

基于电子舌的掺假羊奶快速定量预测模型

Rapid quantitative prediction model of adulterated goat milk based on electronic tongue

韩 慧 王志强 李彩虹 马泽亮 国婷婷 殷廷家

HAN Hui WANG Zhi-qiang LI Cai-hong MA Ze-liang GUO Ting-ting YIN Ting-jia

(山东理工大学计算机科学与技术学院, 山东 淄博 255049)

(School of Computer Science and Technology, Shandong University of Technology, Zibo, Shandong 255049, China)

摘要:为实现对掺假羊奶的快速、客观辨别,模仿人体味觉感知机理研制了一套便携式电子舌检测系统,并建立了一种能够快速鉴别掺假羊奶的新方法。系统检测时,首先对样本溶液进行大幅脉冲扫描,用以获取掺假羊奶的“指纹”信息,然后利用离散小波变换(discrete wavelet transform, DWT)对“指纹”数据中的特征信息进行提取,最后在此基础上,采用主成分分析(principal component analysis, PCA)方法对不同掺假比例的羊奶进行定性辨别。采用粒子群优化极限学习机(Particle swarm optimization extreme learning machine, PSO-ELM)对不同掺假比例的羊奶进行了定量预测。通过试验数据得出,PCA 对 6 种不同掺假比例的羊奶区分达到 100%,区分效果好。PSO-ELM 羊奶纯度预测模型拟合曲线非常接近实测值曲线,因此采用 PSO-ELM 方法建立掺假羊奶纯度定量预测模型具有较高的预测精度。

关键词:电子舌;羊奶掺假;牛奶;主成分分析;粒子群优化极限学习机;预测模型

Abstract: In order to discriminate adulterated goat milk quickly and objectively, a set of portable electronic tongue detection system was exploited, and a new method of fast identification is developed. When detected in the system, the sample solution was first scanned to obtain the "fingerprint" information of adulterated goat milk, and then the discrete wavelet transform (DWT) was used to obtain the characteristics of the "fingerprint" data. On this basis, the principal component analysis (PCA) was used to determine the quality of goat milk with different adulteration ratio. Particle swarm optimization extreme learning machine (PSO-ELM) was applied to quantitatively

predict goat milk with different adulteration proportions. According to the experimental data, PCA could distinguish six kinds of goat milk with different adulteration ratios up to 100%, and it had a good effect on distinguishing adulterated goat milk. In order to realize the quantitative prediction of goat milk with different adulteration ratios, the fitting curve of PSO-ELM goat milk purity prediction model was very close to the measured curve, so the PSO-ELM method was used to establish the quantitative prediction model of goat milk purity with high prediction accuracy. This study might provide new ideas and technical support for qualitative identification and quantitative prediction of adulterated goat milk.

Keywords: electronic tongue; goat milk adulteration; milk; principal component analysis; particle swarm optimization extreme learning machine; prediction model

辨别羊奶质量检测经常使用的分析方法有感官评价分析法,但经常受到人们主观因素的影响,具有重复性差、成本高等缺点。高效气相色谱法、高效液相色谱法这些传统的检测方法所需仪器设备体积大、价格昂贵、分析时间长,无法达到对羊奶快速、客观、全面的检测。电子舌是模仿人体味觉感知系统而设计的一种新型的现代化智能分析检测设备,它检测简易方便、检测时间短、客观性强、成本低。目前,电子舌已广泛应用于环境^[1]、药品^[2-3]、食品^[4-6]等领域。近年来,国内外专家也投入大量精力研究解决羊奶的质量检测问题。例如金嫫等^[7]采用电子鼻检测羊奶中的牛奶掺入;王二丹等^[8]利用非线性化学群集成分分析法对掺杂在羊奶中的牛奶和马奶含量进行了定量分析;贾茹等^[9]利用电子鼻检测系统结合化学计量法对羊奶中蛋白质掺假进行了识别。但至今尚未有利用电子舌对羊奶掺假进行定性定量分析的相关报道。

本研究拟以掺假羊奶为检测对象,利用自行搭建的电子舌检测系统对 6 种混入不同比例牛奶的羊奶进行检测分析。本研究采用离散小波变换 DWT 对响应信号进行特征提取,

基金项目:国家自然科学基金(编号:61473179);山东省自然科学基金(编号:ZR2014FM007);山东理工大学中青年教师海内外访学计划经费资助(编号:2018)

作者简介:韩慧,女,山东理工大学讲师,硕士。

通信作者:王志强(1977—),男,山东理工大学教授,博士。

E-mail: wzq@sdut.edu.cn

收稿日期:2018-09-01

最后在此基础上,采用主成分分析法对不同掺假比例的羊奶进行定性辨别、采用粒子群优化极限学习机对不同掺假比例的羊奶进行定量预测。旨在为掺假羊奶的定性辨别和定量预测提供新的理论依据及技术支持。

1 材料与方法

1.1 样本制备

新鲜羊奶、新鲜牛奶:购于山东省淄博市本地大型超市,均在保质期之内,并在试验前密封保存,防止其氧化变质。

1.2 电子舌系统

USB6002 数据采集卡、自主研发的信号处理模块、传感器阵列、以及配套的上位机组成的电子舌系统如图 1 所示。首先虚拟仪器软件 LabVIEW 产生脉冲激励信号,该信号经过数据采集卡进行模数转换,转换后的信号还需要通过信号处理模块中的恒电位电路进行调理,最后才将处理后的信号施加于传感器阵列。传感器表面会发生电化学反应并产生微弱电流响应信号,由于不同传感器吸附分子能力的不同,得出的电信号也不同,从而导致传感器与标准电极之间电势差发生改变。信号调理模块对该信号进行 I/V(电流/电压)转换和信号放大及其滤波,之后传送到 USB6002 数据采集卡进行 AD 转换,最终送到上位机进行数据预处理和模式识别^[10]。

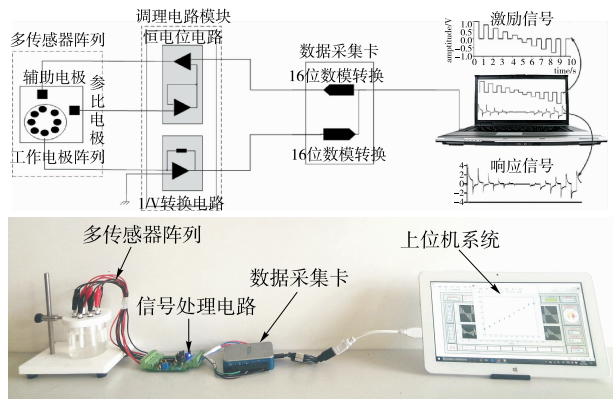


图 1 电子舌结构框图与实物图

Figure 1 The block diagram and physical diagram of electronic tongue

1.3 试验方法

掺假羊奶样本配制:将牛奶按 0%,10%,20%,30%,40%,50%的比例混入羊奶中,配制纯度为 100%,90%,80%,70%,60%,50%的掺假羊奶。配置好溶液后,为防止溶液变质,迅速贴好标签放入冰箱 0~6℃冷藏。检测试验在室温环境下进行。每次测量前后,对 8 个工作电极都进行电化学清洗和抛光处理。从配置好的待检测溶液中依次量取 50 mL 掺假羊奶进行检测。每次量取完之后将配置好的溶液密封好再次放入冰箱冷藏。利用上位机控制传感器阵列对配置好的溶液进行重复检测(20 次),每次检测完需对电极阵列进行电化学清洗,确保每个电极表面光滑如镜面且无残留物,以防止对下次检测结果造成影响。每检测 1 次,可得到 8 000 个数据点,最终可得到 120×8 000 的数据矩阵。

1.4 数据处理与分析

1.4.1 小波分析 在脉冲激励信号的激发下,通过 8 种特定的贵金属电极获取的响应信号数据融合后,除了包含溶液特征信息,还包含有一些冗余和高频噪声信息,对后期模式识别造成巨大干扰^[11]。为了降低系统的处理难度,减少处理时间,应对电子舌的原始数据进行特征提取,因此采用离散小波变换 DWT 对数据进行压缩和降噪预处理。它是时频的局部化分析,最终达到高频处时间细分,低频处频率细分,在去除高频干扰的同时有效地减少了冗余信息,从而提取出特征点,因此有效地压缩了数据,降低了后期模式识别的复杂度,是一种有效的数据预处理方式。

1.4.2 主成分分析 主成分分析是一种分析简化数据、探索数据结构以及进行数据降维的多元统计分析方法。少数几个能充分反映总体信息的指标从原来的多个指标中被组合出来。少数的几个指标就能充分反映总体信息,保留了重要信息且避开了变量间共线性问题^[12]。主成分分析不仅保留了原始变量的主要信息,而且将多个指标问题转换成少数几个综合指标问题,有效地起到降维与简化问题的作用,更容易抓住所研究问题的主要矛盾。主成分分析结果一般通过得分图和载荷因子图表示。主成分得分图中每个点代表 1 个样品,通过点与点之间的距离来反映样品之间特征差异的大小^[13]。

1.4.3 粒子群优化极限学习机

(1) 极限学习机:它(extreme learning machine,ELM)是单隐层前馈神经网络 SLFN 学习算法的一种,它使用简单、有效。人们所熟悉的 BP 神经网络学习算法使用前,需要提前对众多的网络训练参数进行设定,而且容易生成局部最优解^[14]。极限学习机是有监督的机器学习算法,可用来解决回归、二分类、多分类问题。极限学习机在算法的执行过程中不需要调整网络的输入权值以及隐含层偏置,随机初始化生成输入层和隐含层之间的权重和隐含层节点的阈值。只需要设置网络的隐层节点个数,就能获得唯一的最优解。

虽然在大部分情况下,极限学习机 ELM 具有良好的性能,但是极限学习机的精度与偏置阈值 b 、连接权值 W 、隐含层节点数目密切相关。通过计算输入权值矩阵和隐含层的偏置来获得输出权值矩阵,隐含层节点有时候无效,会导致输入权值矩阵和隐含层偏置为零。这样的话就表明实际运用时,为了达到预期的效果,更多的隐含层节点需要被设置。隐含层节点如果设置过多,网络会更加复杂,ELM 的泛化能力会降低,产生拟合现象,造成预测的稳定性不足^[15]。

(2) 粒子群:粒子群算法(PSO)是源于对鸟群捕食行为的研究而设计的一种群智能算法。在区域里只有一块食物。鸟群中的鸟儿们并不知道食物的存放在哪里,最佳的决策策略就是通过搜寻当前离食物最近的鸟的周围区域从而获得食物。粒子群算法(PSO)就是用粒子来模拟鸟群中的鸟儿,将每个优化问题当作是 D 维搜索空间上的一个点,也就是“粒子”。粒子移动的快慢用速度 V 来表示,粒子移动的方向用位置 X 表示。每个粒子在搜索空间中搜寻最优值 P_{best} ,并且将个体极值 P_{best} 共享给粒子群里的其他粒子,最优的个

体极值称之为整个粒子群的当前全局最优值 G_{best} , 粒子们追随当前的最优粒子在空间中搜索^[16]。

(3) 粒子群优化极限学习机:为了解决上文提到的极限学习机的缺点,粒子群优化极限学习机是将粒子群优化和极限学习机结合得到的学习算法。主题思想是极限学习机的输入层权值和隐含层偏置通过粒子群优化算法来进行最优化,以此得到一个最佳网络^[17]。也就是说将极限学习机的输入权值和隐含层偏置当作粒子群算法中的粒子。利用极限学习机算法计算输出权值矩阵,从而减少隐含层节点,依靠较少的隐含层节点获得较高的精度^[18]。

2 结果与讨论

2.1 小波信号预处理

本研究选用 Coiflets、haar、Daubechies、Symlets 作为小波基函数,对采集到的响应信号数据进行预处理,并根据响应信号的几何波形特征确定小波压缩层数。压缩重构信号和原始信号之间的相似程度随着压缩层数的增多而递减,如果压缩层数过多,虽然能够对原始数据进行压缩,但也必然会造成了大量特征信息的丢失,对原始信号特征的表达非常不利^[19]。本研究综合考虑压缩率以及减小数据分析处理难度,采用 MATLAB2014a 软件中 Coiflets、haar、Symlets 等小波函数对采集数据进行多次预处理试验。得到的近似系数重构信号与原始信号进行比较,选择最佳的小波分解层数和小波基函数。通过试验对比,发现选择 sym4 为小波基函数,原始数据经过 6 层压缩效果最佳。可以有效地删除无用的高频干扰和冗余信息,保留原始信号有效成分。将 8 000 个原始数据压缩至 67 个数据,实现数据的有效预处理,如图 2 所示。

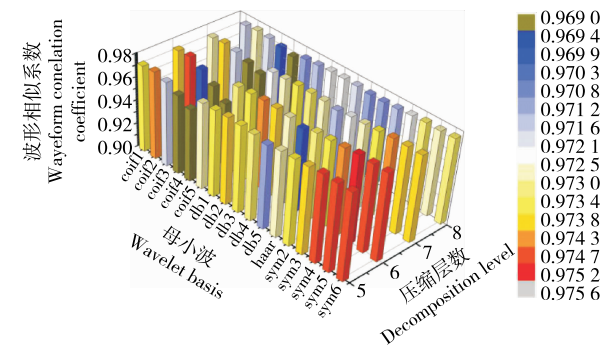


图 2 不同母小波和压缩层数对相似系数的影响

Figure 2 Effects of different mother wavelets and compression layers on similarity coefficient

2.2 掺假羊奶定性辨别

经过小波离散变换预处理后,用 PCA 对数据进行分析。经过 20 次平行检测后,得到其主成分得分值点分布情况如图 3 所示。从图 3 中可以看到,第 1 主成分的贡献率为 72.8%,第 2 主成分的贡献率为 21.1%,累计贡献率为 93.9%,基本能够代表样品的整体信息,说明电子舌的原始特征数据通过主成分分析很好地进行了说明;6 种纯度的羊奶样品分布在不同的区域,相互没有重叠。说明应用主成分分析法 PCA 可以将不同浓度的羊奶准确地区分开来,从而进行了有效的辨识。

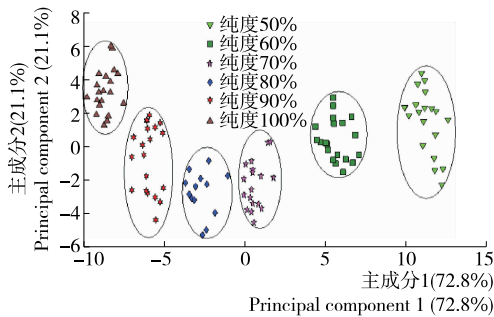


图 3 主成分分析图

Figure 3 The principal component analysis

2.3 掺假羊奶定量预测

为实现掺假羊奶的定量预测,采用 PSO-ELM 方法建立掺假羊奶定量预测模型。选取 90 个样本作为训练集用于模型的建立以及参数的优化。其中 90 个样本包含 6 组不同纯度样本,其中每一组又包含 15 个相同浓度样本。剩余 30 组(6 组纯度,每种纯度 5 个)作为预测集,用于模型的验证。

采用网格搜索法优化极限学习机(GS-ELM)、遗传算法优化极限学习机(GA-ELM)与粒子群优化极限学习机(PSO-ELM)进行比较分析,从而对 PSO-ELM 掺假羊奶的预测模型效果进行验证。分别按式(1)~(3)计算均方根误差(RMSE)、平均绝对误差(MAE)、相关系数(R^2),并对这 3 种 ELM 模型进行评价。

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (1)$$

$$MAN = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (2)$$

$$R^2 = \frac{[\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})]^2}{\sum_{i=1}^N (y_i - \bar{y})^2 \times \sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2}, \quad (3)$$

式中:

$y_i, \hat{y}_i$ —— 分别为样本的目标值和预测值;

$\bar{y}, \bar{\hat{y}}$ —— $y_i, \hat{y}_i$ 的平均值;

N —— 样本量。

RMSE、MAE、 R^2 范围在 $[0, 1]$ 内,并且这 3 个值可以反映模型的预测性和进度。RMSE、MAE 值距离 0 越近越好, R^2 距离 1 越近越好。不同参数优化下的羊奶纯度预测模型如图 4 所示。

从图 4 和表 1 中可以得出,网格搜索法建模集中 RMSE 为 0.022 7、 R^2 为 0.896、MAE 为 0.356,验证集中 RMSE 为 0.028 6、 R^2 为 0.862、MAE 为 0.368,相对于遗传算法及粒子群算法预测效果不好,可能是网格搜索法虽然能够通过逐步搜索寻找出满足精确度的参数组合,但需要不断缩小参数区间且计算量大、效率低所造成的。进一步比较遗传算法及粒子群算法发现,粒子群算法预测效果最好,这是因为遗传算法受适应度函数、初值的影响,而粒子群算法具有速度快、效率高,算法简单等优点。因此相对于网格搜索算法及其遗传算法,PSO 粒子群具有更佳的灵活和适应性,可以快速准确地寻找最优的 ELM 参数组合,因此 PSO-ELM 羊奶纯度预

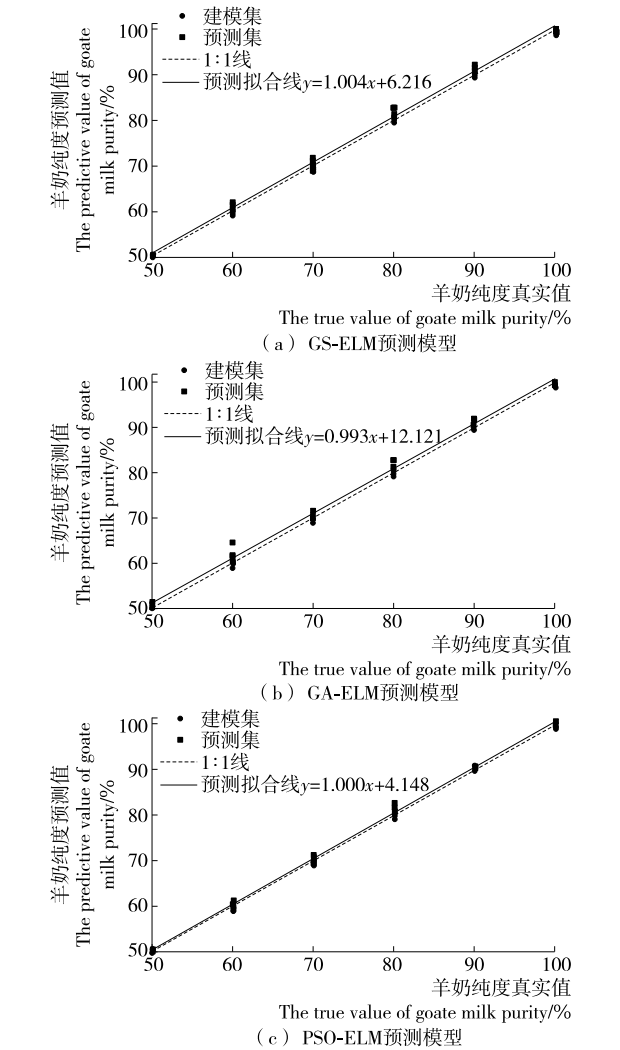


图 4 不同浓度羊奶样本的 ELM 预测图

Figure 4 The ELM prediction of goat milk samples with different concentrations

表 1 ELM 模型性能指标在不同参数优化方法下的对比

Table 1 Performance comparison of PLSR and SVM model based on different parameter optimization methods

预测模型	建模集			验证集		
	RMSE	R ²	MAE	RMSE	R ²	MAE
LOOCV-ELM	0.022 7	0.896	0.356	0.028 6	0.862	0.368
GA-ELM	0.012 3	0.975	0.153	0.027 5	0.964	0.267
PSO-ELM	0.012 2	0.985	0.132	0.023 4	0.968	0.263

测模型精度较高。

3 结论

本研究采用 PCA 对不同掺假浓度的羊奶进行了识别，不同掺假浓度的羊奶各自落在不同的区域，相互之间隔离比较远，因此电子舌可以将不同掺假浓度的羊奶进行有效的区分。采用粒子群优化极限学习机对不同掺假比例的羊奶进行定性预测，PSO-ELM 羊奶掺假比例预测模型具有较高的预测精度。该算法综合了极限学习机 ELM 和粒子群 PSO

的优点，具有参数调整简单、泛化性好的优点。试验结果表明，电子舌能够将不同掺假浓度的羊奶粉很好地区别开来。电子舌作为一种新型的现代化智能感官仪器，下一步在牛奶及其羊奶奶粉的品质以品牌区分以及掺假鉴别等方面具有巨大潜力。

参考文献

[1] 刘大龙, 方昉, 门洪, 等. 用于环境检测的电子舌研究[J]. 传感器学报, 2004(1): 1-6.

[2] 史庆瑞, 马泽亮, 周智, 等. 基于电子舌和模式识别的中成药品质辨识方法研究[J]. 电子测量与仪器学报, 2017(7): 1 081-1 089.

[3] 刘瑞新, 吴子丹, 李学林. 电子舌在药学领域的应用[J]. 中药与临床, 2011, 2(5): 61-64.

[4] BANERJEE R, TUDU B, SHAW L, et al. Instrumental testing of tea by combining the responses of electronic nose and tongue[J]. Journal of Food Engineering, 2011, 110: 356-363.

[5] TIWARI K, TUDU B, BANDYOPADHYAY R, et al. Identification of monofloral honey using voltammetric electronic tongue[J]. Journal of Food Engineering, 2013, 117: 205-210.

[6] 郑海霞, 尹芳缘, 王敏敏, 等. 基于电子鼻的牛奶品质预测方法研究[J]. 中国食品学报, 2013(7): 189-194.

[7] 金螺, 白丽娟, 彭小雨, 等. 采用电子鼻检测羊奶中的牛奶掺入[J]. 食品发酵与工业, 2015(4): 165-168.

[8] 王二丹, 鲁利利, 张泰铭, 等. 非线性化学群集成分分析法测定掺杂在羊奶中的牛奶和马奶含量[J]. 高等学校化学学报, 2015(6): 1 052-1 060.

[9] 贾茹. 电子鼻对羊奶中的蛋白质掺假及多组分混合掺假的识别研究[J]. 食品科学, 2017, 38(8): 308-312.

[10] 史庆瑞, 国婷婷, 殷廷家, 等. 基于电子舌检测的橙汁贮藏品质研究[J]. 食品与机械, 2017(11): 137-142, 203.

[11] 林兆彬, 胡毅, 郑江龙, 等. 小波变换压制噪声在单道地震资料处理中的应用[J]. 应用海洋学学报, 2018(1): 113-119.

[12] 冯艳玲. 一种基于数据挖掘的 HIFU 温度估计方法[J]. 电脑知识与技术, 2011, 7(16): 3 896-3 898.

[13] 张水清, 黄绍敏, 郭斗斗, 等. 基于主成分分析法的土壤肥力评价[J]. 安徽农业科学, 2011, 39(2): 1 096-1 097.

[14] QIU Shan-shan, WANG Jun, TANG Chen, et al. Comparison of ELM, RF, and SVM on E-nose and E-tongue to trace the quality status of mandarin[J]. Journal of Food Engineering, 2015, 166: 193-203.

[15] 陈晓青, 陆慧娟, 郑文斌, 等. 自适应混沌粒子群算法对极限学习机参数的优化[J]. 计算机应用, 2016(11): 3 123-3 126.

[16] 张颖, 李梅. 基于粒子群优化极限学习机的水质评价新模型[J]. 环境科学与技术, 2016(5): 135-138.

[17] 徐大明, 杜永贵, 孙传恒, 等. 基于改进粒子群优化极限学习机的养殖氨态氮含量预测模型[J]. 江苏农业科学, 2017, 45(4): 183-186.

[18] 李婉华, 陈羽中, 郭昆, 等. 基于改进粒子群优化的并行极限学习机[J]. 模式识别与人工智能, 2016(9): 840-849.

[19] 马泽亮, 殷廷家, 国婷婷, 等. 采用电子舌法检测橙汁及酒的品牌及纯度[J]. 食品工业科技, 2018(8): 190-194.