

A Recognition Model of Hand Odor Based on BP Artificial Network*

Chengsheng Long, Xin Wang, Dehua Wu[#], Huidong Zhang, Zhenhua Song

Nanjing Police Dog Research Institute of Public Security Ministry, Nanjing
Email: {longchengsheng, #jqswdh}@163.com

Received: Jan. 18th, 2012; revised: Feb. 25th, 2012; accepted: Mar. 9th, 2012

Abstract: A recognition model has been developed based on Back Propagation Artificial Network and its net parameters were optimized. Hand odors were sampled, concentrated and then analyzed with Chromatography-Mass Spectrometer, resulting in hand odor profiles whose feature variables (47 organic compounds) were extracted and reduced by means of stepwise discriminant analysis. In sequence, a BP network, in which the structure was $47 \times 10 \times 1$, the transfer functions for hidden layer and output layer *tansig* and *logsig*, respectively, and the training function *trainrp*, was proposed. The experiment demonstrated that the network was able to specify whether an odor sample had been from man hand or woman hand.

Keywords: Hand Odor; Pattern Recognition; BP Network

基于 BP 网络的手部气味识别*

龙成生, 王 辛, 吴德华[#], 张汇东, 宋珍华

南京警犬研究所, 南京
Email: {longchengsheng, #jqswdh}@163.com

收稿日期: 2012 年 1 月 18 日; 修回日期: 2012 年 2 月 25 日; 录用日期: 2012 年 3 月 9 日

摘 要: 本文建立了一个基于 BP 人工神经网络的手部气味识别模型, 并对网络参数进行了选择优化。人体手部气味经样品采集、浓缩后, 用气相色谱-质谱联用仪对其进行分析获得手部气味轮廓图, 并利用逐步判别分析法提取了手部气味轮廓图的 47 个特征变量(有机化合物)。在此基础上, 构建了一个 $47 \times 10 \times 1$ 的 BP 网络, 其隐含层和输出层的传递函数分别为 *tansig* 和 *logsig*, 训练函数为 *trainrp*。该网络能正确区分不同性别的手部气味。

关键词: 手部气味; 模式识别; BP 网络

1. 引言

人体气味同指纹一样具有个体特征性, 能用于个体识别^[1,2]。同时, 人体气味在人的社会交际中起着巨大作用^[3]。目前, 大部分研究主要是针对人体气味的化学组成、产生机理和可能的生物功能进行的^[2,4]。让人遗憾的是, 人体气味中有生物功能的化学物质至今

仍不清楚。然而, 训练好的犬能够区别不同人的气味^[5]。警犬气味鉴别是警犬技术的重要组成部分, 能快速准确锁定犯罪嫌疑人。人体气味的个体特征性是警犬气味鉴别的基础, 手部气味是警犬气味鉴别特别关注的一类人体气味^[6,7]。

BP 网络是人工神经网络中应用最广泛的一种神经网络^[8]。在人工神经网络的实际应用, 80%~90%的人工神经网络模型都是采用 BP 网络或它的变化形

*资助信息: 公安部应用创新项目(2011YYCXNJQ164)。

[#]通信作者。

式。如今, BP 网络在模式识别^[9]、图像处理及分析^[10]、控制^[11]等领域均有广泛的应用。

为了探索人体手部气味是否具有区分性别的功能, 本文利用气相色谱 - 质谱联用技术(Chromatography-Mass Spectrometer, GC-MS)对手部气味进行了定性定量分析, 以图谱数据为基础, 建立了基于 BP 神经网络的手部气味识别模型。

2. 手部气味数字化及特征向量提取

手部气味样品利用气相色谱质谱联用技术进行分析^[12]。样品分析完成后, 对总离子流图上的色谱峰进行识别和积分(这由分析仪器所配的工作站 Agilent Chemstation Software 增强版(Windows XP)完成), 提取积分结果中的保留时间及其对应的峰面积, 并将提取的信息导入 Excel 表。导入 Excel 表后的数据形式为: 数据按保留时间升序排列, 每个样品占两列, 第一列为保留时间, 第二列为峰面积, 同一行的峰具有相等的保留时间(差值小于 0.05); 如果样品中上不存在某一保留时间的峰, 则将其峰面积设为 0。数据整理完成后, 将各峰的面积与该样品中所有峰面积的总和相除, 得到各个峰的相对丰度(最高丰度设为 1000),

后续的计算均基于相对丰度进行。

经过上述数据预处理后, 变量集由 135 个色谱峰组成, 即变量集含有 135 个变量。之后, 我们利用逐步差别分析法^[13]提取了特征变量, 特征变量集由 47 个色谱峰组成。图 1 为手部气味数字化及特征向量提取过程的流程图。

3. BP 网络的构建及优化

本研究组建的网络为一单隐含层 BP 网络。此网络结构为 $47 \times H \times 1$, 即输入层节点数为 47(特征变量的数目)、隐含层节点数为 H 、输出层节点数为 1, 如图 2 所示。

网络建立后, 我们优化了隐含层节点数和各层之间的传递函数, 比较了不同训练函数对该网络的影响, 网络参数取值范围如表 1 所示。同时, 我们也对该网络的初始权值和初始阈值进行了设定。特征变量协方差矩阵的前 H 行(H 为隐含层的节点数)为输入层与隐含层之间权值矩阵 $IW\{1,1\}$ 的初始值, 其主对角线为隐含层与输出层之间权值矩阵 $LW\{2,1\}$ 的初始值; 初始阈值均设为 0。

训练集由 40 个手部气味样品组成, 男女各 20 个;

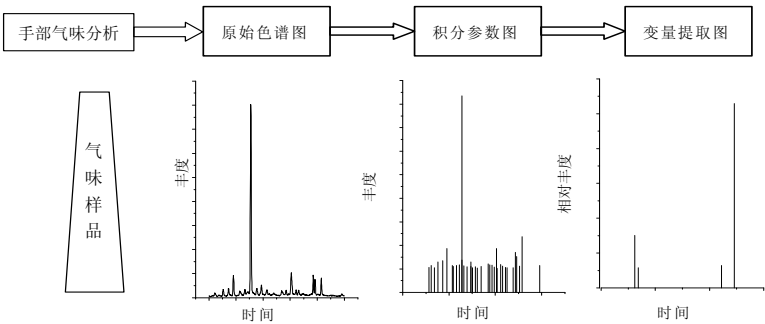


Figure 1. Odor Digitalization and feature variable extraction
图 1. 手部气味数字化及特征向量提取过程示意图

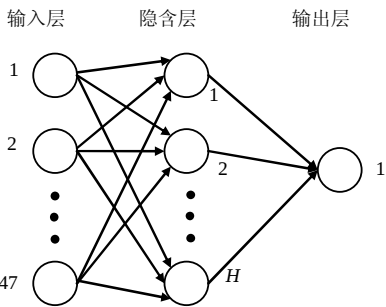


Figure 2. BP network structure
图 2. BP 网络结构

Table 1. Network parameters
表 1. BP 网络设计参数取值范围

参数	取值范围
输入层节点数	47
输出层节点数	1
隐含层层数	1
隐含层节点数	$H \in [7, 17]$
传递函数	<i>tansig</i> , <i>logsig</i> , <i>purelin</i>
训练函数	<i>trainlm</i> , <i>trainscg</i> , <i>trainrp</i>

测试集由 9 个手部气味样品组成，其中男性占 4 个。采用批序模式(batch model)对网络进行训练。所有实验均基于 Matlab (Version 7.0.0.19920(R14))神经网络工具箱进行。

此 BP 网络的主要 Matlab 的代码如下：

1) 参数设置

$P(47 \times 40)$ 为训练集， $T(1 \times 40)$ 为目标向量(0 代表男性，1 代表女性)， f_com 为特征变量的协方差矩阵(47×47)， $test_P(47 \times 9)$ 为测试集， TF_1 和 TF_2 分别为隐含层和输出层的传递函数， TrF 为训练函数，隐含层结点数 $NodeNum = H$ ，输出结点数 $OutNum = 1$ ，最大训练次数 $TrainEpochs = 10000$ 。

2) BP 网络的生成

```
net=newff(minmax(P),[NodeNum, OutNum],...
           {TF_1,TF_2 },TrF);
```

3) 初始权值和阈值设定

```
s_com= f_com(1:NodeNum, :);
com_diag = diag(s_com);
net.IW{1,1} = s_com;
net.LW{2,1}= com_diag(find(com_diag));
net.b{1} = zeros(NodeNum,1);
net.b{2} = 0;
```

4) 网络的训练及测试

```
net.trainParam.goal = 1e-20;
net.trainParam.min_grad = 1e-20;
net.trainParam.epochs = TrainEpochs;
```

```
[net,tr] = train(net, P, T);
test_result = sim(net, test_P).
```

4. 结果与讨论

4.1. 优化结果与讨论

1) 确定隐含层节点数。在讨论隐含层节点数对网络性能的影响过程中，隐含层的传递函数设为“*logsig*”，输出层的传递函数也为“*purelin*”，训练函数为“*trainscg*”，网络性能函数为“*mse*”。实验结果如表 2 所示。从表 2 中数据可以看出，隐含层节点 $NodeNum = 10$ 时网络性能最好。

2) 选择传递函数。在考察传递函数对网络性能的影响时，隐含层的节点数选为 10，训练函数为“*trainscg*”。一般情况下，第一层的传递函数为非线性函数。因此我们将 *tansig*，*logsig*，*purelin* 函数进行分组比较，组合情况如表 3 所示。表 4 为网络优化结果，数据表明隐含层和输出层的传递函数分别为 *tansig* 和 *logsig* 时，网络性能最好。

3) 选择训练函数。在隐含层节点数为 10，隐含层和输出层的传递函数分别为 *tansig* 和 *logsig* 条件下，我们考察了三个训练函数 *trainlm*，*trainscg*，*trainrp* 对网络性能的影响。实验结果如图 3 所示。当误差阈值设为 10^{-20} 时，用 *trainlm* 函数训练网络，其速度要快于 *trainscg* 函数，但比 *trainrp* 函数速度慢。当误差阈值设为 10^{-50} 时，*trainlm* 和 *trainscg* 两个函数均无法达到目标，但是 *trainrp* 仍具有非常好的性能。

Table 2. Results of optimization with various hidden nodes
表 2. 不同隐含层节点数的优化结果

节点数(NodeNum)	7	8	9	10	11	12	13	14	15	16	17
训练次数	3404	3033	3242	2731	3022	3413	3240	4012	3534	3846	3168
网络误差(MSE) $\times 10^{-21}$	9.971	9.989	9.773	8.080	9.982	9.069	8.837	9.968	9.738	9.218	9.947

Table 3. Transfer function groups
表 3. 不同传递函数的组合情况

组合序号	I	II	III	IV	V	VI
隐含层传递函数	<i>tansig</i>	<i>tansig</i>	<i>tansig</i>	<i>logsig</i>	<i>logsig</i>	<i>logsig</i>
输出层传递函数	<i>tansig</i>	<i>logsig</i>	<i>purelin</i>	<i>tansig</i>	<i>logsig</i>	<i>purelin</i>

Table 4. Results of optimization with different transfer functions
表 4. 不同传递函数优化结果

组合序号	I	II	III	IV	V	VI
训练次数	10,000	619	2724	10,000	654	2731
网络误差(MSE)	4.6701e-009	9.6488e-021	9.9106e-021	5.6499e-009	9.6429e-021	8.0797e-021

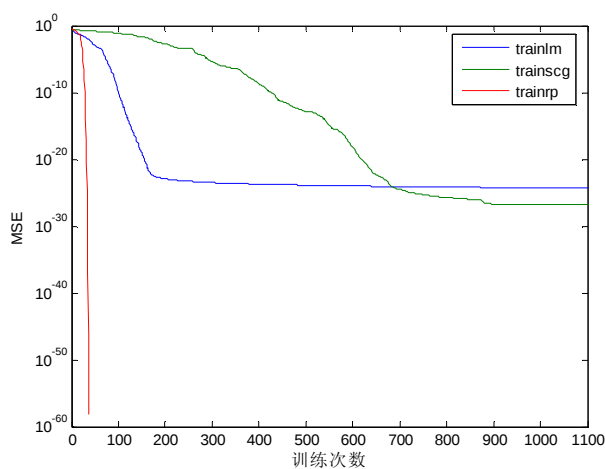


Figure 3. Optimisation results of different training functions
图 3. 不同训练函数优化结果

从图 3 可以看出，*trainlm* 和 *trainscg* 曲线的误差最后趋向于一个极限值，无法达到设定的误差阈值。说明这两个训练函数在此应用中有一定限制，无法达到应用要求。

综上所述，优化后的网络参数为：隐含层节点数 $H = 10$ ，隐含层传递函数为 *tansig*，输出层传递函数为 *logsig*，训练函数为 *trainrp*。

4.2. 训练及测试结果

利用样本数为 40 的训练集对 4.1 中优化好的 BP 网络进行了训练，并用另外 9 个样品对训练好的网络进行了测试，结果如表 5 所示。表中第 i 行第 j 列的数值表示类别 i 的样品被归类为类别 j 的百分数。例如，第 1 行第 1 列的数值 100 表示训练集中女性样品被归类为女性样品的百分比为 100%，即正确率为 100%。

Table 5. Results of the training and test for BP network
表 5. BP 网络训练及测试结果

	训练集样本(O)		测试集样本(O)	
	女性(%)	男性(%)	女性(%)	男性(%)
女性(T)	100	0	100	0
男性(T)	0	100	0	100

5. 结论

人体手部气味经样品采集、仪器分析、特征变量提取后，被转化为可视的手部气味轮廓图。本文利用 BP 人工神经网络实现了人体气味轮廓图的识别，即能够区分不同性别的手部气味轮廓图，此网络对训练集和测试的识别正确率为 100%。在实际应用过程中，此方法可与气味鉴别犬相结合，缩小目标气味的范围，达到快速破案的目的。当然，为了更好地融合仪器分析技术与警犬技术，我们还有大量的工作需要做。例如，开发适合分析人体手部气味的仪器分析技术；进一步探索人体手部气味的数字化方法等等。

参考文献 (References)

[1] D. J. Penn, E. Oberzaucher, K. Grammer, et al. Individual and gender fingerprints in human body odour. *Journal of the Royal Society Interface*, 2007, 4(13): 331-340.

[2] A. P. Kelly, D. P. Wood. *Handbook of hydrocarbon and lipid microbiology*. Berlin: Springer, 2010: 3204-3213.

[3] A. Weller. Human pheromones. *Communication through body odour*. *Nature*, 1998, 392(6672): 126-127.

[4] G. Preti, J. J. Leyden. Genetic influences on human body odor: From genes to the axillae. *Journal of Investigative Dermatology*, 2010, 130(2): 344-346.

[5] 吴德华, 王辛. 警犬鉴别人体气味的发展历程[J]. *警犬*, 2004, 4: 50-51.

[6] A. M. Curran, P. A. Prada and K. G. Furton. The differentiation of the volatile organic signatures of individuals through SPME-GC/MS of characteristic human scent compounds. *Journal of Forensic Sciences*, 2010, 55(1): 50-57.

[7] A. M. Curran, C. F. Ramirez, A. A. Schoon, et al. The frequency of occurrence and discriminatory power of compounds found in human scent across a population determined by SPME-GC/MS. *Journal of Chromatography B*, 2007, 846(1-2): 86-97.

[8] 傅芸璇, 赵红. *MATLAB 神经网络应用设计*[M]. 北京: 机械工业出版社, 2010: 83-97.

[9] 陈学峰, 赵质良, 王金光. 基于 BP 神经网络的齿轮箱故障模式识别[J]. *装备制造技术*, 2010, 10: 16, 27.

[10] Y. D. Zhang, Z. C. Dong, L. N. Wu, et al. A hybrid method for MRI brain image classification. *Expert Systems with Applications*, 2011, 38(8): 10049-10053.

[11] 刘召彤, 李艳萍, 杜光月. 基于 BP 网络的人体穴位分布模型设计[J]. *信息技术与信息化*, 2011, 1: 35-37.

[12] M. Kusano, E. Mendez and K. G. Furton. Development of head-space SPME method for analysis of volatile organic compounds present in human biological specimens. *Analytical and Bioanalytical Chemistry*, 2011, 400(7): 1817-1826.

[13] 龙成生, 王辛, 吴德华等. 基于逐步判别分析的血迹气味识别[J]. *数据挖掘*, 2012, 2: 1-5.