

doi:10.3788/gzxb20134205.0580

## 二维相关近红外谱结合 NPLS-DA 判别 掺杂牛奶的研究

杨仁杰<sup>1,2</sup>, 刘蓉<sup>1</sup>, 徐可欣<sup>1</sup>, 杨延荣<sup>2</sup>

(1 天津大学 精密测试技术及仪器国家重点实验室, 天津 300072)

(2 天津农学院 机电工程系, 天津 300384)

**摘要:**将二维相关近红外谱与多维偏最小二乘判别分析方法结合起来,建立了掺杂牛奶与纯牛奶的判别模型.分别配置掺杂尿素牛奶(1~20 g/L)和掺杂三聚氰胺牛奶(0.01~3 g/L)样品各 40 个,采集纯牛奶及掺杂牛奶样品的近红外光谱.在量化二维相关近红外同步谱的基础上,采用多维偏最小二乘判别分析法分别建立了掺杂尿素、掺杂三聚氰胺及两种掺杂牛奶与纯牛奶的判别模型对未知样品进行判别,其判别正确率分别为 95%、90%和 92.5%,并与偏最小二乘判别和隐变量正交投影判别建模方法进行了比较.结果表明:多维偏最小二乘判别分析法具有更强的预测能力可推广到其它食品的掺杂检测中.

**关键词:**二维相关近红外谱; 多维偏最小二乘判别分析; 掺杂牛奶; 尿素; 三聚氰胺

中图分类号:

文献标识码:A

文章编号:1004-4213(2013)05-0580-6

### Discrimination of Adulterated Milk Using NPLS-DA Combined with Two-dimensional Correlation Near-infrared Spectroscopy

YANG Ren-jie<sup>1,2</sup>, LIU Rong<sup>1</sup>, XU Ke-xin<sup>1</sup>, YANG Yan-rong<sup>2</sup>

(1 State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University,  
Tianjin 300072, China)

(2 Department of Electromechanical Engineering, Tianjin Agricultural University, Tianjin 300384, China)

**Abstract:** In order to develop a rapid, cost-effective and high-throughput analysis method for detecting adulterants in milk, adulterated milk discriminant models are established by combing two-dimensional (2D) correlation near-infrared spectroscopy with multi-way partial least squares-discriminant analysis (NPLS-DA). 40 adulterated milk samples with melamine(0.01~3 g/L)and 40 adulterated milk samples with urea (1~20 g/L)are prepared respectively, then the absorption spectra of all samples are measured. Based on quantization of 2D correlation spectrum, the NPLS-DA models of urea-tainted milk, melamine-tainted milk and two types adulterated milk are constructed. The recognition rates of unknown samples are 95%, 90%, and 92.5% by calibration models individually. At the same time, PLS-DA and OPLS-DA models are established. The results show that NPLS-DA model has better predictive ability than PLS-DA model and OPLS-DA model, and this method can also be applied to other food safety detection areas.

**Key words:** Two-dimensional correlation near-infrared spectroscopy; NPLS-DA; Adulterated milk; Urea; Melamine

基金项目:国家自然科学基金青年基金(Nos. 31201359,60938002,30900275)资助

第一作者:杨仁杰(1978-),男,讲师,博士研究生,主要研究方向为光谱检测技术. Email: rjyang1978@163.com

通讯作者:刘蓉(1978-),女,副教授,博士,主要研究方向为组织光学与光谱应用. Email: rongliu@tju.edu.cn

收稿日期:2012-11-13;录用日期:2012-12-27

## 0 引言

牛奶由于能够提供人体生长发育及维持健康水平的必需营养成分,如蛋白质、脂肪、乳糖和其他固形物等,被称作“居民的血液”。但是,在巨大经济利益的驱使下,乳品的掺假现象越来越严重,牛奶在生活中的需要程度和牛奶安全问题日益突出.因此寻求一种高效快速判定牛奶掺假的方法已成为当今乳品工业的重要工作,具有重要的经济和社会价值.

在国内外,关于采用常规二维光谱的方法对牛奶掺杂研究的报道很多<sup>[1-5]</sup>,但是由于牛奶中掺杂物的多样化、微量化以及掺杂物特征峰与牛奶特征峰相互重叠等因素的影响,致使常规二维谱无法从复杂的牛奶基质中提取待分析的特征信号,无法对复杂的光谱进行解析.二维相关光谱与常规二维谱相比具有更高的分辨率,其主要体现的是随外扰变化的光谱特征<sup>[6-10]</sup>.当任意微扰作用于样品体系时,样品的各种化学组成会发生“独特的”,“选择性”的变化<sup>[6,11]</sup>.若以牛奶中掺杂物浓度为外扰,即将纯牛奶与掺杂牛奶进行相关计算,构建二维相关谱,该相关谱体现的就是牛奶中掺杂物的特征光谱信息,这些信息正是建立判别模型所需要的.

多维偏最小二乘判别分析(Multi-way Partial Least Squares Discriminant Analysis, NPLS-DA)是在偏最小二乘判别分析(Partial Least Squares Discriminant Analysis, PLS-DA)基础上发展起来的一种建模方法.相比二维的 PLS-DA, NPLS-DA 方法处理二维相关谱矩阵,不仅具有减轻特征波数选择工作量的优点,而且还能更有效地提取牛奶中掺杂物的特征光谱信息,所建模型具有较好的预测效果<sup>[12]</sup>.本文将二维相关谱技术与 NPLS-DA 方法结合起来建立掺杂尿素牛奶、掺杂三聚氰胺牛奶及两种掺杂牛奶与纯牛奶的判别模型,实现掺杂牛奶与纯牛奶的正确判别.

## 1 材料、仪器与方法

### 1.1 试剂与样品制备

实验所用的三聚氰胺为天津市赢达稀贵化学试剂厂提供;尿素为中国医药上海化学试剂公司提供.从超市购置三个品牌牛奶:蒙牛全脂灭菌纯牛奶、伊利全脂灭菌纯牛奶和海河全脂灭菌纯牛奶用于试验.用电子天平称取一定质量的掺杂物加入到纯牛奶中,搅拌、摇匀,得到不同质量浓度的掺杂牛奶.准备纯牛奶样本 80 个,按照低浓度分布密、高浓度分布稀疏的原则,分别配置含三聚氰胺质量浓度范围为 0.01-3g/L 的牛奶样本 40 个,含尿素质量浓度范

围为 1-20g/L 的牛奶样本 40 个.

### 1.2 样品光谱的采集

采用美国 PerkinElmer 公司的 Spectrum GX 傅里叶变换红外光谱仪(液氮冷却的 InGaAs 检测器).仪器参量:扫描范围为 4 000~10 000  $\text{cm}^{-1}$ ,分辨率为 4  $\text{cm}^{-1}$ ,扫描间隔为 8  $\text{cm}^{-1}$ ,扫描次数为 8.实验前,对所配置的掺杂牛奶进行均质处理.为了消除仪器漂移的影响,在测量每个样品后再采集蒸馏水的光谱作为背景,用样品光谱扣除相邻背景光谱后作为待分析的光谱数据.

### 1.3 NPLS-DA 方法与量化二维相关近红外谱

NPLS-DA 是基于判别分析基础上的多维偏最小二乘算法<sup>[12]</sup>,并且以  $Y$  变量(二进制变量,类别变量)来取代浓度变量.通过 NPLS-DA 计算相关同步谱矩阵  $X$  与类别矩阵  $Y$  之间的关系,取得  $X$  与  $Y$  的最大协方差  $\text{Cov}(X, Y)$ ,使得所预测的类贡献值的正交性最大.

在本文中,对于掺杂牛奶与纯牛奶的两类判别,  $[1 \ 0]$  表示掺杂牛奶,  $[0 \ 1]$  表示纯牛奶.对于掺杂尿素牛奶、掺杂三聚氰胺牛奶与纯牛奶的三类判别,  $[1 \ 0 \ 0]$  表示掺杂尿素牛奶,  $[0 \ 1 \ 0]$  表示掺杂三聚氰胺牛奶,  $[0 \ 0 \ 1]$  表示纯牛奶.

采用 2D shige(由日本 Kwansai Gakuin 大学 Daisuke Adachi 所编写)进行纯牛奶光谱与第  $k$  个浓度掺杂牛奶(假如有  $s$  个掺杂牛奶样本,即  $k=1, 2, \dots, s$ )光谱相关运算,得到与该浓度对应的二维相关同步谱矩阵  $\Phi_k(v_i, v_j)$ .

$$\Phi_k(v_i, v_j) = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1j} & \cdots & r_{1n} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2j} & \cdots & r_{2n} \\ r_{31} & r_{32} & r_{33} & \cdots & r_{3j} & \cdots & r_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{i1} & r_{i2} & r_{i3} & \cdots & r_{ij} & \cdots & r_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{m1} & r_{m2} & r_{m3} & \cdots & r_{mj} & \cdots & r_{mn} \end{bmatrix}$$

其中  $r_{ij}$  表示样品分别在波数  $v_i$  和  $v_j$  处的吸收相关强度的大小,  $m$  和  $n$  分别表示在两个相关波数区间内采集的吸收的波长数,因此对于  $s$  个掺杂牛奶样品可形成三维矩阵  $(s, m, n)$ ,如果在同一波数区间进行二维相关运算,则  $m=n$ .纯牛奶光谱之间进行二维相关运算,则得到纯牛奶的二维相关同步谱矩阵.而将掺杂牛奶、纯牛奶相关谱矩阵  $X$ 、类别变量  $Y$  与 NPLS-DA 结合起来,则可建立掺杂牛奶与纯牛奶的判别模型.在本文中, NPLS-DA、PLS-DA 和隐变量正交投影判别(Orthogonal Projections to Latent Structures Discriminant Analysis, OPLS-DA)建模主成分都是依据其各自的预测标准偏差

(Root Mean Square Error of Prediction, RMSEP) 最小来进行选取的。

## 2 结果与讨论

### 2.1 掺杂尿素牛奶与纯牛奶判别模型

按浓度梯度法从 40 个掺杂尿素牛奶样品中选出 30 个作为校正集, 剩余 10 个作为预测集; 对应 40 个纯牛奶样品中任意选出 30 个作为校正集, 剩余 10 个作为预测集, 建立 NPLS-DA 模型。图 1 是校正集和预测集中所有样品预测值在掺杂尿素牛奶和纯牛奶空间的散点分布图。对于两类判别: 纯牛奶与掺杂尿素牛奶, 图中已标出判别两类的分界线。纯牛奶类分界线 0.5 以上的样品为纯牛奶, 分界线以下的是掺杂尿素牛奶; 位于掺杂尿素牛奶类分界线 0.5 右侧的样品为掺杂尿素牛奶, 分界线左侧的样品为纯牛奶。

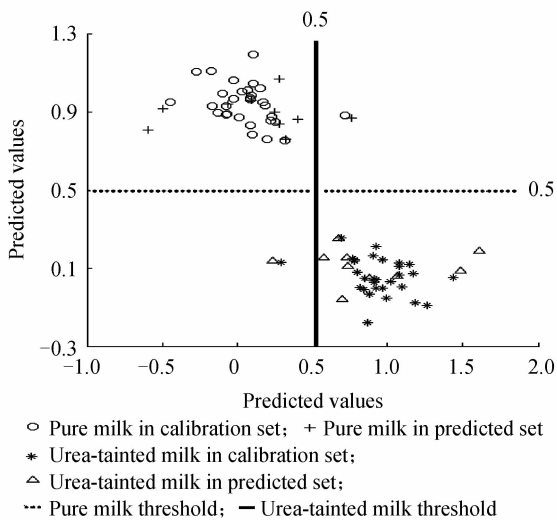
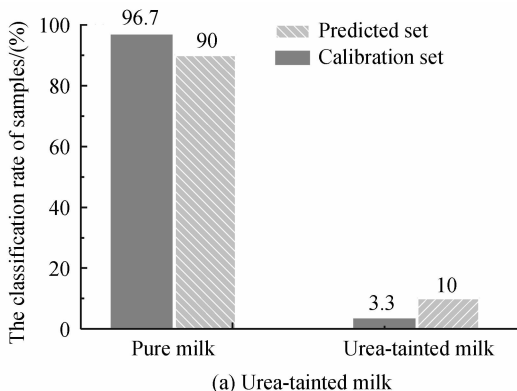


图 1 掺杂尿素牛奶与纯牛奶样品预测值散点图  
Fig. 1 Scatter plot of milk adulterated with urea and pure milk in predicted coordinate space

图 2 是 NPLS-DA 模型对掺杂尿素牛奶和纯牛奶样品正确判别与误判的百分比。图 2(a) 是掺杂尿素牛奶判别结果, 校正集和预测集样品的判别正确率分别为 96.7% 和 90%, 各有 1 个样品被误判为纯



(a) Urea-tainted milk

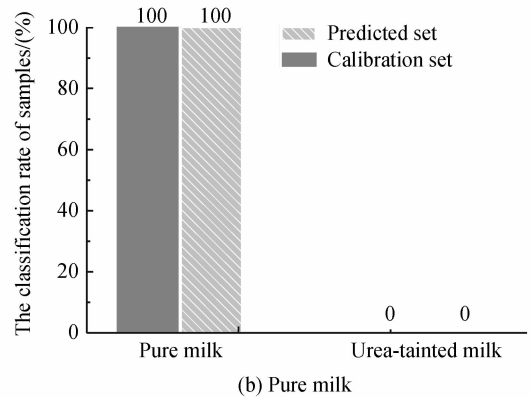


图 2 掺杂尿素牛奶与纯牛奶判别结果

Fig. 2 Discrimination results of urea-tainted milk and pure milk by NPLS-DA

牛奶, 误判率分别为 3.3% 和 10%。图 2(b) 是纯牛奶的判别结果, 校正集和预测集中的纯牛奶判别正确率均为 100%。

为了进一步说明 NPLS-DA 方法的优势, 我们对上述的校正集和预测集, 分别采用了 PLS-DA 和 OPLS-DA 方法建模, 建模参量及对未知样品判别结果见表 1。虽然 NPLS-DA 方法建模所需要的主成分数多于其它方法, 但就校正集和预测集样品的判别正确率而言, NPLS-DA 方法明显优于其它两种方法。三种方法对未知样品的正确判别率分别为 95%、90% 和 90%。

表 1 三种方法对掺杂尿素牛奶与纯牛奶的判别结果

Table 1 Discrimination results of urea-tainted milk and pure milk by NPLS-DA, PLS-DA and OPLS-DA

Models	No of factors	Discriminant accuracy	
		Training set	Test set
NPLS-DA	5	98.3%	95%
PLS-DA	4	93.3%	90%
OPLS-DA	2	98.3%	90%

### 2.2 掺杂三聚氰胺牛奶与纯牛奶判别模型

采用上述同样的方法, 选择 60 个样品 (30 个掺杂三聚氰胺牛奶, 30 个纯牛奶) 为校正集, 剩余 20 个样品 (10 个掺杂三聚氰胺牛奶, 10 个纯牛奶) 为预测集, 分别建立 NPLS-DA、PLS-DA 和 OPLS-DA 模型。图 3 是 NPLS-DA 模型对所有样品的预测值在掺杂三聚氰胺牛奶和纯牛奶空间散点分布图。纯牛奶类分界线 0.5 上方的样品属于纯牛奶, 下方的样品属于掺杂三聚氰胺牛奶; 分布在掺杂三聚氰胺牛奶类分界线 0.5 右侧的样品属于掺杂三聚氰胺牛奶, 分布在分界线左侧的样品为纯牛奶。图 4 是 NPLS-DA 模型对掺杂三聚氰胺牛奶和纯牛奶样品正确判别与误判的百分比。图 4(a) 是掺杂三聚氰胺牛奶判别结果, 在校正集中, 1 个样品被误判为纯牛

奶,判别正确率为 96.7%;在预测集中,2 个样品被误判为纯牛奶,判别正确率为 80%。图 4(b)是纯牛奶的判别结果,校正集和预测集中的纯牛奶判别正确率均为 100%。

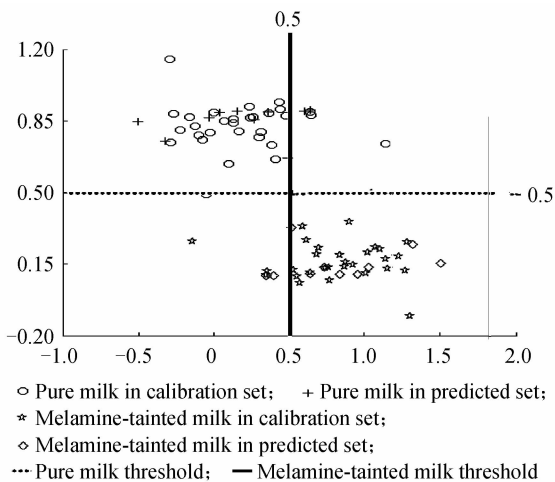


图 3 掺杂三聚氰胺牛奶与纯牛奶样品预测值散点图  
Fig. 3 Scatter plot of milk adulterated with melamine and pure milk in predicted coordinate space

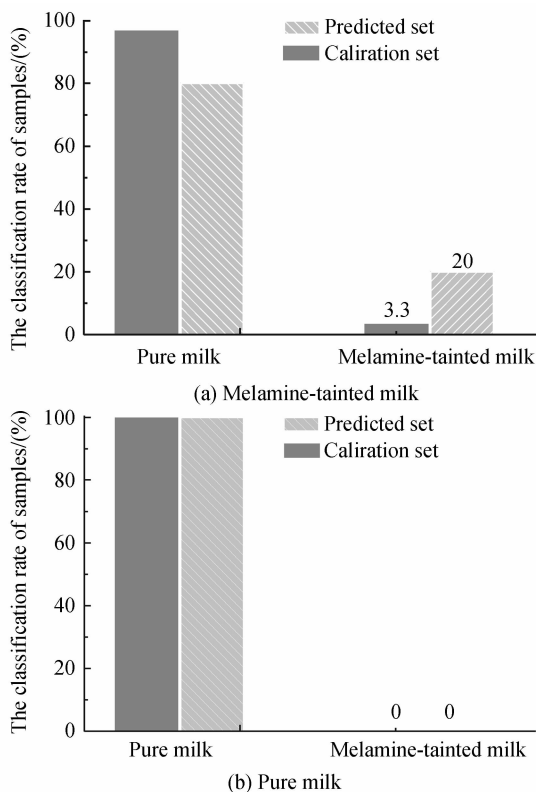


图 4 掺杂三聚氰胺牛奶与纯牛奶判别结果  
Fig. 4 Results of melamine-tainted milk and pure milk by NPLS-DA

表 2 是 NPLS-DA、PLS-DA 和 OPLS-DA 三种方法对掺杂三聚氰胺牛奶与纯牛奶的判别结果,显然,NPLS-DA 方法和 OPLS-DA 方法都取得比较好的建模效果.三种方法对未知样品判别正确率都为 90%。

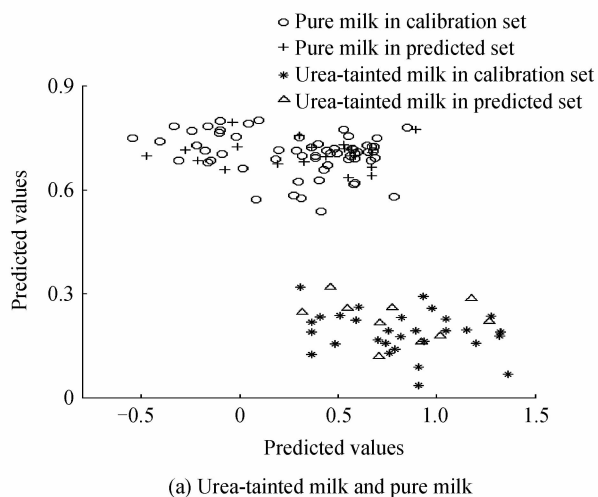
表 2 三种方法对掺杂三聚氰胺牛奶与纯牛奶判别结果  
Table 2 Discrimination results of melamine-tainted milk and pure milk by NPLS-DA, PLS-DA and OPLS-DA

Models	No of factors	Discriminant accuracy	
		Training set	Test set
NPLS-DA	9	98.3%	90%
PLS-DA	4	86.7%	90%
OPLS-DA	2	96.7%	90%

2.3 两类掺杂牛奶与纯牛奶判别模型

为了进一步实现掺杂尿素牛奶、掺杂三聚氰胺牛奶及其纯牛奶之间的判别,将上述掺杂尿素牛奶与掺杂三聚氰胺牛奶两个模型中的校正集样品合并(30 个掺杂尿素牛奶,30 个掺杂三聚氰胺牛奶,60 个纯牛奶),分别建立 NPLS-DA、PLS-DA 和 OPLS-DA 模型.利用所建模型对上述两个模型的预测集样本(10 个掺杂尿素牛奶,10 个掺杂三聚氰胺牛奶,20 个纯牛奶)进行预测。

图 5(a)、(b)分别是校正集和预测集中掺杂尿素牛奶(40 个)、掺杂三聚氰胺牛奶(40 个)与纯牛奶(80 个)预测值在各自空间的散点分布图.从图上可以看出,三类样品:掺杂尿素牛奶、掺杂三聚氰胺牛奶和纯牛奶各自聚集在一起.采用所建立的 NPLS-DA 模型,对校正集中的 120 个样品进行内部预测,其中 6 个掺杂尿素牛奶被误判为纯牛奶;8 个掺杂三聚氰胺牛奶被误判为纯牛奶;60 个纯牛奶都得到正确识别,对校正集所有样品的判别正确率为 88.3%。对预测集中的 40 个未知样品进行判别,其中 2 个掺杂尿素牛奶被误判为纯牛奶;1 个掺杂三聚氰胺牛奶被误判为纯牛奶;20 个纯牛奶都得到正确识别;对未知样品总的判别正确率为 92.5%。图 5(c)是校正集和预测集中 40 个掺杂尿素牛奶与 40 个掺杂三聚氰胺牛奶预测值在各自空间的散点图.很显然,两种掺杂牛奶各自聚集在一起,两类样品完



(a) Urea-tainted milk and pure milk

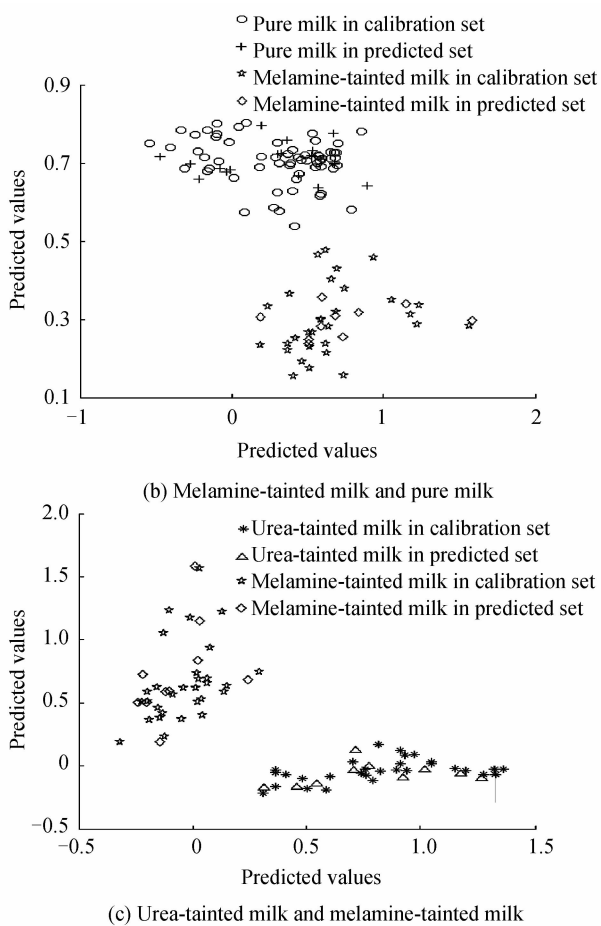


图5 掺杂尿素与纯牛奶、掺杂三聚氰胺与纯牛奶、两种掺杂牛奶预测值散点图

Fig. 5 Scatter plots of two types adulterated milk and pure milk in predicted coordinate space

全分开. 因此通过样品预测值在各自空间的散点图可以实现样品的聚类分析.

对于掺杂尿素牛奶、掺杂三聚氰胺牛奶与纯牛奶间的三类判别, 表3给出了三种方法的判别结果. 从表3中可以看出: PLS-DA方法对未知样品的判别正确率要优于OPLS-DA方法, 其原因可能是高浓度范围的掺杂尿素牛奶和低浓度范围的三聚氰胺

表3 三种方法对两类掺杂牛奶与纯牛奶判别结果  
Table 3 Discrimination results of two types adulterated milk and pure milk by NPLS-DA, PLS-DA and OPLS-DA  
A: urea-tainted milk, B: melamine-tainted milk, C: pure milk

Samples	Models	No of factors	Discriminant accuracy	
			Training set	Test set
A	NPLS-DA	5	80%	80%
B			73.3%	90%
C			100%	100%
A	PLS-DA	4	66.7%	80%
B			50%	50%
C			100%	100%
A	OPLS-DA	2	83.3%	70%
B			76.7%	0
C			70%	95%

牛奶都赋予同一Y值, 再经过OPLS方法处理后, 进一步缩小了掺杂三聚氰胺牛奶与纯牛奶之间的差别, 因此OPLS-DA模型中有更多的纯牛奶误判为掺杂牛奶或掺杂三聚氰胺牛奶误判为纯牛奶. 不过, 无论是校正集, 还是预测集, NPLS-DA的判别正确率都高于其它两种方法. 三种方法对校正集交叉验证判别正确率分别为88.3%、79.2%和82.7%, 对未知样品判别正确率分别为92.5%、84.5%和65%.

### 3 结论

本文利用二维相关谱技术提取牛奶中掺杂物的特征光谱信息, 将NPLS-DA方法用于掺杂牛奶的判定, 分别建立了掺杂尿素牛奶、掺杂三聚氰胺牛奶判别模型及三类判断模型, 这些模型对未知样品的判别正确率分别为95%、90%和92.5%. 并将NPLS-DA与PLS-DA、OPLS-DA模型判别结果进行了比较. 研究表明, NPLS-DA具有更强的预测能力, 对未知样品具有高的判别正确率, 特别是对三类判别. 该方法可推广到其它食品掺杂检测中.

#### 参考文献

- [1] KASEMSUMARAN S, THANAPASE W, KIATSOONTHON A. Feasibility of near infrared spectroscopy to detect and to quantify adulterants in cow milk [J]. *Analytical Sciences*, 2007, **23**(7): 907-910.
- [2] GU Chun-feng, LAN Xiu-feng, YU Yin-shan, et al. Fluorescence spectrum of milk solution [J]. *Acta Photonica Sinica*, 2012, **241**(1): 107-111.  
顾春峰, 兰秀凤, 于银山, 等. 牛奶水溶液的荧光光谱研究 [J]. *光子学报*, 2012, **241**(1): 107-111.
- [3] LU C H, XING B R, HAO G, et al. Rapid detection of melamine in milk powder by near infrared spectroscopy [J]. *Journal Near Infrared Spectroscopy*, 2009, **17**(2): 59-67.
- [4] MAUER L J, CHERNYSHOVA A A, HIATT A, et al. Melamine detection in infant formula powder using near- and mid-infrared spectroscopy [J]. *Journal of Agricultural and Food Chemistry*, 2009, **57**(10): 3974-3980.
- [5] BALABIN R M, SMIRNOV S V. Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: a quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder [J]. *Talanta*, 2011, **85**(1): 562-568.
- [6] NODA I. Advances in two-dimensional correlation spectroscopy [J]. *Vibrational Spectroscopy*, 2004, **36**(2): 143-165.
- [7] CZARNIK M B, MURAYAMA K, TSENKOVA R, et al. Analysis of near-infrared spectra of complicated biological fluids by two-dimensional correlation spectroscopy: protein and fat concentration-dependent spectral changes of milk [J]. *Applied Spectroscopy*, 1999, **53**(12): 1582-1594.
- [8] YANG Ren-jie, LIU Rong, XU Ke-xin. Detection of urea in milk using two-dimensional correlation spectroscopy and partial least square method [J]. *Transactions of the CSAE*, 2012, **28**(6): 259-263.  
杨仁杰, 刘蓉, 徐可欣. 二维相关光谱结合偏最小二乘法测定

牛奶中的掺杂尿素[J]. 农业工程学报, 2012, **28**(6): 259-263.

[9] SLOBODAN S, YUKIHIRO O. Wavelength-wavelength and sample-sample two-dimensional correlation analyses of short-wave near-infrared spectra of raw milk [J]. *Applied Spectroscopy*, 2001, **55**(2): 163-172.

[10] YANG R J, LIU R, XU K X. Determination of melamine of milk based on two-dimensional correlation infrared spectroscopy[C]. *SPIE*, 2012, **8229**: 822918Y.

[11] HU Xin-yao, WANG Qi, SUN Su-qin, et al. New developments of two-dimensional correlation spectroscopy [J]. *Journal of Guangxi Normal University*, 2003, **21**(2): 242-243.

胡鑫尧,王琪,孙素芹,等. 二维相关光谱技术的研究进展 [J]. 广西师范大学学报, 2003, **21**(2):242-243.

[12] DU Shu-xin, SHEN Jin-chang, YUAN Zhi-bao. Multi-way partial least squares modeling for three-dimensional fluorescence spectrometry[J]. *Laser Journal*, 2012, **33**(1): 36-37.

杜树新,沈进昌,袁之报. 三维荧光光谱的多维偏最小二乘建模方法[J], 激光杂志, 2012, **33**(1):36-37.



• 下期预告 •

# 加权非负最小二乘光子相关光谱纳米颗粒 粒径反演方法

单良,孔明

(中国计量学院 信息工程学院, 杭州 310018)

**摘要:**为了降低采用光子相关光谱法进行纳米颗粒测量时噪音对颗粒粒径反演结果的影响,提出了一种基于加权非负最小二乘法的光子相关光谱纳米颗粒粒径计算方法.该方法以光子相关光谱自身作为权值,推导出反演算法的离散模型,避免了接近零点的测量数据波动对测量结果的影响.利用光子相关光谱纳米检测实验平台对 90 nm、190 nm 及混合的乳胶颗粒进行实验研究,并与传统非负最小二乘法反演结果进行了对比.60 s 测量时间的 30 次实验数据表明:对单峰颗粒群进行反演时,该方法多次测量结果与传统非负最小二乘法结果相近,但是多次重复测量的方差较小,证明该方法重复性较好;对多峰颗粒群进行反演时,该方法反演结果更接近颗粒的真实值,而非负最小二乘法其反演结果与真实值有较大偏离.在不同测量时间的实验数据表明:测量较短的情况下,该方法反演结果方差较小,能在更短的采样时间情况下,获得更准确的测量结果.

**关键词:**光子相关光谱法; 非负最小二乘; 加权; 纳米颗粒粒径