

Detection of Cola Using Electronic Nose Based on GA-BP Network^{*}

LU Xiaoli¹, WANG Jun^{1*}, HAI Zheng²

1. College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China;
2. Biology Science Institute Langfang Normal College, Langfang Hebei 065000, China

Abstract : The combination of genetic algorithm and back propagation algorithm for training the neural network is described. It can improve the search efficiency and realize global optimization, and this GA-BP algorithm is employed to detect the cola by electronic nose. Compared with the standard back propagation algorithm and its improved method, the result shows the GA-BP algorithm has good prediction precision, high convergent speed and less running time, and it is a fast and credible method.

Key words : cola; electronic nose; BP neural network; genetic algorithm

EEACC :7230J

基于遗传优化神经网络的电子鼻对可乐的检测^{*}

鲁小利¹, 王 俊^{1*}, 海 铮²

1. 浙江大学生物系统工程与食品科学学院, 杭州 310029;
2. 廊坊师范学院生命科学院, 河北 廊坊 065000

摘 要 : 采用遗传学习算法和误差反向传播 (BP) 算法相结合的混合算法来训练前馈人工神经网络, 从而提高神经网络的收敛质量和收敛速度, 并将此算法运用到电子鼻对可乐的检测上. 与经典 BP 网络及附加动量项 BP 网络的训练与预测进行了比较, 结果显示: 遗传优化 BP 算法具有预测精度高、收敛速度快及运行时间短的优点, 是一种快速、可靠的方法.

关键词 : 可乐; 电子鼻; BP 神经网络; 遗传算法

中图分类号: TP212.6

文献标识码: A

文章编号: 1004-1699(2007)06-1211-04

电子鼻的研究始于 20 世纪 90 年代, 是一种新颖的分析、识别和检测复杂气味与大多数挥发性成分的仪器^[1-2]. 它是根据仿生学原理, 由传感器阵列和自动化模式识别系统所组成. 与普通的化学分析仪器, 如色谱仪、光谱仪等不同, 电子鼻得到的不是被测样品中某种或某几种成分的定性与定量结果, 而是给予样品中挥发性成分的整体信息, 也称“指纹”数据. 基于电子鼻的特点及其方便快速的优越性, 国内外已在食品工业、环境检测、医疗卫生、药品工业、安全保障、公安与军事等方面有不少应用和研究成果.

可乐作为一种液体软饮料, 受到人们特别是广大青少年的喜爱, 相应的质量的稳定越来越受到关注. 目前, 生产厂家和质量监督部门采取的主要检

测方法有化学分析方法和物理化学分析, 这二者都存在一定的局限性: 化学分析方法烦琐, 实时性差; 物理化学分析需要比较贵重的仪器, 且对样品的提取、制备和仪器操作技术要求都较高, 不适应快速分析的需要. 本文以百事可乐、可口可乐及非常可乐这三种常见的可乐为研究对象, 利用电子鼻进行检测. 在电子鼻的模式识别数据处理方面, 神经网络法特别是 BP 网络, 用得比较多^[3-4]. BP 网络的突出优点就是具有很强的非线性映射能力和柔性的网络结构, 但同时也存在已陷入局部极小、收敛速度慢和引起振荡效应等缺点. 由于遗传算法具有很强的宏观搜索能力, 又能以较大的概率找到全局最优解, 所以本文将二者结合起来, 形成遗传优化 BP 算法, 达到优化网络的目的.

基金项目: 国家自然科学基金项目资助(3057746); 教育部新世纪优秀人才支持计划项目资助(NET-04-0544)

收稿日期: 2006-08-24 修改日期: 2006-12-11

1 实验材料、仪器和方法

1.1 材料与仪器

本文中采用超市选购的可口可乐、百事可乐与非常可乐三种饮料为研究对象,所用的实验仪器是德国 AIRSENSE 公司的 PEN2 便携式电子鼻 (Portable Electronic nose).

PEN2 电子鼻包含 10 个金属氧化物传感器阵列,各个传感器的名称及性能描述见表 1. 根据传感器接触到样品挥发物后的电导率 G 与传感器在经过标准活性炭过滤气体的电导率 G_0 的比值进行数据处理和模式识别. 这个由传感器阵列组成的仪器主要包含下面几个部分:传感器通道、采样通道,计算机,其结构如图 1 所示.

表 1 PEN2 的标准传感器阵列

阵列序号	传感器名称	性能描述	备注
1	W1C	芳香成分	甲苯, 10 ml/m ³
2	W5S	灵敏度大,对氮化合物很灵敏	NO ₂ , 1 ml/m ³
3	W3C	氨水,对芳香成分灵敏	苯, 10 ml/m ³
4	W6S	主要对氢气有选择性	H ₂ , 100 ml/m ³
5	W5C	烷烃,芳香成分	丙烷, 1 ml/m ³
6	W1S	对甲烷灵敏	CH ₄ , 100 ml/m ³
7	W1W	对硫化物灵敏	H ₂ S, 1 ml/m ³
8	W2S	对乙醇灵敏	CO, 100 ml/m ³
9	W2W	芳香成分,对有机硫化物灵敏	H ₂ S, 1 ml/m ³
10	W3S	对烷烃灵敏	CH ₄ , 10 ml/m ³

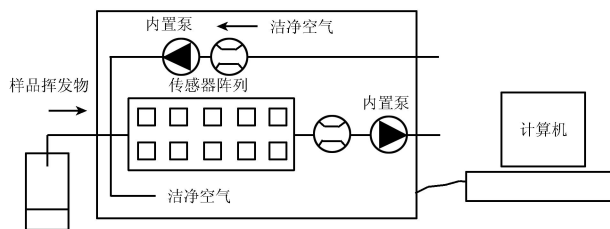


图 1 PEN2 原理图

1.2 实验方法

准确量取样品(可口可乐、百事可乐和非常可乐)2 mL,注入到 150 mL 的烧杯中,并用保鲜膜密封 30 min 后,采用顶空抽样的方法用电子鼻进行检测,检测时间为 60 s,传感器清洗时间为 100 s. 对于每类样品做 30 个重复,共 90 个重复.

2 实验结果与特征提取

PEN2 电子鼻系统所获得的数据是传感器接触到样品气体后的电导率 G 与传感器在经过标准活性炭过滤气体清洗后的电导率 G_0 的比值. 以可口可乐为例,传感器的响应曲线见图 2,每条曲线对应一个传感器,共 10 条曲线. 在测量过程中,每次检测时间为 60 s,电子鼻每秒获取一组(10 个)数据,因

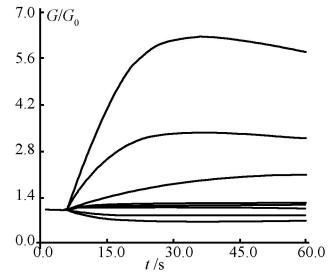


图 2 传感器对可口可乐的响应曲线图

此,对于每个样品,所获得的数据为一 10×60 维的矩阵. 从传感器的响应曲线图分析,选取了 10 s 和 15 s 为传感器响应曲线上上升阶段的响应值、30 s 时响应值为信号最大值以及 45 s、60 s 时响应值来表征信号相对稳定值. 这五个时刻的数据进行下一步的模式识别分析^[12]. 由于考虑到向量维数过多会影响到神经网络的训练速度,因而首先对数据进行 PCA 分析. 经分析发现,前三个主成分,方差贡献率很大,累计贡献率都在 85% 以上,这说明仅取前三个主成分所对应的特征向量决定的三维子空间就能够充分精确地拟合原始数据. 如此,将这些特征参数作为 BP 神经网络的输入特征向量.

3 经典 BP 算法及其改进算法

典型的 BP 网络是三层网络,包括输入层、输出层及隐层,对于网络的训练采用反向传播 (Back Propagation, BP) 算法. 这是一种有导师学习方法,其基本思想是最小二乘算法. 它利用根均方误差和梯度下降法来实现对网络连接权的修正,从而使网络趋向收敛,也即使网络的全局误差趋向极小值.

BP 改进算法中应用比较广泛的是附加动量项的 BP 算法. 经典 BP 算法实质上是一种简单的最速下降静态寻优算法,在修正网络权值时,只是按该时刻的负梯度方式进行修正,而没有考虑以前积累的经验,从而常常使学习过程发生振荡,收敛缓慢. 附加动量项的 BP 算法是网络在修正权值时,不仅考虑误差在梯度上的作用,而且考虑在误差曲面上的变化趋势的影响. Rumelhart, Hinton 和 Williams [1986]建议在权值修正表达式中加上一项“动量项”:

$$w_{ij}(n+1) = D_{ij}(n) + \eta \delta_j(n)$$

其中: $D_{ij}(n) = -\frac{\partial E}{\partial w_{ij}(n)}$ 为 n 时刻的负梯度; η 为学习率, $\eta > 0$; $\delta_j(n)$ 为动量项因子,一般取 0.95 左右.

附加动量法的实质就是将最后一次权值变化的影响,通过一个动量因子来传递,如此防止了 $\delta_j = 0$ 的出现,有助于使网络从误差曲面的局部极小值中跳出来.

4 基于遗传算法的 BP 神经网络

4.1 基本思想

遗传算法^[5-6] (Genetic Algorithm GA) 是一种非导数优化的随机优化方法,可以对一复杂的、多峰的、非线性极不可微的函数实现全局搜索,而 BP 算法对局部搜索比较有效,因此为了使算法能很快地找到满意解,可以先用遗传算法对初始权值进行优化,在解空间中定位出较好的搜索空间,然后用 BP 算法在这些小的解空间中搜索出最优解。

4.2 实现方法及步骤

(1) 编码方案

对 BP 网络中连接权值和阈值进行编码主要有两种方法:一是采用二进制编码方案,另一种是采用实数编码方案。这里采用实数编码。设有三层 BP 网络,其拓扑结构示意图如图 3,输入层 \times 隐含层 \times 输出层为 $n \times k \times m$ 的结构,其中 W_{ij} 为输入层中第 i 个结点与隐含层第 j 个结点的连接权值; o_{ij} 为隐含层中第 j 个结点与输出层第 i 个结点的连接权值。对该 BP 网络编码为: $W_{11} W_{12} \dots W_{1k} W_{21} W_{22} \dots W_{2k} \dots W_{mk} O_{11} O_{12} \dots O_{1n} O_{21} \dots O_{2n} \dots O_{kn}$ 。

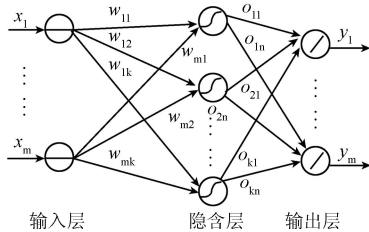


图 3 BP 网络结构示意图

(2) 适应度函数

衡量 BP 网络性能的主要指标是网络的实际输出值与期望输出值之间的误差平方和。该误差平方和小则表示该网络性能好。所以,本文中适应度函数表示为:

$$f = \frac{1}{S^2} = \frac{1}{l \sum_{i=1}^{nl} (y_i - T_i)^2}$$

式中: S 均方根误差; n 输出层神经元个数; l 训练样本数; y_i 对应第 i 个训练样本的网络实际输出值; T_i 对应第 i 个训练样本的网络期望输出值。

(3) 遗传操作

初始种群的群体大小 (POPSIZE) 对计算的搜索空间有很大的影响,本文中取 POPSIZE = 60。确定遗传操作算子:

选择操作:淘汰个体的操作,常用的有轮盘赌选择法和排序选择法。本文采用稳态轮盘赌选择法。具体过程为:将当代种群的个体适应度由大到小进行排序,按下式概率值选择个体: $P_s = \frac{f_i}{\sum_{i=1}^N f_i}$,其中 f_i 为个体 i 的适应度值, N 为染色体数。

交叉操作:依照交叉概率,随即选择的两个父个体按照一定的规则进行某些位置上的字符交换,产生新个体。

变异操作:模拟生物进化过程中的基因突变,常用方法为按位变异。

(4) BP 训练

经过以上的遗传运算,就得到了 BP 神经网络的误差最小的初始权值。将以上得到的初始权值代入 BP 网络进行正常的前馈训练,得到计算样本的实际输出值和期望输出的误差平方和 BP ,若达到了指定的网络的训练精度,则结束,否则转入遗传算法继续优化。

5 在可乐检测中的应用

5.1 经典 BP 算法与附加动量项 BP 算法

在实验中,每类样品做了 30 个重复,从中随机取出 16 个共 $16 \times 3 = 48$ 个样品组成训练集,剩下的 14 个共 $14 \times 3 = 42$ 个样品组成测试集。将从图 2 中提取的 50 个特征参数进行 PCA 分析,所得的 30 个特征向量作为神经网络的输入向量。经多次试验比较后,选择 BP 网络的拓扑结构为 30-12-3。

利用 MatLab7.0 编程实现训练与预测。参数的选择如下:经典 BP 算法学习率 0.9;附加动量项 BP 算法学习率 0.9,动态项因子 0.7;单个样本的最大误差 0.005,最大循环次数为 3000。训练集的回判正确率和测试集的判断正确率见表 3。图 4 所示为两种算法的训练误差平方和随训练步数的变化曲线。

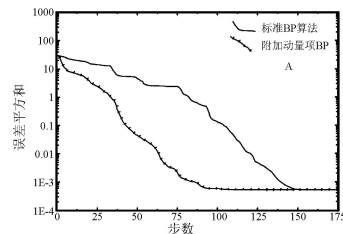


图 4 训练误差平方和随训练步数的变化曲线

表 3 三种算法结果比较

网络类型	网络拓扑结构	训练集的回判正确率	测试集的判断正确率
经典 BP 网络	30-12-3	100 %	83.33 %
附加动量项 BP 网络	30-12-3	100 %	88.10 %
遗传优化 BP 网络	30-12-3	100 %	97.62 %

5.2 遗传优化 BP

训练集和测试集的确定同上. 参数选择为: 遗传算法初始种群 60, 交叉率 0.8, 变异率 0.01, 其中 BP 网络的参数确定同上. 训练集的回判正确率和测试集的判断正确率见表 3. 图 5 所示为该算法的训练误差平方和随训练步数的变化曲线.

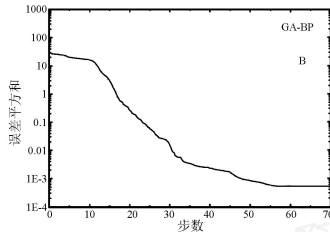


图 5 训练误差平方和曲线

5.3 效果分析

从表 3 可以看出, 虽然对训练集的回判正确率都达到了 100%, 但对测试集的判断率遗传优化 BP 为 97.62%, 比经典 BP 算法的 83.33% 和附加动量项 BP 算法的 88.10% 都要好. 从图 4 的对比中可以看到: 经典 BP 算法和附加动量项 BP 算法分别运行到 147 步和 97 步收敛到指定精度, 而遗传优化 BP 学习算法收敛到相同精度只用了 55 步. 此时的程序运行时间: 前两者分别为 23.874 s 和 17.644 s, 后者为 9.871 s.

为更好的考察与比较三种算法训练的网络的优劣性, 再从每类 30 个重复中随机选取测试集与训练集, 进行交互验证, 共重复了三次, 分别选取的组成测试集样本数为 18, 17, 15, 则对应的组成测试集的样本数分别为 12, 13, 15. 通过验证, 结果说明: 遗传优化 BP 算法比现有的经典 BP 算法和附加动量项 BP 算法在预测精度、收敛速度及运行时间上都取得了较好的效果.

6 结束语

GA 算法能以较快的速度减小搜索空间范围, 而且不易陷入局部极小点; 而 BP 算法则具有局部搜索效率高的特点, 将两者结合从而可以得到比现有的学习算法更好的学习效果, 是一种快速、可靠的方法; GA-BP 算法可以较好地运用到电子鼻的模式识别上. 本文中对可乐测试样本集的判断率为 97.62%, 要提高预测精度, 有待从算法本身的改进上作进一步的研究.

参考文献:

[1] Julian W Gardner, Philip N Bartlett. *Electronic Nose: Principles and Applications* [M]. Oxford University Press, 1999: 245.

- [2] Gardner JW, Bartlett N. A Brief History of Electronic Nose [J]. *Sensors and Actuators B*, 1994(18-19): 211-220.
- [3] 袁承任. 人工神经网络及其应用 [M]. 北京: 清华大学出版社, 1999.
- [4] 高大启. 基于神经网络的模式分类方法 [D]. 浙江大学博士学位论文, 1996.
- [5] David E Goldberg. *Genetic Algorithm in Search, Optimization and Machine Learning* [M]. Reading, MA: Addison Wesley, 1989.
- [6] 黄祖刚, 李建平, 何秀丽等. 用电子鼻鉴别卷烟的方法 [J]. *传感技术学报*, 2004, 23(6): 62-65.
- [7] 邹小波, 方如明, 吴守一等. 遗传算法在智能气体检测装置中的应用研究 [J]. *信号处理*, 2001, 17(5): 463-467.
- [8] 邹小波, 赵杰文, 潘胤飞等. 基于遗传 RBF 网络的电子鼻对苹果质量的评定 [J]. *农业机械学报*, 2005, 36(1): 61-64.
- [9] 李人厚, 张平安译. *精通 MatLab 综合辅导与指南* [M]. 西安: 西安交通大学出版社, 1998.
- [10] 张覃铁, 谢长生, 阳浩等. 电子鼻模式识别方法的对比研究 [J]. *传感技术学报*, 2005, 18(3): 576-579.
- [11] Bahram Ghaffarzadeh K. Using Neural Network and Genetic Algorithm to Enhance Performance in an Electronic Nose [J]. *IEEE Transactions on Biomedical Engineering*, 1999, 46(4): 429-439.
- [12] Cosimo Distanto. On the Study of Feature Extraction Method for an Electronic Nose [J]. *Sensors and Actuators*, 2002, B87: 274-288.



鲁小利(1981-), 女, 硕士研究生, 研究方向为电子鼻的模式识别技术, lxcherry@126.com



王 俊(1965-), 男, 浙江大学教授、博士生导师。“Sensors and Actuators B”等 7 本 SCI 刊物审稿人. 入选教育部新世纪优秀人才支持计划和浙江省“151 人才工程(第一层次)”. 主要从事基于电子鼻技术的农产品品质检测. 先后主持国家高技术研究发展计划(863 计划) 1 项、国家自然科学基金项目 5 项、教育部新世纪优秀人才支持计划 1 项、高校博士学科点基金 2 项、浙江省科技厅重大招标项目 2 项、其它省部级项目 10 余项. 已获国家发明专利 7 项、实用新型专利 12 项. 在国家级学报和境外学术刊物上已发表 100 余篇论文, 其中 SCI 收录 40 余篇、EI 收录 30 余篇. 发表的 SCI 收录论文中被引用 110 余次. jwang@zju.edu.cn

海 铮(1980-), 男, 硕士研究生(通讯联系人), 研究方向为电子鼻的应用技术, sea333355@yahoo.com.cn