

DOI:10.19651/j.cnki.emt.2211300

基于自适应 SA-PSO 改进的 XGBoost 气温预测方法

王 强¹ 秦华旺¹ 齐春帅¹ 王琴琴²

(1.南京信息工程大学电子与信息工程学院 南京 210044; 2.山西师范大学地理科学学院 临汾 041081)

摘要: 提出了一种 SA-PSO-XGBoost 预测模型,基于 2016 年 1 月 1 日~2017 年 12 月 31 日的 ECMWF 气象数据,用于预测南京地区 6 h 后的气温。将气象数据分为训练集和测试集,使用 PCA 降维方法对气象数据特征进行压缩降维,然后应用模拟退火和粒子群优化混合算法对 XGBoost 模型的超参数进行优化,并将测试集数据带入到 SA-PSO-XGBoost、XGBoost、GRU 和 LSTM 神经网络进行对比分析,实验结果表明:SA-PSO-XGBoost 预测 6 h 后的温度模型有更高的准确性和鲁棒性。

关键词: XGBoost;PCA;粒子群优化算法;模拟退火算法;Python

中图分类号: P423;F224-39 **文献标识码:** A **国家标准学科分类代码:** 520.20

Improved XGBoost temperature prediction method based on SA-PSO

Wang Qiang¹ Qin Huawang¹ Qi Chunshuai¹ Wang Qinqin²

(1. Electronics and Information Engineering College, Nanjing University of Information Science and Technology, Nanjing 210044, China;

2. School of Physics and Information Engineering, Shanxi Normal University, Linfen 041081, China)

Abstract: Proposes a SA-PSO-XGBoost prediction model for forecasting the temperature in Nanjing, based on ECMWF meteorological data from January 1, 2016, to December 31, 2017. The meteorological data is divided into training and testing sets. The PCA dimension reduction method is applied to compress and reduce the features of the meteorological data. The SA-PSO-XGBoost model optimizes the hyperparameters using a hybrid algorithm combining simulated annealing and particle swarm optimization. The testing set is then used to compare the performance of the SA-PSO-XGBoost model with XGBoost, GRU, and LSTM neural networks. Experimental results demonstrate that the SA-PSO-XGBoost model outperforms others in terms of accuracy and robustness in predicting the temperature 6 hours ahead.

Keywords: XGBoost;PCA;particle swarm optimization algorithm;simulated annealing algorithm;Python

0 引 言

气温与人们的日常生活、生产生活和科学研究等方面休戚相关,对人们的生活规划具有指导意义。气候千变万化,有时会引起高温,冰雹雷雨等自然灾害,准确的预测气温是我们的当务之急。由于影响气温的因素很多,并且气温的变化机制十分复杂,气温预测的准确性往往不尽人意^[1]。近年来,随着机器学习和数据挖掘等人工智能技术被广泛应用于气象预测中,在预测中最常用的方法是决策树^[2]、遗传算法^[3]、神经网络^[4]等。Jeong^[5]结合双向长短期记忆网络(bi-directional long short-term memory, Bi-LSTM)与卷积神经网络(convolutional neural networks, CNN),预测未来气温,结果显示模型相结合的方式比单

个模型准确率更优;陈岚等^[6]提出了使用灰色关联分析法(grey relation analysis, GRA)解决参数修改的难题,GRA-Conv-BiLSTM 预测模型经过实验结果表明,对于具有时间序列的气象数据,其预测模型可以预报出更加精确的气温;Aliihsan 等^[7]在提前 1 h 和提前一天的气温预测中使用长短期记忆网络(long short-term memory, LSTM)神经网络进行气温预测,其效果比其余机器学习的方法准确率高;杨孟达^[8]使用粒子群算法(particle swarm optimization, PSO)调节 LSTM 神经网络的相关参数,尽可能寻找到全局最佳参数,提高预测气温的精度;杨迎新等^[9]提出了 PSO-GRU 气温预测模型,其模型相对于传统方法而言,收敛速度更快、预测精度更高。

本研究提出了一种准确率较高的 XGBoost 预测气温

方法,首先对数据进行预处理,然后使用 PCA 主成分分析的方法对气象数据进行降维压缩,最后运用模拟退火算法(simulate anneal arithmetic, SA)与粒子群算法改进的 XGBoost 模型对 6 h 后的气温进行预测。实验结果表明,相对于 XGBoost、GRU 神经网络和 LSTM 神经网络传统预测算法,该方法的预测准确精度更高,效果更好,在气温预测领域可以体现出价值。

1 相关背景

1.1 XGBoost 算法

XGBoost 是基于 Boosting 集成学习框架的一个算法工具包,其在并行计算、缺失值处理与预测能力十分强大^[10],在工程应用中通常用作分类和预测问题。XGBoost 将决策树当作弱分类器,通过不断增加新的分类器来优化之前分类器产生的误差,如式(1),最后将不同串行的分类器的值累加求和得到预测结果^[11-13],由此构建出预测模型,定义为:

$$F_m(x_i) = F_{m-1}(x_i) + f_m(x_i) \quad (1)$$

其中, $F_{m-1}(x_i)$ 为前 $m-1$ 个分类器迭代后构建的近似函数, $f_m(x_i)$ 为第 m 个分类器对误差的预测值, $F_m(x_i)$ 是 m 个分类器迭代的预测值。

XGBoost 目标函数,定义为:

$$Obj = \sum_{i=1}^N L[F_m(x_i), \hat{y}_i] + \sum_{j=1}^m \Omega(f_j) \quad (2)$$

其中,可以看出 Obj 由两部分组成, $\sum_{i=1}^N L[F_m(x_i), \hat{y}_i]$ 是 N 个样本点中的真实值与其预测值产生的误差之和,其中 L 为 XGBoost 损失函数, $\sum_{j=1}^m \Omega(f_j)$ 代表各个分类器复杂度之和,定义为:

$$\Omega(f_j) = \xi T + \frac{1}{2} \lambda \sum_{i=1}^T \omega_j^2 \quad (3)$$

其中, ξ 和 λ 是惩罚系数, ξ 的作用是控制分类器叶子节点的数目,叶子节点的个数越多,越容易出现过拟合, λ 的作用是控制每个分类器的权重,其值不可太大,以免后面的分类器没有学习空间, T 为分类器 f_j 叶子节点的个数, ω_j 为叶子节点 j 的输出权值,此式为单个分类器的复杂度。

泰勒公式简化 XGBoost 目标函数,定义为:

$$Obj = \sum_{i=1}^T [G_j \omega_j + \frac{1}{2} (H_j + \lambda) \omega_j^2] + \xi T \quad (4)$$

其中, G_j 为分类器第 j 个叶子节点一阶梯度的总和, H_j 为分类器第 j 个叶子节点二阶梯度的总和,当每个分类器结构确定的时候, G_j 、 H_j 和 T 也随之确定。

想要 XGBoost 达到最佳的性能,需要确定其分类器结构,如果使用穷举法,将分类器可能的结构以枚举方式确定最佳结构使得目标函数值最小,输入样本数据量如果很多,往往需要很大的计算量,这是不现实的。所以 XGBoost 使

用贪心的算法来进行分类器的生成,增益定义为:

$$G = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \xi \quad (5)$$

其中, G 为信息增益,其中 4 项分别为分类器分裂后左边的值、分类器分裂后右的值、分类器为分裂前的值和惩罚系数。选择信息增益最大的分割点作为分类器的结构,如果信息增益是负数,就停止分裂。

在搭建本实验 XGBoost 预测模型时,将数据进行 PCA 降维后的数据作为模型的特征输入,6 h 后的气温数据作为标签,使用 SA-PSO 混合算法对 XGBoost 模型的参数寻优,得到最小的目标函数值,构建出高准确率的气温预测模型。

1.2 粒子群优化算法

粒子群优化算法的思想是模仿鸟群在一定的区域空间内寻找最优解的过程^[14]。如果待搜索空间区域为 L 维,群落有 N 个粒子,其第 i 个粒子的位置表达方式为 $X_{iL}^k = [x_{i1}^k, x_{i2}^k, \dots, x_{iL}^k]$, 其中 $i = 1, 2, \dots, N$, 第 i 个粒子的飞行速度为 $V_{iL}^k = [v_{i1}^k, v_{i2}^k, \dots, v_{iL}^k]$, 在第 i 个粒子迭代过程中,其历史个体最优解为 $P_{iL}^k = [p_{i1}^k, p_{i2}^k, \dots, p_{iL}^k]$, 整个粒子群落的全局最优解为 $P_{gL}^k = [p_{g1}^k, p_{g2}^k, \dots, p_{gL}^k]$, 然后需要更新粒子速度,它是根据个体最优解和全局最优解相对于粒子当前的位置进行运算,从而实现粒子位置的更新。如公定义为:

$$V_{iL}^{k+1} = \omega V_{iL}^k + c_1 r_1 (P_{iL}^k - X_{iL}^k) + c_2 r_2 (P_{gL}^k - X_{iL}^k) \quad (6)$$

$$X_{iL}^{k+1} = V_{iL}^{k+1} + X_{iL}^k \quad (7)$$

其中, ω 表示飞行惯性权重, k 表示群落第 k 迭代, c_1 、 c_2 分别表示个体速度因子和全局速度因子, r_1 、 r_2 是 $0 \sim 1$ 的随机数字。

1.3 模拟退火算法

模拟退火算法^[15]的思想是模拟温度高的固体逐渐降低温度的过程,在模拟退火过程中,使用的是 Metropolis 准则和退火流程 2 个步骤。

Metropolis 准则使用的是不确定的规则,即通过概率来判断是否接收新的状态,定义为:

$$p = \begin{cases} 1, & E(n+1) < E(n) \\ \exp\left(-\frac{E(n+1) - E(n)}{T}\right), & E(n+1) \geq E(n) \end{cases} \quad (8)$$

其中, T 为当前温度,当前状态为 $x(n)$, 下一个状态为 $x(n+1)$, 相对应能量为 $E(n)$ 、 $E(n+1)$ 。当 $E(n) > E(n+1)$, 能量减小,则接受 $x(n+1)$ 状态;当 $E(n) \leq E(n+1)$, 能量虽然增大,但是不能拒绝这种状态,要根据 Metropolis 准则使用概率的方式判断是否接受 $x(n+1)$ 状态,计算 $\exp\left(-\frac{E(n+1) - E(n)}{T}\right)$, 如果其值大于区间 $[0, 1]$ 产生的随机数,则接受 $x(n+1)$ 状态,反之,拒绝 $x(n+1)$ 状态,继续在 $x(n)$ 状态。其核心思想是不好的

状态以一定的概率接受,才有一定的概率可以跳出局部最优解。

算法流程如下:

1)先设置其初始温度 T_0 ,令 $T = T_0$,生成一个随机初始解 $x(n)$,然后计算其相对应的适应函数 $E(n)$,设置区间 $[0,1]$ 的退火率 λ 。

2)令 $T = \lambda T$,使温度下降。

3)对 $x(n)$ 状态进行随机更新扰动,生成 $x(n+1)$ 状态,并计算其对应的适应函数 $E(n+1)$ 。

4)比较 $E(n)、E(n+1)$,使用 Metropolis 准则判断是否接受 $x(n+1)$ 状态。

5)执行步骤 2)后,重复 L 次扰动,并且接受新状态,即执行步骤 3)和步骤 4)。

6)判断是否达到终止条件,若达到,则终止以上过程,若未达到,则返回步骤 2)。

2 基于自适应 SA-PSO-XGBoost 气温预测模型

2.1 数据预处理

本文实验气象数据选自 ECMWF(欧洲中期天气预报中心)南京地区从 2016 年 1 月 1 日 0 时~2017 年 12 月 31 日 12 时的气象数据,数据间隔为 6 h,其经纬度范围为北纬 $31^{\circ}14' \sim 32^{\circ}34'$,东经 $118^{\circ}22' \sim 119^{\circ}12'$ 。该数据集的气象属性包含经度(longitude)、纬度(latitude)、时间(time),还有气象属性:2 m 露点温度(2 m dewpoint temperature)、2 m 气温(2 m temperature)、风的 10 m u 分量(10 m u-component of wind)、风的 10 m v 分量(10 m v-component of wind)、地面气压(surface pressure)、总降水量(total precipitation)。本文将气象数据 6 h 后的 2 m 气温合并到当前时刻的数据中,例如将 6 时刻的 2 m 气温数据放于 0 时刻数据中,处理后部分气象数据如表 1 所示。

表 1 气象数据

经度	纬度	时间	2 m 露点 温度/K	2 m 气温/ K	风的 10 m u 分量/($m \cdot s^{-1}$)	风的 10 m v 分量/($m \cdot s^{-1}$)	地面气压/ Pa	总降 水量/m	2 m 气温_ 6 h/K
118.22	32.34	2016-1-1 0:00	271.371 2	272.666 4	-1.748 9	1.862 6	102 451.930 8	0	284.266 3
118.32	32.34	2016-1-1 0:00	271.363 3	272.531 6	-1.660 2	1.838 0	102 787.731 6	0	284.427 4
118.42	32.34	2016-1-1 0:00	271.503 0	272.608 1	-1.622 9	1.832 8	102 913.357 3	0	284.503 9
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
119.12	31.14	2017-12-31 12:00	274.052 6	276.350 0	-1.981 1	0.887 7	102 532.484 6	0	274.335 5

在气象数据采集下载的过程中,有些数据会出现缺失值、冗余值、异常值等,在本文中每一列的平均值填充其相对应的位置,或者删除多余的数据,为了消除不同气象要素间量纲差异,对气象数据采用标准化的方法进行处理。

2.2 PCA 数据降维

为了保存数据中大部分信息量,并且降低输入数据的规模与维数,去掉数据中无用的噪声^[16],在本实验中对特征数据进行 PCA 主成分分析处理。其主要的思想是将 K 维度的原始数据,重新构造为 N 维数据,其中 $K > N$,由此实现数据压缩的目的。PCA 降维步骤如下:

1)实现数据中心化,将气象数据的每一列属性数据的平均值得出,后每一列的数据减去其对应列的平均值。

2)算出高维原始数据集对应的协方差矩阵,将 K 维 M 个样本的数据集当作矩阵,计算其对应的协方差矩阵 C ,定义为:

$$C = \begin{pmatrix} \text{cov}(x, x) & \text{cov}(x, y) & \text{cov}(x, z) \\ \text{cov}(y, x) & \text{cov}(y, y) & \text{cov}(y, z) \\ \text{cov}(z, x) & \text{cov}(z, y) & \text{cov}(z, z) \end{pmatrix} \quad (9)$$

$$\text{cov}(x, y) = \frac{\sum_{i=1}^M (x_i - \bar{x})(y_i - \bar{y})}{M - 1} \quad (10)$$

3)计算出对应协方差矩阵的特征值 λ 和特征向量 u 。

4)将 K 维气象数据投影到选取的特征向量上,得到 N 维数据,即降维后的气象数据,定义为:

$$Y = UX \quad (11)$$

其中, Y 代表降维后的气象数据, U 代表特征向量, X 代表原始气象数据。

本文将 6 个维度原始数据气象数据集导入 SPSS 软件中,对其进行数据分析实现标准化,然后使用软件降维方法得出主成分分析结果,如表 2。

表 2 PCA 分析结果

成分	初始特征值方差 百分比/%	累积方差贡献率/ %
1	49.610	49.610
2	18.469	68.079
3	17.067	85.146
4	11.749	96.895
5	1.978	98.872
6	1.128	100

表中显示气象数据成分、初始特征值方差百分比和累积方差贡献率。其中某个特征方差越大,则说明此特征在

整个数据集中越重要。表中显示累计方差贡献率大于 85% 的主成分有 3 个,因此本文将 6 维气象数据降为 3 维数据,减小数据规模,由此作为气温预测模型的输入数据,进行以下训练。

2.3 自适应 SA-PSO-XGBoost 算法模型

在本实验中,PSO 算法 ω 、 c_1 和 c_2 因子随着迭代次数不断变化,使 PSO 算法有更好的效果。其中 ω 为飞行惯性权重,其直接控制粒子飞行速度,在 PSO 算法寻找最优值的过程中占有重要影响,在迭代前期,为扩大搜索范围,其值应设置较大;在迭代后期,其值应设置较小,目的是为了局部寻优,因此其值应是逐渐减小的过程。文献[17]经大量实验证明,由 $\omega_{max} = 0.9$ 动态衰减到 $\omega_{min} = 0.4$,PSO 算法的效果最佳。所以在本实验中选取 Sigmoid 非线性函数构造惯性权重函数,定义为:

$$sigmoid(x) = \frac{1}{1 + e^{-x}} \quad (12)$$

$$\omega = \frac{(sigmoid(-4 + 8 \times (k_{max} - k)k_{max}) \times 2 - 1)(\omega_{max} - \omega_{min})}{2} + (\omega_{max} + \omega_{min})/2 \quad (13)$$

其中, k_{max} 为 PSO 算法最大迭代次数, k 为 PSO 算法当前迭代次数。

c_1 为个体速度因子,决定着粒子个体寻优方向, c_2 为全局速度因子,决定着整体搜索空间的寻优方向。PSO 算法前期应该偏向粒子个体寻优,算法后期应该偏向整体最优方向,所以算法前期应 $c_1 > c_2$,后期则 $c_1 < c_2$ 。在本实验中 c_1 与 k 成反比, c_2 与 k 成正比,文献[18]经大量实验证明, $c_{1max} = c_{2min} = 2.5$, $c_{1min} = c_{2max} = 1.25$ 时 PSO 算法效果最佳,定义为:

$$c_1 = c_{1max} - (c_{1max} - c_{1min}) \frac{k}{k_{max}} \quad (14)$$

$$c_2 = c_{2min} + (c_{2max} - c_{2min}) \frac{k}{k_{max}} \quad (15)$$

自适应 SA-PSO 算法的思路是,在 PSO 算法中,针对 ω 、 c_1 和 c_2 因子各自的特性,构建其关于迭代次数 k 的函数,让其自适应变化,使得 PSO 算法在不同的时期效果达到最优。PSO 算法容易陷入局部最优解,模拟退火算法可以增加粒子行为的多样性,可以增大算法收敛的范围^[19],因为在模拟退火算法过程中使用了 Metropolis 准则,可以让粒子有接收差解的可能,有一定的概率跳出局部最优解^[20]。在本文中使用模拟退火算法和因子自适应函数改进粒子群算法。自适应 SA-PSO-XGBoost 算法流程如图 1 所示。

算法具体步骤如下:

1) 设置 XGBoost 调参数的边界,并且设置种群粒子个数、迭代次数 k 、初始温度。

2) 随机生成种群粒子位置和其各自运动的速度,设置 r_1 、 r_2 随机数和退火率 λ 。

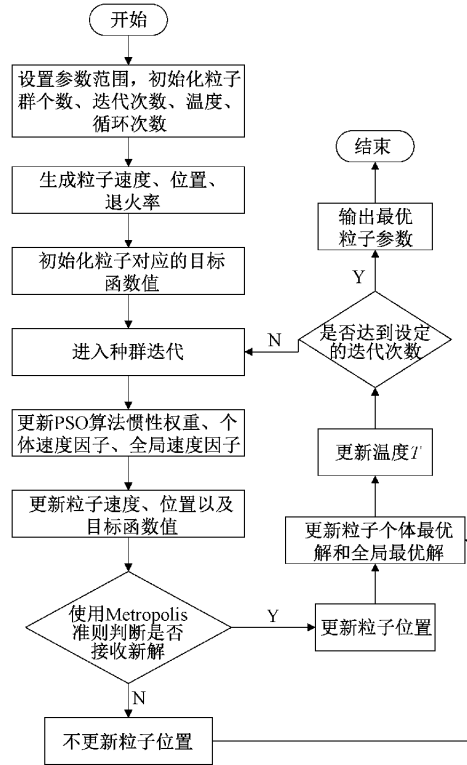


图 1 自适应 SA-PSO-XGBoost 流程

3) 初始化计算各个粒子对应的目标函数值,赋值种群粒子初始位置。

4) 进行种群迭代,根据式(12)~(14)对 ω 、 c_1 和 c_2 更新。

5) 根据式(6)、(7)更新粒子速度和粒子位置,计算目标函数值。

6) 根据式(8) Metropolis 准则计算接收新解的概率 p 。

7) 使用 p 与 $[0, 1]$ 的随机数进行比较,判断是否接受新解。

8) 对 P_{bL} 和 P_{gL} 进行更新。

9) 根据 $T = \lambda T$,更新温度。

10) 判断是否达到设定的迭代次数 k ,如果达到条件则输出最优粒子参数,结束算法;反之,则返回步骤 4)。

3 实验结果与分析

为了验证自适应 SA-PSO-XGBoost 算法模型的准确率和优越性,本文选 ECMWF(欧洲中期天气预报中心)南京地区气象数据,将其分为训练集和测试集,训练集和测试集的比例分为 4 : 1。

3.1 评价指标

本文检测模型的评价指标为平均绝对误差(MAE)、均方根误差(RMSE)和相关系数(R^2),定义为:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (16)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (17)$$

$$R^2 = \left(\frac{\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2 \cdot \sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2}} \right)^2 \quad (18)$$

其中, \hat{y}_i 为预测值, y_i 为真实值。 MAE 、 $RMSE$ 的值反映了预测值与真实值之间误差情况, R^2 反映了预测值与真实值之间的拟合程度。

3.2 SA-PSO-XGBoost 模型结果与分析

为了验证自适应 SA-PSO 算法的正确性与有效性,将 Ackley Function 作为测试函数,其值越小则效果越好,测试函数如式(18)。本小节将 PSO 算法和 SA-PSO 算法进行对比,如图 2。由图 2 可知,自适应 SA-PSO 算法的收敛性优于 PSO 算法,有更好的搜寻最优参数的能力,PSO 算法在第 22 次迭代时,陷入了局部最优解,收敛结果为 7.7045×10^{-7} ,而 SA-PSO 算法收敛于第 11 代,收敛结果为 1.1102×10^{-11} 。

$$F_1(x, y) = -20 \exp\left(-0.2 \sqrt{\frac{x^2 + y^2}{2}}\right) - \exp\left(\frac{\cos 2\pi x + \cos 2\pi y}{2} + 20 + e\right) \quad (19)$$

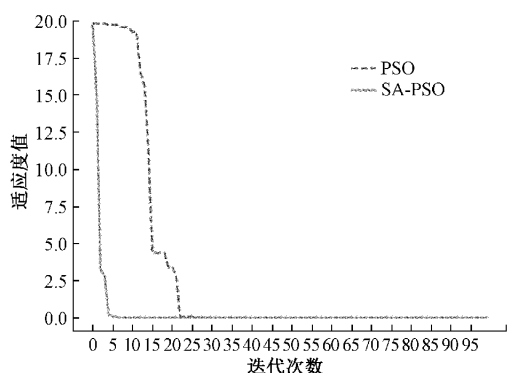


图 2 自适应 SA-PSO 与 PSO 适应值

为了说明 SA-PSO-XGBoost 模型预测能力,本文对比 XGBoost 模型和 SA-PSO-XGBoost 模型的准确度。由于气象数据较多,现选取测试集中 60 个数据点进行对比,其不同模型预测气温对比如图 3 所示,可以看出 SA-PSO-XGBoost 模型与真实值拟合更充分。

对比试验后的两个模型的性能指标,如表 3。

表 3 模型性能对比

预测模型	RMSE	MAE	R-Square
XGBoost	2.447 5	1.806 7	0.933 0
SA-PSO-XGBoost	1.253 2	0.819 3	0.982 4

由表 3 可知,从 $RMSE$ 、 MAE 和 R^2 指标可以看出,与

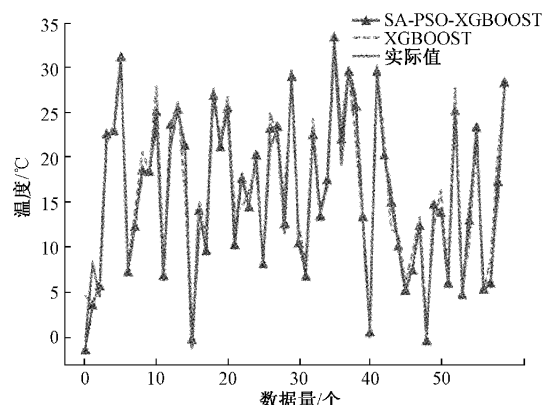


图 3 XGBoost、SA-PSO-XGBoost 模型与真实值对比

前一个模型相比,SA-PSO-XGBoost 模型的 $RMSE$ 下降了 1.194 3, MAE 下降了 0.987 4, R^2 提升了 0.049 4。由此得出 SA-PSO-XGBoost 模型性能更好,适合预测 6 h 气温。

为了证明此模型的优越性,现选取传统气温预测模型 LSTM 神经网络与 GRU 神经网络作为对比模型,选取数据集集中 60 个数据点进行对比,其测试预测气温结果如图 4。

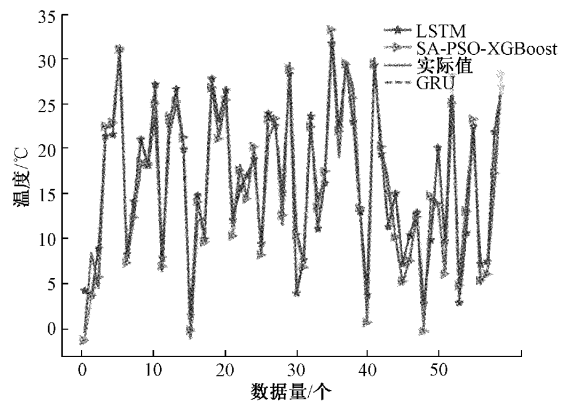


图 4 LSTM、GRU、SA-PSO-XGBoost 模型与真实值对比

由图 4 可知,SA-PSO-XGBoost 算法模型更与实际温度相近,拟合度效果更好,LSTM 模型与 GRU 模型在本次对比实验中发现,与实际值相差很大。SA-PSO-XGBoost 算法模型与 LSTM 模型、GRU 模型性能指标进行对比,如表 4。

表 4 模型性能对比

预测模型	RMSE	MAE	R-Square
LSTM	2.584 4	1.935 7	0.925 3
GRU	3.346 9	2.671 6	0.874 8
SA-PSO-XGBoost	1.253 2	0.819 3	0.982 4

由表 4 可知,从 $RMSE$ 、 MAE 和 R^2 指标可以看出,与前两个模型相比,SA-PSO-XGBoost 模型的 $RMSE$ 分别下

降了 1.331 2、2.093 7, MAE 分别下降了 1.116 4、1.852 3, R^2 分别提升了 0.057 1、0.107 6。由此得出 SA-PSO-XGBoost 模型性能更好,有更理想的预测效果。

4 结 论

气温与人们的日常生活、生产生活和科学研究等领域休戚相关,对人们的生活规划具有指导意义。本文提出一种基于模拟退火与粒子群算法的 XGBoost 预报模型。使用了 ECWMF 南京范围气象数据作为实验数据,对数据进行异常值处理,使用 PCA 主成分分析降低数据噪声,减小气象训练数据量,加快运算速度。粒子群算法会困在局部最优解的情况,因此使用模拟退火和飞行权重系数自适应函数对粒子群算法进行改进,本文以 XGBoost 为主要预测模型,使用模拟退火与粒子群算法对 XGBoost 进行优化改进,最终实验结果显示,SA-PSO-XGBoost 预测模型正确率更高,稳定性更强,有一定的使用价值。

参考文献

- [1] 魏爽,王丽吉,吴书成,等. 基于混合优化 RBF 神经网络的月平均气温预测[J]. 现代计算机,2019(22):3-7.
- [2] 刘亚秋,李海涛,景维鹏. 基于 Hadoop 的海量嘈杂数据决策树算法的实现[J]. 计算机应用,2015,35(4):1143-1147.
- [3] 朱庆生,程柯. 一种基于累积适应度遗传算法的 SVM 多分类决策树[J]. 计算机应用研究,2016,33(1):64-67,74.
- [4] 周玉,朱安福,周林,等. 一种神经网络分类器样本数据选择方法[J]. 华中科技大学学报(自然科学版),2012,40(6):39-43.
- [5] JEONG S. Temperature prediction based on bidirectional long short-term memory and convolutional neural network combining observed and numerical forecast data [J]. Sensors, 2021, 21(3): 941-941.
- [6] 陈岚,文斌,贺南,等. 基于融合模型动态权值的气温预测[J]. 电子测量技术,2022,45(15):7.
- [7] ALIHSAN S, MEHMET B, NIYAZI A, et al. Short-term air temperature prediction by adaptive neuro-fuzzy inference system (ANFIS) and long short-term memory (LSTM) network [J]. Meteorology and Atmospheric Physics, 133(3): 943-959.
- [8] 杨孟达. 基于改进 PSO-LSTM 神经网络的气温预测[J]. 现代信息科技,2020,4(4):110-112.
- [9] 杨迎新,杜景林,武艳. 基于 PCA 和改进 PSO-GRU 神经网络的气温预测[J]. 现代电子技术,2022,45(1):89-94.
- [10] CHEN T Q, CARLOS G. XGBoost: A scalable tree boosting system [C]. Knowledge Discovery and Data Mining. ACM, 2016.
- [11] PRABHA ANJUE ET AL. Design of intelligent diabetes mellitus detection system using hybrid feature selection based XGBoost classifier [J]. Computers in Biology and Medicine, 2021, 136: 104664-104664.
- [12] 宋建,陈广森,陈敬福,等. 基于特征选择和贝叶斯优化 LightGBM 的注塑制品尺寸预测 [J]. 工程塑料应用, 2021, 49(8): 54-60.
- [13] 王坤章,蒋书波,张豪,等. 基于 XGBoost 的回归-分类-回归寿命预测模型 [J/OL]. 中国测试: 1-8 [2022-10-20]. <http://kns.cnki.net/kcms/detail/51.1714.TB.20220616.1537.002.html>.
- [14] 王继寅,薛旭. 利用粒子群算法实现陀螺仪耦合系统动态响应优化 [J]. 电子器件,2021,44(4):980-986.
- [15] 毛志强,马翠红,崔金龙,等. 基于时空双流卷积与 LSTM 的人体动作识别 [J]. 软件,2018,39(9):9-12.
- [16] JI S, HAN X, HOU Y, et al. Remaining useful life prediction of airplane engine based on PCA-BLSTM [J]. Sensors, 2020, 20(16):4537.
- [17] 闫群民,马瑞卿,马永翔,等. 一种自适应模拟退火粒子群优化算法 [J]. 西安电子科技大学学报,2021,48(4):120-127, DOI:10.19665/j.issn1001-2400.2021.04.016.
- [18] TAMO A, BARG A. Bounds on the parameters of locally recoverable codes [J]. IEEE Transactions on Information Theory, 2015, 62(6):3070-3083.
- [19] 孙卫红,童晓,李强. 改进 PSO 优化参数的 LSSVM 燃煤锅炉 NO_x 排放预测 [J]. 数据采集与处理,2015,30(1):231-238.
- [20] 李勇,王建君,曹丽华. 火电厂负荷优化分配的模拟退火粒子群算法 [J]. 电力系统及其自动化学报,2011,23(3):40-44.

作者简介

王强, 硕士研究生, 主要研究方向为气象信息系统, 数据处理。

E-mail: wq1135954658@163.com

秦华旺(通信作者), 博士, 教授, 博导, 主要研究方向为基于 ARM、DSP、FPGA、单片机的嵌入式系统软硬件设计; Windows 平台和 Linux 平台信息系统开发; 复杂信息系统测控; 密码学、量子密码学、信息安全。

E-mail: qin_h_w@163.com

齐春帅, 硕士研究生, 主要研究方向为信息系统。

E-mail: 969355300@qq.com

王琴琴, 本科, 主要研究方向为降水同位素。

E-mail: wqq885590413@163.com