥梁(涵洞)长度( $x_2$ )	0.838	0.771	0.913	0.889	0.922	0.890	0.799	0.959
年平均日当量轴次( $x_3$ )	0.848	0.814	0.958	0.920	0.881	0.864	0.855	0.950
年均降雨量( $x_4$ )	0.812	0.753	0.923	0.883	0.906	0.863	0.817	0.956
年均温度( $x_5$ )	0.847	0.824	0.912	0.887	0.911	0.859	0.808	0.957
车道数( $x_6$ )			0.901					

在灰色关联度分析之前应对其进行量纲一化处理。选取第一列影响因素指标作为标准要素,其余列影响因素指标为比较要素,将比较要素与标准要素相减。

1.2 关联度计算

所谓关联度,实质上是指曲线之间几何形状的差别程度,因此将参考数列所形成曲线和比较数列所形成曲线之间的差值大小作为衡量被解释变量和解释变量之间关联程度的标准<sup>[13]</sup>。各比较数列  $x_i(k)$  与参考数列  $y_i(k)$  在各个时刻即曲线各点的关联系数为  $\zeta(k)$ ,计算公式如式(1)所示。

$$\zeta_i(k)=\frac{\min_i\min_k|y_i(k)-x_i(k)|+\rho\max_i\max_k|y_i(k)-x_i(k)|}{|y_i(k)-x_i(k)|+\rho\max_i\max_k|y_i(k)-x_i(k)|}$$

(1)

式中: $\rho$  为分辨系数, $\rho$  值越小,分辨力越大,取  $\rho$  值为 0.5<sup>[14]</sup>。

对比较序列分别计算其各个指标与参考序列对应元素的关联系数的均值,用于反映各个解释变量与被解释变量的相关程度  $r_{yx_i}$ ,计算公式如式(2)所示。

$$r_{yx_i}=\frac{1}{n}\sum_{k=1}^n\zeta_i(k)$$

(2)

0< $r_{yx_i}$ ≤1 时, $r_{yx_i}$  越大, $y$  与  $x_i$  之间的关联度就越高; $r_{yx_i}$ >0.7 时,说明  $y$  与  $x_i$  相关度较高<sup>[15]</sup>。

为了消除不同年份人工、材料和机械价格因素的影响,选择 2015 年为基准年度,分别将各年度桥涵各构件小修费用通过建筑安装价格指数的方式折算为基准年度的费用值。使用 Python 计算桥涵各构件小修费用与各影响因素指标之间的关联度,结果如表 2 所示。

由于桥涵各构件小修费用与 6 个影响因素指标的关联度均大于 0.7,所以将其作为解释变量对桥涵各构件小修费用进行回归是合理的。

2 基于岭回归、Lasso 回归的桥涵各构件小修费用预测

2.1 桥涵小修工程费用数据预处理

2.1.1 价格指数折算

为了消除不同年份人工、材料和机械价格因素的影响,选择 2015 年为基准年度,采用《陕西统计年鉴 2012》<sup>[16]</sup>中的建筑安装价格指数分别将各年度桥涵各构件小修费用折算为基准年度的费用值  $C_{2015}$ ,如式(3)所示。

$$C_{2015} = \begin{cases} C_n I_{n+1} I_{n+2} \cdots I_{2015} & n < 2015 \\ \frac{C_n}{I_n I_{n+1} \cdots I_{2016}} & n > 2015 \end{cases} \quad (3)$$

式中: $C_n$  为某个构件第  $n$  年的费用值; $I_n$  为第  $n$  年的价格指数。

2.1.2 费用偏度检验

偏度是统计数据分布偏斜方向和程度的度量,是统计数据分布非对称程度的数字特征。偏度是对于分布的标准三阶中心距,如式(4)所示。在统计学中,通常认为偏度的绝对值不大于 2 时的分布为正态分布,可进行模型回归;当偏度大于 2 时,应对具有偏态分布性质的时间序列取自然对数来避免伪回归、消除异方差,在不改变时间序列的性质及相关性的前提下,获得平稳的数据<sup>[17]</sup>。

$$S_k = \frac{\mu_3}{\sigma^3} \quad (4)$$

式中: $S_k$  为偏度; $\mu_3$  为三阶中心距; $\sigma$  为标准差。

使用 Python 中 `stats.mastats.skew()`、`data` 命令计算桥涵小修工程量清单中各个细目的费用偏度值,计算结果如表 3 所示。由于桥涵小修各细目费用偏度值均小于 2,故可直接进行回归,无需取对数处理。

表 3 桥涵小修各细目费用偏度

Table 3 Cost skewness of each item of minor repair of bridge and culvert

细目名称	费用偏度
桥梁混凝土护栏修补	0.4
桥梁防抛网小修	-0.2
桥梁伸缩装置小修	0.3
排水设施小修	0.3
桥梁金属护栏小修	0.1
涵洞渗漏修堵	-0.2
涵洞、通道修补	0.9
桥梁钢扶手小修	0.6

2.1.3 交通量数据预处理

《公路沥青路面设计规范》(JTG D50—2017)<sup>[18]</sup>将车辆类型分为 11 类,本文所调研高速公路的通车类型按车型可分为 1 类车、2 类车、3 类车和 5 类车,根据规范,各类车辆的当量轴载换算系数  $N_1$  计算公式如式(5)所示。

$$N_1 = ADL \sum_{m=2}^{11} (V_m E_m) \quad (5)$$

式中: $A$  为 2 轴 6 轮以上车辆的双向年平均日交通量; $D$  为方向系数; $L$  为车道系数; $m$  为车辆类型分布系数; $V_m$  为  $m$  类车辆类型分布系数; $E_m$  为  $m$  类车辆的当量设计轴载换算系数。

根据规范,以沥青混合料层层底拉应变、沥青混合料层永久变形量作为指标,将交通量转换为年平均日当量轴次。

2.1.4 解释变量数据量纲一化处理

进行解释变量量纲一化处理主要有以下两方面原因:①由于各个解释变量物理意义不同,数据的大小分布有质的差异,不能直接参与回归;②在桥涵小修工程过程中,由于特殊原因的出现,会出现某些极大或者极小的异常数据,数据并不满足正态分布。本文选择标准化对数据进行预处理,使各解释变量具有同等程度的表现力,并处于同一数量级,从而可进行综合对比评价,标准化处理如式(6)~(11)所示。

$$\bar{y}_i = \frac{1}{n} \sum_{k=1}^n y_i(k) \quad k = 1, 2, \cdots, n, i = 1, 2 \quad (6)$$

$$\bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_i(k) \quad k = 1, 2, \cdots, n, i = 1, 2, \cdots, 6 \quad (7)$$

$$s_y = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (y_k - \bar{y})^2} \quad (8)$$

$$s_x = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2} \quad (9)$$

$$y_i(k) = \frac{y_k - \bar{y}}{s_y} \quad (10)$$

$$x_i(k) = \frac{x_k - \bar{x}}{s_x} \quad (11)$$

式中: $s_y$  为被解释变量标准差; $s_x$  为解释变量标准差。

2.2 特征之间的多重共线性检验

在模型回归时,若各个解释变量之间相关性较高,则会导致参数估计量的方差增大,变量的显著性检验失去意义,参数估计量经济含义不合理,从而形成“病态矩阵”,在各解释变量相关性很强的情况下回归的模型很可能忽略重要的解释变量,所以即使

模型拟合度很高,也可能是无效的,因此本文采用皮尔逊相关系数检验法对解释变量的相关性进行检验<sup>[19]</sup>。皮尔逊相关系数  $\rho_{(x_1,x_2)}$  采用两变量之间协方差和标准差的商表示,相关系数用于研究变量之间的相关程度,其取值范围为  $[-1,1]$ ,计算公式如式(12)所示。

$$\rho_{(x_1,x_2)} = \frac{C(x_1,x_2)}{\sqrt{D(x_1)}\sqrt{D(x_2)}} = \frac{E\{[x_1-E(x_1)][x_2-E(x_2)]\}}{\sigma_{x_1}\sigma_{x_2}} \quad (12)$$

式中: $C(x_1,x_2)$ 为变量  $x_1$  与  $x_2$  的协方差; $E(x_1)$ 、 $E(x_2)$ 分别为变量  $x_1$ 、 $x_2$  的期望; $D(x_1)$ 、 $D(x_2)$ 分别为变量  $x_1$ 、 $x_2$  的方差; $\sigma_{x_1}$ 、 $\sigma_{x_2}$ 分别为变量  $x_1$ 、 $x_2$  的标准差。

$\rho_{(x_1,x_2)}$  表征  $x_1$ 、 $x_2$  之间线性关系的紧密程度,  $|\rho_{(x_1,x_2)}|$  越大,  $x_1$ 、 $x_2$  之间的相关性越强。在经济计量学中,通常认为相关系数高于 0.7 时变量之间存在多重共线性,反之,则表示变量之间不存在线性关系<sup>[15]</sup>。

本文使用 Python 计算桥涵小修工程量清单各细目解释变量的皮尔逊相关系数,并绘制出热度图,如图 1~8 所示,颜色越深,代表关联程度越大,反之越小。

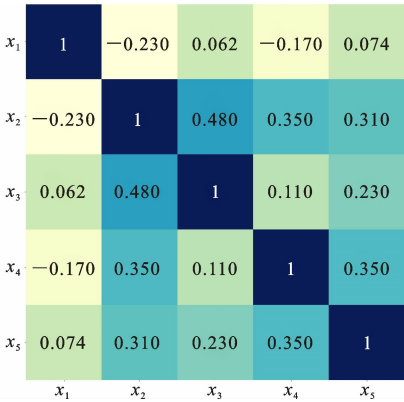


图 1 混凝土护栏修补费用预测解释变量皮尔逊相关系数热度图

Fig. 1 Heat diagram of Pearson correlation coefficient of explanatory variable for prediction of repair cost of concrete guardrail

检验结果表明:

(1)由图 1、2、4、5 和图 8 可知,其各个解释变量数值之间的相关系数均小于 0.7,因此,可将通车年限( $x_1$ )、桥梁长度( $x_2$ )、年平均日当量轴次( $x_3$ )、年均降雨量( $x_4$ )以及年均温度( $x_5$ )同时作为解释变量对其费用模型进行预测。

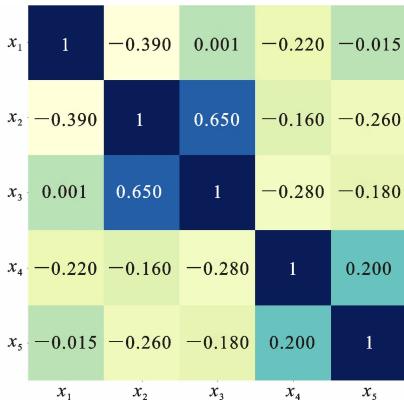


图 2 防抛网小修费用预测解释变量皮尔逊相关系数热度图

Fig. 2 Heat diagram of Pearson correlation coefficient of explanatory variable for prediction of minor repair cost of anti-throwing net

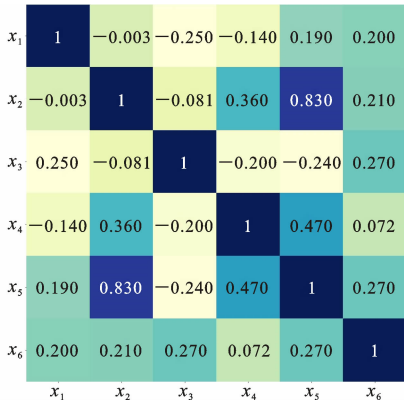


图 3 伸缩装置小修费用预测解释变量皮尔逊相关系数热度图

Fig. 3 Heat diagram of Pearson correlation coefficient of explanatory variable for prediction of minor repair cost of expansion device

(2)由图 3 可知,桥梁长度( $x_2$ )和年均温度( $x_5$ )之间的相关系数大于 0.7,说明两变量在数值上呈高度相关性,根据第 1.2 节分析可知小修费用与桥梁所处地区的年均温度( $x_5$ )之间的灰色关联度大于其与桥梁长度( $x_2$ )之间的灰色关联度,所以应将年均温度( $x_5$ )变量剔除,将桥梁长度( $x_2$ )、通车年限( $x_1$ )、年平均日当量轴次( $x_3$ )、年均降雨量( $x_4$ )以及车道数( $x_6$ )作为解释变量参与伸缩装置小修费用模型预测。

(3)由图 6 可知,涵洞长度( $x_2$ )和年均降雨量( $x_4$ )之间的相关系数大于 0.7,说明两变量在数值上呈高度相关性,根据第 1.2 节分析可知小修费用与桥梁所处地区的年均降雨量( $x_4$ )之间的灰色关联度大于其与涵洞长度( $x_2$ )之间的灰色关联度,所以

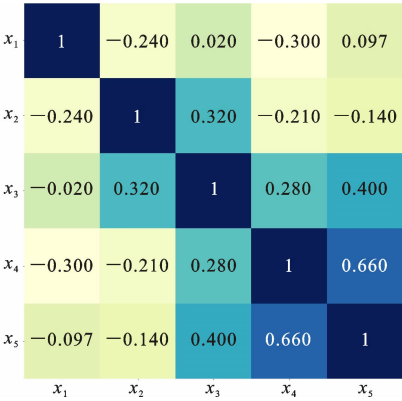


图 4 排水设施小修费用预测解释变量皮尔逊相关系数热度图

Fig. 4 Heat diagram of Pearson correlation coefficient of explanatory variable for prediction of minor repair cost of drainage facilities

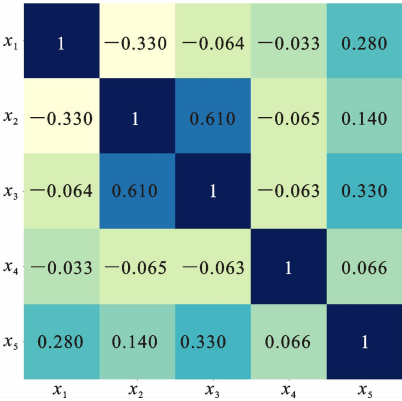


图 5 金属护栏小修费用预测解释变量皮尔逊相关系数热度图

Fig. 5 Heat diagram of Pearson correlation coefficient of explanatory variable for prediction of minor repair cost of metal guardrail

应将涵洞长度( $x_2$ )变量剔除,将年均降雨量( $x_4$ )、通车年限( $x_1$ )、年平均日当量轴次( $x_3$ )以及年均温度( $x_5$ )作为解释变量参与涵洞渗漏修堵费用模型预测。

(4)由图 7 可知,年平均日当量轴次( $x_3$ )和年均温度( $x_5$ )之间相关系数的绝对值大于 0.7,说明两变量在数值上呈高度相关性,根据第 1.2 分析可知小修费用与桥梁所处地区的年均降雨量( $x_3$ )的灰色关联度大于与涵洞长度( $x_5$ )的灰色关联度,因此将年均温度( $x_5$ )变量剔除,把年平均日当量轴次( $x_3$ )、通车年限( $x_1$ )、涵洞长度( $x_2$ )以及年均降雨量( $x_4$ )作为解释变量参与涵洞通道修补费用模型预测。

2.3 基于机器学习的桥涵小修工程费用预测模型

由于岭回归和 Lasso 回归加入正则化的思想,

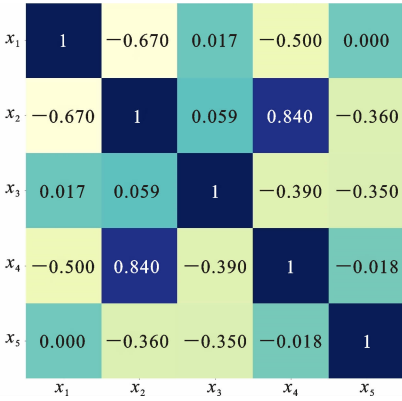


图 6 涵洞渗漏修堵费用预测解释变量皮尔逊相关系数热度图

Fig. 6 Heat diagram of Pearson correlation coefficient of explanatory variable for prediction of repair cost of culvert leakage

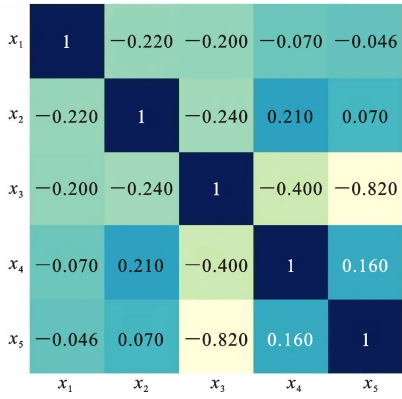


图 7 涵洞、通道修补费用预测解释变量皮尔逊相关系数热度图

Fig. 7 Heat diagram of Pearson correlation coefficient of explanatory variable for prediction of repair cost of culvert and channel

有效改善了最小二乘法过拟合的现象,所以本文采用机器学习的方法,通过 Python 分别对各个细目的费用进行岭回归和 Lasso 回归,通过比较两回归方法的平均绝对百分比误差,从而选取最优模型作为各个细目的费用预测模型<sup>[20]</sup>。

2.3.1 岭回归分析法

岭回归实质上是放弃最小二乘法的无偏性,在损失函数基础上加上一个二范数正则化项<sup>[21]</sup>。通过 Python 对桥涵小修工程各个细目进行岭回归,得到各解释变量的系数统计表(表 4)。

2.3.2 Lasso 回归分析法

Lasso 回归法在建立广义线性模型时,无论因变量是连续的还是离散的,Lasso 都能处理,对数据要求是极低的,所以应用范围较广<sup>[22]</sup>。它通过构造一个惩罚函数,得到一个较为精炼的模型,从而压缩

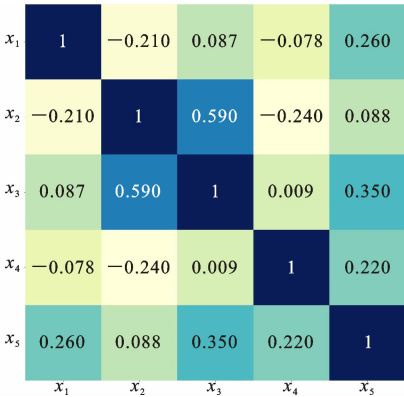


图 8 钢扶手小修费用预测解释变量皮尔逊相关系数热度图

Fig. 8 Heat diagram of Pearson correlation coefficient of explanatory variable for prediction of minor repair cost of steel handrail

一些回归系数,即强制系数绝对值之和小于某个固定值,同时设定一些回归系数为 0。Lasso 回归保留子集收缩的优点,是一种处理具有共线性数据的有偏估计。Lasso 的基本思想同岭回归类似,都是将不适定问题转化为适定问题。Lasso 解决此类问题是在损失函数后加上一个一范数的正则化项。

通过 Python 对桥涵小修工程各个细目进行 Lasso 回归,得到各解释变量的系数统计表,如表 5 所示。

2.3.3 最优模型的选择

岭回归和 Lasso 回归均能解决影响因素之间轻微共线性问题,但由于加入的正则化项不同,所以在对数量不同和平稳程度不同的数据进行回归时,准确度也不同。本文使用测试集中数据对桥涵各构件小修费用预测模型进行检验,选择拟合优度 $R^2$ 较

表 4 桥涵各构件小修费用岭回归系数

Table 4 Ridge regression coefficient of minor repair cost of bridge and culvert components

变量	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	常数项 $b$
桥梁混凝土护栏修补	-222.1	892.4	-351.7	106.2	849.0		2 653.3
桥梁防抛网小修	-176.4	648.0	115.6	190.7	318.0		3 367.1
桥梁伸缩装置小修	2 150.6	-62 688.3	-10 736.8	-21 782.7		40.1	119 286.5
桥梁排水设施小修	8 425.9	-10 967.5	4 020.5	1 124.1	-2 438.4		30 207.6
桥梁金属护栏小修	734.6	-3 796.7	-90.7	370.9	-344.1		7 178.3
涵洞渗漏修堵	-573.2		-1 353.5	10.4	2 669.0		20 210.1
涵洞、通道修补	-6 705.9	10 722.7	-3 239.6	4 666.6			23 829.6
桥梁钢扶手小修	-328.4	1 200.9	2 416.1	-231.6	1 575.7		6 887.0

表 5 桥涵各构件小修费用 Lasso 回归系数

Table 5 Lasso regression coefficient of minor repair cost of bridge and culvert components

变量	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$b$
桥梁混凝土护栏修补	0.0	935.0	0.0	101.9	863.4		2 653.3
桥梁防抛网小修	0.0	899.9	0.0	183.7	364.3		3 367.1
桥梁伸缩装置小修	11 839.5	0.0	0.0	0.0		0.0	119 286.5
桥梁排水设施小修	18 042.9	0.0	1 736.6	0.0	0.0		30 207.6
桥梁金属护栏小修	2 398.1	0.0	0.0	1 112.0	212.6		7 178.3
涵洞渗漏修堵	0.0		0.0	809.0	4 586.8		20 210.1
涵洞、通道修补	0.0	16 625.1	1 654.0	10 609.2			23 830.0
桥梁钢扶手小修	0.0	1 419.8	2 546.1	0.0	15 183.1		6 887.0

大的模型作为最优模型,结果如表 6 所示。

3 桥涵小修工程总费用模型预测

桥梁的设计使用年限较长,高速公路及一级公路为 100 年,其小修工程费用在建成前期较少;栏杆、伸缩缝、支座的设计使用年限约为 15 年,且随通车年限的增加其小修费用呈上升趋势。由于并非每年都对桥涵各构件进行小修,所以在一定年限内,各

构件的小修频率也不尽相同,通过分析所依托的 11 个高速公路桥涵小修工程 2008~2017 年工程量清单,引入小修概率的思想,将各条高速公路各桥涵构件 10 年内小修发生频率作为系数,对桥涵小修工程总费用模型  $w$  进行预测,结果如表 7 所示。

4 模型验证

根据西长高速、靖王高速和商界高速桥涵 2016



表 6 桥涵小修工程各项细目最优模型

Table 6 Optimal model of each item of bridge and culvert minor repair project

桥涵小修工程细目	最优模型
桥梁混凝土护栏修补	$y_1=935.x_2+101.9.x_4+863.4.x_5+2\ 653.3$
桥梁防抛网小修	$y_2=899.9.x_2+183.7.x_4+364.3.x_5+3\ 367.1$
桥梁伸缩装置小修	$y_3=2\ 150.6.x_1-62\ 688.3.x_2-10\ 736.8.x_3-21\ 782.7.x_4+40.1.x_6+119\ 286.5$
桥梁排水设施小修	$y_4=8\ 425.9.x_1-10\ 967.5.x_2+4\ 020.5.x_3+1\ 124.1.x_4-2\ 438.4.x_5+30\ 207.6$
桥梁金属护栏小修	$y_5=734.6.x_1-3\ 796.7.x_2-90.7.x_3+370.9.x_4-344.1.x_5+7\ 178.25$
涵洞渗漏修堵	$y_6=-573.2.x_1-1\ 353.5.x_3+10.4.x_4+2\ 669.x_5+20\ 210.1$
涵洞、通道修补	$y_7=-6\ 705.9.x_1+10\ 722.7.x_2-3\ 239.6.x_3+4\ 666.6.x_4+23\ 829.6$
桥梁钢扶手小修	$y_8=-328.4.x_1+1\ 200.9.x_2+2\ 416.1.x_3-231.6.x_4+1\ 575.7.x_5+6\ 887$

表 7 各条高速公路桥涵小修工程总费用预测模型

Table 7 Prediction model of total cost of minor repair project of bridge and culvert for each expressway

高速名称	最优模型
商界高速	$w=0.3.y_1+0.1.y_2+0.8.y_3+0.3.y_4+0.4.y_5+0.2.y_6+0.1.y_7+y_8$
商漫高速	$w=0.2.y_1+0.5.y_2+0.8.y_3+0.4.y_4+0.2.y_5+0.5.y_6+0.3.y_7+0.5.y_8$
西商高速	$w=0.5.y_1+0.5.y_2+0.4.y_3+0.4.y_4+0.2.y_5+0.5.y_7+0.3.y_8$
绕城高速	$w=0.1.y_1+0.2.y_2+0.5.y_3+0.2.y_4+0.1.y_5+0.1.y_6+0.2.y_7+0.4.y_8$
西长高速	$w=0.3.y_1+0.9.y_2+0.8.y_3+0.3.y_4+y_5+0.2.y_6+0.2.y_7+y_8$
西镇高速	$w=0.3.y_1+0.7.y_2+0.8.y_3+0.2.y_4+0.4.y_5+0.7.y_6+0.6.y_7+0.9.y_8$
吴靖高速	$w=0.2.y_1+y_3+0.4.y_4+0.1.y_7+0.1.y_8$
榆绥高速	$w=0.3.y_3+0.1.y_5$
靖王高速	$w=0.1.y_1+0.2.y_2+0.3.y_3+0.2.y_4+0.2.y_5+0.4.y_8$
神府高速	$w=0.2.y_3+0.2.y_4$
延志吴高速	$w=0.2.y_1+0.3.y_3+0.2.y_4$

年和 2017 年的实际运营数据,使用平均绝对百分比误差和 Wilcoxon 符号秩检验法对桥涵小修工程总费用模型的准确性进行验证,结果如表 8 所示。

检验结果表明:

(1)桥涵小修工程总费用的预测值与实际值之间的平均绝对百分比误差为 4.66%,误差较小,说明模型预测较为准确。

(2)Wilcoxon 符号秩检验结果的显著性为 0.753,大于 0.05,说明原假设成立,即费用预测值与实际值的分布无显著性差异,表明预测较为准确。

综上所述,该模型适用于陕西省陕北、陕南和关中高速公路桥涵资产的小修工程费用预测,且预测

较为准确。

5 结 语

通过分析陕西省部分高速公路桥涵资产小修工程费用历史数据,采用灰色关联度模型分析确定其影响因素有:通车年限、桥梁(涵洞)长度、年平均当量轴次、桥涵所处地区的年均降雨量和温度及车道数。基于小修概率的思想,使用机器学习法预测各条高速公路桥涵小修工程总费用,预测结果与实际费用之间的误差较小,表明该方法适用于高速公路桥涵资产的小修工程费用预测,且预测结果能为合理分配养护费用、提高养护决策水平提供参考。

表 8 桥涵小修工程总费用实际值与预测值汇总及分析

Table 8 Summary and analysis of actual and predicted total cost of minor repair project of bridge and culvert

高速名称	年份	实际值/元	预测值/元	绝对误差/元	绝对百分比误差/%	平均绝对百分比误差/%	Wilcoxon 符号秩检验
西长高速	2016	131 678	135 341	13 663	2.78	4.66	0.753
	2017	165 893	151 206	14 687	8.85		
靖王高速	2016	93 287	91 241	2 046	2.19		
	2017	82 210	80 407	1 803	2.19		
商界高速	2016	530 624	568 009	37 385	7.05		
	2017	508 573	483 614	24 959	4.91		



## 参考文献:

## References:

- [1] 陈树礼,刘永前.城市与公路桥梁技术状况评定方法对比分析[J].建筑科学与工程学报,2018,35(3):95-103.  
CHEN Shuli,LIU Yongqian. Comparative analysis of technical condition evaluation of city bridges and highway bridges[J]. Journal of Architecture and Civil Engineering,2018,35(3):95-103.
- [2] FRANGOPOL D M. Life-cycle performance of deteriorating structures; assessment, design, and management[M]. Reston: American Society of Civil Engineers,2003.
- [3] 牛宏亮.城市桥梁养护费用预测模型与方法研究[D].广州:华南理工大学,2009.  
NIU Hongliang. Research on prediction model and method of urban bridge maintenance cost[D]. Guangzhou:South China University of Technology,2009.
- [4] ALI GHAHARI S, VOLOVSKI M, ALQADHI S, et al. Estimation of annual repair expenditure for interstate highway bridges [J]. Infrastructure Asset Management,2019,6(1):40-47.
- [5] JEBREEN J M I, JOHNSTON D W. Bridge maintenance level of service optimization based on an economic analysis approach[R]. Raleigh: North Carolina State University,1995.
- [6] MOLONEY M, MCKENNA T, FITZGIBBON K, et al. Quality data for strategic infrastructure decisions in Ireland [J]. Infrastructure Asset Management,2017,4(2):40-49.
- [7] 张春霞,王巍.基于费用效益最优的桥梁预防性养护时机确定方法研究[J].公路,2013,58(7):259-262.  
ZHANG Chunxia,WANG Wei. Study on the method of determining the opportunity of bridge preventive maintenance based on cost-benefit optimization [J]. Highway,2013,58(7):259-262.
- [8] WOLDEMARIAM W, MURILLO-HOYOS J, LABI S. Estimating annual maintenance expenditures for infrastructure; artificial neural network approach [J]. Journal of Infrastructure Systems, 2016, 22 (2): 04015025.
- [9] 戚云生.高速公路日常养护工程预算定额及费用标准研究[D].西安:长安大学,2010.  
QI Yunsheng. Study on the budget quota and expenses standard of highway daily maintenance and engineering[D]. Xi'an:Chang'an University,2010.
- [10] 乔建刚,侯会学.基于全寿命周期成本的桥梁改造方案优选的研究[J].中外公路,2014,34(5):93-95.  
QIAO Jiangang,HOU Huixue. Research on optimization of bridge reconstruction scheme based on life cycle cost [J]. Journal of China & Foreign Highway, 2014,34(5):93-95.
- [11] 吴宇蒙,常军.基于养护管理系统的既有桥梁生命周期成本模型研究[J].苏州科技学院学报(工程技术版),2010,23(1):44-49,55.  
WU Yumeng, CHANG Jun. Study on in-service bridge life cycle cost model based on maintenance management system[J]. Journal of Suzhou University of Science and Technology (Engineering and Technology edition),2010,23(1):44-49,55.
- [12] 朱岩.陕西省高速公路桥涵小修保养年公里生产消耗指标研究[D].西安:长安大学,2019.  
ZHU Yan. Research on the annual production and consumption of the maintenance project of the highway bridge-culvert in Shaanxi province [D]. Xi'an: Chang'an University,2019.
- [13] 董金爽,隋龔,薛建阳.传统风格建筑混凝土梁-柱节点承载力影响参数灰色关联分析[J].建筑科学与工程学报,2019,36(6):80-87.  
DONG Jinshuang,SUI Yan,XUE Jianyang. Grey relevance analysis of influencing parameters of bearing capacity of concrete beam-column joints in traditional style buildings [J]. Journal of Architecture and Civil Engineering,2019,36(6):80-87.
- [14] 孙作安,王奇,张铁岩,等.分辨系数的作用及其确定方法[J].辽宁工学院学报,1997,17(4):71-72,84.  
SUN Zuoan,WANG Qi,ZHANG Tieyan,et al. Function by resolution factor and its determination method [J]. Journal of Liaoning Institute of Technology, 1997,17(4):71-72,84.
- [15] 汪冬华.多元统计分析与SPSS应用[M].上海:华东理工大学出版社,2010.  
WANG Donghua. Multivariate statistical analysis and SPSS application[M]. Shanghai:East China University of Science and Technology Press,2010.
- [16] 陕西省统计局,国家统计局陕西调查总队.陕西统计年鉴2012[M].北京:中国统计出版社,2012.  
Shaanxi Provincial Bureau of Statistics,Survey Office of the National Bureau of Statistics in Shaanxi. Shanxi statistical yearbook 2012[M]. Beijing:China Statistics Press,2012.
- [17] 伍德里奇.计量经济学导论:现代观点[M].3版.北京:清华大学出版社,2007.  
WOOLDRIDGE J M. Introductory econometrics: a modern approach[M]3rd ed. Beijing: Tsinghua Uni-

versity Press,2007.

[18] 公路沥青路面设计规范:JTG D50—2017[S]. 北京:人民交通出版社,2017.

Specifications for design of highway asphalt pavement:JTG D50—2017[S]. Beijing:China Communications Press,2017.

[19] 陈永秀. 相关系数含义的理解[J]. 中国考试,2011(7):15-19.

CHEN Yongxiu. Methods for calculating the correlation coefficient[J]. Journal of China Examinations, 2011(7):15-19.

[20] 姚弘智,钟宝荣. 基于机器学习线性回归的正规方程分析[J]. 电脑迷,2018(3):208.

YAO Hongzhi,ZHONG Baorong. Normal equation analysis based on machine learning linear regression[J]. PC Fan,2018(3):208.

[21] 刘 哲,乌 伟,张善文,等. 基于多元岭回归估计小麦种植密度[J]. 中国农业科技导报,2020,22(6):72-80.

LIU Zhe,WU Wei,ZHANG Shanwen,et al. Estimation of wheat planting density based on multivariate ridge regression[J]. Journal of Agricultural Science and Technology,2020,22(6):72-80.

[22] 张沥今,魏夏琰,陆嘉琦,等. Lasso 回归:从解释到预测[J]. 心理科学进展,2020,28(10):1777-1791.

ZHANG Lijin,WEI Xiayan,LU Jiaqi,et al. Lasso regression:from explanation to prediction[J]. Advances in Psychological Science,2020,28(10):1777-1791.