

DOI:10.13869/j.cnki.rswc.2023.03.036.

单振东, 骆汉, 刘顿. 基于机器学习算法的蒸发量模型评估[J]. 水土保持研究, 2023, 30(3): 289-294.

SHAN Zhendong, LUO Han, LIU Dun. Evaporation Model Evaluation Based on Machine Learning Algorithm[J]. Research of Soil and Water Conservation, 2023, 30(3): 289-294.

# 基于机器学习算法的蒸发量模型评估

单振东<sup>1</sup>, 骆汉<sup>1,2</sup>, 刘顿<sup>1</sup>

(1.西北农林科技大学 水土保持研究所, 陕西 杨凌 712100; 2.中国科学院 水利部 水土保持研究所, 陕西 杨凌 712100)

**摘要:** [目的] 探讨合理的气候因子个数, 建立蒸发量模型, 提出基于特征选择算法筛选最优特征集。 [方法] 以陕西榆林、泾河和汉中3个气象站16年(2005-01至2021-03)的逐小时观测资料为研究对象, 利用特征选择函数和遍历循环方法对模型参数、特征变量个数进行优化。基于最佳参数结合随机森林模型和多元线性回归模型两种机器学习算法建立榆林、汉中和泾河地区蒸发量模型, 采用平均绝对误差、均方根误差和平方相关系数三项指标评估模型的预测精度。 [结果] 特征变量和随机森林模型中的决策树个数分别是8, 61时, 模型预测效果最佳。采用优化的随机森林模型、多元线性回归模型评估3个地区的平均绝对误差均为0, 均方根误差除泾河地区相等外, 榆林、汉中地区的均方根误差均小于优化的多元线性回归模型。优化的随机森林模型预测榆林、泾河和汉中地区蒸发量拟合效果分别为0.85, 0.90, 0.86, 优化的多元线性回归模型的拟合效果分别为0.77, 0.83, 0.79。 [结论] 整体而言, 优化的两种模型都具有良好的预测效果且随机森林模型的预测效果优于多元线性回归模型。

**关键词:** 蒸发量; 特征选择; 随机森林模型; 多元线性回归模型

中图分类号: S161.4

文献标识码: A

文章编号: 1005-3409(2023)03-0289-06

## Evaporation Model Evaluation Based on Machine Learning Algorithm

SHAN Zhendong<sup>1</sup>, LUO Han<sup>1,2</sup>, LIU Dun<sup>1</sup>

(1. Institute of Soil and Water Conservation, Northwest A&F University,

Yangling, Shaanxi 712100, China; 2. Institute of Soil and Water Conservation,

Chinese Academy of Sciences and Ministry of Water Resources, Yangling, Shaanxi 712100, China)

**Abstract:** [Objective] The optimal feature set was determined in order to explore the reasonable number of climatic factors to establish an evaporation model, based on filtering by feature selection algorithm. [Methods] 16 years (2005-01 to 2021-03) of Yulin, Jinghe and Hanzhong weather stations of hour-by-hour observation data were taken. The model parameters and the number of characteristic variables were optimized by using feature selection and traversal loop methods. Based on the best parameters combined with two machine learning algorithms, the random forest model and the multiple linear regression model, evaporation model suitable for 3 regions was established, and the three indicators of average absolute error, root mean square error and square correlation coefficient were used to evaluate the model accuracy. [Results] The number of feature variables and decision trees are 8 and 61, the prediction effect of the model is the best. The root mean square error was used to evaluate the prediction effects of the optimized random forest model and the multiple linear regression model, and the error size of the two models was 0. Except for the Jinghe area, the root mean square error of Yulin and Hanzhong was smaller than that of the optimized multiple linear regression model. Predict effects of evaporation in Yulin, Jinghe and Hanzhong fitted by the optimized random forest

收稿日期: 2022-03-07

修回日期: 2022-05-24

资助项目: 陕西省科技重大专项“果园地力提升及有机替代化肥减肥增效关键技术与示范”(2020zdzx03-02-01)

第一作者: 单振东(1994—), 男, 甘肃省定西市人, 硕士研究生, 研究方向为水土保持与荒漠化防治。E-mail: shan18794940428@163.com

通信作者: 骆汉(1985—), 男, 山东省淄博市人, 博士, 博士生导师, 从事水土保持工程研究。E-mail: hanl@ms.iswc.ac.cn

<http://stbcyj.paperonice.org>

model are 0.85, 0.90 and 0.86, and the optimized multiple linear regression model fitting effects are 0.77, 0.83 and 0.79. The optimized random forest model of the average absolute error and the root mean square error are lower than optimized multiple linear regression model. [Conclusion] On the whole, the optimized two models have good prediction effects and the prediction effect of the random forest model is better than the multiple linear regression model.

**Keywords:** evaporation; feature selection; random forest model; multiple linear regression model

机器学习是计算机利用概率论和统计学等知识模拟人类学习,根据现有的数据建立关系,从大量和复杂的数据中抓取有价值的信息,用于预测、预报变化趋势,提高学习效率的一种方法<sup>[1-2]</sup>。蒸发量是在一定时间内,水分经蒸发而散布到空气中的量,因此蒸发将植物、土壤、大气和水分联系在一起,在气候变化研究中具有敏感的指示作用<sup>[3-4]</sup>。目前国内采用红外线热成像技术<sup>[5]</sup>、遥感技术与地面资料结合<sup>[6-7]</sup>、实地监测等<sup>[8-9]</sup>方法测量蒸发量。随着大数据发展,机器学习算法和模型被用于农业预测等<sup>[10]</sup>方面,其中用于预测蒸发量的有随机森林模型<sup>[11]</sup>、多元线性回归模型<sup>[12]</sup>、人工神经网络<sup>[13]</sup>、多层感知器<sup>[14]</sup>、贝叶斯网络等<sup>[15]</sup>模型和算法。其中随机森林模型在实际生活中应用广泛,成为解决非线性问题和回归问题的常用算法,成为建立模型的首选<sup>[16]</sup>。多元线性回归模型是分析多个自变量与单一因变量之间的线性关系的算法,是分析因变量受到多种因素影响的常用模型<sup>[17]</sup>。因此,本文选择随机森林模型和多元线性回归模型预测蒸发量。

陕西地处中国西北内陆,具有温带干旱区气候、暖温带湿润性气候和亚热带大陆性气候3个气候带,以北山和秦岭为界,将陕西省分成陕北黄土高原、关中平原和陕南秦巴山地区,其中榆林、泾河和汉中是各区域的代表性区域<sup>[4]</sup>。因此,基于榆林、泾河和汉中气象站数据资料建立陕西地区蒸发量模型。针对陕西地区蒸发量研究中,基于 $K$ 近邻算法和神经网络算法2种机器学习算法,龙亚星等<sup>[4]</sup>选择了15个气候因子建立蒸发量估算模型,得出 $K$ 近邻算法预测效果优于神经网络算法。但选择的特征变量过多,不但会使模型的训练时间增加,而且特征变量中的无效信息会使模型出现过拟合、过度学习的问题<sup>[18]</sup>。决策树是随机森林模型中的重要参数,合理的决策树数量可以减缓过拟合问题,优化决策树成为随机森林模型中不可缺少的环节<sup>[19]</sup>。鉴于此,为了避免建立蒸发量模型出现过拟合问题,对选择的特征变量和随机森林模型中的决策树个数进行优化,基于最佳特征变量和决策树个数建立榆林、汉中和泾河地区的蒸发量模型,采用3项评价指标评估模型在不

同区域的优越性,为陕西地区蒸发量的预测提供参考依据,并合理配置陕西地区水资源,提高农业灌溉效率具有重大意义。

## 1 资料与方法

### 1.1 资料的来源

气象数据源于中国气象数据网提供陕西汉中、榆林和泾河3个气象站16年(2005年01月—2021年03月)的逐小时观测资料,气象数据中的因子有20—20时降水量、蒸发(大型蒸发皿)、极大风速、极大风速的方向、平均气压、平均2 min风速、平均气温、平均水汽压、平均相对湿度、日照时数、最低气压、最低气温、最高气压、最高气温、蒸发、最大风速、最大风速的方向和最小相对湿度。

### 1.2 数据处理

基于机器学习算法利用Anaconda 3(64-Bit)和Python 3.7(64-Bit)软件,建立蒸发量模型。将20—20时降水量中的微量数据记作0,考虑风向的复杂性,删除最大、极大风速方向的数据,若某天的气象资料中存在缺失值,删除这天所有的气候数据。将气候因子中的20—20时降水量记作( $V_1$ )、平均气压( $V_2$ )、平均2 min风速( $V_3$ )、平均气温( $V_4$ )、平均水汽压( $V_5$ )、平均相对湿度( $V_6$ )、日照时数( $V_7$ )、最低气压( $V_8$ )、最低气温( $V_9$ )、最高气压( $V_{10}$ )、最高气温( $V_{11}$ )、最大风速( $V_{12}$ )、最小相对湿度( $V_{13}$ )和蒸发量(label)。每个气象站的气象数据组成数据集,按照训练集与测试集的比例为3:1划分数据集,将蒸发量作为数据集的标签列(响应变量),其余气象数据作为特征列(特征变量)。气候因子间存在单位和量级差异,为便于后期处理和分析,需要对数据进行归一化处理。

### 1.3 评价指标

为了评估模型的准确精度,引入平均绝对误差(MAE)、均方根误差(RMSE)和平方相关系数( $R^2$ )三项指标,其中平方相关系数越接近1,表示拟合程度越高,预测结果越接近真实值;平均绝对误差和均方根误差数值越小,表示预测模型准确性越强<sup>[20]</sup>。

$$MAE=\frac{1}{n}\sum_{i=1}^n|(y(i)-y''(i))|$$
 (1)

$$RMSE=\frac{1}{n}\sqrt{\sum_{i=1}^n(y(i)-y''(i))^2}$$
 (2)

$$R^2=\frac{\sum_{i=1}^n[y''(i)-\overline{y}(i)]^2}{\sum_{i=1}^n[y(i)-\overline{y}(i)]^2}$$
 (3)

式中: $n$  表示测试集中样本数量; $y(i),y''(i),\overline{y}(i)$  分别表示测试集中蒸发量真实值、预测值和平均值。

2 预测模型

2.1 特征工程及模型简介

2.1.1 特征工程简介 特征工程通过机器学习构建模型,选择相关特征并构成特征子集<sup>[21]</sup>。特征工程按照过滤类型不同分成特征选择和特征排序两类。特征选择是在不改变原始数据的前提下,从高维数据中选择有效特征形成低维度的数据特征,降低维度后的结果仍具有解释作用,达到了简化机器学习算法的目的<sup>[22-23]</sup>。特征排序是与预测算法无关的机器学习算法,是对每个特征变量和预测变量之间建立预测模型,在训练过程中按照每个特征重要性进行排序<sup>[24]</sup>。

2.1.2 模型简介 随机森林依据决策树算法寻找对象属性和对象值之间的映射关系,利用已有的数据构建预测准则,根据变量值对另一个变量进行预测<sup>[25]</sup>。随机森林模型是将特定数量的训练集有放回多次均匀抽取样本,每一组样本组成决策树,形成一个森林,森林中每颗决策树都会产生一个分类结果,将多个分类结果进行投票统计,得票最多的类别为最终的分类结果<sup>[26]</sup>。

多元线性回归含有两个或者两个以上的自变量,且自变量与因变量之间存在线性关系。常见的表达式如下:

$$Y_i=\beta_1X_{1i}+\beta_2X_{2i}+\beta_3X_{3i}+\beta_nX_{ni}+\beta_0$$
  
( $i=1,2,3,\cdots,n$ ) (4)

式中: $n$  为变量的个数; $\beta_0$  为常数; $Y_i$  表示第  $i$  个因变量的值; $\beta_i(1,2,3,\cdots,n)$  表示自变量前的系数<sup>[27]</sup>。

2.2 初始模型

基于榆林、汉中和泾河的气象站数据资料,利用定点气象站数据代表特定区域的气候数据,设置多元线性回归模型和随机森林模型的参数为默认值,选择全部的特征变量建立蒸发量模型,采用三项指标评估模型的准确精度,结果见表 1。采用多元线性回归模型预测榆林和汉中地区的拟合效果为 0.80,预测泾河地区的拟合效果达到 0.86,采用随机森林模型预测榆林、汉中和泾河地区的拟合效果分别为 0.86,0.88,0.91;采用随机森林模型、多元线性回归模型预测泾

河地区的均方根误差、平均绝对误差分别为 0.06,0.01,预测榆林、汉中地区的平均绝对误差值均为 0;随机森林模型预测榆林、汉中地区的均方根误差 0.06,0.07均小于多元回归模型预测的 0.07,0.08。整体而言,采用随机森林模型预测榆林、汉中和泾河地区蒸发量效果优于多元线性回归模型。

表 1 不同地区两种初始模型各项指标对比

地区	随机森林模型			多元线性回归模型		
	RMSE	MAE	$R^2$	RMSE	MAE	$R^2$
榆林	0.06	0.00	0.86	0.07	0.00	0.80
泾河	0.06	0.01	0.91	0.06	0.01	0.86
汉中	0.07	0.00	0.88	0.08	0.00	0.80

2.3 优化模型

2.3.1 特征排序及特征参数优化 为提高随机森林模型、多元线性回归模型在陕西地区的泛化能力,将 3 个气象站的气象数据重新组成新的气象资料数据集。对蒸发量有影响作用的最高气温、最低气温、平均气温、日照时数、平均水气压、最小相对湿度、平均相对湿度、最低气压、平均 2 min 风速和最大风速等气候因子,将气候因子作为特征变量,蒸发量作为预测变量,建立每个特征变量和预测变量的预测模型,调用 Python 中的特征选择函数(SelectKBest),通过卡方检验计算单个气候因子与蒸发量间的卡方值,若卡方值越大,表明两者间相关性越强,依据相关性大小对特征进行排序(图 1)。特征重要性依次是日照时数、最高气温、平均相对湿度、最大风速、平均气温、平均 2 min 风速、最小相对湿度、最低气压、最低气温、平均水气压、最高气压、平均气压、20—20 时降水量。其中前 5 位气候因子的特征重要性均超过 0.08,日照时数是诸多气候因子重要性最高为 0.36;最后 4 位气候因子与蒸发量间的重要程度均低于 0.02,其中 20—20 时降水量是影响蒸发量能力的最弱的气候因子。

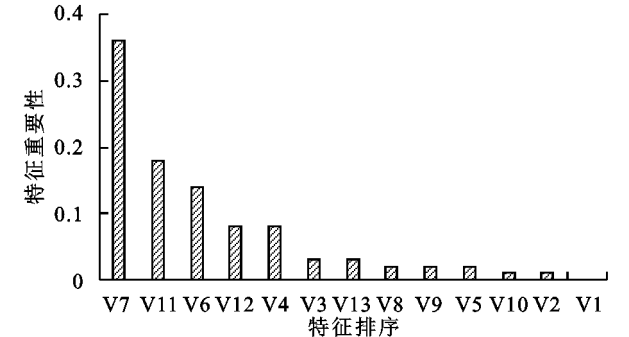


图 1 气候因子重要性排序

根据气候因子个数,设置特征参数个数  $n=1,2,3,4,\cdots,13$ ,按照特征重要性由弱到强顺序逐步减少特征个数,计算不同特征参数个数对应的随机森林模型、多元线性回归模型的模型得分(图 2)。两种模型



的得分随着特征参数个数增加整体呈现先急速增加再缓慢增加后趋于平缓。选取特征个数从1到4时,随机森林模型、多元线性回归模型的模型得分随着特征参数个数的增加呈现快速增加趋势,特征参数个数等于4两种模型的模型得分分别为0.81,0.75;特征参数个数介于5~8,随机森林模型得分随着特征个数增加呈现缓慢增加,模型得分由0.82增加至最大值0.85;特征参数个数介于5~7,多元线性回归模型随着特征参数个数增加呈现缓慢增加,模型得分由0.75增加至最大值0.76。因此,采用机器学习建立蒸发量模型时,选择的特征变量个数为8。

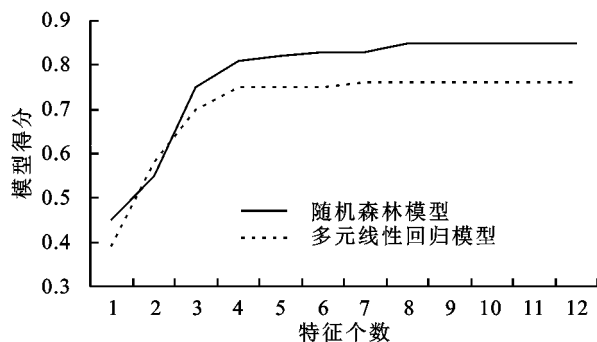


图2 特征参数个数所对应的模型得分

2.3.2 决策树的优化 决策树根据节点属性的分支条件对数据相关属性进行分类<sup>[26]</sup>,树的最大深度和决策树个数是影响随机森林模型的重要参数,合理的决策树数量可以降低过拟合和提升模型的泛化能力<sup>[28]</sup>。基于最佳的特征参数个数,设定决策树个数范围1~100,步长为5,通过遍历循环的方法计算不同决策树个数对应的随机森林模型得分(图3)。决策树个数取值介于1~6,随机森林模型的模型得分随着决策树个数的增加呈现快速增大的趋势,模型得分由0.37增加至0.56;决策树个数取值介于5~31,随着决策树个数的增加模型得分呈增加缓慢的趋势,取值在31~56,61~96之间,随机森林的模型得分保持不变,分别是0.59,0.60;决策树个数取值为61,随机森林模型率先达到最大值0.60。因此,基于机器学习建立预测蒸发量模型中的决策树的个数选择为61。

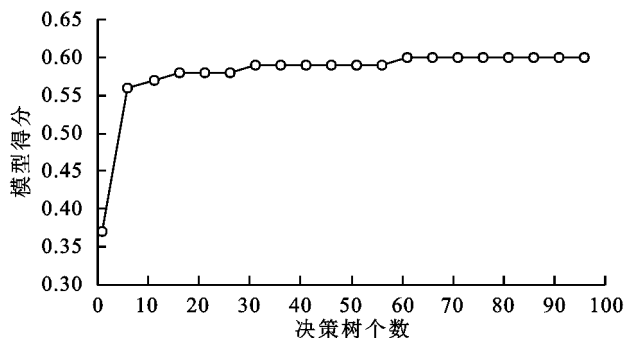


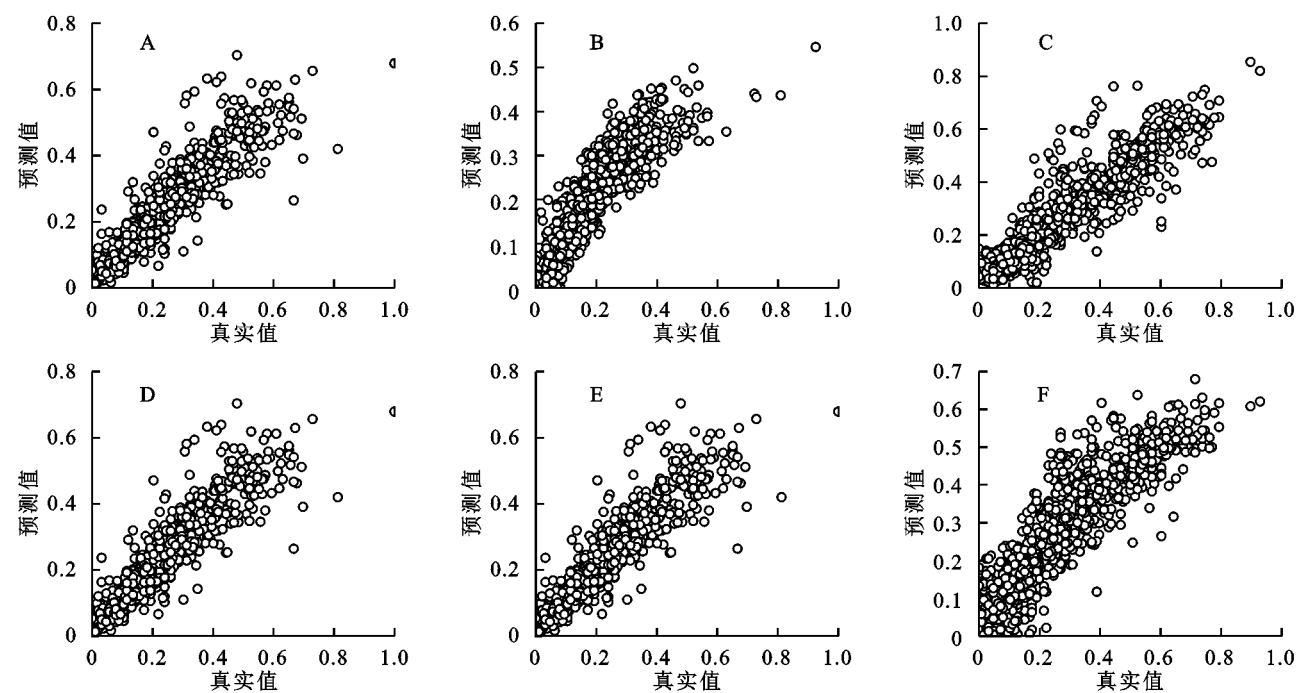
图3 决策树个数所对应模型得分

### 3 模型预测和评估

基于决策树和特征参数的最优个数,利用随机森林模型和多元线性回归模型建立榆林、汉中和泾河地区的优化预测蒸发量模型,采用散点图将真实值与预测值进行展示(图4)和三项指标评估两种优化模型的准确精度(表2)。采用优化的随机森林模型预测汉中、榆林和泾河地区的蒸发量预测效果显著优于优化的多元线性回归模型,优化的随机森林模型预测3个地区的蒸发量的拟合效果均超过0.85,优化的多元线性回归模型预测榆林、汉中地区的蒸发量的拟合效果低于0.80。采用平均绝对误差指标评估优化的2种模型在3个地区的蒸发量的拟合效果均超过0.85,采用优化的多元线性回归模型预测榆林、汉中地区的蒸发量的拟合效果低于0.80。采用平均绝对误差指标评估优化的2种模型在3个地区的预测效果,其值均为0,均方根误差评估2种优化的模型不同地区的蒸发量预测效果,除泾河地区误差大小一致外,汉中、榆林地区优化的随机森林模型的预测效果均优于优化的多元线性回归模型。

### 4 讨论

蒸发量大小受到多种因素的影响,而且因素之间存在复杂的非线性关系。传统的蒸发量测量方法,会受到测量仪器故障、异常天气的影响,造成数据缺失,且无法揭示气候要素之间复杂的关系。随着机器学习算法的发展,多种模型和算法可以解释气候要素之间的关系,并且用于预测蒸发量的变化趋势。通过从众多气候因子中筛选重要因子建立模型,达到简化模型的计算量并且易于推广的效果<sup>[29]</sup>。本文使用特征排序算法对影响蒸发量大小的气候因子进行排序,借助于特征选择函数对建立模型选用的气候因子个数优化,得出特征参数个数为8模型的预测效果最佳,特征因子包含气温(最高气温、平均气温)、日照时数、最低气压、湿度(平均相对湿度和最小相对湿度)和风速(最大风速和平均2 min 风速)。日照时数是影响蒸发量最强的气候因素,因为日照时数反映了太阳辐射量,日照越长,获得辐射越多,蒸发量越大。温度是影响蒸发量能力较强的气候因子,因为在任何温度下,分子都在不断地运动,液体中存在较大的分子能够脱离束缚成为水汽分子,在任何温度下都能蒸发,温度越高,分子的平均动能增加,蒸发速度加快。20—20 mm降水量是影响蒸发量的能力最弱的气候因子,即降水量的大小对蒸发量的影响能力较小,因为降水量过多,会使土壤的湿度增加,蒸发量相应的降低。



注:图 A,B,C 分别表示优化的随机森林模型预测榆林、泾河和汉中的预测效果图;图 D,E,F 分别表示优化的多元线性回归模型预测榆林、泾河和汉中的预测效果图。

图 4 不同地区优化模型预测蒸发量效果

表 2 不同地区两种优化模型各项指标对比

地区	随机森林模型			多元线性回归模型		
	RMSE	MAE	R <sup>2</sup>	RMSE	MAE	R <sup>2</sup>
榆林	0.06	0	0.85	0.08	0	0.77
泾河	0.06	0	0.90	0.06	0	0.83
汉中	0.07	0	0.86	0.08	0	0.79

决策树的数量是影响随机森林模型的重要参数,设置的决策树数量过少,不能保证具有较好的预测效果,设置个数过多,增加模型的训练时间<sup>[30]</sup>。如果人为设定决策数的数量,会使模型出现欠拟合或者过拟合的问题<sup>[31]</sup>。因此,本文设定决策树个数的范围,通过遍历循环的方法寻找最佳的决策树个数,得出随机森林模型得分随着决策树个数的增加整体呈现先快速增加再缓慢增加最后趋于平缓的趋势。当决策树个数为 61 时,随机森林模型的模型得分优先达到 0.60,再继续增加决策树个数,模型得分保持不变,此时决策树个数为最佳的决策树个数。

基于最佳决策树和特征参数个数建立 3 个地区蒸发量模型,采用均方根误差、平均绝对误差和平方相关系数指标,评价随机森林模型和多元线性回归模型的预测效果,得出优化的随机森林模型评估榆林、汉中地区的均方差误差指标均低于优化的多元线性回归模型,平方相关系数高于优化的多元线性回归模型,这与张蓓蓓<sup>[32]</sup>通过特征选择筛选影响顾客满意度的重要特征,采用网格搜索和交叉验证方法对模型参数进行优化,基于均方根误差评估模型优化前后预

测效果得出结论一致。原因是多元线性回归模型模拟人类大脑神经元细胞,拟合准确性程度高,如果学习了数据集的异常值,容易对模型中的其他值产生影响,预测结果出现偏差和模型产生过拟合<sup>[33]</sup>。相比多元线性回归模型,采用随机森林模型预测蒸发量的预测值与真实值更接近,这是由于随机森林模型在训练数据时,对特征和样本进行多次有放回随机抽样,因此在回归预测算法上更加高效。

5 结论

- (1) 气候因子重要性依次是日照时数、最高气温、平均相对湿度、最大风速、平均气温、平均 2 min 风速、最小相对湿度、最低气压、最低气温、平均水气压、最高气压、平均气压、20—20 时降水量。
- (2) 特征变量个数取值为 8,随机森林模型和多元线性回归模型的模型得分最高分别是 0.85,0.76。随机森林模型中的决策树个数为 61 时,随机森林模型得分最高为 0.60。
- (3) 基于最佳的特征参数和决策树的个数,采用随机森林模型预测榆林、汉中和泾河地区的蒸发量的拟合效果分别是 0.85,0.86,0.90,多元线性回归模型的拟合效果分别是 0.77,0.79,0.83;优化的随机森林模型评估 3 个地区的均方根误差整体优于优化的多元线性回归模型。

## 参考文献:

- [1] 陈潮锋,石宇雄,梁锦成,等.基于机器学习算法预测全膝关节置换后住院时长[J].中国组织工程研究,2021,25(27):4300-4306.
- [2] 李昊朋.基于机器学习方法的智能机器人探究[J].通讯世界,2019,26(4):241-242.
- [3] 陈慧,王庆莉. AG2.0 型超声波蒸发传感器遇低温结冰时记录的处理[J].气象水文海洋仪器,2020,37(1):98-101.
- [4] 龙亚星,黄勤,李成伟.基于机器学习技术的蒸发皿蒸发量估算模型[J].气象科技,2021,49(2):166-173.
- [5] 茅威洋,顾峰豪,张铁南,等.基于红外线成像技术的沥青混合料压实度温度监测研究[J].山东交通科技,2019,32(5):50-52.
- [6] 营娜,叶谦,韩战钢,等.全球地表温度大气遥相关路径研究[J].北京师范大学学报:自然科学版,2021,57(3):1-10.
- [7] 赵春江,杨贵军,薛绪掌,等.基于互补相关模型和 IKONOS 数据的农田蒸散时空特征分析[J].农业工程学报,2013,29(8):115-126.
- [8] 周宁.怀柔水库水面蒸发量及变化特征分析[J].工程技术研究,2021,6(4):247-248.
- [9] 刘萌,王善举,樊军,等.利用环刀法快速原位测定土壤蒸发量[J].土壤通报,2021,52(1):55-61.
- [10] 张菊,房世波,刘汉湖.基于微波数据与光学数据集成的机器学习技术在作物产量估算中的应用[J].地球信息科学学报,2021,23(6):1082-1091.
- [11] 任梅芳,庞博,徐宗学,等.基于随机森林模型的雅鲁藏布江流域气温降尺度研究[J].高原气象,2018,37(5):1241-1253.
- [12] 孙英,周金龙,齐子萱,等.日尺度和小时尺度下干旱区潜水蒸发变化特征及成因[J].灌溉排水学报,2020,39(S2):14-19.
- [13] Ali Rahimi Khoob. Artificial neural network estimation of reference evapotranspiration from pan evaporation in a semi-arid environment[J]. Irrigation Science, 2008, 27(1):35-39.
- [14] Ghorbani M A, Ravinesh C. Deo, Zaher Mundher Yaseen, etc. Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran[J]. Theoretical and Applied Climatology, 2018,133(3):1119-1131.
- [15] 段志昌,王建有,王特,等.基于贝叶斯网络的水面蒸发评估模型研究[J].水电能源科学,2021,39(3):1-4.
- [16] 马寰宇.数据挖掘在精准扶贫中的研究及应用[D].哈尔滨:黑龙江大学,2019.
- [17] 李丹.基于规模视角下山东省粮食经营主体种植行为及持续意愿研究[D].北京:中国农业科学院,2019.
- [18] 张建航,牛博威,张祥,等.基于深度信念极限学习机的入侵检测模型[C].2019 互联网安全与治理论坛论文集,2019:111-115.
- [19] 成浩科,沈菲.基于随机森林的河流总磷预测模型及影响因素分析[J].环境保护科学,2021,47(3):62-68.
- [20] 石文兵,苏树智.基于优化随机森林算法的高校餐饮企业营业额预测模型[J].通化师范学院学报,2021,42(2):88-94.
- [21] 夏江流云.基于 Stacking 融合算法的企业贷款违约风险预警模型设计与应用[D].南昌:江西财经大学,2021.
- [22] 张巍,张圳彬.联合图嵌入与特征加权的无监督特征选择[J].广东工业大学学报,2021,38(5):16-23.
- [23] 黄烈佳,杨鹏.基于机器学习的武汉城市圈土地生态安全格局识别与优化策略[J].生态与农村环境学报,2020,36(7):862-869.
- [24] 刘金翰.基于回归分析的汽车测试时间预测[D].兰州:兰州大学,2021.
- [25] 武冠芳.基于大数据的用户特征及活跃度分析[D].北京市:北京邮电大学,2017.
- [26] 刘经纬,陈佳明.创新创业+新工科教学质量提升研究—零基础学会 Python 人工智能[M].北京:首都经济贸易大学出版社,2020.
- [27] 陈建东.多元回归模型在基坑地下水位预测中的应用[J].矿山测量,2021,49(1):52-56.
- [28] 姬正一,陈阳,沈培志,等.决策树集成方法在反舰导弹效能评估中的应用[J].现代防御技术,2021,49(4):15-24.
- [29] Mao S, Zhang X, Han Y, et al. Random forest algorithm for the relationship between negative air ions and environmental factors in an urban park[J]. Atmosphere, 2018,9(12):463.
- [30] 王彤,刘九良,朱劲璇,等.基于随机森林的电力系统暂态稳定评估与紧急控制策略[J].电网技术,2020,44(12):4694-4701.
- [31] 崔云浩,朱军,孟浩.神经网络在农产品销量预测中的应用[J].现代农业科技,2021,28(17):262-266.
- [32] 张蓓蓓,胡敏.基于网格搜索改进随机森林的顾客满意度预测[J].北京信息科技大学学报:自然科学版,2021,36(4):50-54.
- [33] 徐佳,刘峰,吴华勇,等.基于人工神经网络和随机森林学习模型从土壤属性推测关键成土环境要素的研究[J].土壤通报,2021,52(2):269-278.