

# 基于随机森林的沥青路面性能预测

张金喜<sup>1</sup>, 郭旺达<sup>1</sup>, 宋波<sup>2</sup>, 褚永昌<sup>1</sup>, 张阳光<sup>1</sup>

(1. 北京工业大学城市交通学院交通工程北京市重点实验室, 北京 100124;

2. 北京市道路工程质量监督站, 北京 100076)

**摘要:** 为弥补路面性能预测问题中传统预测方法效率差、预测精度低等缺陷, 针对普通公路沥青路面状况指数 (pavement condition index, PCI) 的预测问题, 利用北京市 9 个县区包括路面性能指标、路面结构、交通参数及气象资料在内的 1 249 组观测值, 提出了一种基于随机森林算法的 PCI 预测模型, 并与神经网络、支持向量机模型预测结果进行对比. 研究表明: 通过对比分析不同模型的 3 个定量评价指标 (均方根误差、平均绝对误差和决定系数) 以及可视化散点图, 证明采用随机森林算法的 PCI 预测模型的鲁棒性、准确性要优于神经网络和支持向量机模型, 验证了该模型的有效性和优越性, 可以为后续公路养护预算申请和决策方案制定提供科学依据, 对于提高公路养护的经济效益具有重要意义.

**关键词:** 道路工程; 沥青路面; 路面性能预测; 路面状况指数 (PCI); 机器学习; 随机森林 (RF)

中图分类号: U 418. 6

文献标志码: A

文章编号: 0254 - 0037 (2021) 11 - 1256 - 08

doi: 10. 11936/bjutxb2020020010

## Performance Prediction of Asphalt Pavement Based on Random Forest

ZHANG Jinxi<sup>1</sup>, GUO Wangda<sup>1</sup>, SONG Bo<sup>2</sup>, ZHUO Yongchang<sup>1</sup>, ZHANG Yangguang<sup>1</sup>

(1. Beijing Key Laboratory of Traffic Engineering, College of Metropolitan Transportation,

Beijing University of Technology, Beijing 100124, China;

2. Beijing Board Engineering Quality Supervision Station, Beijing 100076, China)

**Abstract:** To make up the poor efficiency and low prediction accuracy of traditional prediction methods in the pavement performance prediction problem, this study aimed at the prediction of the pavement condition index (PCI) of the ordinary highway asphalt pavement, and then proposed a PCI prediction model based on random forest algorithm by using the performance indexes, pavement structure, traffic parameters and meteorological data of 1 249 group observations of 9 counties in Beijing. The prediction results were compared with the neural network and support vector machine models. Results show that by comparing and analyzing three quantitative evaluation indexes (root mean square error, mean absolute error and  $r$ -square) of different models and visual scatter plots, the robustness and accuracy of the PCI prediction model using random forest algorithm are superior to the neural network and the support vector machine models. It can provide scientific basis for subsequent highway maintenance budget application and decision-making plan formulation, which is of great significance for improving the economic benefits of highway maintenance.

**Key words:** road engineering; asphalt pavement; pavement performance prediction; pavement condition index (PCI); machine learning; random forest (RF)

收稿日期: 2020-02-13

基金项目: 国家自然科学基金资助项目(51778027); 北京市交通行业科技项目(2018-kjc-01-213)

作者简介: 张金喜(1965—), 男, 教授, 博士生导师, 主要从事道路工程、道路与交通安全方面研究. E-mail: zhangjinxi@bjut.edu.cn

目前,我国交通基础设施建设日趋完善,道路工程由“建”转“养”已成为当下发展的必然趋势。路面性能预测是路面养护与管理的重要内容,精确地把握沥青路面关键性能指标的衰变情况,有利于针对最佳的养护时机,制定科学的养护决策方案,以发挥路面最大的服役效益<sup>[1]</sup>。路面状况指数(pavement condition index, PCI)作为路面技术状况评定的关键指标之一,用来评价路面的完好程度。因此如何对沥青路面 PCI 进行更加准确、高效的预测,已经成为路面养护管理领域研究的重要课题。

沥青路面性能衰变是路面在实际服役过程中对引起其原有路面结构或功能改变的因素的反馈。这些因素包括路面结构、交通荷载、自然环境、养护干预等,冗杂繁多,因此表征路面性能的性能指标与其影响因素之间的非线性关系是极其复杂的。针对路面性能预测问题,国内外众多学者研究方法不一。路面性能预测的经典模型是确定型模型,杜艳花<sup>[2]</sup>、张金喜等<sup>[3]</sup>通过总结高速公路路面性能指标随时间的变化特点,采用指数型函数对历史数据进行拟合,分别用于预测国际平整度指数(international roughness index, IRI)和路面 PCI 的衰变情况;肖金平等<sup>[4]</sup>在分析不同路面结构的基础上,建立了路面性能指标的余弦函数型衰变方程,计算结果与实测结果一致性较好。然而经典研究方法是一般只考虑性能指标关于时间序列的经验预测模型,未能建立路面性能指标与其影响因素之间的非线性关系,与实际路面性能衰变情况匹配度相差较大,并且预测精度随着预测周期的推移会逐渐降低。近年来随着机器学习受到越来越多的关注,针对这种非线性的路面性能预测问题便有了新的研究方向。诸如人工神经网络(artificial neural network, ANN)、支持向量机(support vector machine, SVM)等人工智能方法在路面性能预测领域的出现,解决了传统线性预测模型无法应对性能指标突变的问题。Satish 等<sup>[5]</sup>基于人工神经网络建立平整度预测模型,虽然预测结果优于线性回归模型,但是由于样本集不足导致局部出现过拟合的现象;赵静等<sup>[6]</sup>将灰色理论与支持向量机模型相结合,综合考虑路龄、交通荷载等因素建立车辙深度指数预测模型,在项目级预测中验证了模型的精度,考虑到支持向量机模型易于表征小样本情况下的非线性映射关系,对于路网级预测的效果仍有待验证。

Breiman 提出的随机森林(random forest, RF)模型<sup>[7]</sup>,可以有效地处理多因素耦合影响下的非线性回归问题。相较其他机器学习模型,RF 的鲁棒性

好、预测能力强。RF 不仅克服了传统经验模型在处理复杂交互作用的不足,而且与神经网络相比具有更强的抗噪和抗过拟合的能力。此外通过对比选取支持向量机核函数等参数复杂程度,构建 RF 模型的超参数更容易理解。RF 算法已经成功被应用到许多研究领域<sup>[8-12]</sup>。但目前该算法在道路工程领域还未得到充分应用。基于以上分析,本文提出一种基于 RF 的路面性能预测方法,并与神经网络、支持向量机预测模型的预测能力进行对比,以期找到一种路面状况指数的精确预测方法来指导路面养护管理工作。

## 1 研究方法

RF 算法是一种在决策树的基础上建立的集成学习算法,可以用于解决分类和回归问题。由于其泛化能力强、抗噪性能好,因此在地理<sup>[8]</sup>、农业<sup>[9]</sup>、农林学<sup>[10-11]</sup>、生物学<sup>[12]</sup>等领域都已经取得了成功的应用和推广。RF 对多重共线性不敏感,所以缺失数据和不平衡数据对预测结果的不良影响很小,最多可用数千个特征变量进行准确预测。现阶段 RF 算法被视作是构建非线性预测模型的最优算法之一。

### 1.1 决策树回归算法

决策树是 RF 算法的理论基础,在此对决策树算法进行简要介绍。决策树作为识别重要变量和 2 个及以上特征变量关系最快的方法之一,可以用于解决分类和回归问题。通常一棵决策树由根节点出发,经由内部节点逐层生长,最终在叶子节点处触发生长阈值而停止生长。其中根节点代表整个总体,决策树通过使用二分递归分割,将样本集合通过相关策略拆分成 2 个子节点<sup>[13]</sup>。经过逐步拆分,最终由叶子节点输出预测结果。假设存在一个同时包含预测变量和特征变量的训练数据集  $D$ ,构造决策树回归算法流程如下。

1) 遍历全部特征后,找到最优划分特征和分割点 $(x^{(j)}, s)$ ,满足均方误差  $MSE(x^{(j)}, s)$  最小,然后划分为 2 个节点。

2) 对生成的 2 个节点重复调用 1),直至满足模型预设的停止生长条件(比如设置树深、叶子节点包含的最少样本数等)。

3) 根据最终生成的所有叶子节点,得到决策树回归器

$$h(x) = \sum_{m=1}^n c_m I(x \in R_m) \quad (1)$$

式中: $R_m$  为叶子节点的数量; $c_m$  为  $m$  点处的输出

值; $I$ 为指示函数.

### 1.2 RF 基础理论

RF 是以决策树为基学习器通过集成方式构建而成的有监督机器学习方法<sup>[14]</sup>, 进一步在决策树的训练过程引入了随机性, 使其具备优良的抗过拟合以及抗噪能力. RF 分别从样本选取和特征选择 2 个角度体现其随机性<sup>[15]</sup>.

1) 随机选取样本: RF 中每一棵决策树的样本集均是从原始数据集中采用 Bootstrap 策略有放回地抽取、重组形成与原始数据集等大的子集合. 这就意味着同一个子集里面的样本可以是重复出现的, 不同子集中的样本也可以是重复出现的.

2) 随机选取特征: 不同于单个决策树在分割过程中考虑所有特征后, 选择一个最优特征来分割节点. RF 通过在基学习器中随机考察一定的特征变量, 之后在这些特征中选择最优特征<sup>[16]</sup>. 特征变量考察方式的随机性使得 RF 模型的泛化能力和学习能力优于个体学习器.

### 1.3 RF 回归算法

在回归问题中, 随机森林回归 (random forest regression, RFR) 模型在表征一组输入和输出之间高度非线性关系时表现出强大的学习能力. RFR 算法流程如下:

1) 组建训练集  $D$ , 采用 Bootstrap 策略从训练集  $D$  中有放回地重复取样, 组成  $T$  个与训练集  $D$  容量大小相等的新训练集  $D_i (1 \leq i \leq T)$ , 即  $T$  个互相独立的回归决策树.

2) 在回归树  $i$  的  $M$  个特征中随机考察  $m$  个子特征, 遍历每一个抽样子特征后找到最优划分特征和分割点, 满足均方误差 MSE 最小, 然后划分为 2 个节点并且不做剪枝处理.

3) 对生成的 2 个节点重复调用步骤 2), 直至满足模型预设的停止生长条件, 决策树  $h_i$  生长完成.

4) 集结所有独立生长完成的回归决策树随即生成 RF 回归器

$$H(x) = \frac{1}{T} \sum_{i=1}^T h_i(x) \quad (2)$$

## 2 基于 RF 的路面性能预测

RF 作为一种结构复杂、实现简单的集成学习算法, 无须假设变量服从正态性等条件下即可完成对复杂数据集的分类和回归建模<sup>[17]</sup>, 且对数据中的噪声和异常值的包容程度较高, 能够有效地处理复杂的非线性交互作用, 在很多领域的预测问题中都被证明了是一种有效的预测方法, 但是在道路工程中的应用寥寥无几. 由于本研究中预测变量与相关解释变量均为连续型的数值变量, 因此路面性能预测本质上是一种非线性的回归问题. 鉴于此, 本研究基于 PCI 数据组建训练 RFR 模型的样本集, 并通过对比 RFR 模型较路面性能预测中常用机器学习模型 (ANN、SVM) 的预测能力, 探讨 RFR 模型在沥青路面性能预测中的应用. 本文提出技术路线如图 1 所示.

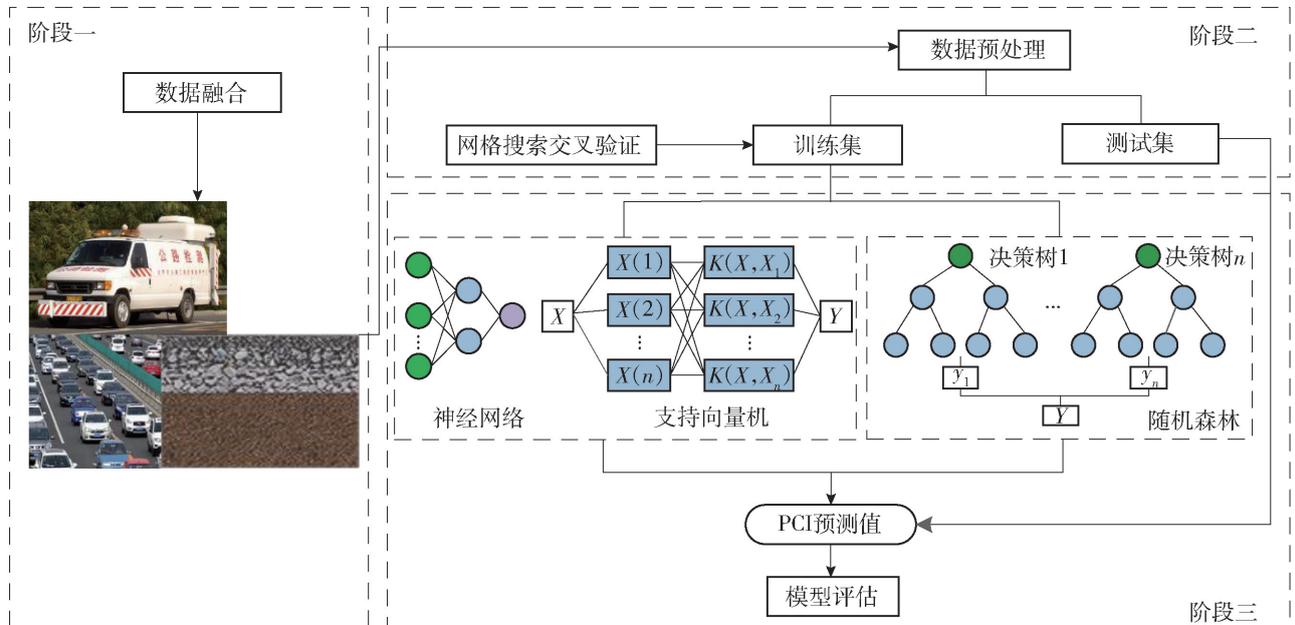


图1 技术路线

Fig. 1 Technical route

## 2.1 数据准备

特征变量相关描述如表 1 所示. 本文以北京市沥青路面为研究对象,从相关部门收集整理了 2010—2016 年的道路检测数据,气候相关数据为北京市包括通州、密云在内的 9 个县区 7 年的气象数据,交通参数、路面结构相关数据来自相关单位建养计划. 通过调查发现,路面性能检测时间可能会与养护工期冲突,因此造成目标年 PCI 值的缺失. 此外在养护干预下路面结构、材料的变化会导致路面性能变化趋势突变. 养护干预对

于路面性能衰变影响重大,为排除此项因素干扰,本研究在进行数据筛选时选取的路段均为预测目标 3 年内发生过养护干预. 本研究将 PCI-1、AADT、AGE、PT、BT、ST、AP、AAT、SD 作为模型的输入变量,PCI 作为模型的目标输出变量. 根据路面性能预测相关数据的收集以及参考国内外文献中进行路面性能预测工作提取的特征变量,确定组建容量为  $10 \times 1\ 249$  的样本集用于构建路面性能预测的 RF 模型,其中训练集与测试集的划分比例为 7:3.

表 1 特征变量相关描述  
Table 1 Description of the feature variables

编号	特征变量	符号	单位	描述
1	路面状况指数	PCI		路面损坏状况评价指标
2	目标前 1 年路面状况指数	PCI-1		观测年前 1 年的路面状况指数
3	年平均日交通量	AADT	pcu/d	全年的日交通量观测结果的平均值
4	路龄	AGE	a	自通车年限起至观测年止
5	面层厚度	PT	mm	沥青路面面层总厚度
6	基层厚度	BT	mm	沥青路面基层总厚度
7	底基层厚度	ST	mm	沥青路面底基层总厚度
8	年降水量	AP	mm	该地区年度降雨量总和
9	年平均气温	AAT	℃	全年各日的日平均气温的算术平均值
10	日照时数	SD	h	地面单日累计承受太阳光线直射的时间

## 2.2 预测模型的构建及调优

本研究基于 RF 的路面性能预测模型的开发在 Python 3.5 平台上实现. 在训练模型过程中,必须定义对应于 RF 算法的超参数. 在 RF 模型的构建过程中,为了使预测模型误差最小,采用网格搜索和十折交叉验证的方法对路面性能预测模型中关键的超参数进行优化,搜索结果如图 2 所示. 其中  $R^2$  平均值和  $R^2$  标准差分别表示十折交叉验证下模型性能平均得分和稳定性,前者越大说明模型性能得分越高,后者越小说明模型稳定性越好. 首先要确定的 RF 中决策树的数量  $n\_estimators$ ,从图 2(a)可以看出随着决策树的数量由 1 增加到 200 时, $R^2$  平均值和  $R^2$  标准差都呈现急剧变化并逐渐趋于稳定. 值得注意的是,如图 3 所示模型的收敛时间由 0.003 s 增加至 0.470 s. 说明较少的树无法很好地刻画非线性关系,导致训练后的 RF 模型拟合不足,当超过一定的临界值后,继续增加

决策树的数量并不会显著提高模型的性能,反而会增加计算时间. 除此之外,RF 模型通过设置决策树的最大深度  $max\_depth$  控制其生长,当达到此阈值时决策树会自动终止训练. 由图 2(b)可知, $max\_depth$  的模型性能曲线整体变化趋势基本与  $n\_estimators$  相同,本文不再进行过多赘述. 通过折衷考虑模型性能得分、稳定性和计算时间,本研究采用  $n\_estimators = 71$ 、 $max\_depth = 51$  作为预测模型的最优超参数.

## 2.3 评价指标

定量地比较不同机器学习模型学习效果的优劣,是机器学习中一个不可或缺且十分重要的环节. 本研究采用均方根误差 (root mean square error, RMSE)、平均绝对误差 (mean absolute error, MAE)、决定系数 ( $R^2$ ) 三个评价指标,评价机器学习模型的性能. 前两者是衡量预测值与实际值的接近程度,值越小说明预测精度越高; $R^2$  是预测值对于实际值

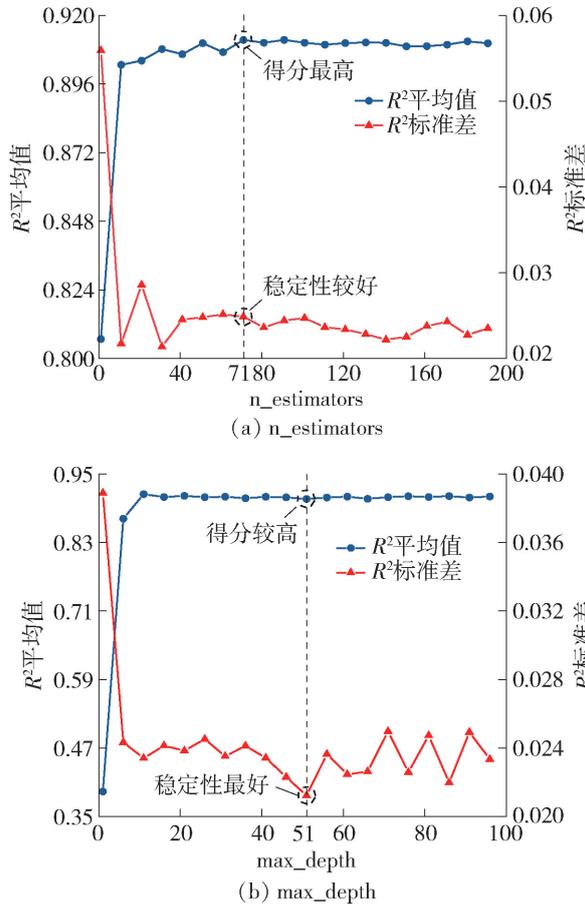
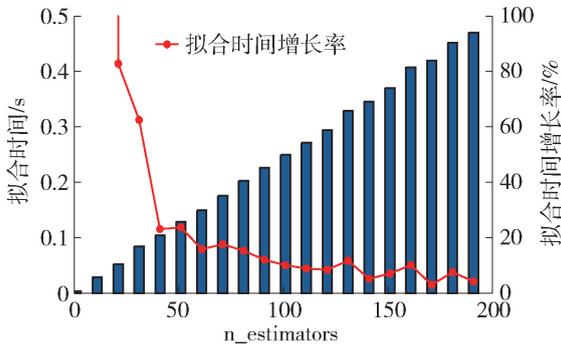


图2 十折交叉验证下关键参数调优结果

Fig. 2 Results of the key parameters optimization using 10-fold cross-validation

图3 收敛时间随  $n\_estimators$  变化趋势Fig. 3 Trend of the convergence time varies with the  $n\_estimators$ 

的表达能力,值越大说明预测精度越高(最大为1). 表征上述指标的数学表达分别为

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (PCI_i - \widehat{PCI}_i)^2} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |PCI_i - \widehat{PCI}_i| \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (PCI_i - \widehat{PCI}_i)^2}{\sum_{i=1}^n (PCI_i - \overline{PCI})^2} \quad (5)$$

式中: $n$  为样本集中  $PCI_i$  的数量; $PCI_i$  为  $PCI$  的实测值; $\widehat{PCI}_i$  为  $PCI_i$  的预测值; $\overline{PCI}$  为  $PCI$  实测值的平均值.

### 3 结果与分析

#### 3.1 特征重要性分析

本研究将基于 RF 的  $PCI$  预测模型的特征重要程度置换为相对特征重要程度. 图 4 给出了 9 种特征变量对于  $PCI$  预测的相对重要性得分. 可以看出,  $PCI$  与目标年上 1 年的  $PCI$  高度相关, 时间序列数据对于路面性能预测至关重要. 年平均日交通量 AADT 对于  $PCI$  预测的重要程度较大, 这点通过交通荷载增加对路面结构综合累计损伤作用的增加不难被解释. 不同路龄状况下路面性能衰变趋势差异性显著, 因此  $PCI$  对于路龄这一特征变量敏感度较高. 与路面自身结构相关的特征变量中面层厚度 PT、基层厚度 BT 这 2 个特征重要性略有差异, 底基层厚度 ST 重要程度略低. 由图 4 可知, 与气象相关的特征变量对于  $PCI$  的影响同样不容忽视, 相对重要程度由大到小排序依次是日照时数 SD、年降水量 AP、年平均气温 AAT, 三者直接影响路面材料性能进而影响路面性能的衰变.

#### 3.2 预测模型验证与评价

本研究为证明基于 RF 的  $PCI$  预测模型的预测能力, 保持与前者相同的训练集和测试集, 分别采用路面性能预测领域成功应用的 2 种机器学习模型 (SVM 和 ANN) 进行对比研究. 按照式 (3) ~ (5) 提出的评价指标, 进一步定量地对比本文提出的 3 种机器学习模型性能.

表 2 给出了不同预测模型的 RMSE、MAE、 $R^2$  值以及计算时间. 从表 2 看出, RFR 模型在测试集和训练集上表现出的学习能力和泛化能力强于 ANN、SVR 模型, 对  $PCI$  预测具有可接受的准确性. 总体来说, 3 种模型在训练集上的 RMSE 分别为 1.042、0.650、0.399, MAE 分别为 0.778、0.455、0.306, RFR 表现出来的学习能力要优于 SVR 和 ANN 模型. RFR 预测模型的泛化能力优越性显著: 相较前两者, RFR 模型的 RMSE 在测试集上分别降低了 42.5% 和 17.7%, MAE 分别下降了 44.1% 和 15.8%, 说明 RF 的预测值

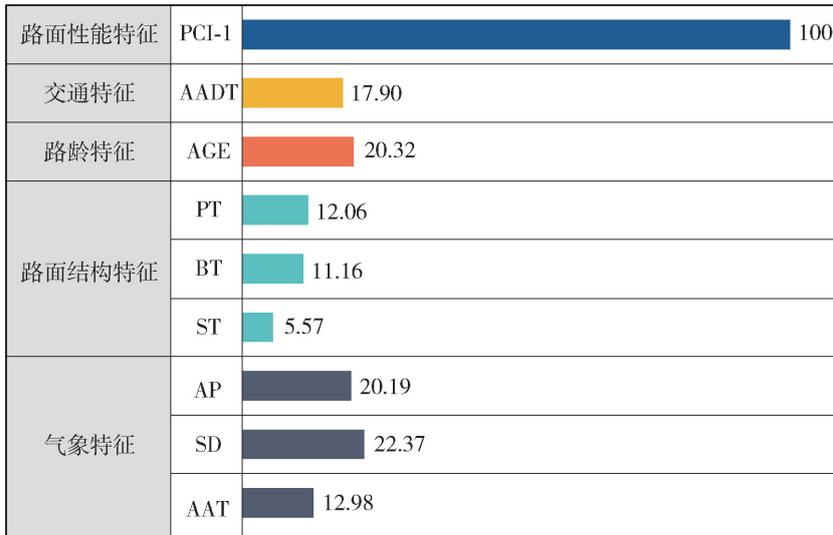


图 4 特征变量重要性

Fig.4 Feature variable importance

表 2 评价指标汇总

Table 2 Summary of the evaluation indices

模型	RMSE		MAE		$R^2$		计算时间/s
	训练集	测试集	训练集	测试集	训练集	测试集	
ANN	1.042	1.246	0.778	0.912	0.832	0.746	0.259
SVR	0.650	0.870	0.455	0.606	0.934	0.866	0.130
RFR	0.399	0.716	0.306	0.510	0.976	0.916	0.174

更加逼近实际值. 此外, 3 种模型计算时间分别为 0.259、0.130、0.174 s, 这表明三者在此数据量大小下的计算效率并未存在显著性差异.

图 5、6 分别给出了评估 3 种模型在训练集和测试集上预测值和实际值对比的散点图. 通过对比训练集、测试集的预测结果, 其中基于 RF 的 PCI 预测

模型的  $R^2$  分别达到 0.976 和 0.916, 预测精度优于其他 2 个模型. 对比图 5、6 可知, 训练集的预测精度普遍高于其对应的测试集的精度, 考虑到  $R^2$  相差不超过 10% 且测试集样本量不大, 所以这种偏差并不严重. 就训练集样本来说, ANN 模型的预测值和实际值较零误差线分散程度最大, SVM 模型次之,

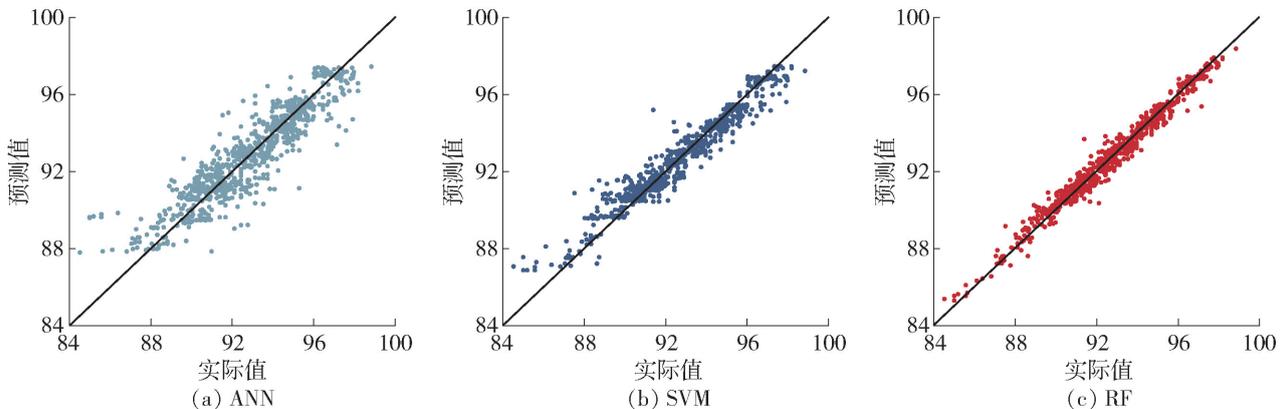


图 5 3 种模型训练集预测值与实际值对比

Fig. 5 Comparison of the predicted and actual training set of three models

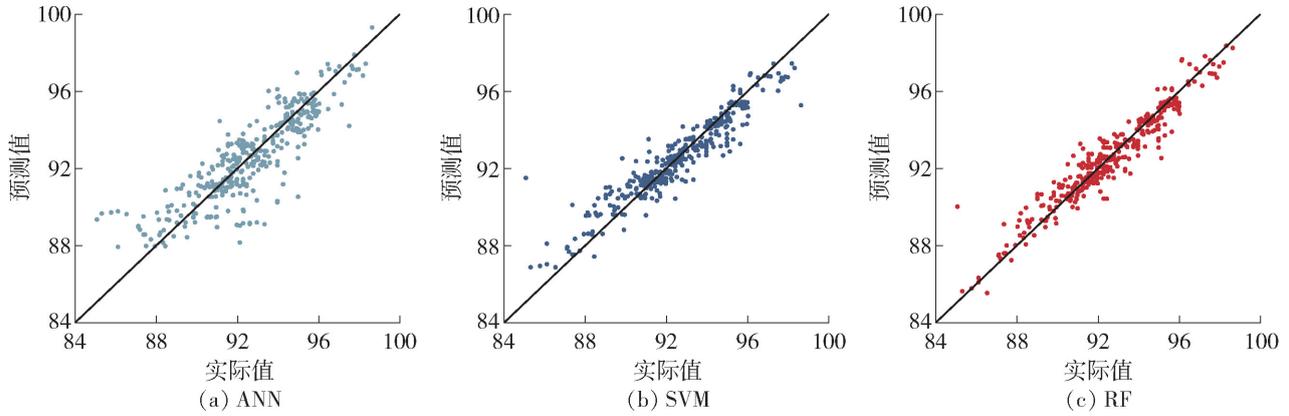


图6 3种模型测试集预测值与实际值对比

Fig. 6 Comparison of the predicted and actual testing set of three models

RFR模型几乎拟合在零误差线附近,说明预测值和实际值基本达到一致。在测试集训练结果中显示,RFR模型的 $R^2$ 达到0.916均高于ANN模型(0.746)和SVM模型(0.866)。

3种模型预测结果存在差异性的原因在于,ANN模型需要依靠大量数据样本不断地修正模型来防止过拟合,提高自身的学习效果。本研究中数据集的数量相较于超量、海量数据集来说,训练不足难以发挥人工神经网络的性能。SVM模型较ANN模型来说可以较好地表征小样本下的非线性映射关系,但是其本身属于弱学习器,而RF属于由一系列弱学习器(决策树)集结在一起的集成算法,并且通过引入随机性增加模型抗噪能力,对于复杂结构的路面性能预测数据集预测精度势必强于单一的弱学习器(ANN,SVM)。综上所述,基于RF的沥青路面状况指数PCI预测模型就路网级预测能力来说,优于人工神经网络和支持向量机预测模型,在处理非线性回归问题上优势显著。

## 4 结论

1) 本研究依据北京地区2010—2016年间路面性能特征、路面结构特征、气象特征以及交通参数特征,以沥青路面PCI为研究对象,采用机器学习中的RF算法建立了北京市沥青路面PCI预测模型。该模型能够准确识别相关特征变量与PCI间复杂的耦合关系,用于预测目标年的PCI有较高的准确率。

2) 本研究引入神经网络和支持向量机2种经典机器学习算法,利用可视化散点图以及参考的3种评价指标,通过对比研究证明了RF模型强大的学习和泛化能力:对学习结果和样本集拟合后,RF模型预测值更为逼近实际值;在训练集和测试集上

的RMSE、MAE和 $R^2$ 都要优于ANN和SVM机预测模型,验证了RF模型的有效性以及优越性。

3) 路面性能预测是路面养护管理的重要内容,基于RF算法的PCI预测模型通过高效、准确的预测可以推动建立现代公路养护管理体系,为后续北京市沥青路面养护规划以及养护资金调配提供指导。

## 参考文献:

- [1] 敬超,张金喜. 沥青路面性能预测研究综述[J]. 中外公路, 2017, 37(5): 31-35.  
JING C, ZHANG J X. Review of asphalt pavement performance prediction[J]. Journal of China & Foreign Highway, 2017, 37(5): 31-35. (in Chinese)
- [2] 杜艳花. 基于时间序列的沥青路面平整度预测方法[J]. 兰州交通大学学报, 2017, 36(1): 41-46.  
DU Y H. Method of prediction of asphalt pavement roughness based on time series method[J]. Journal of Lanzhou Jiaotong University, 2017, 36(1): 41-46. (in Chinese)
- [3] 张金喜,孔祥杰,韩丁丁,等. 不同养护模式下的高速公路沥青路面状况指数预测模型[J]. 北京工业大学学报, 2016, 42(1): 80-86.  
ZHANG J X, KONG X J, HAN D D, et al. Prediction model of asphalt pavement PCI of expressway under different maintenance conditions[J]. Journal of Beijing University of Technology, 2016, 42(1): 80-86. (in Chinese)
- [4] 肖金平,韦慧,赵健,等. 湖南省高速公路路面使用性能衰变模型[J]. 中南大学学报(自然科学版), 2015, 46(7): 2686-2692.  
XIAO J P, WEI H, ZHAO J, et al. Decay model of Hunan expressway pavement performance[J]. Journal of

- Central South University (Science and Technology), 2015, 46(7): 2686-2692. (in Chinese)
- [5] SATISH C, CHALUMURI R S, ANISH K B, et al. Relationship between pavement roughness and distress parameters for Indian highways [J]. Journal of Transportation Engineering, 2013, 139(5): 467-475.
- [6] 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.
- ZHAO J, WANG X C, DING L T, et al. Performance prediction of asphalt pavement based on grey relational analysis and support vector machine regression [J]. Journal of Chongqing University, 2019, 42(4): 72-81. (in Chinese)
- [7] BREIMAN L. Random forests [J]. Machine Learning, 2001, 45(1): 5-32.
- [8] JOHN T F, DANIEL D. M. Predicting hydrologic disturbance of streams using species occurrence data[J]. Science of the Total Environment, 2019, 686: 254-263.
- [9] MAYA G P. S, BHARGAVAL R. A novel approach for efficient crop yield prediction [J]. Computers and Electronics in Agriculture, 2019, 165: 1-9.
- [10] 高若楠,苏喜友,谢阳生,等.基于随机森林的杉木适生性预测研究[J].北京林业大学学报,2017,39(12):40-47.
- GAO R N, SU X Y, XIE Y S, et al. Prediction of adaptability of *Cunninghamia lanceolata* based on random forest[J]. Journal of Beijing Forestry University, 2017, 39(12): 40-47. (in Chinese)
- [11] 欧强新,雷相东,沈琛琛,等.基于随机森林算法的落叶松-云冷杉混交林单木胸径生长预测[J].北京林业大学学报,2019,41(9):9-19.
- OU Q X, LEI X D, SHEN C C, et al. Individual tree DBH growth prediction of larch-spruce-fir mixed forests based on random forest algorithm[J]. Journal of Beijing Forestry University, 2019, 41(9): 9-19. (in Chinese)
- [12] XING J K, LUO K. Estimating biomass major chemical constituents from ultimate analysis using a random forest model[J]. Bio Resource Technology, 2019, 288: 1-7.
- [13] 冯晓龙,高静.基于Spark平台的考勤数据决策树回归分析[J].内蒙古工业大学学报(自然科学版),2018,37(2):53-58.
- FENG X L, GAO J. Attendance data analysis with decision regression tree on Spark [J]. Journal of Inner Mongolia University of Technology (Natural Science Edition), 2018, 37(2): 53-58. (in Chinese)
- [14] 周志华.机器学习[M].北京:清华大学出版社,2016:178-181.
- [15] 任才溶,谢刚.基于随机森林和气象参数的浓度等级预测[J].计算机工程与应用,2019,55(2):213-220.
- REN C R, XIE G. Prediction of  $PM_{2.5}$  concentration level based on random forest and meteorological parameters [J]. Computer Engineering and Applications, 2019, 55(2): 213-220. (in Chinese)
- [16] 王梓杰,周新志,宁芊.基于PCA和随机森林的故障趋势预测方法研究[J].计算机测量与控制,2018,26(2):21-26.
- WANG Z J, ZHOU X Z, NING Q. Study on the fault trend prediction method based on PCA and random forest [J]. Computer Measurement & Control, 2018, 26(2): 21-26. (in Chinese)
- [17] 李欣海.随机森林模型在分类与回归分析中的应用[J].应用昆虫学报,2013,50(4):1190-1197.
- LI X H. Using random forest for classification and regression[J]. Chinese Journal of Applied Entomology, 2013, 50(4): 1190-1197. (in Chinese)

(责任编辑 杨开英)