

## 基于 PLS-GRNN 的豆粕品质近红外光谱检测研究

王立琦<sup>1</sup>, 姚静<sup>1</sup>, 王睿莹<sup>1</sup>, 陈颖淑<sup>1</sup>, 罗淑年<sup>2</sup>, 王伟宁<sup>2</sup>, 张艳荣<sup>1\*</sup>

1. 哈尔滨商业大学计算机与信息工程学院, 黑龙江省电子商务与信息处理重点实验室, 黑龙江 哈尔滨 150028  
2. 哈尔滨商业大学食品工程学院, 黑龙江 哈尔滨 150028

**摘要** 豆粕是大豆浸提取豆油后经适当干燥和热处理所得副产品, 是制作禽畜类饲料的主要原料, 其品质决定营养价值。针对现有豆粕品质检测方法存在着有毒化学试剂使用多、操作复杂、分析时间长、无法满足实际生产线快速检测及调控需求等问题, 提出一种基于近红外光谱分析的豆粕品质多组分检测方法, 以期用于产品质量在线检测及调控。从大豆油脂加工生产线上采集豆粕样品 449 个, 利用 105 °C 烘箱法、凯氏定氮法和索氏提取法分别测定样品的水分、蛋白质和脂肪化学值, 采用瑞士 BuchiNIRMaster 傅里叶变换近红外光谱仪采集样品漫反射光谱。首先利用马氏距离法剔除异常样本, 然后用多种方法对光谱数据进行降噪处理, 对比分析发现小波去噪效果最优。分别采用 KS 和 SPXY 两种算法确定豆粕不同组分的最佳样本分集。为了探讨豆粕组分的近红外吸收特性, 剔除光谱冗余信息, 降低模型计算复杂度, 采用区间偏最小二乘法(iPLS)对 4 000~10 000  $\text{cm}^{-1}$  全谱进行特征提取, 优选出水分、蛋白质和脂肪的特征吸收波段分别为 4 904~5 200, 4 304~4 600 和 4 304~4 600  $\text{cm}^{-1}$ 。最后建立豆粕组分含量的广义回归神经网络(GRNN)预测模型。为了减少网络的输入变量, 缩小网络规模, 提高运行速度, 采用 PLS 对光谱数据降维, 提取主因子得分作为 GRNN 输入变量。通过交叉验证循环法优选网络参数光滑因子 spread 值, 建立豆粕多组分含量 PLS-GRNN 预测模型, 并与经典的 PLS 和 BP 模型对比, 发现 PLS-GRNN 模型效果更优, 其水分、蛋白质和脂肪的预测集  $R^2$  分别为 0.976 9, 0.940 2 和 0.911 1, RMSEP 分别为 0.091 2, 0.383 4 和 0.113 4, RSD 分别为 0.79%, 0.83% 和 8.53%。虽然脂肪的预测误差相对较大, 但也在模型评定标准可用范围之内。实验表明基于 PLS-GRNN 的近红外光谱分析用于豆粕品质检测是可行的, 能够用于实际生产过程中的品质监控。

**关键词** 豆粕品质; 近红外光谱; 区间偏最小二乘; 广义回归神经网络

**中图分类号:** O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)05-1433-06

### 引言

豆粕是大豆浸提取豆油后, 经适当干燥和热处理所得副产品, 产量大, 营养丰富, 主要用于禽畜类饲料, 也是生产化肥、制作食品的辅料, 是重要的期货交易物。豆粕品质的评价指标主要有水分、蛋白质、脂肪、微量物质(如纤维、灰分、氨基酸、维生素、碳水化合物、胡萝卜素等), 其中水分、蛋白质和脂肪占比高, 是衡量豆粕品质的重要指标, 需要在生产过程中不断检测和调控<sup>[1]</sup>。

现有的豆粕品质检测方法包括化学分析法、色谱分析法等, 普遍存在着有毒化学试剂使用多、操作复杂、分析时间

长等问题, 无法满足实际生产过程快速检测及调控的需求。近红外光谱(near-infrared spectroscopy, NIR)技术具有无损、快速、低成本、多组分同时分析、易于实现在线检测等优点, 特别适合生产过程中的质量监控<sup>[2]</sup>。近年来, NIR 在豆粕品质检测方面应用已有报道。庄树华<sup>[3]</sup>、纳嵘<sup>[4]</sup>采用近红外分析豆粕蛋白质含量; Leeson<sup>[5]</sup>等研究了近红外光谱分析法估测豆粕的代谢能; Fontaine<sup>[6]</sup>等采用近红外光谱法预测豆粕中的氨基酸含量; 王红梅<sup>[7]</sup>等利用近红外建立豆粕的蛋白质和总氨基酸预测模型; 杨增玲<sup>[8]</sup>等建立了豆粕含水率、粗蛋白质质量分数的近红外定量分析模型。上述研究所用样品大多是在实验室人工制备, 而实际生产线上的豆粕产品是多组分同时变化, 变量间相互干扰会影响建模效果。

收稿日期: 2021-04-14, 修订日期: 2021-10-14

基金项目: 国家自然科学基金面上项目(32072259), 黑龙江省自然科学基金项目(LH2020C061)资助

作者简介: 王立琦, 1966 年生, 哈尔滨商业大学计算机与信息工程学院教授 e-mail: hsdwlq@163.com

\* 通讯作者 e-mail: zhangyanrong\_5@163.com

本研究直接从大豆油脂加工生产线上采集样品, 根据加工过程中实际检测控制需求, 对豆粕中的水分、蛋白质和脂肪三个主要成分含量建立基于 PLS-GRNN 的近红外分析模型, 以期能用于实际加工过程快速检测, 及时调整工艺参数, 生产出多等级粕、专用蛋白粕和功能大豆油等新产品。

## 1 实验部分

### 1.1 样品化学值测定

本研究目的是实现加工过程中豆粕品质快速检测及调控, 因此直接从大豆油脂生产线上采集 449 个有代表性的豆粕样品, 依据 GB/T6435—1986《饲料水分的测定方法》, 利用 105 °C 烘箱法测定水分含量范围为 9.68%~13.26%; 依据 GB/T5511—2008《谷物和豆类氮含量测定和粗蛋白质含量计算凯氏法》, 利用凯氏定氮法测定蛋白质含量范围为 41.2%~50.9%; 依据 GB/T5009.6—2003《食品中脂肪的测定》, 利用索氏提取法测定脂肪含量范围为 0.43%~3.75%。

### 1.2 光谱数据采集

采用瑞士 BUCHI 公司的 NIRMaster 型傅里叶变换近红外光谱仪扫描豆粕样品, 光谱范围为 4 000~10 000  $\text{cm}^{-1}$ , 扫描频率 4 次  $\cdot \text{s}^{-1}$ , 分辨率 4  $\text{cm}^{-1}$ 。为保证样品扫描均匀性, 每份样品重复扫描 3 次后取平均值, 获得豆粕样品漫反射近红外光谱如图 1 所示。

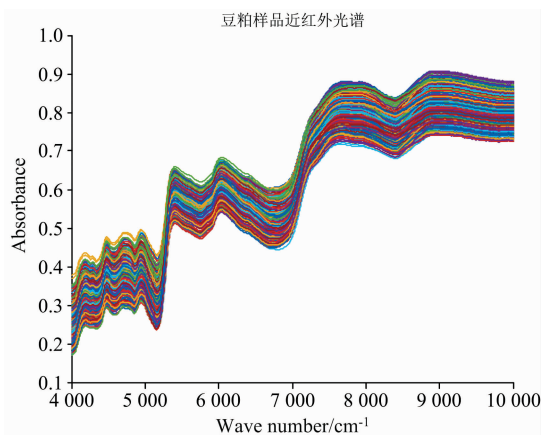


图 1 豆粕样品近红外光谱图

Fig. 1 Near infrared spectra of soybean meal samples

### 1.3 光谱去噪

首先利用马氏距离法从 449 个豆粕样本中剔除 91 个异常样本, 然后采用小波变换对剩余的 358 个样本的光谱数据进行降噪处理。光谱去噪就是在保证光谱数据有用信息的原始真实性前提下, 最大程度地去除各种随机噪声。利用控制参数反复试验法对小波阈值方式、分解尺度和小波基进行筛选<sup>[9]</sup>, 根据去噪信号在原信号中的能量占比和去噪信号与原信号标准差对去噪效果进行评价, 以确保去噪前后信号不失真, 计算公式如式(1)和式(2)

$$\text{perc} = \frac{\sqrt{\sum g^2(n)}}{\sqrt{\sum f^2(n)}} \quad (1)$$

$$\text{err} = \sqrt{\sum f^2(n) - \sum g^2(n)} \quad (2)$$

式(1)和式(2)中,  $f(n)$ 为原始数据,  $g(n)$ 为去噪后的数据。一般  $\text{perc} \propto 1/\text{err}$  越小, 降噪效果越好。然后与移动平均法、多元散射校正和标准正态变量变换三种常规处理方法对比, 发现基于 db6 小波基、2 层分解和 penalty 阈值的小波去噪方式效果最佳, 统计结果如表 1 所示。

表 1 多种降噪方法统计结果对比

Table 1 Comparison of statistical results of various noise reduction methods

去噪方法	perc	err
小波去噪(db6、2层分解、penalty 阈值)	0.999 9	0.000 5
移动平均法	0.999 8	0.001 6
多元散射校正	0.999 7	0.090 4
标准正态变量变换	0.998 5	0.111 0

### 1.4 样本分集

样本集的划分通常有人工选择和计算机识别两种方法。人工选择是将样本化学值顺序排列, 按一定梯度抽取预测集样本。计算机识别是根据光谱特性差异或结合化学值来选择校正集样本。由于本研究为豆粕水分、蛋白质和脂肪多个成分指标同时检测, 实际样品的三个参数是同时变化的, 人工选择样本分集无法做到同时兼顾每个参数指标, 而且经过尝试发现效果很差。计算机识别算法常采用 Kennard Stone (KS) 和 Sample set Partitioning based on joint X-Y distance (SPXY) 两种, KS 算法<sup>[10]</sup>通过计算样本的欧式距离确定校正集, 但它只考虑光谱数据间的关系, 而不考虑与化学值的关系, 因此在预测未知样本时可能缺乏针对性。SPXY 算法<sup>[11]</sup>在 KS 算法基础上改进, 同时兼顾光谱矩阵和浓度矩阵, 以保证最大程度表征样本分布。本研究采用 KS 和 SPXY 两种算法对三个组分的豆粕样本分集, 之后分别对各组分校正集样本建立偏最小二乘(partial least squares, PLS)回归模型, 根据模型的预测效果选择每个组分的最佳样本分集, 有效地避免了人为参与和变量间的相互影响, 两种样本分集方法建模结果如表 2 所示。可以看出, 对于水分和蛋白质, KS 分集法优于 SPXY 分集法, 而对于脂肪则是 SPXY 分集法优于 KS 分集法。

表 2 两种样本分集方法建模结果

Table 2 Modeling results of two sample set partitioning methods

分集方法	指标	校正集		预测集		
		$R^2$	RMSEC	$R^2$	RMSEP	RSD/%
KS	水分	0.922 4	0.239 1	0.917 3	0.242 9	2.11
	蛋白质	0.914 7	0.571 7	0.910 6	0.511 2	1.10
	脂肪	0.875 8	0.154 0	0.839 9	0.178 8	16.83
SPXY	水分	0.906 4	0.243 2	0.896 4	0.251 9	2.20
	蛋白质	0.933 1	0.498 3	0.908 2	0.522 3	1.13
	脂肪	0.852 4	0.161 0	0.891 9	0.145 0	13.10

表 3 为最后的样本分集结果，可以看出，对于水分、蛋白质和脂肪三个参数，其含量化学值的最大值和最小值样本都包含在校正集中，校正集样本和预测集样本的均值和标准差非常接近，说明校正集与预测集的样本分布比较均匀，计算机算法的分集结果符合要求。

**表 3 样本分集结果**  
**Table 3 Sample partition results**

指标	样本集	数目 / 个	最小值 / %	最大值 / %	均值 / %	标准差 / %
水分	校正集	278	9.79	12.82	11.437	0.646 7
	预测集	80	9.97	12.76	11.510	0.603 5
蛋白质	校正集	278	41.2	50.9	46.374	1.494 7
	预测集	80	41.4	49.2	46.436	1.577 7
脂肪	校正集	278	0.44	2.49	1.044	0.352 1
	预测集	80	0.45	2.41	1.107	0.351 8

## 2 结果与讨论

### 2.1 基于 iPLS 的特征波段优选

利用近红外光谱仪采集的波长变量有上千个，其中包含许多与豆粕品质无关的信息，如果用全谱建模，会使模型的计算量增大，稳定性变差，因此在建立预测模型前，有必要进行特征波段优选，剔除光谱中的冗余信息。利用优选出的特征波长变量建模，可以降低模型计算复杂度，提高模型预测性能。

本研究采用区间偏最小二乘法 (interval partial least squares, iPLS) 进行特征波段提取<sup>[12-13]</sup>。将全谱区间分别按 20, 30, 40 和 50 依次等宽均分并对每一个子区间建立 PLS 模型，采用留一交叉验证法计算模型的交互验证均方差 RMSECV 作为评判标准，其最小值对应的子区间即为最佳建模波段。对应水分、蛋白质和脂肪不同子区间数的波段选择结果见表 4、表 5 和表 6。

**表 4 水分 iPLS 不同子区间数波段选择结果**

**Table 4 iPLS waveband selection results of moisture for different subinterval number**

参数指标	子区间数	最佳建模波段/cm <sup>-1</sup>	R <sup>2</sup>	RMSECV
水分	20	4 904~5 200	0.926 9	0.242 4
	30	5 204~5 400	0.924 1	0.246 8
	40	5 216~5 364	0.913 6	0.262 6
	50	4 844~4 960	0.921 5	0.250 9

**表 5 蛋白质 iPLS 不同子区间数波段选择结果**

**Table 5 iPLS waveband selection results of protein for different subinterval number**

参数指标	子区间数	最佳建模波段/cm <sup>-1</sup>	R <sup>2</sup>	RMSECV
蛋白质	20	4 304~4 600	0.939 6	0.510 8
	30	4 404~4 600	0.929 9	0.549 7
	40	4 456~4 604	0.928 7	0.554 0
	50	4 484~4 600	0.930 6	0.546 7

**表 6 脂肪 iPLS 不同子区间数波段选择结果**

**Table 6 iPLS waveband selection results of fat for different subinterval number**

参数指标	子区间数	最佳建模波段/cm <sup>-1</sup>	R <sup>2</sup>	RMSECV
脂肪	20	4 304~4 600	0.875 9	0.170 2
	30	5 604~5 800	0.869 9	0.174 0
	40	4 304~4 452	0.872 3	0.172 7
	50	4 204~4 360	0.872 7	0.172 6

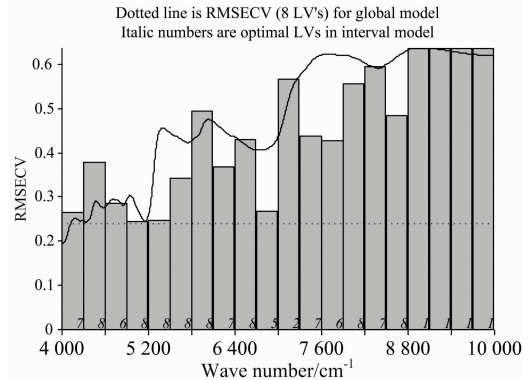


图 2 水分 20 个子区间的 iPLS 建模结果

Fig. 2 iPLS modeling results of 20 subintervals for moisture

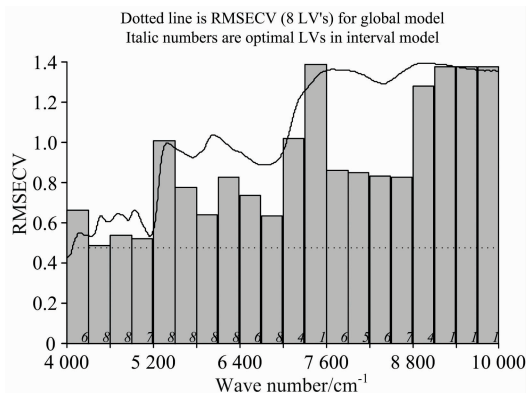


图 3 蛋白质 20 个子区间的 iPLS 建模结果

Fig. 3 iPLS modeling results of 20 subintervals for protein

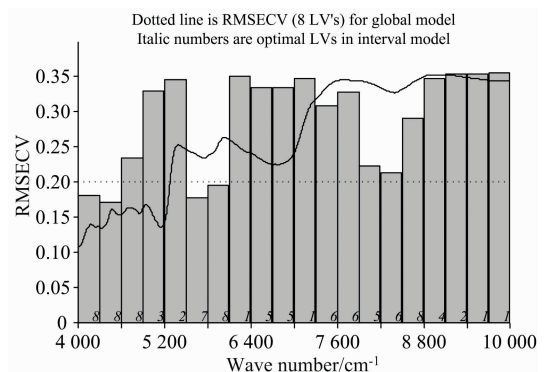


图 4 脂肪 20 个子区间的 iPLS 建模结果

Fig. 4 iPLS modeling results of 20 subintervals for fat

表中可见,无论水分、蛋白质还是脂肪均为划分 20 个子区间时对应的特征吸收波段建模效果最好。图 2、图 3 和图 4 分别展示了水分、蛋白质和脂肪划分 20 个子区间的 iPLS 结果,最终优选出水分、蛋白质和脂肪的特征波段分别为 4 904~5 200, 4 304~4 600 和 4 304~4 600  $\text{cm}^{-1}$ 。

## 2.2 基于 PLS-GRNN 的豆粕组分含量模型建立

### 2.2.1 GRNN 结构

广义回归神经网络 (generalized regression neural networks, GRNN) 是由美国学者 Donald 于 1991 年提出的一种人工神经网络模型<sup>[14]</sup>,是对径向基神经网络 (radial basis function network, RBF) 的改进。GRNN 具有很强的非线性映射能力,在信号过程、结构分析、控制决策系统等领域得到了广泛的应用。但 GRNN 在近红外定量分析方面研究较少,应用在豆粕品质质量检测方面还未见报道。本工作提出建立 PLS-GRNN 联合模型进行豆粕品质多组分含量同步预测,旨在提高豆粕品质预测效率和准确性。

GRNN 结构如图 5 所示,该网络由输入层、模式层、求和层和输出层四层构成<sup>[15]</sup>。输入层神经元数目等于样本中输入向量的维数,各神经元是简单的分布单元,直接将输入变量传递给模式层。模式层神经元数目等于样本的数目,各神经元对应不同的样本。求和层中使用两种类型的神经元进行求和。输出层神经元数目等于样本中输出向量的维数,各神经元将求和层的输出相除。

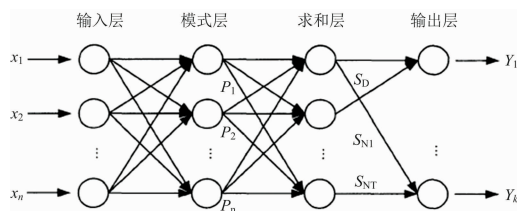


图 5 GRNN 结构

Fig. 5 Structure of GRNN

### 2.2.2 输入变量确定

在光谱分析过程中,光谱数据间可能会存在严重的共线性干扰,因此对输入到网络的光谱数据采取降维处理,不仅可以减少计算量,还可以防止网络陷入局部最小。PLS 不仅可以用于模型的建立,还能用于数据矩阵的分解,提取最佳主因子,从而达到降低数据维度的目的。本研究采用舍-交验证法,根据预测残差平方和 (prediction residual error sum of squares, PRESS) 来确定最佳主因子数,豆粕各组分的 PRESS 值随主因子变化趋势如图 6 所示。可以看出,随着主因子数的增加,PRESS 值明显下降趋势,当主因子数达到一定值时,PRESS 变化趋于平缓,之后基本不再下降,因此可确定水分、蛋白质和脂肪的最佳主因子数分别为 8, 8 和 7,然后将主因子得分作为 GRNN 网络的输入变量用于建模。

### 2.2.3 网络参数优化

GRNN 具有非常简便的网络参数设置功能,整个神经网络只需要设置传递函数中的光滑因子就可以调整网络性能<sup>[16-17]</sup>,而且网络训练过程实际上就是光滑因子的寻优过

程。GRNN 传递函数表示为

$$R(x) = \exp\left(-\frac{\|x-c\|^2}{2\text{spread}}\right) \quad (3)$$

式(3)中,spread 称为光滑因子,决定了训练样本的误差和基函数的形状,其选值大小直接影响模型的预测性能,常用的寻求最佳 spread 的方法是  $k$  折交叉验证循环法。本研究设定 spread 值的范围为 0.1~1,选取 4 折交叉验证来训练 GRNN 网络,spread 寻优曲线如图 7 所示,最小 MSE 所对应的 spread 值即为最优值。图中可见,水分、蛋白质和脂肪的最优 spread 值分别为 0.1, 0.2 和 0.2。

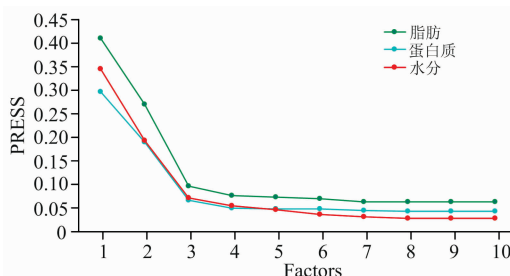


图 6 PRESS 随主因子变化趋势图

Fig. 6 The changing trends of PRESS with the major factors

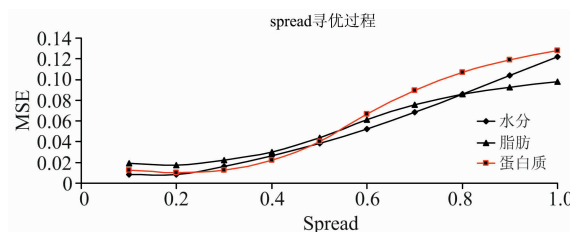


图 7 spread 寻优曲线

Fig. 7 Optimizing curves of spread

## 2.3 PLS-GRNN 模型预测结果及评价

GRNN 网络参数选定后,将 PLS 最佳主因子得分作为网络输入变量,豆粕组分化学值作为输出变量,建立豆粕品质 PLS-GRNN 预测模型,预测效果如图 8 所示。可以看出,水分、蛋白质和脂肪预测样本均在其各自的拟合线附近均匀分布,说明 PLS-GRNN 模型预测效果较好。

最后,将 PLS-GRNN 模型与经典的 PLS 线性模型和 BP 神经网络非线性模型对比,结果如表 7 所示。

由表 7 可见,对于豆粕样品三组分来说,其 PLS-GRNN 模型的预测效果均优于 PLS 模型和 BP 模型,说明 PLS-GRNN 模型的泛化能力更好。其水分、蛋白质和脂肪的预测集决定系数  $R^2$  分别为 0.976 9, 0.940 2 和 0.911 1, 预测均方根误差 RMSEP 分别为 0.091 2, 0.383 4 和 0.113 4, 预测相对标准偏差 RSD 分别为 0.79%, 0.83% 和 8.53%。从实验结果来看,虽然脂肪的 RSD 低于理想要求,但也在模型评定标准可用范围之内,并且从图 8 也可以看出脂肪的拟合效果也很好,分析其原因可能是由于豆粕中脂肪含量低,即使较小的绝对误差也会引起较大的相对误差,而水分和蛋白质含量较高,因此相对误差较小,也可能是三组分间相互影响造成的,关于脂肪的预测精度问题有待进一步研究改善。

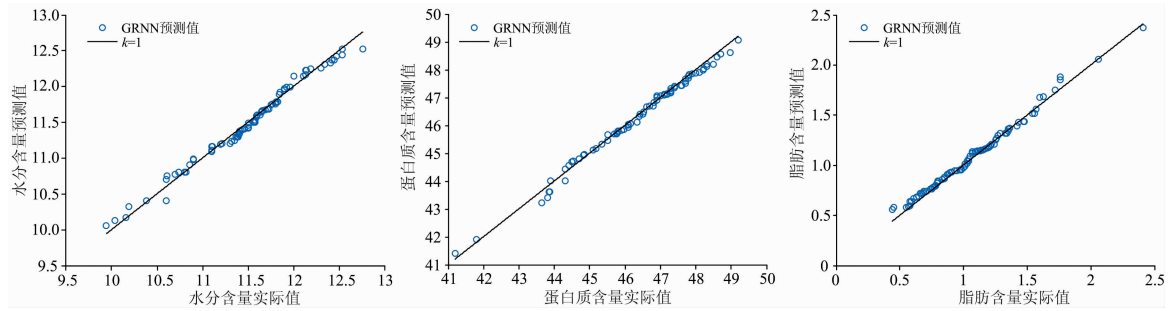


图 8 水分、蛋白质和脂肪的 PLS-GRNN 模型预测效果

Fig. 8 Predictive effects of PLS-GRNN models for moisture, protein and fat

表 7 PLS-GRNN 与 PLS, BP 建模效果对比

Table 7 Comparison of modeling effects between PLS-GRNN and PLS, BP

参数指标	建模方法	$R^2$	RMSEP	RSD/%
水分	PLS	0.917 3	0.242 9	2.11
	BP	0.957 2	0.124 8	1.08
	PLS-GRNN	0.976 9	0.091 2	0.79
蛋白质	PLS	0.910 6	0.511 2	1.10
	BP	0.921 9	0.440 9	0.95
	PLS-GRNN	0.940 2	0.383 4	0.83
脂肪	PLS	0.891 9	0.145 0	13.10
	BP	0.909 3	0.115 3	10.90
	PLS-GRNN	0.911 1	0.113 4	8.53

### 3 结 论

将 PLS-GRNN 应用于豆粕品质多组分含量近红外光谱分析。对小波降噪后的光谱数据进行 iPLS 特征波段提取, 优选出水分、蛋白质和脂肪的最佳建模波段分别为  $4\ 904\sim 5\ 200$ ,  $4\ 304\sim 4\ 600$  和  $4\ 304\sim 4\ 600\ \text{cm}^{-1}$ , 减少了光谱冗余信息, 降低了模型的计算复杂度, 提高了模型效率。建立了豆粕三组分含量的 PLS-GRNN 预测模型, 与 PLS 线性模型和 BP 非线性模型对比, 发现 PLS-GRNN 模型效果最佳, 其水分、蛋白质和脂肪的预测相对标准偏差 RSD 分别为 0.79%, 0.83% 和 8.53%。研究表明基于 PLS-GRNN 的近红外光谱分析用于豆粕品质检测是可行的, 能够用于实际生产过程中的品质监控, 及时调整工艺参数, 以生产出高品质的大豆产品。

### References

[ 1 ] GUO Zhong-yuan, SHEN Shi-lei, ZHOU Xin-qi, et al(郭中原, 慎石磊, 周新奇, 等). Cereal & Feed Industry(粮食与饲料工业), 2020, (4): 46.

[ 2 ] Wei Z, Lin M. Journal of Applied Spectroscopy, 2021, 88(3): 681.

[ 3 ] ZHUANG Shu-hua, ZHANG Guang-li, LU Li-jun(庄树华, 张广利, 卢利军). Oil Crops of China(中国油料), 1991, (2): 77.

[ 4 ] NA Rong(纳 嵘). Animal Husbandry and Feed Science(畜牧与饲料科学), 2017, 38(8): 14.

[ 5 ] Lesson S. Journal of Applied Poultry Research, 1997, (6): 501.

[ 6 ] Fontaine J, Horr J, Schirmer B. Journal of Agricultural and Food Chemistry, 2001, 49: 57.

[ 7 ] WANG Hong-mei, PENG Hai-hong, LIU Jia(王红梅, 彭海宏, 刘 佳). Feed and Animal Husbandry(饲料与畜牧), 2009, (5): 40.

[ 8 ] YANG Zeng-ling, YANG Qin-kai, SHEN Guang-hui, et al(杨增玲, 杨钦楷, 沈广辉, 等). Transactions of the Chinese Society for Agricultural Machinery (农业机械学报), 2019, 50(8): 358.

[ 9 ] Bayer F M, Kozakevicius A J, Cintra R J. Signal Processing, 2019, 162: 10.

[10] Morais Camilo L M, Santos Marfran C D, Lima Kássio M G, et al. Bioinformatics (Oxford, England), 2019, 35(24): 5257.

[11] Tian Han, Zhang Linna, Li Ming, et al. Infrared Physics and Technology, 2018, 95: 88.

[12] Wei Xiao, Zheng Wanqin, Zhu Shiping, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2020, 238: 118453.

[13] Ju Wei, Lu Changhua, Zhang Yujun, et al. Journal of Innovative Optical Health Sciences, 2019, 12(2): 1950005.

[14] Specht D F. IEEE Transactions on Neural Networks, 1991, 2(6): 568.

[15] Kamel A H, Afan H A, Sherif M, et al. Sustainable Computing: Informatics and Systems, 2021, 30: 100514.

[16] Xu Hanxiao, Xu Da, Zhang Naiqian, et al. Journal of Proteome Research, 2021, 20(3): 1657.

[17] Huang Yicheng, Liao Hsienshu. Journal of Intelligent & Fuzzy Systems, 2020, 38(2): 2347.

# Research on Detection of Soybean Meal Quality by NIR Based on PLS-GRNN

WANG Li-qi<sup>1</sup>, YAO Jing<sup>1</sup>, WANG Rui-ying<sup>1</sup>, CHEN Ying-shu<sup>1</sup>, LUO Shu-nian<sup>2</sup>, WANG Wei-ning<sup>2</sup>, ZHANG Yan-rong<sup>1\*</sup>

1. School of Computer and Information Engineering, Harbin University of Commerce, Heilongjiang Provincial Key Laboratory of E-commerce and Information Processing, Harbin 150028, China

2. School of Food Engineering, Harbin University of Commerce, Harbin 150028, China

**Abstract** Soybean meal is a by-product of soybean oil extracted from soybean after proper drying and heat treatment. It is the main raw material for making livestock feed, and its quality determines the nutritional value. There are many problems with existing soybean meal quality detection methods, such as the use of toxic chemical reagents, complex operation, long analysis time, so they cannot meet the needs of rapid detection and control in the production process. This paper proposes a multi-component detection method of soybean meal quality based on near infrared spectroscopy for on-line detection and control of product quality. 449 soybean meal samples were collected from the soybean oil processing line. The chemical values of moisture, protein and fat were determined by 105 °C oven method, Kjeldahl nitrogen determination method and Soxhlet extraction method, respectively. The diffuse reflectance spectra of samples were collected by the Swiss Buchi NIRMaster Fourier Transform near-infrared spectrometer. Firstly, the Mahalanobis distance method was used to remove abnormal samples, and then the spectral denoising was processed by various methods. The results show that the wavelet denoising effect is the best. KS and SPXY algorithms were used to determine the optimal sample partition of different components. In order to investigate the NIR absorption characteristics of soybean meal components, eliminate spectral redundancy and reduce the computational complexity of the model, interval Partial Least Squares (iPLS) was used to extract the features from the whole spectrum of 4 000~10 000  $\text{cm}^{-1}$ . The optimized characteristic absorption bands of moisture, protein and fat were 4 904~5 200, 4 304~4 600 and 4 304~4 600  $\text{cm}^{-1}$ , respectively. Finally, a Generalized Regression Neural Network (GRNN) model was established to predict the component contents of soybean meal. In order to reduce the input variables and the network scale improve the operation speed, PLS was used to reduce the dimension of spectral data, and the principal factor score was extracted as the input variable of GRNN. The PLS-GRNN prediction models of soybean meal multi-component contents were established by optimizing the smooth factor spread through the cross-validation and compared with the classical PLS and BP models. The results show that the PLS-GRNN models are good, the prediction determination coefficients ( $R^2$ ) of moisture, protein and fat are 0.976 9, 0.940 2 and 0.911 1, the Root-Mean-Square Errors of Prediction (RMSEP) are 0.091 2, 0.383 4 and 0.113 4, the Relative Standard Deviations (RSD) of prediction are 0.79 %, 0.83 % and 8.53 %, respectively. Although the prediction error for fat is relatively large, it is also within the available range of the model evaluation criteria. The results show that the near infrared spectroscopy analysis based on PLS-GRNN is feasible to detect soybean meal quality and can be used for quality monitoring in the actual production process.

**Keywords** Soybean meal quality; Near-infrared spectroscopy (NIR); Interval partial least squares (iPLS); Generalized regression neural network (GRNN)

(Received Apr. 14, 2021; accepted Oct. 14, 2021)

\* Corresponding author