全球定位系统
GNSS World of China
• 中国科学引文数据库（CSCD）
• 中文科技期刊数据库
• 中国核心期刊（遴选）数据库
• 日本科学技术振兴机构数据库（JST）
• 中国学术期刊（网络版）（CNKI）
• 中国学术期刊综合评价数据库（CAJCED）
• 中国超星期刊域出版平台

几种典型机器学习算法在短临降雨预报分析研究

池　钦，赵兴旺，陈　健

Short-term rainfall forecast by several typical machine learning algorithm

CHI Qin, ZHAO Xingwang, and CHEN Jian

引用本文:
池钦，赵兴旺，陈健．几种典型机器学习算法在短临降雨预报分析研究[J]．全球定位系统，2022，47(4)：122-128．DOI: 10.12265/j.gnss.2022039
CHI Qin, ZHAO Xingwang, CHEN Jian. Short-term rainfall forecast by several typical machine learning algorithm[J]. Gnss World of China, 2022, 47(4): 122-128. DOI: 10.12265/j.gnss.2022039

在线阅读 View online: https://doi.org/10.12265/j.gnss.2022039

DOI: 10.12265/j.gnss.2022039

几种典型机器学习算法在短临降雨预报分析研究

池钦, 赵兴旺, 陈健

(安徽理工大学 空间信息与测绘工程学院, 安徽 淮南 232001)

摘　要: 针对降雨过程中大气可降水量 (PWV) 和气象参数 (温度 (T)、湿度 (U)、露点温度 (T_d)、气压 (P)) 特征变化情况, 提出基于机器学习算法的短临降雨预报模型. 以北京 (BJFS) 站和武汉 (WUH2) 站 2020 年的 3 h 天顶对流层延迟 (ZTD) 和气象数据为例, 构建随机森林 (RF)、支持向量机 (SVM)、K 近邻 (KNN)、朴素贝叶斯分类器 (NBC) 4 种算法的预报模型, 并引入各自时刻的降雨情况作为新的特征向量, 分别采用 70% 和 80% 训练集的分割方式, 降雨情况作为模型输出, 并利用准确性、精确率和假负率评价模型的适用性. 在取得准确性约 0.92, 精确率约 80%, 假负率约 20% 的结果下, 进一步以时间序列年积日为第 150—200 天的数据为样本, 对 200—250 天的降雨情况进行预报. 实验结果表明: 基于机器学习的短临降雨预报模型可以预报未来 3 h 80% 以上的降雨情况, 且假负率在 20% 以下, 其中 SVM模型的综合性能更优. 与传统的阈值模型相比, 准确率相当, 假负率降低约 50%.

关键词: 机器学习; 天顶对流层延迟 (ZTD); 大气可降水量 (PWV); 气象数据; 短临降雨

中图分类号: P228.4　　**文献标志码:** A　　**文章编号:** 1008-9268(2022)04-0122-07

0　引　言

大气可降水量 (PWV) 是监控气候变化的重要一环. 以全球卫星导航系统 (GNSS) 技术为代表的水汽反演 PWV 方法在时间、空间、速度上占有优势, 在气象学领域中逐渐发挥作用[1]. 而降雨情况与 PWV 的动态特征变化关系, 让不少学者开始利用机器学习模型对降雨进行预报.

降雨预报模型包括降雨信息录入和气象参数因子获取、测试训练集规划确定、降雨预报模型的选择、模型参数的确定、降雨模型训练和建模结果分析等步骤[2]. 在获取准确的降雨信息和气象参数因子等关键数据后, 模型的选择问题是影响降雨预报结果的一个重要因素. 适用的预报模型能够模拟降雨与气象参数因子的数据关系, 利用线性或非线性函数构建两者之间的联系, 这种方法不需要再深入了解降雨发生背后的物理规律, 只需要通过挖掘历史数据 (气象参数、降水信息等) 的变化规律[3].

机器学习模型在降雨预报中表现出了良好的效果[4-5]. LIU 等[6] 基于一种新的空间框架, 将改进的 K 近邻 (KNN) 算法在遥感影像上分析了强降雨下影像的范围. HUANG 等[7] 利用改进的 KNN, 在降雨数据分布不均匀的情况下, 在降雨预报中取得了不错的效果. BOJANG 等[8] 将奇异谱分析与最小二乘支持向量机和随机森林 (RF) 结合, 可用于月降雨量的研究. SHI 等[9] 利用长短期记忆神经网络 (LSTM) 模型引入卫星遥感云图以时间序列建立降雨预报模型, 也取得不错的效果. 然而, 这些研究主要把机器学习算法应用在遥感影像和雷达图像. 因此, 另一批学者在 GNSS PWV 与机器学习的融合应用上进行探索, 尝试利用 GNSS 解算出来的天顶对流层延迟 (ZTD) 通过机器学习算法建立降雨预报模型. 周永江等[10] 利用 BP 神经网络融合气象参数、PWV 和 PM2.5 数据建立时间序列和回归的雾霾预测模型, 时效性达到 3

收稿日期: 2022-03-21

资助项目: 安徽省自然科学基金项目 (2208085MD101); 安徽省自然科学基金项目 (2108085QD171); 安徽高校自然科学研究重点项目 (KJ2021A0443); 安徽理工大学矿区环境与灾害协同监测煤炭行业工程研究中心开放基金 (KSXTJC202006); 安徽理工大学引进人才科研启动基金项目

通信作者: 赵兴旺 E-mail: xwzhao2008@126.com

h. 刘洋等[11]利用反向传播神经网络结合多种气象参数和 PWV 进行短临降雨预报, 比 BP 神经网络拥有更好的性能, 赵庆志等[12]利用最小二乘支持向量机(SVM) 对短临降雨进行预测, 相对传统降雨预测算法具有显著提升.

为了验证机器学习算法在降雨预报中的可靠性能, 本文在上述研究的基础上, 以几种典型机器学习算法构建短临降雨预报模型, 融合 PWV 和气象参数数据, 定量分析和比较这些机器学习算法在相同背景下的降雨预测性能, 研究和评价模型的可行性.

1　理论和数据

1.1　GNSS 获取 PWV

GNSS 信号在传播过程中会受到对流层延迟的干扰, 利用对流层延迟不仅可以改进 GNSS 定位的精度, 同时对水汽的研究有着重要作用. ZTD 可由斜路径方向上的对流层延迟通过映射函数投影在天顶方向上得到. GAMIT 解算的对流层延迟与国际 GNSS 服务 (IGS) 提供的对流层延迟产品具有很好的一致性[13]. 本文使用 IGS ZTD 产品代替 GAMIT 处理的 ZTD 延迟.

ZTD 由天顶对流层静力延迟 (ZHD) 和天顶对流层湿延迟 (ZWD) 两部分组成, 前者是 ZTD 中的主要成分, 可以通过 Saastamoinen 公式求得; 后者通过 ZTD 与 ZHD 之间作差求得. PWV 与 ZWD 之间的转换系数 (π) 由 Bevis 提出, 通过 ZWD 和 π 的乘积可以得到 PWV. 综上, PWV 的计算公式为

$$\begin{cases} \mathrm{ZHD}=\dfrac{0.002277P_0}{1-0.00266\cos(2\phi)-0.00028H} \\ \mathrm{ZHD}=\mathrm{ZTD}-\mathrm{ZHD} \\ \pi=\dfrac{10^6}{\left(\dfrac{k_3}{T_m}+k_2'\right)R_V\rho} \\ \mathrm{PWV}=\pi\,\mathrm{ZHD} \end{cases}. \quad (1)$$

式中: P_0 为地表气压, 单位为 hPa; ϕ 为测站纬度, 单位为度; H 为测站高程, 单位为 km; k_3 和 k_2' 表示经验常数, k_3=3.739×10^5 K$^2\cdot$g$\cdot$hPa^{-1}, k_2'= 16.48 K$\cdot$hPa^{-1}; R_V 表示水汽比气体常数, R_V=461$\left(\mathrm{J\cdot kg^{-1}\cdot K^{-1}}\right)$; ρ 表示液态水的密度, ρ=10^3 kg$\cdot$m^{-3}.

1.2　模型和算法

1.2.1　KNN 算法

KNN 算法是一种通过特征空间中的输入样本寻找 k 个距离最近邻的样本并依据所属类别投票表决的方法[14]. 距离的计算函数有欧几里得距离、巴氏距离和马氏距离等. 常用的欧几里得距离计算的是两个点距离之间的平方差之和的平方根, 计算公式为

$$D(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}. \quad (2)$$

式中, i 表示点 x 和 y 的第 i 个坐标. 通过 KNN 算法对目标进行分类, 输出值是 k 个最近邻样本类别中占比最大的一类. 可以通过手动设置或使用交叉验证结果较为准确的 k 值.

1.2.2　随机森林

随机森林 (RF) 在 Bagging 算法的基础上, 随机选取部分特征向量组成 CART (classification and regression tree) 决策树, 流程如图 1 所示, 重复 m 次建立 m 个决策树模型, 通过多颗决策树联合对结果进行预测.

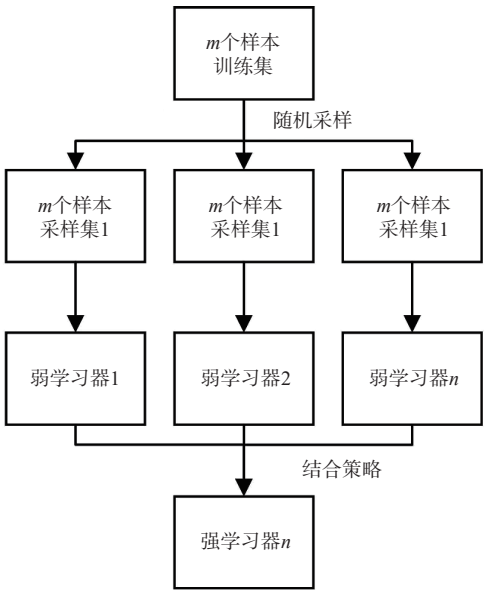

图 1　随机森林示意图

1.2.3　朴素贝叶斯分类器

朴素贝叶斯分类器 (NBC) 是贝叶斯分类器中常用的模型之一. 这种分类器假设特征向量之间独立, 降低了运算的逻辑性和复杂性. 在特征向量为 $\boldsymbol{x}$ 的情况下, 对目标进行归类时, 计算公式为

$$p(y=c_j|\boldsymbol{x})=\frac{p(y=c_j)\,p(\boldsymbol{x}|y=c_j)}{p(\boldsymbol{x})}. \quad (3)$$

对于特征向量的属性是连续性分布的二分类问题, 计算出变量正态分布的均值和方差, 可将公式转换为

$$p(y=+1|\boldsymbol{x})=p(y=+1)\frac{1}{Z}\prod_{j=1}^{n}\frac{1}{\sqrt{2\pi}\sigma_j}\exp\left(-\frac{(x_j-\mu_j)^2}{2\sigma_j^2}\right). \quad (4)$$

式中：Z 表示归归一化因子；μ_j 表示第 j 个特征向量的均值；σ_j 表示第 j 个特征向量的标准差；y =+1 表示样本归为正类的标签.

1.2.4 SVM

SVM 的目的通过寻找一个最具鲁棒性的超平面来将样本进行分类. 这个超平面让不同的样本类别分布在平面两侧, 同时让两侧距离决策边界最近的样本类别有一个极大值. 这个超平面用下面的式子表示：

$$y = \boldsymbol{w}^{\mathrm{T}} \cdot \boldsymbol{x} + b. \tag{5}$$

式中：$\boldsymbol{x}$ 为特征向量；$\boldsymbol{w}$ 表示超平面的归一化方向向量；b 表示阈值.

SVM 可以利用核函数将原始特征向量映射到新空间. 常用的核函数有线性核函数、多项式核函数和高斯核函数等. 在本次实验中, 使用了高斯核函数[15], 如下式所示：

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\gamma \cdot \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2\right). \tag{6}$$

式中：γ 表示高斯核的参数；$\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2$ 表示特征向量的欧几里得距离.

1.3 数据资料

数据选取位于北京 (BJFS) 和武汉 (WUH2) 2 个 GNSS 测站, 其中 ZTD 数据来自 IGS 提供的对流层延迟产品, PWV 由式 (1) 计算得到. 气象数据来自气象网站 rp5.ru, 由英国气象局制作并根据相关资质发布在该网站上, 提供的气象数据有温度 (T)、气压 (P)、相对湿度 (U)、露点温度 (T_d)、每 3 h 降雨量.

2 气象参数特征分析

降雨的发生往往伴随着复杂参数的变化, 研究降水形成过程中 PWV 和多尺度气象参数时间序列的周期性、敏感性等特征, 挖掘降雨的形成机理是有必要的. 图 2~3 分别为 BJFS 站和 WUH2 站降雨及相关其气象参数的时间序列变化. 由图可知, 降雨的发生与 PWV 及其气象参数的变化基本是一致的, 有比较强的相关性. 从全年的数据变化看, 在 PWV 的峰值到来时, 会伴随着降雨的发生；结合气象资料选择降雨较为集中的 180—210 天, 在降雨发生前, 通常伴随着 PWV、T_d 及 U 的上升, T 的下降, P 的陡峭上升；在降雨发生时, 通常伴随着 PWV、P、T_d 及 U 的下降, T 的上升.

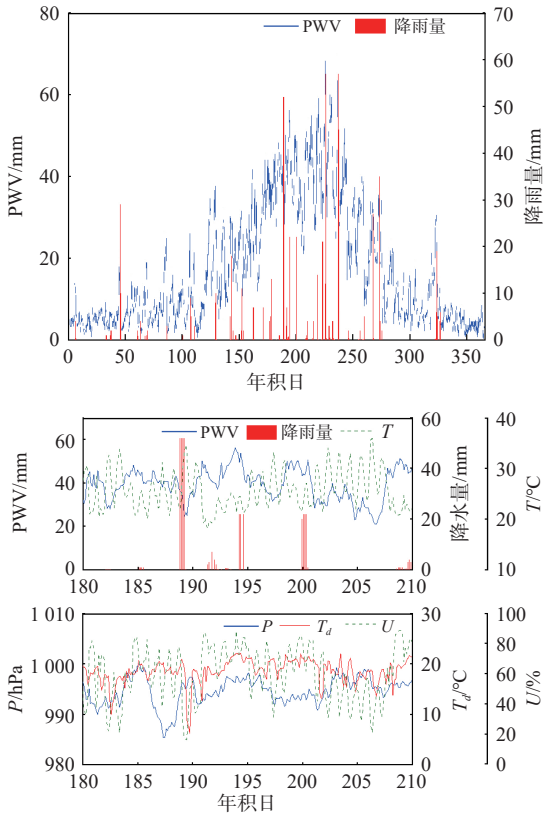

图 2 BJFS 站 2020 年降雨量与 PWV 关系以及 7 月 (年积日第 180—210 天) 降雨量与相关气象参数关系

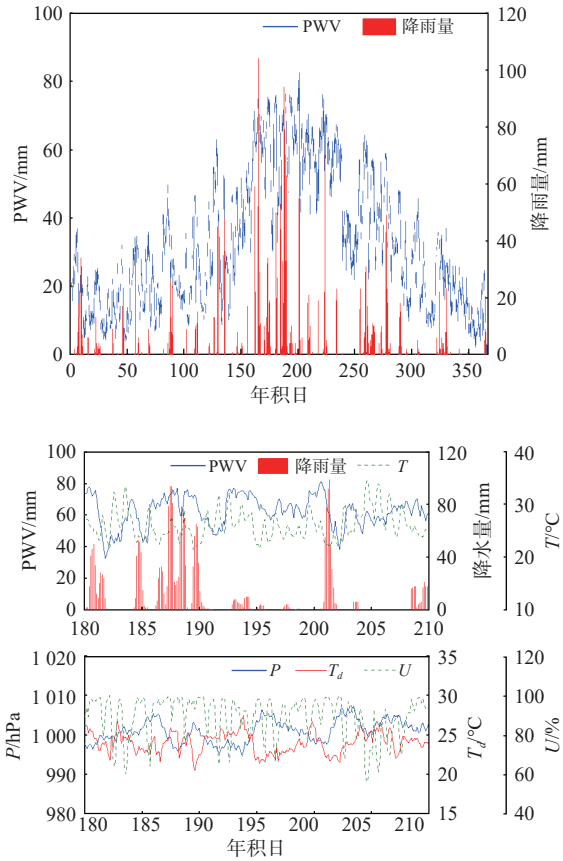

图 3 WHU2 站 2020 年降雨量与 PWV 关系以及 7 月 (年积日第 180—210 天) 降雨量与相关气象参数关系

3　基于机器学习的预报模型构建

3.1　预报流程设计

图 4 展示了区域短临降雨的一般预报框架.

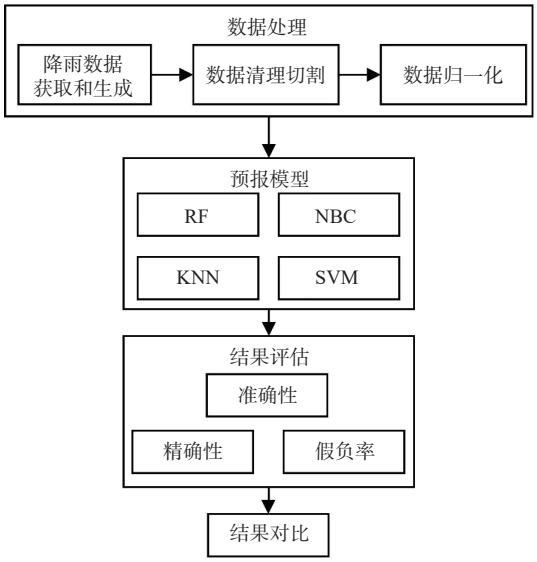

图 4　降雨预报模型流程

以 BJFS 站 2020 年的实验数据为例, 首先对 PWV 和气象参数进行归一化处理. 模型的参数对预报的精度起到重要作用, RF 模型的参数有树的数目和深度, KNN 的参数有权重和距离, SVM 的参数有正则化参数和惩罚参数, 本文利用网格搜索法和交叉验证的方式来确定模型的最优参数. 接着将预报因子 (PWV、T、P、T_d、U) 与降雨情况作为数据集输入模型中, 分别随机将数据集中的 70% 和 80% 作为训练集进行模型训练, 剩下的数据作为测试集进行模型验证, 得到 BJFS 站 2020 年的降雨预报模拟结果. WUH2 站的模拟实验流程与上述流程基本一致.

3.2　结果评价

本文使用准确性 (Accuracy)、精确率 (Precision) 和假负率 (FNR) 来评价降雨预报模型的精度

$$\begin{cases} \text{Accuracy} = \dfrac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \\ \text{Precision} = \dfrac{\text{TP}}{\text{TP}+\text{FP}} \\ \text{FNR} = \dfrac{\text{FN}}{\text{TP}+\text{FN}} \end{cases} . \tag{7}$$

式中: 将降雨预报的分类情况表示为混淆距阵, 具体如表 1 所示. TP 为实际情况降雨, 预报情况为降雨的样本数; TN 为实际情况不降雨, 预报情况为不降雨的样本数; FP 为实际情况不降雨, 预报情况为降雨的样本数; FN 为实际情况降雨, 预报情况为不降雨的样本数.

表 1　降雨预报混淆矩阵

实际值	预报值	
	降雨	不降雨
降雨	TP	FN
不降雨	FP	TN

图 5~7 为 BJFS 站和 WUH2 站 2020 年 100 次的降雨模拟结果, 由图可见, 2 个测站的降雨预报模拟都有不错的效果. BJFS 站 4 种模型不同百分比训练集准确性的平均值均约为 0.96, 精确率的平均值约为 80%, 假负率的平均值约为 21%; WUH2 站 4 种模型不同百分比训练集准确性的平均值约为 0.92, 精确率的平均值约为 86%, 假负率的平均值约为 13%. 而在 4 种模型中, RF 的模型在准确性和精确率上比其他 3 种模型更优一点, SVM 的模型在假负率上比其他 3 种模型更低一点.

传统的阈值方法利用降雨前的 PWV 的变化量和变化率进行短临降雨预报 [16], 表 2 对 BJFS 站和 WUH2 站的 PWV 变化量和变化率进行分析并确定合适的阈值, 模拟 2 个测站的降雨预报效果.

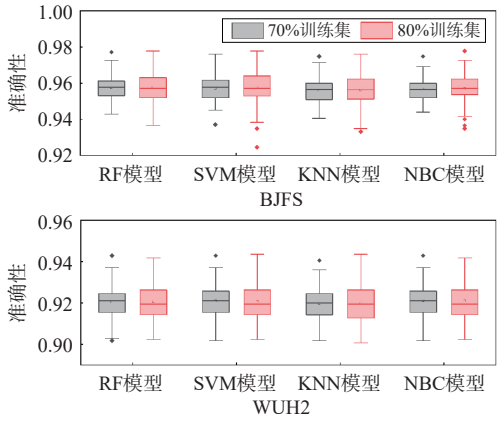

图 5　4 种预报模型的准确性箱图

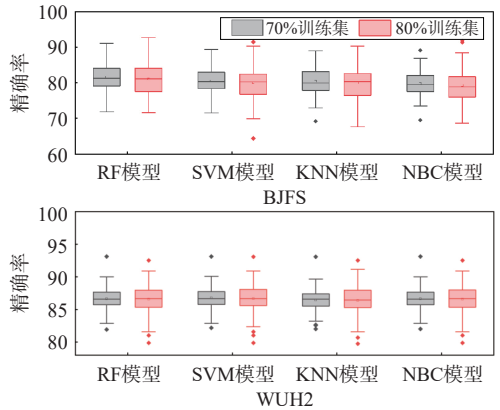

图 6　4 种预报模型的精确率箱图

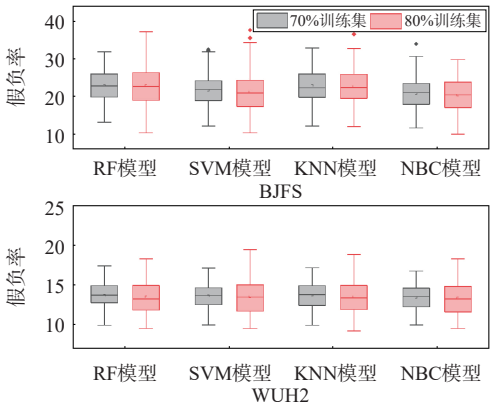

图 7 4 种预报模型的假负率箱图

表 2 BJFS 站和 WUH2 站降雨预报的统计结果

测站名	PWV变化量/mm	PWV变化率/(mm·h^{-1})	精确率/%	假负率/%
BJFS	2.5	0.6	79.2	66.1
WUH2	3.0	0.8	83.3	63.2

由表 2 可以看出, 选择合适的 PWV 变化量和变化率并利用阈值方法对降雨进行预报, 其精确率和假负率约在 80% 和 60%, 说明该方法在一定程度上能对未来短时间进行降雨预报, 但却有着不低的假负率, 对预报的应用存在一定的影响.

综上所述, 4 种模型在 BJFS 站和 WUH2 站的降雨预报都起到了不错的效果, 且漏报率低于传统的阈值方法判断降雨模型.

3.3 预报实验

以 BJFS 站为例, 按时间序列的方式选取年积日为第 150—200 天的数据作为训练集数据, 对数据集进行归一化处理输入预报模型中进行训练, 以 200—250 天的数据作为测试集数据, 预报下一时间段的短临降雨情况. 利用接收器操作特性 (ROC) 曲线和查准率一查全齐 (PR) 曲线对结果进行评估. WUH2 站的预报流程与上述流程基本一致.

图 8~11 为 BJFS 站和 WUH2 站的降雨预报结果. 由图可见, 2 个测站的降雨预报都取得不错的效果, BJFS 站的 ROC 曲线下与坐标轴围成的面积 (AUC) 值最好的是 SVM 模型的 0.923 80, 平均准确率 (AP) 值最好的是 SVM 模型的 0.790 92; WUH2 站的 AUC 值最好的是 SVM 模型的 0.924 30, AP 值最好的是 RF 模型的 0.821 86. 综上所述, SVM 模型的分类器性能略优于 RF 模型, 而 KNN 模型和 NBC 模型也能取得不错的效果. 因此, 本文基于机器学习的短临降雨预报模型对未来 3 h 的降雨预报能达到一个不错的效果, 可以达到 80% 以上的降雨情况, 而假负率在 20% 以下. 相对于传统的阈值预报模型, 在正确率相当的情况下 (其正确率约为为 80%), 假负率降低了 50% 左右 (其假负率约为 70%).

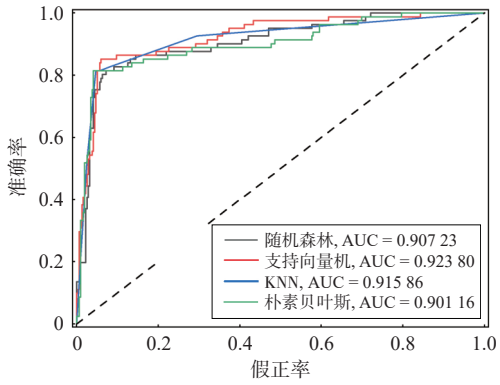

图 8 BJFS 站的 ROC 和 AUC 曲线

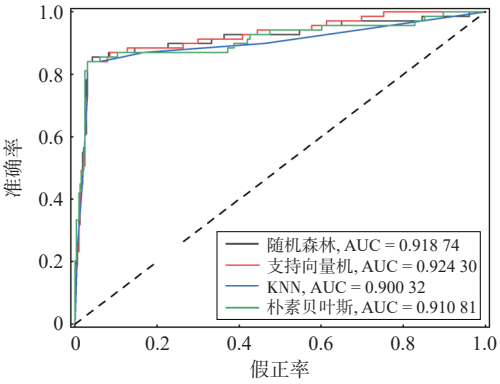

图 9 WUH2 站的 ROC 和 AUC 曲线

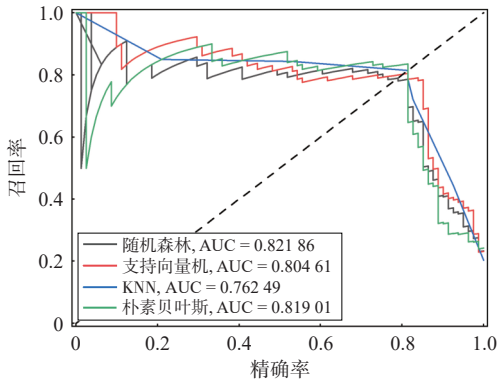

图 10 BJFS 站的 PR 和 AP 曲线

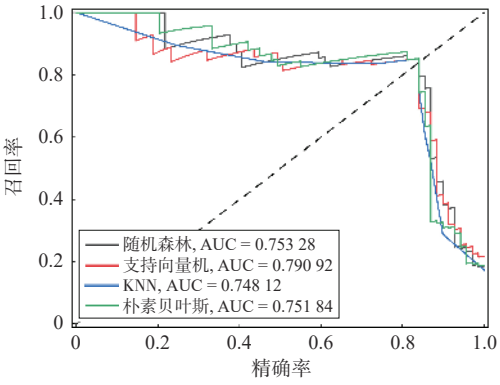

图 11 WUH2 站的 PR 和 AP 曲线

4　结　论

1) 通过分析降雨发生前后与 PWV 和多种气象参数 (T、P、T_d、U) 的一种非线性变化关系得出, 在降雨发生前, 会有 PWV、T_d、U 和 P 的上升过程, T 的下降, 而在降雨发生时, 这些参数发生相反的态势.

2) 利用不同的机器学习算法, 分别对测站整年的降雨数据划分不同的训练集构建短临降雨预报模型, 结果表明 4 种模型均能取得不错的效果, 准确性在 0.9 以上, 精确率在 80% 以上, 假负率在 25% 以下, 而 RF 模型在准确性和精确率上更优, SVM 的模型在假负率上更优.

3) 以时间序列构建的短临降雨预报模型的结果表明, 4 种模型对未来 3 h 的 80% 以上降雨情况可以很好的预报, 假负率在 20% 以下, 相较传统的阈值方法, 假负率降低了约 50%, 有了很大的改进. 其中 SVM 模型的综合性能略优, 在 BJFS 和 WUH2 测站上的 AUC 最好, BJFS 的 AP 最好, 其次是 RF 模型, 最后 KNN 模型和 NBC 模型也能取得不错的效果. 综上, 4 种典型机器学习构建的短临降雨预报模型具有不错的可行性.

致谢: 感谢 IGS 提供的 GNSS 数据, 感谢 rp5.ru 网站提供的气象数据.

参考文献

[1] HE Q, ZHANG K F, WU S Q, et al. Real-time GNSS-derived PWV for typhoon characterizations: a case study for super typhoon Mangkhut in Hong Kong[J]. Remote sensing, 2019, 12(1): 104. DOI: 10.3390/rs12010104

[2] FAYAZ S A, ZAMAN M, BUTT M A. Knowledge discovery in geographical sciences —a systematic survey of various machine learning algorithms for rainfall prediction[C]// International Conference on Innovative Computing and Communications, 2021: 593-608. DOI: 10.1007/978-981-16-2597-8_51

[3] 王江波. 长短期记忆网络在短临降雨中的应用[D]. 南京: 南京信息工程大学, 2021.

[4] AHMED K, SACHINDRA D A, SHAHID S, et al. Multi-model ensemble predictions of precipitation and temperature using machine learning algorithms[J]. Atmospheric research, 2020(236): 104806. DOI: 10.1016/j.atmosres.2019.104806

[5] YANG M X, WANG H, JIANG Y Z, et al. GECA proposed ensemble –KNN method for improved monthly runoff forecasting[J]. Water resources management, 2020, 34(11): 849-863. DOI: 10.1007/s11269-019-02479-2

[6] LIU S, LIU R, TAN N Z. A spatial improved-KNN-based flood inundation risk framework for urban tourism under two rainfall scenarios[J]. Sustainability, 2021, 13(5): 2859. DOI: 10.3390/su13052859

[7] HUANG M, LIN R, HUANG S, et al. A novel approach for precipitation forecast via improved K-nearest neighbor algorithm[J]. Advanced engineering informatics, 2017(33): 89-95. DOI: 10.1016/j.aei.2017.05.003

[8] BOJANG P O, YANG T-C, PHAM Q B, et al. Linking singular spectrum analysis and machine learning for monthly rainfall forecasting[J]. Applied sciences, 2020, 10(9): 3224. DOI: 10.3390/app10093224

[9] SHI X J, CHEN Z R, WANG H, et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems, 2015(1): 802-810. DOI: 10.48550/arXiv.1506.04214

[10] 周永江, 姚宜斌, 颜笑, 等. 融合 GNSS 气象参数的 BP 神经网络雾霾预测研究[J]. 大地测量与地球动力学, 2019, 39(11): 1148-1152.

[11] 刘洋, 赵庆志, 姚顽强. 基于多隐层神经网络的GNSS PWV和气象数据的降雨预测研究[J]. 测绘通报, 2019(S1): 36-40.

[12] 赵庆志, 刘洋, 姚顽强. 利用最小二乘支持向量机的短临降雨预测模型构建[J]. 大地测量与地球动力学, 2021, 41(2): 152-156.

[13] BYUN S H, BAR-SEVER Y E. A new type of troposphere zenith path delay product of the international GNSS service[J]. Journal of geodesy, 2009, 83(3): 367-373. DOI: 10.1007/S00190-008-0288-8

[14] HUANG S, HUANG M M, LYU Y J. An improved KNN-based slope stability prediction model[J]. Advances in civil engineering, 2020(11): 1-16. DOI: 10.1155/2020/8894109

[15] WANG H, ASEFA T, SARKAR A. A novel non-homogeneous hidden Markov model for simulating and predicting monthly rainfall[J]. Theoretical and applied climatology, 2021, 143(7): 627-638. DOI: 10.1007/s00704-020-03447-2

[16] 姚宜斌, 赵庆志, 李祖锋, 等. 基于全球导航卫星系统资料的短时降水预报[J]. 水科学进展, 2016, 27(3): 357-365.

作者简介

池钦 (1998—), 男, 硕士, 研究方向为 GNSS 数据处理与 GNSS 气象学.

赵兴旺 (1982—), 男, 博士, 教授, 研究方向为精密单点定位模型及实时应用研究.

陈健 (1993—), 男, 博士, 研究方向为 GNSS 精密定位与数据处理.

Short-term rainfall forecast by several typical machine learning algorithm

CHI Qin, ZHAO Xingwang, CHEN Jian

(*School of Spatial Informatics and Geomatics Engineering*, *Anhui University of Science and Technology*, *Huainan* 232001, *China*)

Abstract: According to the characteristic changes of precipitable water vapor and meteorological parameters (temperature (T), humidity (U), dew point temperature (T_d), surface pressure (P)) during the rainfall process, it is possible to establish a short-term rainfall forecast model based on machine learning algorithms. This paper uses the 3-hour zenith tropospheric delay and meteorological data of the bjfs station and wuh2 station in 2020 as examples to construct the prediction model of the four algorithms: random forest (RF), support vector machine (SVM), K-nearest neighbor (KNN), and naive bayes classifier (NBC), and introduces the rainfall events at each time as the new feature vector, adopts the segmentation method of 70% and 80% training sets respectively, takes the rainfall events as the model output, and the applicability of the model is evaluated by the accuracy, precision rate and false negative rate. After obtaining the accuracy is about 0.92, the precision rate is about 80%, and the false negative rate is about 20%, the data of 150—200 days in the time series are further used as samples to predict the rainfall of 200—250 days. The results indicate that The short-term rainfall forecast model based on machine learning can predict more than 80% of the rainfall events in the next 3 hours, and the false negative rate is below 20%, among which the SVM model has better comprehensive performance. Compared with the traditional threshold model, the accuracy rate is equivalent, and the false negative rate is decreased by about 50%.

Keywords: machine learning; zenith tropospheric delay; Precipitable Water Vapor (PWV); meteorological data; short-term rainfall