

可见光/近红外光谱分析的木材树种与密度同时预测

赵鹏, 李悦

东北林业大学信息与计算机工程学院, 黑龙江 哈尔滨 150040

摘要 光谱分析已经在木材特性参数(例如木材树种、气干密度、强度、含水率、表面粗糙度等)检测中得到应用,但是,现有的木材检测研究都只是针对上述某一项参数做数学建模和预测。如果需要检测木材多项参数,那么需要进行多次建模,并且每次建模预测时使用的数学模型类型(例如神经网络的类型)和内部结构参数一般各不相同。为了提高木材质量检测效率,提出了一种基于可见光/近红外光谱的木材树种和密度同时预测方法,它只需要一次建模和预测就可以实现这两项参数的同时输出。对东北5种常见木材(杨木、桦木、樟子松、白松和落叶松)进行检测,首先,采用K/S算法划分样本集,保证了训练集和预测集具有一定的代表性。然后,使用主成分分析和小波变换两种光谱降维方法,分别与BP神经网络和偏最小二乘支持向量机相结合建立了4种木材树种和密度同时预测模型和预测精度对比。采用美国海洋公司的Ocean Optics USB2000-VIS-NIR微型光纤光谱仪采集样本的可见光/近红外光谱并进行预测处理,光谱范围为350~1100 nm。结果表明,这四种模型都可实现对木材树种和密度的同时预测,其中小波变换降维方法结合偏最小二乘支持向量机所建立的模型预测效果相对较好,树种正确识别率为100%,训练集密度的 R 为0.9734,预测集密度的 R 为0.9408,训练集密度的RMSE为0.02613,预测集密度的RMSE为0.03846,它为同时对木材多项特性参数进行预测的便携式多功能一体化木材光谱检测仪器的开发奠定了理论基础。此外,还采用该公司生产的另一款光谱范围为900~1650 nm的FLAME-NIR型微型光纤光谱仪进行了同样的实验。对比发现,利用FLAME-NIR型光谱仪所得出的结果整体比利用USB2000-VIS-NIR型光谱仪所得到的结果好,但是相差并不是很大。这说明该方法可用于对木材种类与密度的同时预测,而且具有一定的稳定性和精度,也节约了仪器的成本。

关键词 光谱分析;木材树种;木材密度;同时预测

中图分类号: O439 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)11-3525-08

引言

实木木材是基本的木材工业产品,它是家具建材行业的主要生产原料,不同品质的木材具有不同的特性参数(例如木材树种、密度、强度、含水率、表面粗糙度等),导致其用途、物理性能及价格差别很大。因此,对木材特性参数进行检测是木材质量检测的重要内容,现在主流的检测方法是无损检测法,例如图像处理法、光谱分析法等。

可见光/近红外光谱分析技术具有成本低、效率高、速度快、无损、检测方便、测试重现性好等优点,已被广泛应用于各个领域的产品质量定性和定量研究^[1-4]。到目前为止,已有研究表明利用近红外光谱实现木材树种的识别及木材密

度等参数的预测是可行的^[5-7]。但是,这些光谱分析方法都是只针对木材树种或者密度等中的某一项参数进行建模和预测,不能够对这些参数进行同时建模预测。如果需要对木材多项参数进行检测,那么需要进行多次建模和预测。

从仪器设计和测量的角度看,多参数同时并行测量是当前仪器科学的主流方向。此外,在化学领域也有研究表明利用光谱和BP神经网络可实现物质多属性的同时预测^[8-9],例如,Zhang等通过对聚丙烯酸酯水性涂料性质的研究,根据三种单体(丙烯酸丁酯、甲基丙烯酸甲酯和苯乙烯)的用量和两种颜料(二氧化钛和碳酸钙)的用量同时预测四个目标值即硬度、附着力、耐冲击性和反射率,总体预测准确率达到98%,证明了BP神经网络对涂料的多种性能进行同时预测是可行且准确的^[9]。

收稿日期:2018-09-20,修订日期:2019-01-10

基金项目:国家自然科学基金面上项目(31670717),中央高校基本科研业务费专项基金项目(2572017EB09),黑龙江省自然科学基金面上项目(C2016011)资助

作者简介:赵鹏,1972年生,东北林业大学信息与计算机工程学院教授 e-mail: bit_zhao@aliyun.com

但是,到目前为止还未有研究利用近红外光谱实现木材树种、气干密度、强度等多项参数的同时预测。本工作采用两种光谱降维方法和两种数学建模进行预测和效果对比,只需一次建模和预测就可实现对木材树种和气干密度的同时预测,在一定程度上提高了建模预测的效率。

BP神经网络是一种按误差逆向传播算法训练的多层前馈网络,是目前应用最为广泛的神经网络模型之一。BP神经网络由输入层、隐含层和输出层三部分组成,因其在建立模型时具有很好的学习、预测及非线性映射能力和柔性的网络结构,已广泛用于定性和定量分析^[6, 10]。最小二乘支持向量机是由 Suykens 等在标准的支持向量机基础上提出的,可用于解决模式识别和回归等问题^[11-12]。LSSVM 利用结构风险最小原则,有效提高了求解速度和抗噪声能力,在解决近红外光谱的非线性问题上表现出色^[13]。和传统的 SVM 相比,LSSVM 以求解线性方程组的简便方式替代了求解二次规划问题的复杂方法,实现了约束条件由不等式约束到等式约束的转变。

以东北常见五种树种杨木、桦木、樟子松、白松和落叶松作为研究对象,将主成分分析(principal component analysis, PCA)和小波变换(wavelet transform, WT)两种降维方法分别与反向传播人工神经网络(BP)和最小二乘支持向量机(least squares support vector machines, LSSVM)两种模型相结合,建立了对木材树种和气干密度的同时预测模型并对模型的预测结果进行了对比分析。

1 实验部分

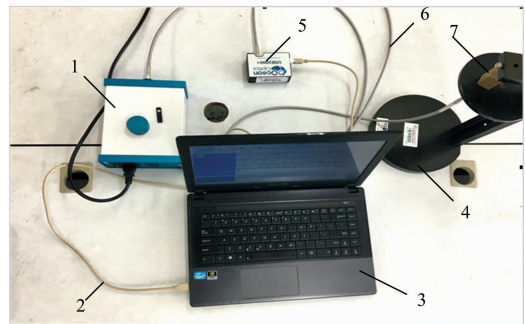
1.1 样本制备与密度测定

实验用的是东北常见的杨木(*Populus davidiana*)、桦木(*Betula platyphylla*)、樟子松(*Pinus sylvestris*)、白松(*Picea jezoensis*)、落叶松(*Larix gmelinii*)5种木材,将其加工成 2 cm×2 cm×2 cm 的木块,每个树种 80 个木块,共 400 个木块作为实验样本。这 5 种树种木材购置于大庆市木材市场,产地是郎乡林业局下属林场。5 种木材气干密度的测量按照《木材物理力学性质试验方法》(GB1933—1991)进行。

1.2 近红外光谱的采集

实验装置由计算机、光谱仪、卤光灯、光纤、支架等组成,光谱采集用美国海洋公司的 Ocean Optics USB2000-VIS-NIR 微型光纤光谱仪,待测样本放在光谱仪的支架上,待测面朝下,光纤探头距待测物体表面约 5 mm,圆形视场角的直径为 6.35 mm,实验装备示意图如图 1 所示。近红外光谱的采集采用漫反射方式,谱区范围为 350~1 100 nm,分辨率约为 1.5 nm。采集条件:室温(22±2)℃,湿度 25%±3%。采集光谱前,需进行光谱白板校正和暗校正,白板校正时使用标准聚四氟乙烯白板为背景,暗校正则是以遮光方式进行,同时需打开电子暗噪声校正和杂散光校正按钮。在光谱仪器配套的软件 SpectraSuite 中参数设置如下:积分时间为 1 ms,重复扫描 900 次取平均,平滑度为 5。计算机显示平均光谱,并将其保存,每测完 5 个样本进行一次标准白板的校正,光谱数据以 excel 格式导出进行处理。有研究表明,

用近红外光谱分析测量木材密度时,从木材横切面、径切面、弦切面取得的结果中以从横切面采集到的光谱建立的预测模型效果最好^[14],所以本实验采取的是在每个样本的横切面上随机选取 4 个测试点进行光谱采集,取其平均值作为一个样本的结果,并将其保存到指定位置。



1.卤光灯; 2.USB连接线; 3.计算机; 4.支架;
5.微型光谱仪; 6.光纤; 7.待测样本;

图 1 实验设备连接示意图

1: 卤光灯; 2: USB 连接线; 3: 计算机; 4: 支架;
5: 微型光谱仪; 6: 光纤; 7: 待测样本

Fig. 1 Diagram of experimental equipment

1: Halogen lamp; 2: USB bus; 3: Computer; 4: Holder;
5: Fiber spectrometer; 6: Optics fiber; 7: Test sample

1.3 光谱数据降维方法

由于近红外光谱包含的信息量较大、光谱重叠严重且处理起来较为困难,若直接使用原始光谱数据进行建模往往会导致模型稳定性差、精度低,同时也会增加计算的复杂度。分别选取了主成分分析和小波变换两种方法对光谱数据进行降维处理。

主成分分析是多元统计中的一种变换方法,是目前最常用的一种光谱特征信息提取方法,在损失很少信息的前提下,主成分分析能将多个指标转化为少数综合指标,代替原来较多的变量。小波变换是一种高效的数据压缩技术,将信号分解成一系列小波函数的叠加,具有时频分析能力。利用小波变换方法对预处理后的光谱进行压缩,既可起到压缩数据的作用还可以实现去噪和提取特征信息的作用。光谱经小波分解后分为低频和高频两个部分,低频部分对应近似系数,高频部分对应细节系数,低频系数含有原光谱信号绝大部分有效信息,而以噪声为代表的冗余信息多存在于高频系数中。将小波变换后的低频小波系数代替原来的光谱数据,大大降低了输入数据量,提高了建模效率。

1.4 模型建立

由于是对树种和密度的同时预测,因此选取的模型须具有多输出的特点。选取了 BP 神经网络和最小二乘支持向量机 LSSVM 作为对树种和密度进行同时预测的模型。以 BP 神经网络为例,给出模型的应用体系结构如图 2 所示。模型的输入为经过降维处理后的光谱数据,模型的前 5 个输出为树种预测值,第 6 个输出为密度预测值。五个树种分别将其编号为如表 1 所示的单位矩阵编码,预测结束后,将每位的预测值都转换为二进制编码。当某位的预测值<0.5 且与 0 的偏差(这里指绝对值,下同)<0.5 时,则将该位的预测

值视为 0；当某位的预测值 ≥ 0.5 且与 1 的偏差 < 0.5 时，将该位的预测值视为 1。当某位的偏差 ≥ 0.5 时，说明判别不稳定，则视为无法识别/拒识（另外，预测编码含有多个 1 例如 11000，也算作拒识）。只有当 5 位编码的预测值和真实值完全相同时，视为识别正确（例如预测树种编码和真实树种编码都是 10000，二者一致），否则视为误识（例如，预测编码是 10000，真实编码 01000，二者不一致）。模型第 6 位的输出值为密度预测值，保留其原始值作后续分析处理。此外，LSS-VM 模型输出比较简单，它的前 5 位是 0 或 1，不需要做判断和转换。所有的数据处理及模型的建立都是在软件 Matlab 2015a 中完成。

表 1 树种对应编号
Table 1 Coding of tree species

树种	编码
杨木	[1 0 0 0 0]
桦木	[0 1 0 0 0]
樟子松	[0 0 1 0 0]
白松	[0 0 0 1 0]
落叶松	[0 0 0 0 1]

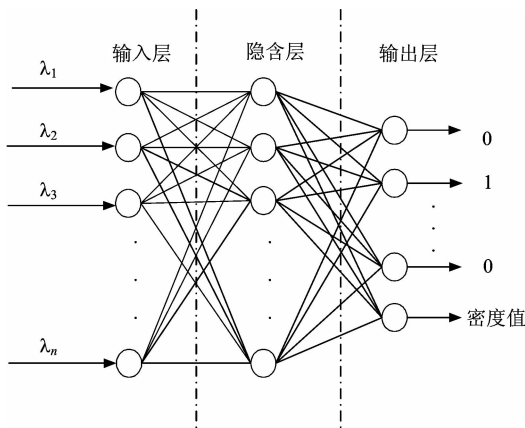


图 2 BP 神经网络应用体系结构

Fig. 2 Architecture of the BP neural network applied

1.5 模型评价

通过树种正确识别率(correct recognition rate, CRR)、密度的相关系数(r)和均方根误差(root mean square error, RMSE)来评价模型拟合精度和模型的预测能力。CRR 越高, r 值越接近 1, RMSE 的值越小且训练集的 RMSE 和预测集的 RMSE 较为接近, 所建模型性能越好。

2 结果与讨论

2.1 近红外光谱分析及预处理

由于系统误差使光谱曲线在首尾两端产生的噪声较大, 信噪比低, 不具有分析价值, 直接影响了实验的准确性, 所以将首尾的波段去掉, 只研究 450~900 nm 范围内的反射光

谱, 以此提高实验的精度。由图 3、图 4 可知, 五种木材的光谱曲线重叠严重, 不同树种木材的光谱曲线之间有较小的差异, 同一树种木材不同密度样本的光谱曲线较为相似, 因此要通过光谱数据完成对木材种类的识别以及密度的预测, 需要利用一定的预处理方式及数学模型。为消除样本不均匀、基线漂移、高频随机噪声、光散射等影响, 对原始光谱采用 Savitzky-Golay 平滑法, 平滑点数为 10, 再进行标准化(standard normal variate, SNV)处理。

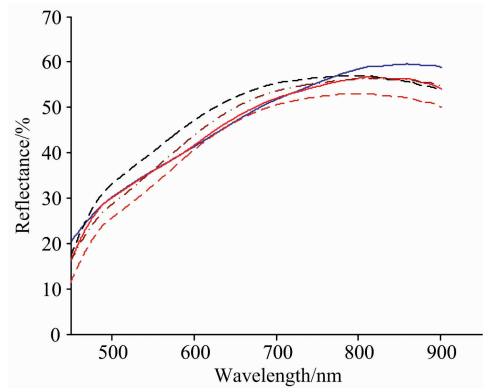


图 3 随机选取的五个不同树种样本的原始光谱

Fig. 3 Original spectra of randomly selected different tree species

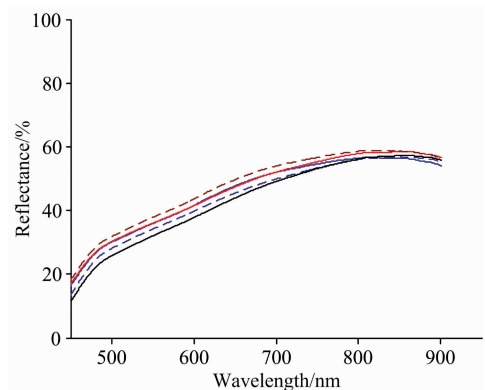


图 4 随机选取的同一树种不同密度的 5 个样本的原始光谱

Fig. 4 Original spectra of 5 specimens with different densities randomly selected from the same tree species

2.2 样本集划分

Kennard-Stone(K/S)算法是通过计算样本间的欧氏距离来划分样本集的, 将光谱差异大的样本选入训练集, 其余的样本归入预测集, 通过此方法可提高预测模型的精度和泛化能力^[14]。400 个样本中采用 K/S 法从每个树种中挑出 60 个样本作为训练集, 20 个样本作为预测集, 即训练集中共有 300 个样本, 预测集中共有 100 个样本(训练集和预测集没有交集)。图 5 为采用 K/S 法挑出的训练集和预测集的第一主成分和第二主成分分布图, 可以看出预测集样本包括在训练集样本的分布空间中, 说明此预测集具有代表性。全部样本的气干密度分布参见表 2。

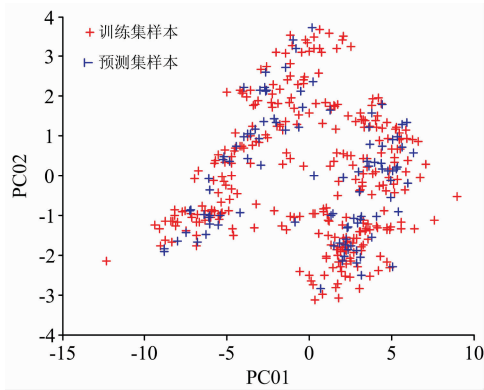


图 5 样本的第一主成分和第二主成分分布图

Fig. 5 Distribution of the first principal component and the second principal component of the sample

表 2 样本的气干密度汇总

Table 2 Statistical information on the gas dry density of the sample

样本集类	样本数	最大值/ ($\text{g} \cdot \text{cm}^{-3}$)	最小值/ ($\text{g} \cdot \text{cm}^{-3}$)	平均值/ ($\text{g} \cdot \text{cm}^{-3}$)
全部样本	400	0.750 9	0.314 0	0.468 3
训练集	300	0.750 9	0.314 0	0.469 4
预测集	100	0.744 9	0.314 8	0.464 8
杨木	80	0.472 9	0.323 2	0.371 4
桦木	80	0.750 9	0.535 8	0.618 2
樟子松	80	0.567 8	0.387 4	0.466 2
白松	80	0.398 2	0.314 0	0.342 9
落叶松	80	0.642 6	0.409 2	0.542 8

2.3 基于主成分分析降维的建模结果分析

由表 3 可得, 前 3 个主成分的累计贡献率达到了 99%, 结合图 6 可知提取的前 3 个主成分有一定的分类能力, 定性描述了五种树种木材的特征差异。从图 6 可以看出边缘样本区分不明显, 落叶松、白松和樟子松得分相对比较开, 但仍有少许样本和其他树种混在一起, 而杨木和桦木两者比较难区分。由表 3 可以看出, 经主成分分析光谱数据后, 训练集及预测集的前 7 个主成分的累计贡献率均已达到了 99.9%, 因此将前 7 个主成分代替每个样本的光谱数据作为

表 3 训练集和预测集前 7 个主成分贡献率及累计贡献率

Table 3 The contributions and cumulative contributions of first 7 principal components of training and forecasting sets

主成分	训练集 贡献率/%	训练集累计 贡献率/%	预测集 贡献率/%	预测集累计 贡献率/%
PC01	83.938 1	83.938 1	86.391 4	86.391 4
PC02	13.053 9	96.992 0	10.611 2	97.002 6
PC03	2.304 8	99.296 9	2.421 9	99.424 5
PC04	0.282 7	99.579 6	0.262 7	99.687 3
PC05	0.195 7	99.775 3	0.150 4	99.837 6
PC06	0.084 8	99.860 1	0.075 3	99.913 0
PC07	0.071 2	99.931 2	0.040 8	99.953 7

模型的输入。

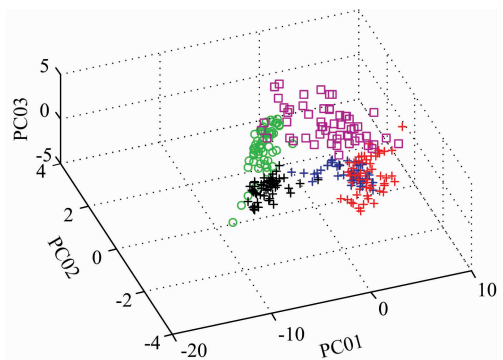


图 6 5 个树种的前 3 个主成分散点图

Fig. 6 Scatter plot of first three principal components of five tree species

2.3.1 PCA-BP 模型预测结果

BP 神经网络的输入层节点数为 7, 经过反复试验数次, 确定隐含层节点数为 14, 因此最佳的网络结构为 7(输入层)—14(隐含层)—6(输出层)三层神经网络模型。输入层和隐含层之间的传递函数选取为 tansig, 隐含层和输出层之间的传递函数为 purelin, 迭代次数设定为 1 000 次, 误差目标为 0.001, 学习率为 0.01, 训练完成后对预测集进行预测。图 7 为密度预测的散点分布图, 图 8 为对树种进行预测的结果图, 由图中可以看出, 有一个白松样本被视为无法识别。

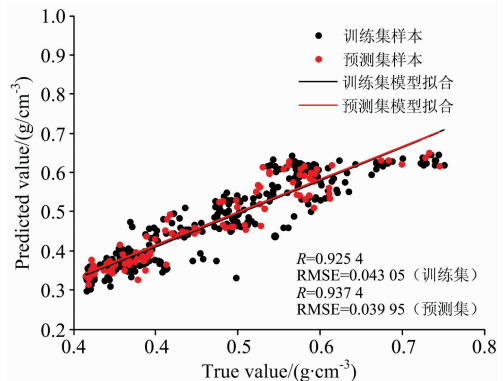


图 7 PCA-BP 密度预测值与真实值的相关性图

Fig. 7 The diagram of the correlation between the predicted value and the true value of the density of the PCA-BP model

2.3.2 PCA-LSSVM 模型预测结果

LSSVM 常用的 4 种核函数(线性核、多项式核、径向基函数核和高斯核)中径向基核函数(RBF)应用最为广泛, 它只有惩罚因子(c)和 RBF 核函数的宽度(g)两个参数, 因此选取 RBF 核函数作为 LSSVM 模型的核函数。传统的 LSSVM 模型中的 c 和 g 两个参数是任意给定或者凭经验选取的, 存在主观性和随意性。为了提高 LSSVM 学习和泛化能力, 需要对 c 和 g 两个参数进行优化, 目前常用的有 k -折交叉验证法、网格搜索法、粒子群优化算法以及遗传算法等, 本文采取了网格搜索法与 5-折交叉验证相结合的方法。根据

交叉验证树种平均正确识别率和密度平均相关系数的和 (CR_{Rev}+R_{cv})最大的原则选取最优参数,优化 LSSVM 模型。可知当 c 和 g 分别为 22.627 4 和 1 时,建立的模型预测性能最好,利用最优的 PCA-LSSVM 模型对未参与建模的预测集样本进行预测分析。PCA-LSSVM 模型的密度预测结果如图 9 所示,图 10 为对树种进行预测的结果图,由图中可以看出,有 1 个落叶松样本被视为无法识别。

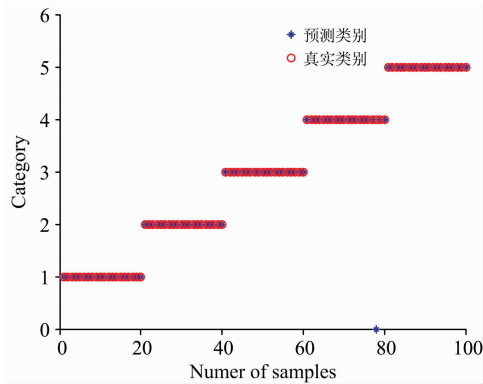


图 8 PCA-BP 树种预测结果

Fig. 8 The wood species prediction result of the PCA-BP model

且低频系数重构光谱图保留了原光谱 99.99% 的能量,同时与原光谱的相关系数达到了 0.999 9,说明了小波变换没有改变光谱信息。对光谱进行小波变换后,将 1 328 维的原始光谱减少到了 29 维,用低频小波系数代替原来的光谱变量作为模型的输入。

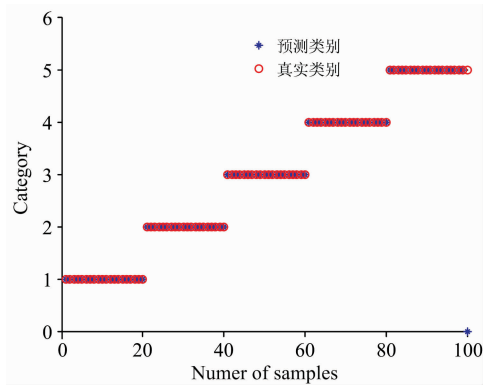


图 10 PCA-LSSVM 树种预测结果

Fig. 10 The wood species prediction result of the PCA-LSSVM model

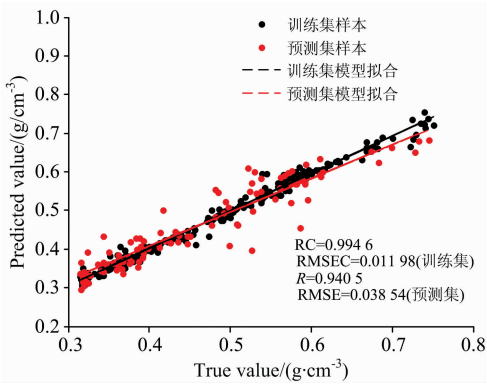


图 9 PCA-LSSVM 密度预测值与真实值的相关性图

Fig. 9 The diagram of the correlation between the predicted value and the true value of the density of the PCA-LSSVM model

2.4 基于小波变换降维的建模结果分析

在进行小波变换时,模型的预测能力受小波基函数和分解尺度的影响。选择常用的 Symlet 小波函数中的 sym5 为小波基。Daubechies 所提出的 Symlet 小波函数是对 db 函数的一种改进,具有较好的正则性。Symlet 小波函数在连续性、滤波器长度、支集长度等方面和 dbN 小波是相一致的,但 symN 小波函数比 dbN 小波函数具备更好的对称性,在对信号进行分析和重构时能在一定程度上减少相位失真。分解时每层采用不同的阈值并使用软阈值策略,对预处理后的光谱信号进行 6 尺度的分解,然后用分解得到的低频系数进行重构光谱。如图 11 所示,分解层数为 0 处的光谱为预处理后的原始光谱,当分解层数为 6,光谱的形状没有发生太大变化,

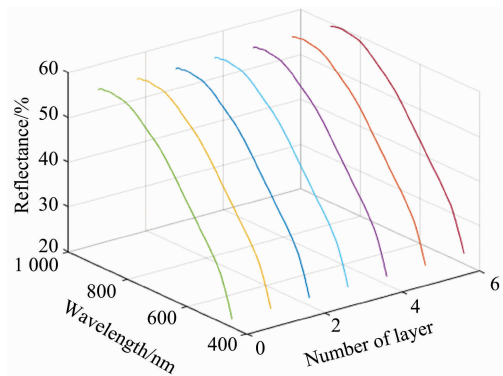


图 11 预处理后光谱与低频小波系数重构光谱对比图

Fig. 11 Comparison between the pre-processed spectrum and the reconstructed spectrum

2.4.1 WT-BP 神经网络模型预测结果

将 6 层分解后的 29 维低频小波系数作为 BP 神经网络的输入,经过反复实验确定隐含层节点数为 26,因此最佳的网络结构为 29(输入层)—26(隐含层)—6(输出层)三层神经网络模型。输入层和隐含层之间的传递函数选取为 tansig,隐含层和输出层之间的传递函数为 purelin,迭代次数设定为 1 000 次,误差目标为 0.000 1,学习率为 0.01,训练完成后对预测集进行预测。图 12 为密度预测的散点分布图,图 13 为对树种进行预测的结果图,由图中可以看出,有 1 个樟子松样本和一个白松样本被视为无法识别。

2.4.2 WT-LSSVM 模型预测结果

采用网格搜索法与 5-折交叉验证相结合的方法对 WT-LSSVM 模型中的参数 c 和 g 进行优化,可知当 c 和 g 分别为 724.077 3 和 16 时,建立的模型预测性能最好,利用最优的 WT-LSSVM 模型对未参与建模的预测集样本进行预测分

析。WT-LSSVM 模型的密度预测结果如图 14 所示, 图 15 为对树种进行预测的结果图, 由图中可以看出, 没有错分与无法识别的样本。

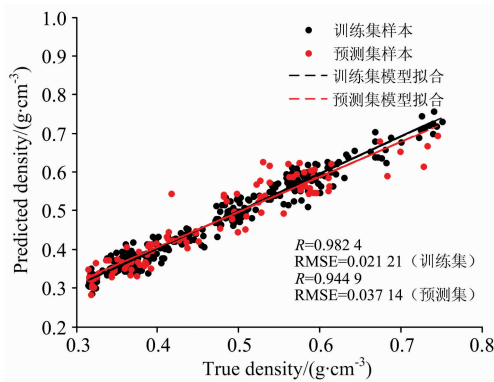


图 12 WT-BP 密度预测值与真实值的相关性图

Fig. 12 The diagram of the correlation between the predicted value and the true value of the density of the WT-BP model

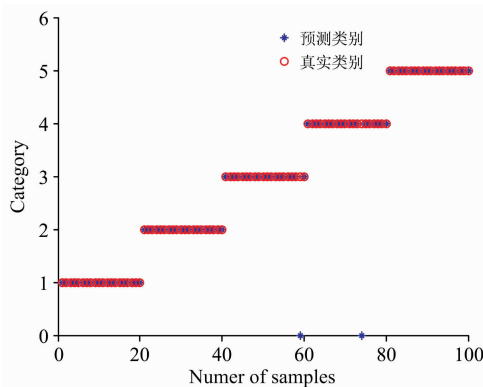


图 13 WT-BP 树种预测结果

Fig. 13 The wood species prediction result of the WT-BP model

2.5 各模型预测结果对比分析

对比以上建立各模型分别对预测集样本进行预测的结果, 如表 4 所示, PCA-BP 模型、PCA-LSSVM 模型、WT-BP 模型和 WT-LSSVM 模型都可实现对木材树种和密度的同时

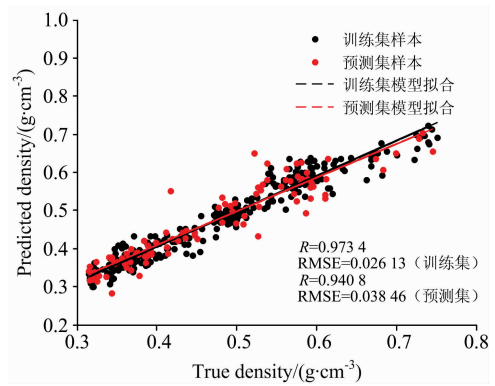


图 14 WT-LSSVM 密度预测值与真实值的相关性图

Fig. 14 The correlation diagram between the predicted and the true density values of the WT-LSSVM model

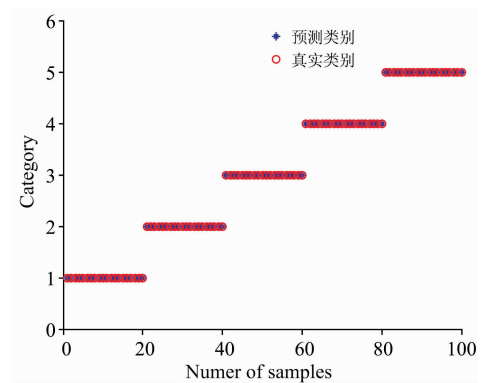


图 15 WT-LSSVM 树种预测结果

Fig. 15 The wood species prediction result of the WT-LSSVM model

预测, 且效果都相差不是很大, 其中 WT-BP 模型对密度的预测效果为最好, 预测集密度的 R 达到了 0.9449, 但对树种的预测不是最佳, 只有 98%。整体上看, 基于小波变换降维的模型略优于基于主成分分析降维的模型。这是因为木材近红外光谱信息和木材属性之间存在着非线性关系, 而主成分分析是一种线性降维方法, 因此该主成分分析方法的降维效果稍差。经过综合比较, WT-LSSVM 模型的预测效果相对较好, 预测集密度的 R 为 0.9408, 只比 WT-BP 模型低了

表 4 各模型预测结果对比

Table 4 Comparison of prediction results

模型	350~1100 nm/USB2000-VIS-NIR spectrometer						900~1650 nm/FLAME-NIR spectrometer					
	密度				树种		密度				树种	
	训练集		预测集		训练集	预测集	训练集		预测集		训练集	预测集
	R	RMSE	R	RMSE			R	RMSE	R	RMSE		
PCA-BP	0.9254	0.04305	0.9374	0.03995	98.67	99	0.9332	0.04103	0.9442	0.03787	98.67	100
PCA-LSSVM	0.9946	0.01198	0.9405	0.03854	98.33	99	0.9867	0.02178	0.9453	0.03321	98.33	99
WT-BP	0.9824	0.02121	0.9449	0.03714	100	98	0.9811	0.02111	0.9531	0.03534	100	99
WT-LSSVM	0.9734	0.02613	0.9408	0.03846	99.33	100	0.9828	0.02482	0.9564	0.03112	98.67	100

0.41%，但是它的 CRR 是四种模型中最高的，为 100%，因此 WT-LSSVM 模型对树种和密度的预测能同时达到较好的效果。

为了验证利用可见光和近红外光谱实现木材种类与密度同时预测的稳定性，我们采用美国海洋光学公司生产的另一款波长为 900~1 650 nm，光谱波长分辨率为 2.0 nm 的 FLAME-NIR 型微型光纤光谱仪进行了同样步骤的实验，所用仪器如图 16(b)所示，实验结果如表 4 所示。经过对比发现，利用 FLAME-NIR 型光谱仪所得出的结果整体比利用 USB2000-VIS-NIR 型光谱仪所得到的结果好，但是相差并不是很大，说明本研究所使用的可见光近红外光谱可用于对木材种类与密度的同时预测，且能保证一定的稳定性、满足一定的精度要求。由于图 16(a)中的 USB2000-VIS-NIR 型光

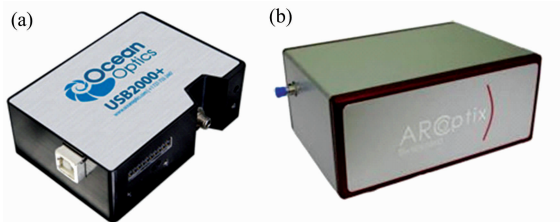


图 16 美国海洋公司的两种光谱仪

(a): USB2000-VIS-NIR 型微型光纤光谱仪;

(b): FLAME-NIR 型微型光纤光谱仪

Fig. 16 Two spectrometers produced by USA Ocean Optics Co.

(a): USB2000-VIS-NIR miniature fiber optic spectrometer;

(b): FLAME-NIR miniature fiber optic spectrometer

谱仪的价格比图 16(b)中的 FLAME-NIR 型光谱仪低很多，所以与此同时节约了开发的成本。

3 结 论

实现了基于光谱分析的木材树种及密度这两项参数的同时预测，较为准确地预测出了未知木材的种类及气干密度。建模过程中，通过采用 K/S 算法对样本集进行划分，保证了训练集和预测集具有一定的代表性。通过主成分分析和小波分析两种光谱降维方法与人工神经 BP 网络和最小二乘支持向量机分别结合建立了四种模型。研究表明，四种模型都可实现对木材树种和密度的同时预测，其中小波变换降维和最小二乘支持向量机相结合的模型预测效果相对较好，CRR 达到了 100%，训练集密度的 R 为 0.973 4，预测集密度的 R 为 0.940 8，训练集密度的 RMSE 为 0.026 13，预测集密度的 RMSE 为 0.038 46。此外，将利用 350~1 100 nm 的光谱所得出的结果和利用 900~1 650 nm 的光谱所得出的结果进行对比，可知前者略差于后者，但是两者之间的差异很小，说明利用可见光近红外光谱对木材种类与密度进行同时预测，能保证一定的稳定性，且能满足一定的精度要求，同时也节约了开发的成本。

后期研究中，我们将使用美国海洋公司的 USB2000-VIS-NIR 微型光纤光谱仪开发便携式的木材多项参数光谱检测仪器，该款光谱仪体积小重量轻，整套光谱采集设备(含笔记本电脑)是 RMB 3 万元左右。本研究为同时对木材多项参数(例如树种、密度、强度、含水率等)进行预测的便携式在线检测光谱仪器的研发奠定了理论基础。

References

- [1] Mendoza F A, Cichy K, Lu R, et al. Food and Bioprocess Technology, 2014, 7(9): 2666.
- [2] Li J, Huang W, Zhao C, et al. Journal of Food Engineering, 2013, 116(2): 324.
- [3] Li X L, Yi S L, He S L, et al. Precision Agriculture, 2016, 17(3): 365.
- [4] Wang J X, Fan L F, Wang H H, et al. Biosystems Engineering, 2017, 162: 40.
- [5] Hauksson J B, Bergqvist G, Bergsten U, et al. Wood Science and Technology, 2001, 35(6): 475.
- [6] Nisgoski S, Oliveira A A, Muñiz G I B. Wood Science and Technology, 2017, 51(4): 929.
- [7] ZHAO Rong-jun, XING Xin-ting, LÜ Jian-xiong, et al(赵荣军, 邢新婷, 吕建雄, 等). Scientia Silvae Sinicae(林业科学), 2012, 48(6): 106.
- [8] CHEN Zhen-ning(陈振宁). Chinese Journal of Analytical Chemistry(分析化学), 2001, 25(11): 1322.
- [9] Tao Z H, Yuan Z, Ping C. Journal of Mathematical Chemistry, 2009, 46(4): 1050.
- [10] ZHANG Yao, LI Min-zan, ZHENG Li-hua, et al(张 瑶, 李民赞, 郑立华, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2015, 31(9): 121.
- [11] SUN Xu-dong, DONG Xiao-ling(孙旭东, 董小玲). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2013, 29(14): 262.
- [12] WANG Jing, JIANG Gang, CHEN Zhong-jie(王 静, 蒋 刚, 陈中杰). Application Research of Computers(计算机应用研究), 2013, 30(12): 3597.
- [13] JIN Ye, YANG Kai, WU Yong-jiang, et al(金 叶, 杨 凯, 吴永江, 等). Chinese Journal of Analytical Chemistry(分析化学), 2012, 40(6): 925.
- [14] GUO Wen-chuan, LIU Da-yang(郭文川, 刘大洋). Transactions of the Chinese Society of Agricultural Machinery(农业机械学报), 2014, 45(9): 230.

Simultaneous Prediction of Wood Density and Wood Species Based on Visible/Near Infrared Spectroscopy

ZHAO Peng, LI Yue

College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China

Abstract Spectral analysis has been widely used in wood physical feature parameter detection such as wood species, density, strength, surface roughness and humidity. However, the current wood detection is used to predict the single wood parameter. If the multiple wood parameter detections are required, the single wood detection needs to be performed some times. In order to improve the wood parameter detection's efficacy, we propose a simultaneous prediction scheme for wood species and wood density parameters with only one prediction. First, the K/S algorithm is used to divide the training and prediction sets to make them representative. Then, two dimensionality-reduction methods of principal component analysis and wavelet transform are combined with BP neural network and least squares support vector machine to establish four prediction models that can predict both wood species and density. In experiments, a small fiber spectrometer of USA Ocean Optics USB2000-VIS-NIR is used to acquire the visible/near infrared spectral curves with a spectral interval of 350~1 100 nm. The results show that all four models can achieve simultaneous prediction of wood species and density, and the model established by wavelet transform dimensionality-reduction method combined with least squares support vector machine is relatively better. The correct recognition rate of wood species based on the combination of wavelet transform and partial least squares support vector machine is 100%, the density correlation coefficient of training set is 0.973 4, the density correlation coefficient of prediction set is 0.940 8, the density training root mean square error is 0.026 13, and the prediction root mean square error is 0.038 46. It lays a theoretical foundation for the development of portable real-time on-line detection instruments that can simultaneously predict several parameters of wood physical feature. Moreover, another spectrometer of FLAME-NIR with a spectral interval of 900~1 650 nm is also used to perform the same prediction experiments. By comparisons, we find that the prediction results with the FLAME-NIR model are slightly superior to those with the USB2000-VIS-NIR model. Therefore, our simultaneous prediction of wood species and wood density is practical with a definite stability, accuracy, and a low instrumentation cost.

Keywords Spectral analysis; Wood species; Wood density; Simultaneous prediction

(Received Sep. 20, 2018; accepted Jan. 10, 2019)