

doi : 10. 16473/j. cnki. xblykx1972. 2020. 05. 019

## 基于集群算法优化 BP 神经网络的 NIRS 树种识别研究<sup>\*</sup>

明曼曼, 陈芳, 孙恺琦, 顾崎岩, 吴思齐, 王学顺  
(北京林业大学, 北京 100083)

**摘要:** 为探究基于近红外光谱分析技术的木材树种准确、快速识别新方法, 并研究光谱波段范围对识别的影响, 以大叶桉、杉木、落叶松、马尾松、樟子松 5 种木材样品为研究对象, 针对 3 种光谱波段范围, 分别建立未优化的 BP 神经网络模型 (BP-ANN)、粒子群算法优化 BP 神经网络模型 (PSO-BP) 和人工蜂群算法优化 BP 神经网络模型 (ABC-BP), 对比模型识别准确率和运行时间。研究结果显示, 波段越长, 模型的识别准确率越高; PSO-BP 与 ABC-BP 模型的识别准确率均高于 BP-ANN, 其中 ABC-BP 模型的识别效果最好, 可达到 95.333%; ABC 算法较于 PSO 算法优化 BP 神经网络模型用于木材近红外光谱树种识别时间更短。基于集群算法优化 BP 神经网络模型能有效应用于树种识别研究, 具有一定应用前景。

**关键词:** 近红外光谱技术; 木材树种识别; BP 神经网络; 人工蜂群算法; 粒子群算法

**中图分类号:** S 781; O 657.3    **文献标识码:** A    **文章编号:** 1672-8246 (2020) 05-0124-05

## NIRS Tree Species Identification Based on Cluster Algorithm Optimized BP Neural Network

MING Man-man, CHEN fang, SUN Kai-qi, GU Qi-yan, WU Si-qi, WANG Xue-shun  
(Beijing Forestry University, Beijing 100083, P. R. China)

**Abstract:** In order to explore a new method for accurate and rapid identification of wood species based on near-infrared spectroscopy analysis technology, and study the impact of spectral band range on identification, an unoptimized BP neural network model (BP-ANN) and a particle swarm optimization algorithm were established for the three spectral band ranges to optimize the BP neural network. Five kinds of wood samples, namely eucalyptus, fir, larch, masson pine and scotch pine were selected as research objects. The network model (PSO-BP) and artificial bee colony algorithm optimize the BP neural network model (ABC-BP), and compare the model recognition accuracy and running time. The research results showed that the longer the band, the higher the recognition accuracy of the model. The recognition accuracy of the PSO-BP and ABC-BP models is higher than that of BP-ANN, of which the ABC-BP model has the best recognition effect, which can reach 95.333%. Compared with the PSO algorithm, the ABC algorithm in the optimized BP neural network model can be used to identify tree species in the wood near-infrared spectrum in a shorter time. In general, optimized BP neural network model based on the cluster algorithm can be effectively applied to the research of tree species identification and has certain application prospects.

**Key words:** near infrared spectroscopy; timber species identification; BP neural network; artificial bee colony algorithm; particle swarm optimization

<sup>\*</sup> 收稿日期: 2020-04-14

基金项目: 北京林业大学大学生创新创业训练项目 (G201910022066), 中央高校基本科研业务费专项基金 (2015ZCQ-LY-01)。

第一作者简介: 明曼曼 (1988-), 女, 主要从事数据挖掘、人工智能研究。E-mail: Mingman\_0808@163.com  
通信作者简介: 王学顺 (1959-), 男, 教授, 博士, 主要从事数据挖掘、人工智能研究。E-mail: Wangxueshun6688@sina.com

木材是一种重要的可再生资源。随着我国经济、社会的快速发展,国民对木材的需求日益增加,木材市场不断扩大,但木材供需不平衡、木材造假等问题也相继出现。木材识别是根据木材构造、材色、纹理与气味等特征确定木材树种名称的工作<sup>[1]</sup>。木材识别对合理产材、合理用材、提高林木培育质量、寻找代用材、规范木材市场、打击木材造假等起到了很重要的作用<sup>[2]</sup>。木材识别技术种类繁多,其中基于近红外光谱的木材识别技术相比于传统解剖识别<sup>[3]</sup>、木材数据库检索<sup>[4]</sup>、计算机木材图像识别<sup>[5-6]</sup>、DNA 识别<sup>[7-10]</sup>等其他木材识别技术,不仅更加快速、准确,同时还具有无损、无污染、高效率的特点。

近红外光谱分析技术 (near infrared reflectance spectroscopy, NIRS) 是 20 世纪 70 年代兴起的一种新的成分分析技术。近红外光谱主要源自于有机物中含氢集团等的倍频及合频吸收,通过漫反射方式得到近红外区 (波长范围为 780~2 526 nm) 的吸收光谱<sup>[11]</sup>。木材是由纤维素、半纤维素和木质素三种主要成分构成,其结构中含有大量的羟基、碳氢键等含氢基团。通过采集木材近红外光谱,可以得到木材的物质结构信息;并且将近红外光谱技术与化学计量学方法结合,可有效应用于木材识别,如杨忠等<sup>[12]</sup>通过近红外光谱技术结合化学计量学中主成分分析方法对八类红木的近红外光谱进行分析,通过绘制三维 PCA 得分图有效识别八类红木;马明宇等<sup>[13]</sup>使用 BP 神经网络与广义回归神经网络建立了 NIRS 树种识别模型;庞晓宇等利用近红外光谱结合化学计量学中 BP 神经网络对三种人工林木材进行识别,探讨了隐含层神经元个数、光谱预处理方法等对 BP 神经网络模型的影响<sup>[14]</sup>。

反向传播 (back propagation, BP) 人工神经网络<sup>[15]</sup>作为一种化学计量学方法,按照误差逆向传播原理通过对已知试样训练,来预测未知试样。由于 BP 算法易陷入局部最优,且随机初始化的权值和阈值易使结果波动较大,故常引用自然启发式的元启发式算法进行权值和阈值的优化。本研究选择基于群体的元启发式算法进行优化:一是粒子群算法<sup>[16]</sup> (particle swarm optimization, PSO),模拟鸟群飞行觅食的行为,通过鸟之间的集体协作使群体达到最优,此算法基于群体迭代,群体在解空间中追随最优粒子进行搜索;二是人工蜂群算法<sup>[17]</sup> (Artificial Bee Colony, ABC),模拟蜜蜂采蜜行为,

通过角色分工和转换实现群体协作,以寻找最优解。

本研究对桉树 (大叶桉) (*Eucalyptus robusta* Smith)、杉木 (*Cunninghamia lanceolata*)、落叶松 (*Larix gmelinii* (Rupr.) Kuzen. )、马尾松 (*Pinus massoniana* Lamb.)、樟子松 (*Pinus sylvestris* var. *mongolica* Litv.) 这 5 个树种的树木样本的近红外光谱数据按照 3 种不同波段进行主成分降维处理,分别将所得结果输入 BP 神经网络模型、粒子群算法优化的 BP 神经网络模型 (PSO-BP) 和人工蜂群算法优化的 BP 神经网络模型 (ABC-BP),比较 3 种模型所得结果的准确率以及运行时间,为木材识别的研究提供科学参考。

## 1 材料与方法

### 1.1 试验材料

试验材料为 274 个木材近红外光谱样本,其中桉树 46 个、杉木 60 个、落叶松 64 个、马尾松 62 个、樟子松 42 个。样本数据均由中国林业科学研究院木材工业研究所提供。

### 1.2 试验仪器

本研究所用到的试验仪器主要包括: Field Spec 近红外光谱仪 (美国 ASD 分析光谱仪器公司)、光谱分析软件 The Unscrambler v9.7。

### 1.3 试验方法

在光谱实验室温度 ( $22 \pm 1.5$ ) °C,湿度 ( $50 \pm 3$ ) %RH 的采集环境下,使用美国 ASD 公司的 Field Spec 近红外光谱仪进行木材光谱数据的采集,光谱仪的波长范围为 350~2 500 nm,每条光谱包含 2 151 个数据点。光谱采集时,使用光纤探头采集样品表面的近红外光谱,光谱插空白校准采用商用聚四氟乙烯制成的白色材料,照射光源用 DC 直流灯,在样品上方与样品成 60° 夹角,并与样品纵轴保持平行。见图 1、图 2。



图 1 近红外光谱仪

Fig. 1 Near Infrared Spectrometer

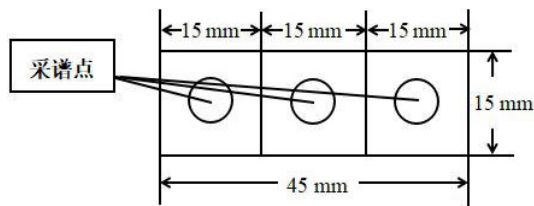


图 2 样品采谱位置图

Fig. 2 Location of wood samples for NIR

274 个样本分别以每秒 30 次扫描结果取平均值为最终样本光谱数据。光谱采集结束后, 光谱数据通过光谱分析软件 The Unscrambler v9.7 以 Unscrambler R 文件形式保存。

1.4 数据预处理

对大叶桉、杉木、落叶松、马尾松、樟子松 5 种木材的近红外光谱数据进行建模, 光谱范围为 780~2 500 nm。

数据从波段选择和主成分降维两方面进行预处理。先分别选取 780~1 100、1 100~2 500 和 780~2 500 nm 3 种波段建模, 然后通过主成分分析方法 (principal component analysis, PCA) 对 3 种不同波段的光谱数据分别进行降维处理, 以达到排除木材光谱数据所包含的众多化学信息共存中相互重叠信息的目的<sup>[14]</sup>。3 种波段的光谱数据分别经过主成分降维后, 原始光谱均降至 3 维时, 累计贡献率达到 99%, 将预处理后的综合变量作为模型的输入变量储存在 Excel 中。

对于 BP 神经网络模型的输出数据, 先分别对不同树种按顺序编号, 大叶桉、杉木、落叶松、马尾松、樟子松分别对应编号 1、2、3、4、5。再对 BP 神经网络的期望输出采用 0、1 分类, 除对应编号位置记为 1 外, 其余位置记为 0。本研究取每种木材随机排列后前 2/3 的光谱数据合为训练样品集 (183 个), 剩下 1/3 的光谱数据合为测试样品集 (91 个), 各种木材对应的期望输出和训练集、测试集数目见表 1。

表 1 5 种木材样品的相关信息

Tab. 1 Information on five wood samples

编号	树种名称	期望输出	训练集 / 个	测试集 / 个
1	大叶桉	(1 0 0 0 0)	31	15
2	杉木	(0 1 0 0 0)	40	20
3	落叶松	(0 0 1 0 0)	43	21
4	马尾松	(0 0 0 1 0)	41	21
5	樟子松	(0 0 0 0 1)	28	14

注: 大叶桉 (*E. robusta*)、杉木 (*C. lanceolata*)、落叶松 (*L. gmelinii*)、马尾松 (*P. massoniana*)、樟子松 (*P. sylvestris* var. *mongolica*)。

本研究中所建立的 BP 神经网络模型均为单隐含层 BP 神经网络模型, 根据试验得到最优的隐含层节点数为 7 个, 神经网络的结构确定为 3-7-5, 输入层到隐含层的传递函数为 Tansig 函数, 隐含层到输出层的传递函数采用 Purlin 函数, 随机赋予 BP 神经网络初始权值和阈值。模型设置学习率为 0.1, 最大训练次数为 1 000 次, 目标误差为 0.000 01, 采用均方误差为网络误差。当样品对应位置的输出值为输出的 5 个数中最大值时, 则判定该样品识别正确, 否则判定为识别错误。

2 结果与分析

2.1 基于 BP 神经网络的树种识别模型

本研究数据处理工具为 MATLAB R2018a 软件。将数据随机排列, 选取前 2/3 为训练集, 剩余为测试集, 对不同波段的光谱数据建立网络识别模型, 并记录预测准确率。重复 10 次上述操作, 结果取平均值 (表 2)。

表 2 不同光谱波段选择的 BP 神经网络模型

Tab. 2 BP network models with different spectral band selections

平均准确率 / %	BP Model 1 (780~2 500 nm)	BP Model 2 (1 100~2 500 nm)	BP Model 3 (780~1 100 nm)
大叶桉	100	100	100
杉木	100	99.000	99.500
落叶松	94.286	90.000	92.857
马尾松	94.286	78.571	83.809
樟子松	75.714	82.857	68.571
总准确率 / %	92.857	90.086	88.948

由表 2 可知, 3 种 BP 神经网络模型分类结果准确率均达到 88% 以上。其中, 大叶桉的识别准确率最好, 均为 100%, 杉木的识别准确率次之, 除个别几次实验之外, 识别结果均为 100%; 落叶松的识别准确率稍差, 但仍可达到 90% 以上, 而樟子松与马尾松识别准确率较低, 尤其在波段范围为 780~1 100 nm 的模型判别中, 马尾松的识别准确率仅达到 68.571%。从波段角度分析, 波段范围为 780~2 500 nm 的 BP 神经网络模型对 5 种树木的识别准确率最高, 达到 92.857%。同时在 BP 神经网络模型中, 波段范围越长, 建立的 BP 神经网络模型识别准确率越高。

2.2 基于粒子群算法 (PSO) 优化 BP 神经网络的

树种识别模型 (PSO-BP)

初始化 PSO 算法参数, 设定粒子群算法迭代次数为 1 000, 种群规模为 20, 进化代数 为 80, 以上各参数均由程序反复实验而获得最优值。模型输入与实验软件同 2.1, 记录识别准确率以及算法运行时间。

表 3 不同光谱波段选择的 PSO-BP 模型  
Tab. 3 PSO-BP models with different spectral band selections

平均准确率 /%	PSO-BP Model 1	PSO-BP Model 2	PSO-BP Model 3
大叶桉	100	100	100
杉木	100	99.5	99.5
落叶松	94.761	88.095	96.667
马尾松	94.286	86.667	83.809
樟子松	83.571	90.714	77.143
总准确率 /%	94.524	92.995	91.424
平均运行时间 /s	223.555	212.494	210.208

注: Model 1 为 780~2 500 nm; Model 2 为 1 100~2 500 nm; Model 3 为 780~1 1000 nm。下表同。

由表 3 可知, PSO-BP 模型下, 大叶桉的识别准确率均为 100%, 杉木、落叶松、马尾松的识别准确率均在 83% 以上, 樟子松的识别准确率较差, 其中在光谱波段范围为 780~1 100 nm 的 PSO-BP 模型中, 优化后樟子松的识别准确率仅为 77.143%。从波段角度分析, 波段范围为 780~2 500 nm 的 PSO-BP 神经网络模型对五种树木识别准确率仍是最高, 达到 94.524%。同时也可以看出, 在 PSO-BP 模型中, 波段范围越长, 建立的 PSO-BP 模型识别准确率越高。但 PSO-BP 模型运行时间较长, 其平均运行时间均在 210 s 以上。

2.3 基于人工蜂群算法 (ABC) 优化 BP 神经网络的树种识别模型 (ABC-BP)

设定蜂群规模为 20, 蜜源数量为 10, 最大迭代次数为 50, 个体最大更新次数为 100。维数 D 取 58, 参数范围为  $[-3, 3]$ , 以上各参数均由程序反复实验而获得最优值。模型输入与实验软件同 2.1。基于人工蜂群算法的 BP 神经网络的适应度变化见图 3。

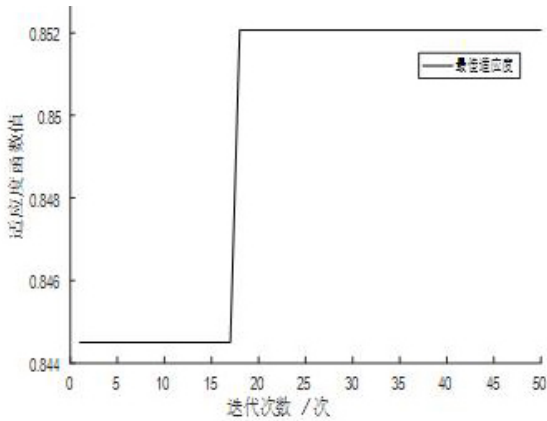


图 3 ABC 适应度变化曲线

Fig. 3 The fitness curve of ABC

由适应度变化曲线可看出, 迭代至第 17 代时, 适应度值大幅度猛烈上升至 0.852, 均方误差迅速减小, 说明 ABC 算法性能较好, 收敛较快。记录识别准确率以及算法运行时间。

表 4 不同光谱波段选择的 ABC-BP 模型  
Tab. 4 ABC-BP models with different spectral band selections

平均准确率 /%	ABC-BP Model 1	ABC-BP Model 2	ABC-BP Model 3
大叶桉	100	100	100
杉木	100	99.023	100
落叶松	95.238	91.904	96.667
马尾松	92.857	87.619	89.047
樟子松	88.571	91.428	78.571
总准确率 /%	95.333	93.995	92.857
平均运行时间 /s	146.705	142.931	141.578

由表 4 可知, 人工蜂群优化的 BP 神经网络识别模型的识别准确率较高, 3 种 ABC-BP 模型识别准确率均达到 92% 以上。其中, 大叶桉的识别准确率为 100%, 除两次试验外, 杉木的识别准确率均为 100%, 落叶松的识别准确率在 91% 以上, 樟子松的识别准确率仍为最低。在光谱波段范围为 780~1 100 nm 的 ABC-BP 模型中, 优化后樟子松的识别准确率仅为 78.571%。从波段角度分析, 波段范围为 780~2 500 nm 的 ABC-BP 神经网络模型对 5 种树木识别准确率仍是最高, 达到 95.33%。同时也可以看出, 在 ABC-BP 模型中, 波段范围越长, 建立的 ABC-BP 模型识别准确率越高。此外, ABC-BP 模型平均运行时间约为 140s。



## 2.4 模型比较

将波段范围为 780~2 500 nm 的 BP 神经网络模型、PSO-BP 神经网络模型以及 ABC-BP 神经网络模型的识别准确率结果进行对比,图 4 为 3 种模型平均识别准确率对比图,可以看出,人工蜂群算法相比于粒子群算法更能够有效提高 BP 神经网络的识别准确率,此外,由表 2~表 4 可以看出,在运行时间方面,ABC-BP 模型运行时间也远少于 PSO-BP 模型,模型效率更高。

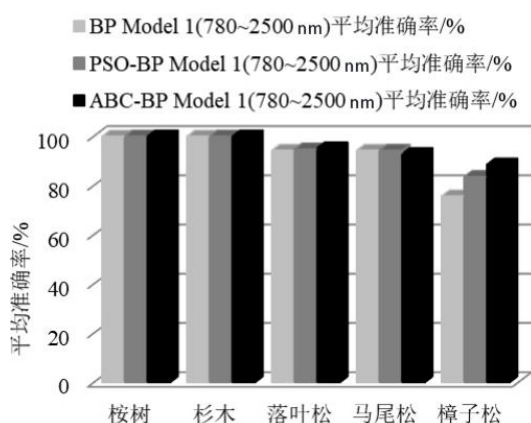


图 4 3 种模型平均识别准确率对比

Fig. 4 Comparison of average recognition accuracy of 3 models

## 3 结语

本文针对大叶桉、杉木、落叶松、马尾松、樟子松 5 种木材的近红外光谱数据,利用集群算法优化 BP 神经网络进行了木材树种的识别研究。分别通过利用粒子群算法和人工蜂群算法计算出 BP 神经网络的最优权值和阈值,以建立新的模型。各模型处理结果显示:波段方面,波段越长,所建立的 BP 神经网络识别模型判别准确率越高;准确率方面,3 种模型中 ABC-BP 模型的识别准确率最高,达到 95.333%,对于集群算法优化的 BP 神经网络模型,人工蜂群算法优化 BP 神经网络较粒子群算法优化 BP 神经网络效果更好;运行时间方面,ABC-BP 模型运行时间明显比 PSO-BP 模型运行时间更短,但 ABC-BP 模型的运行时间明显长于没有优化的 BP 神经网络模型,还需进一步改进。综上所述,在适当的光谱波段范围内,将 BP 神经网络结合近红外光谱用于树种识别研究具有很好的研究价值,而基于集群算法优化 BP 神经网络用于树种识别可以进一步提高树种识别的准确率,其中人工

蜂群算法优化 BP 神经网络用于树种识别其准确率显著提高,具有一定的应用前景。

## 参考文献:

- [1]汪杭军.基于纹理的木材图像识别方法研究[D].北京:中国科学技术大学 2013.
- [2]姜笑梅,殷亚方,刘波.木材树种识别技术现状、发展与展望[J].木材工业 2010 24(4):36-39.
- [3]王宪,沈华杰,于清琳,等.基于 IAWA 的 3 种简易木材识别方法探究[J].西南林业大学学报(自然科学版),2019 39(6):167-172.
- [4]孙书冬,甘雪菲,霍丽平,等.基于 Visual Basic 的木材识别系统[J].木材工业 2013 27(5):21-25.
- [5]戴天虹,王克奇,白雪冰,等.基于神经网络和颜色特征对木材进行分级的分析[J].森林工程 2006 22(1):18-20.
- [6]刘子豪,汪杭军.基于 PCA+FisherTrees 特征融合的木材识别[J].林业科学 2013 49(6):125-131.
- [7]Tnah L H, Lee S L, Ng K K S, et al. DNA extraction from dry wood of *Neobalanocarpus heimii* (Dipterocarpaceae) for forensic DNA profiling and timber tracking [J]. Wood Science and Technology 2012 46(5):813-825.
- [8]Jiao L, Yin Y, Xiao F, et al. Comparative analysis of two DNA extraction protocols from fresh and dried wood of *Cunninghamia lanceolata* (Taxodiaceae) [J]. Iawa Journal 2012 4(33):441-456.
- [9]夏攀,曹禹,王创,等.利用 DNA 条形码技术鉴别植物[J].中国司法鉴定 2014(3):50-52.
- [10]余敏.降香黄檀与越南黄檀木材 DNA 条形码筛选与鉴定研究[D].合肥:安徽农业大学 2015.
- [11]赵雅欣,王红英.近红外光谱分析技术在饲料工业中的应用进展[J].饲料工业 2005 26(21):37-41.
- [12]杨忠,江泽慧,吕斌.红木的近红外光谱分析[J].光谱学与光谱分析 2012 32(9):2405-2408.
- [13]马明宇,王桂芸,黄安民,等.人工神经网络结合近红外光谱用于木材树种识别[J].光谱学与光谱分析 2012, 32(9):2377-2381.
- [14]庞晓宇,杨忠,吕斌.基于近红外光谱与误差反向传播神经网络技术的三种人工林木材识别研究[J].光谱学与光谱分析 2016 36(11):3552-3556.
- [15]许禄,邵学广.化学计量学方法[M].北京:科学出版社 2004:130-138 279-312.
- [16]纪震,廖惠连.粒子群算法及应用[M].北京:科学出版社 2009:13-47.
- [17]沈夏炯,王龙,韩道军.人工蜂群优化的 BP 神经网络在入侵检测中的应用[J].计算机工程 2016 42(2):190-194.

(编辑:胡光辉)