

基于 ARIMA 和 SVM 模型的沥青路面 使用性能预测比较

张丽娟¹, 杨源¹, 梅诚¹, 许薛军²

(1. 华南理工大学 土木与交通学院, 广东 广州 510640; 2. 广东省交通运输规划研究中心, 广东 广州 510101)

摘要:通过构建时间序列 ARIMA 和支持向量机(SVM)模型对沥青路面的 *PCI* 值进行预测,并比较两种模型的预测结果,建立适宜的沥青路面使用性能预测模型。分别采用线性核、多项式核、径向基核和 Sigmoid 核函数构建 SVM 模型,根据交叉验证(Cross-Validation)方法确定核函数及其参数,对比不同核函数的 *PCI* 值预测误差精度。最后利用选取的多项式核函数 SVM 算法预测广东省普通国省道沥青路面使用性能。结果表明:采用多项式核、径向基核函数 SVM 算法的预测结果较 ARIMA 模型预测结果精度高。SVM 算法中核函数对预测结果影响显著,4 种核函数 SVM 预测模型中,基于多项式核函数模型预测精度最高,其预测沥青路面使用性能的平均绝对百分误差(MAPE)在 1% 以下,其次为径向基核函数,而 Sigmoid 核函数预测精度最低。采用线性核函数 SVM 算法预测结果与 ARIMA 模型预测结果接近,预测精度满足要求。

关键词:道路工程; 沥青路面使用性能; 预测; ARIMA; SVM; 核函数

随着行车荷载急速增长和使用年限不断增加,沥青路面的损坏愈来愈严重,准确预测和评价沥青路面使用性能衰变趋势,是判断养护时机、制定养护规划的前提^[1]。目前中国公路管理部门构建了多个公路养护信息管理平台,进一步完善了中国公路基础数据库,应利用大数据理论及技术对这些数据进行挖掘分析,建立有效的沥青路面使用性能预测模型,可为科学制定路面养护管理决策提供理论支撑。

路面使用性能数据是定期检测所获取的按时间先后顺序排列所形成的时间序列,通过曲线拟合和参数估计建立时间序列模型,可以定量分析和预测路面使用性能发展趋势。武建民等^[2]、法鲁克·铁来克^[3]分别采用基于时间序列分析法的 ARIMA 模型对路面行驶使用性能指数 *RQI*、沥青路面损坏状况指数 *PCI* 进行预测,发现 ARIMA 模型预测精度在 5% 以内,预测精度较高。

沥青路面使用性能受自然环境、交通荷载、结构、材料、维修养护等多种非线性因素综合影响。支持向量机(SVM)是在处理小样本非线性问题方面具有特

有优势的一种机器学习算法,SVM 算法所采用的核函数及其参数直接影响到回归预测精度^[4]。黄冰等^[5]采用多项式核函数建立了路面破损状况 *PCI* 的 SVM 预测模型;刘黔会等^[6]利用径向基核函数建立沥青路面使用性能的 SVM 预测模型;董锴等^[7]建立采取径向基核函数的 SVM 路面评价模型进行路面性能综合评价。

针对沥青路面使用性能具有样本小、维数高与非线性等特点,有学者将 SVM 和其他方法相结合以提高 SVM 预测精度。李巧茹等^[8]结合主成分分析法和支撑向量机理论,建立了 PCA-SVM 沥青路面使用性能评价模型;李海莲等^[9]构建了改进萤火虫算法和支撑向量机相结合的 IFA-SVM 预测模型;赵静等^[10]提出了基于灰色关联度分析和支撑向量机回归(GRA-SVR)的沥青路面使用性能预测模型。

时间序列模型和支撑向量机模型预测技术各有其特点,应根据路面性能变化规律采用适宜的沥青路面性能预测模型。杨彦海等^[11]分别利用 BP 神经网络时间序列模型和支撑向量机模型对就地热再生沥青路面

收稿日期:2020-03-20

基金项目:国家自然科学基金资助项目(编号:51808228);广东省交通运输厅科技项目(编号:科技 2017-02-003);广东省公路事务中心科研课题(编号:粤公研 2017-15)

作者简介:张丽娟,女,博士,副教授。E-mail:tczljuan@scut.edu.cn

中沥青老化指标衰变进行预测,发现在有限数据量情况下支持向量机模型较BP神经网络时间序列模型的预测精度高。

该文分别构建时间序列ARIMA和支持向量机(SVM)模型,且SVM模型分别采用线性核、多项式核、径向基核和Sigmoid核4种核函数对沥青路面的PCI值进行预测,对比分析ARIMA和SVM模型及SVM模型所采用不同核函数的预测结果,并将所选择的预测模型应用于广东省普通国省道沥青路面使用性能预测中。

1 ARIMA和SVM模型预测原理

1.1 ARIMA模型预测原理

差分自回归移动平均模型(ARIMA模型)是最常用的一种时间序列分析法模型。ARIMA(p, d, q)模型中,AR、MA分别为自回归模型、移动平均模型, p 、 q 为相应的自回归模型阶数、移动平均模型阶数, d 为将非平稳时间序列进行平稳化处理的差分阶数。ARIMA模型是差分运算和ARMA模型的结合,其用一定的数学模型来近似描述预测对象随时间变化又相互关联的数据序列,模型检验合理后,就可以通过时间序列的过去值和现在值来预测未来某时段的价值。

设 $\{x_t, t \in T\}$ 为时间序列, B 为后移算子,即 $Bx_t = x_{t-1}$,则ARIMA(p, d, q)模型可表示为^[2,12]:

$$\Phi_p(B)(1-B)^d x_t = \theta_q(B)a_t \quad (1)$$

其中:

$$\Phi_p(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p \quad (2)$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (3)$$

式中: Φ 和 θ 分别为自回归系数和移动平均系数; $\Phi_p(B)$ 为自回归算子; $\theta_q(B)$ 为移动平均算子; a_t 为白噪声时间序列。

ARIMA模型的建模步骤一般为^[13]:

(1) 时间序列平稳性检验。观察初始时间序列的线性趋势、周期性、方差齐性等特点,采用ADF检验方法,对初始时间序列进行单位根检验。将非平稳时间序列进行 d 阶差分处理为平稳时间序列。

(2) 对平稳时间序列进行模式识别,确定模型的阶数。采用自相关系数和偏自相关系数初步确定阶数 p 、 q 可能值,根据BIC准则(即BIC值越小越好)得到相对较优的阶数 p 和阶数 q 。

(3) 模型的检验。通过Q统计量检验残差是否为白噪声,判断预测模型的合理性。

(4) 模型预测。根据检验合理的ARIMA模型进行预测分析。

1.2 SVM模型预测原理

支持向量机(SVM)是一种通过有限的样本信息进行数据挖掘、建立逼近真实问题模型的机器学习算法,其在小样本、非线性、高维数的回归预测领域优势明显,有效地避免了过学习、欠学习等现象。

SVM将样本空间中的非线性数据通过内积核函数映射到高维特征空间,并在新的高维特征空间中构造回归函数 $f(x)$,从而实现样本空间中的非线性回归分析。

设样本集: $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l), x_i \in R^n$ 为输入变量, $y_i \in R$ 为输出值, R 为实数集。寻求的支持向量机回归函数为^[5,9]:

$$f(x) = \omega \cdot \phi(x) + b \quad (4)$$

式中: ω 为权重向量; b 为偏置量; $\phi(x)$ 为映射函数。

根据结构风险最小化原则,在最小目标 $\frac{1}{2} \|\omega\|^2$

中加入惩罚项 $C \sum_{i=1}^l (\zeta_i + \zeta_i^*)$,形成目标函数:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\zeta_i + \zeta_i^*) \quad (5)$$

$$\omega \cdot \phi(x_i) + b - y_i \leq \epsilon + \zeta_i, i = 1, 2, \dots, l$$

$$\text{s. t. } \{y_i - \omega \cdot \phi(x_i) - b \leq \epsilon + \zeta_i^*, i = 1, 2, \dots, l$$

$$|\zeta_i, \zeta_i^*| \geq 0$$

(6)

式中: C 为惩罚系数且大于0; ϵ 为拟合精度且大于0; ζ_i, ζ_i^* 为松弛变量。

采用拉格朗日乘子法将式(5)的最优化问题转换成对偶问题进行求解,即:

$$L = \frac{1}{2} \omega \times \omega + C \sum_{i=1}^l (\zeta_i + \zeta_i^*) - \sum_{i=1}^l \alpha_i [\zeta_i + \epsilon - y_i + f(x_i)] - \sum_{i=1}^l \alpha_i^* [\zeta_i^* + \epsilon - y_i + f(x_i)] - \sum_{i=1}^l (\zeta_i \eta_i + \zeta_i^* \eta_i^*) \quad (7)$$

$$\left[\max \left[-\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(x_i \cdot x_j) - \right. \right.$$

$$\left. \sum_{i=1}^l (\alpha_i + \alpha_i^*) \epsilon + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i \right]$$

$$\text{s. t. } \left\{ \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \right.$$

$$[\alpha_i, \alpha_i^* \in [0, C]$$

(8)

式中: $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ 为 Lagrange 因子,且都大于 0。

引入核函数,得到变形后的回归函数方程为:

$$f(x)=\omega \cdot \phi(x)+b=\sum_{i=1}^l(\alpha_i-\alpha_i^*)K(x_i,x)+b$$

(9)

式中: $K(x_i,x)$ 为核函数。

常用的满足 Mercer 条件的核函数有^[4-5]:

(1) 线性(Linear)核函数: $K(x_i,x)=(x_i \cdot x)$

(2) 多项式(Polynomial)核函数:

$$K(x_i,x)=[\gamma(x_i \cdot x)+r]^d$$

(3) 径向基(RBF)核函数:

$$K(x_i,x)=\exp(-\gamma\|x_i-x\|^2)$$

(4) Sigmoid 核函数:

$$K(x_i,x)=\tanh[\gamma(x_i \cdot x)+r]$$

式中: γ, r, d 为待定参数。

在支持向量机预测过程中,核函数和惩罚参数 C 的取值直接影响预测精度。支持向量机学习算法本质上是基于核函数的学习算法。

2 ARIMA 模型和 SVM 模型预测 PCI 值对比

2.1 ARIMA 模型预测 PCI 值

广东佛山一环北延线路段采用沥青路面结构,表 1 为该路段在 2011—2015 年的 PCI 、 RQI 、 RDI (路面车辙深度指数)、 SRI 等沥青路面使用性能评价指标数据。

表 1 佛山一环北延线路段路面使用性能评价数据

评价年份/年	PCI	RQI	RDI	SRI
2011	86.7	88.4	85.8	87
2012	83.5	86.1	83.2	85
2013	82.1	84.3	82.2	83
2014	81.2	82.6	81.4	82
2015	80.6	80.2	79.5	81

以表 1 中的 PCI 值预测对比分析 ARIMA 模型和 SVM 模型在沥青路面使用性能预测中的应用。

建立 ARIMA 模型进行数据分析时,需要通过调整模型参数以满足精度要求。针对表 1 的 PCI 指标数据,采用 SPSS 软件建立了 ARIMA(1,0,1)模型,表 2、3 为模型拟合参数。

从表 2、3 可见:ARIMA(1,0,1)模型平稳的 R^2 、 R^2 分别为 0.913 和 0.914,说明 ARIMA(1,0,1)模型

表 2 ARIMA(1,0,1)模型拟合参数

拟合统计量	PCI 值			
	平稳的 R^2	R^2	均方根误差 RMSE	标准的 BIC
均值	0.913	0.914	1.428	1.999
最大值	0.913	0.914	1.428	1.999
最小值	0.913	0.914	1.428	1.999

表 3 ARIMA(1,0,1)模型输出参数

类别	估计	标准误差 SE	t	Sig.
PCI 值—模型_(1,0,1)	9.349	0.090	103.564	0.006
1 阶滞后项	0.222	3.248	0.068	0.956
0 阶滞后项	-0.085	0.036	-2.382	0.253

适用于拟合 PCI 数据序列,但均方根误差 $RMSE$ 较大,说明 ARIMA(1,0,1)模型的拟合精度不够高,且 t 统计(Sig. 列)的显著性不明显。因此,需调整 ARIMA(1,0,1)模型的移动平均数,将模型调整为 ARIMA(1,0,0),表 4、5 为改进后的 ARIMA(1,0,0)模型参数。

表 4 ARIMA(1,0,0)模型拟合参数

拟合统计量	PCI 值			
	平稳的 R^2	R^2	均方根误差 RMSE	标准的 BIC
均值	0.892	0.892	1.129	1.208
最大值	0.892	0.892	1.129	1.208
最小值	0.892	0.892	1.129	1.208

表 5 ARIMA(1,0,0)模型输出参数

类别	估计	标准误差 SE	t	Sig.
PCI 值—模型_(1,0,0)	87.062	1.121	77.690	0.000
1 阶滞后项	-0.197	0.846	-5.233	0.037
0 阶滞后项	-1.433	0.055	-4.039	0.056

从表 4、5 可以看出:ARIMA(1,0,0)模型保持了较大、较平稳的 R^2 ,均方根误差 $RMSE$ 减小,模型拟合度高,且 t 统计(Sig. 列)的显著性明显。ARIMA(1,0,0)模型的 BIC 值也较 ARIMA(1,0,1)模型的 BIC 值小,一般而言,选取 BIC 值小的模型为最佳模型。因此,选用 ARIMA(1,0,0)模型作为预测模型,佛山一环北延线沥青路面的 PCI 值预测结果如表 6 所示。

表6 ARIMA(1,0,0)模型PCI预测值及置信区间

路龄/ 年	PCI 实测值	PCI 预测值	置信上 限UCL	置信下 限LCL	相对误差 率/%
1	86.7	85.63	90.54	80.72	1.23
2	83.5	83.98	88.80	79.17	0.57
3	82.1	82.90	87.71	78.08	0.97
4	81.2	81.46	86.27	76.64	0.32
5	80.6	79.92	84.74	75.11	0.84

从表6可见:ARIMA模型预测PCI值的精度比较高,相对误差率在1.23%以下。

2.2 SVM模型预测PCI值

以佛山一环北延线路段沥青路面结构的PCI指标数据(表1)为例,利用Matlab系统和libSVM软件包进行SVM回归预测仿真分析。核函数的选择是构造具有良好性能SVM模型的关键^[4],通常可以利用专家的经验预先选定核函数;或从试用的不同核函数中选取误差最小的核函数作为最佳核函数,即Cross-Validation方法。

采用均方误差MSE和相关系数平方SCC指标来评价预测精度,从线性核函数、多项式核函数、径向基核函数和Sigmoid核函数中,通过Cross-Validation方法选取最优的核函数构建SVM预测模型。

表7、8为4种核函数的SVM模型PCI值预测结果及其精度评价。

表7 SVM模型PCI值预测结果

路龄/ 年	PCI 实测值	PCI 预测值			
		线性核 函数	多项式 核函数	径向基 核函数	Sigmoid 核函数
1	86.7	85.133	86.300	86.300	67.388
2	83.5	83.900	83.900	83.900	74.870
3	82.1	82.667	81.916	82.500	82.100
4	81.2	81.433	80.799	81.600	83.019
5	80.6	80.200	81.000	81.000	83.136

表8 SVM模型PCI值预测精度评价

参数	线性 核函数	多项式 核函数	径向基 核函数	Sigmoid 核函数
MSE	0.630	0.135	0.160	977.659
SCC	0.891	0.993	0.994	0.877

从表7、8可以看出:选择不同核函数会产生不同的预测结果,核函数对预测结果影响显著。沥青路面

使用性能是总体平稳、局部波动的序列,4种核函数中,多项式核函数预测的PCI值变化趋势与实测值最接近,预测效果最佳;径向基核函数预测结果与多项式核函数预测结果相近,预测结果较理想;线性核函数预测效果基本满足预测精度要求;Sigmoid核函数预测精度最差,表明Sigmoid核函数在沥青路面使用性能小样本的识别效果比较差。

2.3 ARIMA模型和SVM模型预测结果对比

选择平均绝对误差(MAE)、均方根误差(RMSE)和平均绝对百分误差(MAPE)作为对比ARIMA模型和SVM模型预测精度指标。根据表6、7的PCI值预测结果,计算出预测误差指标值见表9,图1为ARIMA模型PCI预测结果与线性核函数、多项式核函数的SVM预测结果对比。

表9 ARIMA和SVM模型预测结果评价

预测方法	MAE	RMSE	MAPE/%
ARIMA模型	0.658	0.713	0.789
线性核函数	0.633	0.794	0.752
SVM模型 多项式核函数	0.357	0.367	0.431
高斯核函数	0.400	0.400	0.483

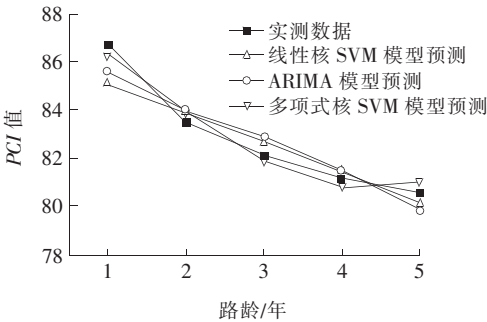


图1 ARIMA和线性核函数、多项式核函数SVM模型PCI预测结果对比

从表9和图1可以看出:ARIMA模型与线性核函数SVM模型预测的PCI变化趋势几乎一致,预测精度非常接近。ARIMA模型本质上是基于线性的模型^[13],路面使用性能PCI值的时间序列数据往往包含非线性关系,因此ARIMA模型与线性核函数SVM模型预测的PCI值预测精度相对不高;多项式核函数、径向基核函数预测的PCI值比ARIMA模型预测结果精度高,多项式核函数预测的PCI值变化趋势更接近于实测值变化趋势,故可采用多项式核函数SVM模型进行路面使用性能预测。

2.4 SVM 算法在沥青路面使用性能预测中的应用

针对广东省多条不同自然气候条件、交通量、结构层厚度及材料的普通国省道干线,从中统计出具有代表性的典型沥青路面结构路段在 2011—2015 年间的 130 组 *PCI*、*RQI*、*RDI*、*SRI* 等路面使用性能评价分项指标检测数据,利用多项式核函数 SVM 模型分别预测对应的 *PCI*、*RQI*、*RDI*、*SRI* 值。

表 10 为各评价指标值预测结果评价表。

表 10 ARIMA 和 SVM 模型预测结果评价

评价指标	MAE	RMSE	MAPE/%
<i>PCI</i>	0.555	1.488	0.605
<i>RQI</i>	0.387	1.148	0.437
<i>RDI</i>	0.376	0.515	0.427
<i>SRI</i>	0.429	1.002	0.491

从表 10 可见:*PCI*、*RQI*、*RDI*、*SRI* 值预测结果平均绝对误差(MAE)、均方根误差(RMSE)和平均绝对百分误差(MAPE)最大值分别为 0.555、1.488 和 0.605%,满足预测精度要求,体现了多项式核函数 SVM 模型在处理小样本非线性预测中的优越性。

3 结论

分别对比分析了基于时间序列 ARIMA 和支持向量机(SVM)模型的沥青路面使用性能预测结果,得出结论:

(1) 沥青路面使用性能数据是定期检测所获取的时间序列,采用时间序列 ARIMA 模型预测 *PCI* 值的精度比较高,相对误差率在 1.23%以下。采用线性核函数的 SVM 算法预测结果与 ARIMA 模型预测结果很接近。

(2) 沥青路面使用性能数据少、影响路面性能衰减趋势的因素具有不确定和非线性^[14],在有限数据量情况下,采用多项式核、径向基核函数 SVM 算法的 *PCI* 值预测结果较 ARIMA 模型预测结果精度高,体现出 SVM 模型在处理小样本非线性问题方面具有特有的优势。

(3) SVM 算法中核函数对预测结果影响显著。4 种核函数 SVM 预测模型中,基于多项式核函数模型预测精度最高,其次为径向基核函数,而 Sigmoid 核函数预测精度最低。采用多项式核函数 SVM 算法预测

广东省普通国省道沥青路面使用性能的平均绝对百分误差在 1%以下,满足工程实际需求。

(4) 影响沥青路面使用性能的因素非常复杂,应结合其他算法,建立考虑多因素、长周期的 SVM 预测综合模型,为准确预测沥青路面使用性能、制定科学的沥青路面养护决策提供参考。

参考文献:

[1] 周育名,李金明,李平,等.考虑公路等级的沥青路面性能衰变预测方法[J].中外公路,2019,39(1):34—39.

[2] 武建民,刘大彬,李福聪,等.基于时间序列分析法的沥青路面使用性能预测[J].长安大学学报(自然科学版),2015,35(3):1—7.

[3] 法鲁克·铁来克.基于 ARIMA 模型的橡胶沥青路面状况衰变预测研究[J].北方交通,2019(9):79—82.

[4] 宋晖,薛云,张良均.基于 SVM 分类问题的核函数选择仿真研究[J].计算机与现代化,2011(8):133—136.

[5] 黄冰,颜可珍,张邹.基于支持向量机的沥青路面使用性能评价[J].公路工程,2012,37(2):19—21.

[6] 刘黔会,张挣鑫,黄方林,等.基于支持向量机的沥青路面使用性能预测探究[J].公路工程,2018,43(2):201—205.

[7] 董锴,王瑜.基于 SVM 的道路路面性能评估分析与应用[J].公路工程,2019,44(1):179—183.

[8] 李巧茹,郭知洋,王耀军,等.基于 PCA—SVM 的高速公路沥青路面使用性能评价[J].北京工业大学学报,2018,44(2):283—288.

[9] 李海莲,林梦凯,王起才,等.基于 IFA—SVM 的高速公路沥青路面使用性能预测[J].公路交通科技,2019,36(12):8—14.

[10] 赵静,王选仓,丁龙亭,等.基于灰色关联度分析和支持向量机回归的沥青路面使用性能预测[J].重庆大学学报,2019,42(4):72—81.

[11] 杨彦海,张东旭,董帅,等.BP 时间序列与支持向量机对就地热再生沥青路面沥青指标老化预测对比分析[J].公路,2018,63(2):231—237.

[12] 赵鹏,李璐.基于 ARIMA 模型的城市轨道交通进站量预测研究[J].重庆交通大学学报(自然科学版),2020,39(1):40—44.

[13] 罗志丹,刘英,郭伟.基于跟踪微分器的泰勒展开与 ARIMA 混合模型对股票价格的短期预测[J].数学的实践与认识,2019,49(23):67—77.

[14] 杨三强,饶奇,史向英.不同老化程度下沥青流变力学性能[J].长沙理工大学学报(自然科学版),2019,16(1):22—27.