

苹果渣多酚提取工艺的神经网络建模与遗传算法优化

赵武奇, 仇农学, 王 宏
(陕西师范大学食品工程系, 陕西 西安 710062)

摘 要: 本实验建立了苹果渣多酚提取工艺的人工神经网络模型, 研究了提取工艺神经网络模型的遗传算法优化技术。结果表明, 结构为 4-9-1 的神经网络能较为精确的拟合输入的样本数据, 其对测试样本的输出与实验结果的相关系数为 0.985; 遗传算法优化出的最佳提取工艺参数为乙醇浓度为 62%、乙醇溶液的体积与苹果渣的质量之比为 14:1, 温度为 69.7℃, 提取时间为 5.9h, 该工艺参数下的提取率明显大于单因素试验和二次组合试验的结果, 比最好的大 16.9%。用神经网络模型描述提取工艺参数与提取率之间的关系, 用遗传算法优化工艺参数, 能设计出最佳的提取工艺参数。

关键词: 神经网络; 遗传算法; 苹果渣多酚; 优化; 模型

Modeling and Optimization of Polyphenol Extraction Process from Apple Pomace Based on Artificial Neural Network and Genetic Algorithm

ZHAO Wu-qi, QIU Nong-xue, WANG Hong
(Department of Food Engineering, Shaanxi Normal University, Xi'an 710062, China)

Abstract: The artificial neural network (ANN) model for polyphenol extraction from apple pomace was studied and ANN model was optimized by using genetic algorithm (GA) in this paper. Results showed that the structure of network is of four neurons of input layer with one neuron of output layer, one hidden layer and nine neurons. The trained network has a high generalization and the correlation coefficient between simulating outputs of the BP network and the experiments result is 0.985. The optimum

收稿日期: 2006-05-19

基金项目: 国家“十五”重大专项(2001BA501A21)

作者简介: 赵武奇(1965-), 男, 副教授, 博士, 研究方向为食品加工新技术。

G200 凝胶柱进一步分离, 自动收集馏分, 逐管检测, 结果如图 1 所示。从图 1 可以看出, 存在两个大的洗脱尖峰, 且这两个峰的 A_{280nm} 、 A_{620nm} 的最大值恰好重叠, 表明这两个组份可能是蛋白聚糖。分别命名为 F1、F2。将 F1、F2 组份收集, 冷冻干燥, F1、F2 的分子量有待进一步深入研究。

2.4 硫酸根的测定

经柱层析后的羊栖菜多糖, 测得 F1 的 SO_4^{2-} 含量为 8.95%, F2 的 SO_4^{2-} 含量为 3.92%。说明羊栖菜多糖为含硫酸基的蛋白多糖。

2.5 纯度检查

聚丙烯酰胺凝胶电泳检测结果显示, F1 和 F2 均为单一的紫红色条带, 表明 F1、F2 为单一多糖。

参考文献:

- [1] 许忠能, 王朝晖, 孙立. 羊栖菜药用价值的研究进展[J]. 中草药, 2000, 31(11): 876-878.
- [2] 阮积惠. 羊栖菜的药用功能研究现状[J]. 中国野生植物资源, 2001, 20(6): 8-10.
- [3] 张惟杰. 糖复合物生化研究技术[M]. 2版. 杭州: 浙江大学出版社, 1999: 80.
- [4] 周鹏, 谢明勇, 傅博强, 等. 茶叶粗多糖的提取及纯化研究[J]. 食品科学, 2001, 22(11): 46-47.
- [5] 陈绍瑗, 莫为民, 潘远江, 等. 海洋药物研究(III)—羊栖菜多糖[J]. 兰州大学学报, 1998, 34(4): 110-113.
- [6] 过非, 许时婴. 羊栖菜粗多糖的提取研究[J]. 食品工业, 2002(3): 11-13.
- [7] 郑小乐, 曾玉旻. 羊栖菜澄清汁饮料[J]. 食品与机械, 2001(6): 17-18.
- [8] 周峙苗, 邹玲玲, 李国兴. 羊栖菜多糖饮料的试制[J]. 食品工业科技, 2001, 22(3): 60-62.

performance, 16.9% greater than the greatest value in experiments, is obtained when alcohol concentration is 62% while ratio of alcohol solution volume to apple pomace weight is 14:1, temperature was 69.7 °C, and treatment time 5.9 hours. The optimal process parameters can be obtained with ANN model to describe the relationship between process parameters and target and GA so as to optimize process parameters.

Key words artificial neural network; genetic algorithm; polyphenols from apple pomace; optimization; model

中图分类号: R943.41

文献标识码: A

文章编号: 1002-6630(2007)12-0138-05

苹果渣是果汁厂榨汁后产生的下脚料。苹果渣中含有多酚类化合物^[1], 药理研究表明, 苹果多酚具有预防龋齿、高血压、过敏反应, 抗肿瘤、突变, 阻碍紫外线吸收等多种保健功能; 同时还具有抗氧化、消臭、保鲜、护色、防止维生素损失等作用^[2]。因此从果渣中提取多酚对于提高资源的利用率具有重要的作用。提取的工艺条件决定提取率的大小, 为了获得最大的提取率, 需要设法找出使提取率尽可能高的提取工艺条件, 而提取率与提取工艺条件之间是一种多输入的复杂非线性关系。

人工智能理论的发展为解决非线性系统的建模与优化问题提供了新的思路。其中人工神经网络(ANN)技术以其具有辨识和逼近任意复杂非线性系统的能力在建模与控制领域得到广泛应用。Mittal^[3]用ANN简单、方便、准确地预测了食品的冷冻时间, 预测了片状食物油炸过程的传热和传质过程。Boilereaux等^[4]采用滑动窗ANN方法准确地在线控制了食品的解冻过程。蔡宇杰^[5]等人利用两层神经网络优化了生淀粉酶的发酵培养基。方柏山^[6]等人利用两层神经网络得到了木糖醇的最佳发酵培养基配方。高海燕^[7]等人利用神经网络准确地预测了异亮氨酸的发酵趋势。黄明志^[8]探索了动态BP网络和RBF网络在红霉素发酵过程状态预估中的应用。蒋益虹等^[9]利用神经网络获得了红曲杨梅果酒糖化和发酵的最优工艺。赵丽丽等^[10]运用BP神经网络对L-缬氨酸的发酵过程进行模拟。银建中等^[11]用人工神经网络技术精确的对超临界萃取过程进行模拟和预测。

遗传算法是基于自然选择和群体遗传机理的随机优化算法, 是一种适用于复杂形态函数的全局寻优方法。张宇飞^[12]建立了中药滴丸制剂过程工艺参数与成品率的神经网络模型, 并利用遗传算法对模型输入参数空间进行寻优, 得出了成品率最高时所对应的工艺参数。

因此, 本实验在用乙醇从苹果渣中提取多酚实验的基础上, 应用人工神经网络技术, 建立提取工艺条件与提取率函数关系, 并在此基础上利用遗传算法进行全局寻优, 从而获得最优的提取工艺条件。

1 材料与方法

1.1 材料、试剂与仪器

红富士苹果渣(含苹果籽)由陕西恒兴果汁饮料有限公司提供, 粉碎, 过60目筛。

单宁酸、钨酸钠、钼酸钠、硫酸锂、无水乙醇等均为国产分析纯。

AB204-N型电子天平 德国METTLER公司;
WFJ2000型分光光度计 尤尼柯(上海)有限公司;
SHZ-82水浴恒温振荡器 江苏丹阳门科教仪器厂;
SZ-93型自动双重纯水蒸馏器 上海亚荣生化仪器厂。

1.2 苹果渣多酚的提取试验

分别进行乙醇浓度(A)、乙醇溶液的体积与苹果渣的质量之比(B)、温度(C)及提取时间(D)的单因素试验及二次旋转组合试验, 目标为提取率(E)。表1为二次旋

表1 试验因子水平编码表
Table 1 Factors and levels of orthogonal test

水平	A 乙醇浓度(%)	B 乙醇:苹果渣(V/W)	C 温度(°C)	D 时间(h)
+r	80	14	70	6
+1	70	12	60	5
0	60	10	50	4
-1	50	8	40	3
-r	40	6	30	2

转组合试验因子水平编码表。

提取过程为取苹果渣5.000g, 以一定浓度、一定体积的乙醇溶液恒温振荡提取, 提取液经300目滤布过滤, 40℃下真空旋转蒸发回收乙醇, 剩余液3500r/min离心10min, 取上清液定容至100ml, 取1ml以FC法进行总酚测定^[11]。

1.3 建模方法

以乙醇浓度、料液比、提取时间及温度为神经网络四个输入单元。提取率作为待建立网络的输出单元。由于二次旋转组合设计具有均衡搭配性和整齐可比性, 试验点分布合理, 试验次数少, 样本具有广泛的代表性, 因此, 神经网络的训练数据来自二次旋转组合设计试验, 部分单因素试验数据为训练测试样本, 其余单因素试验数据为测试样本, 采用尺度变换法将输入输出数据转换在-1~1范围内, 隐含层采用双曲正切S型变换函数, 输出层采用纯线性变换函数, 用提前停止法训练网络, 以确定网络结构。经过验证, 确定神经网络模型。

1.4 优化方法

用长度为8位的二进制码串分别表示乙醇浓度、料液比、提取时间及温度四个变量。将这4个8位二进制

表2 转化后网络模型训练数据
Table 2 Training data of network model after change

组号	A	B	C	D	E	组号	A	B	C	D	E
1	0.5	0.5	0.5	0.5	0.23	14	-0.5	-0.5	0.5	-0.5	0.088
2	0.5	0.5	0.5	-0.5	0.164	15	-0.5	-0.5	-0.5	0.5	0.014
3	0.5	0.5	-0.5	0.5	-0.008	16	-0.5	-0.5	-0.5	-0.5	-0.07
4	0.5	0.5	-0.5	-0.5	-0.062	17	-1	0	0	0	-0.086
5	0.5	-0.5	0.5	0.5	0.132	18	1	0	0	0	-0.18
6	0.5	-0.5	0.5	-0.5	0.064	19	0	-1	0	0	-0.024
7	0.5	-0.5	-0.5	0.5	-0.048	20	0	1	0	0	0.194
8	0.5	-0.5	-0.5	-0.5	-0.094	21	0	0	-1	0	-0.058
9	-0.5	0.5	0.5	0.5	0.262	22	0	0	1	0	0.266
10	-0.5	0.5	0.5	-0.5	0.198	23	0	0	0	-1	-0.02
11	-0.5	0.5	-0.5	0.5	0.152	24	0	0	0	1	0.194
12	-0.5	0.5	-0.5	-0.5	0.096	25	0	0	0	0	0.076
13	-0.5	-0.5	0.5	0.5	0.172						

码串连接在一起,组成一个32位长的二进制编码串,这样就构成了优化问题的染色体编码。随机产生 100×32 位编码字符,组成100个初始个体,构成初始群体。采用赌盘选择机制从群体中选择性能优良的个体作为父代进行繁殖。随机配对后,通过单点式杂交(交叉概率 $p_c=0.9$)交流各自的优秀基因及单点变异(变异概率 $p_m=0.001$)。产生新的基因型和种群。对新种群进行评价,判断其是否满足算法停止准则,若不满足,则继续遗传、变异操作,直到满足为止。

1.5 程序设计

MATLAB 软件包是近年来推出的大量用于工程与数值计算的优秀软件。它具有快速的矩阵运算能力,它的主包和工具箱含有很多函数,其中包括神经网络工具箱,能极大地减少用户的工作量。因此,本实验中的编程工作均基于MATLAB6.0版本。

2 结果与分析

2.1 提取工艺参数与提取率的关系

二次旋转组合试验的数据经过尺度变换后得出的网络训练数据见表2。

2.2 网络训练

理论已经证明,三层BP网络可以逼近任何有理函数,而且已成功应用于多个领域。但是在本实验中,由于实验样本数据存在误差,在网络训练过程中很容易发生过度拟合现象。表面上看,训练好的网络与训练样本拟合非常好,其实并不表示真实的映射关系,用未经训练的新样本对网络模型进行检验和评价时,往往发现有较大误差。为了解决过度拟合问题,本实验采用了提前停止法训练网络。其基本思想为,将所有样本数据划分为训练样本集、验证集和测试样本集,训练集用于对神经网络进行训练,验证集用于在神经网络训练的同时监控网络的训练过程。在训练初始阶段,验

证集形成的验证误差通常会随着网络训练误差的减小而减小,但是当网络开始进入“过度训练”时,验证误差就会逐渐增大,当验证误差增大到一定程度时,网络训练会提前停止,这时训练函数会返回验证误差取最小值时的网络对象。图1是本实验网络训练的样本误差曲线。训练样本的误差随着学习过程的进行迅速下降,收敛很快。验证和测试样本的误差在学习的初始阶段与训练样本相似,迅速下降,这表示BP网络在整体上迅速逼近真实的输入输出映射关系。但是验证和测试样本的误差很快就达到最低点而不再下降,并有逐步上升的趋势,这表示网络开始出现过度拟合,此时程序自动停止学习。可见提前停止法能有效防止网络过度拟合。

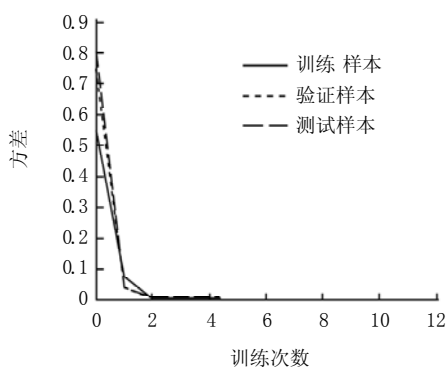


图1 网络学习误差平方和曲线
Fig.1 Squared error curves of network training

2.3 网络参数选择

使用MATLAB语言编制BP网络训练和分析程序,并利用该程序对神经网络有关参数进行优选。表3列出了不同隐层节点数时的训练结果,可以看出,隐层节点数较少,对工艺参数与性能之间关系的模拟不够精确;隐层节点数大于9时,网络性能变差。隐层节点数为9时,网络输出和目标之间的误差平方和

表3 不同隐含层神经元数的BP网络比较
Table 3 Comparison of BP networks with different numbers of neuron in hidden layer

网络结构	均方差	相关系数		
		训练	检验	评价
4-3-1	0.0069275	0.878	0.9	0.881
4-6-1	0.0066688	0.938	0.919	0.915
4-8-1	0.0061787	0.952	0.973	0.943
4-9-1	0.0018842	0.998	0.987	0.985
4-10-1	0.0044539	0.996	0.952	0.916
4-12-1	0.0049546	0.993	0.926	0.913

为0.0018842,网络对训练、验证和测试样本的输出与目标输出的相关系数在0.985以上,满足试验要求。

2.4 苹果渣多酚提取工艺的神经网络模型

根据以上的试验结果,得出苹果渣多酚提取工艺的神经网络结构为4-9-1。图2~4分别表示训练样本、验证样本和测试样本的BP网络仿真输出值和回归直线。由图可见,本实验建立的BP网络能够较好地仿真验证样本和测试样本数据,模型准确可信,能够用来实现苹果渣多酚提取工艺参数与提取率之间关系的映射。

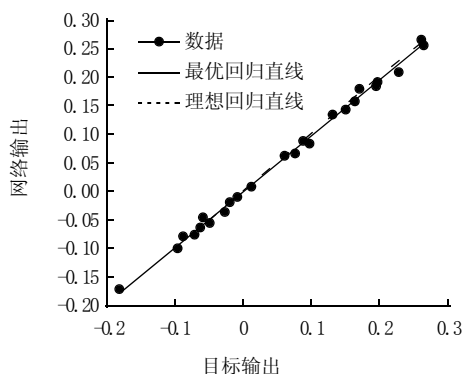


图2 训练样本网络输出回归直线
Fig.2 Linear regression between network output and corresponding targets

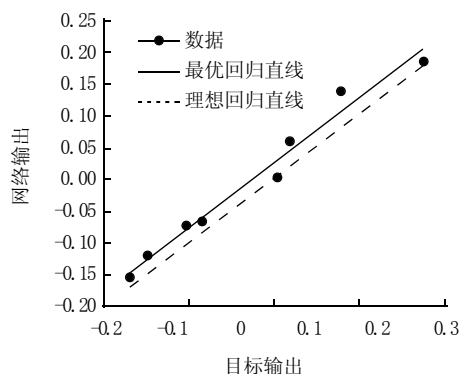


图3 检验样本网络输出回归直线
Fig.3 Linear regression between network outputs of test set and corresponding targets

2.5 遗传算法寻优结果

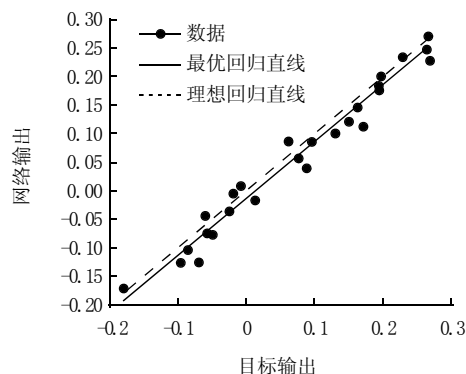


图4 评价样本网络输出回归直线
Fig.4 Linear regression between network outputs of validation set and corresponding targets

使用MATLAB语言实现神经网络技术和遗传算法结合的苹果渣多酚提取工艺的优化系统的开发。系统的技术路线见图5。

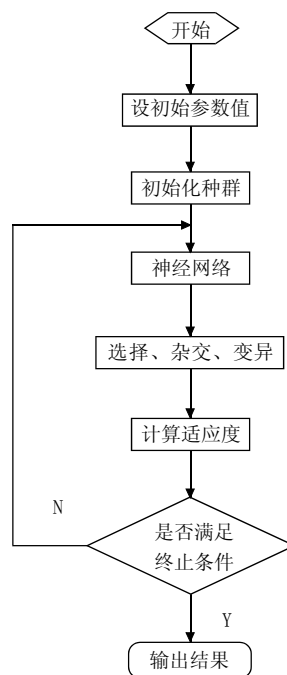


图5 遗传算法结合神经网络的示意图
Fig.5 Combination of genetic algorithms and neural networks

图6为该优化系统寻优的历代适应度变化曲线。图中虚线代表历代中最大适应度的变化曲线,实线表示每代中最大适应度的变化曲线,点线表示每代中适应度平均值的变化曲线。得到的最大的提取率为3773.6mg/kg,对应的最佳的工艺参数为乙醇浓度为62%、乙醇溶液的体积与苹果渣的质量之比为14:1,提取温度为69.7℃,提取时间为5.9h。

3 讨论

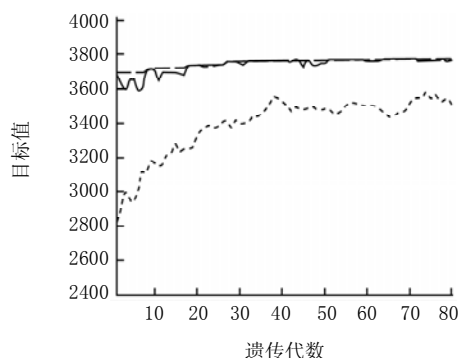


图6 目标值变化曲线

Fig.6 Effects of target value and generations

优化的结果明显大于单因素试验和二次组合试验的结果,比最好的(3227.760mg/kg)大16.9%。该工艺参数下提取率的预测值和实验值也基本相符,完全可以满足实际需要。苹果渣多酚提取工艺参数与提取率之间是多输入、复杂的非线性关系,机理分析法和传统的系统辨识法对这种问题适应性差,难以得到实用的提取工艺模型,人工神经网络技术具有辨识和逼近任意复杂非线性系统的能力,能实现苹果渣多酚提取工艺参数与提取率之间关系的映射,所建的结构为4-9-1的BP网络能较好地模拟验证样本和测试样本数据。神经网络建立的工艺参数与提取率之间的函数具有高度的非线性,是一个有约束的非线性优化问题,在寻找这些问题的全局最优解时,传统的优化方法有着局限性,遗传算法是基于生物学中遗传变异、杂交等的一种自适应全局优化概率搜索算法,能适用复杂形态函数的全局寻优,本实验将其用于人工神经网络模型的寻优,得到了苹果渣多酚提取的最佳工艺参数。

4 结 论

4.1 BP神经网络能适用于苹果渣多酚提取工艺参数与提取率之间的非线性建模。结构为4-9-1的神经网络能较为精确的拟合输入的样本数据,模型准确可信,可以代替试验,该模型的建立为工艺参数的优化打下基础。

4.2 苹果渣多酚提取的最佳工艺参数为乙醇浓度为62%,乙醇溶液的体积与苹果渣的质量之比为14:1,温度为69.7℃及提取时间为5.9h。该工艺参数下的提取率明显大于单因素试验和二次组合试验的结果,比最好的大16.9%。

4.3 建立苹果渣多酚提取工艺参数与提取率之间的神经网络模型,并利用遗传算法对模型输入参数空间进行寻优,这种将人工神经网络技术和遗传算法结合起来进行提取工艺参数优化是完全可行的。

参考文献:

- [1] 李志西,李元瑞,葛蕾,等. 苹果渣多酚超声波提取工艺及其抗氧化性研究[J]. 西北农林科技大学学报:自然科学版, 2005, 33(8): 131-134.
- [2] ALONSO M G, TERESA S P, BUELGA C S, et al. Evaluation of the antioxidant properties of fruits[J]. FC, 2004, 84: 13-18.
- [3] MITTAL G S, ZHANG J X. Prediction of freezing time for food products using a neural network[J]. Food Research International, 2000, 33: 557-562.
- [4] BOILLEREAUX L, CADET C, LE B A. Thermal properties estimation during thawing via real time neural network learning[J]. Journal of Food Engineering, 2003, 57: 17-23.
- [5] 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]. 无锡轻工业大学学报, 2001, 20(4): 416-421.
- [6] 方柏山, 陈宏文, 谢晓兰, 等. 基于神经网络与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): 640-648.
- [7] 高海燕, 薄亚明, 高志宇. L-异高氨酸发酵过程的神经网络预测控制[J]. 基础自动化, 2001, 8(3): 10-18.
- [8] 黄明志, 杭海峰, 储炬, 等. 人工神经网络在红霉素发酵过程状态预估中的应用[J]. 华东理工大学学报, 2000, 26(2): 162-170.
- [9] 蒋益虹, 沈益民. 人工神经网络用于红曲杨梅果酒生产工艺的优化[J]. 浙江大学学报:农业与生命科学版, 2003, 29(3): 275-279.
- [10] 赵丽丽, 陈宁, 熊明勇, 等. 利用神经网络对L-缬氨酸发酵建模[J]. 无锡轻工业大学学报, 2003, 22(2): 44-47.
- [11] 银建中, 修志龙, 毕明树, 等. “人工神经网络”方法用于超临界流体萃取模拟[J]. 高校化学工程学报, 2002, 16(6): 691-695.
- [12] 张宇飞, 邵秀丽, 雷建军. 基于神经网络和遗传算法的中药滴丸制剂过程建模与优化[J]. 计算机工程与应用, 2005(2): 192-216.
- [13] GRDINIC V, VUKOVI J. Prevalidation in pharmaceutical analysis Part I: Fundamentals and critical discussion[J]. Journal of Pharmaceutical and Biomedical Analysis, 2004, 35: 489-512.