

油脂安全

DOI: 10.12166/j.zgyz.1003-7969/2020.01.009

核磁共振氢谱 – PCA – SVM 回归法用于稀奶油中 植脂奶油掺假定量分析

李 玮, 杨红梅, 王 浩, 贾婧怡, 刘 琪

(北京市食品安全监控和风险评估中心, 北京 100094)

摘要:基于核磁共振氢谱 – PCA – SVM 回归方法, 建立稀奶油中掺假植脂奶油的快速定量方法。核磁共振氢谱数据经分段积分、归一化等数据预处理后, 利用 PCA 进行数据降维, 采用交叉验证的方法对 SVM 中的参数进行优化, 然后使用最优参数建立稀奶油中掺假植脂奶油比例的定量校准模型。将校正结果与 PLS 和 SVM 算法比较, 并将所建立的 3 个模型用于测试集样本的预测。结果表明: 基于 PCA – SVM 算法定量模型的 RMSECV 为 3.69, RMSEP 为 5.87, 训练集和测试集的 R^2 分别为 0.987 5 和 0.974 3, 模型的稳定性、准确性以及模型的预测能力均优于 PLS、SVM 算法, 且运行速度明显快于 SVM 算法。研究所建立的 PCA – SVM 模型表现出较好的模型稳定性和预测精度, 结合核磁共振氢谱技术可以快速、准确地测定稀奶油中掺假植脂奶油含量, 为规范市场上奶油蛋糕等烘焙制品的质量监管提供技术支持。

关键词:核磁共振氢谱(^1H NMR); 主成分分析; 支持向量机; 稀奶油; 植脂奶油; 掺假定量

中图分类号: TS201. 2; TS225 文献标识码: A 文章编号: 1003-7969(2020)01-0038-06

Determination of cream adulterated with non-dairy whip topping using ^1H NMR combined with principal component analysis and support vector machines

LI Wei, YANG Hongmei, WANG Hao, JIA Jingyi, LIU Qi

(Beijing Municipal Center for Food Safety Monitoring and Risk Assessment, Beijing 100094, China)

Abstract: The adulteration of non-dairy whip topping in cream was analysed using ^1H NMR combined with principal component analysis (PCA) and support vector machines (SVM). ^1H -NMR was subsection integrated and normalized, and spectral dimension was also reduced through PCA. The cross validation was applied to optimize the parameters of PCA – SVM. The calibration model of determination of cream adulterated with non-dairy whip topping was established using the optimal parameters of PCA – SVM. The performance of PCA – SVM models was compared with partial least squares (PLS) and SVM, and the feasibility of these three methods was examined on the testing set. The results showed that the RMSECV and RMSEP obtained for PCA – SVM were 3.69 and 5.87 respectively, and the determination coefficients of the training set and testing set were 0.987 5 and 0.974 3 respectively. The stability, accuracy and prediction ability of the model were better than PLS and SVM algorithm, and the running speed was faster than SVM algorithm. In conclusion, a combination of ^1H NMR with PCA – SVM method could quickly and accurately determined the content of adulterated non-dairy whip topping in cream, which could provide technical support for regulating the quality supervision of bakery products such as cream cake on the market.

收稿日期: 2019-04-22; 修回日期: 2019-09-20

基金项目: 国家重点研发计划(2018YFC1602304)

作者简介: 李 玮(1984), 女, 正高级工程师, 博士, 研究方向为食品营养与安全(E-mail) liwei@bjmu.edu.cn。

Key words: ^1H nuclear magnetic resonance spectroscopy (^1H NMR); principal component analysis (PCA); support vector machines (SVM); cream; non-dairy whip topping; quantification of adulteration

稀奶油(cream)是以动物乳为原料的乳制品,是制作裱花蛋糕等西式烘焙食品的重要原料^[1]。植脂奶油(non-dairy whip topping)是以食用氢化油、甜味剂等为主要原料的人造奶油^[2],与稀奶油在外观、特性与用途上都极为相似,但价格却远低于稀奶油。在实际烘焙制品应用中,不良商家常以植脂奶油掺假于稀奶油中来替代纯稀奶油。但目前国内有关烘焙制品中奶油种类及掺假定量的检测标准尚未出台,加之相关国家标准的缺失,给有效监管带来了难度,而且随着网络烘焙店的兴起和奶油蛋糕等西式烘焙食品消费的逐步增长,这些问题将会越来越突出。

目前,对乳脂掺假检测的传统方法有感官鉴别、气相色谱等方法,但是这些方法前处理繁琐、检测时间长,且测定结果受人为因素影响大^[3-4]。近年来,亦有学者采用红外、拉曼、荧光等光谱技术结合化学计量学方法针对油脂含量在80%以上的黄油进行掺假定量鉴别研究^[5-7]。但这些光谱技术受其检测原理所限,其图谱对乳脂等复杂混合物体系内各成分解释能力有限,难以实现对建模样本的真实性判断,具有引入不真实的样本造成校正模型的真实性降低或(和)将特殊真实样本排除在建模样本外造成校正模型的普适性减弱的风险。为有效解决烘焙市场上的问题,应以稀奶油和植脂奶油为基础,采用接近实际情况的方式配制掺假的混合奶油制品并开发适于稀奶油检测的前处理方法。

核磁共振(nuclear magnetic resonance, NMR)是有机物结构分析的强有力工具,与液相色谱、气相色谱法等传统分析方法相比,具有高通量、重现性好、操作简便、结构信息丰富等优点,对食品等复杂体系的整体分析具有优势^[8-9],已被用于多种植物油及其他脂类成分的分析中^[10-11]。由于通过NMR所得的样品信息具有数据信息复杂、多维数据矩阵内各变量之间具有高度的相关性等特点,需要结合化学计量学等算法对数据进行分析。支持向量机(SVM)^[12-13]是近年来机器学习研究的一项重大成果,其主要思想是建立一个分类超平面作为决策曲面,使得正例和反例之间的隔离边缘被最大化,错误概率的上界最小化。与传统的人工神经网络相比,支持向量机结构简单,且泛化能力明显提高。

本研究采用市售稀奶油、植脂奶油样本,建立了基于核磁共振氢谱-PCA(主成分分析)-SVM回归法的稀奶油中植脂奶油掺假含量测定模型,利用该模型可实现裱花奶油产品中稀奶油掺假植脂奶油

的含量测定。与传统的PLS(偏最小二乘法)及单纯SVM回归算法建立的定量模型相比,基于PCA-SVM算法的定量模型的稳定性、准确性以及模型的预测能力均优于PLS、SVM算法,可为规范市场上奶油蛋糕等烘焙制品的质量监管提供技术支持。

1 材料与方法

1.1 实验材料

1.1.1 原料与试剂

市场上采集了25个不同品牌、产地的奶油样品(见表1),其中稀奶油(脂肪含量35.0%~38.0%)14个,植脂奶油(脂肪含量16.5%~37.5%)11个。稀奶油4℃存放;植脂奶油-20℃存放,使用前常温融化,避免反复冻融。所有样品均在保质期前使用。

氘代氯仿(CDCl_3 ,氘代度99.8%),美国CIL公司。

表1 用于 $^1\text{H-NMR}$ 分析的25个奶油样品信息

编号	产地	种类	编号	产地	种类
C-1	法国	稀奶油	C-14	意大利	稀奶油
C-2	意大利	稀奶油	N-1	中国台湾	植脂奶油
C-3	德国	稀奶油	N-2	中国苏州	植脂奶油
C-4	中国青岛	稀奶油	N-3	中国天津	植脂奶油
C-5	新西兰	稀奶油	N-4	中国佛山	植脂奶油
C-6	法国	稀奶油	N-5	中国盐城	植脂奶油
C-7	英国	稀奶油	N-6	中国河南	植脂奶油
C-8	法国	稀奶油	N-7	中国苏州	植脂奶油
C-9	丹麦	稀奶油	N-8	中国盐城	植脂奶油
C-10	德国	稀奶油	N-9	韩国	植脂奶油
C-11	澳大利亚	稀奶油	N-10	中国佛山	植脂奶油
C-12	法国	稀奶油	N-11	中国江门	植脂奶油
C-13	爱尔兰	稀奶油			

1.1.2 仪器与设备

Bruker AVANCE 600 MHz超导傅里叶变换核磁共振仪(配有BBO探头和Topspin3.2处理软件),瑞士Bruker公司;Norell 5 mm核磁管,美国Norell公司;XS204电子天平,瑞士Mettler Toledo公司;Centrifuge 5424R离心机,德国Eppendorf公司;TARGIN多管涡旋振荡器,北京踏锦科技有限公司;Tissuelyser II均质器,德国Qiagen公司。

1.2 实验方法

1.2.1 掺假样品的制备

按稀奶油中掺假植脂奶油质量分数分别为0%、5%、15%、25%、35%、45%、55%、65%、75%、85%、95%、100%比例称取2种奶油,将称好的2种奶油分别置于12个15 mL的塑料瓶中,每个比例点称取的2种奶油的总质量为10 g。放入直径约为5

mm 的不锈钢研磨珠,用盖子密封后放入多管涡旋振荡器中振荡 1 min,使瓶中的两种奶油充分混合。

1.2.2 样品溶液的制备

称取约 500 mg 1.2.1 中制备的奶油样品于 2 mL EP 管中,加入 1 mL CDCl_3 ,置于均质器中均质 40 s(30 Hz)后放入离心机中,4℃条件下离心 10 min (8 000 r/min)。移取 600 μL 上清液于 5 mm 核磁管中,待测。

1.2.3 训练集与测试集样本的获得

在 Matlab 中用 Randperm 函数随机将 14 个稀奶油分为 2 组,每组 7 个样品;同样,将 11 个植脂奶油分为 2 组,一组 6 个样品,一组 5 个样品。取稀奶油中第 1 组 7 个样品和植脂奶油中第 1 组 6 个样品,按照 1.2.1 掺假样品制备方法制备样品,共 213 个样本作为训练集(training set)样本;取稀奶油中第 2 组 7 个样品和植脂奶油中第 2 组 5 个样品,按照 1.2.1 掺假样品制备方法制备样品,共 112 个样本作为测试集(testing set)样本。

1.2.4 仪器条件

^1H NMR 测定条件:使用 Bruker 标准脉冲序列 noesyig1d,检测温度为 297 K,谱宽 SWH 为 6 002.40 Hz,中心频率 O1P 为 2 400.52 Hz,脉冲延迟时间 D1 为 10 s,混合时间 D8 为 0.01 s,扫描次数 NS 为 32。

1.2.5 谱图处理

测得的 ^1H -NMR 谱使用 Bruker Topspin3.2 软件处理,变换点数为 64 K,LB 为 1.00 Hz,用指数窗函数处理,基线和相位校正均采用手动方式进行,TMS 为内标信号(δ 0.00)。处理后的图谱以 δ 0.40~7.00 进行分段积分并进行面积归一化处理。

1.2.6 模型建立及评价

以图谱的分段积分值作为输入自变量,以掺假样本中所含稀奶油脂肪的相对含量作为拟合输出因变量,建立模型。按照下式计算因变量。

$$Y = aK / (aK + bJ)$$

式中: Y 为掺假样本中所含稀奶油脂肪的相对含量; a 为稀奶油样品商品标签标识的脂肪含量; b 为植脂奶油样品商品标签标识的脂肪含量; K 为掺假样本中稀奶油的质量分数; J 为掺假样本中植脂奶油的质量分数。

PCA、PLS 模型的建立:将上述所得数据进行 PCA 和 PLS 分析,分别选用中心化法(Ctr)、单位方差法(UV)的数据标度换算方式。PLS 模型采用交叉验证(CV) Q^2 值、排列实验对所建立的拟合模型

进行评判。

PCA-SVM 回归模型的建立:首先用 PCA 分析方法将训练集数据自变量降维,得到新的特征变量,交叉验证优化模型的 c 、 g 值,然后用最优参数训练 SVM 回归模型,模型类型为 epsilon-SVR,核函数为径向基(RBF)函数。用得到的模型预测测试集。模型评价:采用训练集交互验证均方根误差($RMSECV$)、测试集的判定系数(R^2)、预测均方根误差($RMSEP$)作为回归模型的评价指标。 $RMSECV$ 用于评价建模方法的可行性及所得模型的预测能力, $RMSEP$ 用于评价所建模型对外部样本的预测能力,这两个值越小,表明模型的准确度越高,预测能力越好; R^2 越接近于 1,说明期望值和预测值之间相关性越好。

2 结果与分析

2.1 稀奶油和植脂奶油 ^1H -NMR 图谱分析

图 1 为植脂奶油和稀奶油 CDCl_3 提取物的 ^1H -NMR 图谱。根据文献[14~15]报道,图谱中 δ 0.75~0.99 信号为脂肪链末端甲基质子信号, δ 1.11~1.47 信号为长链脂肪酸一般性亚甲基质子信号, δ 1.57~1.68 信号为脂肪链上与羰基相隔一个亚甲基的亚甲基上的质子信号, δ 1.92~2.14 信号为与脂肪链上双键相连的亚甲基质子信号, δ 2.26~2.38 信号为脂肪链上与羰基直接相连的亚甲基质子信号, δ 2.75~2.85 信号为脂肪链上两个双键之间亚甲基质子信号, δ 4.10~4.38 和 δ 5.24~5.29 信号分别为甘油三酯中丙三醇的亚甲基和次甲基质子信号, δ 5.30~5.43 信号为非共轭脂肪酸不饱和质子信号。

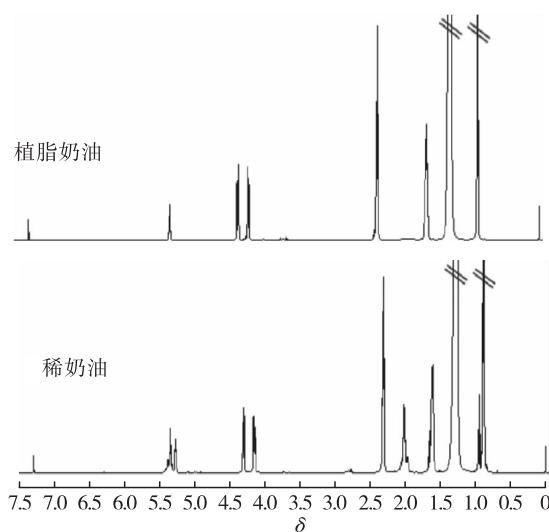


图 1 稀奶油和植脂奶油 CDCl_3 提取物的 ^1H -NMR 图谱

从图 1 可以看出,与稀奶油相比,植脂奶油的氢

谱质子信号明显缺少 δ 0.99(甘油三酯丁酸脂肪链末端甲基质子信号)、 δ 1.92~2.14(与甘油三酯脂肪链上双键相连的亚甲基质子信号)、 δ 5.30~5.43(甘油三酯非共轭脂肪酸不饱和质子信号)这3组质子信号。这是因为,稀奶油来源于天然乳脂,其脂肪酸组成与牛乳类似,含有牛乳中所特有的脂肪酸(丁酸)和较多的饱和脂肪酸;而植脂奶油现多以氢化棕榈仁油为原料,所以其中的脂肪酸以饱和脂肪酸为主。

2.2 PCA 分析

为检查定量掺假实验所用奶油样品是否存在异常样品,对25个奶油样品 CDCl_3 提取物 ^1H -NMR 谱图进行PCA分析,结果见图2。

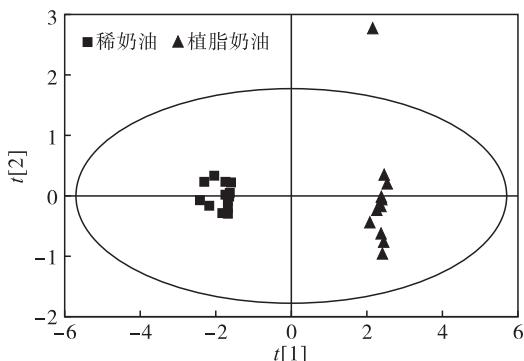


图2 稀奶油、植脂奶油 CDCl_3 提取物 ^1H -NMR 图谱
PCA 分析 PC1/PC2 得分图

由图2可见,植脂奶油组出现了一个明显远离其他样品的异常样品,代表该样品的化学成分与其他植脂奶油样品具有显著性差异。通过查阅原始 ^1H -NMR图谱并结合多种2D-NMR图谱发现,与其他植脂奶油样品相比,此样品中含有香兰素,因而在PCA得分图上会表现出异常偏离的趋势。香兰素是一种具有奶香气息的可食用香料,在奶油、糖果等食品中使用广泛,因此此样品虽然在PCA得分图中远离其他植脂奶油样品,但其却代表了一类含有香兰素的植脂奶油样本的特征,因此后续的定量掺假实验中应保留此样品,不应当做异常样本剔除。

2.3 PLS 掺假定量模型

以采用训练集数据建立PLS稀奶油中植脂奶油掺假定量分析模型,并经内部交叉验证和排列实验对所建立的拟合模型进行评判。在潜在变量为4时,交叉验证 Q^2 为0.977。在此之后,通过排列实验随机多次($n=200$)改变分类变量(y)的排列顺序得到相应不同的随机 Q^2 和 R^2 值对模型有效性做进一步的检验,结果如图3所示。

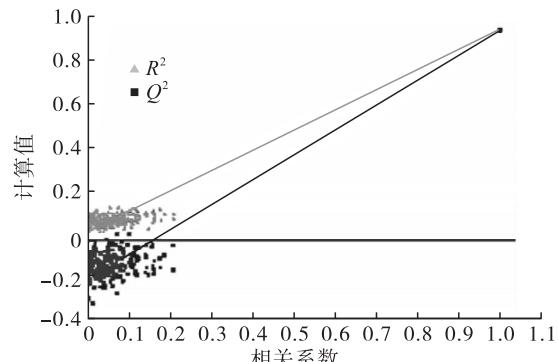


图3 PLS 掺假定量模型排列实验

由图3可知,图中所有的 Q^2 均在 R^2 之下,且 Q^2 的回归直线与 y 轴的交点在负半轴,说明该模型未过度拟合。以模型可预测指标 Q^2 并结合排列实验验证结果说明建立的定量模型成立。

2.4 PCA-SVM、SVM 掺假定量模型

PCA-SVM 掺假定量模型: 在本研究中,采用了径向基函数作为SVM的核函数,以降维得到的新的特征变量作为输入自变量,以掺假样本中所含稀奶油脂肪的相对含量作为拟合输出因变量,通过交叉验证的方法优化 c, g 值,优化后的最佳 c 值为256,最佳 g 值为0.0625,此时的 RMSECV 为3.69。以优化得到的 c, g 值作为模型参数,采用 svmtrain 函数训练降维后的训练集数据,建立 PCA-SVM 掺假定量模型。

SVM 掺假定量模型: SVM 掺假定量模型的建立过程与 PCA-SVM 模型相似,只是不进行自变量数据的降维。

2.5 PLS、SVM 与 PCA-SVM 模型建模效果、预测结果对比

分别采用建立的PLS、SVM与PCA-SVM掺假定量模型对112个测试集样本进行预测,模型预测值和实际值相关性见图4~图6。从图4~图6可以看出,SVM模型的预测性要优于PLS模型。

模型相关参数结果见表2。由表2可知:SVM模型的RMSECV 小于 PLS 模型,而训练集 R^2 大于 PLS 模型,说明 SVM 模型的拟合精度优于 PLS 模型;同时,SVM 模型的 RMSEP 小于 PLS 模型,而测试集 R^2 大于 PLS 模型,说明 SVM 模型对外部样本的预测能力高于 PLS 模型。同是 SVM 模型,经过 PCA 降维的 SVM 模型 RMSECV、RMSEP、测试集 R^2 均优于没有降维的 SVM 模型,说明 PCA 降维能够在保证原变量信息不变的前提下,将数据降维,使原来的多维问题大大简化,有效缩短运行时间,提高预测精度。

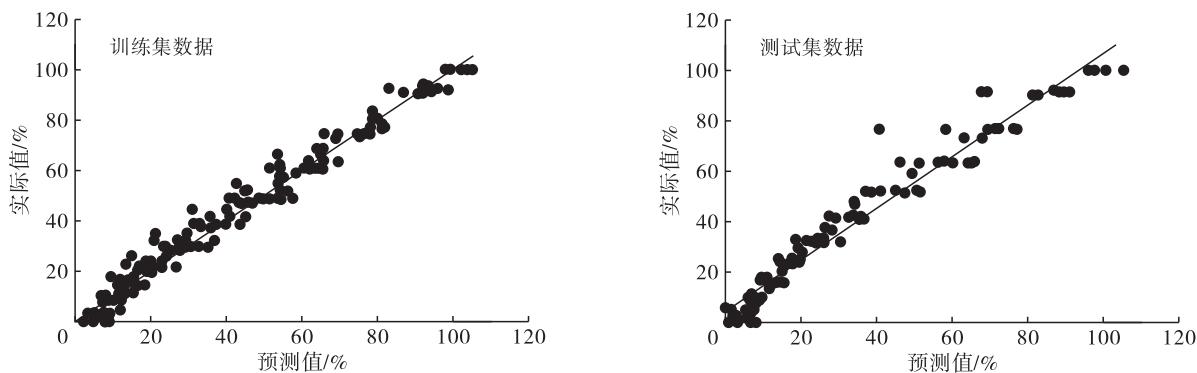


图4 PLS掺假定量模型预测值和实际值相关性

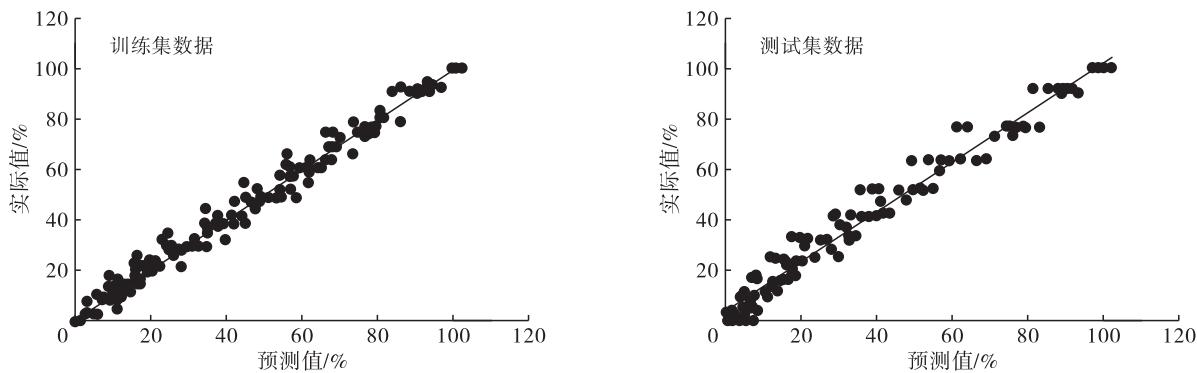


图5 SVM掺假定量模型预测值和实际值相关性

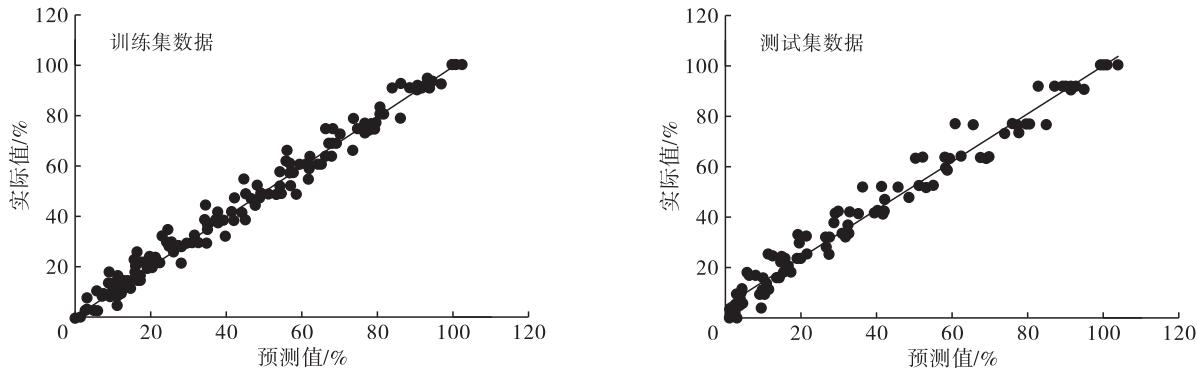


图6 PCA-SVM掺假定量模型预测值和实际值相关性

表2 PLS、SVM与PCA-SVM模型相关参数

模型	RMSECV	RMSEP	测试集 R^2	训练集 R^2
PLS	4.25	8.20	0.957 4	0.980 4
SVM	3.88	5.95	0.971 6	0.987 5
PCA-SVM	3.69	5.87	0.974 3	0.987 5

3 结论

本研究将PCA-SVM回归法与核磁共振技术相结合,利用PCA法对数据进行降维处理,建立稀奶油中植脂奶油掺假定量分析模型,并与PLS及SVM建模结果进行对比。结果表明,PCA-SVM模型具有稳定性好、预测准确度高、外推能力强及不存在过度学习现象等优点。

参考文献:

[1] GB 19646—2010, 食品安全国家标准 稀奶油、奶油和无

- 水奶油[S]. 北京: 中国标准出版社, 2010.
- [2] SB/T 10419—2007, 植脂奶油[S]. 北京: 中国标准出版社, 2007.
- [3] KRAUSE A, MIRACLE R, SANDERS T, et al. The effect of refrigerated and frozen storage on butter flavor and texture[J]. J Dairy Sci, 2008, 91(2):455–465.
- [4] GUTIERREZ R, VEGA S, DIAZ G, et al. Detection of non-milk fat in milk fat by gas chromatography and linear discriminant analysis[J]. J Dairy Sci, 2009, 92(5):1846–1855.
- [5] NTAKATSANE M, LIU X, ZHOU P. Short communication: rapid detection of milk fat adulteration with vegetable oil by fluorescence spectroscopy [J]. J Dairy Sci, 2013, 96(4):2130–2136.
- [6] HEUSSEN P, JANSEN H, SAMWEL I. The use of multi-

(下转第 114 页)

- and method for using the same: US6458379[P]. 2002 - 10 - 01.
- [25] 吴素萍. 米糠油功能特性及其制取的研究现状[J]. 江西科学, 2007, 25(4): 421 - 424.
- [26] BERNARDI D S, PEREIRA T A, MACIEL N R, et al. Formation and stability of oil - in - water nanoemulsions containing rice bran oil: in vitro and in vivo assessments [J]. J Nanobiotechnol, 2011, 9: 1 - 9.
- [27] 许仁溥. 谷维素研制三十余年历程[J]. 粮食与油脂, 2003(3): 10 - 13.
- [28] LEMUS C, ANGELIS A, HALABALAKI M, et al. γ -Oryzanol: an attractive bioactive component from rice bran [M]//Wheat Rice Disease Prevention and Health. USA: Academic Press, 2014:409 - 430.
- [29] PATEL M, NAIK S N. *Gamma*-oryzanol from rice bran oil—a review[J]. J Sci Ind Res, 2004, 63: 569 - 578.
- [30] 许仁溥, 许申. 谷维素在皮肤领域应用开发[J]. 粮食与油脂, 1995(4): 32 - 36.
- [31] 孙开进, 仓金顺. 谷维素的生产和应用[J]. 中小企业科技, 2002, 18(5): 17.
- [32] 刘瑞菊, 褚建波, 陈启云. 谷维素的研究进展[J]. 河南科技大学学报(医学版), 2015, 33(3): 237 - 240.
- [33] 陈思强, 缪丽华. 米糠油的精炼工艺及应用研究进展
- [J]. 广东化工, 2013, 40(13):104 - 105.
- [34] 潘亚萍. 米糠的开发与应用[J]. 中国油脂, 2010, 35(6): 52 - 54.
- [35] 郭天一, 罗非君. 二十八烷醇生理功能研究进展[J]. 粮食与油脂, 2017(3): 26 - 30.
- [36] 毛佳, 陈建华, 黄少烈. 二十八烷醇的研究进展[J]. 广东化工, 2007, 34(3): 56 - 57.
- [37] 刘纯友, 马美湖, 靳国锋, 等. 角鲨烯及其生物活性研究进展[J]. 中国食品学报, 2015, 15(5): 147 - 156.
- [38] 曹蕊. 米糠提取物与化妆品中的应用——米糠营养成分分析及其应用研究[D]. 江苏 无锡: 江南大学, 2008.
- [39] 王丰玲, 张英锋, 郑向美, 等. 天然维生素 E 的制备、抗氧化机理及应用前景[J]. 化学教育, 2007(12): 10 - 12.
- [40] 李志江, 戴凌燕. 米糠植酸提取的优化研究[J]. 饲料工业, 2005, 26(19): 16 - 18.
- [41] 徐浩, 张海玲, 顾广东, 等. 米糠加工及综合利用研究进展[J]. 粮油食品科技, 2017, 25(5): 37 - 41.
- [42] 李长乐, 方冬冬, 师园园, 等. 稻谷加工副产品米糠的综合利用现状分析[J]. 粮食加工, 2017(3): 27 - 30.
- [43] 朱沛沛, 王斌斌, 彭青, 等. 植酸的应用研究进展[J]. 杭州化工, 2013, 43(3): 12 - 14.

(上接第 42 页)

variate modelling of near infra - red spectra to predict the butter fat content of spreads[J]. Anal Chim Acta, 2007, 595(1/2):176 - 181.

- [7] REYHAN S, ISMAIL H, HÜSEYIN E, et al. Determination of butter adulteration with margarine using Raman spectroscopy[J]. Food Chem, 2013, 141(4):4397 - 4403.
- [8] MARIA R M, COLNAGO L A, FORATO L A, et al. Fast and simple nuclear magnetic resonance method to measure conjugated linoleic acid in beef[J]. J Agric Food Chem, 2010, 58(11):6562 - 6564.
- [9] SOPELANA P, ARIZABAleta I, IBARGOITIA M, et al. Characterisation of the lipidic components of margarines by ^1H nuclear magnetic resonance[J]. Food Chem, 2013, 141(4):3357 - 3364.
- [10] ELISABETTA S, GABRIELLA P, GIULIO C, et al. Identification of production chain of Asiagod ' Allevo

cheese by nuclear magnetic resonance and principal component analysis[J]. J Agric Food Chem, 2008, 56(16):7208 - 7214.

- [11] 李玮, 姜洁, 路勇, 等. 氢谱定量测定奶酪中总共轭亚油酸的含量[J]. 食品科学, 2015, 36(10):134 - 138.
- [12] 顾亚祥, 丁世飞. 支持向量机研究进展[J]. 计算机科学, 2011, 38(2):14 - 17.
- [13] 梅虎, 梁桂兆, 周原. 支持向量机用于定量构效关系建模的研究[J]. 科学通报, 2005, 24(6):1703 - 1711.
- [14] GUILLEN M, RUIZ A. Rapid simultaneous determination by proton NMR of unsaturation and composition of acyl groups in vegetable oils[J]. Eur J Lipid Sci Technol, 2003, 105(11):688 - 696.
- [15] SACCHI R, ADDEO F, PAOLILLO L. ^1H and ^{13}C NMR of virgin olive oil. An overview[J]. Magn Reson Chem, 1997, 35(13):S133 - S145.