

## 中红外光谱联合模式识别鉴别奶粉中三聚氰胺

庞佳烽<sup>1</sup>, 汤 湛<sup>1</sup>, 李艳坤<sup>1, 2\*</sup>, 徐崇然<sup>1</sup>, 卞希慧<sup>3</sup>

1. 华北电力大学(保定)环境科学与工程系, 河北 保定 071003
2. 河北省燃煤电站烟气多污染物协同控制重点实验室, 河北 保定 071003
3. 天津工业大学化学与化工学院, 天津 300387

**摘 要** 奶粉中蛋白质的含量是决定奶粉品质的重要指标。奶粉可以通过掺杂三聚氰胺以虚假提高奶粉中蛋白质的检测值, 严重危害了消费者的身体健康。采用傅里叶变换中红外光谱(FTIR)联合化学计量学的模式识别方法(模型)对奶粉中的三聚氰胺进行快速鉴别, 借助模式识别技术实现了对中红外光谱客观量化的解析, 克服了谱图比对鉴别的局限性、复杂性及主观性。分别配置纯奶粉样品和具有不同质量浓度(0.01%~0.2%)三聚氰胺的掺伪奶粉样品。扫描得到样品的中红外透射光谱数据后, 首先对原始光谱数据进行归一化预处理, 然后采用包括无监督(聚类)和有监督(判别)的多种模式识别(分类)方法进行综合分析比较。其中, 传统的主成分分析(PCA)、距离判别法(欧式距离和皮尔逊相关系数)和非负矩阵分解(NMF)无监督模式识别方法均无法准确鉴别出纯奶粉和掺三聚氰胺的奶粉样品; 采用有监督的偏最小二乘判别分析(PLS-DA)模型识别掺伪奶粉样品时, 识别灵敏度与特异度也较低; 最后采用线性判别分析方法(LDA)和非相关线性判别分析方法(ULDA), 成功地实现了含三聚氰胺掺伪奶粉的鉴别, 识别灵敏度和特异度均达到100%。尤其 ULDA 方法最大化了两类样本之间的距离, 筛选出包含最佳分类信息的特征变量, 仅用一个判别矢量便可对样本进行区分。利用 ULDA 方法进行了红外光谱重要变量(特征波长)的筛选, 考察了保留变量与识别准确率的关系, 在保留较少变量数目下实现了纯奶粉与掺三聚氰胺奶粉的鉴别, 对于奶粉中三聚氰胺的定性识别浓度可低至0.01%。因此, 提出基于中红外光谱快捷、准确地识别奶粉中掺入低含量三聚氰胺的模式识别模型, 相比传统的化学分析方法有优势, 为奶粉的掺伪识别与质量控制提供了有效的途径, 并可拓展应用到其他食品的真伪优劣的鉴别中。

**关键词** 奶粉; 三聚氰胺; 中红外光谱; 模式识别; 非相关线性判别分析

**中图分类号:** O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)10-3235-06

### 引 言

奶粉中蛋白质的含量是决定奶粉品质的一个重要指标。然而通过在奶粉中添加三聚氰胺以虚假提高奶粉中的表观蛋白质含量, 从而获取更多的利润。三聚氰胺掺杂物严重危害了消费者的身体健康和消费权益。2008年三鹿奶粉污染事件发生之后, 国家质量监督检验检疫总局颁布了《原料乳与乳制品中三聚氰胺检测方法》(GB/T 22388—2008): 高效液相色谱法、液相色谱-质谱/质谱法和气相色谱-质谱联用法; 之后卫生部又公布三聚氰胺限量值: 婴幼儿食品中为  $1 \text{ mg} \cdot \text{kg}^{-1}$ , 其他普通食品中为  $2.5 \text{ mg} \cdot \text{kg}^{-1}$ 。上述三种实

验方法虽可以精确测定奶粉中的三聚氰胺, 但是需要对奶粉进行消解、萃取等预处理, 检测手段较繁琐, 费用较高。同时也出现了一些应用近红外光谱和拉曼光谱对含有三聚氰胺的奶粉进行鉴别。其中, 近红外光谱仪器普及性相对较差, 并不适于低含量三聚氰胺的识别<sup>[1]</sup>; 而采用拉曼光谱检测时, 荧光背景干扰往往导致有些区域的拉曼光谱信号较弱, 需要采用表面增强拉曼光谱技术<sup>[2-5]</sup>。

傅里叶变换中红外光谱法(FTIR)是一种常用的分析方法, 可以对复杂体系进行整体宏观鉴定, 并且快捷、分辨率高。由于奶粉成分多样、谱峰重叠引起的谱图特征减弱; 由于生产工艺和配方的不同, 奶粉组成、颜色及颗粒大小等差异较大, 因此将化学计量学技术<sup>[6]</sup>引入中红外光谱以实现客

收稿日期: 2019-11-04, 修订日期: 2020-03-15

基金项目: 中央高校基本科研业务费(2017MS135), 国家自然科学基金项目(21305043)资助

作者简介: 庞佳烽, 1995年生, 华北电力大学(保定)环境科学与工程系硕士研究生 e-mail: jiafengpang@163.com

\* 通讯联系人 e-mail: liyankun@ncepu.edu.cn

观量化的解析,克服通过比对谱图及谱图参数(峰位、峰高、峰面积等)差异的传统鉴别方法的局限性、复杂性及主观性。本文采用计量学技术中的多种模式识别(分类)方法对奶粉及掺低含量三聚氰胺奶粉的中红外光谱进行快速识别。通过比较解析,发现采用非相关线性判别分析方法(ULDA)成功地识别了掺伪奶粉。并且利用 ULDA 方法进行了红外光谱重要变量(特征波长)的筛选,考察了保留变量与识别准确度的关系,在保留较少变量数目下实现了纯奶粉与掺三聚氰胺奶粉的鉴别,识别灵敏度和特异度均达到 100%。为奶粉的掺伪鉴别、质量控制提供了有效的途径。

## 1 实验部分

### 1.1 仪器

德国布鲁克公司傅里叶变换红外光谱仪(Tensor2.0)。

### 1.2 样品

于大型超市购买国产品牌奶粉,包括伊利(女士奶粉、全脂甜奶粉、全脂奶粉和学生奶粉);蒙牛(全脂甜奶粉、学生高钙高铁奶粉、女士高钙高铁奶粉、全家高钙高铁奶粉和多维高钙高铁奶粉);飞鹤(加锌加钙奶粉和全脂甜奶粉);贝因美健力学配方奶粉;维恩(加钙奶粉和早餐奶粉);维维(维他型豆奶粉、儿童豆奶粉和蔗糖豆奶粉);三元全家甜奶粉、完达山全脂甜奶粉、永和豆奶粉、大庆老奶粉、乐福记豆奶粉和古城全脂加糖奶粉。

样品配置:(1)纯奶粉样品:上述纯奶粉每种制样两个,共制备 46 个样品;(2)掺三聚氰胺奶粉样品:选取伊利、飞鹤、贝因美、三元、完达山品牌奶粉与三聚氰胺混合,采用逐级稀释法在纯奶粉中混合少量三聚氰胺,每种奶粉在 0.01%, 0.05%, 0.01%, 0.05%, 0.1% 和 0.2% (质量)浓度下分别制样两个,共制备 60 个样品。

### 1.3 光谱采集

使用溴化钾与奶粉进行混合研磨压片后,采用傅里叶变换红外光谱仪采集透射光谱。参数:波数范围 4 000~400  $\text{cm}^{-1}$ ,分辨率约 1.4  $\text{cm}^{-1}$ 、扫描次数 32。参比为空气,环境温度 25  $^{\circ}\text{C}$ ,每个样品重复扫描 3 次,取其平均光谱作为原始光谱(2 524 个变量)。

### 1.4 计算

采用 Kennard-Stone 分组法划分训练集与预测集。46 个纯奶粉样本中选取 23 个做训练集,其余样本做预测集;60 个掺三聚氰胺奶粉样本中选取 30 个做训练集(三聚氰胺质量浓度为 0.01%~0.2%),其余样本做预测集(三聚氰胺质量浓度为 0.01%~0.2%)。

对原始光谱数据首先进行最大最小归一化预处理(离差标准化)。通过对原数据的线性变换,将结果映射到集合[0, 1]里。

$$x_{\text{new}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \quad (1)$$

式(1)中,  $x_{\text{max}}$  是样本数据最大值,  $x_{\text{min}}$  为样本数据最小值。

使用 MATLAB R2014a 对数据进行建模分析。识别灵敏

度和特异度分别为

$$\text{SEN} = A/(A + C) \times 100\% \quad (2)$$

$$\text{SPE} = B/(B + D) \times 100\% \quad (3)$$

$$\text{ACC} = [1 - (C + D)/(A + B)] \times 100\% \quad (4)$$

式(2)~式(4)中, SEN 为灵敏度, SPE 为特异度, ACC 为准确度,  $A$  为真实掺伪样数目,  $B$  为真实纯样数目,  $C$  为预测假掺伪样数目,  $D$  为预测假纯样数目。

## 2 方法与原理

主成分分析(principal component analysis, PCA)是寻找最小均方误差意义下代表原始数据投影的方法;非负矩阵分解(non-negative matrix factorization, NMF)是用两个非负矩阵的乘积表示原一个非负矩,从而进行数据降维的方法;距离判别分析(distance discriminant analysis, DDA)是以代判样品到各总体的距离远近为判据的一种直观判别方法;偏小二乘判别分析(partial least squares discrimination analysis, PLS-DA)是将定量 PLS 用于判别分析的一种策略;线性判别分析(linear discriminant analysis, LDA)寻找分类的有效投影方向,投影后两类样本均值之间的距离尽可能大<sup>[7-9]</sup>。

非相关线性判别分析(uncorrelated linear discriminant analysis, ULDA)方法<sup>[10-12]</sup>是在线性判别分析基础上,对变换矩阵列向量间的不相关性加以考虑,以此来减少数据在降维后的冗余度。与传统 Fisher 判别分析不同, ULDA 的非相关判别矢量(uncorrelated discriminant vector, UDV)之间互不相关,因此可以保留更多信息。ULDA 可以使样本的类内距离最小化,类间距离最大化,从而有效地获取最佳分类特征。

## 3 结果与讨论

### 3.1 光谱特征分析

将纯奶粉和掺三聚氰胺奶粉按照上述实验方法进行测定,原始红外光谱分别如图 1(a)和(b)所示。图 2 中曲线  $a$ ,  $b$  分别为随机选取的纯奶粉和掺三聚氰胺奶粉(质量浓度为 0.05%)的谱图比较。从图 1 可以看出,不同品牌奶粉的红外光谱近似。1 746  $\text{cm}^{-1}$  出现油脂的羰基特征峰, 2 925  $\text{cm}^{-1}$  处为油脂的亚甲基特征峰, 2 854  $\text{cm}^{-1}$  对应油脂的甲基特征峰, 1 655  $\text{cm}^{-1}$  为奶粉中蛋白质的羰基特征峰, 1 159  $\text{cm}^{-1}$  为碳水化合物的羰基特征峰, 3 383  $\text{cm}^{-1}$  是碳水化合物的羟基特征峰<sup>[13]</sup>, 峰位均标在了图 2 中。从图 2 可以看出,纯奶粉和掺三聚氰胺奶粉光谱图的形状大体相同,并且吸光度值相近。这是由于纯奶粉和掺少量三聚氰胺奶粉所包含的主要成分复杂多样且相同、质量也相近,而中红外光谱只能识别基团特征峰,无法识别含相同基团的其他物质的干扰。因此,单纯以肉眼观察比对谱图参数(峰位、峰高等)的差异来区分不同类样本具有复杂性及主观性,很难实现客观的解析。因此,将归一化后的数据分别输入几种有监督和无监督的模式识别模型进行综合比对分析。

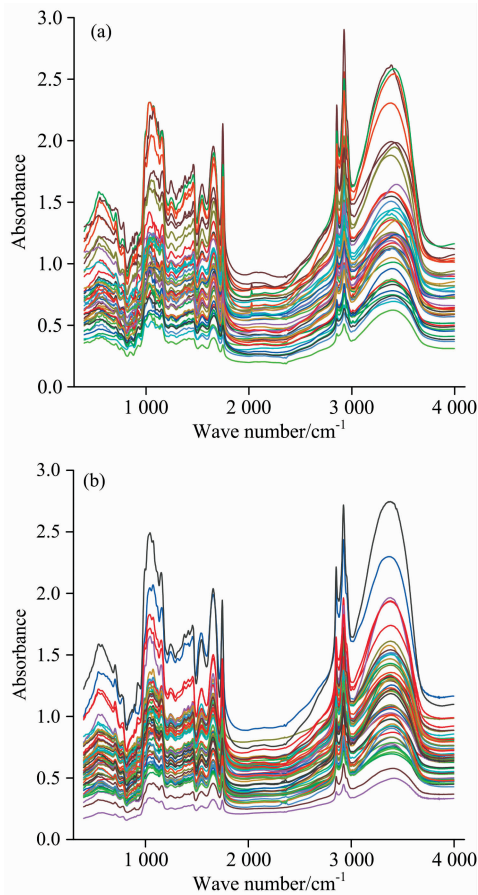


图 1 奶粉红外原始光谱叠加图

(a): 纯奶粉; (b): 掺入三聚氰胺的奶粉

Fig. 1 Raw Infrared spectra of milk powder

(a): Pure milk powder;

(b): Melamine-containing milk powder

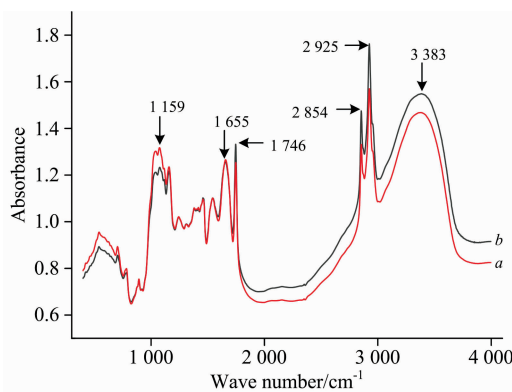


图 2 纯奶粉和掺三聚氰胺奶粉的光谱例图

a: 掺入三聚氰胺的奶粉; b: 纯奶粉

Fig. 2 Spectra examples of pure milk powder and melamine-containing milk powder

a: Melamine-containing milk powder; b: Pure milk powder

### 3.2 无监督模式识别方法

#### 3.2.1 主成分分析(PCA)和非负矩阵分解(NMF)

对光谱数据进行主成分分析,前两个主成分累计方差贡献率为 98.6%,因此将两类样本的第一和第二主成分做图,如图 3 所示。掺三聚氰胺奶粉和纯奶粉样本没有各自的聚集区域,分布杂乱。采用非负矩阵分解法,得到两类奶粉的分类图(图 4)。纯奶粉和掺三聚氰胺奶粉仍然没有各自聚集区域,两类样本交叉重叠严重,无法分开。PCA 对样品分类,有时无法识别即使是简单的线性组合特征,这可能是导致 PCA 无法对纯奶粉和掺三聚氰胺奶粉准确分类的原因。非负矩阵分解在分解过程中加入矩阵元素均为非负的约束条件,即所得矩阵中的元素都为非负,原矩阵用两个非负矩阵的乘积表示。虽然 NMF 具有收敛速度快、左右非负矩阵存储空间小等优点,但在多组分体系光谱数据降维过程中可能会导致原特征信息的丢失,给聚类分析带来困难。

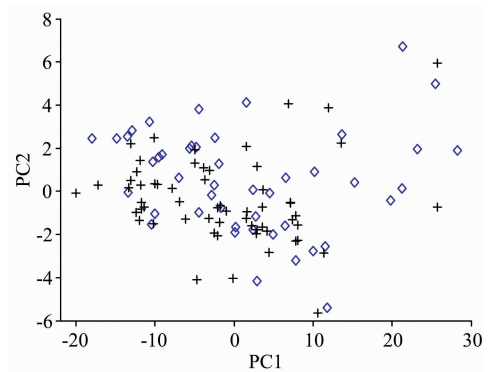


图 3 PCA 分类结果

+ : 掺入三聚氰胺的奶粉; ◇ : 纯奶粉

Fig. 3 The classification results of PCA

+ : Melamine-containing milk powder;

◇ : Pure milk powder

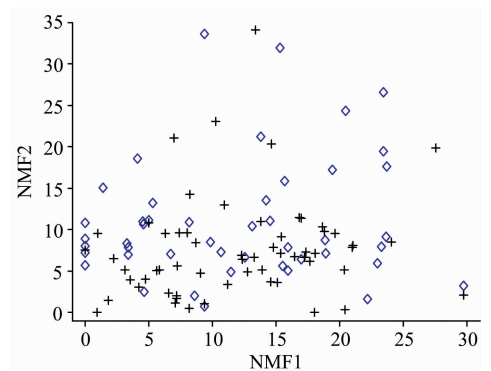


图 4 非负矩阵分解的分类结果

+ : 掺入三聚氰胺的奶粉; ◇ : 纯奶粉

Fig. 4 The classification results of NMF

+ : Melamine-containing milk powder;

◇ : Pure milk powder

#### 3.2.2 距离判别分析(DDA)

采用距离判别分析方法中的欧氏距离(Euclidean Distance)和皮尔逊相关系数(pearson correlation coefficient, PCC)方法。欧氏距离分析对纯奶粉和掺三聚氰胺奶粉样本分类的灵敏度为 78.9%(60/76)、特异度为 74.2%(46/62);

皮尔相关系数方法对两类样品的分类的灵敏度为 71.4% (60/84)、特异度为 65.7%(46/70)。聚类结果表明,这两种距离未能有效地体现出纯奶粉、掺三聚氰胺奶粉之间的真实差异。因此,无法对奶粉进行准确分类。

以上无监督聚类方法均未能实现两类样本的区分识别,因而采用有监督判别方法进行分析。

### 3.3 有监督模式识别方法

#### 3.3.1 偏最小二乘判别分析(PLS-DA)

纯奶粉标记为 [1, 0], 掺三聚氰胺奶粉标记为 [0, 1], 与对应光谱之间建立 PLS-DA 模型, 然后对预测集分类, 结果如图 5(a,b)所示。纯奶粉中有 12 个样品分类错误, 掺三聚氰胺奶粉有 10 个样品分类错误。预测灵敏度为 71.4% (30/42), 特异度为 69.7%(23/33)。PLS-DA 应变量反映自变量的隶属类型, 通过寻找光谱矩阵和样品分组信息的最大协方差, 从而在新的低维坐标系中对样品重新排序。PLS-DA 虽然可以减少变量间多重共线性产生的影响, 但自变量相关程度过高或难以拟合非线性, 可能导致模型丢失最佳分类特征。

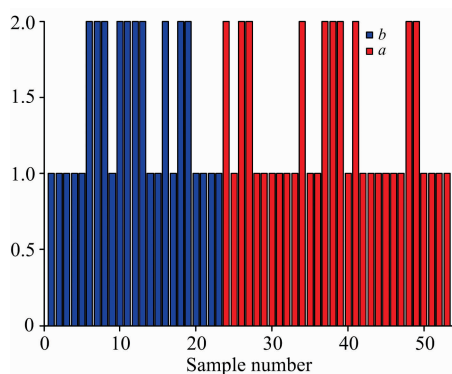


图 5 偏最小二乘判别分析分类结果

(a): 掺入三聚氰胺的奶粉; (b): 纯奶粉

Fig. 5 The classification results of PLS-DA

(a): Melamine-containing milk powder;

(b): Pure milk powder

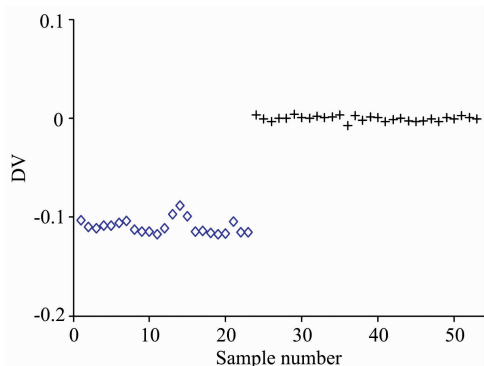


图 6 线性判别分析分类结果

+ : 掺入三聚情胺的奶粉; ◇: 纯奶粉

Fig. 6 The classification results of LDA

+ : Melamine-containing milk powder;

◇: Pure milk powder

#### 3.3.2 线性判别分析(LDA)和非相关线性判别分析(ULDA)

LDA 方法用训练集建模, 对预测集样本进行识别, 结果如图 6 所示。掺假奶粉和纯奶粉的判别矢量(discriminant vector, DV)值有差异, 能很好地将两者分开。

同样采用 ULDA 对预测集样本进行分类, 如图 7 所示。从非相关判别矢量(UDV)可以发现, 纯奶粉、掺三聚氰胺奶粉通过 ULDA 算法得到了很好的区分, 具有 100%的灵敏度和 100%的特异性。对比图 6 发现, ULDA 使两类样本间的距离(类间距)更远, 分类效果更好。由于 ULDA 在 Fisher 判别准则基础上, 转换矩阵中的任意两个列向量相互需满足“S-正交”, 这也是 LDA 与 ULDA 的主要区别。在这一约束下, 原始数据转换成的新的矩阵中的变量(UDV)是不相关的, 这对于特征提取非常有利, 可以使新变量的信息冗余度最小。其中的线性转换矩阵为  $G$ , 图 8 给出了光谱全部 2 524 个变量的  $G$  的绝对值图。 $G$  可以看作 ULDA 的载荷,  $G$  绝对值越大, 相应变量对分类的重要性越高。但保留的变量数目过少, 在摒弃冗余信息的同时, 会导致携带有效信息变量的丢失, 因此考察了预测准确度随着保留变量数目的变

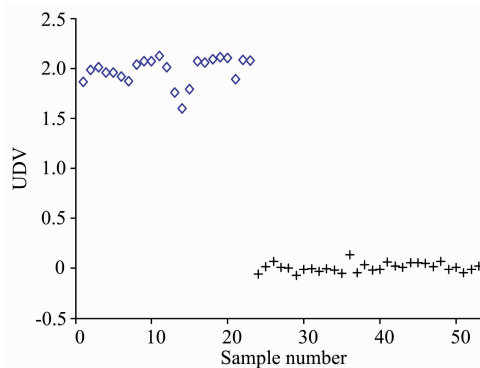


图 7 非相关线性判别分析分类结果

+ : 掺入三聚氰胺的奶粉; ◇: 纯奶粉

Fig. 7 The classification results of ULDA

+ : Melamine-containing milk powder;

◇: Pure milk powder

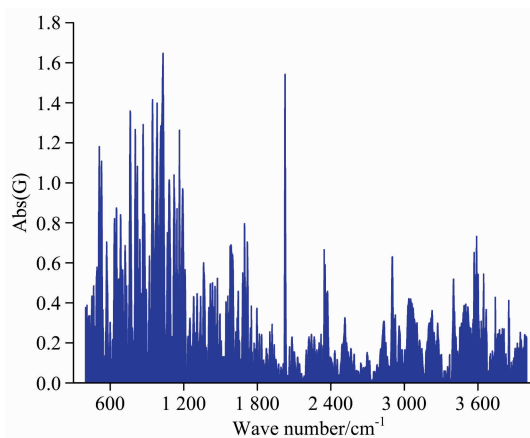


图 8 ULDA 的转换矩阵图

Fig. 8 The transformation vector plot for ULDA

化情况。考察了  $G$  绝对值从 0.68~1.34(间隔 0.03)取值对应保留变量数目在 1 189~592(占原变量数目的 47.1%~23.5%)范围内所得到的识别准确率值,发现预测准确度大体随变量数目减少呈现下降趋势,准确度逐渐从 100%降低到 52.8%。而且当只有两类样本时,ULDA 仅用一个判别矢量实现了对样本的区分,结果直观。ULDA 算法为准确鉴别低含量三聚氰胺奶粉提供了一种非常有效的途径。

## 4 结 论

采用包括无监督(聚类)和有监督(判别)的多种模式识别

方法对奶粉及含低含量三聚氰胺奶粉的中红外光谱进行综合分析比较,提出快捷、准确地定性识别奶粉中低含量三聚氰胺的最优 ULDA 模型,识别灵敏度和特异度均达到 100%。并且利用 ULDA 方法挑选了红外光谱中表征分类信息的特征变量,考察了保留变量与识别准确度的关系,在保留较少特征变量数目下实现了纯奶粉与掺三聚氰胺奶粉的鉴别区分。为奶粉的掺伪识别提供了非常有效的途径,并有望拓展应用到其他食品真伪优劣的筛查中。

## References

- [1] ZHANG Qing-qing, SHEN Xiao-fang, CHAI Jun-yu, et al(张青青,沈晓芳,柴俊宇,等). The Food Industry(食品工业), 2018, 39(12): 135.
- [2] Karunathilaka S R, Farris S, Mossoba M M, et al. Food Additives and Contaminants Part A: Chemistry Analysis Control Exposure & Risk Assessment, 2016, 34(2): 170.
- [3] Dingari N C, Barman I, Saha A, et al. Journal of Biophotonics, 2013, 6(4): 371.
- [4] Ma P, Liang F, Sun Y, et al. Microchimica Acta, 2013, 180(11-12): 1173.
- [5] Dhakal S, Chao K L, Qin J W, et al. Journal of Food Measurement and Characterization, 2016, 10(2): 374.
- [6] Shao X G, Cui X Y, Yu X M, et al. Talanta, 2018, 183: 142.
- [7] Heider E C, Mujumdar N, Campiglia A D. Analytical and Bioanalytical Chemistry, 2016, 408(28): 7935.
- [8] Liu P, Wang J, Li Q, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 206: 23.
- [9] Gupta A, Barbu A. Pattern Recognition, 2018, 78: 215.
- [10] Li Y K, Ma X P, Huang K N, et al. Indian Journal of Biochemistry & Biophysics, 2019, 56: 53.
- [11] LIANG Yi-zeng, XU Qing-song(梁逸曾,许青松). Instrumental Analysis of Complex Systems-With, Gray and Black Analytical Systems and Their Multivariate Methods(复杂体系仪器分析—白、灰、黑分析体系及其多变量解析方法). Beijing: Chemical Industry Press(北京:化学工业出版社), 2012. 525.
- [12] Li Y K, Zeng X C. Analytical Methods, 2016, 8: 183.
- [13] Chen J B, Zhou Q, Sun S Q. Journal of Molecular Structure, 2016, 1124: 262.

# Identification of Melamine in Milk Powder by Mid-Infrared Spectroscopy Combined With Pattern Recognition Method

PANG Jia-feng<sup>1</sup>, TANG Chen<sup>1</sup>, LI Yan-kun<sup>1, 2\*</sup>, XU Chong-ran<sup>1</sup>, BIAN Xi-hui<sup>3</sup>

1. North China Electric Power University, Department of Environmental Science and Engineering, Baoding 071003, China

2. Hebei Key Lab of Power Plant Flue Gas Multi-Pollutants Control, Baoding 071003, China

3. School of Chemistry and Chemical Engineering, Tiangong University, Tianjin 300387, China

**Abstract** The content of protein in milk powder is an important factor to determine the quality of milk powder. The value of protein in milk powder can be falsely increased by adulterating melamine, which endangers the health of consumers seriously. In this study, melamine in the milk powder was identified quickly by Fourier transform mid-infrared (FTIR) spectroscopy combined with pattern recognition method (model). The objective analysis of the infrared spectrum is achieved with the help of pattern recognition technology, which overcomes the limitation, complexity and subjectivity of identification through spectra comparison. Pure milk powder samples and adulterated milk powder samples with different mass concentrations (0.01%~0.2%) of melamine were prepared respectively. After collecting the mid-infrared transmission spectrum data of the samples, the original data were firstly normalized and then comprehensively analyzed using multiple pattern recognition (classification) models including unsupervised (clustering) and supervised (discriminating) methods. Among them, traditional principal component analysis (PCA), distance discrimination (Euclidean distance and Pearson correlation coefficient) and non-negative matrix factorization

(NMF) unsupervised pattern recognition methods can not accurately identify pure milk powder and melamine-containing milk powder samples. The recognition sensitivity and specificity of the supervised model of partial least squares discriminant analysis (PLS-DA) method were also low. Finally, linear discriminant analysis (LDA) and non-correlated linear discriminant analysis (ULDA) were used, the identification of melamine-containing milk powder was successfully achieved, and the sensitivity and specificity of recognition were both 100%. In particularly, the ULDA method maximizes the distance between the two types of samples, selects the feature variables containing the best classification information, and distinguishes the samples with only one discriminant vector. Furthermore, the ULDA method was used to select the important variables (characteristic wavelengths) of the infrared spectra. The relationship between the retained variables and recognition accuracy was investigated. The identification of pure milk powder and melamine-containing milk powder was realized with fewer variables. The recognition concentration of melamine in milk powder can be as low as 0.01%. Therefore, it is proposed to identify a small amount of melamine in the milk powder with simple and rapid way based on MIR spectroscopy, which has advantages over traditional chemical analysis methods and provides an effective way for the adulteration recognition and quality control of milk powder. The method is expected to extend to apply in the identification of other foods.

**Keywords** Milk powder; Melamine; Mid-infrared spectroscopy; Pattern recognition; Uncorrelated linear discriminant analysis

(Received Nov. 4, 2019; accepted Mar. 15, 2020)

\* Corresponding author

---

## 敬告读者——《光谱学与光谱分析》已全文上网

从 2008 年第 7 期开始在《光谱学与光谱分析》网站([www.gpxygpfx.com](http://www.gpxygpfx.com))“在线期刊”栏内发布《光谱学与光谱分析》期刊全文,读者可方便地免费下载摘要和 PDF 全文,欢迎浏览、检索本刊当期的全部内容;并陆续刊出自 2004 年以后出版的各期摘要和 PDF 全文内容。2009 年起《光谱学与光谱分析》每期出版日期改为每月 1 日。

《光谱学与光谱分析》期刊社