

文章编号:1001-5078(2011)09-0986-05

· 红外技术 ·

基于红外光谱的烟叶自动分级研究

刘剑君¹, 申金媛², 张乐明², 刘润杰²

(1. 河南省烟草公司郑州分公司, 河南 郑州 450001; 2. 郑州大学信息工程学院, 河南省激光与光电信息技术重点实验室, 河南 郑州 450001)

摘要: 烟叶的自动分级一直是国内外学者智能化研究的一个重要方向。通过分析烟叶的主要组成成分和烟叶光谱信息特征, 发现烟叶红外光谱可作为烟叶分级特征, 并通过神经网络模型验证了红外光谱作为其分级特征的可行性, 通过对比分析选取最佳红外光谱间隔、光谱范围以及最必要的光谱预处理方法。利用概率神经网络对9个等级的烟叶进行分组分级, 首先对选光谱数据进行减均值的预处理以消除基线漂移, 然后将其作为神经网络的输入模式, 相应的等级或组份作为理想输出训练网络。选择近半数的样本作为训练样本, 其余为测试样本; 网络对于训练样本的正确吻合率为100%, 测试样本的平均正确吻合率91%以上。结果表明烟叶的红外光谱可以作为烟叶的分级特征, 概率神经网络可以用于烟叶自动分级, 为烟叶的自动分级提供了新方法。

关键词: 烟叶分级; 红外光谱概率; 神经网络

中图分类号: TP183 **文献标识码:** A **DOI:** 10.3969/j.issn.1001-5078.2011.09.010

Study on automatic classification of tobacco based on infrared spectrum

LIU Jian-jun¹, SHEN Jin-yuan², ZHANG Le-ming², LIU Run-jie²

(1. Zhengzhou Filiale, Tobacco Company of Henan Province, Zhengzhou 450001, China; 2. School of Information Engineering, Zhengzhou University, Henan Key Laboratory of Laser and Opto-electric Information Technology, Zhengzhou 450001, China)

Abstract: The automatic classification of tobacco leaves has been an important intelligent research field. By studying on the components of tobacco leaves and analyzing the spectrum of tobacco leaves, the infrared spectra of tobacco leaves can be used as stamp of the tobacco grading. This is verified by the neural network model. The best intervals of the infrared spectra and the necessary preprocessing algorithm are studied and decided. 400 infrared spectra of tobacco leaves are adopted to group or grade leaves by neural network model. Almost half leaves are used as training samples and others as test samples. The mean rates of correct fit classification are 100% and 91% for training and test samples respectively. The grouping and grading results of neural network model based on infrared spectra show that the infrared spectra can be applied to grade tobacco leaves automatically. This indicates it is a new method for the automatic grading tobacco leaves.

Key words: tobacco leaves grading; infrared spectrum; probabilistic neural network

1 引言

因烟叶中所含碳、氢基团(如 C-O, C-H 基团)的红外光谱的合频与被频吸收^[1-4], 烟叶的红外光谱与烟草内在特征和化学成分都密切相关。烟叶等级不同, 其红外光谱信号不同。本文利用烟叶红外

光谱结合概率神经网络对烟叶进行分组分级研究。

基金项目: 河南省烟草专卖局科学计划与计划开发项目(No. 2008)资助。

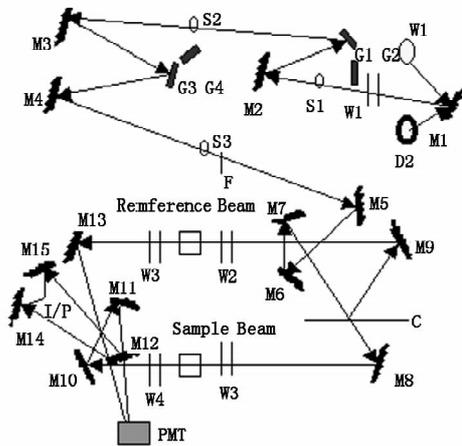
作者简介: 刘剑君(1972-), 男, 助农师, 研究方向为烟叶分级。

收稿日期: 2011-03-04; **修订日期:** 2011-04-19

目前烟叶分级基本上都停留于人工分级,我国烟叶分级方法也一样,依靠专家依据大量的经验,通过眼,鼻,手等从感观上对烟叶等级进行确认。可见这种传统的“师傅带徒弟”式的分级方式不仅每年要花费很多的人力物力,而且具有很强的主观随意性,不同的人对同一片烟叶很可能有不同的判断。很容易造成国家经济损失或与烟农产生纠纷,影响烟农积极性。所以对客观,公正的机器智能分级的需求就愈发迫切。现国内外有关烟叶智能分级的研究基本上都着重集中在图像处理领域。提取图像特征^[5-8],如长度,面积,周长,长宽比,残损比,平均色调、亮度和饱和度等,大都采用最近邻算法利用这些图像特征对烟叶进行分级。显然有些与烟叶等级密切相关的特征如厚度、叶片结构、含油分无法用图像特征准确描述,相反,这些因素可以由烟叶的红外光谱很好地被体现。

2 光谱数据获取及预处理

本实验的光谱数据由岛津 UV-3600 分光光度计获取。其基本原理光路如图 1 所示。



D2, W1 氙灯 钨灯; S1 入射狭缝; S2 中层狭缝; S3 出射狭缝; F 滤镜; G1, G2 衍射光栅(第一级分光); G3, G4 衍射光栅(第二级分光); C. H. 分光镜; W1 - W3 石英窗(30 mm); W4 - W5 石英窗(40 mm); M1 - M20 反光镜; I/P InGaAs/PBS; PBS pbs 单元; InGaAs InGaAs 光电二极管; PMT 光电倍增管; Reference \ Sample Beam 参比 \ 样本端

图 1 获取光谱数据原理图

其工作过程为:在分光光度计中,灯是不产生位移的,当所需某一个波段时,驱动灯反光镜的步进马达连接传感器,驱动反光镜到合适位置与该波段相对应的灯成合适角度,经过 M1 反射出所需要的光。从光源部分出来的光通过一个很窄的狭缝,这个狭缝通常被称为入射狭缝,这个入射狭缝固定光的宽度和高度。通过狭缝后,光进入了前单色仪,通过 M2 和 G1 或者 G2 的反射和分光, G1/G2 分离和识

别不同波长。经过前单色仪的光的谱带进入了中间狭缝 S2,通过 S2 进入到主单色仪,这个光谱带经过主单色仪的分光再次被分光,通过出射狭缝 S3 后进入滤镜 F,经过反光,通过一个斩光器将光分为相同的两束光,一个为样本光,一个为参比光。两束光经过 M10 ~ M15 被送入检测器。光栅和出射狭缝的特征,以及光栅和出射狭缝的位置决定了仪器的分辨率。这里所谓的分辨率是指一个仪器分辨不同波长的能力。要区分两个不同的波长的光,关键就是仪器的分辨率必须小于这两个波长之差的十分之一。即光源决定分光光度计的波长测量范围,衍射光栅决定其最小采样波长间隔,光电探测器决定其测量精度。将获得的原始光谱数据做预处理,为了消除基线漂移和机器噪声,对原始光谱数据进行了减均值的预处理。

3 烟叶特征谱选择

UV-3600 的测量范围覆盖了从紫外到红外光谱,跨度较大。其有效测量谱所包含的信息量非常巨大,这样就很有必要选取合适的吸收谱。本文选取烟叶的红外吸收谱作为其分组分级特征,理论基础如下:

(1) 烟叶等级划分的 7 大因素中的成熟度与烟叶内在的化学成分与结构有着密切的关系。众所周知,物质的谱信息能反映物质的特性和组分。红外光谱其实反映的是 C-O, C-H 基团的和频与倍频吸收;而烟叶作为一种植物,其内在的化学成分中含有 C-O, C-H 基团。研究表明,烟草的化学成分可分为以下几类: C-H 化合物、C-O-N 的化合物、酶类、有机酸、酚类化合物、色素。烟叶成熟度其实就是烟叶的这些化学成分及含量在叶表的反应,也是 7 大因素中唯一的综合性因素,是判定烟叶等级的首选因素。烟叶的成熟度分为:未熟(仍处于旺盛生长期)、假熟(生长发育已接近生理成熟)、初熟(已经成熟,但成熟度不够)、适熟(工艺成熟期)、完熟(完全成熟)、过熟(成熟过度)。研究表明:烟叶成熟度越好,还原糖、挥发酸、K+ 成分越高,挥发碱含量就越低^[9]。烟叶中总糖含量为 18% ~ 24%,总氮(N)含量为 1.5% ~ 3.5%,碱含量为 1.5% ~ 3.5%。对于成熟期的烟叶来说,其总糖含量为 12.89%,总 N 含量为 2.08%,总碱含量为 2.76%;对于初熟期的烟叶来说,其总糖含量为 12.43%,总 N 含量为 1.95%,总碱含量为 2.53%;对于未熟期的烟叶来说,其总糖含量为 16.02%,总 N 含量为 1.36%,总碱含量为 1.22%。从以上三种含量的数

据变化中可以看出:不同成熟度的烟叶,不仅其内在成分含量不同,而且叶组织结构不同,所表现出的等级自然也不相同。因其含量成分的不同,故红外光谱信号的强度会不同,峰值位置也会不同。比如烟叶从初熟到适熟的过程中上部(B)、中部(C)烟叶的厚度会变薄,含水率也开始下降。这样烟叶的化学成分的比例会开始变化,导致红外光谱信号变化。

(2)影响烟叶等级划分的油分因素与红外光谱密切相关。烟叶的油分是叶片组织细胞内含有的一种或半液体物质,在烟表特征表现为油润(叶片生长发育良好,柔软液体多,叶表油感足)、丰满(叶片生长发育尚好,柔软液体比较少,叶表尚有油感)、枯燥(叶片生长发育不好,柔软液体少,叶表无油感)的程度。不同产地的油份特征不一样,比如河南产区的烟叶含油量较重,云南产区的烟叶含油量较小。这里说的油分并不是指含油的多少,而是指在一定水含量情况下,眼的观感和手的触感。油分多的烟叶吸湿性较强。红外光谱法就是基于对这种半液体物质所含的官能团在红外区有特征吸收峰,不同波长处的吸收峰是由不同的官能团的不同振动引起的^[10]。

(3)跟分级有关的因素中的身份因素也与红外光谱密切相关。烟叶的身份具体表征为烟叶的厚度。烟叶的身份可划分为五个档次:中等(叶片生长发育好,化学成分含量丰实,厚薄适宜),稍厚(多为B级烟,叶片生长发育尚好,化学成分含量较多,比较重),稍薄(适中的等级档次,叶片生长发育尚好,化学成分含量丰实),厚(叶片生长发育不好,比较重),薄(叶片生长发育不好,化学成分含量较少)。烟叶身份的不同,其表征颜色也会不同,身份越厚,则颜色越深。C-N化合物的含量就越高,这样身份的五个档次与红外光谱的关系就显而易见了,烟叶越厚,则烟叶的红外吸收谱越大。烟叶厚度,一般随部位的上升而出现从薄到厚变化。不同档次的烟叶身份特征就对应了不同的红外吸收谱特征。比如同一级烟叶F级烟叶一般比L级烟叶身份厚,C3F身份就比C3L身份厚,C3F对应的红外吸收谱就比C3L对应的红外吸收谱峰值高。

(4)烟叶等级划分的7大因素中的色度其实就是烟叶内部叶主要化学成分以及挥发油、树脂含量的反映,其与红外光谱也密切相关。在烟叶分级中,色度划分为5个档次:浓、强、中、弱、淡。色度其实是叶表色彩的均匀度,饱和度,光泽度的综合体现。色度的浓淡度与油分多少呈正相关。所以,色度与

烟叶分级关系和油分基本相同。色度浓的烟叶总糖含量较高,总N、总碱等含量较低,比例也比较协调,吸湿性也较强,随着色度的从浓到淡减弱,烟叶品质等级也逐渐降低。色度不同,烟叶化学成分含量不同,则红外光谱吸收峰值不同。

除此之外,国外已有成熟的类似检测仪器面世。2010年,日本发明了基于红外光谱的茶叶分级检测装置。这使得本文更有理由选取红外光谱作为特征谱。

4 概率神经网络分组、分级结果

PNN是由Specht博士提出来的^[11-12],它由前馈型径向基函数神经网络发展而来的,是径向基神经网络的一种拓展应用^[13-15],常用于进行模式识别或分类。

现行国家烤烟标准将烟叶部位划分为下部(X)、中部(C)和上部(B);将颜色划分为柠檬黄(L)、桔黄(F)、红棕(R)三个基本色。

在分组PNN网络中,为更好的讨论模型的分组能力,采用的烟叶样本来源于2009年度河南省产地实时收购上来的烟叶,取B、C、L组的各36片做训练样本,其余的126片做测试样本,其中B组中有3个等级,C组中有3个等级,L组中有3个等级。假设3个组别烟叶的颜色均为已知。分组PNN网络的输入维数为671,对应于1660~3000nm波长范围的吸收光谱,光谱选取间隔为2nm;输出个数为3个:分别对应上部(B),中部(C),下部(X);本文做了光谱间隔的选取对比,选取间隔10nm时,则分组PNN网络的输入维数为135。表1,表2分别给出了间隔为2nm,10nm时的PNN网络分组吻合率。在分级识别中,采用的烟叶样本来源于2010年度河南生产地实时收购上来的烟叶,为B2F\B3F\B4F。取B2F,B3F,B4F三个等级共85片做训练样本,其余的111片做测试样本。由于采用相同的光谱信号,所以分级模型的输入与分组网络相同。为更好地说明问题,本文做分级时,假设烟叶的色和组已经确定,所以只在同组同色的烟叶中进行分级识别。模型跟分组模型一致。表3给出了分级吻合率。

表1 PNN分组的吻合率

样本吻合率	B(上部)等级	C(中部)等级	X(下部)等级
训练样本吻合率/%	100	100	100
测试样本吻合率/%	100	90.9	93.1
测试样本平均吻合率/%	93.3		

表2 PNN 分组的吻合率

样本吻合率	B(上部)等级	C(中部)等级	X(下部)等级
训练样本吻合率/%	100	100	100
测试样本吻合率/%	92.3	90	92.7
测试样本平均吻合率/%	92.1		

从表2中可以看出,光谱间隔为10 nm时,分组吻合率依然可以接受。故本文认为可以选取间隔为10 nm,这样就降低了网络的输入维数,节省了时间。

表3 PNN 分级的吻合率

样本吻合率	B2F 等级	B3F 等级	B4F 等级
训练样本吻合率/%	100	100	100
测试样本吻合率/%	93	92	91.4
测试样本平均吻合率/%	91.9		

结果表明:烟叶分组训练样本的吻合率均为1,烟叶分组测试样本的平均吻合率达到0.921,烟叶各个等级分级训练样本的吻合率均为1,分级测试样本的平均吻合率为0.919。

5 结论与展望

(1)利用烟叶红外光谱信息特征和相对简单的减均值预处理方法,结合PNN模型建立了烟叶自动分级模型,分级吻合率均在0.91以上,证明了红外光谱是烟叶自动分级的一个很好的特征。

(2)本文着手于烟叶的分组和分级研究,对分色研究没有涉及,分色于颜色有关,是否选用红外光谱是下一步研究的重点。

(3)本文对红外光谱的光谱间隔进行了合理选取,是否能选取更大的间隔也是下步工作中的重心。

(4)烟叶的红外光谱范围以及最有效的光谱预处理方法也是以后需要进一步研究的方面之一。

(5)最后,可以用来进行分级的方法很多,如最近邻和多种神经网络模型以及均值聚类等等,需要通过研究选择花时最少,吻合率高兼顾复杂性的模型。

参考文献:

[1] Xue Lei, Wang Yiqun, Fan Jianlian. Study on the spectroscopy of yellow serpentine jade[J]. Laser & Infrared, 2009, 29(3): 267-239. (in Chinese)
薛蕾,王以群,范建良.黄色蛇纹石玉的谱学特征研究[J].激光与红外,2009,29(3):267-239.

[2] Zhang Jun, Jiang Feihon. Spectral identification of infrared clouds using A practical multi-target discrimination system

[J]. Laser & Infrared, 2000, 30(4): 95-97. (in Chinese)

张骏,江飞虹.污染云团红外光谱多目标识别系统[J].激光与红外,2000,30(4):95-97.

[3] Xiong Wei, Fang Yonghua. Pollutant gas cloud spectrum real-time simulation based on gauss dispersion mode[J]. Laser & Infrared, 2007, 37(7): 637-640. (in Chinese)
熊伟,方勇华.基于高斯扩散模式的污染气体云团红外光谱实时仿真[J].激光与红外,2007,37(7):637-640.

[4] Fan Jianliang, Guo Shouguo, Liu Xueliang, et al. Study on spectroscopy of natural and treated jadeite [J]. Laser & Infrared, 2007, 37(8): 769-772. (in Chinese)

范建良,郭守国,刘学良,等.天然与处理翡翠的光谱学研究[J].激光与红外,2007,37(8):769-772.

[5] Zhang Jianping, Wu Shouyi, Fang Ruming. Computer aided quality inspecting and grading for agricultural products(1st half) quantitative inspection of external quality characters of tobacco leaf[J]. Transactions of the Chinese Society of Agricultural Engineering, 1996, (9): 158-161. (in Chinese)

张建平,吴守一,方如明.烟叶外观品质特征的定量检验[J].农业工程学报,1996,(9):158-161.

[6] Zhang Fan, Fang Ruming, Cai Jianrong. Study of getting tobacco leaf weight based on neural network technology [J]. Transactions of the Chinese Society of Agricultural Machinery, 2000, 11: 61-65. (in Chinese)

张帆,方如明,蔡健荣.基于人工神经网络技术提取烟叶质量特征值[J].农业机械学报,2000,11:61-65.

[7] Han Liqun, He Wei, Duan Zhengang, et al. Intelligent technologies in automatic grading of flue-cured tobacco [J]. Transactions of the Chinese Society of Agricultural Engineering, 2002, 11: 173-175. (in Chinese)

韩力群,何为,段振刚,等.烤烟烟叶自动分级的智能技术[J].农业工程学报,2002,11:173-175.

[8] Han Liqun, He Wei, Su Weijun, et al. Grading flue-cured tobacco leaf based on artificial brain intelligent system [J]. Transactions of the Chinese Society of Agricultural Engineering, 2008, (8): 137-140. (in Chinese)

张帆,方如明,蔡健荣.基于拟脑智能系统的烤烟烟叶分级研究[J].农业工程学报,2008,(8):137-140.

[9] Cai Xianjie, Wang Xinmin, Yin Qisheng. Study on the quantitative relationship between maturity and quality of tobacco leaf [J]. Acta Tabacaria Sinica, 2005, 11(4): 42-46. (in Chinese)

蔡宪杰,王信民,尹启生.成熟度与烟叶质量的量化关系研究[J].中国烟草学报,2005,11(4):42-46.

[10] Han Sufang, Tan Jichun, Zeng Xiaoying, et al. An infrared

- spectrum method to measure the concentration of mineral oil in water. [J]. *Environmental Monitoring in China*, 1998, 14(6):13-15. (in Chinese)
- 韩素芳, 谭吉春, 曾晓英, 等. 红外光谱法监测水中矿物油含量[J]. *中国环境监测*, 1998, 14(6):13-15.
- [11] Zheng Yali, Wang Kang. Classifying and filtering spam-mail based on probabilistic neural network[J]. *Computer and Modernization*, 2008, 1(1):8-10. (in Chinese)
- 郑亚莉, 王康. 基于概率神经网络的垃圾邮件分类[J]. *计算机与现代化*, 2008, 1(1):8-10.
- [12] Simon Haykin. *Neural networks*[M]. 2nd edition. Ye Shiwei, Shi Zhangzhi, et al. transl. Beijing: China Machine Press, 2006:1-22. (in Chinese)
- Simon Haykin. *神经网络原理*[M]. 叶世伟, 史忠植, 译. 第2版. 北京:机械工业出版社, 2006:1-22.
- [13] Wang Chengru, Wang Jinjia, Li Jin. MCE-based PNN training algorithm for speaker identification[J]. *Chinese Journal of Scientific Instrument*, 2002, 10:153-156. (in Chinese)
- 王成儒, 王金甲, 李静. 一种用于说话人辨认的概率神经网络的 MCE 训练算法[J]. *仪器仪表学报*, 2002, 10:153-156.
- [14] Zhao Wenbo, Huang Deshuang, Guo Lin. Genetic optimization of the structure of radial basis probabilistic neural networks[J]. *Journal of University of Science and Technology of China*, 2003, 12:733-741. (in Chinese)
- 赵温波, 黄德双, 郭磷. 径向基概率神经网络结构的遗传优化[J]. *中国科学技术大学学报*, 2003, 12:733-741.
- [15] Yang Liu, Ren Changming, Zhou Tong, et al. The research of stochastic pattern classifiers using Parzen window algorithm[J]. *Henan Science*, 2005, 2:97-99. (in Chinese)
- 杨柳, 任长明, 周铜, 等. 采用 Parzen 窗法的随机模式分类器研究[J]. *河南科技*, 2005, 2:97-99.