

鱼粉、豆粕样本-样本二维相关近红外光谱判别^{*}

吕程序¹ 陈龙健¹ 杨增玲¹ 韩鲁佳^{1,2}

(1. 中国农业大学工学院, 北京 100083; 2. 中国农业大学动物营养学国家重点实验室, 北京 100091)

【摘要】 为了探讨利用二维相关近红外光谱法快速判别鱼粉和豆粕饲料原料的可行性,共收集了52个鱼粉、60个豆粕样品,其中34个鱼粉和40个豆粕样品作为定标集,其余18个鱼粉和20个豆粕样品作为独立验证集,用Spectrum400型近红外光谱仪扫描获取光谱,并对其特征峰进行解析。采用样本-样本二维相关分析方法定性判别鱼粉、豆粕,当预处理方法为一阶导数时,判断正确率为100%。与偏最小二乘判别分析方法相比,二者均可正确检测鱼粉、豆粕,其中二维相关方法的建模及验证算法简单、计算量小、运算速度快。

关键词: 鱼粉 豆粕 二维相关光谱 近红外光谱 判别

中图分类号: O657.33; S816.17 文献标识码: A 文章编号: 1000-1298(2012)12-0141-05

Discriminant Analysis of Fish Meal and Soybean Meal Using Sample-sample 2-D Correlation Near-infrared Spectroscopy

Lü Chengxu¹ Chen Longjian¹ Yang Zengling¹ Han Lujia^{1,2}

(1. College of Engineering, China Agricultural University, Beijing 100083, China)

2. State Key Laboratory of Animal Nutrition, China Agricultural University, Beijing 100091, China)

Abstract

In order to investigate the feasibility of sample-sample 2-D correlation near infrared diffuse reflectance spectroscopy method for different protein feeds rapid identification, 52 fish meal samples and 60 soybean meal samples were collected, including 74 calibration samples and 38 independent validation samples. Samples were measured by spectrometer. Calibration model was established by sample-sample 2-D correlation spectroscopy method with the pretreatment of 3-point first derivative. This method performed 100% of correctly classified samples. Compared with partial least squares discriminant analysis method, both of them can correctly identify the fish meal and soybean meal. But the sample-sample 2-D correlation spectroscopy method is fast for calculation.

Key words Fish meal, Soybean meal, 2-D correlation spectroscopy, Near infrared spectroscopy, Discriminant

引言

蛋白饲料营养丰富,蛋白质含量高,利于饲养动物的吸收利用,是畜禽优质的饲料原料。鱼粉、豆粕均为主要蛋白饲料原料,因其营养价值、动物机体吸收利用效果不同,市场价值存在差异。豆粕为大豆

提取豆油后的产物,因大豆原料种类、压榨工艺不同,豆粕的性状存在一定的差异。鱼粉是鱼或其加工的下脚料经蒸煮、压榨、干燥、粉碎制得的粉末状制品,是一种极其复杂的混合物,不同鱼粉组成、含量差异较大。利用常规的实验室化学分析方法进行鱼粉、豆粕定性判别费时、费力。因此,迫切需要研

究开发可进行现场快速鉴别的方法。

近红外光谱 (near infrared spectroscopy, 简称 NIRS) 法具有快速、无损、操作简便、成本低等优势, 近年来被广泛应用于饲料检测行业。现有研究表明近红外光谱结合传统的化学计量学方法可以定性、定量检测鱼粉中的豆粕^[1-2], 但传统化学计量学方法算法复杂, 耗时多, 不能适应大数据量光谱批量处理的需求。

二维相关光谱 (2-D correlation spectroscopy, 简称 2DCOS) 源于核磁共振领域^[3]。1986 年 Noda 提出了二维红外相关光谱概念^[4], 因其具有高效分离复杂光谱中重叠峰等诸多优点得到迅速发展^[5-6]。而后发展为广义二维相关光谱^[7], 可应用于更广阔的光谱范围, 如二维相关近红外光谱等, 同时, 简化了算法, 使二维相关光谱计算更为简易。Šašić 等^[8-9]进一步将传统的变量-变量二维相关光谱扩展到样本-样本二维相关光谱, 使二维相关光谱从获取变量信息转变为获取样本信息, 表明样本-样本二维相关同步谱可以表征样本外扰信息, 如浓度等^[10], 这为该方法进行快速定量、定性判别分析提供了可能。

本研究拟针对鱼粉、豆粕的快速判别分析要求, 开展基于样本-样本二维相关近红外光谱的可行性研究。

1 材料与方法

1.1 样本采集与制备

收集鱼粉样品 52 个和豆粕样品 60 个。使用 ZM100 型旋风磨 (德国 Retsch GmbH 公司) 粉碎样品, 过 1.0 mm 标准筛。所有鱼粉样本均经过显微镜检测确证其真实性^[11]。随机选取 34 个鱼粉和 40 个豆粕为定标集, 其余 18 个鱼粉和 20 个豆粕为独立外部验证集。

1.2 近红外光谱采集与分集

光谱采集仪器为 Spectrum400 型光谱仪 (英国 Perkin Elmer 公司), 使用积分球附件、旋转样品台及 $\phi 10$ cm 高性能样品杯。数据采集软件为随机软件 Spectrum 6.3.5。扫描参数: 光谱扫描波数范围 10 040 ~ 3 960 cm^{-1} , 采集数据点波数范围 10 000 ~ 4 000 cm^{-1} , 分辨率 16 cm^{-1} , 数据采集间隔 8 cm^{-1} , 样品扫描 32 次, 每个样品重复装样 3 次, 获取样本平均光谱。背景选用内置聚四氟乙烯背景底板, 扫描 100 次。

1.3 数据处理

1.3.1 数据处理软件

所有的数据处理均在 Matlab 7.0 软件 (The

MathWorks, 美国) 中实现。其中光谱预处理以及偏最小二乘判别分析 (partial least squares discriminant analysis, 简称 PLS-DA) 的建模、验证使用了 PLS-Toolbox 5.5 (Eigenvector Research, 美国)。

1.3.2 预处理方法

为消除光谱信号的基线漂移、放大光谱信号差异, 选用加权最小二乘二次基线校正、Savitzky Golay 一阶、二阶导数方法; 为规范数据, 使用了归一化方法; 为消除样本粒度不均引起的散射, 使用多元散射校正、变量标准化、变量标准化和去散射处理组合方法, 分别表示为 MSC、SNV、SNVD。

1.3.3 样本-样本二维相关光谱计算

体系受外扰 t 引起的一系列光谱 $X_{v,t}$, 经中心化处理得到动态谱 $\tilde{X}_{v,t}$ ^[7] 为

$$\tilde{X}_{v,t} = \begin{cases} X_{v,t} - \bar{X}_v & (T_{\min} \leq t \leq T_{\max}) \\ 0 & (\text{其他}) \end{cases} \quad (1)$$

式中 \bar{X}_v ——参考谱, 是定标集样本的平均谱

v ——光谱变量 t ——外扰水平

T_{\max} 、 T_{\min} ——最大、最小外扰水平

样本-样本二维相关同步谱 ϕ 为^[8]

$$\phi = \frac{1}{n_v - 1} \tilde{X}_{v,t} \tilde{X}_{v,t}^T \quad (2)$$

式中 n_v ——光谱变量数

Šašić 等^[10] 研究表明样本-样本二维相关切线谱与样本浓度存在线性关系。对于任意样本 i 切线谱表达式为

$$\phi_i = \frac{1}{n_v - 1} \tilde{X}_{v,i} \tilde{X}_{v,i}^T \quad (3)$$

式中 $\tilde{X}_{v,i}$ ——样本 i 的动态谱

本文研究对象鱼粉、豆粕光谱可以看作为一系列浓度外扰组成的光谱, 其中鱼粉中豆粕的质量分数为 0%, 豆粕质量分数为 100%, 即二维相关切线谱包含了鱼粉中豆粕的浓度信息。因此, 通过设定合理的阈值, 即可对其进行定性判别。

为消除选取不同样本 i 的切线谱对定性判别的影响, 选用一类样本切线谱的平均谱 ϕ_A 进行判别, 即

$$\phi_A = \frac{1}{n_A} \sum_{a=1}^{n_A} \frac{1}{n_v - 1} \tilde{X}_{v,a} \tilde{X}_{v,a}^T \quad (4)$$

式中 n_A ——选用的一类样本在定标集中数目

$\tilde{X}_{v,a}$ ——选用的一类样本在定标集中任意样本 a 的动态谱

本文选用的是定标集鱼粉样本 34 条切线谱的平均谱, 即 n_A 为定标集鱼粉光谱条数, $\tilde{X}_{v,a}$ 为定标集

中鱼粉样本 a 的动态谱。

建模时,根据式(1)计算定标集动态谱,根据式(4)计算定标集所有样本-定标集鱼粉样本切线谱的平均谱,该平均谱为 74×1 的行向量 φ_{AC} 。设定鱼粉、豆粕样本类别属性分别为 0、1,定标集样本的类别属性向量即为一条 0、1 表示的 74×1 行向量 M_C 。切线谱的平均谱 φ_{AC} 与类别属性向量 M_C 进行一元线性回归,求得回归方程为

$$M_C = k \varphi_{AC} + s \quad (5)$$

式中 k, s ——参数,由 Matlab 进行线性回归求得

定标过程中选用全交互验证,根据交互验证标误差 (E_{CV}),计算定性判别阈值^[2]

$$Q_1 = 0.5 - 1.25E_{CV} \quad Q_2 = 0.5 + 1.25E_{CV}$$

综合定标集的判断正确率、误判率、不确定率和 E_{CV} 等指标,选取较优的预处理方式。

验证模型时,根据式(1)、(4)计算验证集动态谱以及与验证样本-定标集鱼粉样本切线谱的平均谱 φ_{IV} ,并根据建模时建立的回归方程拟合。拟合结果大于或等于 Q_2 的样本判别为豆粕,拟合结果小于或等于 Q_1 的样本判别为鱼粉,拟合结果在 Q_1, Q_2 之间的样本不能判别其属性。

2 结果与讨论

2.1 样本光谱分析

图 1 所示为 52 个鱼粉样本和 60 个豆粕样本的近红外光谱图。从图中可见鱼粉、豆粕原始近红外光谱差异不显著,因此需要适宜的预处理方法来消除基线漂移、旋转,以获得更清晰的光谱轮廓变化。

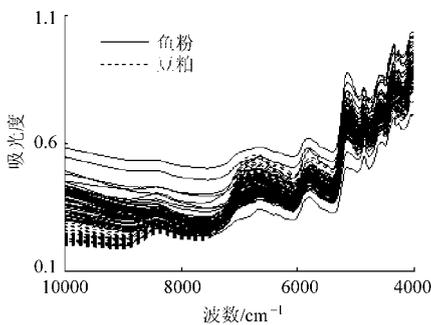


图 1 鱼粉、豆粕近红外光谱图

Fig. 1 Spectra of fish meal and soybean meal

图 2 为 52 个鱼粉和 60 个豆粕样品的平均 Savitzky Golay 5 点二阶导数光谱。鱼粉、豆粕光谱在 $5900 \sim 5700 \text{ cm}^{-1}$ 存在差异,其中 5869 cm^{-1} 、 5787 cm^{-1} 为 C—H 键一倍频伸缩振动吸收峰; 5755 cm^{-1} 为蛋白质 α -螺旋二级结构的二阶导数特征峰^[12](图 2b)。鱼粉、豆粕二阶导数光谱在 $4650 \sim 4520 \text{ cm}^{-1}$ 存在差异,其中 $4650 \sim 4610 \text{ cm}^{-1}$ 为 C—H 键合频振动特征峰; 4590 cm^{-1} 为蛋白质中二酰胺合

频振动特征峰; $4540 \sim 4525 \text{ cm}^{-1}$ 为蛋白质 β -折叠二级结构二阶导数特征峰。 $4400 \sim 4200 \text{ cm}^{-1}$ 为脂肪特征吸收峰谱带,其中 4329 cm^{-1} 为 C—H 键弯曲振动二倍频吸收峰; 4314 cm^{-1} 、 $4270 \sim 4230 \text{ cm}^{-1}$ 为 C—H 合频振动特征峰^[12-13](图 2c)。

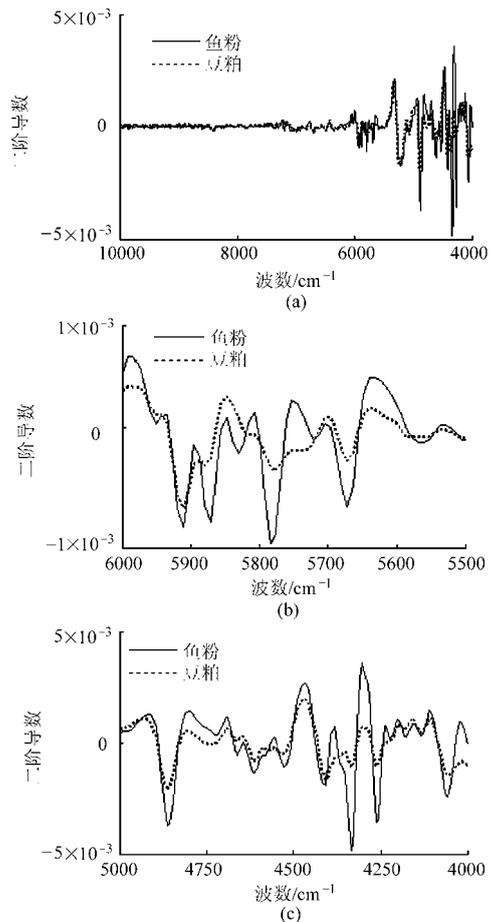


图 2 鱼粉、豆粕平均近红外二阶导数光谱图
Fig. 2 Second derivative spectra of fish meal and soybean meal

2.2 样本-样本二维相关光谱分析

2.2.1 预处理方法选择

使用定标集 34 个鱼粉光谱、40 个豆粕光谱进行样本-样本二维相关计算,得到同步谱如图 3 所示。其中样本编号 1~34 为鱼粉样本,35~74 为豆粕样本。可见同步谱图关于自动峰对称分布,其中,鱼粉-鱼粉、豆粕-豆粕二维同步相关值为正,鱼粉-豆粕二维同步相关值为负。同种类样本表现为同步正相关性;不同种类样本表现为同步负相关性,因此可以利用样本-样本二维相关光谱法进行定性判别分析。

对不同方法预处理的光谱进行样本-样本二维相关计算,判别如表 1 所示。可见利用原始光谱进行定性判别,定标集中有 3 个鱼粉样本不能判断其属性。经预处理后,除归一化,其他方法均可正确定性判别样本属性,其中 Savitzky Golay 一阶、二阶导

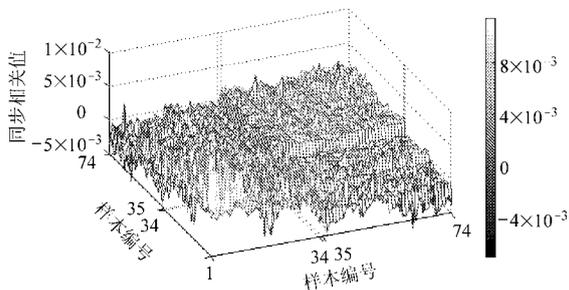


图3 定标集原始光谱的样本-样本二维相关同步谱

Fig.3 Sample-sample 2-D correlation synchronous of calibration spectra

数预处理后, E_{CV} 最小, 其二维相关同步谱图如图4所示, 可见较原始光谱可更清晰地反映样本类别信息。这主要是因为经导数处理后, 消除了低频噪声的影响, 同时光谱信号被放大, 有利于有效信息的检出。

表1 不同光谱预处理方法对样本-样本2DCOS定性判别结果的影响

Tab.1 Results of sample-sample 2-D correlation spectroscopy analysis using different pretreatments

预处理方法	定标集		E_{CV}	验证集	
	判断正确率/%	不确定率/%		判断正确率/%	不确定率/%
无	95.9	4.1	0.027	100	0
基线校正	100	0	0.018	100	0
归一化	100	0	0.026	97.1	2.9
MSC	100	0	0.020	100	0
SNV	100	0	0.020	100	0
SNVD	100	0	0.019	100	0
一阶导数	100	0	0.015	100	0
二阶导数	100	0	0.015	100	0

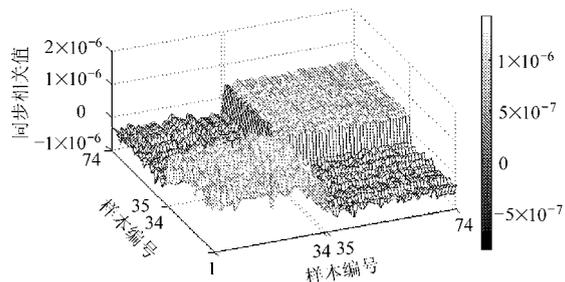


图4 定标集一阶导数光谱样本-样本二维相关同步谱图

Fig.4 Sample-sample 2-D correlation synchronous of first derivative spectra

2.2.2 基于一阶导数光谱的二维相关分析

利用实验获得的光谱进行 Savitzky Golay 3 点一阶导数处理。通过二维相关计算, 得到 34 个切线谱的平均谱如图5所示。类别属性中 0 表示鱼粉, 1 表示豆粕。平均切线谱和类别属性向量进行一元线性回归, 回归结果如图6所示。采用全交互验证,

E_{CV} 为 0.014 7, 进而得出定性判别阈值分别为 0.482、0.518。而后, 利用验证集动态谱进行验证, 判定结果如图7所示, 可见, 样本-样本二维相关近红外光谱可准确的对独立外部验证集进行判定, 该方法可用于鱼粉、豆粕的定性判别分析。

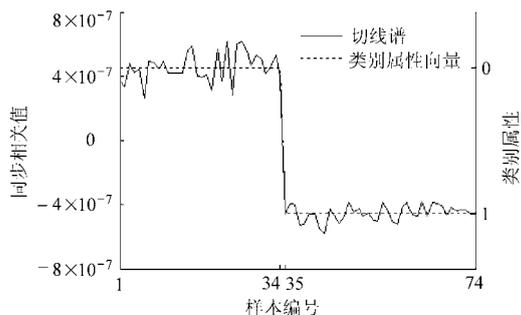


图5 定标集样本-定标集鱼粉切线谱平均谱

Fig.5 Average slice spectra of calibration sample-fish meal 2-D correlation synchronous

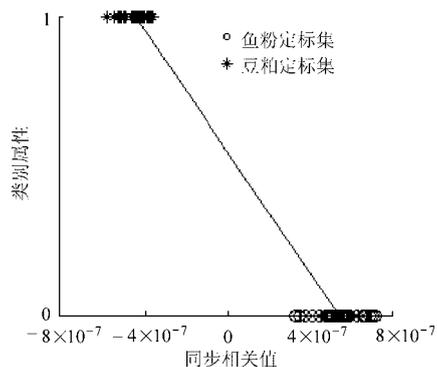


图6 切线谱的平均谱与类别属性向量回归结果

Fig.6 Regression result of average slice spectra and category matrix

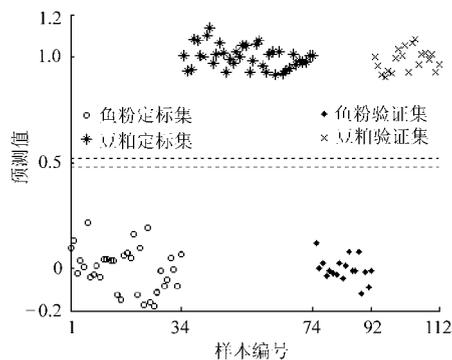


图7 一阶导数光谱样本-样本二维相关定性判别结果

Fig.7 Qualitative results of sample-sample 2-D correlation first derivative spectroscopy analysis

2.3 2DCOS 和 PLSDA 比较分析

PLSDA 分析过程中, 对鱼粉、豆粕光谱分别赋值 -1、1, 定标过程中设定最大主成分因子数 20, 采用全交互验证方法, 最优的主成分因子数根据 E_{CV} 最小的原则选出, 并以此避免模型的过拟合。根据预测集标准误差最小来优选预处理方法。研究结果表明 Savitzky Golay 15 点一阶导数预处理后的光谱,

选用 13 个主成分因子建模效果最好, 内部验证、外部验证判断正确率均为 100%。

样本-样本 2DCOS 分析和 PLSDA 方法均可以正确定性判别鱼粉、豆粕。传统的最小二乘回归方法, 在参数估计式 $\hat{\theta} = (X^T X)^{-1} X^T Y$ 中, $(X^T X)$ 必须是可逆矩阵才有意义, 所以当模型变量多重相关, 或者当样本数少于变量个数时, 参数估计式一般就会失效。为解决上述问题, 在偏最小二乘回归方法提取主成分后再进行回归处理, 并反复迭代计算直至满足建模精度为止。利用 PLS_Toolbox 进行 PLSDA 分析并计算其程序运行时间, 循环 300 次取平均, 建模、验证的时间分别为 1.095 4、0.028 0 s; 利用自编程序进行 2DCOS 定性判别分析并计算程序运行时

间, 循环 300 次取平均, 建模、验证的时间分别为 0.209 7、0.004 6 s。研究结果表明样本-样本 2DCOS 分析方法建模、验证计算时间短。该方法算法简单、运算量小、建模效率高、识别速度快。适用于数据量较大的光谱处理, 如显微光谱以及高光谱图像; 同时在建模速度要求较高的在线建模中也有巨大的应用潜力。

3 结束语

本研究表明, 基于样本-样本二维相关近红外光谱技术, 可以自动、准确、快速地识别鱼粉、豆粕饲料原料。该算法简洁, 运算速度快, 在大数据量运算中有着巨大的应用潜力。

参 考 文 献

- 1 陆月青. 近红外光谱分析技术在饲料分析中的应用研究[D]. 南宁: 广西大学, 2007.
Lu Yueqing. Application of NIR technology in the feed analysis[D]. Nanning: Guangxi University, 2007. (in Chinese)
- 2 石光涛, 韩鲁佳, 杨增玲, 等. 鱼粉中掺杂豆粕的可见和近红外反射光谱分析研究[J]. 光谱学与光谱分析, 2009, 29(2): 362 ~ 366.
Shi Guangtao, Han Lujia, Yang Zengling, et al. Methods of analyzing soybean meal adulteration in fish meal based on visible and near infrared reflectance spectroscopy [J]. Spectroscopy and Spectral Analysis, 2009, 29(2): 362 ~ 366. (in Chinese)
- 3 Aue W P, Bartholdi E, Ernst R R. Two-dimensional spectroscopy. Application to nuclear magnetic resonance [J]. Journal of Chemical Physics, 1976, 64(5): 2 229 ~ 2 246.
- 4 Noda I. Two-dimensional infrared (2-D IR) spectroscopy of synthetic and biopoly-mer [J]. Bull. Am. Phys. Soc., 1986, 31: 520 ~ 522.
- 5 Noda I. Two-dimensional infrared spectroscopy [J]. J. Am. Chem. Soc., 1989, 111(21): 8 116 ~ 8 118.
- 6 Noda I. Two-dimensional infrared (2D IR) spectroscopy: theory and applications [J]. Appl. Spectrosc., 1990, 44(4): 550 ~ 561.
- 7 Noda I. Generalized two-dimensional correlation method applicable to infrared raman and other types of spectroscopy [J]. Appl. Spectrosc., 1993, 47(9): 1 337 ~ 1 342.
- 8 Šašić S, Muszynski A, Ozaki Y. A new possibility of the generalized two-dimensional correlation spectroscopy 1 sample-sample correlation spectroscopy [J]. J. Phys. Chem. A, 2000, 104(27): 6 380 ~ 6 387.
- 9 Šašić S, Muszynski A, Ozaki Y. A new possibility of the generalized two-dimensional correlation spectroscopy 2 sample-sample and wavenumber-wavenumber correlations of temperature-dependent near-infrared spectra of oleic acid in the pure liquid state [J]. J. Phys. Chem. A, 2000, 104(27): 6 388 ~ 6 394.
- 10 Šašić S, Ozaki Y. Wavelength-wavelength and sample-sample two-dimensional correlation analyses of short-wave near-infrared spectra of raw milk [J]. Appl. Spectrosc., 2001, 55(2): 163 ~ 172.
- 11 Commission Regulation (EC) No 152/2009. Laying down the methods of sampling and analysis for the official control of feed [S]. 2009.
- 12 Workman J, Weyer L. Practical guide to interpretive near-infrared spectroscopy [M]. New York: CRC Press, 2007.
- 13 Piraux F, Dardenne P. Feed authentication by near-infrared microscopy [C] // Davies A M C, Giangiacomo R. Proc. of the 9th International Conference on Near-infrared Spectroscopy, Italy, 1999.