

doi:10.3969/j.issn.1007-7162.2015.03.017

一种基于电子鼻的辛味中药材的分类鉴别方法研究

范丹君, 骆德汉, 于昊

(广东工业大学 信息工程学院, 广东 广州 510006)

摘要: 为获取新的气味识别方法以提高智能传感器模式分类识别准确率和速度, 使用了内置 10 个传感器的便携式电子鼻 PEN3 对辛味中药材进行气味采集检测. 将辛味中药材在烧杯中进行密封静置待其形成稳定的气味顶空环境时, 运用电子鼻对其进行检测采样得到样品高维气味数据信息. 与传统的线性数据分析方法不同, 针对气味蕴含多种诸如浓度、各种挥发性物质成分等特征, 可知气味非线性的本质特征, 在本次分析中采用了流形算法中的非线性的局部线性嵌入 (Locally Linear Embedding, LLE) 算法对非线性的气味数据进行特征提取与降维, 再采用基于 Fisher 的线性判别分析 (Linear Discriminant Analysis, LDA) 实现对特征子空间的模式聚类与分类, 通过多次实验优化 LLE 算法的参数, 得到了最佳的辛味中药材的模式识别效果. 分析结果表明, 运用 LLE 和 LDA 相结合的算法 (即 LLE_LDA) 可以很好地完成不同种类辛味中药材的挥发性气味信息的模式分类, 为深层次地分析基于电子鼻的气味数据信息提供了一种新方法.

关键词: 模式识别; 气敏传感器; 局部线性嵌入_线性判别分析; 分类鉴别; 非线性降维

中图分类号: TP212.9

文献标志码: A

文章编号: 1007-7162(2015)03-0091-06

Classification of Pungent Chinese Herbs by Using Electronic Nose

Fan Dan-jun, Luo De-han, Yu Hao

(School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: In this approach, the intelligent portable electronic nose with a gas sensor array containing ten gas sensors was used to collect the odor information of different varieties of pungent Chinese herbs. In order to get a new sensor data processing solution to pattern recognition of the Chinese herbs accurately, a manifold algorithm named Locally Linear Embedding (LLE) was used to finish the dimensionality reduction and feature extraction of the odor data. And then the Fisher Linear Discriminant Analysis (LDA) was adopted to complete the identification via clustering. After many times of experiments to optimize the parameters of LLE, the classification results are quite perfect. The results demonstrate that by using this new combined method, the electronic nose can not only distinguish 6 types of Chinese herbs, but also identify the 3 different batches and the 3 different denominations of origin accurately, which provides a new method for further analysis.

Key words: pattern recognition; gas sensors; locally linear embedding_linear discriminant analysis (LLE_LDA); classification; nonlinear dimensionality reduction

电子鼻 (即 Electronic Nose), 是通过模拟生物的嗅觉器官并通过机器学习来实现鼻子生理功能的一种仪器^[1-2]. 传统的电子鼻模式识别方法主要有主成分分析 (Principal Components Analysis, PCA)、线性判别分析 (Linear Discriminant Analysis, LDA)、多

元线性回归 (Multiple Linear Regression, MLR)、偏最小平方 (Partial Least Squares, PLS) 等^[3-4]. 这些方法都是建立在响应数据正态分布这一假设上. 在实际应用中, 获取的高维数据往往呈现非线性特性或者线性不可分的情况, 很多研究表明, 气味数据是一种

收稿日期: 2014-01-07

基金项目: 广东省自然科学基金重点资助项目 (S2011020002906)

作者简介: 范丹君 (1987-), 女, 硕士研究生, 主要研究方向为仿生嗅觉/仿生味觉的智能传感器信号处理与分析.

具有内部特征如浓度、化学成分、体积分数等的高维非线性结构数据^[5-6]. 直接使用线性的分析方法, 可能会丢失原始数据的非线性信息, 不能反映数据的真实结构. 于是一些改进的算法被提出来, 如以核函数为基础的核主成分分析法(Kernel PCA)、加权核 Fisher 分析方法^[7]、支持向量机(Support Vector Machine, SVM)^[8]、核自组织映射聚类算法(Kernel SOM)^[9]、BP 神经网络(BP Neural Network)^[10-11]等. 但是核方法的效果取决于如何选择的核函数, 而且核选择通常不能体现数据内在结构. SOM 算法容易产生局部极小, 神经网络过分依赖于模型的选择, SVM 目前只给出解决二分类算法, 对于多分类问题, 需要通过多个二类组合解决, 效率不高.

在电子鼻获取的数据信息的处理中, 所采用的方法众多, 各有特点, 关键是要将数据的特点与模式识别方法组合起来, 以提高电子鼻气味识别效率和运算速度.

针对传统算法的不足, 文中以辛味中药材为研究对象, 采用非线性降维方法——局部线性嵌入 LLE (Locally Linear Embedding) 对基于电子鼻的气味数据进行分析. 与传统的线性降维方法相比, 非线性降维最显著优点是在进行数据分析时考虑到数据集内部的真实结构. 因此, 非线性降维方法在分析高维非线性数据, 如运用 LLE 来分析人脸数据时能够非常好地实现对高维数据的降维和特征提取^[12-13].

非线性的流形算法在图象处理领域获得了较成功的应用, 而且非线性的流形算法可以展现数据真实的内在结构. 但是由于 LLE 对稀疏矩阵很敏感, 较少运用于除图像分析以外的其他领域, 所以在电子鼻气味数据分析方面的应用非常少. 由计算分析结果知, 电子鼻采集的高维气味数据并非稀疏矩阵, 同时数据谱图分析类似于图像分析, 因此, 运用 LLE 方法来分析电子鼻采集到的高维气味数据具有可行性.

文中采用电子鼻系统对 3 组样本的辛味中药材进行气味检测与采集, 再采用 LLE_LDA 相结合的方法以 6 种不同种类辛味中药材、3 种不同生产批次(生产时间)的广东鸡骨草、3 种不同产地的何首乌为研究对象, 对样品的气味数据信息进行分析, 实现了对训练样本的分类鉴别.

1 实验材料、仪器和方法

1.1 实验材料

本实验选用的辛味中药材样本由本基金项目合

作单位广东药学院提供, 共 3 组样本:

(1) 6 种气味差异较大的不同种类的辛味中药材: 干姜(气辛辣)、广藿香(气芳香)、薄荷(气芳香)、肉桂(气辛、甜)、鱼腥草(茎叶揉之有浓烈鱼腥气)、鸡骨草(气微).

(2) 3 种同种类同产地不同生产批次的广东鸡骨草(*Herba Abri*, HA): 鸡骨草 20110720(简称 HA1107)、鸡骨草 20111124(简称 HA1111)、鸡骨草 20120519(简称 HA1205).

(3) 3 种不同产地的何首乌(*Radix Polygoni Multiflori*, RPM): 河南何首乌(RPM-HN)、广东何首乌(RPM-GD)、安徽何首乌(RPM-AH).

1.2 实验仪器

本实验采用德国 ARISENSE 公司生产的 PEN3 (Portable Electronic Nose 3) 型电子鼻, 该电子鼻在只有 1.8 mL 的体积内放置了 10 种不同的金属氧化物传感器. 所用的传感器对硫化化合物、甲烷、氢、乙醇和烃类物质具有很好的选择性. 同时使用 EDU 对气味进行富集与浓缩处理. 10 种金属氧化物传感器分别称为: $S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9, S_{10}$.

该电子鼻设备具有自动调整、自动校准及自动气味富集的功能. 在采样过程中的响应信号记录为某传感器接触到样品稳定挥发性气流后的电导率 G 与该传感器接触经过标准活性炭过滤的基准气体后的电导率 G_0 的比值, 即 G/G_0 .

1.3 实验方法/过程

称 10 g 样品, 置于 250 mL 的烧杯中, 采用静置顶空抽样的方法进行样品气味信息的采集, 顶空生成时间为 60 min, 即将样品装入烧杯后密封静置 1 h. 将 PEN3 电子鼻与电脑连接后, 运行其配置软件 WinMuster. 设定电子鼻的各检测参数如下: 采样间隔为 1 s, 采样时间为 120 s, 预采样准备时间为 5 s, 零标校正时间为 10 s, 清洗时间为 120 s, 样本进气流量为 150 mL/s, 每个样品采样 15 次.

1.4 数据分析/处理方法

根据电子鼻的气体阵列传感器采集的原始数据, 分别采用经典流形算法中的局部线性嵌入算法 LLE 以及 LLE_LDA 相结合的算法进行分析比较.

2 LLE_LDA 传感器智能信息处理方法基本原理

文中采用 LLE_LDA 的分析方法, 构造数据样本集的低维特征空间, 实现对不同属性气味数据的模式分类.

非线性降维的目的就是寻找高维数据集中的内在低维结构. LLE 是一种非线性降维方法,它能够使降维后的数据保持原有拓扑结构,而且对原始数据进行处理时不需要进行数据的预处理^[14-16].

其基本思想是认为每个数据点可以由它的 K 个邻居点线性组合而成,当所有高维数据样本点降到低维空间后,能保持其在原高维空间中的邻居关系不变.但是直接对数据进行降维可能会使原始数据有效信息丢失比较多,因此可能会丢失一些有用的信息.

LDA 是一种线性分析方法,它通过找到一个最佳投影,使各样本点间类间距离最大化,类内距离最小化. LDA 常用来设计线性分类器,用于多类别数据的分类判别.

结合两者的特点,目前有学者使用 LLE 与 LDA 相结合的方法用于人脸模式识别领域,并进行深入探讨^[17-18].

传统的 LLE 方法分以下 3 步:

- (1) 寻找每个样本点的 K 个近邻点.
- (2) 由每个样本的近邻点计算出该样本点的局部重建权值矩阵.
- (3) 是将所有的样本点映射到低维空间中.

在 LLE 算法中,设 $X = \{x_1, x_2, \dots, x_N\}$ 是 D 维欧氏空间中的 N 个数据点的集合. 假设数据点都位于一个维数为 $d (d \in D)$ 的非线性流形上, LLE 算法通过对每个数据点 x_i , 选择其 K 个邻居点 $N_i = \{x_{mi}, \dots, x_{mik}\}$, 通过求解一个有约束的最小二乘问题求出它们的最佳重构权重 $\{w_{i1}, \dots, w_{ik}\}$ (即线性组合系数), 在保持重构权重不变的情况下, 最小化代价函数:

$$\min \varepsilon(Y) = \sum_{i=1}^N \left| y_i - \sum_{j=1}^k w_j^i y_{ij} \right|^2.$$

通过最小化上述代价函数式,可以得到一个稀疏矩阵 M , 通过求解 M 的 $d+1$ 个特征向量, 取其 $2 \sim d+1$ 间的特征向量作为输出结果, 得到全局的低维嵌入坐标 $y = \{y_1, \dots, y_N\}$, 完成了数据的降维与特征提取, 得到了其特征子空间.

在本过程中, LLE 算法需要设置两个重要参数, 邻居点个数 K 和嵌入空间的维数 d . 要对参数进行优化, 选择合适的 d 值, 将其降低到合适的低维空间, 得到其最佳特征子空间.

最后, 将运用 LLE 得到的低维嵌入坐标作为 LDA 的输入, 在满足 LDA 算法不会出现小样本问题的前提下 (即训练样本数大于输入数据维数 d), 使

用 LDA 中的 fisher 判别准则使特征子空间映射到低维判别子空间完成特征子空间的分类. 调整 K 值和 d 值进行反馈分析, 找到最优参数, 从而实现对多类别电子鼻智能信号的分类.

3 结果与讨论

3.1 仿生嗅觉对辛味中药材的信号响应图

为了直观地表示出所选择的第一组 6 种辛味中药材样品 (干姜、广藿香、薄荷、肉桂、鱼腥草、鸡骨草) 的综合挥发性气味信息, 图 1 给出了 6 种样本在 10 ~ 60 s 时间内的可视化电子鼻传感器响应信号极坐标图谱. 由图 1 可知, 对不同的中药材品种, 传感器的响应曲线不大一样, 即各种药材具有自己的气味特异性且随着测量时间的延长其响应值趋于稳定. 图 2 给出了同一产地 3 个不同生产批次鸡骨草 (广东鸡骨草) 极坐标图谱, 对于不同批次的鸡骨草其响应特性最大差异体现在少数传感器上 (如 S_6, S_8); 图 3 给出了 3 种不同产地的何首乌的极坐标图谱, 对于不同产地的何首乌, 其响应特性差异只有在 S_2 和 S_{10} 上比较小, 在其他的几个气敏传感器上其响应差异比较明显.

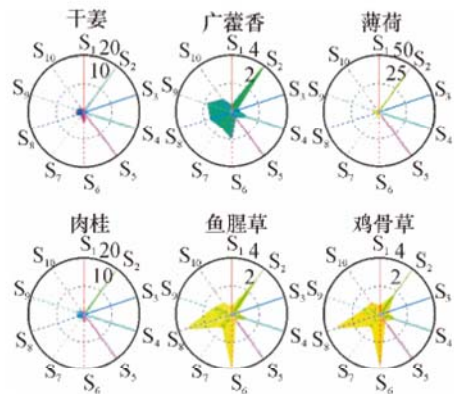


图 1 6 种辛味中药材气味响应极坐标图谱
Fig. 1 Polar maps of six Chinese Herbs

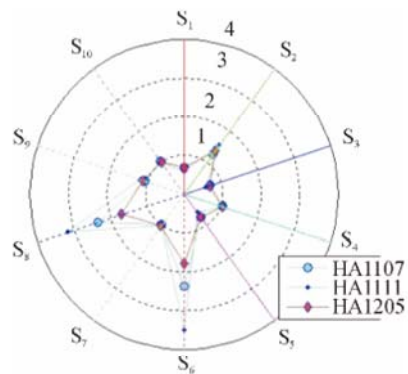


图 2 3 种鸡骨草的极坐标图谱
Fig. 2 Polar maps of three Herba Abri

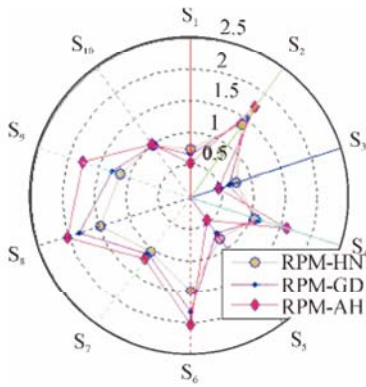


图3 3种何首乌的极坐标图谱

Fig.3 Polar maps of three Radix Polygni Multiflori

3.2 6种不同种类的中药材的分类鉴别

本组实验样品为:干姜、广藿香、薄荷、肉桂、鱼腥草、鸡骨草。分别运用LLE与LLE_LDA对6种不同种类的辛味中药材气味数据信息进行分析。图4为对气味数据单独使用LLE得到的分类结果图,图5为对气味数据使用LLE_LDA得到的分类结果图。由图4可知,虽然直接对电子鼻获取的辛味中药材气味信息使用线性嵌入算法LLE,其分类鉴别效果不是很好,样本交叉主要体现在干姜、薄荷上。但是其分类结果从另一个方面显示,使用LLE可以从某种程度上进行辛味中药材的气味数据的分类。

对图4和图5进行比较分析可知,若直接使用LLE进行中药材气味数据的分类,得到的结果是各样本点类内距离较大,样本点比较发散,容易引分类误判;若加入LDA进行聚类,则取得了较好的效果。但是从总体上看,各种类之间类间距离大、类内距离小,聚类效果好,分类效果比直接使用LLE好。

在本部分中,单独使用LLE进行分类模式识别时,LLE的参数设置为 $K = 15, d = 2$;使用LLE_LDA进行分类模式识别时,LLE的参数设置为 $K = 15, d = 20$ 。

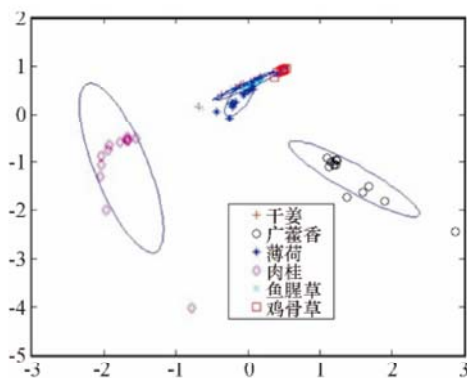


图4 6种辛味中药材LLE分类图

Fig.4 Classification of six Herbals by using LLE

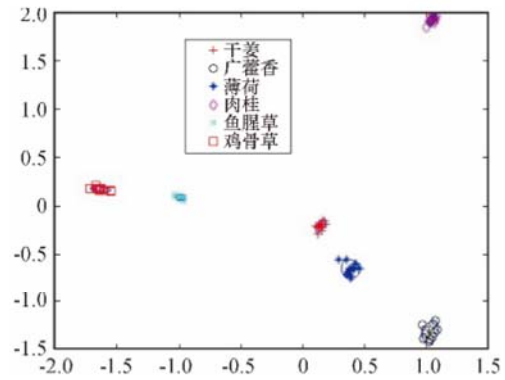


图5 6种辛味中药材的LLE_LDA分类图

Fig.5 Classification of six Herbals by using LLE_LDA

3.3 同种类同产地不同生产批次的中药材的分类鉴别

本组实验样品均为广东鸡骨草,生产批次分别是HA1107,HA1111,HA1205。

分别运用经典LLE与改进型LLE对3种不同生产批次的广东鸡骨草气味数据信息进行分析。图6为单独使用经典LLE时得到的分类效果图,图7为使用改进型LLE的分类效果图。比较图6和图7可知,若使用改进型LLE来进行实验样品的分类,可实现非常理想的效果,在实现数据的降维之后,类与类之间完全分开,类内样本点聚拢,达到了模式识别的最佳效果。

在对本组3个不同生产批次的广东鸡骨草实验样品进行分类鉴别时,每个批次的样品挑选10个样本数据,每个样本数据为 120×10 维,数据样本集为 $120 \times 10 \times 3$ 维。

使用经典LLE进行分类模式识别时,LLE的最优参数为 $K = 12, d = 2$;使用改进型LLE进行分类模式识别时,LLE的最佳参数设置为 $K = 12, d = 20$ 。

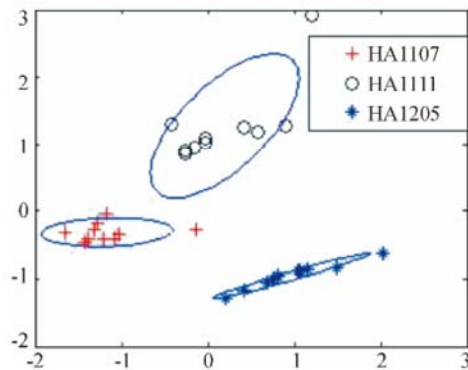


图6 3种不同生产批次的广东鸡骨草LLE分类图

Fig.6 Classification of three HAs by using LLE

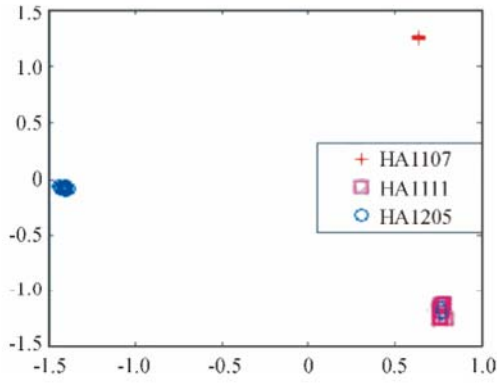


图7 3种不同生产批次的广东鸡骨草 LLE_LDA 分类图
Fig.7 Classification of three HAs by using LLE_LDA

3.4 同种类不同产地的中药材的分类鉴别

本组样品为3个不同产地的何首乌:河南何首乌(RPM-HN)、广东何首乌(RPM-GD)、安徽何首乌(RPM-AH)。

分别运用LLE与LLE_LDA对3种不同产地的何首乌气味数据信息进行分析。图8为单独使用LLE时得到的分类效果图,图9为综合使用LLE_LDA的分类效果图。比较图8和图9可知,运用LLE_LDA的效果明显好于单独使用LLE且模式识别效果非常好。

在参数设置上,在对本组辛味中药材样品进行分类鉴别时,单独使用LLE进行分类模式识别时,LLE的最优参数为 $K=12, d=2$;使用LLE_LDA进行分类模式识别时,LLE的最佳参数设置为 $K=12, d=20$ 。

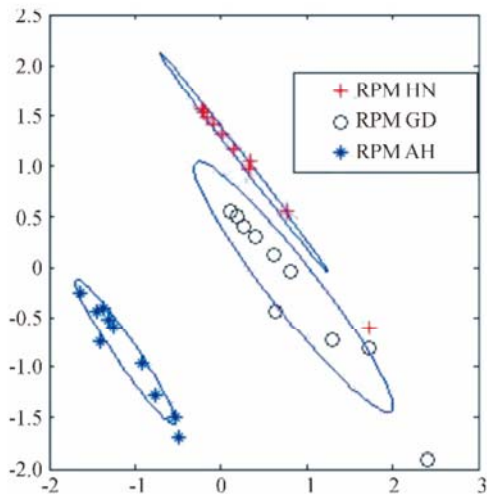


图8 3种不同产地的何首乌的LLE分类结果

Fig.8 Classification of three RPMs by using LLE

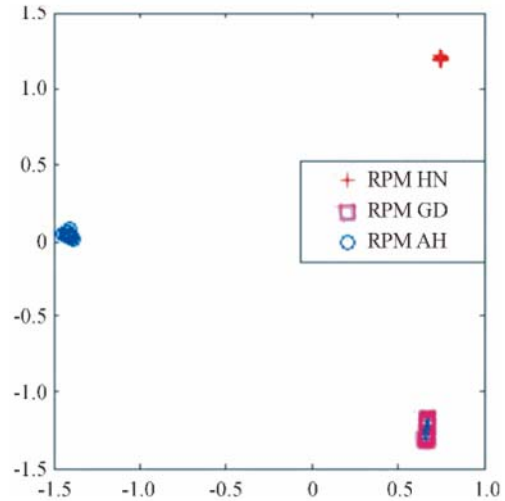


图9 3种不同产地的何首乌的LLE_LDA分类结果

Fig.9 Classification of three RPMs by using LLE_LDA

4 结束语

使用电子鼻对辛味中药材的高维气味信息进行采集,采用非线性方法LLE对高维非线性气味数据进行维数简约,最大程度还原气味数据真实的内部结构与局部特性,结合LDA进行聚类。结果表明,采用LLE_LDA算法对数据进行分析,成功实现了对6种不同种类辛味中药材以及3种不同批次的广东鸡骨草和3个不同产地的何首乌的分类鉴别。

由于LLE_LDA算法的结果很大程度上跟参数 K 和 d 的选择有关,在原始的LLE算法中,有两个参数很重要:

- (1) 每个样本点的邻域个数 K ;
- (2) 嵌入空间的维数 d 。

对于 K 值的选取,在LLE算法中是关键的一步,如果选择 K 值过大,就会丢失原始数据所处的流形的局部信息,如果选择的 K 值太小,会导致原来连续的流形分裂成互不相连的子流形。假如 d 过高,输出数据则易受到影响,反之,不能正确地提取样本数据的固有特征。因此两个参数的正确选择对于算法的执行起着重要的作用,如何选取合适的参数,值得进一步探讨。

总体而言,为基于电子鼻采集的气味信息完成中药材种类分类鉴别提供了一种新的思路与方法。

参考文献:

[1] Gardner J W, Bartlett P N. A brief history of electronic noses [J]. Sensors and Actuators B, 1994, 18 (19): 211-

- 220.
- [2] 张文娜, 秦国军, 胡鸩庆. 人工嗅觉系统关键技术研究进展 [J]. 传感器与微系统, 2011, 30 (8) : 1-4.
Zhang W N, Qin G J, Hu N Q. Research development of artificial olfactory system key technology [J]. Transducer and Microsystem Technologies, 2011, 30 (8) : 1-4.
- [3] 刘宁晶, 史波林, 赵镭, 等. 电子鼻检测技术研究进展 [J]. 食品科技, 2012, 37 (10) : 248-252.
Liu N J, Shi B L, Zhao L, et al. Progress in detection of e-electronic nose [J]. Food Science and Technology, 2012, 37 (10) : 248-252.
- [4] 刘红秀, 骆德汉, 张泽勇. 机器嗅觉系统气味识别算法 [J]. 传感技术学报, 2006, 19 (6) : 2518-2522.
Liu H X, Luo D H, Zhang Z Y. Odour recognition algorithms for machine olfaction system [J]. Chinese Journal of Sensors and Actuators, 2006, 19 (6) : 2518-2522.
- [5] 谷瑞军. 基于流形学习的高维空间分类器研究 [D]. 无锡: 江南大学物联网工程学院, 2008.
- [6] Michael C B, Doleman B J, Schaffer A, et al. Assessing the ability to predict human percepts of odor quality from the detector responses of a conducting polymer composite-based electronic nose [J]. Sensors and Actuators B, 2001 (72) : 149-159.
- [7] 刘晓亮, 王福龙, 黄诚, 等. 一种加权的核 Fisher 鉴别分析在人脸识别中的应用 [J]. 广东工业大学学报, 2009, 26 (4) : 65-69.
Liu X L, Wang F L, Huang C, et al. The application of a weighted kernel fisher discriminant analysis applied in face recognition [J]. Journal of Guangdong University of Technology, 2009, 26 (4) : 65-69.
- [8] Luo D H, Wang J, Chen Y M. Classification of Chinese Herbal Medicine Based on SVM [C] // 2012 IET International conference on Information Science and Control Engineering (ICISCE 2012). Shenzhen: [s. n.], 2012: 1191-1195.
- [9] 蒋玉玲, 杨宜民. 基于 SOM 算法的机器视觉颜色识别 [J]. 广东工业大学学报, 2011, 28 (2) : 40-42.
Jiang Y L, Yang Y M. Color recognition of machine vision based on SOM algorithm [J]. Journal of Guangdong University of Technology, 2011, 28 (2) : 40-42.
- [10] 梁慧冰, 李梅. 人工神经网络在预测领域中的应用 [J]. 广东工业大学学报, 1998, 15: 82-87.
Liang H B, Li M. Application of artificial neural networks for the prediction [J]. Journal of Guangdong University of Technology. 1998, 15: 82-87.
- [11] 卢萍, 金朝永. PID 神经网络的研究和改进 [J]. 广东工业大学学报, 2011, 28 (4) : 55-58.
Lu P, Jin C Y. Research and improvement of PID neural network [J]. Journal of Guangdong University of Technology, 2011, 28 (4) : 55-58.
- [12] 熊明, 王汝言, 唐琳. 基于局部线性嵌入与主成分分析的人脸识别方法 [J]. 重庆邮电大学学报: 自然科学版, 2009 (1) : 92-94, 114.
Xiong M, Wang R Y, Tang L. Face recognition based on locally linear embedding and principal component analysis [J]. Journal of Chongqing University of Posts and Telecommunications: Natural Science Edition, 2009 (1) : 92-94, 114.
- [13] 陈高曙, 曾庆宁. 基于 LLE 算法的人脸识别方法 [J]. 计算机应用研究, 2007, 24 (10) : 176-177, 187.
Chen G S, Zeng Q N. Face recognition method based on LLE algorithm [J]. Application Research of Computers, 2007, 24 (10) : 176-177 + 187.
- [14] Lin T, Zha H B, Lee S U. Riemannian manifold learning for nonlinear dimensionality reduction [J]. ECCV, 2006, 3951 (1) : 44-55.
- [15] Pan Y Z, Ge S S, Abdullah A M. Weighted locally linear embedding for dimension reduction [J]. Pattern Recognition, 2009 (24) : 791-811.
- [16] Rong Z, Min Y. Image feature optimization based on nonlinear dimensionality reduction [J]. Journal of Zhejiang University: Science A, 2009, 10 (12) : 1720-1737.
- [17] 薛清福, 李小丽, 陈雅芳. 融合 LLE 与 LDA 特征的人脸识别方法 [J]. 电脑与信息技术, 2010, 18 (3) : 5-8.
Xue Q F, Li X L, Chen Y F. Face recognition of fusing LLE and LDA [J]. Computer and Information Technology, 2010, 18 (3) : 5-8.
- [18] 马祥, 王映卓, 樊强. 基于 LLE 与 Fisher 线性判别的人脸识别算法 [J]. 现代电子技术, 2012 (8) : 64-66.
Ma X, Wang Y Z, Fan Q. Face recognition algorithm based on LLE and Fisher linear discrimination [J]. Modern Electronics Technique, 2012 (8) : 64-66.