饲料中粗脂肪和粗纤维含量的近红外光谱快速分析

郝 勇,吴文辉,商庆园

华东交通大学机电与车辆工程学院, 江西 南昌 330013

摘 要 采用近红外光谱(NIRS)结合偏最小二乘(PLS)方法,实现对饲料中粗脂肪和粗纤维的快速定量分 析。采用 Norris-Williams 平滑求导(NW)和多元散射校正(MSC)方法对光谱进行预处理; 蒙特卡罗无信息变 量消除法(MCUVE)、变量组合集群分析法(VCPA)和区间变量迭代空间收缩法(iVISSA)用于光谱变量选择 和优化; PLS 用于光谱校正模型的建立,采用校正集相关系数(Rc)、交互验证均方根误差(RMSECV)、预测 集相关系数(R。)和预测集均方根误差(RMSEP)评价模型。光谱预处理中经 MSC 处理后的光谱模型优于其 他预处理方法, 其 RMSECV 和 RMSEP 值都减小, R。和 R。值都增大。脂肪定量分析中, 原始光谱模型的 RMSECV 和 R_c 为 0.21 和 0.87, RMSEP 和 R_c 为 0.20 和 0.88, 变量数(V_n)为 1 501; 经 MCUVE 方法选择 变量后建立的定量模型,其 RMSECV 和 R_c 为 0.17 和 0.92, RMSEP 和 R_c 为 0.19 和 0.89, V_n 为 400 个; 经 VCPA 选择变量建立 PLS 定量模型, 其 RMSECV 和 R_c 为 0. 21 和 0. 87, RMSEP 和 R_c 为 0. 25 和 0. 81, V_n 为 12; 经 iVISSA 选择变量后的模型, 其 RMSECV 和 R_c 为 0.21 和 0.86, RMSEP 和 R_c 为 0.20 和 0.87, V_n 为 20。粗纤维定量分析中,原始模型的 RMSECV 和 R_c 为 0.28 和 0.91, RMSEP 和 R_o 为 0.25 和 0.95, V_n 为 1 501; 经 MCUVE 选择后的模型, 其 RMSECV 和 R_c 为 0.23 和 0.95, RMSEP 和 R_c 为 0.23 和 0.94, V_n 为 740; 经 VCPA 选择变量后的模型, 其 RMSECV 和 R_c 为 0.27 和 0.91, RMSEP 和 R_o 为 0.30 和 0.91, V_n 为 11; 经 iVISSA 选择后变量的模型, 其 RMSECV 和 R_c 为 0.29 和 0.90, RMSEP 和 R_c 为 0.27 和 0.93, V_n为 20。结果表明, MSC 方法可以有效提高光谱质量, 消除光谱平移误差; MCUVE 变量选择方法可以简 化模型提高模型精度和稳定性,建立最优模型。在粗脂肪的定量分析模型中, MSC 处理后的光谱经过 MCUVE 选择后剩余 400 个变量, R_e 和 R_p 相较于全谱模型提高了 0.05 和 0.01, RMSECV 和 RMSEP 分别 降低到了 0.17 和 0.19; 经 VCPA 和 iVISSA 选择变量的模型其结果与全谱模型相似, 但其变量分别只有 12 和 20 个。在粗纤维模型中, 经 MCUVE 选择后 740 个变量用于建立 PLS 模型, 其 R。和 R。为 0.95 和 0.94, RMSECV 和 RMSEP 分别为 0.23 和 0.23, VCPA 和 iVISSA 分别运用 11 和 12 个变量建立回归模型, 但结 果都比 MCUVE 模型差。利用饲料近红外光谱建立 MSC-MCUVE-PLS 模型可以有效对饲料粗脂肪和粗纤 维进行定量分析。

关键词 近红外光谱;饲料;蒙特卡罗无信息变量消除法;变量组合集群分析法;区间变量迭代空间收缩法中图分类号: S816.2 文献标识码: A **DOI**: 10.3964/j. issn. 1000-0593(2020)01-0215-06

引言

饲料营养的精准配制不仅可以提高饲料的质量,还可以 节约成本。粗脂肪和粗纤维是饲料主要成分,其含量检测是 饲料精准配制的重要环节,传统的饲料成分方法为湿化学检 测方法其灵敏度和精确度高,但操作复杂、检测时间长、成 本高。近红外技术作为一种快速无污染的检测方法广泛应用 于农业、林业、化工、制药等行业。如 Sagrario Modrono 等运用便携式光谱仪实现了饲料中蛋白质、粗纤维和淀粉等组分的实时测定^[1];Patrica A. Harris 等运用近红外光谱分析方法准确分析了马蹄草的营养成分^[2];Lusia Mandrile 等运用近红外光谱进行饲料中虫类蛋白含量的检测等^[3]。

本文利用近红外光谱实现对饲料的粗脂肪和粗纤维的含量检测,在利用近红外光谱建立含量分析回归模型前,采用光谱预处理方法和变量选择方法来简化模型,以提高模型分

收稿日期: 2018-11-21, 修订日期: 2019-03-19

基金项目: 国家自然科学基金项目(21265006),国家质量监督检验检疫总局科技计划项目(2015IK160),江西省自然科学基金项目(2015ZBAB201003)资助

析精度。

1 实验部分

1.1 样本

收集了不同厂家不同时期生产的饲料样本 325 个,包括 乳猪饲料、中猪饲料和大猪饲料。在采集近红外光谱前先将 样本利用高速万能粉碎机(FW135)粉碎,过 40 目筛,制成粉末状样本。饲料的粗脂肪和粗纤维的实际含量按照国家标准 GB/T 6433—2006 和 GB/T 6434—2006 确定。

1.2 近红外光谱的采集

使用瑞士步琦公司 NIR-Flex N-500 傅里叶近红外光谱 仪,其波数范围为 $10~000\sim4~000~cm^{-1}$ 。为了保证光谱仪的 正常运行和所获光谱的一致性,饲料样本的厚度统一为 4~mm,实验室温度为 $10\sim30~$ °C,湿度为 30%RH $\sim70\%$ RH。本文共获取 352条饲料光谱,每条光谱由 1~501个变量组成。

1.3 近红外光谱预处理方法

建立饲料粗脂肪和粗纤维定量识别模型时,352个样本光谱根据 xy 联合距离划分样本集算法(sample set partitioning based on joint x-y distance, SPXY)按照 2:1 比例划分为校正集和预测集,校正集有 235 个样本,预测集含有 117个样本^[4]。

为消除背景噪声的影响需要对原始光谱进行预处理。常用的光谱预处理方法主要分为两大类:光谱导数法和散射校正法,其中 Norris-Williams 导数法(Norris-Williams derivation, NW)和多元散射校正法(multiplicative scatter correction, MSC)有较好的光谱预处理效果[5-6]。

1.4 近红外光谱变量选择方法

每条饲料的近红外光谱有 1 501 个变量,其包含了饲料的全部信息。变量选择方法通过选择有用变量去除冗余信息来简化模型,提高模型精度。变量选择方法有很多,本文采

用蒙特卡罗无信息变量消除法^[7-9] (Monte carlo based uninformative variable elimination, MCUVE)、变量组合集群分析法^[10-12] (variables combination population analysis, VCPA) 和区间变量迭代空间收缩法^[13-14] (interval variable iterative space shrinkage approach, iVISSA)进行变量选择并比较结果,从而确定最优的变量选择方法。

1.5 模型的评价

光谱经过预处理和变量选择后,利用校正集样品建立饲料粗脂肪和粗纤维含量模型,再利用预测集样品来对模型进行验证,确定模型能否达到粗脂肪和粗纤维的含量分析要求。采用校正集相关系数 R_c 、预测集相关系数 R_p 、交叉验证均方根误差 RMSECV、预测集均方根误差 RMSEP等指标来评价模型的预测精度和稳健性[$^{15-17}$]。

2 结果与讨论

2.1 光谱预处理方法的结果分析

饲料中粗脂肪和粗纤维的实际含量统计信息如表 1 所示。

表 1 饲料中粗脂肪和粗纤维含量统计表(单位:%)

Table 1 Statistical table of crude fat and corasefiber content in feed (unit of content: %)

类别	最大值	最小值	均值	标准差
粗脂肪	4.18	2.62	3.46	0.42
粗纤维	4.66	1.91	3.19	0.68

采用 NW 导数和 MSC 对光谱进行预处理。其中 NW 导数包括平滑法 (NW^{0th}) 、一阶导数法 (NW^{1st}) 和二阶导数法 (NW^{2nd}) 三种方法,同时还利用 NW^{1st} -MSC 方法来进行预处理。如表 2 所示为不同预处理方法的饲料粗脂肪和粗纤维含量模型结果。

表 2 不同预处理方法的饲料粗脂肪和粗纤维含量模型结果

Table 2 The results of crude fat and corase fiber contents in feed with different pretreatment methods

预处理方法	粗脂肪			粗纤维				
	RMSEP	$R_{ m p}$	RMSECV	$R_{ m c}$	RMSEP	$R_{ m p}$	RMSECV	$R_{ m c}$
Original	0.50	0.90	0.74	0.83	0.20	0.88	0.21	0.87
$NW^{0\mathrm{th}}$	0.61	0.89	0.80	0.80	0.21	0.87	0.22	0.86
NW^{1st}	0.59	0.90	0.80	0.81	0.21	0.87	0.22	0.87
NW^{2nd}	0.71	0.86	0.89	0.76	0.23	0.84	0.23	0.85
MSC	0.51	0.92	0.64	0.87	0.19	0.87	0.21	0.87
$NW^{1\mathrm{st}}\text{-}MSC$	0.54	0.91	0.71	0.85	0.20	0.88	0.21	0.87

通过表 2 可以看出,相比与原始光谱的 PLS 模型,经 NW^{0th} , NW^{1st} 和 NW^{2nd} 三种方法处理后的模型都没起到明显的效果。而 MSC 与 NW^{1st} -MSC 方法对饲料噪声去除有明显效果,对粗脂肪模型精度有明显的提升。在粗脂肪模型中,相比于 NW^{1st} -MSC 方法,单独使用 MSC 预处理方法能够更好的实现光谱的信息提取,校正集的 RMSECV 由 0.74 下降到 0.64, R_c 由 0.84 提高到 0.87。在粗纤维模型中,MSC 的

RMSEP 相比于未经处理的有着 0.01 的提升,而 RMSECV 与 R_c 基本相同; NW^{1st} -MSC 处理后的效果与原始光谱模型一致。因此,针对饲料的粗脂肪和粗纤维含量分析的光谱预处理采用 MSC 方法。图 1 为饲料的原始近红外光谱和经过 MSC 预处理后的光谱。MSC 修正了光谱间的相对基线平移和偏移,使光谱更为紧密。

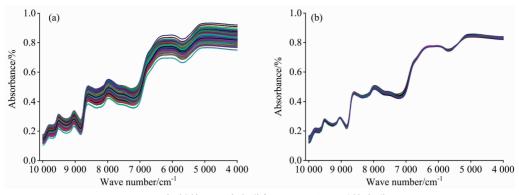


图 1 饲料的近红外光谱与 MSC 预处理后的光谱

Fig. 1 Original NIRS of feed and spectra after pretreatment with MSC

2.2 饲料粗脂肪模型分析

饲料光谱经过 MSC 预处理后,分别利用 MCUVE, VC-PA 和 iVISSA 三种变量选择方法结合 PLS 建立饲料粗脂肪含量分析模型。图 2 为不同变量选择方法的脂肪含量分析PLS 模型校正和测试相关图。

图 2(a) 为利用经过 MSC 预处理的光谱建立的 PLS 模型,其校正集和测试集的回归直线基本重合,相关系数 R_o 和 R_o 分别为 0.87 和 0.88。模型的 RMSECV 和 RMSEP 分别为 0.21 和 0.20。图 2(b) 为 MSC-MCUVE-PLS 模型,图

中可以看出校正集与测试集相比与 MSC-PLS 模型分布更为 紧密, R_c 为 0.92,RMSECV 为 0.17。相比于全谱的模型其校正模型更为稳定,精度更高。同时, R_p 相较于 MSC-PLS 模型提高了 0.01,且 RMSEP 降低了 0.01,变量数 V_n 为 400,大幅度降低。图 2(c) 为 MSC-VCPA-PLS 模型,在校正模型中 R_c 为 0.87,RMSECV 为 0.21 与全谱的 PLS 模型相同,而 R_p 为 0.81,RMSEP 为 0.25,与 MSC-PLS 模型相比起预测结果相差较多。 V_n 为 12 个,小于全谱变量数和 MCUVE筛选后的变量数。图 2(d) 为 MSC-iVISSA-PLS模

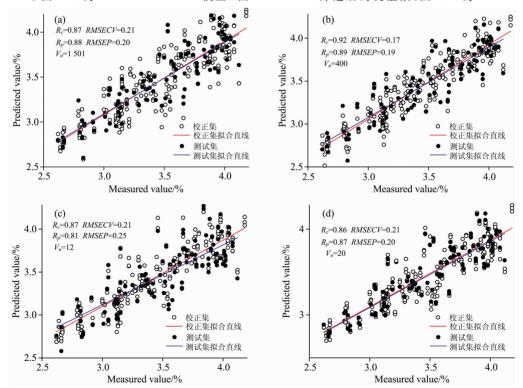


图 2 不同变量选择方法的饲料粗脂肪含量分析 PLS 模型校正和测试相关图 (a): 多元散射模型; (b): 多元散射蒙特卡罗模型;

(c): 多元散射变量组合集群模型; (d): 多元散射区间变量迭代空间收缩模型

Fig. 2 Calibration and test correlation diagrams of PLS model for quantitative analysis of feed fat content with different variable selection methods

(a): MSC-PLS; (b): MSC-MCUVE-PLS; (c): MSC-VCPA-PLS; (d): MSC-iVISSA-PLS

型,在此模型中 R_c 为 0. 86, RMSECV 为 0. 21 与全谱的校正集模型相同,同时预测集的 R_p 为 0. 87, RMSEP 为 0. 20 且与全谱的预测值相同, V_n 为 20 与 VCPA 选择的变量数近似。

通过图 2 可以看出三种变量选择方法中,经 MCUVE 方法变量选择后,其模型的各个参数都优于全谱模型。经 VC-PA 和 iVISSA 方法选择后都以极少的变量建立检测模型且与全谱模型几乎相同。若精度要求不高时,可以选择 VCPA和 iVISSA 方法进行变量选择,建立更简洁的模型;若要提升精度,MCUVE则是最好的选择。

2.3 饲料粗纤维模型分析

如图 3 为经不同变量选择方法的粗纤维定量分析模型相关图。图 3(a) 为经 MSC 预处理后由全谱 1501 个变量建立的 PLS 模型,其校正集相关系数 R_c 为 0.91,RMSECV 为 0.28,在测试集中 R_p 为 0.95,RMSEP 为 0.25。图 3(b) 为经过 MCUVE 变量选择后的模型,选择后剩余 740 个变量用于模型分析,其 R_c 和 R_p 分别为 0.95 和 0.94,较全谱模型的校正集 R_c 提高了 0.04。同时 RMSECV 为 0.23,较全谱模型

降低了 0.05, RMSEP 降低了 0.02, V_n 为 740, 由此可知 MCUVE 变量选择法有效地选择了有用变量,提高了模型的 精度。

图 3(c)为经过 VCPA 变量选择后建立的 PLS 模型,由图可知,经过 VCPA 选择了 11 个变量来建立模型,相比与全谱和 MCUVE 方法,VCPA 用极少的变量实现了几乎相同的结果,其校正集 R_c 为 0.91,RMSECV 为 0.27 略好于全谱校正模型,预测集 R_p 和 RMSEP 都差于全谱。图 3(d) 为 iVISSA 变量选择后的饲料粗纤维检测模型,模型的建立基于 20 个变量,其 R_c 为 0.90,RMSECV 为 0.29, R_p 为 0.93,RMSEP 为 0.27 相比与全谱和 MCUVE 模型较差,但较于 VCPA 模型更好。

对比三种变量选择方法,可以看出 MCUVE 变量选择方法校正集模型参数优于全谱模型,而预测集的参数与全谱预测集参数几乎一致,因其变量数少于全谱,则模型更为简洁。利用 VCPA 和 iVISSA 进行变量选择后,都剩余极少变量,部分有用信息变量被剔除,因而其模型参数较全谱都稍差一些,其优点是模型简单。

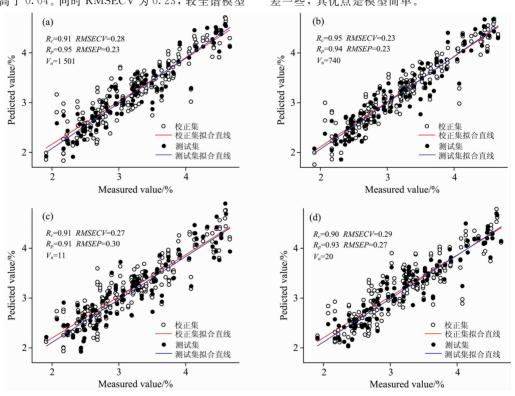


图 3 不同变量选择方法的饲料粗纤维含量分析 PLS 模型校正和测试相关图 (a):多元散射模型;(b):多元散射蒙特卡罗模型;

(c): 多元散射变量组合集群模型; (d): 多元散射区间变量迭代空间收缩模型

Fig. 3 Calibration and test correlation diagrams of PLS model for quantitative analysis of corase fiber content with different variable selection methods

(a): MSC-PLS; (b): MSC-MCUVE-PLS; (c): MSC-VCPA-PLS; (d): MSC-iVISSA-PLS

3 结 论

近红外光谱分析方法用于饲料中粗脂肪和粗纤维的定量

分析,采用不同的光谱预处理方法和光谱变量选择方法对光谱模型进行优化。结果表明,在饲料粗脂肪定量分析中,MSC 预处理方法结合 MCUVE 变量选择方法可以建立最优粗脂肪分析模型,模型变量数由 1501 降低到 400, RMSEP

和 RMSECV 分别降低为 0.19 和 0.17, R_p 和 R_c 提高到 0.89 和 0.92; 在饲料粗纤维定量分析中,利用 MSC 预处理方法 和 MCUVE 变量选择方法可以实现最优的粗纤维定量分析模型,模型变量数为 740,模型的 RMSEP 和 RMSECV 分别

为 0. 23 和 0. 23, R_p 和 R_c 分别为 0. 94 和 0. 95。因此,基于近红外光谱建立的 MSC-MCUVE-PLS 模型可以有效的进行饲料的粗脂肪和粗纤维的定量分析,其精度高,稳定性好且更简洁。

References

- [1] Sagrario M, Ana S, Adela MF, et al. Talanta, 2017, 162: 597.
- [2] Patrica A H, Sarah N, Harry B C, et al. Journal of Equine Veterinary Science, 2018, 71: 13.
- [3] Luisa M, Ilaria F, Giuseppina A, et al. Food Chemistry, 2018, 267: 240.
- [4] Kelton S S, Anderson S S, Telma W L, et al. Journal of Computer Science, 2015, 11(4): 621.
- [5] Hideyuki S, Junji M. Spectrochimica Acta Part A; Molecular and Biomolecular Spectroscopy, 2018, 192; 236.
- [6] Asmund R, Frans B, Soren B E. Trebds in Analytical Chemistry, 2009, 28(10): 1201.
- [7] Li C, Zhao T L, Li C, et al. Food Chemistry, 2017, 221: 990.
- [8] Liu X W, Cui X Y, Yu X M, et al. Chinese Chemical Letters, 2017, 28: 1447.
- [9] WU Jing-zhu, WANG Feng-zhu, WANG Li-li, et al(吴静珠,汪凤珠,王丽丽,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2015, 35(1): 99.
- [10] Yun Y H, Wang W T, Deng B C, et al. Analytica Chimica Acta, 2015, 862: 14.
- [11] Yang Y N, Xie F F, Yan B, et al. Chemometrics and Intelligent Laboratory Systems, 2017, 170: 102.
- [12] ZHAO Huan, HUAN Ke-wei, ZHENG Feng, et al(赵 环, 宦克为, 郑 峰, 等). Journal of Changchun University of Science and Technology(长春理工大学学报·自然科学版), 2016, 39(5): 51.
- [13] Deng BC, Yun YH, Liang YZ, et al. Analyst, 2014, 139: 4836.
- [14] Deng BC, Yun YH, MaP, et al. Analyst, 2015, 140: 1876.
- [15] Ripoll G, Lobon S, Joy M. Meat Science, 2018, 143: 24.
- [16] Fien D L, Elisabeth P, Hasna D, et al. Journal of Pharmaceutical and Biomedical Analysis, 2018, 151; 274.
- [17] Pedro S S, Andreia S, Ana C, et al. Food Chemistry, 2018, 242: 196.

The Research on Quantitative Analysis of Feed Crude Fat and Corase Fiber Based on Near Infrared Spectroscopy and Variables Selection Methods

HAO Yong, WU Wen-hui, SHANG Qing-yuan

School of Mechatronics & Vehicle Engineering, East China Jiaotong University, Nanchang 330013, China

Abstract Near infrared spectroscopy (NIRS) combined with partial least squares (PLS) method was used to achieve rapid quantitative analysis of crude fat and corase fiber in feed. The norris-williams derivation (NW) and multiplicative scatter correction (MSC) methods were used to pretreat the spectrum, and the monte carlo based uninformative variable elimination (MCUVE), variables combination population analysis (VCPA) and interval variable iterative space shrinkage approach (iVISSA) were used to select and optimize the variables of the spectrum. PLS was used for the establishment of the spectral calibration model, and the parameters of calibration set correlation coefficient (R_c) , root mean square error of cross validation (RMSECV), prediction set correlation coefficient (R_p) and root mean square error of prediction (RMSEP) were used to evaluate the models. Compared with other pretreatment methods, the RMSECV and RMSEP values of the spectral model after MSC treatment decreased, while the R_c and R_p values increased. In the crude fat quantitative analysis model, the RMSECV and R_c of the original spectral model were 0.21 and 0.87, RMSEP and R_p were 0.20 and 0.88, and the number of variables (V_n) was 1501. After selecting variables by MCUVE method, RMSECV and R_c were 0.17 and 0.92, RMSEP and R_p were 0.19 and 0.89, and V_n was 400. For VCPA-PLS model, the RMSECV and R_c were 0.206 and 0.87, RMSEP and R_p were 0.25 and 0.81, and V_n was 12. For iVISSA-PLS model, the RMSECV and R_c were 0. 21 and 0. 86, RMSEP and R_v were 0. 20 and 0. 87, and V_n was 20. In the corase fiber model, the RMSECV and $R_{\rm c}$ of the original PLS model were 0.28 and 0.91, RMSEP and $R_{\rm p}$ were 0.23 and 0.95, and $V_{\rm n}$ was 1501. After MCUVE selection, the RMSECV and R_c of the model were 0.23 and 0.95, RMSEP and R_p were 0.25 and 0.94, and V_n was 740. After VCPA selection, the RMSECV and R_c of the model were 0.27 and 0.91, RMSEP and R_p were 0.30 and 0.91, and V_n was 11. After iVISSA selection, the RMSECV and R_c of the model were 0.29 and 0.90, RMSEP and R_p were 0.27 and 0. 93, and V_n was 20. The results showed that the MSC method could effectively improve the spectral quality and eliminate the spectral translation error; the MCUVE variable selection method could simplify the model to improve the model accuracy and stability, and establish the optimal model. In the crude fat quantitative analysis model, after the MSC-processed spectrum was selected by MCUVE, the remaining 400 were used to establish the PLS model, R_c and R_p were improved by 0. 05 and 0. 01 compared to the full-spectrum model, and the RMSECV and RMSEP were reduced to 0. 17 and 0. 19; The model selected by VCPA and iVISSA had almost the same result as the full-spectrum model, and its greatest feature was that only 12 and 20 variables were selected. In the corase fiber model, 740 variables selected by MCUVE were used to establish the PLS model with R_c and R_p of 0. 95 and 0. 94, RMSECV and RMSEP of 0. 23 and 0. 23, respectively; VCPA and iVISSA used 11 and 12 variables to establish the regression model, but its model results were all worse than the MCUVE model. The establishment of MSC-MCUVE-PLS quantitative analysis model using feed near-infrared spectroscopy could effectively quantify crude fat and corase fiber in feed.

Keywords Near infrared spectroscopy (NIRS); Feed; Monte carlo based uninformative variable elimination (MCUVE); Variables combination population analysis (VCPA); Interval variable iterative space shrinkage approach (iVISSA)

(Received Nov. 21, 2018; accepted Mar. 19, 2019)

本刊声明

近期以来,一些不法分子假冒《光谱学与光谱分析》期刊社名义,以虚假网站等形式欺骗广大作者、读者。这些虚假网站 公然假冒《光谱学与光谱分析》期刊名义进行大肆的征稿并骗取作者的审稿费和版面费。经部分作者及读者举报,现有关部门 已就此介入调查。我刊将通过法律途径向假冒者追究相应的责任,维护本刊权利。

本刊官方网站已正式开通, 网址为

http://www.gpxygpfx.com/

在此郑重声明,本网址为《光谱学与光谱分析》期刊唯一开通运行的官方网站。本刊从未授权任何单位或个人以任何形式(包括网上网下)代理本刊征稿、审稿等项业务。

希望广大读者和作者切实维护好自身的合法权益,防止受骗上当。

《光谱学与光谱分析》期刊社 2019年3月15日