

基于改进的 Adaboost-BP 模型在降水中的预测

王 军^{1,2}, 费 凯^{1*}, 程 勇²

(1. 南京信息工程大学 计算机与软件学院, 南京 210044; 2. 南京信息工程大学 信息化建设与管理处, 南京 210044)

(* 通信作者电子邮箱 feikainiust@163.com)

摘 要: 针对目前分类算法对降水预测过程存在着泛化能力低、精度不足的问题, 提出改进 Adaboost 算法集成反向传播 (BP) 神经网络组合分类模型。该模型通过构造多个神经网络弱分类器, 赋予弱分类器权值, 将其线性组合为强分类器。改进后的 Adaboost 算法以最优归一化因子为目标, 在提升过程中调整样本权值更新策略, 以此达到最小化归一化因子的目的, 从而确保增加弱分类器个数的同时降低误差上界估计, 通过最终集成的强分类器来提高模型的泛化能力和分类精度。选取江苏境内 6 个站点的逐日气象资料作为实验数据, 建立 7 个降水等级的预报模型, 从对降雨量有影响的众多因素中, 选取 12 个与降水相关性较大的属性作为预报因子。通过多次实验统计, 结果表明基于改进的 Adaboost-BP 组合模型具有较好的性能, 尤其对 58259 站点的适应性较好, 总体分类精度达到 81%, 在 7 个等级中, 对 0 级降雨的预测精度最好, 对其他等级的降雨预测有不同程度的精度提升, 理论推导及实验结果证明该种改进可以提高预测精度。

关键词: 分类器; 改进 Adaboost; BP 神经网络; 组合模型; 权值调整; 归一化因子

中图分类号: TP399 **文献标志码:** A

Prediction of rainfall based on improved Adaboost-BP model

WANG Jun^{1,2}, FEI Kai^{1*}, CHENG Yong²

(1. College of Computer and Software, Nanjing University of Information Science and Technology, Nanjing Jiangsu 210044, China;

2. Information Construction and Management Department, Nanjing University of Information Science and Technology, Nanjing Jiangsu 210044, China)

Abstract: Aiming at the problem that the current classification algorithm has low generalization ability and insufficient precision, a combination classification model combining Adaboost algorithm and Back-Propagation (BP) neural network was proposed. Multiple neural network weak classifiers were constructed and weighted, which were linearly combined into a strong classifier. The improved Adaboost algorithm aimed to optimize the normalization factor. The sample weight update strategy was adjusted during the lifting process, to minimize the normalization factor, increasing the number of weak classifiers while reducing the error upper bound estimate was ensured, and the generalization ability and classification accuracy of the final integrated strong classifier was improved. A daily precipitation model of 6 sites in Jiangsu province was selected as the experimental data, and 7 precipitation models were established. Among the many factors influencing the rainfall, 12 attributes with large correlation with precipitation were selected as the forecasting factors. The results show that the improved Adaboost-BP combination model has better performance, especially for the site 58259, and the overall classification accuracy is 81%. Among the 7 grades, the prediction accuracy of class-0 rainfall is the best, and the accuracy of other types of rainfall forecast is improved. The theoretical derivation and experimental results show that the improvement can improve the prediction accuracy.

Key words: classifier; improved Adaboost; BP neural network; combined model; weight adjustment; normalization parameter

0 引言

随着社会经济发展以及人类本身对气象服务要求不断提升, 气象领域的气象数据采集渠道日益丰富, 数据规模不断增加, 且其具有的空间属性、高维性、不稳定性, 为研究传统气象预报模式增加巨大难度, 尤其在研究各气象要素之间内部联

系时, 尤为乏力, 从而导致获取到的大量气象资料并未有效利用, 对于推动气象模式预报发展并未有实质性作用。

气象数据是典型的时间序列数据, 天气系统的内部相互影响条件错综复杂, 对于大量的采集数据进行分析处理时, 传统的气象研究模式无法发现其隐含的价值, 而数据挖掘技术为研究大量气象数据提供了新途径, 为气象领域中发现各属

收稿日期: 2017-03-17; 修回日期: 2017-05-18。 基金项目: 国家自然科学基金资助项目(61402236, 61373064); 江苏省农业气象重点实验室开放基金资助项目(KYQ1309); 江苏省“六大人才高峰”项目(2015-DZXX-015, 2013-DZXX-019); 江苏省产学研前瞻性联合研究项目(BY2014007-2); 公益性行业(气象)科研专项(GYHY201106037)。

作者简介: 王军(1970—), 男, 安徽铜陵人, 教授, 硕士, CCF 会员, 主要研究方向: 无线传感网、大数据; 费凯(1990—), 男, 江苏连云港人, 硕士研究生, 主要研究方向: 大数据气象领域应用; 程勇(1980—), 男, 重庆人, 高级工程师, 博士, CCF 会员, 主要研究方向: 无线传感网、大数据。

性联系发挥着重要作用,分类挖掘技术通过有监督的学习探索历史气象数据中潜在的规律,可以提高气象预报模式的准确率^[1]。

近些年,多种分类挖掘技术被应用于气象领域,如决策树、关联规则、贝叶斯等^[2-3]。文献[4]将支持向量机(Support Vector Machine, SVM)分类方法应用于降水预测中,结合多个相关因子,利用四川盆地降水量数据建立分类预测模型,在短气候预测中有着不错的性能。文献[5]研究通过决策树对与降水密切相关的湿度、温度、大气压、风速和露点温度等众多气象要素的采集数据进行预报建模,改进后的模型具有较好的适应性;文献[6]将朴素贝叶斯分类器与降雨量问题相结合,将贝叶斯模型用于降雨量分类预测研究,以郑州市 34 年气象数据建立分类预测模型,预测精度略好于短期气候中常用的回归分析;文献[7]提出一种改进 SVM 算法的二进制分类核函数(Binary and Ordinal classification Kernel method, BIORK)模型,将该模型用于降雨等级预测,并且证明该模型在气象要素合适的情况下有较好的预测精度,尤其是无降雨天气预测。但是这些模型共同点是基于较小的数据集为前提才获取到较好的效果,不适用于规模较大和多站点的数据集。文献[8]运用 SVM 分类算法研究局部气象数据,提出基于 Adaboost 算法集成 SVM 分类器的多分类模型,通过多个分类器的协同组合多个基分类器,以此提高单分类模型在降雨中的分类预测精度;文献[9]提出以 Adaboost 算法集成 BP 神经网络模型预测了湖北某风场风速,对每个弱学习器赋予权值,最终线性组合成强预测器对风速预测,通过该多集成方法提高了学习器的预测能力。相对单个弱学习器而言,数据集较大时,Adaboost 集成弱学习器模型具有一定的优势,但随着迭代次数的增加,分类器的性能开始下降,出现了性能退化问题,很难达到理想中的要求。

针对上述问题,本文提出一种改进 Adaboost-BP(可简称为 Ada-BP)算法,该算法集成 BP 神经网络,使得最终的强分类预测模型泛化能力增强,通过破坏样本分布的均匀性缓解性能退化问题。该算法核心是在 Adaboost 算法中线性组合多个神经网络分类器得到强分类器,并利用该分类器完成分类预测,通过在迭代过程中改变权值分布策略,以优化归一化因子,降低误差上界,从而提高强分类器的分类能力^[10]。

1 BP 神经网络

BP 神经网络是误差反向传播算法,可以逼近任意连续的函数,具有很强的非线性映射能力,它由输入层、隐含层、输出层构成,隐含层是介于输入层和输出层之间^[11]。隐含层可以由多层构成,但隐含层的增加导致模型的复杂性增加,文献[12]研究发现三层结构的 BP 神经网络可以近似任何多元多项式函数,因此,在本文研究过程中使用只包含一层隐含层的 BP 神经网络来建立模型,其模型拓扑结构如图 1 所示,其中 BP 神经网络的学习过程分为两个部分:正向传播和反向传播。当正向传播的时候,信号通过神经元传递到下一层,如果输出层不能获得所需的输出,误差信号由各神经元反向传播来调整各神经元参数。通过这两个交替的过程完成信息提取和存储,使得输出层值逼近期望的值。

其中 x_1, x_2, \dots, x_q 是网络输入层各神经元的输入值, w_{ij} 为输入层到隐含层神经元间的连接权值, w_{jk} 为隐含层到输出层的连接权值, y_i 为神经网络的输出值。

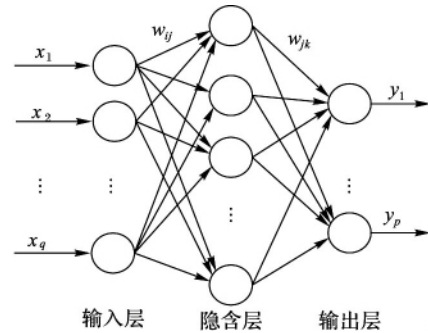


图 1 BP 神经网络拓扑结构

Fig. 1 Topological structure of BP neural network

2 基于改进的 Adaboost-BP 模型

Adaboost 算法是一种迭代算法,常用于解决分类问题。其核心思想是针对同一个训练集训练不同的弱分类器,后将多个弱分类器集成,构成最终的强分类器^[13]。该算法迭代的过程是一个对弱分类算法提升过程,重视被错误分类的样本和性能较高的分类器,提高训练中错误样本的权值和性能较好的分类器权值,反之则降低,分类的结果由多个性能较好的弱分类器决定;同时,在训练中,对训练效果较好的样本降低权重,而训练效果较差的增加其权重,通过多次训练迭代,得到由多个弱分类器线性组合的强分类器,以此提高对数据的分类精度。

本文将 BP 神经网络作为弱分类器,选取神经网络中的合适的激励函数与神经元参数来构造 BP 神经网络分类器,利用 Adaboost 算法多次提升训练,最终获得强分类器,其模型结构如图 2。

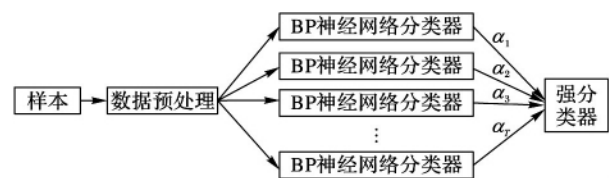


图 2 Adaboost-BP 模型结构

Fig. 2 Structure of Adaboost-BP model

通过研究分析系列 Adaboost 算法以及原理推导,每次迭代过程中,算法会对错误样本进行加权,随着迭代次数的增加,当多个易错样本被多次错误分类后,易错样本具有较大的权值,反之样本的总体误差的出现了上升趋势。对于该种退化问题,只能选取弱分类器空间中的唯一参照标准 Z_t 作分析^[14]。对 Z_t 作如式(1)推导:

$$\begin{aligned}
 Z_1 * Z_2 * Z_3 * Z_4 * \dots * Z_T &= \\
 Z_1 * Z_2 * Z_3 * Z_4 * \dots * Z_{t-1} * \sum_{i=0}^N D_t(x_i) * e^{-\alpha_t y_i C(x_i)} &= \\
 Z_1 * Z_2 * \sum_{i=1}^N D_t(x_i) \prod_{i=3}^N e^{-\alpha_t y_i C(x_i)} &= \\
 \sum_{i=0}^N D_1(x_i) * e^{-\alpha_1 y_i C(x_i)} * \sum_{i=1}^N D_2(x_i) * e^{-\alpha_2 y_i C(x_i)} * \sum_{i=0}^N D_3(x_i) \prod_{i=3}^N e^{-\alpha_t y_i C(x_i)} &= \\
 D_1(x_i) * \sum_{i=1}^N D_2(x_i) \prod_{i=2}^N e^{-\alpha_t y_i C(x_i)} &= \frac{1}{N} \sum_{i=1}^N e^{-y_i H(x_i)} \quad (1)
 \end{aligned}$$

由此可知,归一化因子虽然具有使样本呈现概论分布的

同时,还可以左右整个分类器的误差上界,如上面公式推导,实质上归一化因子的积就是整个分类器的误差上界。理论上随着分类器的增加误差上界将应该下降,但实际上随着迭代的增加,归一化因子在后期在持续增长趋势,从而导致随着分类器数量的增加性能反而降低的情况。对强分类器的性能优化实质上可以回归到是对归一化因子 Z_t 优化。 Z_t 的大小决定着误差下降的速度,对 Z_t 的优化使得其在每次迭代中都小于 1,当分类器的个数达到无穷时,分类器的误差上界理论值可以达到无限接近 0。对于分类器而言,归一化因子 Z_t 的公式分解为如式(2):

$$\sum_{i=1}^N D_t(x_i) * e^{-\alpha_t y_i C(x_i)} = \sum_j \left[\sum_{y=1}^N D_t(x_i) * e^{-\alpha_t} + \sum_{y=-1}^N D_t(x_i) * e^{\alpha_t} \right] = 2 \sum_j \left[\sqrt{p_t^i * q_t^i} \right] \quad (2)$$

其中: $p_t^i = \sum_{y=-1}^N D_t(x_i)$, $q_t^i = \sum_{y=1}^N D_t(x_i)$ 。通过将分类器代入归一化因子 Z_t 的式中整理推导,其公式可以被衍生到对极小值的求解问题,由于 $p_t^i + q_t^i = 1$ 是定值,对 Z_t 的优化思路可以被视为对 Z_t 的极小化,对于和为定值的两数之积研究分析,两数的之差越小,则积越大,对 Z_t 的极小化也就可以转换为两数之积的极小值。 p_t^i 和 q_t^i 是实质也就是错分与正分样本权值的总和。分析发现,当迭代次数的增加,错分的样本权值持续的增加,正确分类的样本权值总和减少,从而导致 p_t^i 和 q_t^i 两数之差减小,而这两个数值的大小直接决定因素是权值更新式(3)如下:

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} * \begin{cases} e^{\alpha_t}, & C_t(x_i) \neq y_i \\ e^{-\alpha_t}, & C_t(x_i) = y_i \end{cases} \quad (3)$$

其中: α_t 为弱分类器被赋予的权值,弱分类器权值公式满足关系是 $\alpha_t > 0$,由此可知弱分类器上的权值只有在大于 0 的情况下才有意义,由此可推导: $e^{\alpha_t} > 1 \Rightarrow e^{\alpha_t} + 2e^{\alpha_t} > 1 + 2e^{\alpha_t} \Rightarrow e^{\alpha_t} > (1 + 2e^{\alpha_t})/3 \Rightarrow (D_t(i) * e^{\alpha_t})/Z_t > (D_t(i) * (1 + 2e^{\alpha_t})/3)/Z_t$,由推导可知,将错误样本权值更新公式调整为 $D_{t+1}(i) = D_t(i) * (1 + 2e^{\alpha_t})/(3 * Z_t)$,通过该方式跟新样本权值,既满足分类器权值大于 0 的条件,也可以使得错分样本的权值小于原更新方式,抑制错误样本权值累积过大,从而抑制 p_t^i 的增长。同样,对于正确分类的权值更新使其满足关系 $\alpha_t > 0$ 的条件且增大其权值, $\alpha_t > 0 \Rightarrow 0 > -\alpha_t \Rightarrow 1 > e^{-\alpha_t} \Rightarrow 1 + 2e^{-\alpha_t} > e^{-\alpha_t} + 2e^{-\alpha_t} \Rightarrow (1 + 2e^{-\alpha_t})/3 > e^{-\alpha_t} \Rightarrow (D_t(i) * e^{-\alpha_t})/Z_t < D_t(i) * [(1 + 2e^{-\alpha_t})/3]/Z_t = D_{t+1}(i)$ 。相比原更新公式的赋值方式,改进后的更新公式对正确样本权值赋予相对提高,所以,将样本在迭代过程中的权值更新方式调整为式(4):

$$D_t(i) = \frac{D_t(i)}{3 * Z_t} * \begin{cases} 1 + 2e^{\alpha_t}, & C_t(x_i) \neq y_i \\ 1 + 2e^{-\alpha_t}, & C_t(x_i) = y_i \end{cases} \quad (4)$$

以减小学习误差为最终目的研究发现,样本的误差上界与归一化因子以及样本权值分布存在着一定联系,一般的 Adaboost 算法对样本分布调整时总是期望样本呈现均匀分布,但随着迭代过程中分类器的增加,性能开始退化,公式推导证明该种现象是由后期迭代过程中归一化因子的增大而导致总体误差的上升。而本文改进的 Ada-BP(Adaboost-BP) 算

法通过调整样本权值更新方式来打破现有算法中样本的分布均匀性,由式(4)在迭代过程中对样本调整权值更新策略,与原权值更新方式相比,增大了正确样本的权值,减小了错误样本的权值,目的是使得 p_t^i 和 q_t^i 两数差距拉大,从而直接优化归一化因子 Z_t ,使得迭代过程中归一化因子总是呈现小于 1 的趋势,以确保每次增加 BP 神经网络分类器时都能降低误差的上界,最终达到降低总体误差目的,从而提高改进的 Ada-BP 性能。

算法详细步骤如下:

步骤 1 设置初始化神经网络的权值和阈值。

步骤 2 数据归一化处理,使样本值分布在 [0, 1]。

步骤 3 训练神经网络调整隐含层节点数,设定学习率。

步骤 4 给出气象数据样本集 $(x_1, y_1), \dots, (x_m, y_m)$ (y_1, y_2, \dots, y_m 是样本的期望输出)。

步骤 5 初始化气象数据样本权值 $D(i) = 1/N$ (N 是样本的数量)。

步骤 6 for $t = 1, 2, \dots, T$ (T 是提升的次数)。

步骤 7 $\varepsilon_t = \frac{\sum_{i=1}^N D_t[C_t(x_i) \neq y_i]}{\sum_{i=1}^N D_t(x_i)}$ (计算弱分类器错误率 y_i 为期望值)。

步骤 8 if $\varepsilon_t > 0.5$ then 转步骤 5

步骤 9 End if: 权值计算 $\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right)$ 。

步骤 10 通过式(4)更新样本权值。

步骤 11 跳转第步骤 6,直至结束迭代。

步骤 12 $H(x) = \arg \max \left(\sum_{t=1}^T \alpha_t C_t(x_i) \right)$ (得到最终的强分类器)。

3 实验及分析

降水形成过程涉及的因素众多,其中基本气象要素包括温度、高度、纬向风、经向风、垂直速度等。利用这些基本气象要素通过动力诊断获取与降水相关的平流项、梯度项、涡度、水汽通量、水汽通量散度等近 100 多个的气象物理量,在众多的物理量中选择与降水相关因子,将直接决定降水分类模型的性能效果。将降水所涉及的所有因子作为属性将大大削弱模型的性能且增加模型的复杂度,选取与区域降水相关性较大的因子作为降水样本的属性集有利于提高分类预报效果。

一般对降水因子的选择通过逐步回归的方法,利用 Pearson 相关系数法对属性重要性进行排序,在众多因子中选择与降水相关性最好的属性。本文从众多因子排序中,选择包括最小相对湿度、平均相对湿度、极大风速、最大风速、日照时数、平均水气压、日最低温度、平均风速、平均气温、蒸发量、相对湿度、露点温度作为样本属性中的重要因子,所以在降水分类预测中只选择 12 个相关性较大因子。

3.1 降水模型建立

本研究的内容是对降水等级预测,将所选的 12 个属性作为网络的输入向量,通过所选的重要属性对降水等级分析,将降水的等级作为网络的输出向量,按国家气象局颁布的降水等级划分标准,建立 7 个等级区域降雨预报模型,表 1 为样本

降水等级统计表。

表 1 降水等级划分
Tab. 1 Level of precipitation

降雨等级	24 h 降水量/mm	等级	降雨等级	24 h 降水量/mm	等级
无雨	<0.1	0	暴雨	50~99.9	4
小雨	0.1~9.9	1	大暴雨	100~249.9	5
中雨	10.0~24.9	2	特大暴雨	>250	6
大雨	25.0~49.9	3			

实验数据来源于江苏省气象局,该数据集为 2001 年—2010 年 10 年间的实况观测数据,实验中选取江苏境内包括南通(58259)、徐州(58027)、盐城(58154)、淮安(58141)、泰州(58246)、扬州(58245)共 6 个站点逐日降水资料作为降水实验数据,实况数据数据集共包含 19863 条记录,将实验数据按比例分为训练数据与预测数据,同时对降水数据中各类别按比例划分训练数据与测试数据。由于气象数据存在的特点,经过预处理后的各个站点的数据量并不相同,各等级间的数据非均匀分布。

其中整个网络采用 12-N-7 的三层网络结构,输入层分别将上述的 12 个特征因子作为网络输入的特征向量,其中的 N 为隐含层神经元数目, N 的值结合经验公式,在一定的范围内采取试算的原则,输出层设置 7 个神经元,将输出层中 7 个降水等级作为神经网络的输出向量,神经网络选用的传递函数为 tansig 函数,学习规则采用梯度下降法,在神经网络训练前,通常会对数据进行归一化处理,在训练模型时先将数据预处理归一化到 0~1^[15]。

3.2 实验结果分析

为了验证本算法的先进性,通过 Matlab 编制程序,将本文提出的改进 Ada-BP 与传统 Ada-BP 算法作实验对比,说明改进的 Ada-BP 优于传统的 Ada-BP;同时,与性能较好的文献[7]中所用的 BIORK 算法做对比分析来说明本文提出的方法先进性。为保证实验可比性,实验中采用相同的数据集,将实验获得的结果作统计分析。

经过对 BP 神经网络训练最终确定单个性能较好的神经网络参数,选择较好的 BP 神经网络模型作为训练过程中的弱分类器。首先,对比改进后的 Adaboost 算法与改进前的 Adaboost 算法提升不同次数准确率之间的统计比较,如图 3 所示,随着弱分类器个数的增加,强分类器对样本的分类精度也随之提高,改进后的 Adaboost 算法弱分类器个数达到 20~30 时,性能趋于稳定,当 $T=20$ 时,性能达到最高,改进后的 Adaboost 算法的分类精度稳定地高于传统 Adaboost 算法 3%,但两者在随着弱分类器数量增加以后同时出现性能饱和情况。

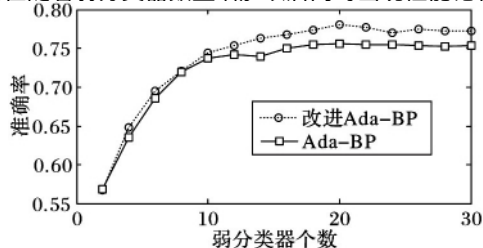


图 3 弱分类器上精度比较
Fig. 3 Comparison accuracy of prediction with different number of weak classifiers

图 4 分别统计了改进后 Ada-BP、Ada-BP、BIORK 三种模型在不同样本集容量下的性能:改进后的 Ada-BP 分类器性能随着样本的增加而上升,最终性能达到饱和,稳定在 77%~79%;Ada-BP 模型随着样本的增加,性能呈上升趋势,最终也到达饱和水平,但随着样本数量的增加,传统 Adaboost 集成的强分类器模型性能出现退化,低于性能峰值较多,最终稳定在 74%~76%;BIORK 单分类器模型的性能在样本集较少时,拥有明显的优势,随着样本的增加,BIORK 模型的性能逐渐趋于稳定,最终饱和于 75% 左右。从统计结果看,随着样本集数量的增加,改进后 Ada-BP 算法在总体分类精度上具有明显优势。

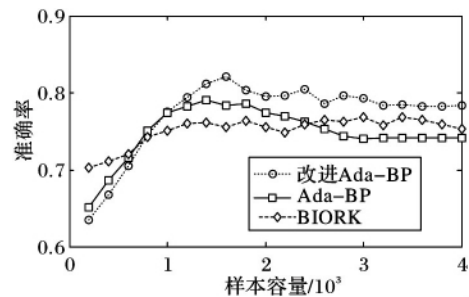


图 4 样本容量与性能关系比较
Fig. 4 Comparison accuracy of prediction with different capacity of samples

图 5 统计了三种算法对每个类样本的分类预测精度,从总体趋势显示:改进的 Ada-BP 算法对所有等级的分类预测性能都好于传统 Ada-BP 和 BP 两种算法;与其他等级相比,三种算法对 0 级降水同时表现出较好的性能,对 6 级的降水表现的性能最差,但改进后的 Ada-BP 算法在降水为 0 级优势较明显,降水等级为 2、3 时略低于 BIORK 分类模型。由于样本集数据所占比重不同,导致分类器对每个等级呈现不同的性能,但三种模型同时表现出对 0 级的样本预测精度最高,6 级的预测精度最低,由于等级为 6 的数据占比最少,且大暴雨级别及以上等级的降雨形成原因复杂,不能仅仅根据十几个相关性较强的因子作为判断依据,从而导致样本预测效果较差。

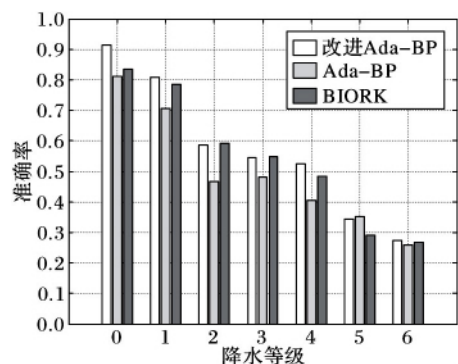


图 5 各降水等级样本预测性能对比
Fig. 5 Comparison accuracy of prediction with different grades of rainfall

图 6 统计了三种算法对各区站点的预测精度:改进的 Ada-BP 在各个站点中的预测准确率较为稳定,总体性能在 78% 以上,其中主要对南通站(58259)的预测精度较高,对扬州站(58245)的预测精度略低;BIORK 模型在淮安站(58141)与扬州站(58245)略高于改进的 Ada-BP 模型;但改进的 Ada-BP 模型对其余站点的适应性优于其他两种模型,所有站点的预测精度有一定程度的提高,其中改进的模型在南通站

(58259) 与徐州站的优势最大, 明显高于 Ada-BP 与 BIORK 两种模型。从各站点实验结果表明, 改进的 Ada-BP 模型优于其他两种模型。

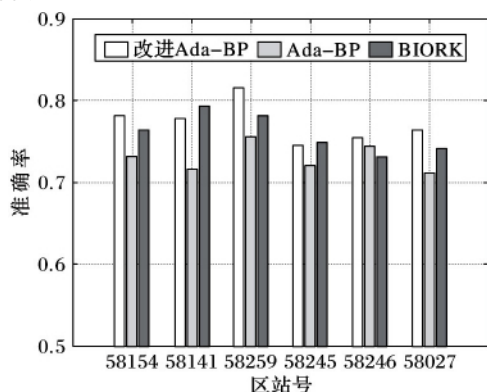


图6 各车站预测精度对比

Fig. 6 Comparison accuracy of prediction for different stations

通过上述实验表明, 本文提出的改进 Adaboost-BP 算法性能高于另外两种算法, 在降水等级预测中取得良好效果, 但在不同的站点和降水等级中表现出差异性, 在一定程度上改进了传统 Adaboost 算法集成 BP 神经网络和现有的分类预测算法对气象数据挖掘的局限性。

4 结语

本文提出的改进 Adaboost-BP 模型, 在集成迭代过程中调整样本权重更新方式, 以最小化归一化因子为目标, 减轻强分类器的性能退化现象。实验结果表明, 该模型优于现有的降雨预测模型, 一定程度上克服气象分类挖掘算法精度较低的缺陷, 提高了对不同样本的分类预测能力, 但对复杂的、降雨量大的天气预测精度较低。对算法作改进来提高其预测的精度有较大难度, 将成为以后降水预测模型的研究方向。但总体而言, 本文提出的模型确实提高了预报的准确率, 具有较高的工程价值。

参考文献 (References)

- [1] OLAIYA F, ADEYEMO A B. Application of data mining techniques in weather prediction and climate change studies [J]. International Journal of Information Engineering & Electronic Business, 2012, 4(1): 304-310.
- [2] GUO Z Y, DAI X Y, LIN H. Application of association rule in disaster weather forecasting [J]. Geographic Information Sciences, 2004, 10(1): 68-72.
- [3] CANO R, SORDO C, GUTIÉRREZ J M. Applications of Bayesian networks in meteorology [M]// Advances in Bayesian Networks. Berlin: Springer, 2004: 309-328.
- [4] 杨淑群, 芮景析, 冯汉中. 支持向量机(SVM)方法在降水分类预测中的应用[J]. 西南农业大学学报(自然科学版), 2006, 28(2): 252-257. (YANG S Q, RUI J X, FENG H Z. Application of Support Vector Machine (SVM) in rainfall categorical forecast [J]. Journal of Southwest Agricultural University (Natural Science), 2006, 28(2): 252-257.)
- [5] PRASAD N, KUMAR P, NAIDU M. An approach to prediction of precipitation using Gini index in SLIQ decision tree [C]// Proceedings of the 2013 4th International Conference on Intelligent Systems, Modelling and Simulation. Washington, DC: IEEE Computer Society, 2013: 56-60.
- [6] 何伟, 孔梦荣, 赵海青. 基于贝叶斯分类器的气象预测研究[J]. 计算机工程与设计, 2007, 28(15): 3780-3782. (HE W, KONG M R, ZHAO H Q. Research on meteorological prediction with Bayesian classifier [J]. Computer Engineering and Design, 2007, 28(15): 3780-3782.)
- [7] SÁNCHEZ-MONEDERO J, SALCEDO-SANZ S, GUTIÉRREZ P A, et al. Simultaneous modelling of rainfall occurrence and amount using a hierarchical nominal-ordinal support vector classifier [J]. Engineering Applications of Artificial Intelligence, 2014, 34(3): 199-207.
- [8] 滕少华, 樊继慧, 陈潇, 等. SVM 多分类器协同挖掘局域气象数据[J]. 广西大学学报(自然科学版), 2014(5): 1131-1137. (TENG S H, FAN J H, CHEN X, et al. Application of SVM-based multi-classifiers in mining cooperatively local area meteorological data [J]. Journal of Guangxi University (Natural Science Edition), 2014(5): 1131-1137.)
- [9] 吴俊利, 张步涵, 王魁. 基于 Adaboost 的 BP 神经网络改进算法在短期风速预测中的应用[J]. 电网技术, 2012, 36(9): 221-225. (WU J L, ZHANG B H, WANG K. Application of Adaboost-based BP neural network for short-term wind speed forecast [J]. Power System Technology, 2012, 36(9): 221-225.)
- [10] ZHU J, ROSSET S, ZOU H, et al. Multi-class AdaBoost [EB/OL]. [2017-01-03]. <https://web.stanford.edu/~hastie/Papers/samme.pdf>.
- [11] 张丹, 韩胜菊, 李建, 等. 基于改进粒子群算法的 BP 算法的研究[J]. 计算机仿真, 2011, 28(2): 147-150. (ZHANG D, HAN S J, LI J, et al. BP algorithm based on improved particle swarm optimization [J]. Computer Simulation, 2011, 28(2): 147-150.)
- [12] JIANG C. Review of back propagation neural network applied to athletics [C]// Proceedings of the 2012 24th Chinese Control and Decision Conference. Piscataway, NJ: IEEE, 2012: 2371-2375.
- [13] LI X, WANG L, SUNG E. AdaBoost with SVM-based component classifiers [J]. Engineering Applications of Artificial Intelligence, 2008, 21(5): 785-795.
- [14] 廖红文, 周德龙. AdaBoost 及其改进算法综述[J]. 计算机系统应用, 2012, 21(5): 240-244. (LIAO H W, ZHOU D L. Review of AdaBoost and its improvement [J]. Computer System & Applications, 2012, 21(5): 240-244.)
- [15] LI J, WU X H, QIN C H, et al. The design of image compression with BP neural network based on the dynamic adjusting hidden layer nodes [J]. Advanced Materials Research, 2012, 433/434/435/436/437/438/439/440: 3797-3801.

This work is partially supported by the National Natural Science Foundation of China (61402236, 61373064), Opening Foundation of Jiangsu Key Laboratory of Agricultural Meteorology (KYQ1309), Six Talent Peaks Project in Jiangsu Province (2015-DZXX-015, 2013-DZXX-019), the Prospective Joint Research Project of Jiangsu Province (BY2014007-2), Special Fund for Meteorological Research in Public Interest (GYHY201106037).

WANG Jun, born in 1970, M. S., professor. His research interests include wireless sensor network, big data.

FEI Kai, born in 1990, M. S. candidate. His research interests include application of big data in meteorological field.

CHENG Yong, born in 1980, Ph. D., senior engineer. His research interests include wireless sensor network, big data.