

文章编号:1673-1689(2009)03-0424-05

基于神经网络和遗传算法培养基优化的发酵经济学

罗剑飞, 林炜铁*

(华南理工大学 生物科学与工程学院, 广东 广州 510006)

摘要: 利用目前较为先进的神经网络(ANN)和遗传算法结合(GA)的优化方法,通过以6种培养基组成为输入,亚硝酸盐氧化菌的活性与培养基成本之比的“性价比”为输出,构建结构为6-8-1 BP神经网络的非线性的非结构模型,并以该模型为遗传算法的目标函数,进行遗传算法的全局寻优,优化得到具有最高“性价比”的培养基组成。结果表明,基于神经网络和遗传算法能较好地对培养基进行优化,并实现对发酵经济学的初步研究。通过优化,得到亚硝酸盐氧化菌的最佳培养基组成为: NaNO_2 2.390 g/L, KH_2PO_4 1.355 g/L, MgSO_4 0.019 g/L, NaCl 0.031 g/L, NaHCO_3 4.373 g/L, FeSO_4 0.005 g/L,其最优“性价比”为8.705,比初始的6.835提高了27.36%。

关键词: 发酵经济学;BP神经网络;遗传算法;“性价比”;亚硝酸盐氧化菌

中图分类号:TQ 920.1

文献标识码:A

The Economics of Fermentation for Medium Optimizing by Neural Networks Coupling Genetic Algorithms

LUO Jian-fei, LIN Wei-tie*

(School of Biological Science and Engineering, South China University of Technology, Guangzhou 510006, China)

Abstract: The aim of study the economics of fermentation is to control the cost of products and increase the rate of product market share by reducing the cost and raising the performance—price ratio. In this study, the economics of fermentation medium was evaluated by applying of neural networks coupling with genetic algorithms. An optimal medium composition of nitrite oxidizing bacteria: NaNO_2 2.390 g/L, KH_2PO_4 1.355 g/L, MgSO_4 0.019 g/L, NaCl 0.031 g/L, NaHCO_3 4.373 g/L, FeSO_4 0.005 g/L, and the optimal performance price ratio were achieved at 8.705, higher 27.36% than that of the corresponding value of the initial (6.835) in a limited time and more less of experiments.

Key words: The economics of fermentation, BP neural networks, Genetic algorithms (GA), Performance-price ratio, Nitrite oxidizing bacteria (NOB)

发酵经济学是发酵工程与技术经济学交叉而成的学科,主要致力于选择适当的发酵原料、方法及工艺,降低发酵产品的成本^[1]。对产品成本的控制包括选育优良菌株、培养基成分分析和发酵方式

的选择等。在工业生产中,确定生产菌株后,培养基成为发酵产品特别是传统发酵产品生产成本的决定性环节,而培养基各个组分的成本对产品成本有很大影响,占38%~72%。因此,对于发酵经济

收稿日期:2008-06-12

* 通讯作者: 林炜铁(1964-),男,广东平远人,副教授,硕士生导师,主要从事环境微生物理论和应用等研究。

Email:13902202933@163.com

学所研究的重点来说,在最大限度地利用培养基生产产品的同时,还需要通过优化培养基成分来降低生产成本^[2-3]。为了提高产品的市场占有率,在降低生产成本的同时需要保证产品的性能质量,因此要解决好产品质量和生产成本之间的关系,提高产品的性价比,即提高产品性能质量与产品价格之比。

神经网络是一种“黑箱”模型,具有很强的非线性映射能力,可用于反映非线性条件下,难以用常规的数学模型描述的问题^[4-5]。遗传算法是以自然选择和遗传理论为基础,将生物进化过程中适者生存规则与群体内部染色体的随机信息交换机制相结合的高效全局寻优搜索算法^[6-7]。神经网络和遗传算法的结合可以通过引入非线性的模型来描述各因素间复杂的关系,并在遗传算法的基础上,通过全局寻优找出最佳值。国内外已经有不少学者通过神经网络和遗传算法来优化发酵培养基^[8-15]。

利用神经网络结合遗传算法的优化方法,优化亚硝酸盐氧化菌的培养基组成,以提高该菌的活性,并在遵循发酵经济学原理的基础上,通过提高亚硝酸盐氧化菌的活性与培养基成本之比的“性价比”,来研究发酵经济学。

1 材料与方法

1.1 菌种、培养基及培养条件

菌种为作者所在实验室分离得到的亚硝酸盐氧化菌(NOB)^[16]。培养基组成及其价格为:NaNO₂ 17元/kg, KH₂PO₄ 30元/kg, MgSO₄ 41元/kg, NaCl 15元/kg, NaHCO₃ 24元/kg, FeSO₄ 19元/kg。初始培养基质量浓度为 NaNO₂ 0.5 g/L, KH₂PO₄ 0.136 g/L, MgSO₄ 0.14 g/L, NaCl 0.3 g/L, NaHCO₃ 1.6 g/L, FeSO₄ 0.03 g/L。NaNO₂ 和 NaHCO₃ 分别是 NOB 的唯一氮源和碳源;并且 NaNO₂ 是 NOB 的唯一能量来源,它是通过将亚硝酸氮氧化为硝酸氮产生的 ATP 为 NOB 的生长和代谢提供能量。NOB 的培养是将种子液接种到 250 mL 的三角瓶中,在 30 ℃、180 r/min 摇床培养 7 d 后,再将其按 10% 的接种量(体积分数)接种到 15 L 的气升式反应器中分批补料培养。

1.2 实验分析方法

由于硝化细菌平板计数困难、代时长等特点^[17],因此对硝化细菌的研究采用在 600 nm 可见光下测定 A₆₀₀ 来对其计数,这是基于 A₆₀₀ 与混合液悬浮固体量(MLSS)的转换关系实现的。其反应式如下:

$$MLSS = 1.271 \times A_{600}$$

式中,1.271 即为实验得到的 A₆₀₀ 与 MLSS 线性关系的斜率。

NOB 对环境的贡献是通过对亚硝酸氮的降解能力来显现的,因此以亚硝酸盐氧化速率为对象研究 NOB 的活性。亚硝酸氮在 A_{520 nm} 的可见光下有最大吸收值,通过测定 520 nm 下的吸光值即可知道溶液中亚硝酸氮的量,其测定方法参考 GB 7493-87。亚硝酸氮氧化速率的计算如下:

$$NOR = \frac{(A_{520(0)} - A_{520(3)}) \times m}{MLSS \times 3} \times 24$$

式中,NOR 为亚硝酸氮氧化速率,单位为 g/(g·d) (g NO₂⁻ - N/g MLSS · d); A₅₂₀ 是在可见光 520 nm 下的吸光度,下标括号内为测定时间;m 是亚硝酸氮标准曲线斜率;3 意指反应时间为 3 h,24 为 1 d 的降解量。通过研究亚硝酸盐氧化菌的活性与培养基成本之比的“性价比”,来研究发酵经济学,“性价比”的计算公式为:

$$\text{“性价比”} = \frac{\text{亚硝酸氮氧化速率(NOR)}}{\text{培养基的总成本}}$$

1.3 实验设计方法

1.3.1 神经网络(ANN)建模 采用 MATLAB 7.4 进行神经网络的构建。由于 BP 神经网络结构简单,具有较强的非线性映射能力,是应用最为广泛的一类神经网络形式,因此这里采用 BP 网络进行建模。以培养基中的 6 种成分 NaNO₂、KH₂PO₄、MgSO₄、NaCl、NaHCO₃、FeSO₄ 的初始质量浓度为输入值,以“性价比”为输出值,通过选择输入层神经元数、隐含层神经元数和输出层神经元数,确定网络的结构,并对所确定的网络进行训练,根据 BP 网络模型的输出结果,检测网络的精度,并检测网络的泛化能力。

1.3.2 遗传算法(GA)寻优

GA 寻优的运算过程如下:

1) 编码:遗传算法在进行搜索之前先将解空间的解数据表示成遗传空间的基因型串结构数据,这些串结构数据的不同组合就构成不同的点;

2) 初始种群的生成:随机产生 N 个初始串结构数据,每个串结构数据称为一个个体,N 个个体构成一个种群;

3) 适应度函数值的评价检测:将产生的种群代入构建好的神经网络中,作为网络的输入,该输出即为目标函数值,然后把目标函数值映射为适应度值;

4) 新种群的产生:按照遗传策略,运用选择、交叉、变异算子作用于初始种群,并形成下一代种群;

5) 终止条件判断:判断种群性能是否满足,或完成最大迭代次数,若不满足,则转到步骤2),或修改遗传策略再回到步骤2);

6) 最优解的输出。

1.3.3 神经网络和遗传算法的结合 神经网络和遗传算法结合的方式见图1。图中,通过适应度函数设定培养基各组分的取值范围,由GA在这范围内随机产生48组实验组合。以40组实验结果为训练样本集训练神经网络模型,并以剩下的8组实验结果为测试样本集对训练好的模型进行测试,由此构建神经网络模型。并以该模型的输出为GA的目标函数,通过遗传算法的全局寻优,找到具有最高“性价比”的培养基组成。

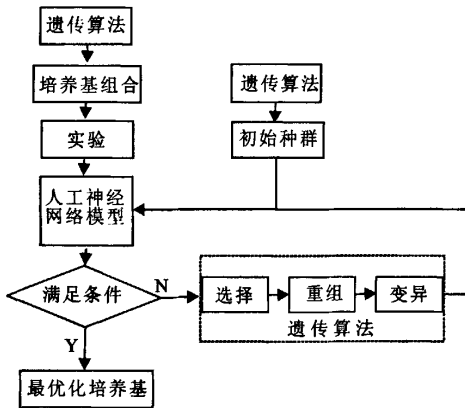


图1 ANN与GA结合的框图

Fig. 1 Schema of the working principle for ANN coupling GA

2 结果与讨论

2.1 BP神经网络模型

2.1.1 网络结构的确定 根据BP网络的设计,一般的非线性映射都可以通过单隐层的网络实现。由于输入向量有6个元素,所以网络输入层的神经元有6个,而输出向量只有1个,所以输出层的神经元数为1个。隐层节点数选择的恰当与否对网络的整个性能来说(如容错性、学习速率和泛化能力等)具有决定意义。一般的,隐层节点数太少,网络从样本中获取信息的能力就差,容错性也差,不能识别以前没有遇到的样本;如果隐层节点数太多,网络就有可能将样本的噪声也一并学习,造成学习的过度拟合,降低网络的泛化能力,同时也增加计算量。因此需要选择一个适合的隐层节点数。

从GA产生的48组实验数据中选取40组实验数据及其试验得到的“性价比”作为学习样本,以Sigmoid为网络传递函数、Levenberg-Marquardt

为训练函数、learnngdm为学习函数、MSE为性能函数,进行网络训练。从图2可以看出,当神经元数为8时,均方误差(MSE)接近0,因此选择隐含层的神经元数为8。这样,构建一个拓扑结构为6-8-1的BP神经网络模型,见图3。

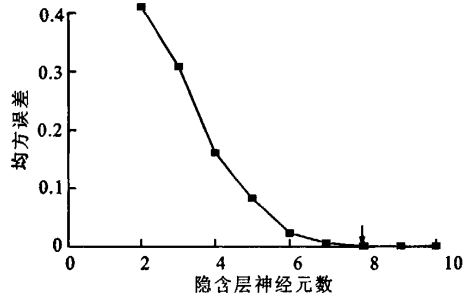


图2 神经网络隐含层神经元数不同取值时的均方误差

Fig. 2 Mean square error (MSE) of different number of neurons in the hidden layer

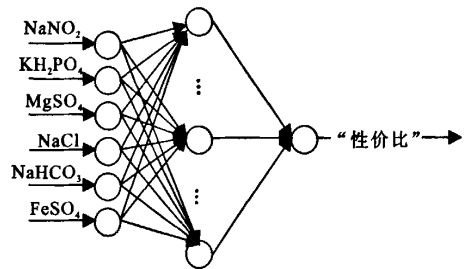


图3 BP神经网络结构

Fig. 3 Architecture of BP neural network

2.1.2 网络模型测试 将40组试验数据代入到构建好的神经网络中,其模拟值与试验值的拟合结果见图4,线性拟合度达0.9996。这说明该神经网络对网络输入和输出之间非线性拟合是精确的。

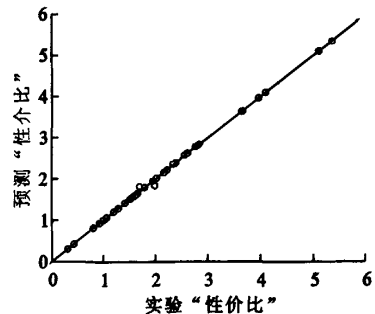


图4 实验得到的“性价比”和模型预测得到的“性价比”的比较

Fig. 4 Relationship of experimental performance-price ratio and performance-price ratio predicted by ANN model

从图4可见,网络模拟值和试验值接近1。这是因为测试样本来自训练样本,由于训练精度非常

高,网络模型对每组数据都高精度地拟合了。为了更准确地测试网络的性能,必须从训练样本以外的数据样本选取测试样本。因此,将 GA 产生的 48 组试验数据中没有进行训练的剩余 8 组试验数据作为测试样本,代入神经网络模型中进行测试,其测试值和试验值线性拟合结果见图 5: $P_{ANN-test} = 0.3765 + 0.8242P_{exp}$ ($R=0.9271$)。可见,两者具有较好的相关性。这说明构建的模型具有较好的泛化能力,可对试验结果进行很好的预测。

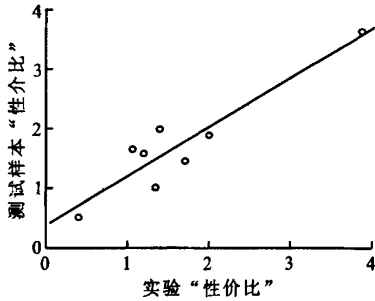


图 5 实验得到的“性价比”与测试样本的“性价比”的比较

Fig. 5 Relationship of experimental performance-price ratio and performance-price ratio tested by ANN model

2.2 GA 全局寻优设计

2.2.1 初始种群(培养基组成)的生成 为了提高寻优概率和保证一定的精度,取每个变量的子串长度 l 为 20,则 6 个变量组合形成的个体(染色体)长度 L 为 120。以表 1 为各变量的取值范围,随机产生的 48 个个体即为初始种群。

表 1 各变量参数的取值范围

Tab. 1 Value range of each variable

g/L					
NaNO ₂	KH ₂ PO ₄	MgSO ₄	NaCl	NaHCO ₃	FeSO ₄
0.05~0.055	0.0136~0.014	0.014~0.016	0.03~0.035	0.16~0.165	0.003~0.0035
5	1.36	1.4	3	16	0.3

2.2.2 目标函数的求解与适应度函数的转换 将 GA 随机产生的个体代入到构建好的神经网络模型中预测,ANN 输出的结果即为 GA 相应的目标函数值。由于目标函数可能有正有负,为了求解得到最大值,需要将目标函数与适应度函数进行变换。且 MATLAB 自带的 GA 工具箱为最小寻优,因此要对适应度函数进行线性变换,即:

$$f' = -f$$

f 、 f' 为转换前后的适应度函数。

2.2.3 遗传策略 通过常用的基于秩适应度计

万方数据

算,评价初始种群,并选择适应度高的个体遗传到下一代;将选择的个体进行重组,随机配对后,通过单点交叉(交叉概率 0.7),使优良的基因进行交流;最后通过变异运算产生新的基因型和种群。

2.3 ANN-GA 全局寻优

以神经网络的输出为 GA 的目标函数,通过 GA 全局寻优,优化得到了最优解为 8.834,其最佳的培养基组成为: NaNO₂ 2.390 g/L, KH₂PO₄ 1.355 g/L, MgSO₄ 0.019 g/L, NaCl 0.031 g/L, NaHCO₃ 4.373 g/L, FeSO₄ 0.005 g/L。GA 在寻优过程中解的变化和种群均值的变化见图 6。

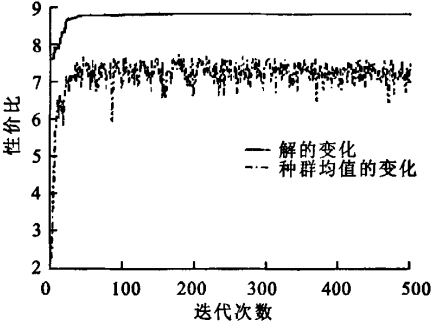


图 6 500 次迭代后最优解的变化和种群均值的变化 (MATLAB 生成)

Fig. 6 Curve of the optimal value and the mean of population after iterating 500 times (MATLAB)

2.4 优化结果的验证

将优化得到的培养基进行试验,得出 3 个平行值为 8.431、8.782、8.903 的“性价比”,对其进行单样本 t 检验,得出试验得到的结果与优化预测得到的结果没有显著差异,说明基于神经网络和遗传算法能较好地优化得到最优值。

3 结 语

结构为 6-8-1 的 BP 神经网络具有较强的学习能力和泛化能力,能够对实验结果进行较好地拟合,并构建得到一个模型。以该神经网络模型为目标函数,通过 GA 的全局寻优,优化得到最佳的培养基组成为 NaNO₂ 2.390 g/L, KH₂PO₄ 1.355 g/L, MgSO₄ 0.019 g/L, NaCl 0.031 g/L, NaHCO₃ 4.373 g/L, FeSO₄ 0.005 g/L,并通过试验得出最佳“性价比”为 8.705,比初始的 6.835 提高了 27.36%。

研究表明,神经网络具有高度的非线性映射能力,能较好地进行非线性拟合。遗传算法对于非线性问题的求解具有高效、鲁棒性强等特点。在对发酵经济学的研究中,基于神经网络和遗传算法的优

化方法,可以对发酵培养基成分进行较好地优化,通过减少原料成本和提高“性价比”,从而达到工业生产中对产品成本的控制。另外,神经网络和遗传

算法的结合也可以用于发酵条件优化,通过优化选择最佳的培养方式,以减少生产中的成本消耗。

参考文献(References):

- [1] 许喜林. 发酵经济学[J]. 现代食品科技, 1989(2): 1-2.
XU Xi-lin. Economics of the Fermentation[J]. *Modern Food Science and Technology*, 1989(2): 1-2. (in Chinese)
- [2] 姚汝华. 微生物工程工艺原理[M]. 广州: 华南理工大学出版社, 2005.
- [3] 任雁, 叶柯, 张惟广. 浅谈发酵经济学[J]. 四川食品与发酵, 2005, 41(3): 1-7.
REN Yan, YE Ge, ZHANG Wei-guang. Discussion on the economics of the fermentation[J]. *Sichuan Food and Fermentation*, 2005, 41(3): 1-7. (in Chinese)
- [4] Montague G, Morris J. Neural-network contributions in biotechnology[J]. *TIBTECH*, 1994(12): 312-24.
- [5] Franco-Lara E, Link H, Weuster-Botz D. Evaluation of artificial neural networks for modeling and optimization of medium composition with a genetic algorithm[J]. *Process Biochemistry*, 2006, 41: 2200-2206.
- [6] Desai K M, Akolkar S K, Badhe Y P, et al. Optimization of fermentation media for exopolysaccharide production from *Lactobacillus plantarum* using artificial intelligence-based techniques[J]. *Process Biochemistry*, 2006, 41: 1842-1848.
- [7] 雷英杰, 张善文, 李续武, 等. MATLAB 遗传算法工具箱及应用[M]. 西安: 电子科技大学出版社, 2005.
- [8] 方柏山, 陈宏文, 谢晓兰, 等. 基于神经网络和遗传算法的木糖醇发酵培养基优化研究[J]. 生物工程学报, 2000, 16(5): 648-650.
FANG Bai-shan, CHEN Hong-wen, XIE Xiao-lan, et al. The medium optimization of xylitol fermentation based on neural networks and genetic algorithms[J]. *Chinese Journal of Biotechnology*, 2000, 16(5): 648-650. (in Chinese)
- [9] 蔡宇杰, 诸葛斌, 张锡红, 等. 遗传算法与神经网络耦联法优化生淀粉酶发酵培养基[J]. 无锡轻工大学学报, 2001, 20(4): 421-423.
CAI Yu-jie, ZHUGE Bin, ZHANG Xi-hong, et al. Optimization of medium of raw starch enzyme fermentation with neural networks and genetic algorithms[J]. *Journal of Wuxi University of Light Industry*, 2001, 20(4): 421-423. (in Chinese)
- [10] 王军峰, 杜吉泉, 储炬, 等. 应用神经网络和遗传算法优化利福霉素 B 发酵培养基[J]. 中国抗生素杂志, 2006, 31(5): 278-280.
WANG Jun-feng, DU Ji-quan, CHU Ju, et al. Optimization of the fermentation medium for rifamycin B production by using artificial neural networks and genetic algorithms[J]. *Chinese Journal of Antibiotics*, 2006, 31(5): 278-280. (in Chinese)
- [11] 周景文, 徐建, 陈守文, 等. 基于径向基神经网络和遗传算法的聚- γ -谷氨酸发酵培养基优化[J]. 食品科学, 2006, 27(10): 288-292.
ZHOU Jing-wen, XU Jian, CHEN Shou-wen, et al. Optimization of Poly- γ -Glutamate fermentation medium based on radius basis function networks and genetic algorithm[J]. *Food Science*, 2006, 27(10): 288-292. (in Chinese)
- [12] 张慧, 王健, 陈宁. L-缬氨酸发酵培养基的神经网络建模与遗传算法优化[J]. 生物技术通讯, 2005, 16(2): 156-158.
ZHANG Hui, WANG Jian, CHEN Ning. Neural networks modeling and genetic algorithms for optimization fermentation medium of L-Valine[J]. *Letters in Biotechnology*, 2005, 16(2): 156-158. (in Chinese)
- [13] Fang B S, Chen H W, Xie X L, et al. Using genetic algorithms coupling neural networks in a study of xylitol production: medium optimization [J]. *Process biochemistry*, 2003, 38: 979-985.
- [14] Moreira G A, Micheloud G A, Beccaria A J, et al. Optimization of the *Bacillus thuringiensis* var. *kurstaki* HD-1 δ -endotoxins production by using experimental mixture design and artificial neural networks [J]. *Biochemical Engineering Journal*, 2007, 35: 48-55.
- [15] Desai K M, Vaidya B K, Singhal R S, et al. Use of an artificial neural network in modeling yeast biomass and yield of β -glucan[J]. *Process Biochemistry*, 2005, 40: 1617-1626.
- [16] 任杰, 林炜铁, 罗小春, 等. 硝化菌保藏特性及衰减动力学研究[J]. 中国生物工程杂志, 2007, 27(12): 61-65.
REN Jie, LIN Wei-tie, LUO Xiao-chun, et al. Study on the conservative property and decay kinetic of nitrite oxidizing bacteria[J]. *China Biotechnology*, 2007, 27(12): 61-65. (in Chinese)
- [17] Wilen B M, Jin B, Lant P. The influence of key chemical constituents in activated sludge on surface and flocculating properties[J]. *Water Research*, 2003, 37: 2127-2139.

(责任编辑: 秦和平)