

文章编号:1673-1689(2010)06-0921-06

基于 BP 神经网络的洋葱伯克霍尔德菌脂肪酶 发酵软测量建模

汪小锋, 付强, 申旭光, 杨江科, 闫云君*

(华中科技大学 生命科学技术学院, 教育部分子生物物理重点实验室, 湖北 武汉 430074)

摘要: 为建立洋葱伯克霍尔德菌(*Burkholderia cepacia*)脂肪酶发酵过程的软测量模型, 运用 BP 神经网络对洋葱伯克霍尔德菌脂肪酶的发酵过程进行软测量建模, 并利用遗传算法对神经网络的初始权值和阈值进行优化, 实现模型加快收敛速度, 达到全局最优解效果。该模型能够比较精确地模拟菌体生长、底物消耗以及发酵产酶的过程动态, 具有良好的泛化能力, 说明 BP 神经网络结合遗传算法在洋葱伯克霍尔德菌脂肪酶发酵过程的模拟与预测中是一种高效快速的方法。

关键词: 洋葱伯克霍尔德菌脂肪酶; BP 神经网络; 遗传算法; 发酵; 模型

中图分类号: Q 939

文献标识码: A

Soft Sensor Modeling of *Burkholderia cepacia* Lipase Fermentation Process Based on BP Neural Network

WANG Xiao-feng, FU Qiang, SHEN Xu-guang, YANG Jiang-ke, YAN Yun-jun*

(Key Laboratory of Molecular Biophysics of Ministry of Education, College of Life Science and Technology, Huazhong University of Science & Technology, Wuhan 430074, China)

Abstract: The aim of this study was to establish a soft sensor model of the *Burkholderia cepacia* lipase fermentation process. Procedure and method: for this, the backpropagation neural network was applied to rebuild a soft sensor model for the fermentation process of *Burkholderia cepacia* lipase, and genetic algorithms was further employed to optimize the initial weights and thresholds of neural networks, so as to speed up the convergence rate and achieve a comprehensive optimum result. Result: The model could imitate precisely the process of bacteria growth, culture medium consumption and lipase production, which suggests that the model possesses good reliability and generalization ability. The backpropagation neural network combined with genetic algorithms is an effective and quick method for simulating and predicating fermentation process of *B. cepacia* lipase.

Key words: *Burkholderia cepacia* lipase, BP neural network, genetic algorithm, fermentation, modeling

收稿日期: 2009-12-30

基金项目: 国家 863 计划重点项目(2006AA020203, 2007AA100703, 2009AA03Z232); 国家 863 计划专题项目(2007AA05Z417); 教育部新世纪优秀人才支持计划项目(NCET-07-0336); 武汉市科技攻关计划重点项目(200720422138)。

作者简介: 汪小锋(1983—), 男, 湖北安陆人, 微生物学博士研究生。Email: wxf7907@163.com。

* **通信作者:** 闫云君(1969—), 男, 湖北罗田人, 理学博士, 教授, 主要从事能源生物技术方面的研究。Email: yanyunjun@tom.com

脂肪酶(EC 3. 1. 1. 3)是一类特殊的酯键水解酶,能在油水界面上催化脂类物质分解、合成和酯交换等反应。作为重要的工业用酶,脂肪酶在食品、制革、饲料、洗涤、油酯化工等传统工业领域应用十分广泛,近年在生物能源、有机合成、药物手性拆分和工具酶等新型应用领域前景也十分广阔^[1-2]。洋葱伯克霍尔德菌脂肪酶(*B. cepacia* Lipase, BCL)对多种有机溶剂(醇)、热、氧化剂、表面活性剂、去污剂、蛋白酶等有较好抗性,是目前在有机合成、洗涤剂添加剂和非水相催化中应用最为广泛的脂肪酶之一^[3]。

发酵过程是具有高度的时变性、非线性、带约束的多输入多输出的复杂系统,基本生物量难以在线测量^[4],发酵过程的建模和控制十分复杂,对发酵过程实施有效的在线控制和优化已成为目前发酵工程的重要研究课题。人工神经网络(Artificial neural networks, ANN)是一种模仿动物神经网络行为特征,进行分布式并行信息处理的数学模型算法。它由多个简单处理单元按照某种规则相互连接而成,具有自学习、记忆、联想和计算功能,通过调整内部大量节点之间相互连接的关系,进行大规模非线性自适应信息处理,自动调节不同类型的非线性响应从而达到处理信息的目的^[5]。近年来,神经网络在生物工程方面的应用越来越广泛,主要用于发酵培养基优化、发酵过程仿真、发酵过程控制优化、故障诊断和软测量等方面^[6-7]。遗传算法(Genetic algorithm, GA),是模拟达尔文的遗传选择和自然淘汰的生物进化过程的计算模型,具有简单通用、鲁棒性强、适于并行处理等特点^[7]。作者采用软测量技术,利用可实时测量的参数(pH值、溶氧和时间)估计不能在线测量的参数(脂肪酶活力、残糖浓度和生物量),结合遗传算法和BP神经网络,建立洋葱伯克霍尔德菌脂肪酶发酵软测量模型。

1 材料与方法

1.1 菌株

洋葱伯克霍尔德菌(*B. cepacia*) PCL-3,作者所在实验室筛选、鉴定并保藏。

1.2 发酵罐和电极

BIOTECH-10JGZ型10L搅拌发酵罐;上海保兴生物设备工程有限公司制造;pH电极、溶氧电极:购于瑞士Mettler Toledo公司。

1.3 培养基

斜面培养基(g/dL):蛋白胨 1,酵母膏 0.5,

NaCl 1.0,琼脂 2.0;

种子培养基(g/dL):葡萄糖 0.5,蛋白胨 1, K_2HPO_4 0.1, $MgSO_4 \cdot 7H_2O$ 0.05;

发酵产酶培养基(g/dL):葡萄糖 1,工业蛋白胨 2, K_2HPO_4 0.2, $MgSO_4 \cdot 7H_2O$ 0.05,尿素 0.14;橄榄油乳化液体积分数 2%,初始 pH 8.3~8.4。

1.4 10 L 发酵罐培养条件

将活化后的菌种接种到 75 mL (500 mL 三角瓶)的种子培养基中,28 ℃、250 r/min 培养 12 h。10 L 发酵罐装液量为 60%,121 ℃灭菌 30 min,冷却至 30 ℃接种,接种体积分数 3%,搅拌转速 300 r/min,通风量 1.0 vvm,通过流加泡敌消泡。从第 0 小时开始,每隔 4 h 取一次样,发酵周期为 48 h。在线测量数据有 pH 值、溶氧,离线化验数据有酶活力、菌体质量浓度、残糖质量浓度。

1.5 分析方法

1.5.1 菌体干重测定 取 10 mL 发酵液于预先干燥好的离心管中,5 000 r/min 离心 10 min,弃上清液,并用无菌水离心洗涤 2 次,弃上清液,于 80 ℃干燥至恒重,用电子天平称重。

1.5.2 残糖浓度的测定 采用 3,5-二硝基水杨酸法^[8]。

1.5.3 脂肪酶活力的测定 脂肪酶活性测定采用 pNPP 法^[9]。将 pNPP 溶解在乙腈中,终浓度为 10 mmol/L,依次加入乙醇和 50 mmol/L Tris-HCl 缓冲液,使最终 $V_{乙腈} : V_{乙醇} : V_{Tris-HCl} = 1 : 4 : 95$ 。将发酵液于 12 000 r/min 离心 1 min 后,取适当稀释后的 50 μ L 酶液加入到 4 mL 经 45 ℃预热 5 min 后的 pNPP 底物中,反应 5 min 后,加入 50 μ L 0.5 mol/L EDTA,立即放冰水中中止反应。用紫外可见分光光度计(405 nm)读取吸光度值,根据对硝基苯酚的浓度计算酶活。一个单位的脂肪酶活力定义为:45 ℃下,pH 8.5 时每分钟分解 pNPP 并释放出 1 μ mol 对硝基苯酚所需的酶量。

2 结果与分析

2.1 BP 模型的构造方法

发酵过程的一些重要参数如菌体浓度、基质浓度、产物浓度等是不能在线测量的,而这些参数的实时监测对于整个发酵过程的控制又是极其重要的。因此,作者采用软测量技术,利用可以实时测量的参数(如时间、pH 值、溶氧)估计不能在线测量的参数,实现对整个发酵过程的有效模拟。采取分批发酵的方法,选取 11 批分批发酵样本数据,其中

10 批分批发酵数据为 BP 神经网络的训练样本,1 批分批发酵数据为验证样本。模型输入: $u=[t; \text{pH}; \text{DO}; \text{pH}_1; \text{DO}_1]^T$,全部为在线可测参数。 $t, \text{pH}(t), \text{DO}(t)$ 分别为发酵时间(小时)、pH 值、DO, pH_1, DO_1 分别表示前一时刻的 pH, DO。模型输出:酶活力(U/mL)模型输出为酶活力估计值,菌体模型输出为菌体质量浓度(g/L)估计值,残糖模型输出为残糖质量浓度(g/L)估计值。

从理论上讲,任一非线性结构在满足一定的隐节点数的条件下,总可以通过一个三层神经网络进行任意逼近,故选取三层的神经网络结构即可满足要求的拟合精度和模型的泛化性。将获得的数据进行归一化处理得到训练样本集,输入和输出神经元数目均为已知,采用构造法确定隐含层神经元数目,从而确定 BP 神经网络的结构,然后用遗传算法工具箱对 BP 神经网络的初始权值和阈值进行优化,再将优化后的权值和阈值赋给神经网络训练,最后将训练好的网络用于测试样本的测试以考察其泛化能力。测试的评价标准为仿真输出 y 与实际的输出 T 之间的相对误差。

相对误差公式: $E=(y-T)/T$

2.2 脂肪酶活力的神经网络模型 I

将归一化处理好的 5 个输入变量的值作为 BP 网络的输入,目标向量是离线测量得到的酶活力值。设定隐层数区间 $[16, 40]$,采用构造法逐一测试,测试的精度采用均方误差值 ϵ ,设定 $\epsilon=10^{-5}$,网络的最大训练次数为 5 000,学习速率为 0.05,确定隐层数为 29 时训练次数较低,因此 BP 神经网络的结构为:5-29-1。BP 神经网络在进行学习之前的权值和阈值是随机产生的,初始的权值和阈值会直接影响神经网络的学习速率,增加网络的学习时间。用遗传算法工具箱对 BP 神经网络的初始权值和阈值进行优化可以很好的解决上述问题。设定遗传算法的初始种群为 50,遗传代数 100,神经网络的学习速率为 0.05,网络训练精度为 0,最大训练次数为 5 000。

遗传算法和神经网络训练网络的过程见图 1, 2。由图 3 可知遗传算法优化后的权值和阈值经过神经网络学习后能较为精确的拟合输入的测试样本数据,其最大相对误差不超过 1.5%,说明模型准确可信,泛化能力较好。

2.3 菌体生长的神经网络模型

将归一化处理好的 5 个输入变量的值作为 BP 网络的输入,目标向量是离线测量得到的生物量的数据。采用模型 I 的方法确定隐层数为 27,因此

BP 神经网络的结构为:5-27-1。采用模型 I 的方法用遗传算法工具箱优化神经网络的权值和阈值,并对网络进行训练和测试。遗传算法和神经网络训练网络的过程见图 4,5。由图 6 可知,模型 II 能较为精确的拟合输入的测试样本数据,其最大相对误差不超过 0.06%,说明模型精度较高,泛化能力较好。

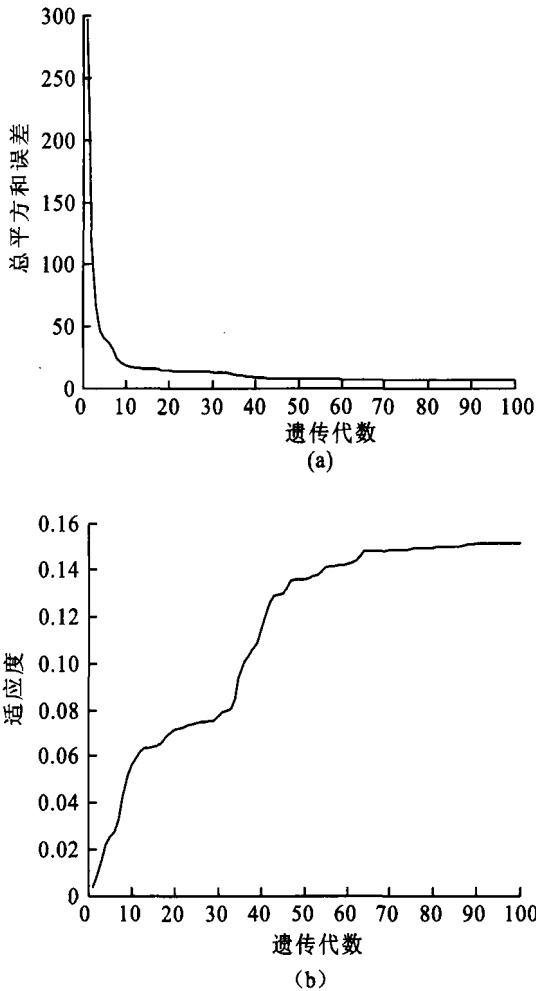


图 1 模型 I 的遗传算法训练曲线 (a) 误差平方和曲线, (b) 适应值曲线
Fig. 1 Training curve for genetic algorithm of model I ((a), Sum-Squared Error curve; (b) Fitness curve)

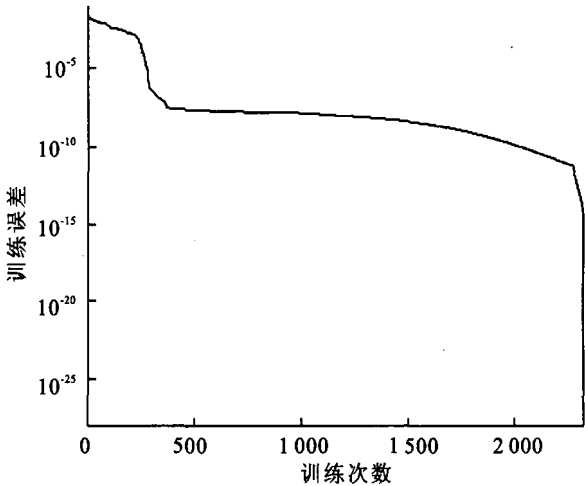


图 2 模型 I 的神经网络的训练曲线
Fig. 2 Training curve for neural network of model I

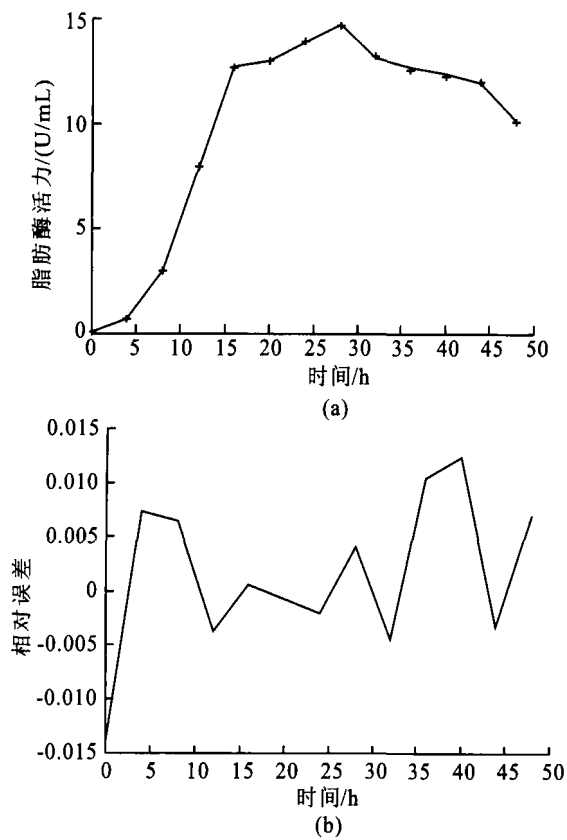


图 3 酶活力模型的测试曲线和相对误差曲线
Fig. 3 Test and relative error curve of enzyme activity model II ((a) Sum-Squared Error curve; (b) Fitness curve)

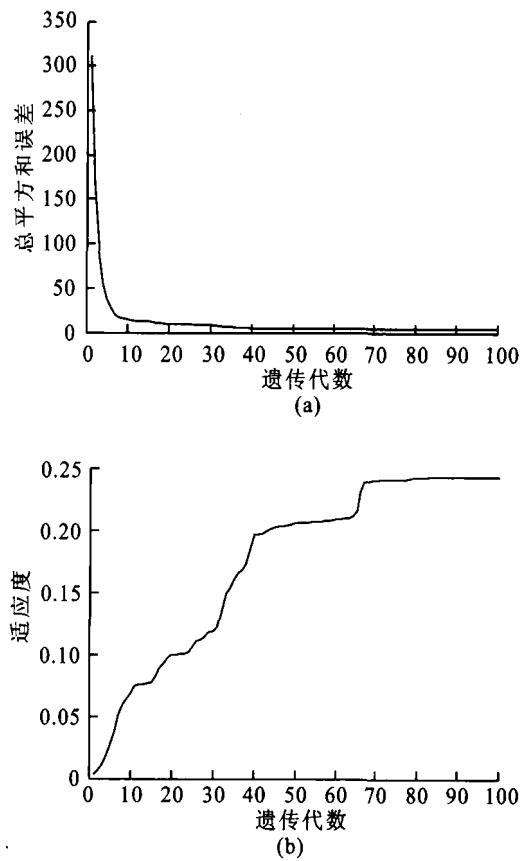


图 4 模型 II 的遗传算法训练曲线((a) 误差平方和曲线,(b) 适应值曲线)
Fig. 4 Training curve for genetic algorithm of model II

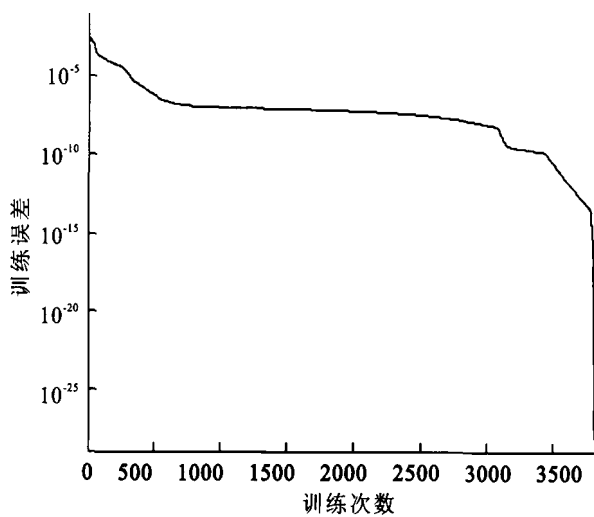


图 5 模型 II 的神经网络的训练曲线
Fig. 5 Training curve for neural network of mode II

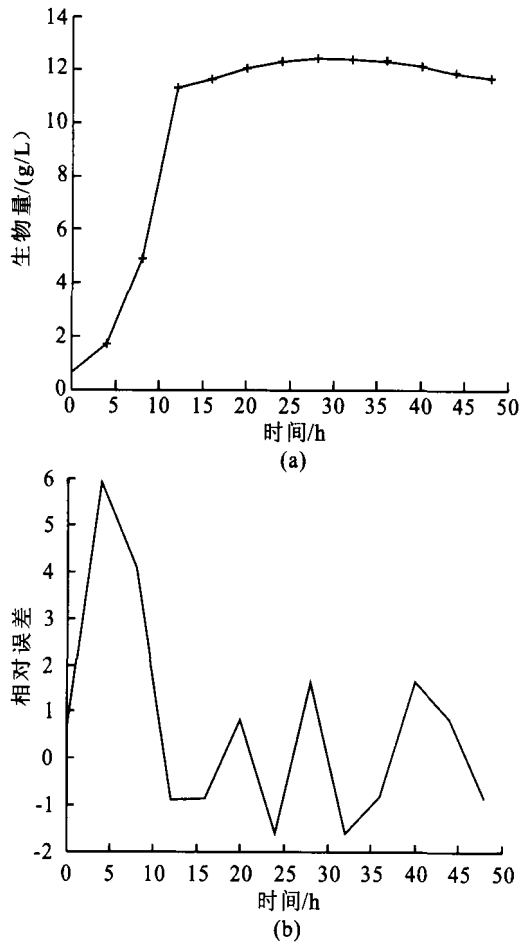


图 6 菌体生长模型的测试曲线和相对误差曲线
Fig. 6 Test and relative error curve of cell growth model

2.4 基质消耗的神经网络模型 III

将归一化处理好的 5 个输入变量的值作为 BP 网络的输入,目标向量是离线测量得到的残糖浓度的数据。采用模型 I 的方法确定隐层数 36,因此 BP 神经网络的结构为:5-36-1。采用模型 I 的方法用遗传算法工具箱优化神经网络的权值和阈值,并对网络进行训练和测试。遗传算法和神经网络训练网络的过程见图 7,8。由图 9 可知,模型 II 能较为精确的拟合输入的测试样本数据,其最大相对误

差不超过 0.3%,说明模型精度较高,泛化能力较好。

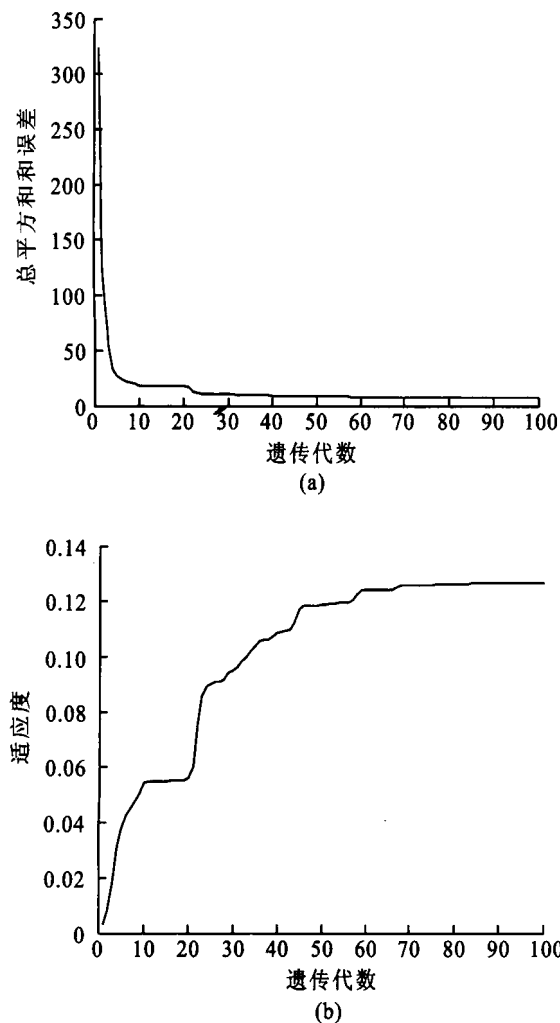


图 7 模型 III 的遗传算法训练曲线(A 误差平方和曲线,B 适应值曲线)
Fig. 7 Training curve of genetic algorithm of model III (A,Sum-Squared Error curve; B, Fitness curve)

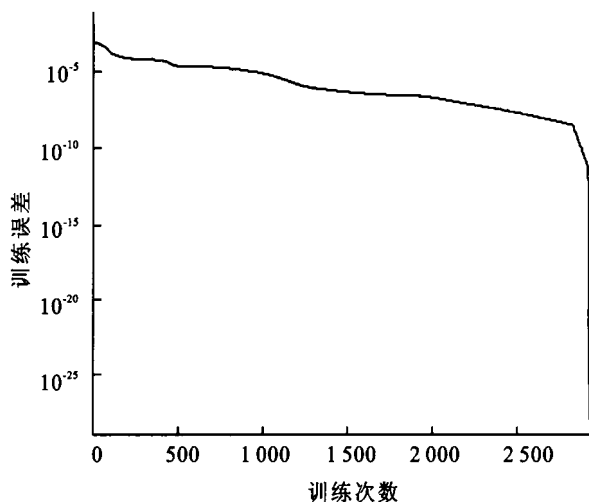


图 8 模型 III 的神经网络的训练曲线
Fig. 8 Training curve for neural network of model III

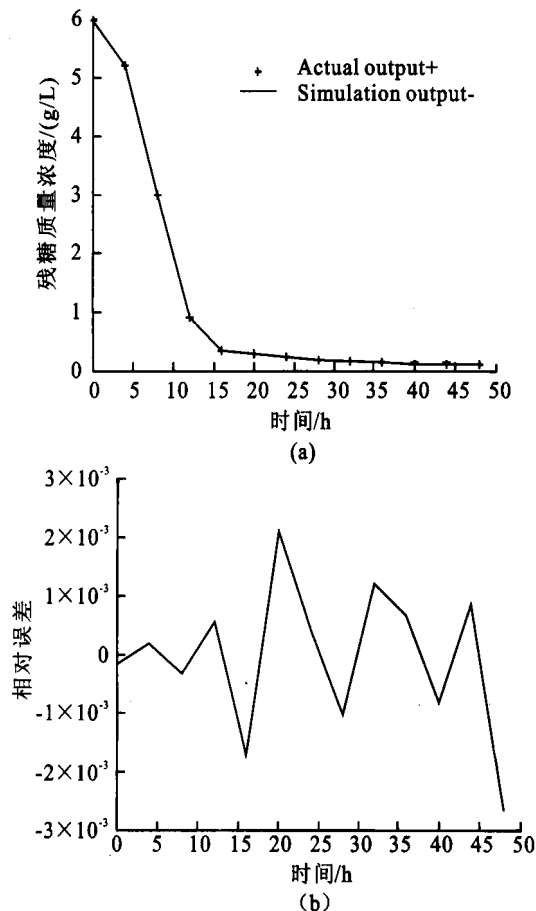


图 9 基质消耗模型的测试曲线和相对误差曲线
Fig. 9 Test and relative error curve of substance consuming model

3 结 语

综上所述,运用遗传算法和 BP 神经网络相结合的方法对洋葱伯克霍尔德菌脂肪酶的发酵过程进行软测量建模,增强了网络的学习和泛化能力,较好地解决了 BP 神经网络容易陷入局部最优,运算速度慢的缺点。所建立的模型能够比较精确地模拟发酵产酶、菌体生长以及底物消耗 3 个过程的动态变化,训练好的神经网络对测试样本进行拟合的精度较高,泛化能力较好,能够为洋葱伯克霍尔德菌脂肪酶的发酵生产过程提供动态模拟。而且,通过以上 3 个模型还能实时获得难以在线测量的关键参数,准确地预测发酵趋势,为有效监控发酵过程并进行优化操作奠定了技术基础。

但是,遗传算法和 BP 神经网络相结合的方法本身也存在一定的缺陷。BP 神经网络在确定隐含层神经元个数时,根据构造法去判别,具有较大的随机性,需要多次测试后凭经验判断,再通过实验验证,比较耗时费力;由于初始化的权值和阈值是随机的,BP 神经网络每次结果不一样,需要多次训练才有可能找到比较理想的结果;BP 神经网络和遗传算法中某些参数的选择也是按照经验进行的,

然后通过实验的方式不断改进,样本数量的多少和质量好坏将直接影响模型的预测效果。随着遗传算法和BP神经网络算法的不断完善和改进,相信

遗传算法和BP神经网络相结合的方法在微生物发酵的软测量建模方面将会得到越来越广泛的应用。

参考文献(References):

- [1] Jaeger K E, Eggert T. Lipases for biotechnology[J]. **Curr Opin Biotechnol**, 2002, 13(4): 390—397.
- [2] Hasan F, Shah A A, Hameed A. Industrial applications of microbial lipases[J]. **Enzyme and Microb Techn**, 2006, 39(2): 235—251.
- [3] 舒正玉, 杨江科, 黄瑛, 等. 生物柴油生产用脂肪酶资源及研发现状[J]. 湖北农业科学, 2007, 46(6): 1027—1031.
SHU Zheng-yu, YANG Jiang-ke, HUANG Ying, et al. Resources and current state of lipases used in biodiesel production [J]. **Hubei Agricultural Sciences**, 2007, 46(6): 1027—1031. (in Chinese)
- [4] Becker T, Enders T, Delgado A. Dynamic neural networks as a tool for the online optimization of industrial fermentation [J]. **Bioproc Biosyst Eng**, 2002, 24(6): 347—354.
- [5] 罗剑飞, 林伟铁. 基于神经网络和遗传算法培养基优化的发酵经济学[J]. 食品与生物技术学报, 2007, 28(3): 424—428.
LUO Jian-fei, LIN Wei-tie. The economics of fermentation for medium optimizing by neural networks coupling genetic algorithms[J]. **Journal of Food Science and Biotechnology**, 2007, 28(3): 424—428. (in Chinese)
- [6] Garcíá O F, Gómez C E. Estimation of oxygen mass transfer coefficient in stirred tank reactors using artificial neural networks[J]. **Enzyme and Microb Techn**, 2001, 28(6): 560—569.
- [7] 许哲军, 何云, 郭美锦, 等. 重组毕赤酵母表达期菌体浓度的软测量模型[J]. 食品与生物技术学报, 2006, 25(5): 28—34.
XU Zhe-jun, HE Yun, GUO Mei-jin, et al. Soft sensor modeling for predicting biomass concentration at recombinant pichia pastoris expression phase[J]. **Journal of Food Science and Biotechnology**, 2006, 25(5): 28—34. (in Chinese)
- [8] Sarkar D, Modak J M. Optimisation of fed-batch bioreactors using genetic algorithms[J]. **Chem Eng Sci**, 2003, 58(11): 2283—2296.
- [9] 王俊刚, 张树珍, 杨本鹏, 等. 3,5-二硝基水杨酸(DNS)法测定甘蔗茎节总糖和还原糖含量[J]. 甘蔗糖业, 2008(5): 45—49.
WANG Jun-gang, ZHANG Shu-zhen, YANG Ben-peng, et al. Application of 3,5-Dinitrosalicylic acid (DNS) method to test the reducing sugar and water-soluble total sugar content in sugarcane internodes[J]. **Sugarcane and Canesugar**, 2008 (5): 45—49. (in Chinese)
- [10] Eom G T, Song J K, Ahn J H, et al. Enhancement of the efficiency of secretion of heterologous lipase in Escherichia coli by directed evolution of the ABC transporter system[J]. **Appl Environ Microbiol**, 2005, 71(7): 3468—3474.

(责任编辑:李春丽)