

doi:10.3969/j.issn.1000-6362.2022.07.002

蔡淑芳,林营志,吴宝意,等.利用线性和非线性耦合方式建立温室温湿度预测模型[J].中国农业气象,2022,43(7):527-537

## 利用线性和非线性耦合方式建立温室温湿度预测模型<sup>\*</sup>

蔡淑芳, 林营志, 吴宝意, 郑东海, 雷锦桂<sup>\*\*</sup>

(福建省农业科学院数字农业研究所, 福州 350003)

**摘要:** 基于蔬菜种植试验温室内温度、相对湿度和光照强度的实测数据, 根据 ARIMA 模型和 RBF 神经网络对线性和非线性问题的预测能力差异, 构建 ARIMA-RBF 神经网络权重组合的温湿度预测模型, 对温室内温度和湿度的动态变化进行预测, 并比较各模型预测精度。结果表明: 温室内温湿度分别具有更明显的线性和非线性变化特征, 对应预测性能较好的单一模型分别为 ARIMA 模型和 RBF 模型。相较单一模型, ARIMA-RBF 神经网络权重组合模型的预测精度更高、稳定性更好。最佳温度组合模型的 MAE、MAPE 和 RMSE 分别为 1.04℃、2.95%和 1.21℃; 最佳湿度组合模型的 MAE、MAPE 和 RMSE 分别为 0.35 个百分点、0.36%和 0.55 个百分点。权重组合模型通过适当的加权策略充分发挥了单一模型对数据不同特征的处理能力, 能较好地评估温室内温湿度状态, 可为建立更具普适性的温室环境因子模型提供参考。

**关键词:** 温室; 温度; 湿度; 模型; ARIMA; RBF

## Greenhouse Temperature and Humidity Prediction Models Based on Linear and Nonlinear Coupling Methods

CAI Shu-fang, LIN Ying-zhi, WU Bao-yi, ZHENG Dong-hai, LEI Jin-gui

(Institute of Digital Agriculture, Fujian Academy of Agricultural Sciences, Fuzhou 350003, China)

**Abstract:** In order to provide reference for vegetable growth management and environmental optimization regulation, the dynamic changes of greenhouse temperature and humidity were predicted by using ARIMA model and RBF neural network. In this study, the data from July 28 to 29 were taken as the validation set, and the data from the previous 5 to 25 days were taken as the sample set. Three input variables were set to explore the temperature and humidity model prediction effects. Based on the measured data of air temperature, relative humidity and light intensity in the vegetable greenhouse, and according to the difference in the predictive ability of ARIMA model and RBF neural network for linear and nonlinear problems, temperature and humidity prediction models based on weight combination of ARIMA and RBF neural network were constructed. The results showed that the temperature and humidity in the greenhouse had more obvious linear and nonlinear variation characteristics, respectively, and the single models with better prediction effects were the ARIMA model and the RBF model. The optimal single temperature model was the model ARIMA(0,1,1)(1,1,1)<sub>24</sub> with 25 days sample set. The optimal single humidity model was the RBF model (3-6-2) with 25 days sample set and input variable A (temperature, humidity and light intensity). Compared with the optimal single models, the ARIMA-RBF neural network weight combination model had higher prediction accuracy and better stability. The MAE, MAPE and RMSE were 1.04℃, 2.95%, 1.21℃ for

<sup>\*</sup> 收稿日期: 2021-09-28

基金项目: 福建省自然科学基金项目(2020J011377); 福建省科协项目(闽科协学(2021)9号); 福建省农业科学院项目(CXTD2021013-1)

<sup>\*\*</sup> 通讯作者: 雷锦桂, 研究员, 研究方向为数字农业, E-mail: 71906244@qq.com

第一作者联系方式: 蔡淑芳, E-mail: csf2019@qq.com

the optimal temperature combination model and 0.35 percentage point, 0.36%, 0.55 percentage point for the optimal humidity combination model. The weight combination model gives full play to the ability of single models to process data from different characteristics through an appropriate weighting strategy, and can better evaluate the temperature and humidity status in the greenhouse, which provides a reference for establishing a more universal greenhouse environmental factor model.

**Key words:** Greenhouse; Temperature; Humidity; Model; ARIMA; RBF

温室是人为建造的封闭设施,通过对温室局部范围内环境和气象因素的调控,把最适宜的温度、湿度、光照、水分等环境条件提供给作物生长,使作物达到最佳生长状态,以获得高产和优质产品<sup>[1]</sup>。日光温室的相对封闭性决定了其内部环境参数不同于外界气候。目前,蔬菜温室环境调控方式大多依靠农业生产者的经验,通过对当前采集的监测点位环境数据进行分析,进而做出调控决策,其数据采集的反馈与调控措施的生效都存在滞后性和片面性<sup>[2]</sup>。

常见的环境因子预测模型主要有 2 类,即机理模型和经验模型。机理模型根据能量守恒和质量守恒定律对温室环境进行探究,模型中每个参数代表某个实际的物理量,具有明确的物理意义,方便人们的理解,但存在建模复杂、未知参数众多、测量代价高和模型自适应性较差等问题。经验模型是以现代计算理论为基础,直接根据输入输出数据所提供的信息来建立过程模型,无需考虑温室耗散、热辐射等因素的影响。一些学者将易于测量的温室外界环境因子同温室内部环境因子建立关联映射进行求解预测,回归预测法、时序分析法、支持向量机、神经网络等模型均在该领域有应用研究<sup>[3-4]</sup>。但对于一些复杂的预测系统,采用单一的预测方法,往往难以获得满意的结果。

近年来,为提高预测精度,一些研究者在经验模型的基础上对预测方法进行优化,包括对不同经验模型的组合预测<sup>[5-6]</sup>、经验模型的参数优化<sup>[7-8]</sup>,经验模型的分类预测<sup>[9]</sup>等。其中,经验模型的组合预测方法是,利用不同预测函数或算法对数据特征的提取能力差异,通过从单一模型中提取有效的预测信息,使不同预测模型并行工作,达到提高预测精度和增强预测稳定性的目的。组合预测方法在风速、风电功率、水文领域研究上表现出较高的准确性和适用性,在温湿度领域研究方面相对较少。田东等<sup>[5]</sup>将组合方法引入食用菌温室温度预测并取得了较好

的预测效果,其具体应用的是经验模型的残差优化组合方法,而经验模型的权重优化组合方法<sup>[10]</sup>在温室温湿度预测的研究中未见报道。

本研究以生产型果蔬温室的历史环境数据为基础,将温室温湿度视为是蕴含线性和非线性分量的动态叠加的时间序列,利用 ARIMA 模型和 RBF 神经网络分别对线性问题和非线性问题的预测能力,提出 ARIMA-RBF 神经网络权重组合预测方法建立温湿度预测模型。通过设置不同样本集和不同维度输入变量,系统性探讨样本选择、输入参数对模型温度和湿度预测性能的影响,探究权重组合模型对预测性能的优化能力,以期对温室环境调控和果蔬高效生产提供科学依据。

## 1 资料与方法

### 1.1 数据来源及预处理

试验数据采自福建省农业科学院中以示范农场薄膜温室。温室(26°7'43"N, 119°20'0"E)长 115.2m,宽 68m,肩高 4m,顶高 6.5m,钢架结构,钢架上方覆盖聚乙烯棚膜,温室配备环流风机、喷雾、内遮阴、顶开窗、水肥一体化等设备。一年两茬蔬菜基质栽培,夏季和冬季分别种植黄瓜、番茄。试验数据采集装置使用威海某公司所提供的温度、湿度、光照度传感器自动采集记录,温度、相对湿度和光照强度的观测精度分别为 $\pm 0.5^{\circ}\text{C}$ 、 $\pm 3$  个百分点、 $\pm 0.126\%\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ 。传感器位于栽培区域离地面高约 1.5m 处,采集频率为每 30s 采集一次。

考虑到月度温湿度变化具有较高的相似性<sup>[9]</sup>,基于相似数据集的模型训练有利于提高预测性能;同时,试验温室位于福州,当地一年中以高温高湿气候为主,尤其是在 7 月等夏季高温时段对温室通风降温功能的需求较高。因此,以 2020 年 7 月数据为例进行模型研究,不同月份可根据本研究提出的方法进行分别建模。鉴于样本量对模型预测精度会产生影响<sup>[3]</sup>,将 7 月 28-29 日的数据作为验证集,前 5d (7 月 23-27 日)、10d (7 月 18-27 日)、15d (7

月 13–27 日)、20d (7 月 8–27 日) 和 25d (7 月 3–27 日) 的数据作为样本集, 探讨不同样本集的温湿度模型预测效果。因输入变量的高自相关可能出现过拟合而影响模型预测精度<sup>[11]</sup>, 对影响温湿度的温度、相对湿度和光照强度的数据进行筛选, 探讨不同输入变量对模型预测效果的影响。

采用拉依达法则<sup>[10]</sup>即当测量值与平均值差的绝对值大于 3 个标准差时, 认为该测量值为异常值, 将该值替换为异常点两侧数据的平均值。由于 1h 内温室环境数据变化不大, 因此将采集频率为每 30s 一次的温室环境数据按小时平均进行平滑。经测算, 7 月温室内温度变化范围为 30.16~45.95℃, 相对湿度变化范围为 86.17%~99.80%, 光照强度变化范围为 0~744.07 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ 。为避免某一维数据值过大而导致训练错误, 提高预测的准确性, 对温室环境数据进行归一化处理, 在预测分析结束后再通过反归一化恢复。

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

式中,  $x$  为温度、相对湿度和光照强度,  $x_{\max}$  为最大值,  $x_{\min}$  为最小值,  $x^*$  为归一化值。

## 1.2 建模方法

### 1.2.1 ARIMA 模型

自回归移动平均模型(Autoregressive Integrated Moving Average Model, ARIMA)是将自回归过程(AR)和移动平均过程(MA)相结合的时间序列组合模型<sup>[12]</sup>。其基本思想是, 利用时间序列取得的变量自身历史数据, 通过曲线拟合和参数估计来建立数学模型, 对变量未来发展状况进行推断和估计。温室内温度和相对湿度每 24h 呈现周期性变化特征, 可用季节 ARIMA 模型进行分析。

设温度和相对湿度的时间序列  $\{Y_t\}$  的周期为  $S$ , 延迟算子记为  $B$ , 即  $BY_t = Y_{t-1}$ ; 季节差分算子  $\nabla_s = 1 - B^s$ , 则有

$$\nabla_s Y_t = (1 - B^s) Y_t = Y_t - Y_{t-s} \quad (2)$$

对于  $D$  阶季节差分, 则有

$$\nabla_s^D Y_t = (1 - B^s)^D Y_t = \nabla_s^{D-1} Y_t - \nabla_s^{D-1} Y_{t-s} \quad (3)$$

如果存在非负整数  $d$  和  $D$  使得  $\{Y_t\}$  满足差分方程式

$$\varnothing(B)\varphi(B)\nabla^d\nabla_s^D Y_t = \theta(B)\vartheta(B)e_t \quad (4)$$

式中,  $\varnothing(B)$  为非季节性自回归模型参数,  $\varphi(B)$  为季节性自回归模型参数,  $\nabla$  为非季节性差分算子,  $\nabla = 1 - B$ ;  $d$  表示非季节性差分的阶数,  $\theta(B)$  为非季节性移动平均模型参数,  $\vartheta(B)$  为季节性移动平均模型参数,  $e_t$  为白噪声序列, 称  $\{Y_t\}$  为季节周期  $S$ , 非季节阶数为  $p$ 、 $d$ 、 $q$ , 季节阶数为  $P$ 、 $D$ 、 $Q$  的季节 ARIMA 模型, 记为  $\text{ARIMA}(p,d,q)(P,D,Q)_S$ 。

其中

$$\varnothing(B) = 1 - \varnothing_1 B - \varnothing_2 B^2 - \dots - \varnothing_p B^p \quad (5)$$

$$\varphi(B) = 1 - \varphi_1 B^s - \varphi_2 B^{2s} - \dots - \varphi_P B^{Ps} \quad (6)$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (7)$$

$$\vartheta(B) = 1 - \vartheta_1 B^s - \vartheta_2 B^{2s} - \dots - \vartheta_Q B^{Qs} \quad (8)$$

式中,  $p$  表示非季节性自回归过程滞后的阶数,  $q$  表示非季节性移动平均过程滞后的阶数,  $P$  表示季节性自回归过程滞后的阶数,  $Q$  表示季节性移动平均过程滞后的阶数。

ARIMA 模型比较简单, 对资料的要求单一, 只需变量本身的历史数据, 且对数据线性关系特征的提取能力较强, 在实际中有着广泛的适用性。其建模的主要思路为<sup>[12]</sup>, 通过观察温度和相对湿度序列的自相关函数(ACF)、偏自相关函数图(PACF)和散点图、方差、ADF 单位根检验等方式, 判断时间序列的平稳性。对非平稳的时间序列进行差分处理、对数转换等使其平稳化。通过判断时间序列 ACF 和 PACF 的拖尾性和截尾性, 确定模型的阶数。以 BIC 贝叶斯信息准则对阶数的不同组合进行筛选, 得到最优的 ARIMA 模型。通过模型参数和残差白噪声检验, 判断模型的合理性。应用所建立的模型预测下一时刻温湿度值。

### 1.2.2 RBF 神经网络

RBF 神经网络属于 3 层前向网络类型, 具有结构简单、计算收敛速度快、结果稳定且逼近精度高的特点。其基本思想是用 RBF 作为隐单元的“基”构成隐藏层空间, 隐藏层对输入矢量进行变换, 将低维的模式输入数据变换到高维空间内, 使得在低维空间内的线性不可分问题在高维空间内线性可分<sup>[13]</sup>。温室内温度和相对湿度除了受自身历史值的影响, 还受到光照强度等环境因素的影响, 即温湿度时间序列是由线性部分和非线性部分构成的时间序列, 其非线性部分可应用 RBF 神经网络进行建模预测。

RBF 神经网络表达式为

$$Y = W^T R(x^*) \quad (9)$$

$$R_g = \exp\left(-\frac{x^* - c_g^2}{2\sigma_g^2}\right) \quad (10)$$

式中,  $Y$  为神经网络输出;  $W$  为权值向量;  $g$  为隐含层节点;  $R_g$  为高斯函数;  $c_g$  为神经网络中心向量;  $\sigma_g$  为高斯基函数的宽度,  $x^*$  为神经网络输入向量。

$t$  时刻的隐含层节点线性组合输出为

$$Y_{gt} = w_1 R_1 + w_2 R_2 + \dots + w_g R_g \quad (11)$$

网络的学习指标函数为

$$\varepsilon_t = \frac{(Y_t - Y_{gt})^2}{2} \quad (12)$$

式中,  $Y_t$  为  $t$  时刻的实际值。

RBF 神经网络具有良好的非线性映射能力, 在不使用先验知识的情况下, 能快速对模型进行无线逼近, 是应用较为广泛的环境因子预测模型。其建模的主要思路为<sup>[13]</sup>, 根据预测温度和相对湿度的建模目的和输入变量选择, 确定输入层、输出层节点数; 以误差平方和最小为约束目标, 确定隐含层节点数; 使用聚类特征树和距离聚类相结合的两步聚类法确定径向基函数的中心和宽度; 采用普通最小二乘法求取隐含层至输出层的权值向量; 利用确定的 RBF 模型预测下一时刻温湿度值。

### 1.2.3 ARIMA-RBF 权重组合模型

因温度和相对湿度是蕴含线性分量和非线性分量的时间序列, ARIMA 和 RBF 神经网络分别具有较高的线性和非线性特征数据的提取能力, 因此有必要将二者结合。ARIMA-RBF 权重组合模型是从单一模型中提取有效的预测信息, 通过适当的加权, 集成不同模型预测性能, 达到提高预测精度和稳定性, 提高预测时间序列发展趋势能力的目的。

利用 ARIMA 和 RBF 神经网络对温度和湿度数据序列的预测分析结果, 采用最优加权法求出组合模型的权重, 得到温度、湿度最终预测值  $Y_{3t}$ , 表示为

$$Y_{3t} = k_1 Y_{1t} + k_2 Y_{2t} \quad (13)$$

式中,  $Y_{1t}$ 、 $Y_{2t}$  为基于 ARIMA 和 RBF 神经网络的温度、湿度预测值。  $k_1$ 、 $k_2$  为 ARIMA 和 RBF 模型的权重, 其值可通过最优加权法即“组合预测结果误差平方和最小”原则来求取<sup>[10]</sup>。该方法为, 假设温度和湿度数据序列  $Y_t$  有  $m$  组数据,  $E_{1t}$  和  $E_{2t}$  分别

为基于 ARIMA 和 RBF 神经网络的温度和湿度预测的误差值。

$$E_{jt} = Y_t - Y_{jt} \quad (t=1, 2, \dots, m; j=1, 2) \quad (14)$$

则组合预测结果误差  $E_t$  为

$$E_t = k_1 E_{1t} + k_2 E_{2t} \quad (k_1 + k_2 = 1, k_1 \geq 0, k_2 \geq 0) \quad (15)$$

通过求取最小  $E_t^2$  获得组合模型权重, 代入式 (13) 获得下一时刻温度、湿度最终预测值。

### 1.3 模型评价

采用平均绝对误差 MAE、平均绝对百分比误差 MAPE、均方根误差 RMSE 对模型进行评价。

$$MAE = \frac{1}{m} \sum_{t=1}^m |Y_t - Y_{jt}| \quad (16)$$

$$MAPE = \frac{100}{m} \sum_{t=1}^m \left| \frac{Y_t - Y_{jt}}{Y_t} \right| \quad (17)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{t=1}^m (Y_t - Y_{jt})^2} \quad (j=1, 2, 3) \quad (18)$$

### 1.4 数据处理

采用 Origin、SPSS 等软件对模型进行作图和处理。

## 2 结果与分析

### 2.1 利用 ARIMA 方式建立预测模型

#### 2.1.1 模型识别与诊断

应用季节模型 ARIMA(p,d,q)(P,D,Q)<sub>s</sub> 对室内温度和湿度序列数据进行分析。以验证集的前 5d、10d、15d、20d 和 25d 数据作为样本集。经过序列平稳性检验以及 ACF 和 PACF 对差分序列定阶, 根据 BIC 值, 分别确定不同样本集的最优温度和湿度 ARIMA 模型, 结果见表 1。对所有模型参数进行 T 检验, 不同样本集所建模型各参数均通过了 0.05 水平的显著性检验, 说明模型参数有效; 残差白噪声检验的统计量均未通过 0.05 水平的显著性检验, 表明残差符合随机分布, 且没有离群值出现, 反映所建模型合理。

#### 2.1.2 模型预测结果及分析

应用所建立的不同样本集的温湿度 ARIMA 模型对 2020 年 7 月 28–29 日共 48h 的室内温湿度进行预测, 结果见图 1。由图 1a 可见, 5d 样本集模型温度预测结果与实际值具有明显的偏差; 10d 样本集模型对 7 月 29 日的温度预测结果与实际值偏差较大。相对地, 15d、20d 和 25d 样本集模型温度预测

表 1 不同样本集的温湿度 ARIMA 模型及 48h 预测误差统计结果

Table 1 Temperature and humidity ARIMA models with different sample sets and statistical results of 48 hours prediction errors								
建模对象 Model object	样本集 Sample set	模型形式 Model form	模型统计 Model statistics		模型参数显著 性 Model parameters Sig.	48h 预测误差统计结果 Statistical results of 48hours prediction errors		
			Ljung-Box Q(18) Sig.	离群值 Nr. of outliers		平均绝对误差 MAE	平均绝对百分比误差	均方根误差
						MAPE(%)	比误差	RMSE
温度 Temperature (°C)	25d	ARIMA(0,1,1)(1,1,1) <sub>24</sub>	0.132	0	<0.001	1.08	2.94	1.35
	20d	ARIMA(0,1,1)(1,1,1) <sub>24</sub>	0.287	0	<0.002	1.17	3.09	1.52
	15d	ARIMA(0,1,1)(0,1,1) <sub>24</sub>	0.746	0	<0.004	1.18	3.08	1.61
	10d	ARIMA(1,1,0)(0,1,0) <sub>24</sub>	0.312	0	<0.001	2.38	6.66	2.88
	5d	ARIMA(1,1,0)(0,0,1) <sub>24</sub>	0.584	0	<0.001	3.64	9.45	4.68
湿度 humidity (%)	25d	ARIMA(0,0,3)(1,0,1) <sub>24</sub>	0.333	0	<0.001	0.69	0.71	1.27
	20d	ARIMA(2,0,0)(1,0,1) <sub>24</sub>	0.993	0	<0.001	0.71	0.74	1.27
	15d	ARIMA(2,0,0)(1,0,1) <sub>24</sub>	1.000	0	<0.001	0.39	0.41	0.91
	10d	ARIMA(1,0,0)(1,0,1) <sub>24</sub>	1.000	0	<0.001	0.63	0.65	1.26
	5d	ARIMA(0,0,2)(1,0,0) <sub>24</sub>	0.991	0	<0.041	0.82	0.85	1.42

Note: MAE is mean absolute error, MAPE is mean absolute percentage error, RMSE is root mean square error. The same as below.

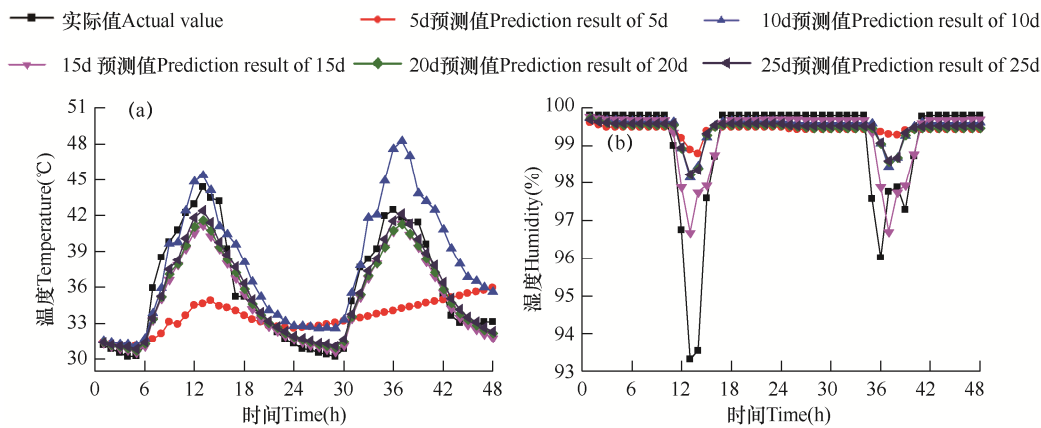


图 1 不同样本集的温度(a)和湿度(b)ARIMA 模型 48h 预测结果

Fig. 1 48 hours prediction results of temperature (a) and humidity (b) ARIMA models with different sample sets

结果与实际值较为接近，其中，25d 预测结果与实际值最为接近。由图 1b 可见，几组预测模型在湿度最低时段均有较为明显的偏差，尤其是在 7 月 28 日的湿度最低时段，几组预测模型的预测结果偏差均较大。从总体上看，5d 样本集模型湿度预测结果与实际值的偏差最为明显；10d、20d 和 25d 样本集模型的预测结果与实际值的偏差次之；相对地，15d 样本集模型湿度预测结果与实际值更为接近。

2.1.3 模型预测性能评价

由表 1 可见，使用不同样本集的温湿度 ARIMA 模型，其预测效果不尽相同。温度预测结果的 3 项误差指标 MAE、MAPE 和 RMSE 分别为 1.08~3.64℃、2.94%~9.45%和 1.35~4.68℃；湿度预测结果的 3 项误差分别为 0.39~0.82 个百分点、0.41%~0.85%、

0.91~1.42 个百分点。随着样本集天数从 25d 到 5d 的逐步减少，ARIMA 模型温度预测结果的 3 项误差逐渐变大，预测精度和稳定性逐渐变差。湿度预测结果 3 项误差从低到高排序的样本集为 15d、10d、25d、20d 和 5d，表明太少天数样本集的预测效果最差，中等天数样本集的预测效果最佳。综上所述，25d 样本集温度预测和 15d 样本集湿度预测模型精度最高。

2.2 利用 RBF 神经网络方式建立预测模型

2.2.1 模型求解

为分析周围环境对温湿度的影响，且避免输入变量的高自相关可能出现的过拟合问题，设置 3 种输入变量以进行预测效果比较。输入变量 A 为前 a 个时刻温度、湿度和光照强度；输入变量 B 分为 B<sub>1</sub>

和  $B_2$  两种,  $B_1$  为前  $b$  个时刻湿度和光照强度 (温度建模时),  $B_2$  为前  $b$  个时刻温度和光照强度 (湿度建模时); 输入变量  $C$  为前  $c$  个时刻光照强度。为适当减少冗余参数和简化模型, 初始赋值  $a=b=c=6$ , 通过经验调优, 当  $a=b=c=1$  时, RBF 模型预测效果最佳。因此, 最终将对应的前 1 个时刻历史值确定为输入变量  $A$ 、 $B$ 、 $C$ 。模型输出为  $t$  时刻温度、湿度预测

值。因输入变量  $B$  不同, 据此分别训练温度、湿度模型; 其余变量则合并训练模型。模型训练时, 将样本集前 60% 的数据作为训练样本, 后 40% 的数据作为测试样本。以误差平方和最小为约束目标, 确立隐含层节点数; 经两步聚类法确定径向基函数的中心和宽度; 采用普通最小二乘法进行计算, 建立相应的温湿度 RBF 预测模型, 结果见表 2。

表 2 不同输入变量和样本集的温湿度 RBF 模型及 48h 预测误差统计结果

Table 2 Temperature and humidity RBF models with different input variables and sample sets and statistical results of 48 hours prediction errors

建模对象 Model object	输入变量 Input variable	样本集 Sample set	模型形式 Model form	48h 预测误差统计结果 Statistical results of 48 hours prediction errors		
				平均绝对误差 MAE	平均绝对百分比误差 MAPE(%)	均方根误差 RMSE
温度 Temperature (°C)	A	25d	3-6-2	1.98	5.92	2.68
		20d	3-9-2	2.16	6.44	2.83
		15d	3-10-2	2.44	7.17	3.1
		10d	3-5-2	2.84	8.25	3.44
		5d	3-10-2	2.53	6.97	3.36
	$B_1$	25d	2-3-1	1.80	5.14	2.22
		20d	2-10-1	1.41	4.11	1.75
		15d	2-8-1	1.63	4.71	1.96
		10d	2-3-1	2.06	5.84	2.53
		5d	2-5-1	1.74	4.96	2.17
	C	25d	1-7-2	1.74	4.99	2.14
		20d	1-5-2	1.60	4.56	2.02
		15d	1-7-2	1.67	4.79	2.05
		10d	1-8-2	1.70	4.92	2.05
		5d	1-9-2	1.80	5.07	2.21
湿度 Humidity (%)	A	25d	3-6-2	0.49	0.50	0.72
		20d	3-9-2	0.55	0.56	0.8
		15d	3-10-2	0.67	0.68	1.16
		10d	3-5-2	0.56	0.58	1.19
		5d	3-10-2	0.67	0.69	1.33
	$B_2$	25d	2-1-1	1.00	1.02	1.24
		20d	2-3-1	1.10	1.12	1.71
		15d	2-4-1	0.76	0.78	1.57
		10d	2-10-1	0.63	0.65	1.26
		5d	2-7-1	0.70	0.72	1.44
	C	25d	1-7-2	0.82	0.84	1.60
		20d	1-5-2	0.97	0.98	1.81
		15d	1-7-2	0.72	0.74	1.33
		10d	1-8-2	0.50	0.52	1.06
		5d	1-9-2	0.59	0.61	1.40

Note: A is temperature, humidity and light intensity,  $B_1$  is humidity and light intensity,  $B_2$  is temperature and light intensity, C is light intensity. The same as below.

### 2.2.2 模型预测结果及分析

应用所建立的不同输入变量和样本集的 RBF 模型对 2020 年 7 月 28–29 日的温湿度进行预测。以输入变量  $B_1$  的 RBF 模型温度预测和输入变量 A 的 RBF 湿度预测为例, 预测结果见图 2。由图 2a 可见, 输入变量  $B_1$  的不同样本集的 RBF 模型温度预测结果与实际值均有不同程度的偏差, 尤其是在 2 日温度最低时段的偏差明显, 各模型均未能较好地捕捉低温变化特征。由图 2b 可见, 输入变量 A 的 RBF 模型湿度预测结果与实际值均有一定偏差。其中, 在对 7 月 28 日的湿度最低时段预测上, 几组模型之间预测结果差别明显; 25d 样本集 RBF 模型湿度预测结果与实际值最为接近, 20d 样本集模型次之, 其余样本集模型预测结果与实际值偏差较大。

### 2.2.3 模型预测性能评价

由表 2 可见, 使用不同输入变量和样本集的温湿度 RBF 模型, 其预测性能表现不一。温度预测结果的 3 项误差 MAE、MAPE 和 RMSE 分别为 1.41~2.84℃、4.11%~8.25%和 1.75~3.44℃; 湿度预测结果的 3 项误差范围为 0.49~1.10 个百分点、0.50%~1.12%和 0.72~1.81 个百分点。当输入变量为 A、B、C 时, 温度预测结果与实际值 3 项误差最小的为 25d、20d 和 20d 样本集模型; 湿度预测结果为 25d、10d 和 10d 样本集模型。当样本集一致时, 温度预测结果的 3 项误差大小为输入变量  $B_1$  模型<A 模型, 输入变量 C 模型<A 模型; 湿度预测结果的 3 项误差大小为输入变量 C 模型< $B_2$  模型, 输入变量 A 模型< $B_2$  模型。总体来看, 输入变量  $B_1$  的 20d 样本集模型的温度预测以及输入变量 A 的 25d 样本集模型的

湿度预测 3 项误差最小, 预测精度和稳定性最好。

### 2.3 利用 ARIMA 与 RBF 组合方式建立预测模型

#### 2.3.1 构建组合预测模型

为最大限度提高 ARIMA-RBF 权重组合模型预测精度, 拟以最优单一 ARIMA 及 RBF 模型为基准建立组合模型。采用 25d 样本集 ARIMA 温度预测和输入变量  $B_1$  的 20d 样本集的 RBF 温度预测, 建立温度权重组合模型; 采用 15d 样本集 ARIMA 湿度预测和输入变量 A 的 25d 样本集的 RBF 湿度预测, 建立湿度权重组合模型。求解线性规划问题使得误差平方和最小, 得到温湿度权重组合模型加权系数向量。据此得到温度组合模型的预测值为  $Y_{3t} = 0.63 Y_{1t} + 0.37 Y_{2t}$ , 湿度组合模型的预测值为  $Y_{3t} = 0.39 Y_{1t} + 0.61 Y_{2t}$ 。

#### 2.3.2 组合模型预测结果及分析

由图 3a1、3a2 可见, 单独使用 RBF 神经网络的温度预测结果误差相对较大, 尤其是在低温时段附近误差明显, ARIMA 模型在低温时段附近的预测结果与实际值更接近, 表明温度变化受时序影响较大。除低温时段外, ARIMA-RBF 模型温度预测曲线与实际值曲线更贴近, 误差更小。由图 3b1、3b2 可见, ARIMA 模型的湿度预测效果较差, 特别是湿度低值阶段误差最大, 表明湿度此时可能受到非线性因素的影响较大。由于 RBF 在捕捉数据序列非线性特征上效果较好, 因此, 其在湿度低值阶段的预测曲线与实际值曲线相对更吻合。总体上看, ARIMA-RBF 模型湿度预测结果与实际值最接近, 预测误差曲线最靠近 0 值线。

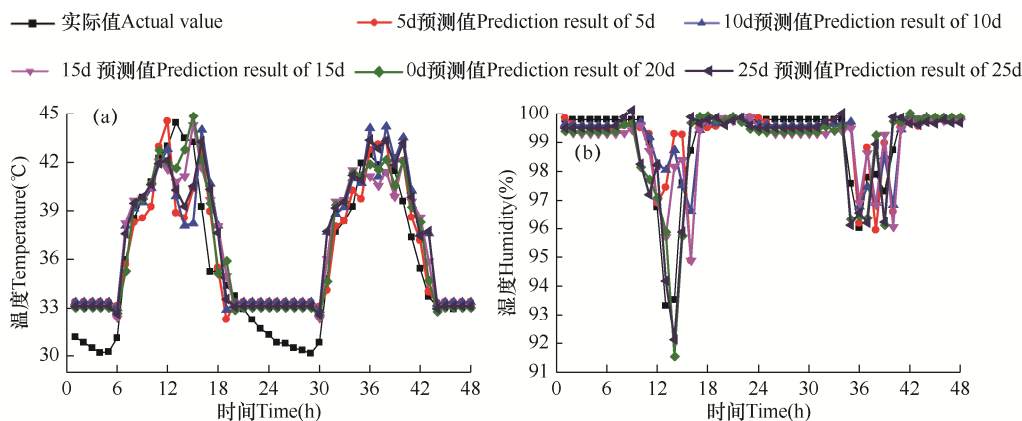


图 2 输入变量  $B_1$  的温度 RBF 模型(a)和输入变量 A 的湿度 RBF 模型(b)48h 预测结果

Fig. 2 48 hours prediction results of temperature RBF models with input variable  $B_1$ (a) and humidity RBF models with input variable A(b)



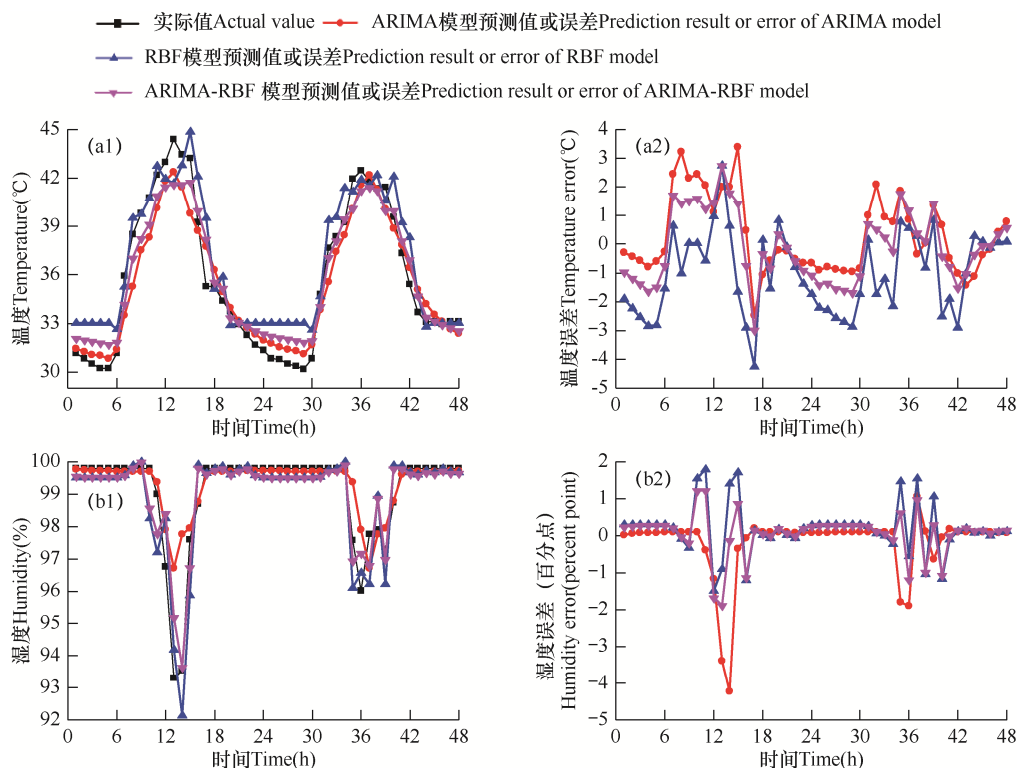


图 3 温度(a)和湿度(b)ARIMA、RBF、ARIMA-RBF 模型 48h 预测结果(1)及误差(2)对比

Fig. 3 48 hours prediction results(1) and errors(2) comparison of temperature(a) and humidity(b) ARIMA, RBF and ARIMA-RBF models

### 2.3.3 组合模型预测性能评价

由表 3 可见, ARIMA-RBF 模型温度预测结果

MAPE 略高于 ARIMA 模型, 除此之外, ARIMA-RBF 模型温湿度预测结果 3 项误差 MAE、MAPE 和 RMSE

表 3 温湿度最佳 ARIMA、RBF、ARIMA-RBF 模型及 48h 预测误差统计结果

Table 3 Optimal temperature and humidity ARIMA, RBF and ARIMA-RBF models and statistical results of 48 hours prediction errors

预测时间 Predicted time	输入变量 Input variable	样本集 Sample set	建模对象 Model object	模型类别 Model type	48h 预测误差统计 Statistical results of 48 hours prediction errors		
					平均绝对误差 MAE	平均绝对百分比误差 MAPE(%)	均方根误差 RMSE
2020-07-28 —29	B <sub>1</sub>	25d	温度 Temperature	ARIMA	1.08	2.94	1.35
		20d		RBF	1.41	4.11	1.75
				ARIMA-RBF	1.04	2.95	1.21
	A	15d	湿度 Humidity	ARIMA	0.39	0.41	0.91
		25d		RBF	0.49	0.50	0.72
				ARIMA-RBF	0.35	0.36	0.55
2019-07-30 —31	C	20d	温度 Temperature	ARIMA	0.89	2.43	1.53
		20d		RBF	1.06	2.84	1.56
				ARIMA-RBF	0.85	2.31	1.21
	B <sub>2</sub>	15d	湿度 Humidity	ARIMA	2.37	2.70	4.33
		5d		RBF	2.60	3.00	3.68
				ARIMA-RBF	2.30	2.65	3.21



均较 ARIMA 和 RBF 模型有不同程度的降低,表明 ARIMA-RBF 模型温湿度预测性能最佳。温度 ARIMA 模型预测结果 3 项误差均小于 RBF 模型,表明温度 ARIMA 模型预测性能优于 RBF 模型。湿度 ARIMA 模型预测结果 MAE、MAPE 小于 RBF 模型, RMSE 大于 RBF 模型;在 3 项误差指标中, RMSE 反映的是模型预测结果异常值情况, RMSE 越小表明异常值越少,预测性能越稳定,可以认为模型综合预测效果更佳。因此,温度预测模型总体效果排序为 ARIMA-RBF>ARIMA>RBF,湿度预测模型总体效果排序为 ARIMA-RBF>RBF>ARIMA。相同方法在 2019 年 7 月 30–31 日为验证集的温湿度建模与分析中也得出了类似结论。

### 3 结论与讨论

#### 3.1 结论

(1) 应用 5~25d 的样本集对蔬菜温室内温湿度的 ARIMA 模型预测表明, 25d 样本集温度模型 ARIMA(0,1,1)(1,1,1)<sub>24</sub> 的预测精度最高,其预测结果的 MAE、MAPE 和 RMSE 分别为 1.08℃、2.94%和 1.35℃; 15d 样本集湿度模型 ARIMA(2,0,0)(1,0,1)<sub>24</sub> 的预测精度最高,其 3 项误差指标分别为 0.39 个百分点、0.41%和 0.91 个百分点。

(2) 应用 5~25d 的样本集和 3 种输入变量对温湿度的 RBF 模型预测表明,输入变量 B<sub>1</sub>(湿度、光照强度)的 20d 样本集温度 RBF 模型(2-10-1)的预测精度最高,其预测结果 MAE、MAPE 和 RMSE 分别为 1.41℃、4.11%和 1.75℃;输入变量 A(温度、湿度、光照强度)的 25d 样本集湿度 RBF 模型(3-6-2)的预测精度最高,其 3 项误差指标分别为 0.49 个百分点、0.50%和 0.72 个百分点。

(3) 最优 ARIMA 模型和最优 RBF 模型的权重组合模型预测表明,温度预测模型总体预测效果排序为 ARIMA-RBF>ARIMA>RBF,湿度预测模型总体预测效果排序为 ARIMA-RBF>RBF>ARIMA。温度 ARIMA-RBF 模型预测结果 MAE、MAPE 和 RMSE 分别为 1.04℃、2.95%、1.21℃;湿度 ARIMA-RBF 模型 3 项误差指标分别为 0.35 个百分点、0.36%和 0.55 个百分点。

#### 3.2 讨论

本研究发现, 15~25d 样本集的 ARIMA 温度模型预测结果与实际值较为接近,其对温度最高时段和最低时段的预测结果也较为理想,表明温度具有

明显的时间序列特征,其受到自身历史值的影响较为明显,这与田东等<sup>[5]</sup>认为温室温度与自身历史温度数据有较大关联,以及徐映梅等<sup>[12]</sup>认为 ARIMA 模型在预测简单时间序列数据更具优势的观点类似。随着样本集天数的逐渐减少, ARIMA 模型温度预测结果的 3 项误差逐渐增大,表明较多的训练数据更有利于捕捉温度时间序列的数据特征。ARIMA 湿度模型在湿度最低时段具有较为明显的偏差,表明湿度的非线性变化特征更为明显,而 ARIMA 模型对非线性关系的挖掘较欠缺<sup>[10]</sup>,其并不能准确地刻画非线性特征,因此预测误差偏大。

前人研究表明,神经网络法受输入参数、样本选择的影响很大<sup>[14]</sup>。本研究发现,输入变量 B<sub>1</sub>(湿度、光照强度)和 C(光照强度)的温度 RBF 预测模型优于输入变量 A(温度、湿度、光照强度)模型,输入变量 A 和 C 的湿度 RBF 预测模型优于输入向量 B<sub>2</sub>(温度和光照强度)模型。说明输入变量温度、湿度和光照强度间相互耦合的复杂关系会对预测效果产生影响,模型训练时需要输入参数进行筛选<sup>[15]</sup>,具有较多维度输入数据特征的神经网络不一定占优势<sup>[16]</sup>。总体上,输入变量 B<sub>1</sub>的 20d 样本集模型的温度预测以及输入变量 A 的 25d 样本集模型的湿度预测精度最高,表明训练数据较多,模型泛化能力更好。已有研究表明,样本集与测试集相似性越高,越有利于提高预测精度<sup>[9]</sup>;本研究的温湿度预测最佳样本集结果说明,预测日温、湿度可能分别与前 20d、前 25d 的样本集相似度较高,表明以 1 个月为时间跨度对温湿度进行建模可能比较合理,这与田东等<sup>[5]</sup>设定的建模周期类似。

RBF 模型对低温阶段预测较差和低湿阶段预测更佳,印证了 ARIMA 模型中对于温度的线性特征和湿度的非线性特征的判断;这与程俊等<sup>[11]</sup>认为,擅长捕捉非线性特征的神经网络,对具有线性特征的时间序列可能存在过度拟合,继而导致预测结果不够精确的观点一致。预测误差统计结果表明,湿度 RBF 模型的 RMSE 较 ARIMA 模型低,但 MAE 和 MAPE 均比后者高,表明湿度可能受到温光湿以外的非线性因素影响较大。温室内部水肥灌溉系统的运行可能对湿度产生了较大的影响, Jung 等<sup>[3]</sup>的研究也得出了类似的结论。本研究仅利用室内温光湿数据,探讨其对温湿度变化的影响,暂未考虑室外环境及调控措施与温湿度变化的关联性。为进一步提

高预测性能,后期需要补充灌溉历史等相关数据进行模型的再次拟合与筛选。

权重组合模型是以预测误差平方和最小为目标函数的权重系数最优解计算方法,该模型通过适当的加权平均形式得出预测结果,综合利用了各种单项预测模型所提供的信息,充分考虑了预测目标的不同数据特征,因此,可以降低信息的不完备性,从而能够更加有效地提高预测精度<sup>[17-18]</sup>。温室温湿度数据具有线性的趋势性和季节性特征;同时受到周围环境参数的影响,具有非线性特征<sup>[6]</sup>。本研究结果表明,ARIMA-RBF模型通过ARIMA、RBF模型分别对线性和非线性数据特征的提取,能够更好地拟合温室环境影响因子与温湿度之间的关系,总体预测效果更加理想,泛化性能更好,与温室内实际温湿度较为接近,能够更好地评估温室温湿度状态。考虑到预测过程类似,本研究仅展示了7月的预测结果,针对其他月份的温室环境数据,可以利用本研究提出的预测方法分别训练不同的模型,以达到预期的效果。本研究提出的组合预测模型权重系数在预测各个时期是固定不变的,然而针对单一ARIMA和RBF模型,其在不同时刻预测精度并不固定,因此,针对单一预测模型在各个预测时期的拟合精度高低进行有序赋权的变权组合预测模型<sup>[19]</sup>还有待于进一步研究。

#### 参考文献 References

- [1] 蔡淑芳,吴宝意,雷锦桂.基于光温效应的大白菜生理特性及营养品质动态模拟效果[J].中国农业气象,2021,42(1):34-43.  
Cai S F,Wu B Y,Lei J G.Dynamic simulation effect of physiological characteristics and nutritional quality of Chinese cabbage based on light and temperature function[J].Chinese Journal of Agrometeorology,2021,42(1):34-43.(in Chinese)
- [2] 牟双.日光温室气温预测及顶部通风调控贝叶斯模型[D].沈阳:沈阳农业大学,2018.  
Mou S.Bays model of solar greenhouse temperature prediction and roof ventilation regulation[D].Shenyang:Shenyang Agricultural University,2018. (in Chinese)
- [3] Jung D H,Kim H S,Jhin C,et al.Time-serial analysis of deep neural network models for prediction of climatic conditions inside a greenhouse[J].Computers and Electronics in Agriculture,2020,173:1-11.
- [4] 韦婷婷,杨再强,王琳,等.玻璃温室和塑料大棚内逐时气温模拟模型[J].中国农业气象,2018,39(10):644-655.  
Wei T T,Yang Z Q,Wang L,et al.Simulation model of hourly air temperature inside glass greenhouse and plastic greenhouse[J].Chinese Journal of Agrometeorology,2018,39(10):644-655.(in Chinese)
- [5] 田东,韦鑫化,王悦,等.基于MA-ARIMA-GASVR的食用菌温室温度预测[J].农业工程学报,2020,36(3):190-197.  
Tian D,Wei X H,Wang Y,et al.Prediction of temperature in edible fungi greenhouse based on MA-ARIMA-GASVR[J].Transactions of the CSAE,2020,36(3):190-197.(in Chinese)
- [6] 赵全明,宋子涛,李奇峰,等.基于CNN-GRU的菇房多点温湿度预测方法研究[J].农业机械学报,2020,51(9):294-303.  
Zhao Q M,Song Z T,Li Q F,et al.Multi-point prediction of temperature and humidity of mushroom based on CNN-GRU[J].Transactions of the Chinese Society for Agricultural Machinery,2020,51(9):294-303.(in Chinese)
- [7] 于学儒.日光温室关键环境因子变化规律与监控系统研究[D].泰安:山东农业大学,2019.  
Yu X R.Study on change regulation of key environmental factors and monitoring and control system in solar greenhouse[D].Tai'an:Shandong Agricultural University,2019.(in Chinese)
- [8] 张坤鳌,赵凯.基于改进CFA PSO-RBF神经网络的温室温度预测研究[J].计算机应用与软件,2020,37(6):95-99.  
Zhang K A,Zhao K.Greenhouse temperature prediction based on improved CFA PSO-RBF neural network[J].Computer Applications and Software,2020,37(6):95-99.(in Chinese)
- [9] 陈昕,唐湘璐,李想,等.二次聚类与神经网络结合的日光温室温度二步预测方法[J].农业机械学报,2017,48(S1):353-358.  
Chen X,Tang X L,Li X,et al.Two-steps prediction method of temperature in solar greenhouse based on twice cluster analysis and neural network[J].Transactions of the Chinese Society for Agricultural Machinery,2017,48(S1): 353-358. (in Chinese)
- [10] 刘春红,杨亮,邓河,等.基于ARIMA和BP神经网络的猪舍氨气浓度预测[J].中国环境科学,2019,39(6):2320-2327.  
Liu C H,Yang L,Deng H,et al.Prediction of ammonia

- concentration in piggery based on ARIMA and BP neural network[J].China Environmental Science,2019,39(6):2320-2327.(in Chinese)
- [11] 程俊.基于 ARIMA-LSTM 混合模型的机械传动件制造企业销售预测方法研究与应用[D].成都:电子科技大学,2018.
- Cheng J.Research and application of sales forecasting method for mechanical transmission parts manufacturing enterprises based on ARIMA-LSTM hybrid model[D].Chengdu:University of Electronic Science and Technology of China,2018.(in Chinese)
- [12] 徐映梅,陈尧.季节 ARIMA 模型与 LSTM 神经网络预测的比较[J].统计与决策,2021(2):46-50.
- Xu Y M,Chen Y.Comparison between seasonal ARIMA model and LSTM neural network forecast[J].Statistics and Decision,2021(2):46-50.(in Chinese)
- [13] 关鹏,焦玉勇,段新胜.基于 RBF 神经网络的土体导热系数非线性预测[J].太阳能学报,2021,42(3):171-178.
- Guan P,Jiao Y Y,Duan X S.Non-liner prediction of soil thermal conductivity based on RBF neural network[J].Acta Energiæ Solaris Sinica,2021,42(3):171-178.(in Chinese)
- [14] 林海涛,蒋传文,任博强,等.基于模糊聚类的风速短期组合预测[J].华东电力,2010,38(2):295-299.
- Lin H T,Jiang C W,Ren B Q,et al.Short term combined prediction of wind speed based on fuzzy clustering[J].East China Electric Power,2010,38(2):295-299.(in Chinese)
- [15] 杨雷,张宝凤,朱均超,等.基于 PCA-PSO-LSSVM 的温室大棚温度预测方法[J].传感器与微系统,2018,37(7):52-55.
- Yang L,Zhang B F,Zhu J C,et al.Temperature prediction method of greenhouse based on PCA-PSO-LSSVM[J].Transducer and Microsystem Technologies,2018,37(7):52-55.(in Chinese)
- [16] 赵琦琳,邱飞,杨健.NARX 神经网络模型在昆明市环境空气质量预测中的应用[J].中国环境监测,2019,35(3):42-48.
- Zhao Q L,Qiu F,Yang J.Application of NARX neural network model in environmental air quality prediction in Kunming[J].Environmental Monitoring in China,2019,35(3):42-48.(in Chinese)
- [17] 施佳.基于 ARIMA-BP 组合模型的某餐饮 O2O 企业订单预测研究[D].北京:北京交通大学,2018.
- Shi J.Research on the order forecast of a restaurant O2O enterprise based on the ARIMA-BP combination model[D].Beijing: Beijing Jiaotong University,2018.(in Chinese)
- [18] 王冬冬.基于 ARIMA 和 BP 神经网络模型的重庆市 GDP 实证分析[D].桂林:广西师范大学,2019.
- Wang D D.Empirical analysis of Chongqing GDP based on ARIMA and BP Neural Network Model[D].Guilin: Guangxi Normal University,2019.(in Chinese)
- [19] 钱超.高速公路 ETC 数据挖掘研究与应用[D].西安:长安大学,2013.
- Qian C.Research and application of data mining on expressway ETC system[D].Xi'an:Chang'an University, 2013.(in Chinese)