

模式识别技术在生物杀虫剂苏云金杆菌生产中的应用

张 怡

(清华紫光集团总公司工程化学部)

聂峰光

(中国科学院化工冶金研究所生化工程国家重点实验室)

摘 要: 采用 BP 模型, 以苏云金杆菌固态发酵为模型对象, 研究了基于人工神经网络的模式识别方法在固态发酵工艺条件辨别中的应用。网络自学习结果表明, 发酵工艺条件分类全部正确, Cross Validation 方法考察网络预测能力也得到满意的结果。说明人工神经网络在生物发酵工程中有广泛的应用前景。

关键词: 模式识别; 人工神经网络; 苏云金杆菌; 固态发酵

生物反应系统属于自适应、高度复杂的非线性体系, 迄今为止, 对于生物反应过程深层次的了解一直是科学工作者追求的目标。尤其象固态发酵体系, 其在线监控手段少, 有关发酵机理、传质特性的研究报道均较少。在目前的固态发酵工艺优化研究中, 正交实验目前仍是重要的研究手段。利用回归统计的方法可使正交实验和其它条件实验数据信息得到较充分的利用, 建立的统计模型可预测发酵条件变化时的发酵情况。但回归统计法需要对过程的规律有一定的认识, 并要给定模型的解析表达式, 而且需要根据回归情况, 人为地调整解析表达形式, 直至符合数据的变化规律。对于非线性、复杂性高、监控手段少的固态发酵系统, 用这种方法需要作大量的工作, 而且常常很难给出一个较好的解析表达形式。因此, 运用现代化手段, 进行有效的数据处理和信息收集是十分必要的。

神经网络(ANN)是非线性并行分布处理方法^[1], 它能够联想、概括而不需要规则, 具有模拟人的部分形象思维的能力, 特别适合于联想、估价及模式识别, 已用于计算机视觉、语音识别、天气预报、股市分析及医学诊断等领域。在化学化工中, 神经网络也得到不少成功的应用, 如糖类化合物氢核磁共振谱图和质谱图的识别^[2,3]、结构-活性关系的研究及化工故障诊断等^[4]。本文选择生物农药苏云金杆菌的固态发酵体系作为模型对象, 讨论模式识别方式在生物发酵领域中的应用。

收稿日期: 2000/04/20

张 怡, 工程师, 北京市清华紫光大楼 420 室 清华紫光集团总公司工程化学部, 100084 Email: zhangyi@chn@yahoo.com

1 模式识别方法

1.1 模型对象的选择

菌种: *Bacillus thuringiensis* var. HD21, (简称 Bt), 由中国科学院动物所保存。发酵条件: 固态发酵, 条件参见文献[5]。芽孢数测定: CFU 平板活菌计数法。

1.2 工艺特征选取和分类

选取发酵工艺的 5 个关键条件作为工艺特征, 碳源、氮源、pH 调控因子、含水率及接种量。发酵水平可由产品的活芽孢数表征, 因此以单位产品的活芽孢数由小到大依次将发酵情况分为 4 种模式:

芽孢数 < 100 亿ög; 100 亿ög < 芽孢数 < 150 亿ög; 150 亿ög < 芽孢数 < 200 亿ög; 芽孢数 > 200 亿ög。

1.3 网络结构

针对生化过程变量连续性的特点, 本文选择一个隐含层的 BP (Back Propagation) 网来建立发酵质量模型, 其输入层 5 个节点, 对应发酵工艺的 5 个关键条件, 输出层 4 个节点, 对应不同水平的发酵情况(如图 1), 上述 4 种模式, 网络输出期望值依次为 (1 0 0 0)、(0 1 0 0)、(0 0 1 0) 和 (0 0 0 1)。

神经元输出函数为 Sigmoid 函数

$$f(E_i) = y_i = \frac{1}{1 + e^{-E_i}}$$

式中 i ——节点, 可处于隐含层或输出层; E_i ——节点的总输入值; y_i ——节点 i 的输出值; $f(E_i)$ ——Sigmoid 函数。

1.4 网络的自学习及检验

为使各样本输入参数所起的作用大致均衡, 以

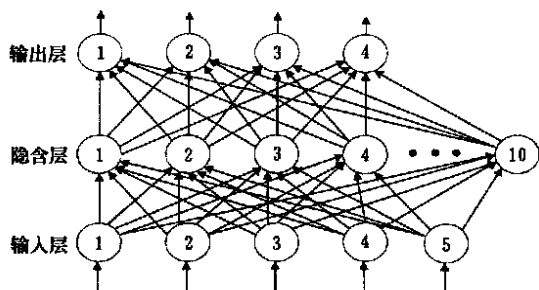


图 1 BP 神经元网络拓扑结构示意图

Fig 1 The topological structure of BP algorithm

防过分夸大或减小某些因素的作用,对输入样本参数归一化。大量计算表明,输入样本归一化到区间 [0, 1, 0.9] 较为合适。输入样本参数按下式归一化处理

$$Z_j = \frac{0.8}{Z_{max} - Z_{min}} Z_j + \frac{0.1Z_{max} - 0.9Z_{min}}{Z_{max} - Z_{min}}$$

式中 Z_j ——输入层第 j 个输入样本值; Z_{max} , Z_{min} ——输入层最大值和最小值。

本文采用改进的 GRD 算法^[6],节点 i 处所有的输入 x_j 首先被加权表示为

$$E_i = W_{ij}x_j + H$$

式中 j ——节点,可处于输入层或隐含层; x_j ——节点 i 处输入值; W_{ij} ——从节点 j 到节点 i 的权重;

H ——节点 i 的阈值。

为清楚地了解神经网络的性质及其分类、预测能力,采用 Cross Validation 方法评估 ANN 的预测能力。即从固定样本中随机删除几个样本,用剩下的样本建立网络模型,反过来预报被删除的样本,重复该过程直至所有的样本都被删除一次且只被删除一次。本文用选择的 42 个样本对上述网络进行训练,通过训练,调整各节点之间的权重 W_{ij} 和阈值 H ,使得实际网络输出不断逼近期望输出。当误差函数值小于 0.002,认为训练完毕,自学习过程结束。

2 结果与讨论

2.1 神经网络的分类和预测能力

经大量计算,选定的 10 个隐节点的网络对训练集的分类全部正确,这说明神经网络经过自适应、自组织的学习过程,可将有关的知识(分类信息)完全掌握。

在采用 Cross Validation 方法对网络预测未知样本的能力进行检验时,42 个样本依次作为被预测的样本,排除在训练集之外,其余 41 个作为训练集训练网络,然后用所得到的网络参数对被预测的样本进行预报,这样循环 42 次,其结果见表 1。

表 1 Cross Validation 检验预测结果

Tab 1 The predicted and verified results by Cross Validation

样本	发酵条件 ³					芽孢数	分类		输出值			
	1	2	3	4	5		希望值	计算值	1	2	3	4
1	50	15	0	60	40	45.5	1	1	0.9996	0.0000	0.0818	0.0000
2	50	15	10	60	40	166.3	3	3	0.0000	0.0011	0.9985	0.0001
3	50	15	5	60	40	191.3	3	3	0.0001	0.0000	1.0000	0.0000
4	50	25	0	60	40	74.6	1	1	0.9866	0.0000	0.5459	0.0000
5	50	25	10	60	40	141.1	2	2	0.0000	0.5200	0.4507	0.0000
6	50	25	11	60	40	141.1	2	2	0.0000	0.9994	0.0028	0.0000
7	50	25	13	60	40	132	2	2	0.0000	1.0000	0.0002	0.0000
8	50	25	3	60	40	195	3	4	0.3851	0.0000	0.0000	0.9295
9	50	25	5	51.4	40	211.7	4	4	0.0005	0.0204	0.0000	0.9998
10	50	25	5	54.8	40	218.1	4	4	0.0004	0.0029	0.0005	0.9993
11	50	25	5	57	40	250	4	4	0.0004	0.0008	0.0020	0.9979
12	50	25	5	59.9	40	229	4	4	0.0000	0.0000	0.3316	0.9602
13	50	25	5	60	20	183	3	3	0.0000	0.0000	0.9999	0.0000
14	50	25	5	60	30	198	3	3	0.0000	0.0000	0.6515	0.4834
15	50	25	5	60	50	209	4	4	0.0053	0.0000	0.0545	0.9695
16	50	25	5	60	80	194	3	4	0.0000	0.0000	0.0000	1.0000
17	50	25	5	66.0	40	185	3	3	0.0007	0.0000	1.0000	0.0000
18	50	25	5	67.4	40	174.2	3	3	0.0006	0.0000	1.0000	0.0000
19	50	25	7	60	40	189	3	3	0.0000	0.0000	0.9840	0.0196
20	50	25	9	60	40	159	3	2	0.0000	0.8894	0.2029	0.0006
21	50	35	10	60	40	147.1	2	2	0.0000	0.9999	0.0002	0.0047
22	50	35	5	60	40	231.7	4	2	0.0013	0.0188	0.0032	0.0012
23	60	15	0	60	40	45.8	1	1	0.9994	0.0000	0.2371	0.0002
24	60	15	10	60	40	133.0	2	3	0.0000	0.0002	0.9852	0.0000

样本	发酵条件 ³					芽孢数	分类		输出值			
	1	2	3	4	5		希望值	计算值	1	2	3	4
	25	60	15	5	60		40	168.3	3	3	0.0096	0.0000
26	60	25	0	60	40	65.9	1	1	0.9999	0.0000	0.0073	0.0000
27	60	25	10	60	40	119	2	2	0.0000	0.9935	0.0627	0.0002
28	60	25	5	60	40	238.3	4	3	0.0003	0.0000	0.9114	0.1772
29	60	35	0	60	40	45.0	1	1	0.9944	0.0000	0.0000	0.3677
30	60	35	10	60	40	112.0	2	2	0.0000	1.0000	0.0000	0.0093
31	60	35	5	60	40	129.2	2	2	0.0000	0.9981	0.0000	0.7949
32	70	15	0	60	40	68.8	1	1	1.0000	0.0000	0.0002	0.0000
33	70	15	10	60	40	167.1	3	3	0.0000	0.0059	0.9767	0.0000
34	70	15	5	60	40	200.4	4	4	0.0000	0.0004	0.0000	0.9772
35	70	20	5	60	40	254.5	4	4	0.0031	0.0000	0.0000	0.9998
36	70	25	5	60	40	214	4	4	0.0016	0.0238	0.0000	0.9913
37	70	25	10	60	40	151.7	3	2	0.0000	0.9928	0.0008	0.0004
38	70	35	0	60	40	48.0	1	1	0.9947	0.0000	0.0000	0.0030
39	70	35	10	60	40	106.1	2	2	0.0000	0.9979	0.0089	0.0000
40	70	35	5	60	40	138.3	2	2	0.0002	0.0294	0.0000	0.0048
41	75	20	5	60	40	213.5	4	4	0.0078	0.0002	0.0000	0.9998
42	75	25	5	60	40	211.2	4	4	0.0269	0.0030	0.0000	0.9902

3 发酵条件: 1 为碳源的用量(以麸皮计), g; 2 为氮源的用量(以棉籽饼粉计), g; 3 为 pH 调节剂的用量(以草木灰计), g; 4 为初始含水率, %; 5 为接种量, %。

由表 1 知, 共有 7 个样本分类错误。分析这 7 个样本的特点, 其中 8、16、20 和 37 号都被误判为相邻类别, 这是因为实验结果本身就是处于分类的边界, 因此被判断为相邻类别概率较大。24、28 号样本也被判断为相邻类别。22 号样本判断的误差较大, 这说明该样本或者在预测集的分类中具有决定作用, 或者可能是样本本身的实验误差较大, 造成判断结果的差距。

2.2 隐含节点数的影响

神经网络的隐含层节点数对网络性能有很大的影响, 但目前尚无成熟的理论可供指导。

计算表明, 隐节点数小于或等于 5, 网络无法实现对原始训练集全部正确分类; 隐节点数为 6 或 7 时, 网络对原始训练集分类全部正确, 但对个别 Cross Validation 检验训练集分类有误; 隐节点数在 8 以上, 网络对原始训练集和 Cross Validation 检验训练集全部正确分类。

隐节点数对网络预测能力的影响见表 2。增加隐节点数可提高网络分类的准确率, 当节点数增加至 10 时, 网络的行为达到最佳。再增加隐节点数, 网络的预测能力没有改善, 这说明体系提供的信息已被充分利用, 再增加隐节点数只会使计算费用增加。

表 2 隐含层节点数与准确率的关系

Tab 2 The relationship between the number of internal nodes and the rate of accuracy

节点数	7	8	9	10	15	18
准确率/%	64.3	69.1	81.0	83.3	83.3	83.3

3 结 语

人工神经网络能很好地用于 Bt 固态发酵工艺条件分类, 并能给出满意的工艺预报结果, 选择合适的隐节点数是网络成功地分类和预报的关键。神经网络具有高度的非线性处理能力, 能够预料, 在生物反应过程中将会有广阔的应用前景。对模型对象 Bt 的固态发酵过程模式识别描述, 为神经网络模式识别技术在淀粉酶、单细胞蛋白、酿造行业等监控、描述困难的固态发酵体系中的推广应用提供了值得借鉴的依据。

【参 考 文 献】

- [1] 焦季成著. 神经网络系统理论. 西安: 西安电子科技大学出版社, 1990
- [2] Thom sen J U, Meyer B. Pattern recognition of H NMR spectra of sugar alditol using a neural network. J Magn Reson, 1989, 84: 212~ 217
- [3] Sellers J, York W, Albersheim P, et al. Identification of mass spectra of partially methylated alditol acetates by artificial neural networks. Carbohydrate Research, 1990, 207: c1~ c5
- [4] Venkatasubramanian V, Vaidyanathan R, Yamamoto Y. Process fault detection and diagnosis using neural networks—I. Steady-state processes. Computers Chem Eng, 1990, 14: 699~ 712
- [5] Burges H D and Hussey N W. Microbial Control of insects and mites. San Diego: Academic Press, 1971
- [6] Leonard J, Kramer M A. Improvement of the back2propagation algorithm for training neural networks. Computers Chem Eng, 1990, 14: 337~ 341

Application of Neural Network Pattern Recognition to Discriminating the Production Conditions of *Bacillus Thuringiensis*

Zhang Yi

(Tsinghua Unisplendour Group, Dept of Chemical Engineering, Beijing 100084)

Nie Fengguang

(Institute of Chemical Metallurgy, Chinese Academy of Sciences)

Abstract: Artificial neural network pattern recognition was applied to discriminating the solid state fermentation conditions of *Bacillus thuringiensis*. The neural network was constitute of three layers and the back propagation algorithm was used. To evaluate the performance of the networks, the Cross Validation strategy was employed and satisfactory results was obtained. It was showed that neural network pattern recognition approach was quite promising in the optimization of fermentation conditions.

Key words: pattern recognition; artificial neural network; *Bacillus thuringiensis*; solid state fermentation

《农业工程学报》2000 年所刊论文即将进入 CSA 检索系统(附件 2)

Dear Dr Wang Yingkuan,

I wanted to let you know that we have received issues 1- 5 of your Transactions. **We are currently processing them and they should appear in our databases very shortly.** We look

forward to the opportunity to receive and review future issues as well! Thank you for your contributions and participation with CSA.

Sincerely,

Karen A. Skiba

(王应宽)

《农业工程学报》被 CSA 收录的有关细节(附件 3)

Dear Wang Yingkuan,

Thank you for your reply and your interest in our abstracts services. To be included in our abstracts databases, all we ask of you is that you supply to us free of charge your publication(s) on a regular basis. Our staff does the rest. There is no cost or fee to you. Our editors will browse your publications and select articles, which will be of interest to our customers. These articles will be abstracted for inclusion in our databases. Articles need not be in English, as we do have language translators to write or translate abstracts from other languages into English.

CSA does not require any formal written agreement, nor do we issue certificates. If you would like to 'advertise' the fact that your

publications are referenced in our databases, you may certainly include this information in your journal. You may send your journals or publications either way you prefer, whichever is more convenient or cost efficient. Although, the sooner we receive your publications, the sooner the abstracts will appear in our databases.

So, in summation, all we ask of you is that you send your publications to us on a complimentary basis; we'll do the rest. I hope that I have successfully answered all of your questions. Please feel free to contact me if you need any further information. CSA looks forward to the opportunity to receive and review your fine publication.

Sincerely,

Karen A. Skiba

(王应宽)