

doi:10.6041/j.issn.1000-1298.2015.S0.021

近红外多组分分析中异常样本识别方法*

尹宝全^{1,2} 史银雪¹ 孙瑞志¹ 王文狄¹

(1. 中国农业大学农业部农业信息获取技术重点实验室, 北京 100083; 2. 中国农业大学烟台研究院, 烟台 264670)

摘要: 近红外光谱分析中,异常样本的存在严重影响定标模型的预测性能和适配性。基于 X/Y 联合的 ODXY 异常样本识别和剔除方法,提出并证明了一种专用于多组分分析的 MODXY 异常样本识别方法。实验采用 80 组玉米近红外光谱数据,利用不同异常样本识别方法剔除异常样本后建立玉米含水率、含油率、蛋白质含量和淀粉含量 4 种组分的偏最小二乘预测模型,采用预测均方差和决定系数作为评价指标比较所建模型的性能,检验 MODXY 方法在多组分分析中的异常样本识别能力。实验结果表明:在近红外多组分分析中,MODXY 方法在大多数情况下具有更好的异常样本识别能力;MODXY 方法和 ODXY 方法均有一定的适用范围,它们更适用于相应组分化学值的相对标准偏差较大的情况。

关键词: 异常样本 近红外光谱 多组分分析 马氏距离 偏最小二乘回归

中图分类号: S132;O657.3 **文献标识码:** A **文章编号:** 1000-1298(2015)S0-0122-06

Outlier Samples Detection Method for NIR Multicomponent Analysis

Yin Baoquan^{1,2} Shi Yinxue¹ Sun Ruizhi¹ Wang Wendi¹

(1. Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture, China Agricultural University, Beijing 100083, China 2. Yantai Academy, China Agricultural University, Yantai 264670, China)

Abstract: Near infrared spectroscopy is currently a highly versatile tool used in diverse fields. However, outlier samples strongly affect the performance of the prediction model in near infrared spectroscopy analysis. Therefore, detecting and eliminating the outlier samples is a major and important procedure in near infrared spectroscopy analysis. Using the outlier samples detection based on joint $X-Y$ distances (ODXY) method, a special outlier samples detection method for multicomponent analysis was proposed and proved, termed as MODXY method. Experimental data was derived from the near infrared spectra of 80 corns. Based on these, the PLS models of moisture content, oil content, protein content and starch content were constructed by eliminating outlier samples using different outlier detection methods. The obtained models were compared in terms of performance by the predictive root mean square error (RMSEP) and the coefficient of determination (R^2). The results showed that in most cases the MODXY method had better outlier sample recognition capability in NIR multicomponent analysis compared with other methods. Both ODXY method and MODXY method had their own suitable range, and they were more effective when the relative standard deviation of components was large enough.

Key words: Outlier samples NIR spectroscopy Multicomponent analysis Mahalanobis distance PLS regression

引言

近红外光谱 (Near infrared spectroscopy, NIR)

分析技术具有快速、无损、可多组分同时检测、易于在线分析等特点,被广泛应用于农产品、食品安全等多个领域^[1-3]。另外,NIR 光谱具有信噪比低、倍

收稿日期: 2015-10-28 修回日期: 2015-11-19

* 新疆生产建设兵团科技支疆计划资助项目(2014AB037)

作者简介: 尹宝全, 博士生, 中国农业大学烟台研究院讲师, 主要从事农业信息化技术、数据挖掘算法研究, E-mail: yinbaoquan@126.com

通讯作者: 孙瑞志, 教授, 博士生导师, 主要从事计算机网络及应用、农业信息化技术研究, E-mail: sunrz_cn@sina.com.cn

频、谱峰重叠严重等特点,如何从复杂、重叠、变动的背景中提取弱信号是 NIR 光谱分析的难点,异常样本的存在进一步加大了光谱分析的难度。因此,有效地识别和剔除异常样本是建立定标模型之前的一个重要步骤。

所谓异常样本是指在建模过程中加入某些样本后导致所建模型的预测精度将大幅降低,这些样本称为异常样本。常用的异常样本识别方法是根据该样本的化学值 Y 或光谱值 X 是否严重偏离样本的总体分布进行判别,将偏离总体分布超过设定范围的值视为异常值。根据异常值的不同,可分为以下两种情形: X 或 Y 异常; X 和 Y 均异常。其中,第 1 类样本值异常的原因包括实验误操作、仪器异常、波长漂移、环境波动等因素的影响,应视为异常样本,建模时需剔除;第 2 类样本值异常的原因除包括上述原因外,还可能由于该样本本身属性的不同导致化学值和光谱值与其他样品相比具有特殊性。对于前者,该样本应视为异常样本进行剔除;对于后者,建模时加入这些样本虽然可能会降低模型对一般样本的预测精度,但可显著提高模型的适配范围^[4],该样本应予以保留。

NIR 光谱分析中异常样本的识别方法大多只考虑 NIR 单一组分分析的情况。多组分分析时,则将每一组分分别与光谱关联分析或只考虑光谱的异常情况^[5-7]。专门针对多组分分析的异常样本识别方法未见报道。本文从目前常用的异常样本识别方法入手,探究一种专门针对多组分同时分析时的异常样本识别方法,并通过实验进行验证。

1 材料与方法

1.1 实验样本集

实验采用 Eigenvector Research, Inc. 发布的玉米 NIR 光谱分析数据集,该数据集可从 <http://www.eigenvector.com/data/Corn/corn.mat> 下载。实验选用该数据集中利用 M5 光谱仪测量的 80 个玉米样本的 NIR 光谱及其相应玉米样本的含水率、含油率、蛋白质含量和淀粉含量 4 种组分的化学值进行分析。

数据集中所用 M5 型光谱仪的扫描范围为 1 100 ~ 2 498 nm,扫描间隔为 2 nm,每条光谱 700 个采样点。

数据集中玉米样本的近红外反射光谱如图 1 所示。

1.2 异常样本识别方法

1.2.1 常用的异常样本识别方法

常用的异常样本识别方法通常通过判断样本的

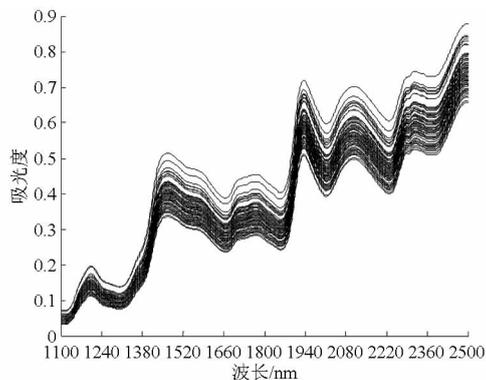


图 1 玉米样本的近红外光谱

Fig. 1 NIR spectra of corn samples

光谱值 X 或化学值 Y 是否异常进行判别,如马氏距离 (Mahalanobis distance, MD) 法^[8-9]、半数重采样 (Resampling by half-mean, RHM) 法^[10]、杠杆值与学生 T 检验法等。MD 法的优点是算法简单,运行速度较快,对第 1 类样本值异常具有较好的识别能力,其缺点是将第 2 类样本值异常的情形均视为异常样本,易造成某些边界样本被误剔除,最终影响所建模型的适配范围。为此,蒙特卡洛采样 (Monte Carlo sampling, MCS) 法^[11-12]综合考虑了 X 、 Y 之间的关联联系,利用预测残差对异常样本敏感的特性,在一定程度上实现了对第 2 类样本值异常的正确划分。缺点是算法中需多次对样本集进行划分并进行建模,使算法运行速度较慢。

为实现对第 2 类情形的正确判别,并提高算法的运行速度,提出了一种基于 X/Y 联合的异常样本识别和剔除 (Outlier samples detection based on joint X - Y distances, ODXY) 方法。

为方便描述,定义如下: $X(n \times m)$ 为光谱数据。 n 为样本数; m 为每条光谱的采样点数; $Y(n)$ 为样本的化学值。

设第 i 个样本的光谱为一维向量 X_i ;其化学值为 Y_i ,且存在线性关系

$$Y_i = AX_i + b \quad (1)$$

对所有样本取平均得

$$\bar{Y} = A\bar{X} + b \quad (2)$$

其中, \bar{X} 为平均光谱; \bar{Y} 为所有样本化学值的平均值。

由式(1)、(2)可得

$$Y_i - \bar{Y} = A(X_i - \bar{X}) \quad (3)$$

对于单组分分析, $Y_i - \bar{Y}$ 是一个数,式(3)两边可同除以 $Y_i - \bar{Y}$ 得

$$\frac{A(X_i - \bar{X})}{Y_i - \bar{Y}} = 1 \quad (4)$$

两边左乘 A^{-1}

$$\frac{X_i - \bar{X}}{Y_i - \bar{Y}} = A^{-1} = C \quad (5)$$

其中, C 为常数。

两边取 L2 范数, 并设 $dY_i = \|Y_i - \bar{Y}\|$; $dX_i = \|X_i - \bar{X}\|$, 有

$$\frac{dX_i}{dY_i} = C \quad (6)$$

式(6)的物理意义为: 近红外光谱分析中, 任意样本的光谱到平均光谱的距离与该样本的化学值到其平均值的距离成正比。因此, 可将 dX_i/dY_i 的值超出设定范围的样本视为异常样本, ODXY 法就是基于以上原理设计的。

ODXY 法是联合考虑了化学值 Y 与光谱值 X 之间的关系提出的, 可较好地解决第 2 类异常样本的识别问题。与 MCS 方法相比, ODXY 法具有一定的理论基础, 且算法简单, 运行速度更快。

1.2.2 多组分分析的异常样本识别方法

由上节所述可见, 无论传统的异常样本识别方法还是 ODXY 识别方法均没有考虑一组光谱同时用于多个组分化学值的分析(即多组分分析)时各化学值和光谱值之间的关联关系。在多组分分析中, 一组光谱同时包含所有组分化学值的信息, 各组分的化学值和该光谱之间存在着必然的联系。如果在多组分分析时, 综合考虑它们之间的关系, 将有可能提高异常样本辨别能力。

对于多组分分析, 定义如下: $Y(n \times l)$ 为多组分分析时样本对应的化学值。其中, l 为每条光谱可同时分析的组分数。例如, 实验样本集中的玉米近红外光谱被同时用于玉米的含水率、含油率、蛋白质含量和淀粉含量 4 种组分的分析, 则组分数 $l = 4$ 。

设第 i 个样本的光谱为一维向量 X_i ; 该样本的第 j 个组分的化学值为 $Y_i(j)$, 其中, $j = 1, 2, \dots, l$; 化学值 $Y_i(j)$ 与光谱值 X_i 之间的关系为

$$Y_i(j) = A_j X_i + b_j \quad (7)$$

根据 ODXY 法的原理推导, 可得

$$\frac{dX_i}{dY_i(j)} = C_j \quad (8)$$

其中, C_j 为常数; $dY_i(j) = \|Y_i(j) - \bar{Y}(j)\|$; $\bar{Y}(j)$ 为第 j 个组分化学值的平均值。

式(8)两边取倒数, 得

$$\frac{dY_i(j)}{dX_i} = \frac{1}{C_j} \quad (9)$$

因为各组分对应于同一光谱, 所以 dX_i 与 j 无关。对式(9)两边按 j 进行求和得

$$\sum_{j=1}^l \frac{dY_i(j)}{dX_i} = \sum_{j=1}^l \frac{1}{C_j} = C_m \quad (10)$$

其中, C_m 为常数。

式(10)的物理意义为: 近红外多组分分析中, 任意一个样本所对应的所有组分的化学值到其平均值间的距离之和与该样本光谱到平均光谱之间的距离成正比。

基于此, 本文提出了一种专用于多组分分析的异常样本识别和剔除方法 (ODXY method for multicomponent analysis, MODXY)。

1.3 数据处理与分析方法

为检验 MODXY 方法在多组分分析中对异常样本的识别能力和适配范围, 实验中采用以下数据处理过程和分析方法。

(1) 光谱预处理方法的选择: 由于高频噪声、基线漂移、散射光等因素的影响使 NIR 光谱的信噪比较低。要建立稳定、可靠的预测模型, 需首先对光谱数据进行预处理, 不同的数据预处理方法也将影响所建预测模型的精确度和适配范围。由于本次实验目的是实现异常样本识别方法的对比, 实验中选择相同的 S-G 平滑 + 标准正态变量变换 (SNVT) + 去趋势化处理的预处理方法。

(2) 异常样本识别和剔除方法的选择: 传统的异常样本识别方法虽未考虑各组分与光谱间的关联, 但也可用于多组分分析, 只是将各组分与光谱分别独立考虑。为检验 MODXY 方法在多组分分析中的性能, 本文选用有代表性的 MD 法、ODXY 法与 MODXY 法进行对比实验。

(3) 训练集和验证集划分方法的选择: 数据建模过程中需将样本集采用一定的方法划分为训练集和验证集。常用的划分训练集和验证集的方法有随机 (RS) 法、KS 法^[13] 和 SPXY 法等。SPXY 算法在联合考虑了光谱值和化学值的基础上对数据集进行划分, 该方法划分出的训练集更具代表性, 所建模型的泛化能力更强^[14]。因此, 实验中样本集的划分采用 SPXY 法。

(4) 建模算法的选择: 偏最小二乘法 (Partial least squares regression, PLS) 克服了光谱数据间的高度相关性, 较好地解决了样本个数少于变量个数等问题, 是光谱分析中最常用的建模方法^[15-18]。实验选用 Matlab 2013 提供的 PLS 算法进行建模。

(5) 异常样本识别方法的评价指标选择: 为实现不同异常样本识别方法的对比, 首先在预处理后的训练集上建立玉米含水率、含油率、蛋白质含量和淀粉含量的 PLS 预测模型。然后, 利用该预测模型对验证集的样本进行检验, 通过比较验证集中样本各组分的预测值与实测值之间的关系, 计算预测均方差 (Root mean square error of prediction, RMSEP)

和预测决定系数 R^2 , 以 RMSEP 和 R^2 作为评价指标, 实现对异常样本识别方法的性能比较。通常具有较小 RMSEP 和较大 R^2 的预测模型所采用的异常样本识别方法具有更好的辨别能力^[19]。

以上所有算法均通过 Matlab 2013 编程实现。

2 结果分析与讨论

2.1 多组分分析中基于 X/Y 变量联合的异常样本识别方法实现

根据多组分分析中各组分化学值与光谱之间的关系, MODXY 异常样本识别方法实现如下:

(1) 首先对近红外光谱数据采用主成分分析 (Principal components analysis, PCA) 进行降维。

(2) 计算各光谱到平均光谱的距离 dX_i 。

(3) 计算各组分化学值到所有化学值平均值之间的差值 dY_i 。

(4) 计算各样本 dX_i/dY_i 的值 D_i 。

(5) 判别样本是否正常的阈值范围计算公式为

$$D_i = \bar{D} \pm e\sigma \quad (11)$$

式中 \bar{D} —— D_i 的平均值

σ —— D_i 的标准差

e ——调整阈值范围的权重系数

将 D_i 超出阈值范围的样本判定为异常样本。通过调整权重系数 e 的值, 可以改变异常样本识别的灵敏度。 e 值过小可能导致正常样本被误判为异常, e 值过大则可能导致某些异常样本未被正常识别。在进行异常样本识别和剔除时, 应通过设定不同 e 值后建立 PLS 模型, 根据对比 PLS 模型的预测结果选取最佳 e 值。

另外, 为消去光谱距离 dX_i 和化学值的距离 dY_i 之间因权重不同造成的影响, 可分别对其作除以各自最大值的处理; 为消除异常值的影响, 计算阈值范围时可用中位数替代平均值。

2.2 不同异常样本识别方法的对比实验结果

分别利用 MD、ODXY、MODXY 法剔除异常样本, 然后采用 1.3 节中所述的数据处理与分析方法对剔除异常样本后的玉米样本数据建立含水率、含油率、蛋白质含量和淀粉含量的 PLS 预测模型。其中, MODXY 法综合考虑了所有组分与光谱之间的关联关系, 所以对于所有组分, MODXY 法所识别异常样本的结果相同; MD 法只考虑光谱的异常情况, 对于不同组分所识别异常样本的结果也相同; 但对于 ODXY 法而言, 需根据含水率、含油率、蛋白质含量、淀粉含量和光谱之间的关系分别判别, 所得异常样本的结果也不尽相同。

为检验异常样本识别方法在不同数据集中的性

能, 本实验在选定 SPXY 法作为样本集划分方法的前提下, 通过改变训练集与验证集的样本数量划分比例来实现样本集的改变。实验中训练集与验证集的样本数量划分比例分别采用 3:1、4:1、5:1、6:1、7:1、8:1、9:1 和 10:1。得到不同的训练集后, 分别建立不同训练集下的玉米含水率、含油率、蛋白质含量和淀粉含量的 PLS 预测模型, 计算预测均方差 (RMSEP) 和决定系数 R^2 。然后分别利用 RMSEP 和 R^2 作为评价指标, 实现对不同异常样本识别方法的性能比较, 如图 2~5 所示。图中, 横坐标是为获得不同训练集所采用的样本数量划分比例; 纵坐标是分别采用 MD、ODXY、MODXY 法剔除异常样本后所建 PLS 预测模型的预测均方差或预测决定系数 R^2 。

图 2a、2b 为分别采用 MD、ODXY、MODXY 3 种方法剔除异常样本后在不同训练集和验证集下建立的玉米含水率预测模型的预测均方差 (RMSEP) 和预测决定系数 R^2 。由图 2a 可以看出, 除样本数量划分比例为 5:1 以外, 利用 MODXY 法和 ODXY 法剔除异常样本后所得的预测模型 RMSEP 要小于或明显小于 MD 法; 采用 MODXY 法所得到的 RMSEP 小于等于采用 ODXY 法所得到的 RMSEP。由图 2b 可以看出, 以 R^2 作为评价指标时, 采用 MODXY 法所得的 R^2 值均大于或明显大于采用 MD 法时得到的 R^2 值, 大于或等于采用 ODXY 法得到的 R^2 值。

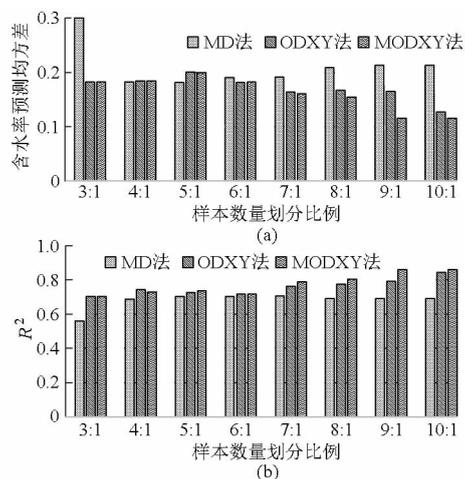


图 2 玉米含水率预测性能比较

Fig. 2 Performance comparison of prediction models for corn moisture content

图 3a、3b 为分别采用 MD、ODXY、MODXY 3 种方法剔除异常样本后在不同训练集和验证集下建立的玉米含油率预测模型的 RMSEP 和 R^2 。由图 3a 可以看出, 以 RMSEP 为评价指标时, 采用 MODXY 法时所得到的 RMSEP 要明显小于采用 MD 法所得到的预测 RMSEP; 除样本数量划分比例为 6:1 时采

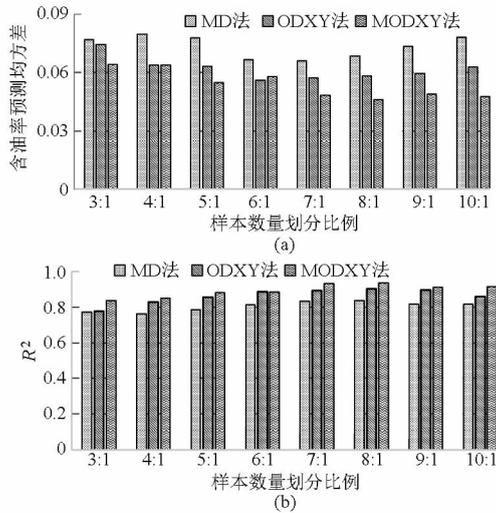


图3 玉米含油率预测性能比较

Fig. 3 Performance comparison of prediction models for corn oil content

用 MODXY 法所得到的 RMSEP 略大于采用 ODXY 法外,其他情况下均小于等于 ODXY 法。由图 3b 可以看出,采用 MODXY 法时所得的含油率预测模型的 R^2 值明显大于采用 MD 法时得到的 R^2 值,略大于采用 ODXY 法时得到的 R^2 值。

图 4 为分别采用 MD、ODXY、MODXY 3 种方法剔除异常样本后在不同训练集和验证集下建立的玉米蛋白质含量预测模型的 RMSEP 和 R^2 。由图 4a 可以看出,以 RMSEP 为评价指标时,采用 MODXY 法时所得到的 RMSEP 要明显小于采用 MD 法所得到的预测 RMSEP;除样本数量划分比例为 9:1 外,采用 MODXY 法时所得到的 RMSEP 均小于 ODXY 法。由图 4b 可以看出,采用 MODXY 法时所得的蛋白质含量预测模型的 R^2 值大于采用 MD 法和 ODXY 法时得到的 R^2 值;ODXY 法和 MD 法相比,不同样本

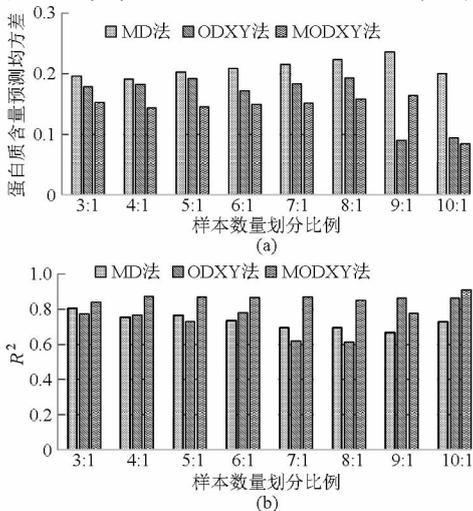


图4 玉米蛋白质预测性能比较

Fig. 4 Performance comparison of prediction models for corn protein content

集下所得 R^2 值的大小关系没有明显规律。

图 5 为分别采用 MD、ODXY、MODXY 3 种方法剔除异常样本后在不同训练集和验证集下建立的玉米淀粉含量预测模型的预测均方差和预测决定系数 R^2 。由图 5a 可以看出,除样本数量划分比例为 3:1 和 4:1 以外,利用 MODXY 法所得的 RMSEP 要小于采用 MD 法;ODXY 法和 MD 法相比,所得的 RMSEP 没有明显的大小规律。由图 5b 可以看出,采用 MODXY 法和 MD 法所建立的玉米淀粉含量预测模型的 R^2 基本相等,采用 ODXY 法所得的 R^2 在某些样本集下略小。

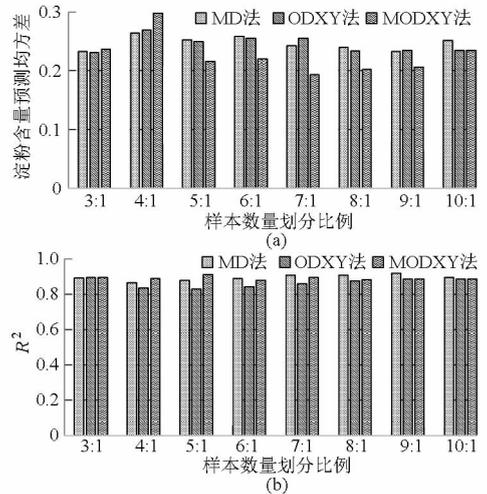


图5 玉米淀粉含量预测性能比较

Fig. 5 Performance comparison of prediction models for corn starch content

综上所述,利用 ODXY 法剔除样本所建立的含水率、含油率和蛋白质含量的预测模型在大多数情况下较 MD 法剔除样本后所建立的 PLS 模型具有更小的 RMSEP 和更大的 R^2 ;MODXY 法剔除样本后所建立的预测模型较 MD 法和 ODXY 法剔除样本后所建立的预测模型具有更小的 RMSEP 和更大的 R^2 。结果表明:在本次实验的玉米含水率、含油率和蛋白质含量预测建模过程中,MODXY 法优于 MD 法和 ODXY 法,ODXY 法在大多数情况下优于 MD 法。

在对玉米淀粉含量的分析过程中,MODXY 方法和 ODXY 方法剔除预测样本后所建模型较 MD 方法剔除样本后所建模型并不总是具有更小的 RMSEP 和更大的 R^2 值。结果表明:在对玉米淀粉含量的预测建模过程中,MODXY 法并没有表现出明显优于 MD 法和 ODXY 法的性能。

2.3 MODXY 异常样本识别方法的适应性分析

为进一步研究图 5 中所得实验结果产生的原因,对所有样本的各组分化学值进行分析,计算其相对标准偏差,其结果分别为:样本集中的玉米含水率

的相对标准偏差为 0.04, 含油率的相对标准偏差为 0.05, 蛋白质含量的相对标准偏差为 0.06, 淀粉含量的相对标准偏差为 0.01。由此可见, 该玉米样本集中的淀粉含量的相对标准偏差远小于其它 3 种组分的相对标准偏差。结合 ODXY 法和 MODXY 法的原理, 二者均采用 dY 与 dX 的比值作为判定指标, 当 dY 较小时, dY/dX 的值也较小, 利用 dY/dX 辨别异常样本的灵敏度也将减低。当 dY/dX 的值小于 dY/dX 的随机误差波动时, 将无法继续采用 dY/dX 来实现对异常样本的辨别。

因此, MODXY 法和 ODXY 法更适合于相应组分化学值相对标准偏差较大的样本集中的异常样本

识别。

3 结论

(1) MODXY 法是专用于近红外多组分分析的一种异常样本识别方法, 它联合考虑了光谱值 X 和各化学值 Y 之间的关联关系, 对于光谱值 X 和化学值 Y 均异常的情形具有较好的辨别能力。

(2) ODXY 法和 MODXY 法更适合于相应组分化学值相对标准偏差较大的情况。

(3) 在近红外多组分分析中, 当所分析组分的化学值相对标准偏差较大时, MODXY 法相对于其他方法具有更好的异常样本识别能力。

参 考 文 献

- 1 Talens P, Mora L, Morsy N, et al. Prediction of water and protein contents and quality classification of Spanish cooked ham using NIR hyperspectral imaging [J]. *Journal of Food Engineering*, 2013, 117(3):272-280.
- 2 Teixeira A, Oliveira A, Paulos K, et al. An approach to predict chemical composition of goat *Longissimus thoracis et lumborum* muscle by near infrared reflectance spectroscopy [J]. *Small Ruminant Research*, 2015, 126:40-43.
- 3 Mourot B P, Gruffat D, Durand D, et al. Breeds and muscle types modulate performance of near-infrared reflectance spectroscopy to predict the fatty acid composition of bovine meat [J]. *Meat Science*, 2015, 99:104-112.
- 4 严衍禄. 近红外光谱分析的原理、技术与应用[M]. 北京:中国轻工业出版社, 2013.
- 5 张丙芳, 苑立波, 张丙秀. 近红外光纤光谱技术检测西红柿果浆总酸及可溶性糖的研究[J]. *光谱学与光谱分析*, 2014, 34(2):350-353.
- 6 Pereira C M F S, Alves A M F B, Vieira A C F G A. Fourier transform near-infrared spectroscopy as a reliable method for quick real time analysis of complex samples in industry[J]. *Ind. Eng. Chem. Res.*, 2013, 52(32):10873-10879.
- 7 谈爱玲, 毕卫红. 基于近红外光谱法的多组分复杂溢油源定量建模分析[J]. *光谱学与光谱分析*, 2012, 32(12):3203-3207.
- 8 Reeves J B, Kessel J S V. Near-infrared spectroscopic determination of carbon, total nitrogen, and ammonium-N in dairy manures [J]. *Journal of Dairy Science*, 2000, 83(8):1829-1836.
- 9 刘强, 罗长兵, 陈绍江, 等. 近红外光谱分析青贮玉米 NDF 中判别异常光谱的研究[J]. *光谱学与光谱分析*, 2007, 27(8):1514-1518.
- 10 赵振英, 林君, 张怀柱. 近红外光谱法分析油页岩含油率中异常样品识别和剔除方法的研究[J]. *光谱学与光谱分析*, 2014, 34(6):1707-1710.
- 11 Wu S, Liu C, Wu J, et al. Outlier sample analysis on near infrared spectroscopy determination for flour ash[C]// 2014 Sixth International Conference on Measuring Technology and Mechatronics Automation IEEE, 2013:781-784.
- 12 刘智超, 蔡文生, 邵学广. 蒙特卡洛交叉验证用于近红外光谱奇异样本的识别[J]. *中国科学:化学*, 2008, 38(4):316-323.
- 13 Wang J, Cheng J, Liu H, et al. Optimization of informative spectral regions in FT-NIR spectroscopy for measuring the soluble solids content of apple[J]. *Intelligent Automation & Soft Computing*, 2015, 21(3):355-370.
- 14 Galvao R K H, Araujo M C U, José G E, et al. A method for calibration and validation subset partitioning [J]. *Talanta*, 2005, 67(4):736-740.
- 15 Reis M M, Rosenvold K. Early on-line classification of beef carcasses based on ultimate pH by near infrared spectroscopy [J]. *Meat Science*, 2014, 96(2):862-869.
- 16 Alamprese C, Casale M, Sinelli N, et al. Detection of minced beef adulteration with turkey meat by UV-vis, NIR and MIR spectroscopy[J]. *LWT—Food Science and Technology*, 2013, 53(1):225-232.
- 17 Kamruzzaman M, Sun D W, Elmasry G, et al. Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis[J]. *Talanta*, 2013, 103(2):130-136.
- 18 傅霞萍, 应义斌. 基于 NIR 和 Raman 光谱的果蔬质量检测研究进展与展望[J]. *农业机械学报*, 2013, 44(8):148-164.
- 19 Fu X P, Ying Y B. Application of NIR and Raman spectroscopy for quality and safety inspection of fruits and vegetables: a review [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2013, 44(8):148-164. (in Chinese)
- 19 Noha M, Da-Wen S. Robust linear and non-linear models of NIR spectroscopy for detection and quantification of adulterants in fresh and frozen-thawed minced beef [J]. *Meat Science*, 2013, 93(2):292-302.