

基于 GA-BP 神经网络的雨生红球藻生长趋势预测

崔世钢, 石兰婷, 张永立, 何林, 李欣颀, 张靖宇 (天津职业技术师范大学, 天津 300222)

摘要 虾青素具有极强的抗氧化功能, 然而在自然界中, 能够在相同时间内积累最多虾青素的生物主要为雨生红球藻。但其生长周期较长且对生长环境依赖性强, 因此其生长环境参数与其生长、繁殖状况之间存在着非线性关系。BP 神经网络模型运用了梯度下降法的基本思想, 可以较好逼近复杂的非线性关系。但由于 BP 神经网络容易陷入局部最小, 从而影响预测的结果, 所以采用遗传算法对 BP 神经网络进行优化。根据影响雨生红球藻生长的主要环境(pH), 建立基于遗传算法的 BP 神经网络模型, 并对雨生红球藻生长状况进行试验验证, 准确率可以达到 90% 以上。

关键词 雨生红球藻; BP 神经网络; 遗传算法; 预测; pH

中图分类号 TP 183 文献标识码 A

文章编号 0517-6611(2022)20-0235-05

doi: 10.3969/j.issn.0517-6611.2022.20.060



开放科学(资源服务)标识码(OSID):

Prediction of Growth Trend of *Haematococcus pluvialis* Based on GA-BP Neural Network

CUI Shi-gang, SHI Lan-ting, ZHANG Yong-li et al (Tianjin University of Technology and Education, Tianjin 300222)

Abstract Astaxanthin has a strong antioxidant function. However, in nature, the organisms that can accumulate the most astaxanthin at the same time are mainly *Rhodococcus pluvialis*. However, its growth cycle is long and strongly dependent on the growth environment, so there is a nonlinear relationship between its growth environment parameters and its growth and reproduction. BP neural network model used the basic idea of gradient descent method, which could better approach the complex nonlinear relationship. However, because BP neural network was easy to fall into local minimum, which affected the prediction results, genetic algorithm was used to optimize BP neural network. According to the main environment (pH) affecting the growth of *Haematococcus pluvialis*, a BP neural network model based on genetic algorithm was established, and the growth status of *Haematococcus pluvialis* was experimentally verified. The accuracy could reach more than 90%.

Key words *Haematococcus pluvialis*; BP neural network; Genetic algorithm; Forecast; pH

虾青素是一种酮类胡萝卜素^[1], 比其他天然抗氧化药物具有更强的抗氧化活性。虾青素能够对多种疾病起到辅助治疗的作用, 能够抑制肿瘤生长、保护中枢神经系统、预防心脑血管疾病等, 因此广泛用于食品、医药和化妆品^[2-4]。雨生红球藻在特定的胁迫条件下可以快速积累虾青素, 最大积累量可达其干细胞重量的 1%。近年来, 国内外对于大量生产虾青素的研究已经逐渐向培育雨生红球藻靠拢^[5]。

侯奇等^[6]建立了预测李斯特菌生长的二级 BP 神经网络模型。由于雨生红球藻生长周期长, 而且与大多数微生物不同, 其生长状态分为 2 个阶段, 不同阶段生长条件也不相同, 因此很少有可供参考的客观数据。特别是在试验过程中需要消耗大量的时间对雨生红球藻的生长状况进行观测, 但其生长趋势依旧很难预测, 导致反复试验消耗大量的人力物力依旧不容易得到好的培养效果。因此建立雨生红球藻生长趋势的预测模型, 对提高雨生红球藻生产效率的研究提供重要的参考。鉴于此, 根据影响雨生红球藻生长的主要环境(pH), 建立基于遗传算法的 BP 神经网络模型, 并对雨生红球藻生长状况进行试验验证。

1 雨生红球藻

雨生红球藻(*Haematococcus pluvialis*)是一种生活在淡水中的单细胞绿藻, 其生长状态分为细胞大量增殖阶段和虾青素积累阶段。在环境适宜营养充足的条件下, 该细胞主要呈现绿色并处于游动状态, 此时繁殖迅速, 虾青素含量少。当生存环境恶劣时, 为了保证正常生长游动细胞的鞭毛退化,

细胞停止游动, 为了保证正常生存开始大量积累虾青素, 细胞逐渐变为红色^[7-8]。

根据雨生红球藻的生长特性, 国内外的学者提出了 2 阶段生产方式的生产模式: 第 1 阶段是对藻细胞进行大量的增殖(也被称为细胞增殖阶段); 第 2 阶段是胁迫雨生红球藻细胞快速积累虾青素(也被称为虾青素积累阶段)。大量研究表明, 雨生红球藻的生长需要很多适宜的条件, 如温度、光照强度、生存环境的营养成分、pH 等, 其中 pH 是关键因素之一。雨生红球藻培养过程中, pH 的范围将直接影响藻类细胞的生长和虾青素的积累^[9]。相关研究表明, 当扩培阶段的藻液处于弱碱性(pH=7.0~9.0)时, 光合作用效率大幅提高, 藻细胞大量增殖。当诱导阶段的藻液 pH 为 7.0~8.0^[10]时, 虾青素积累最快^[11]。

2 遗传算法以及 BP 神经网络

BP 神经网络(Back Propagation Neural Network)是一种多层前馈神经网络, 可以通过误差的反向传播来修正层间的参数。通过输入和期望输出, 学习需预测的数据样本。通过输入层将数据传递给隐藏层中的各个神经元, 通过计算得到预测的输出值, 若预测值与实际值相减的误差不能满足预测的精度, 则输入层将所得的误差返回, 从而修改各层之间的权阈值, 使得到的预测输出与实际输出之间的误差逐渐减小, 直至满足预测所需的精度。其学习规则采用梯度下降法, 以此来得到目标函数的最小值, 常用于解决复杂的非线性模型。

雨生红球藻的生长环境参数与生长情况之间存在着非线性关系, 因此采用对非线性函数具有最佳逼近性和全局最优性能的神经网络来预测雨生红球藻的生长条件。该研究将雨生红球藻的培养时的 pH 作为神经网络的输入量, 其他

基金项目 国家重点研发计划项目(2017YFB0403904)。

作者简介 崔世钢(1963—), 男, 天津人, 教授, 博士, 博士生导师, 从事智能控制研究。

收稿日期 2021-11-21; **修回日期** 2021-12-28

环境变量保持不变,将试验测得的生长状态数据作为神经网络的输出量,建立雨生红球藻在不同 pH 生长状态预测模型。其拓扑结构如图 1 所示。

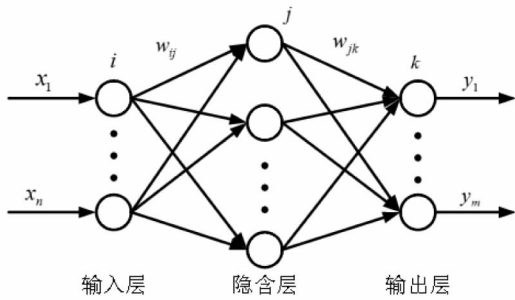


图 1 BP 神经网络结构

Fig.1 Structure of BP neural network

遗传算法 (Genetic Algorithm, GA) 起源于生物进化论,它是模拟自然界生物进化过程中优胜劣汰的一种优化算法。将种群和 DNA 的概念引入需要优化的参数中,形成参数种群编码,按照适应度函数对遗传信息进行选择、交叉、变异等操作,从而实现对个体的筛选。保留具有优秀基因的个体,淘汰适应度差的个体。种群不断地迭代、循环直至得到满足条件的最优解^[12]。

BP 神经网络主要由信息前向传播和误差反向传播组成。信息前向传播指的是待预测数据经过输入层传递到隐藏层中,在隐藏层通过对阈值和权值进行计算得到输出结果。将预测结果和实际值做差得到预测的误差值,将误差值传回隐藏层,并通过误差函数对权值和阈值进行更新,不断重复上述 2 个过程,最终得到理想的参数^[13-15]。

3 GA-BP 神经网络模型

由于 BP 神经网络是采用梯度下降法来对相应的参数进行优化,因此在优化的过程中会因为初值的选取出现局部最小值的问题,这样不仅训练时间长,收敛速度也慢。因此,由于遗传算法具有良好的全局寻优能力,故采用其对 BP 神经网络进行优化,从而找到最优的权值和阈值,提高预测模型的准确率。

3.1 BP 神经网络的建立

3.1.1 设定初始参数。网络的训练方法采用梯度下降法。最大训练次数 epochs 为 2 000,学习率 $\mu=0.005$,目标误差 0.000 01。

3.1.2 确定隐含层节点数。BP 神经网络选取结构为 3 层,分别是 1 个输入神经元、5 个隐含神经元和 1 个输出神经元,分别对细胞和数量进行预测。其中,隐含层神经元数量根据式(1)确定为 5 个,并经过验证可达到最佳的预测效果。

$$m = \sqrt{n+l} + \alpha \quad (1)$$

式中, n 、 l 、 m 分别为输入、输出以及隐含层节点数量, α 取 1~10 的整数。

3.1.3 转移函数的选取。隐含层的传递函数采用 Tanh 函数,值域为 $(-1, 1)$ 。其定义为:

$$\text{Tanh}(x) = 2\sigma(2x) - 1 = \frac{2}{1+e^{-2x}} - 1 \quad (2)$$

输出层的传递函数采用 Sigmoid 函数。其函数定义见式

(3):

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (3)$$

3.2 遗传算法优化 BP 神经网络 将遗传算法的 3 类遗传算子与 BP 神经网络的权值与阈值结合以来,通过对数据进行训练,找到最优的参数,进而达到对雨生红球藻的生长趋势的精准预测。其模型建立步骤如下:

3.2.1 种群初始化。种群的个体是 BP 神经网络的所有权值和阈值,对个体采用实数编码。编码采用二进制表示,长度为:

$$S = n \times m + m + m \times l + l \quad (4)$$

种群的规模要根据实际问题进行选取,才能使遗传算法的优化性能达到最好。因此,初始种群为 15。

3.2.2 适应度函数。通过编码个体得到 BP 神经网络的初始权值和阈值,以 BP 神经网络训练得到的输出与实际输出之间的绝对误差 E 推出个体适应度值 f_i 。

$$E_i = \text{abs}(y_i - o_i)$$

$$f_i = \frac{1}{\left(\sum_{i=1}^n E_i\right)} \quad (5)$$

式中,BP 神经网络的预期输出和实际输出分别为 y_i 、 o_i 。根据适应度函数,期望输出和实际输出的误差越小,适应度函数越大。

3.2.3 选择操作。从种群中选择优秀个体作为繁殖后代的母体。方法采用轮盘赌选择法。主要思想是每个个体被选中的概率与适合度成正比。即:

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j} \quad (6)$$

3.2.4 交叉操作。交叉操作是遗传基因重组的过程,以便将当前的最优基因传到下一代并获得新的个体。从群体中随机选择 2 个个体进行配对,根据交叉概率 $p_c=0.2$ 交换 2 个个体的部分基因。从而获得 2 个具有新基因的个体。将随机选取的个体 ω_1 、 ω_2 在第 k 位基因的交换如下:

$$\begin{cases} g_{\omega_1,k} = g_{\omega_1,k} a + g_{\omega_2,k} (1-a) \\ g_{\omega_2,k} = g_{\omega_2,k} a + g_{\omega_1,k} (1-a) \end{cases} \quad (7)$$

式中, b 为 $[0, 1]$ 的随机数。

3.2.5 变异操作。变异操作相当于生物学上的基因突变,根据变异概率 $p_m=0.1$ 对第 i 个个体的第 j 个基因 g_{ij} 进行变异,随机产生优秀的生物个体,丰富种群的多样性。其变异方法如下:

$$g_{ij} = \begin{cases} g_{ij} + (g_{ij} - g_{\max}) b_1 (1 - d/D_{\max}) & b_1 > 0.5 \\ g_{ij} + (g_{\min} - g_{ij}) b_1 (1 - d/D_{\max}) & b_1 \leq 0.5 \end{cases} \quad (8)$$

式中, b_1 为随机数, d 为当前迭代次数, D_{\max} 为最大进化次数, g_{\max} 为基因 g_{ij} 的上界, g_{\min} 为基因 g_{ij} 的下界。

GA-BP 神经网络算法流程图见图 2。

4 试验与数据采集

4.1 试验方法与设计 雨生红球藻藻种 (*Haematococcus pluvisialis*, FACHB-712) 购于上海光语生物科技有限公司藻种

库。藻种培养基选用 BG11 培养基,按照藻液:培养基=1:2 的比例进行配比。采用直流驱动的单色光照,光照黑暗时间对比为 12 h:12 h,每天测量一组数据。

细胞增殖阶段培养基 pH 分别为 7.5、8.0、8.5,将配制好的藻液放置到光照强度为 $20 \mu\text{mol}/(\text{m}^2 \cdot \text{s})$ 且温度为室温的光生物培养架上。

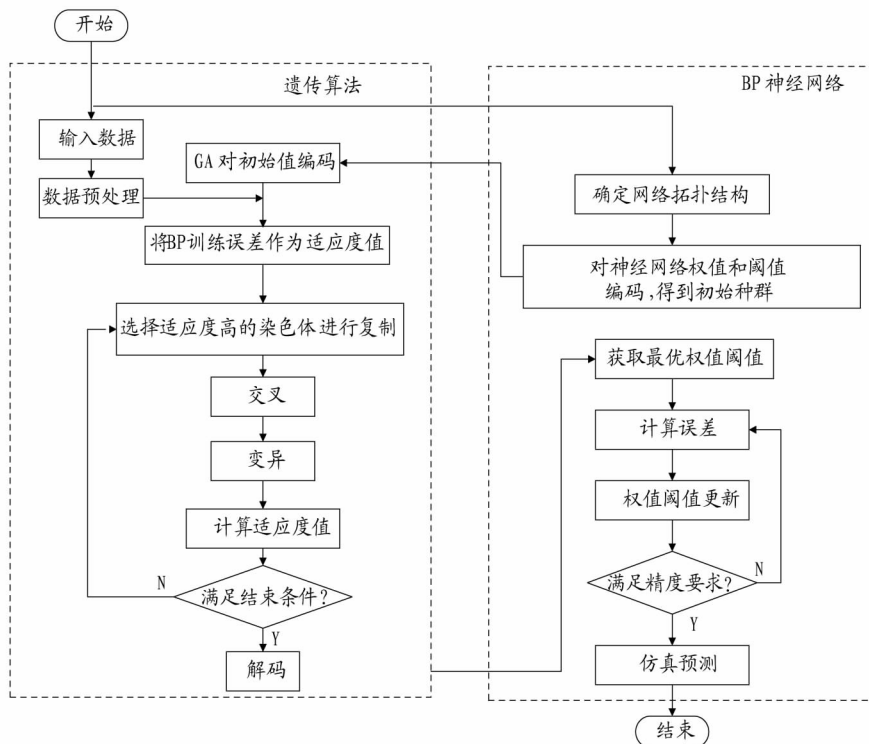


图 2 GA-BP 神经网络算法流程

Fig.2 GA-BP neural network algorithm flow

4.2 样本数据采集 每次从试验样本中取出 15 mL 藻液,在超净工作台进行取样以及数据测量,共测量 70 d 增殖试验的藻细胞半径和数量。测量第 35 天的数据后,在锥形瓶中加入 200 mL BG11 培养基,为藻的进一步增殖提供所需的营养物质。

4.3 数据预处理 由于需要预测的藻细胞数量和半径具有不同的数量级且差异较大,会影响网络的预测结果,因此需要对原始数据进行归一化处理。采用 Matlab 的 $\text{mapminmax}(x)$ 函数对数据进行归一化处理,使处理后的数据范围在 $[-1, 1]$, 其转换公式为:

$$y = \frac{(y_{\max} - y_{\min}) \times (x - x_{\min})}{x_{\max} - x_{\min}} + y_{\min} \quad (9)$$

式中, x 为藻细胞数据的初始值, x_{\max} 为藻细胞数据的最大值, x_{\min} 为藻细胞数据的最小值, y_{\max} 、 y_{\min} 为参数,可自己设定也可默认为 $[-1, 1]$ 。经过归一化处理的数据,在预测结束后需要进行反归一化,将数据恢复到正常值,该研究反归一化采用 postmnmx 函数。

5 结果与分析

采用 Matlab 作为预测的平台,利用 Matlab 自带的函数实现神经网络的搭建。将收集到的 70 d 试验数据作为预测的训练集和测试集,前 60 d 为训练集,后 10 d 为测试集。为了得到不同 pH 下细胞平均半径和细胞数量变化趋势,采用 BP 和 GA-BP 神经网络分别对 pH=7.5、8.0、8.5 进行预测。预测结果如图 2~4 所示,部分预测数据如表 1、2 所示。

表 1 BP 神经网络部分预测结果

Table 1 Partial prediction results of BP neural network

pH	细胞半径 Cell radius				细胞数量 Number of cells			
	实际值 Actual value μm	预测值 Estimate μm	绝对误差 Absolute error// μm	相对误差 Relative error//%	实际值 Actual value 万个/mL	预测值 Estimate 万个/mL	绝对误差 Absolute error 万个/mL	相对误差 Relative error//%
7.5	3.31	3.32	0.01	0.30	349	365	16	4.58
	3.32	3.33	0.01	0.30	375	351	24	6.40
8.0	3.20	3.21	0.01	0.31	354	350	4	1.13
	3.21	3.23	0.02	0.62	364	362	2	0.55
8.5	3.22	3.23	0.01	0.30	337	330	7	2.08
	3.23	3.26	0.03	0.93	349	356	7	2.01

表 2 GA-BP 神经网络部分预测结果

Table 2 Partial prediction results of GA-BP neural network

pH	细胞半径 Cell radius				细胞数量 Number of cells			
	实际值 Actual value μm	预测值 Estimate μm	绝对误差 Absolute error//μm	相对误差 Relative error//%	实际值 Actual value 万个/mL	预测值 Estimate 万个/mL	绝对误差 Absolute error 万个/mL	相对误差 Relative error//%
7.5	3.31	3.31	0	0	349	335	14	4.01
	3.32	3.31	0.01	0.3	375	355	20	5.33
8.0	3.20	3.24	0.04	1.25	354	355	1	0.28
	3.21	3.25	0.04	1.24	364	360	4	1.10
8.5	3.22	3.23	0.01	0.31	337	320	17	5.04
	3.23	3.25	0.02	0.62	349	340	9	2.59

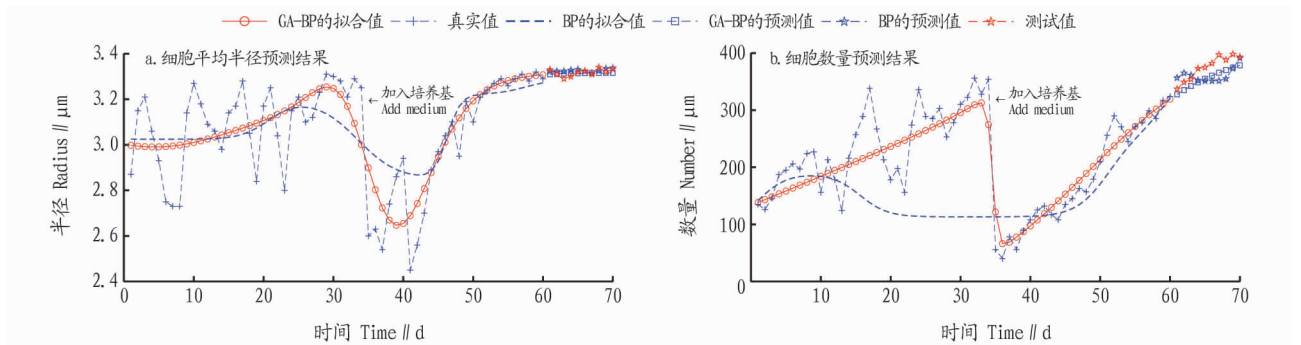


图 3 pH 7.5 时细胞增殖阶段藻细胞的生长趋势

Fig.3 Growth trend of algal cells in cell proliferation stage at pH 7.5

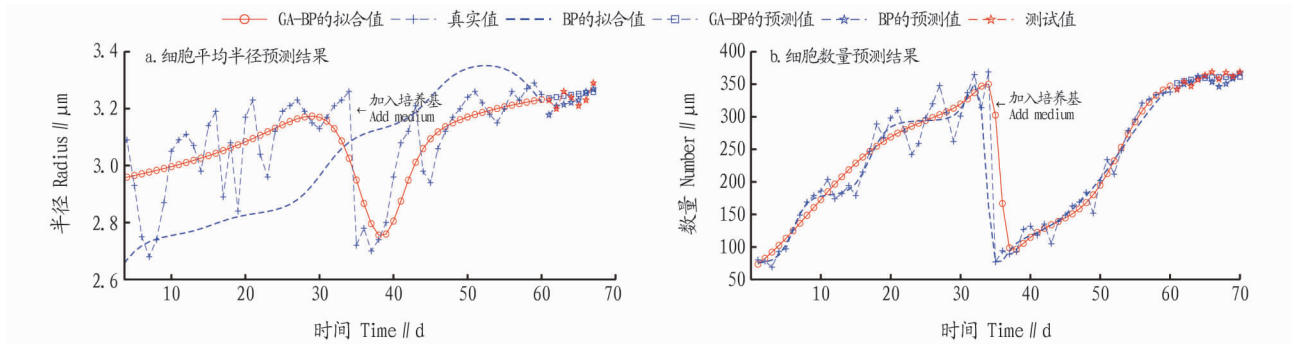


图 4 pH 8.0 时增殖培养阶段藻细胞的生长趋势

Fig.4 Growth trend of algal cells in cell proliferation stage at pH 8.0

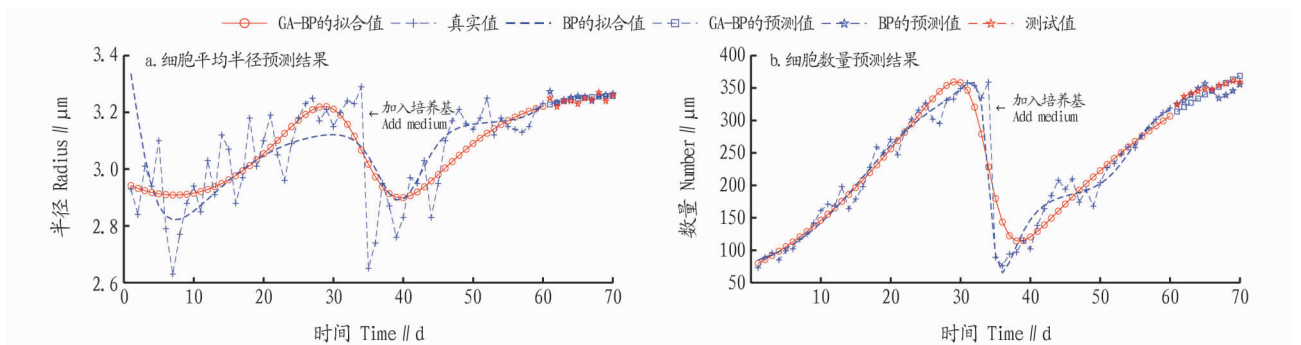


图 5 pH 8.5 时增殖培养阶段藻细胞的生长趋势

Fig.5 Growth trend of algal cells in cell proliferation stage at pH 8.5

采用平均绝对百分比误差 (mean absolute percentage error, MAPE) 和均方根误差 (root mean square error, RMSE) 对预测模型性能进行评价。

$$MAPE = \frac{1}{M} \sum_{i=1}^M \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (10)$$

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M (\hat{y}_i - y_i)^2} \quad (11)$$

式中, M 为需要预测样本的数量, y_i 为数据的真实值, \hat{y}_i 为数据的预测值。

其评价结果如表 3、4 所示。从上述试验可以看出, GA-BP 神经网络的 MAPE 和 RMSE 均低于 BP 神经网络。因此, 可以得出 GA-BP 神经网络预测的准确率较高。

表 3 GA-BP 和 BP 神经网络预测半径评价结果

Table 3 Evaluation results of predicted radius of GA-BP and BP neural network

神经网络 Neural network	pH 7.5		pH 8.0		pH 8.5	
	MAPE %	RMSE	MAPE %	RMSE	MAPE %	RMSE
BP	0.40	0.016	0.78	0.030	0.39	0.015
GA-BP	0.37	0.015	0.60	0.024	0.31	0.013

表 4 GA-BP 和 BP 神经网络预测数量评价结果

Table 4 GA-BP and BP neural network prediction quantity evaluation results

神经网络 Neural network	pH 7.5		pH 8.0		pH 8.5	
	MAPE %	RMSE	MAPE %	RMSE	MAPE %	RMSE
BP	5.91	25.36	1.82	8.68	2.34	10.52
GA-BP	4.98	19.90	1.46	6.27	2.27	9.47

6 结论

通过对雨生红球藻在不同 pH 下的生长状况数据进行训练, 得到了根据时间预测藻细胞生长趋势的模型。从训练结果看, 在不同 pH 下, GA-BP 神经网络对 10 d 数据的预测结果精确度高于 90%, 且较合理地预测出了藻细胞的生长趋势, 说明网络架构已达到预期效果。

该模型的使用一方面可以快速得出雨生红球藻在一定

时间内的生长状况, 克服其周期长的问题; 另一方面能够更节省人力物力, 在试验前通过模拟预测, 可以得到单变量环境对雨生红球藻生长状况的影响, 方便研究人员更好地找到最适合藻细胞生长的环境值, 在培养过程中能够更节省时间避免过多的重复试验, 也能够大大降低试验的失败率。

参考文献

- [1] 高桂玲, 成家杨, 马炯. 雨生红球藻和虾青素的研究[J]. 水产学报, 2014, 38(2): 297-304.
- [2] 彭永健, 吕红萍, 王胜南, 等. 天然虾青素的研究进展[J]. 中国食品添加剂, 2017(4): 193-197.
- [3] 赵英源, 刘俊霞, 陈姝彤, 等. 虾青素生理活性的研究进展[J]. 中国海洋药物, 2020, 39(3): 80-88.
- [4] 李艳, 高静. 雨生红球藻天然虾青素提取研究进展[J]. 食品与机械, 2020, 36(12): 211-217.
- [5] 杨慧敏. 雨生红球藻高产虾青素的培养条件研究[D]. 湛江: 广东海洋大学, 2016.
- [6] 侯奇, 刘静, 管骁. 基于神经网络的微生物生长预测模型[J]. 食品与机械, 2018, 34(2): 120-123.
- [7] 陶云莹. 四种环境因子对雨生红球藻 (*Haematococcus pluvialis*) 生长、虾青素及内源激素积累的影响[D]. 青岛: 中国海洋大学, 2015.
- [8] 李颖逾. 雨生红球藻细胞转化及虾青素积累与培养因子关系研究[D]. 青岛: 中国科学院研究生院(海洋研究所), 2005.
- [9] 黄韵祺. 基于神经网络的藻液 pH 值控制系统研究[D]. 天津: 天津职业技术师范大学, 2019.
- [10] 吴娇, 刁庆宇, 白雪, 等. 雨生红球藻的培养及虾青素的提取与检测[J]. 大连民族大学学报, 2019, 21(5): 406-411.
- [11] 侯冬梅. 雨生红球藻高产虾青素的光诱导工艺研究[D]. 上海: 华东理工大学, 2014.
- [12] 高玉明, 张仁津. 基于遗传算法和 BP 神经网络的房价预测分析[J]. 计算机工程, 2014, 40(4): 187-191.
- [13] 姜晨, 刘喜波. 基于 GA-BP 神经网络模型的 P2P 网贷借款人信用风险预测研究[J]. 商展经济, 2021(1): 49-51.
- [14] WU X Z, NIU H, LI X J, et al. A study on the GA-BP neural network model for surface roughness of basswood-veneered medium-density fiberboard[J]. Holzforschung, 2020, 74(10): 979-988.
- [15] LI Y F, REN C, ZHAO H W, et al. Investigating long-term vehicle speed prediction based on GA-BP algorithms and the road-traffic environment[J]. Science China information sciences, 2020, 63(9): 1-3.

(上接第 229 页)

- [25] 辛承松, 董合忠, 唐薇, 等. 棉花盐害与耐盐性的生理和分子机理研究进展[J]. 棉花学报, 2005, 17(5): 309-313.
- [26] 蒋玉蓉, 吕有军, 祝水金. 棉花耐盐机理与盐害控制研究进展[J]. 棉花学报, 2006, 18(4): 248-254.
- [27] 魏俊梅. 24 个棉花种质资源耐盐性差异比较[J]. 中国棉花, 2014, 41(5): 14-16.
- [28] 孙小芳, 刘友良. 棉花品种耐盐性鉴定指标可靠性的检验[J]. 作物学

报, 2001, 27(6): 794-801.

- [29] 辛承松, 罗振, 吴振美. 抗虫棉不同类型品种苗期耐盐化特性差异研究[J]. 棉花学报, 2012, 24(5): 406-413.
- [30] ASHRAF M, AHMAD S. Influence of sodium chloride on ion accumulation, yield components and fibre characteristics in salt-tolerant and salt-sensitive lines of cotton (*Gossypium hirsutum* L.) [J]. Field crops research, 2000, 66(2): 115-127.
- [31] 刘伟, 崔淑芳, 张海娜, 等. 棉花对盐胁迫的响应机制及缓解措施的研究进展[J]. 农学学报, 2015, 5(11): 10-16.

(上接第 234 页)

- [10] 艾瑞咨询. 2018—2019 中国农村电商行业分析及商业模式研究报告 [EB/OL]. [2021-09-23]. <https://www.iimedia.cn/c400/63427.html>.
- [11] 刘海宏, 吴睿辉. 新时代背景下农业电商发展模式研究[J]. 中国物流与采购, 2019(8): 75-80.
- [12] 刘伟. 新时代背景下农业电商发展模式探讨[J]. 农业经济, 2018(9): 130-132.
- [13] 智研咨询. 2022-2028 年中国农村电商行业供需态势分析及竞争格局

预测报告 [EB/OL]. [2020-12-15]. <https://www.chyxx.com/research/202010/903661.html>.

- [14] 洪涛. 2020 中国农产品电商发展报告(摘要)[J]. 农业工程技术, 2020, 40(9): 25-28.
- [15] 金恩焘. 数字乡村战略下农产品流通体系创新研究[D]. 南昌: 南昌大学, 2020.
- [16] 陶甄. 农产品电商的战略成本管理问题分析: 以深圳农产品股份有限公司为例[D]. 南昌: 江西财经大学, 2020.