

不同物理形态烟叶样品的模型转移研究

谭观萍^{1a,1b}, 易克², 段志超², 张发明³, 宾俊^{1a,1b}, 范伟^{1a,1b}, 周冀衡^{1a,1b*}, 王子维^{1a,1b}

(1.湖南农业大学 a.烟草研究院; b.生物科学技术学院, 湖南 长沙 410128; 2.湖南中烟工业有限责任公司, 湖南 长沙 410014; 3.云南省烟草公司保山市公司, 云南 保山 678000)

摘 要: 采用近红外光谱仪采集整烟叶、卷烟丝和烟叶粉末光谱数据, 并采用流动分析仪测定烟碱含量, 利用偏最小二乘法建立烟叶烟碱的光谱预测模型, 再通过斜率截距算法(SBC)、分段直接标准化算法(PDS)和典型相关分析算法(CCA)3 种模型转移算法, 将整烟叶、卷烟丝和烟叶粉末便携式近红外光谱转移到粉末状烟叶傅立叶近红外光谱模型上, 比较分析预测均方根误差值(RMSEP)。结果表明: 烟叶粉末烟碱近红外光谱预测模型经 SBC、PDS 和 CCA 算法模型转移后的 RMSEP 值分别为 0.741 0、0.736 5、0.298 2, 卷烟丝的 RMSEP 值分别为 0.725 0、0.513 2、0.222 2, 整烟叶的 RMSEP 值分别为 0.712 6、0.446 6、0.333 9, CCA 算法模型转移优于 SBC 与 PDS 算法。

关 键 词: 烟草; 近红外光谱; 模型转移; 典型相关分析算法

中图分类号: S572.01

文献标志码: A

文章编号: 1007-1032(2017)02-0151-05

Calibration transfer on tobacco NIR spectra with different physical states

TAN Guanping^{1a, 1b}, YI Ke², DUAN Zhichao², ZHANG Faming³, BIN Jun^{1a, 1b}, FAN Wei^{1a, 1b}, ZHOU Jiheng^{1a, 1b*}, WANG Ziwei^{1a, 1b}

(1.a.Institute of Tobacco; b.College of Bioscience and Biotechnology, Hunan Agricultural University, Changsha 410128, China; 2. Hunan Zhongyan Industrial Co., Ltd, Changsha 410014, China; 3.Tobacco Company of Baoshan City, Yunnan, 678000, China)

Abstract: In order to use a portable near infrared spectrometer preferably and quickly to conduct the real-time on scene test in tobacco production, the measured spectrum was usually moved to the spectral model already established in laboratory by mathematical method to directly conduct the forecast analysis, which saves cost compared with rebuilding a model. In this paper, slope intercept algorithm (SBC), piecewise direct standardization (PDS) and canonical correlation analysis (CCA) were used and compared in transfer of portable near infrared spectrum of the whole leaf, cigarette silk and powder to the powdered tobacco on Fourier near-infrared spectroscopy model, followed by predicting values of root mean square error (RMSEP). The results showed that RMSEP value of near infrared spectrum of the powder, cigarette silk and whole leaf calibration transferred by SBC algorithm were 0.736 5, 0.798 8, 0.298 6 respectively, by PDS were 0.889 2, 0.640 9, 0.300 7 respectively, by CCA were 0.716 5, 0.761 3, 0.536 9, respectively. RMSEP results after calibration transfer with CCA were better than those with SBC and PDS, and better the RMSEP predicted by rebuilt model using the portable near infrared spectrum directly.

Keywords: tobacco; near-infrared spectroscopy; calibration transfer; canonical correlation analysis

从近红外光谱中提取化学信息^[1-3]常采用主成分回归(PCR)^[4]和偏最小二乘回归(PLS)^[5]的多元校正技术。目前, 烟草化学成分的近红外光谱模型大多数是基于傅立叶近红外光谱仪采集的烟叶粉末光谱所建^[6], 所建模型稳健, 但因仪器较大, 移动可能影响仪器精度且使用不便, 难以实现现场检

测。便携式近红外光谱仪在使用过程中, 需对采集到的烟叶光谱信息重新建立模型, 如若通过化学计量学方法将便携式近红外光谱仪采集的光谱转移至傅立叶近红外光谱模型上, 将有可能节省人力物力, 加快实时分析检测速度。Y. QIN 等^[6]通过混合建模和模型转移技术, 将整烟叶光谱转移到烟叶粉

收稿日期: 2016-12-01

修回日期: 2017-01-04

基金项目: 中国烟草总公司云南省公司科技项目(2015YN22)

作者简介: 谭观萍(1992—), 女, 湖南株洲人, 硕士研究生, 主要从事烟草品质化学研究, tguanping22@163.com; *通信作者, 周冀衡, 博士, 教授, 主要从事烟草生物科学与工程技术研究, jhzhou2005@163.com

末光谱模型中;苏虹^[7]对同一台近红外光谱仪采集整烟叶和烟叶粉末光谱进行混合建模后,再对混合光谱进行模型转移;杨凯等^[8]利用 2 台仪器对原烟的近红外光谱进行了模型转移;郑开逸等^[9]通过模型转移不同仪器和不同时间测量的整烟叶近红外光谱。

笔者尝试在已有傅立叶近红外光谱仪模型数据的基础上,运用便携式近红外光谱仪采集整烟叶、卷烟丝和烟叶粉末的烟叶光谱。通过模型转移技术对烟叶烟碱进行在线分析,以便更好地使便携式近红外光谱仪应用到生产现场,实现实时快速精确检测分析烟叶的烟碱含量。

1 材料与方法

1.1 材料

烟叶品种为 K326。选取烘烤后的完整烟叶、经 QS-2A 烟叶切丝机处理的卷烟丝、经高速粉样机处理的烟叶粉末样品各 75 份。

1.2 方法

1.2.1 烟碱含量测定

参考文献[10],用流动分析仪 PULSE 3000 测定烟叶粉末样品的烟碱含量。

1.2.2 近红外光谱数据的采集

仪器 A 为 Antaris II 傅立叶变换近红外光谱仪(主仪器 A),光谱采集方式为积分球漫反射^[11-13]。

采集烟叶粉末样品的傅立叶变换光谱数据,共测量 3 次,光谱数据取平均值。

仪器 B 为 B&WTEK i-Spec 近红外光谱仪(从仪器 B),光谱采集方式为探头漫反射^[14]。采集 3 种烟叶样品(整烟叶、卷烟丝、烟叶粉末)的便携式近红外光谱数据,测量 5 次,光谱数据取平均值。

1.2.3 模型的建立与转移

模型的建立采用偏最小二乘法(PLS)^[15];模型转移采用斜率截距算法(SBC)^[16-18]、分段直接标准化法(PDS)^[19-21]和典型相关分析法(CCA)^[9,22-25]进行。

1.2.4 模型转移评价指标

模型转移的效果采用预测均方根误差值(RMSEP)^[26]进行评价。

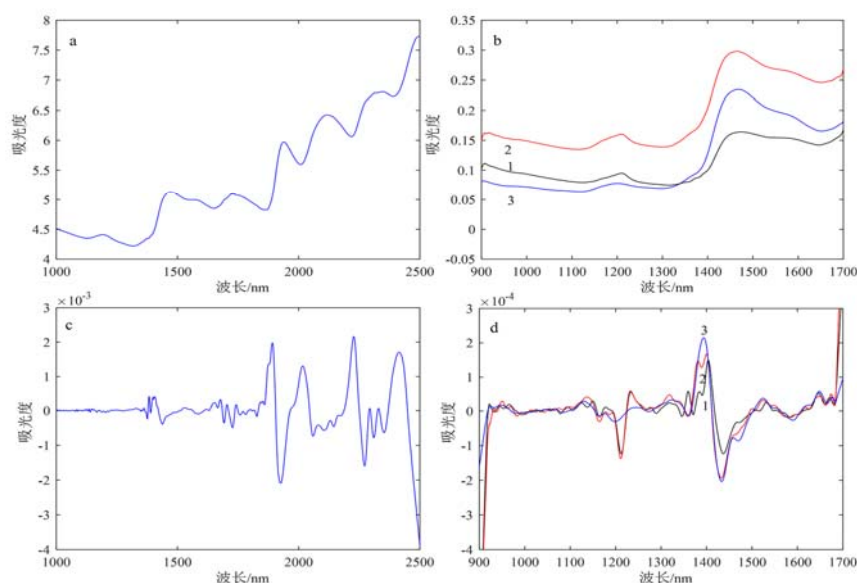
2 结果与分析

2.1 烟叶烟碱含量

测定结果表明,烟叶粉末样品的烟碱含量为 1.742 5%~3.052 3%,均值为 2.120 3%,标准偏差为 0.580 3。

2.2 烟叶光谱预处理

利用 Kennard-Stone 样本划分方法,将 60 个样本划作训练集,15 个样本划作验证集。烟碱近红外光谱信息经过平滑、微分求导^[27]的预处理后,结果如图 1。从图 1-a、图 1-b 中可以观察到,2 台不



a 主仪器 A 烟叶粉末原始光谱; b 从仪器 B 烟叶原始光谱, 1 烟叶粉末; 2 卷烟丝; 3 整烟叶; c 主仪器 A 烟叶粉末预处理后的光谱; d 从仪器 B 烟叶预处理后的光谱, 1 烟叶粉末, 2 卷烟丝, 3 整烟叶。

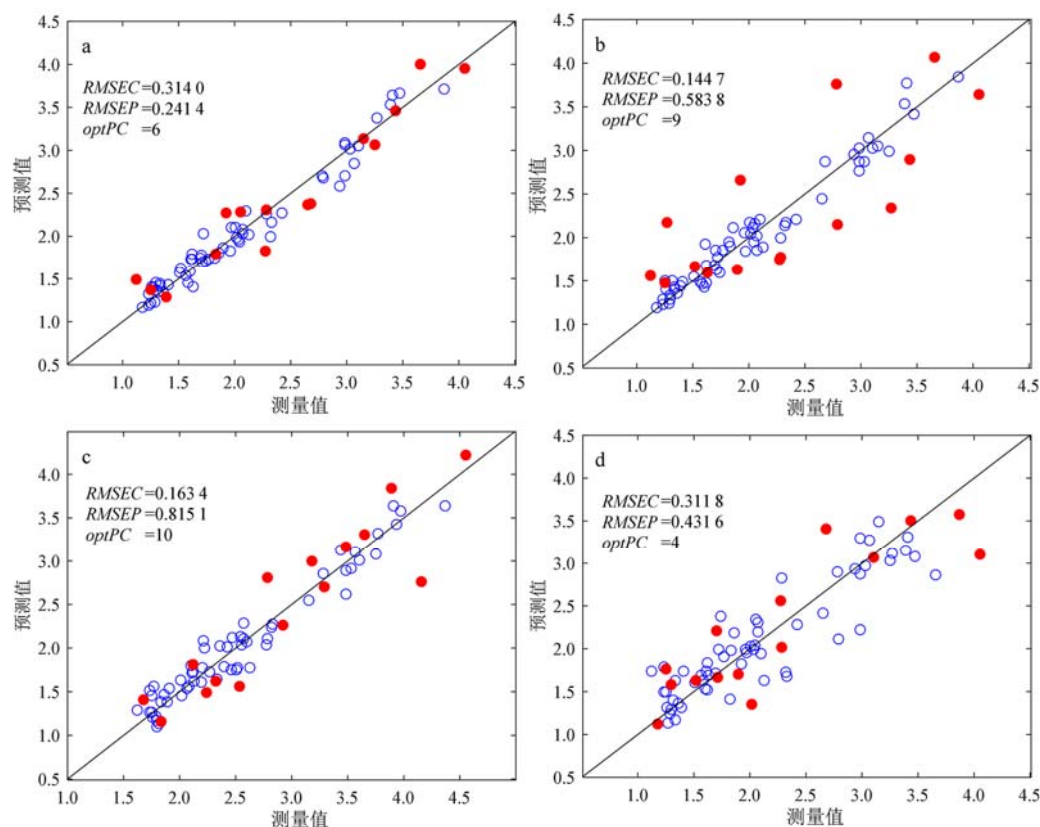
图 1 烟叶样品近红外光谱

Fig.1 The spectra information of tobacco

同型号的光谱仪器采集到的光谱范围不同,光谱之间存在明显差异;从图 1-c、图 1-d 中可以看出,适当的预处理虽然可以减少光谱基线漂移的影响,但不能帮助消除样品状态造成的光谱偏差。考虑到 2 种仪器模型传递研究需基于同一波段,因此截取 1 000~1 700 nm 波段予以研究。通过 3 次样条插值拟合这段光谱,使得便携式近红外光谱仪与傅立叶近红外光谱仪光谱波长点数相同。

2.3 烟叶样品烟碱的近红外光谱预测模型

不同物理形态的烟叶样品通过不同型号的近红外光谱仪采集烟碱光谱后,建立各自对应的模型预测误差如图 2 所示,可以看出,对于烟叶粉末从仪器建立的 pls 模型效果不如主仪器。可能是从仪器包含烟叶的信息不如主仪器完整,因而从仪器模型预测效果不佳。



a 主仪器 A 光谱模型; b、c、d 分别从仪器 B 烟叶粉末、卷烟丝和整烟叶光谱模型。

图 2 烟叶烟碱的近红外光谱模型的预测误差

Fig.2 The nicotine spectral model prediction error

2.4 烟叶样品的模型转移

表 1 为从仪器 B 烟叶粉末、卷烟丝、整烟叶经 SBC、PDS、CCA 算法模型校正的均方根。表中 N 为标准样品个数。从列表结果可以看出,标准样品个数对于 PDS 方法的影响小。对于 SBC 与 CCA 算法,随着标准样品个数的增加,它们的预测均方根误差(RMSEP)先增加后减小再增加,但 CCA 的整体预测均方根误差小于 SBC。从仪器 B 卷烟丝的 PDS 方法标准样品个数为 18,而对于 SBC 与 CCA 算法

分别在标准光谱为 7、9,但是 CCA 算法的整体预测均方根(0.222 2)优于其他转移算法和便携式近红外光谱重新建立模型的预测均方根误差(0.359 8)。从仪器 B 整烟叶的 PDS 方法标准样品个数为 12,而对于 SBC 与 CCA 算法分别在标准光谱为 7、2,但是 CCA 算法的预测均方根(0.333 9)优于其他转移算法和便携式近红外光谱重新建立模型的预测均方根误差(0.431 6)。

表 1 从仪器 B 的烟叶样品烟碱近红外光谱预测模型转移的预测均方根

Table 1 The result of tobacco powder, cigarette silk whole leaf calibration transfer of Slave B

标准光谱	烟叶粉末 <i>RMSEP</i>			卷烟丝 <i>RMSEP</i>			整烟叶 <i>RMSEP</i>		
	SBC	PDS	CCA	SBC	PDS	CCA	SBC	PDS	CCA
<i>N</i> =2	0.915 8	0.956 7	0.392 7	1.890 5	1.404 5	0.374 3	0.985 1	1.389 3	0.333 9
<i>N</i> =3	0.739 4	1.202 0	0.373 9	0.804 1	1.886 2	0.381 1	0.795 7	2.097 5	0.364 8
<i>N</i> =4	0.793 4	0.799 6	0.298 2	0.734 0	1.313 3	0.375 0	0.783 6	1.568 6	0.380 2
<i>N</i> =5	0.791 8	1.379 1	0.346 7	0.744 8	0.853 1	0.409 1	0.753 4	0.770 6	1.844 3
<i>N</i> =6	0.791 4	0.929 2	0.615 3	0.725 1	0.794 9	0.349 6	0.782 6	0.686 7	0.860 5
<i>N</i> =7	0.790 8	0.864 7	0.661 0	0.725 0	0.759 9	0.296 7	0.786 2	0.623 5	0.864 9
<i>N</i> =8	0.737 0	0.796 5	0.667 1	0.814 0	1.209 8	0.332 7	0.728 1	0.531 8	0.767 9
<i>N</i> =9	0.756 6	0.807 0	0.734 3	0.805 6	0.655 4	0.222 2	0.727 9	0.513 0	0.730 7
<i>N</i> =10	0.760 3	0.827 2	0.471 3	0.816 5	0.660 9	0.235 9	0.712 7	0.476 5	0.551 1
<i>N</i> =11	0.762 6	0.810 9	0.456 8	0.792 9	0.631 0	0.245 6	0.712 6	0.490 4	0.580 1
<i>N</i> =12	0.745 4	0.880 5	0.462 9	0.787 5	0.569 4	0.266 9	0.714 3	0.446 6	0.583 9
<i>N</i> =13	0.741 0	0.894 9	0.438 1	0.838 7	0.560 7	0.268 0	0.716 5	0.463 6	0.624 3
<i>N</i> =14	0.739 5	0.879 8	0.346 7	0.825 5	0.573 8	0.287 0	0.717 4	0.503 7	0.548 8
<i>N</i> =15	0.738 3	0.847 7	0.361 0	0.825 5	0.550 1	0.318 0	0.718 3	0.547 8	0.593 3
<i>N</i> =16	0.736 5	0.798 8	0.298 6	0.889 2	0.532 7	0.300 7	0.716 5	0.494 5	0.536 9
<i>N</i> =17	0.736 4	0.810 9	0.301 8	0.873 7	0.518 4	0.289 9	0.716 7	0.495 2	0.533 8
<i>N</i> =18	0.738 5	0.842 3	0.301 9	0.828 4	0.513 5	0.288 3	0.718 9	0.494 6	0.844 2
<i>N</i> =19	0.737 1	0.819 7	0.307 3	0.782 6	0.539 0	0.278 9	0.7211	0.531 9	0.831 8
<i>N</i> =20	0.737 4	0.822 0	0.347 9	0.789 4	0.536 6	0.232 9	0.721 6	0.532 9	0.849 2
<i>N</i> =21	0.741 0	0.805 0	0.344 8	0.786 5	0.526 7	0.224 9	0.723 4	0.525 7	1.117 3

从模型转移的结果来看,CCA 算法转移的结果优于 SBC、PDS 算法,因为 CCA 算法不需要将傅立叶近红外光谱仪和便携式近红外光谱仪 2 台仪器采集到的光谱通过 3 次样条插值拟合,使波长点数相同;由于 SBC 算法是对预测结果进行简单的一元校正^[7],PDS 算法计算量大,难以确定其各波段的主成分回归时的主成分数^[18],而 CCA 算法可以将不同光谱仪采集的光谱进行转移^[22-25],因此,CCA 算法是较适合将便携式近红外光谱仪采集的光谱转移到傅立叶近红外光谱仪建立的模型上的方法。

3 讨论

比较 SBC 算法、PDS 算法和 CCA 算法将便携式近红外光谱仪采集的整烟叶、卷烟丝和烟叶粉末的光谱转移到傅立叶近红外光谱仪烟粉光谱模型上,结果表明,对于烟叶样品的 3 种物理形态,CCA 算法模型转移效果都优于另 2 种,模型转移后预测均方根误差(*RMSEP*)低于从仪器 B 重新建模。便携式近红外光谱仪经傅立叶近红外光谱仪已经建立的模型进行转移,可快速得出预测值,节约重复建

模的时间。

虽然 CCA 算法的结果优于 SBC 和 PDS 算法,但对于整烟叶光谱在标准光谱多于 5 时的预测均方根误差值仍大于自身重新建模的值,可能是样本量少的缘故,因而需要对模型进一步优化,增加整烟叶烟碱的样本量;同时也需对烟叶中的总糖、还原糖和总氮含量的光谱模型进行转移,以实现烟叶化学成分的全面快速分析检测。

参考文献:

- [1] 严衍祿. 近红外光谱分析基础与应用[M]. 北京:中国轻工业出版社, 2005: 113-266.
- [2] LIP DU G, CAI W, et al. Rapid and nondestructive analysis of pharmaceutical products using near-infrared diffuse reflectance spectroscopy[J]. Journal of Pharmaceutical & Biomedical Analysis, 2012, 70(21): 288-294.
- [3] LIN J, LO S C, BROWN C W. Calibration transfer from a scanning near-IR spectrophotometer to a FT-near-IR spectrophotometer[J]. Analytica Chimica Acta, 1997, 349(3): 263-269.
- [4] 张学博, 冯艳春, 胡昌勤. 近红外多元校正模型传递的进展[J]. 药物分析杂志, 2009, 29(8): 1390-1399.
- [5] BROWN C W, LYNCH P F, OBREMSKI R J, et al.

- Matrix representations and criteria for selecting analytical wavelengths for multicomponent spectroscopic analysis[J]. *Analytical Chemistry*, 2002, 54(9): 1472–1479.
- [6] QIN Y, GONG H. NIR models for predicting total sugar in tobacco for samples with different physical states[J]. *Infrared Physics & Technology*, 2016, 77: 239–243.
- [7] 苏虹. 不同物理状态的烟叶样品建模和模型转移的研究与分析[D]. 青岛: 中国海洋大学, 2014: 37–53.
- [8] 杨凯, 刘鹏, 王维妙, 等. 原烟在线近红外光谱模型转移研究[J]. *中国烟草学报*, 2012, 18(6): 27–31.
- [9] 郑开逸, 范伟, 吴婷, 等. 应用典型相关分析法校正不同仪器和不同时间测量的近红外光谱[J]. *计算机与应用化学*, 2013, 30(3): 246–250.
- [10] YC/T160—2002 烟碱含量测定[S].
- [11] 李其禄, 王斐, 王金凤, 等. 不同近红外光谱仪用于酒石酸美托洛尔片厂家鉴别的对比研究[J]. *光谱学与光谱分析*, 2015, 35(10): 2924–2929.
- [12] 周竹, 李小昱, 高海龙, 等. 漫反射和透射光谱检测马铃薯黑心病的比较[J]. *农业工程学报*, 2012, 28(11): 237–242.
- [13] 文韬, 郑立章, 龚中良, 等. 基于近红外光谱技术的茶油脂脂肪酸含量的快速检测[J]. *湖南农业大学学报(自然科学版)*, 2016, 42(6): 676–681.
- [14] 谭观萍, 宾俊, 范伟, 等. 模型集群分析——随机森林方法在烟叶分类中的应用[J]. *江西农业学报*, 2017, 29(1): 69–74.
- [15] 吴金卓, 孔琳琳, 李颖, 等. 近红外光谱法测定土壤全氮和碱解氮含量[J]. *湖南农业大学学报(自然科学版)*, 2016, 42(1): 91–96.
- [16] BOUVERESSE E C, HARTMANN A, MASSART D L, et al. Standardization of near-infrared spectrometric instruments[J]. *Analytical Chemistry*, 1996, 68(6): 982–990.
- [17] 周林峰. 黄酒主要品质的近红外光谱检测模型建立与转移的研究[D]. 杭州: 中国计量学院, 2013: 36–81.
- [18] 刘贤, 董苏晓, 韩鲁佳, 等. 青贮饲料近红外光谱分析模型转移研究[J]. *农业机械学报*, 2009, 40(5): 153–157.
- [19] ALAMAR M C, MOLTÓ E, BOBELYN E, et al. Calibration transfer between NIR diode array and FT-NIR spectrophotometers[J]. *Proceedings of Frutic Information & Technology for Sustainable Fruit & Vegetable Production*, 2005(5): 12–16.
- [20] 黄承伟. 仪器间的光谱模型传递及谱图标准化[D]. 杭州: 浙江大学, 2012: 64–89.
- [21] 张学博. 通用性近红外模型的验证和维护[D]. 北京: 中国协和医科大学, 2009.
- [22] FAN W, LIANG Y, YUAN D, et al. Calibration model transfer for near-infrared spectra based on canonical correlation analysis[J]. *Analytica Chimica Acta*, 2008, 623(1): 22–29.
- [23] 范伟. 基于分子光谱的多元校正模型优化与转移方法及其应用研究[D]. 长沙: 中南大学, 2013: 72–82.
- [24] 罗峻, 聂凤明, 吴淑焕, 等. 纺织品快速无损检测校正模型的转移与共享研究[J]. *中国纤检*, 2016, 10(1): 79–81.
- [25] 宗鹏飞, 王志斌, 张记龙, 等. 基于红外被动测距的基线拟合算法研究[J]. *激光技术*, 2013, 37(2): 174–177.
- [26] 宾俊, 范伟, 周冀衡, 等. 近红外技术结合 SaE-ELM 用于烤烟烘烤关键参数的在线监测[J]. *烟草科技*, 2016, 49(9): 50–56.
- [27] 张民, 李银花, 袁晴春, 等. 基于高光谱技术的茶鲜叶茶多酚含量的估算模型[J]. *湖南农业大学学报(自然科学版)*, 2015, 41(4): 450–454.
- 责任编辑: 罗慧敏
英文编辑: 罗维