

文章编号: 1672-8785(2017)11-0044-05

纸张含水量的傅里叶变换 中红外光声光谱分析

马赵扬¹ 杜昌文²

(1. 中国国家图书馆, 北京 100081 ;

2. 中国科学院南京土壤研究所, 江苏南京 210008)

摘 要: 随着科学技术的不断发展, 社会正在向智能化演变。实现纸质文献的实时监测以达到保存和保护文献的目的是未来发展的必然趋势。为了给库房文献的实时监测提供依据, 结合化学计量学方法, 利用傅里叶变换中红外光声光谱技术快速测定了纸张含水量。基于中红外光声光谱, 分别采用主成分回归 (Principal Components Regression, PCR) 方法和偏最小二乘回归 (Partial Least Squares Regression, PLSR) 方法建立了纸张含水量的定量模型, 并通过交叉验证选择了最佳主成分数。结果表明, PLSR 方法的建模结果优于 PCR 方法 (决定系数 $R^2: 0.3681 > 0.3532$)。通过增加主成分数可以使模型预测变好, 但也存在过拟合风险。未来拟采集更多的纸张样本, 以期建立稳定的纸张性质检测模型, 为实现纸质文献的红外光谱实时监测奠定基础。

关键词: 纸张含水量; 红外光声光谱; 主成分回归; 偏最小二乘回归

中图分类号: G250.7; O657.3 **文献标志码:** A **DOI:** 10.3969/j.issn.1672-8785.2017.11.009

Determination of Paper Moisture Content by Infrared Photoacoustic Spectroscopy in Fourier Transform

MA Zhao-yang¹, DU Chang-wen²

(1. National Library of China, Beijing 100081, China;

2. Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China)

Abstract: With the progress of science and technology, society is evolving toward intelligence. To use real-time paper literature monitoring to realize the preservation of paper literature will be an inevitable development trend in the future. In order to provide the basis for the real-time monitoring of library literature, Fourier transform medium infrared photo-acoustic spectroscopy is used to determine the water content in paper in combination with stoichiometry. A quantitative model of paper water content is established by Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR) respectively on the basis of medium-infrared photoacoustic spectroscopy. The optimum primary component number is selected by cross-validation. The result shows that the modeling result of partial least squares regression is better than that of the principal component regression (the determination coefficient is $0.3681 > 0.3532$). By increasing the number of principal component, the prediction of the models will be better. However, there are some risks of fitting. In the future, more paper samples will be collected so as to establish a stable paper quality detection model and lay a basis for the real-time infrared spectrum

收稿日期: 2017-07-24

基金项目: 中国国家图书馆青年项目 (NLC-KY-2015-36)

作者简介: 马赵扬 (1987-), 女, 河南焦作人, 馆员, 主要从事历史文献学和红外光谱研究。

E-mail: mzyllxiaoma@126.com

monitoring of paper literature.

Key words: paper moisture content; infrared photoacoustic spectroscopy; principal component regression; partial least squares regression

0 引言

纸张含水量是纸质文献保存过程中一个非常重要的参数,其大小关乎纸张寿命的长短,即纸质文献的保存年限。目前,文献库房的湿度一般保持在 45% ~ 65% 之间。纸张含水量通过影响纸张的伸缩率来改变其理化性质^[1]。含水量过高,纸张容易发霉腐烂;含水量过低,纸张韧性下降,导致发脆易碎。在相同的空气相对湿度下,不同类型的纸张具有不同的含水量^[2]。同样,纸张老化也是一个不可避免的过程,而且不同老化程度纸张的含水量也有所不同^[3]。在纸张的老化过程中,纸张的吸湿性降低,含水量也有所下降。

目前,纸张含水量测量方法包括重力法、红外吸收法和微波损耗法等^[4]。国家标准规定的水分检测方法是烘箱法(GB/T 462-2008 纸、纸板和制浆分析试样水分的测定)即重力法。这三种方法在纸张含水量的测定方面已经有了很好的应用。其中,烘箱法是最传统的测定方法。本文采用的对比方法即为烘箱法。

传统的纸张含水量测定方法具有一定的局限性,不仅需要提供纸张样本(成品文献需裁剪样本),而且在检测过程中还会对纸张造成一定程度的破坏。作为一种无损的光谱检测技术,中红外光声光谱技术通过扫描测定纸张大部分的物质信息,包括纸张的含水量及纤维素等物质组成信息。因此,本文利用中红外光声光谱对纸张含水量进行测定,以判断傅里叶中红外光声光谱技术在纸张性能检测方面是否具有有良好的应用前景,从而为利用该技术分析纸张的其他性能奠定基础,并为实现纸质文献的红外光谱实时监测作好准备。

1 实验部分

1.1 材料

供试纸张为新闻纸,采集于国家图书馆期刊保存本库房及私人藏书,总共 42 个样本。其中包含 16 个民国纸张样本和 26 个 1990 年以后的纸张样本。另外,本次试验样本还包含不同程度的破损类别(五级破损均有)以及无破损纸张;纸张类型有新闻纸、铜版纸、书皮纸和宣纸等。

1.2 烘箱法

将试样放入已烘干至恒重的铝盒中并标记序号,然后合上盖子称总重量。打开容器的盖子,并将两者一起放入 105±2 °C 的烘箱中进行烘干。当烘干结束后,立即将容器盖好并移入干燥器中;冷却 30 min 后重新称重。计算烘干前后样本减少量与原质量的百分比,测试结果准确至 0.1%。天平采用千分之一天平。

1.3 纸张样本含水量的统计分析

分别利用 SPSS 软件中的 Q-Q 图法和非参数检验方法检验数据是否符合正态分布。

1.4 红外光声光谱

本次红外光谱扫描采用美国 Thermo Fisher Scientific 公司生产的 Nicolet 380 型傅里叶变换红外光声光谱仪以及 DTG 检测器;采用美国 MTEC 公司生产的 PA 300 型光声池;动镜速率为 0.3165 cm/s,扫描分辨率为 4 cm⁻¹,扫描次数为 32 次。将纸张样品放入直径为 10 mm、高度为 5 mm 的圆柱形不锈钢样品池中。每隔半个小时以碳黑作为一次背景,扫描范围为 4000 ~ 400 cm⁻¹ 中红外波段。在光谱预处理方面,采用 Matlab R2009a 软件进行平滑去噪及标准化处理。

1.5 模型建立与预测

利用主成分分析简化数据,并提取光谱信息。分别采用 PCR 方法和 PLSR 方法建立纸张含水量的红外光谱模型。这两种模型均采用 Matlab 多元校正工具箱建立^[5]。

1.5.1 主成分分析

主成分分析(Principal Component Analysis, PCA)是 Hotelling 于 1933 年提出的一种多元统

计分析技术^[6]。它通过变换原始数据的特征空间,生成维数更低且新生成的各变量之间相互独立的特征空间。原始变量与新生成的变量之间线性相关。PCR 方法利用主成分分析作回归分析。

1.5.2 偏最小二乘回归

PLSR 是一种针对待测变量的主成分分析方法,它有针对性地对光谱矩阵进行主成分分析。根据携带待测变量信息的多少,各个主成分从大到小依次排列。由于 PLSR 也是一种线性分析方法,各个统计变量之间线性相关^[7]。

1.5.3 交叉验证法

交叉验证是一种用于建立模型的方法。通过交叉验证选择最佳主成分数,可以达到建立最优模型的效果^[8]。本文实验选用 10 折交叉验证。

1.6 模型评估指标

为了比较不同方法所建模型的精度,采用实测值和预测值的线性决定系数 (R^2)、预测的标准误差 (Root Mean Square Error, RMSE) 以及残留预测偏差 (Ratio of Performance to Deviation, RPD) 作为评价指标。RPD 是样本实测值的标准偏差与样本预测值的标准预测误差的比值,其数值及模型优劣情况见表 1。

表 1 模型评价指标

RPD 值	模型精度
<1.0	模型或预测极差
1.0 ~ 1.4	模型或预测较差
1.4 ~ 1.8	模型或预测可用
1.8 ~ 2.0	模型或预测较好
2.0 ~ 2.5	模型或预测很好
>2.5	模型或预测极好

2 结果与分析

2.1 用烘箱法测得的纸张含水量

表 2 列出了供试样本的含水量统计特征。其中,含水量最低为 2.03%,最高为 10.53%。利用 SPSS 软件中的 Q-Q 图法以及非参数检验方法检测了样本的分布情况,其结果呈近似正态分布。

表 2 样本的含水量统计

样本数	最小值	最大值	平均值	标准差
42	2.03%	10.53%	5.03%	2.09%

2.2 样本的中红外光声光谱特征

图 1 为所有供试样本的中红外光声光谱图。其中, 3400 cm^{-1} 处的吸收主要由纤维素分子中 O-H 的伸缩振动产生; 2900 cm^{-1} 处的吸收主要由 C-H 的伸缩振动产生; $1650\sim 1700\text{ cm}^{-1}$ 的吸收由木质素分子中 C=O 的伸缩振动产生^[9]。而水的特征吸收主要在 3400 cm^{-1} 和 1600 cm^{-1} 附近。

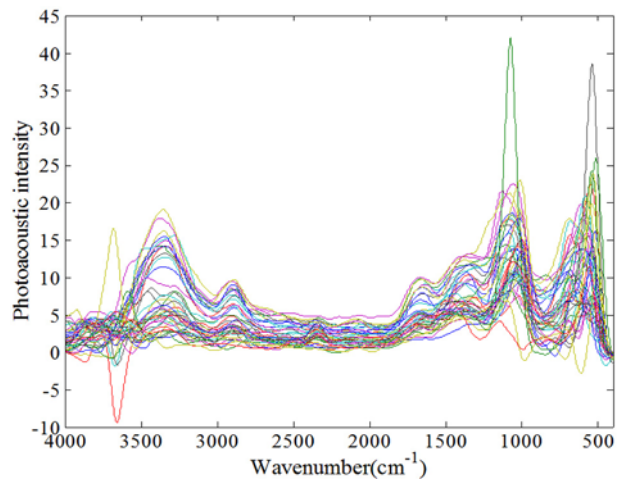


图 1 纸张样本的中红外光声光谱图

2.3 利用 PCR 方法建立模型

基于红外光谱扫描数据,利用 PCR 方法建立模型。这里的建模数据为该波段的全部数据,即 $4000\sim 400\text{ cm}^{-1}$ 范围内的所有数据信息。通过交叉验证,选取最佳主成分数为 6。利用前 6 个主成分建立模型。图 2 所示为通过主成分回归所建模型的纸张含水量的实测值与预测值的相关性。所建模型的 R^2 值为 0.3532, $RMSE$ 值为 1.72%, RPD 值为 1.22。结果表明,该模型的预测效果较差。

中红外区全波段的数据信息中包含许多跟纸张含水量无关的信息。因此,我们选择与纸张含水量相关性较强的两个区域 ($2600\sim 3700\text{ cm}^{-1}$ 和 $1500\sim 2000\text{ cm}^{-1}$) 重新建立模型。图 3 所示为所建模型的纸张含水量的实测值与预测值的相

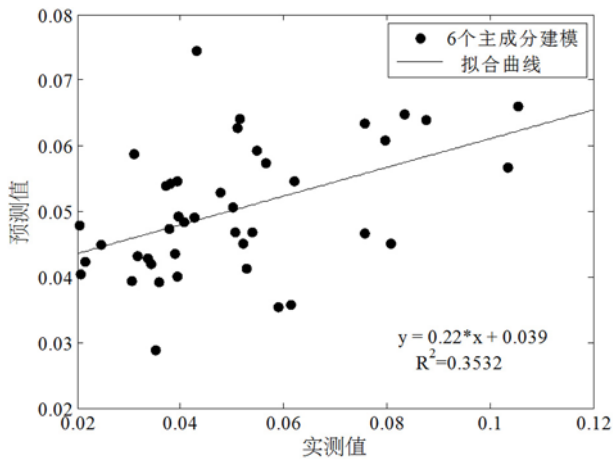


图 2 纸张含水量的实测值与预测值的相关性

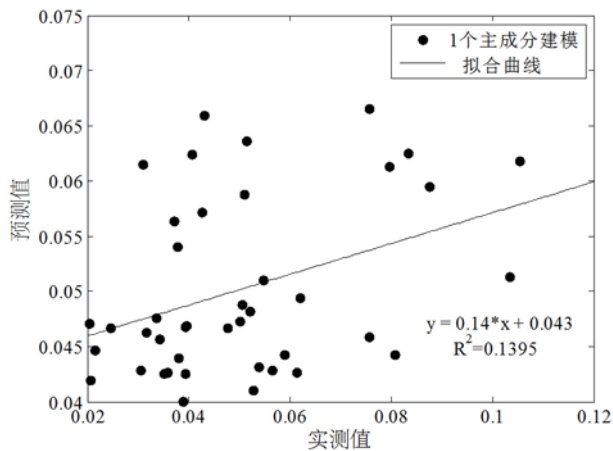


图 3 纸张含水量的实测值与预测值的相关性

关性。通过交叉验证选择最佳主成分数为 1。所建模型的 R^2 值为 0.14, $RMSE$ 值为 1.99%, RPD 值为 1.06。结果表明, 该模型的预测效果较差。与全波段建模结果相比, 选择部分区域进行建模的结果更差。这可能是由于在建模区域的选择上有所偏差, 说明与纸张含水量明显相关的波段无法替代全波段作为预测波段; 另外也表明 PCR 方法在本次试验中的建模效果较差。

2.4 利用 PLSR 方法建立模型

由于用 PCR 方法所建模型的结果较差, 我们采用 PLSR 方法建立模型。该方法是一种针对待测变量的主成分分析方法, 其主成分根据与待测变量的相关性由大到小依次排列。用 PLSR 方法所建模型的针对性更强。为了使模型的预

测线性结果和误差结果能够同时达到最优, 我们使用交叉验证方法选择最佳主成分数为 2, 前两个主成分携带 36.81% 的光谱信息。利用前两个主成分建立模型。图 4 所示为所建模型的纸张含水量的预测值与实测值的相关性。该模型的 R^2 值为 0.37, $RMSE$ 值为 1.7%, RPD 值为 1.23。结果表明, 所建模型的预测效果较差。随着用于建立模型的主成分数的增加, 模型的 R^2 值逐渐增大甚至达到 95% 以上, $RMSE$ 值逐渐减小。但模型的适用性随之变差, 因此存在过度拟合的风险。

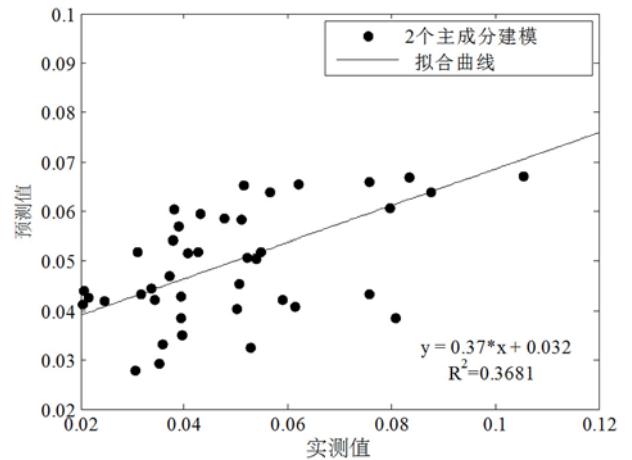


图 4 纸张含水量的实测值与预测值的相关性

通过对比 PCR 方法和 PLSR 方法的建模结果可知, 基于后者 2 个主成分建立的模型优于基于前者 6 个主成分建立的模型, 表明用 PLSR 方法所建模型的针对性更强。

2.5 模型的预测分析

随机抽取 75% 的样本并将其作为校正样本, 同时将其余 25% 作为测试样本。表 3 列出了校正样本和测试样本的纸张含水量的统计结果。

利用 PLSR 方法建模, 并通过交叉验证选取最佳主成分数为 2。所建模型的 R^2 值为 0.4566, $RMSE$ 值为 1.61%, RPD 值为 1.33。将测试样本的光谱数据代入模型当中, 计算测试样本的预测结果。其 $RMSE$ 值为 2.15%, 实测值与预测值的 R^2 值为 0.02。结果表明, 该模型的预测效果较差。

表 3 校正样本和测试样本的
纸张含水量的统计结果

含水量 (%)	样本数	最小值	最大值	均值	标准差
校正样本	32	2.15%	10.53%	5.10%	0.02
测试样本	10	2.03%	8.06%	4.73%	0.02

3 结束语

通过对比 PCR 方法和 PLSR 方法的建模结果可知, 后者所建模型的效果更好。无论是 R^2 值 ($0.3681 > 0.3555$) 还是 $RMSE$ 值 ($1.7\% < 1.72\%$), 它都要优于前者所建模型, 且只用了前两个主成分。PLSR 方法对纸张含水量的针对性更强。但这两种方法所建模型的 RPD 值均较低, 表明其模型效果都不是很好, 只是 PLSR 方法所建模型的 RPD 值稍高于 PCR 方法所建模型。

随机抽取 75% 的校正样本建立模型, 并对其余 25% 的测试样本进行了预测。结果表明, 该模型的预测效果较差, 其测试样本的 R^2 值仅为 0.02。

究其原因, 可能有两点: (1) 供试样本之间差异较大。本次试验的供试纸张有 17 个民国纸张样本, 其破损程度较高; 另有 26 个 1990 年以后的纸张样本。由于样本之间年代相差较大, 不同时代的造纸方法的差异使得纸张本身的组成就大不相同, 且均有不同程度的破损。(2)

供试样本的数量偏少。本次试验的样本数为 42 个。较少的样本数量不利于模型的建立。

我们未来拟采集更多的纸张样本, 并对包括纸张含水量在内的更多的纸张性质进行分析, 以期建立稳定的纸张性质检测模型, 为实现纸质文献的红外光谱实时监测奠定基础。

参考文献

- [1] 李志宏, 黄可嘉. 从纸张的化学组成谈图书、档案资料的保护 [J]. 吉林省教育学院学报, 2005, 21(2): 12-15.
- [2] 侯刚健, 赵军. 浅析纸张的老化因素与高校图书馆应采取的措施 [J]. 科技情报开发与经济, 2010, 20(12): 50-51.
- [3] 张志清. 文献纸张的自然老化及其研究 [J]. 档案学研究, 1998, 12(2): 60-61.
- [4] Anderson J G. 纸张含水量测量 [J]. 国外计量, 1992, 20(5): 4-5.
- [5] 熊磊, 于东伟. 酸处理后纤维素分子结构的显微红外光谱分析 [J]. 纤维素科学与技术, 2013, 21(2): 59-62.
- [6] 许禄, 邵学广. 化学计量学方法 (第二版) [M]. 北京: 科学出版社, 2004.
- [7] 王惠文. 偏最小二乘回归方法及应用 [M]. 北京: 国防科技出版社, 1996.
- [8] 范永东. 模型选择中的交叉验证方法综述 [D]. 山西: 山西大学, 2013.
- [9] McCarty G W, Reeves J B. Comparison of Near Infrared and Mid Infrared Diffuse Reflectance Spectroscopy for Field-scale Measurement of Soil Fertility Parameters [J]. Soil Science, 2006, 171(2): 94-102.
- [8] 杜召杰. 光纤声光调制测振技术的研究 [D]. 淄博: 山东理工大学, 2016.
- [9] 所睿, 范志军, 李岩, 等. 双频激光干涉仪技术现状与发展 [J]. 激光与红外, 2004, 34(4): 251-253.
- [10] 杜召杰, 王辉林. 基于光纤声光调制的振动位移检测技术研究 [J]. 激光杂志, 2015, 36(12): 85-87.
- [11] 王辉林, 王迎春. 双光束激光多普勒测速实验系统的不确定度分析 [J]. 计量学报, 2013, 34(1): 1000-1158.
- [12] 邹思成, 王刚. 温度等因素对 CCD 数据采集电路性能影响的研究 [J]. 天文研究与技术, 2006, 3(3): 275-282.
- [13] 朱文营. 加工中心主轴在线监测技术研究 [D]. 淄博: 山东理工大学, 2012.
- [14] 费业泰. 误差理论与数据处理 [M]. 北京: 机械工业出版社, 2010.

(上接第 38 页)

- [2] 黄清明, 徐小萍, 白文华. 全跳动综合检查仪的设计与误差分析 [J]. 机械工程师, 2010, 20(4): 51-54.
- [3] 刘来超. 某型传动轴外径及圆跳动测量机的设计 [D]. 秦皇岛: 燕山大学, 2014.
- [4] Lu D F, Yuan F, Cheng Z. Measurement of Roughness inside Aperture [C]. Harbin: The Second International Symposium on Istruction Science and Technology, 2002.
- [5] 高传梅, 王辉林. 基于光纤声光调制的轮廓检测技术研究 [J]. 应用光学, 2013, 34(6): 1062-1066.
- [6] 浦昭邦, 赵辉. 光电测试技术 (第 2 版) [M]. 北京: 机械工业出版社, 2013.
- [7] 戴高良. 双频激光纳米干涉测量技术的研究 [D]. 北京: 清华大学, 1998.