



DOI:10.12404/j.issn.1671-1815.2402909

引用格式:管雪梅,张威,杨渠三.基于改进的多表达式编程算法的木材染色配方预测[J].科学技术与工程,2025,25(7):2865-2873.

Guan Xuemei, Zhang Wei, Yang Qusan. Prediction of wood dyeing formula based on improved multi expression programming algorithm[J]. Science Technology and Engineering, 2025, 25(7): 2865-2873.

轻工业、手工业、生活服务业

基于改进的多表达式编程算法的 木材染色配方预测

管雪梅,张威*,杨渠三

(东北林业大学计算机与控制工程学院,哈尔滨 150000)

摘要 由于珍贵木材日益稀缺以及过度开发导致的严重环境问题,有必要通过对普通木材进行染色来模仿珍贵木材的外观。在本研究中采用计算机辅助染色技术,实现对普通木材的高精度染色,从而创造出外观类似珍贵木材的替代品,减少对它们的依赖。首先,基于基因表达编程(gene expression programming, GEP)的概念,提出了一种多表达式编程(multi-expression programming, MEP)算法来预测染料配比,考虑到多种染料之间的复杂相互作用,采用多基因表达,MEP算法能够处理这些复杂的多种染料之间的相互作用,从而得到更直观的函数表达式。为了提高MEP的函数挖掘准确性,自适应调整突变和重组算子的概率,并使用并行编程来增强函数挖掘效率。与基因表达编程的结果相比,MEP深入挖掘了函数关系,并在颜色预测中获得了0.113的相对偏差结果。

关键词 木材染色;基因表达编程;多表达式编程;计算机颜色匹配;遗传算法;光谱反射率

中图分类号 TS65; **文献标志码** A

Prediction of Wood Dyeing Formula Based on Improved Multi Expression Programming Algorithm

GUAN Xue-mei, ZHANG Wei*, YANG Qu-san

(Northeast Forestry University, Haerbin 150000, China)

[Abstract] Due to the increasing scarcity of precious woods and the severe environmental issues caused by overexploitation, it is necessary to mimic the appearance of precious woods by dyeing ordinary wood. Computer-assisted dyeing technology was utilized to achieve high-precision dyeing of ordinary wood, thus creating substitutes that resemble precious woods and reducing dependence on them. Initially, based on the concept of gene expression programming (GEP), a multi-expression programming (MEP) algorithm was proposed to predict dye ratios. Considering the complex interactions among various dyes, multi-gene expression was employed. The MEP algorithm can handle these complex interactions between multiple dyes, resulting in more intuitive functional expressions. To enhance the function mining accuracy of MEP, the probabilities of mutation and recombination operators were adaptively adjusted, and parallel programming was employed to boost function mining efficiency. Compared to gene expression programming results, MEP delves deeper into functional relationships and achieves a relative deviation of 0.113 in color prediction.

[Keywords] wood staining; gene expression programming; multi-expression programming; computer color matching; genetic algorithms; spectral reflectance

随着全球气候变化、人口压力和不可持续的伐木行为加剧,珍贵木材的稀缺性日益显现。世界自然基金会报告称,全球每年约有1 000万 hm^2 的森林面积消失,最高速度发生在热带地区,这些地区

是一些最有价值的木材树种的家园。这种情况给建筑、家具和装饰艺术等行业带来了重大挑战,这些行业对桃花心木、柚木和紫檀等珍贵木材的需求量很大。例如,国际自然保护联盟将300种用于木

收稿日期:2024-04-21 修订日期:2024-07-30

基金项目:国家自然科学基金面上项目(32171691);黑龙江省自然科学基金联合引导项目(LH2020C37);哈尔滨市科技局制造业创新人才项目(CXRC20221110393)

第一作者:管雪梅(1980—),女,汉族,黑龙江哈尔滨人,博士后,教授。研究方向:智能决策与木材科学。E-mail:gxm_maomao1980@163.com。

* 通信作者:张威(2000—),男,汉族,黑龙江海伦人,硕士研究生。研究方向:智能决策与木材科学。E-mail:3303993983@qq.com。

投稿网址:www.stae.com.cn

材贸易的树木列为受威胁物种,突显了这些资源的紧急状态^[1-2]。

此外,这也引发了关于环境可持续性和生物多样性保护的重要讨论。联合国估计,森林支持着80%的陆地生物多样性,失去这些森林意味着失去对生态系统健康和人类生存至关重要的物种。在这种背景下,科学家和工程师正在转向更环保和可持续的方法,将普通木材转化为在视觉和触感上与珍贵木材相似的替代品。根据美国环境保护局(Environmental Protection Agency, EPA)的报告,使用来源负责且可持续染色的木材,与使用原始珍贵木材相比,可以将碳足迹减少多达30%。在这种转变中,木材染色和色彩匹配技术的作用变得尤为关键。据估计,高精度染色技术可以将对珍贵木材的需求减少多达50%,为保护全球森林及其维持的生物多样性提供了一个重要解决方案^[3-4]。

木材染色需要高水平的技术精度和科学知识。本研究的目标是通过计算机辅助染色技术准确模拟珍贵木材的颜色,从而能够更精确地染色普通木材,进而创建外观上无法区分的珍贵木材替代品,而不会产生高昂的环境成本。

近年来,神经网络和深度学习已被用于计算机辅助染色中的颜色预测。Chen等^[5]整合了卷积神经网络(convolutional neural networks, CNN)、多层感知机(multilayer perceptron, MLP)和深度残差网络(ResNet)3种类型的神经网络模型来预测纺织系统的染色,达到了令人满意的颜色差异评估性能。Furferi等^[6]将库贝尔卡-蒙克理论与人工神经网络结合起来,用于预测织物的反射率,从而提高了库贝尔卡-蒙克理论的性能。Wang等^[7]利用粒子群优化对反向传播(back propagation, BP)神经网络进行了优化,用于颜色预测,从而显著提高了胶囊颜色配置的精度。李文峰^[8]通过改进的极限学习机(extreme learning machine, ELM)算法预测了木材染色配方,取得了优越的色彩匹配效果。黄焯^[9]成功地将BP神经网络和遗传规划应用到色纱线的颜色预测中,取得了优异的结果。Chaouch^[10]提出了一个新模型,使用遗传算法来解决颜色配方预测问题,计算了建议配方中预测浓度与实际浓度之间的误差,该算法的有效性已经得到评估和证实。Zhang等^[11]利用基于改进的循环神经网络的深度学习算法建立了光谱反射率和染色配方之间的模型,该模型可以提供高度的预测准确性。Li等^[12]的研究基于特征加权支持向量回归和粒子群优化,研究了一种预测织物染色配方的方法,该方法利用粒子群优化改进了特征加权支持向量回归,并表现出良好的

结果,目标颜色与再现颜色之间存在轻微的平均色差。

李博文等^[13]提出一种基于遗传算法优化单类支持向量机的方法对油田离心泵注水站进行异常检测,得到极高的精度。由遗传算法(genetic algorithm, GA)和遗传规划(genetic programming, GP)发展而来的基因表达式编程(gene expression programming, GEP),由于其简单的编码和处理复杂问题的能力,逐渐被应用于预测问题中。GEP吸收了遗传算法和遗传规划的优点,可以在更小的搜索空间中获得更复杂的解,并且可以自由改变基因的长度和数量,从而更好地适应问题的复杂性^[14]。由于其稳健和灵活的特性,GEP可以处理复杂问题,使其能够应对色彩匹配中涉及的复杂情况^[15-17]。该研究在伪装色彩匹配中应用多元回归分析,利用色差计算公式建立了非线性函数模型。这确保了光谱曲线符合相同颜色、相同光谱的要求^[18]。

多表达式编程(multi-expression programming, MEP)是对GEP的改进,使单个染色体能够包含多个表达式,与GEP相比,这显著提高了代码的可重用性和预测准确性^[19]。现基于染色木板的光谱反射率来预测3种红色染料的浓度。3种染料对光谱反射率的影响相互作用,造成了一个复杂的情况,其中存在多个目标变量之间的相互作用或相互作用^[19]。多表达式编程算法可以有效解耦这些纠缠的问题^[20]。

迄今为止,还没有学者使用MEP来预测使用大量数据的木材染色的染料配方。本研究旨在填补这一空白,并充分利用MEP来解耦染料配方和光谱反射率之间的关系,获得它们之间关系的表达式。这使得可以基于所获得的贵重材料的光谱反射率,用简单的数学表达式来预测染料配方。

1 数据采集

对于木材染色配方,需要大量的数据训练来确保预测的准确性。在数据获取方面,本研究使用三种类型的染料进行木材染色,并从染色木样本中获取光谱反射数据。

1.1 实验材料和设备

樟子松胶合板(樟子松经旋切加工成尺寸为40 mm × 70 mm × 1 mm的染色胶合板);漂白剂(H₂O₂溶液);反应性红色(X-3B)染料、反应性黄色(X-RG)染料、反应性蓝色(X-3G)染料;渗透剂(水性JFC溶液);固色剂(无水碳酸钠);染料促进剂(NaCl);来自芬兰的SPECIM高光谱分析仪;数字显示恒温水浴(上海立辰邦熙仪器技术有限公司);电子天平;烧杯;202-I型电动恒温干燥箱。

1.2 实验步骤

(1) 漂白:漂白可以去除木材中的一些提取物,从而提高染料的吸收率。准备浓度 4% 的 H_2O_2 500 mL 漂白溶液,并加入 0.5 g Na_2SiO_3 和 Na_3PO_4 。将准备好的染色单板放入装满漂白溶液的烧杯中。将烧杯放入 65 °C 的恒温水浴中,设置水浴比为 1:20。经过 2 h 后,取出单板,并用清水冲洗,去除表面残留物(图 1)。

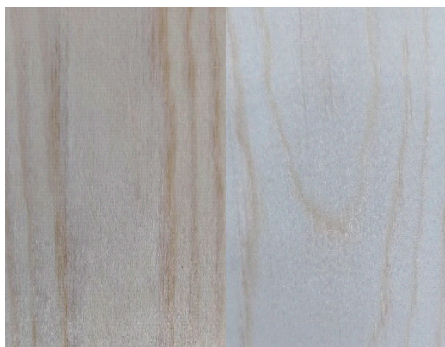


图 1 漂白前和漂白后的木质胶合板

Fig. 1 Wood veneer before and after bleaching

(2) 染色:根据预先设置的 800 组染料配比数据,对木材进行染色。根据每组染料的浓度,准备 500 mL 的染料溶液,并添加 0.5 mL JFC(脂肪醇聚氧化稀醚)溶液和 15 g/L 的 NaCl。将染色单板放入烧杯中,然后将烧杯放入 65 °C 的恒温水浴中。加热 2 h 后,添加 20 g/L 的无水碳酸钠进行固定,30 min 后,取出木板并用清水冲洗,去除表面残留物。

(3) 干燥:将染色后的染色单板放入干燥箱中,直至湿度达到 6% ~ 8%。

(4) 光谱数据获取:使用芬兰 SPECIM 高光谱分析仪拍摄已干燥的染色胶合板的照片,并使用 ENVI (environment for visualizing images) 提取每块板的光谱反射率。为确保数据的准确性,随机选择板的五个均匀染色部分。提取 400 ~ 700 nm 波段的平均光谱反射率,每个波长间隔约为 2.6 nm。从某一染色木样本中提取的光谱反射率如表 1 所示(省略 426.49 nm 之后的数据)。

表 1 染色木胶合板部分波长的光谱反射率数据

Table 1 Partial wavelength spectral reflectance data for a stained wood veneer

波长/nm	光谱反射率	波长/nm	光谱反射率
397.66	0.363	416.00	0.188
400.28	0.322	418.62	0.177
402.90	0.283	421.24	0.168
405.52	0.251	423.86	0.163
408.13	0.226	426.49	0.155
410.75	0.206	⋮	⋮
413.37	0.198		

2 数据特征提取

特征提取在处理高维数据时是至关重要的步骤。它有助于减少数据的维度,同时保留数据的主要结构和信息,从而提高机器学习模型的性能。本研究中测量的光谱数据是在 397 ~ 700 nm 可见光波段之间的光谱反射率。数据集包含大量特征,但并非所有特征对于预测任务都是有用的。本研究使用了随机森林模型来评估每个特征的重要性,通过计算每个特征在所有决策树中降低得不纯度的平均值来实现的^[21-22]。

对特征重要性的计算基于决策树的分裂过程。光谱反射率的特征提取实际上是一个回归问题。对于回归问题,假设有 m 棵树,每棵树 t 都有其节点的分裂过程。特征 i 带来的均方误差的平均减少可以表示为

$$MSER_{i(t)} = \sum (N_{node} / N_{total}) (MSE_{parent} - MSE_{node}) \quad (1)$$

式(1)中: N_{node} 为在每个节点分裂后获得的样本数量; N_{total} 为总样本数; MSE_{parent} 为父节点的均方误差; MSE_{node} 为在每个节点分裂后获得的子节点的均方误差。 $MSER_{i(t)}$ 与特征 i 的重要性成正比。

然后,本研究将数据集分为 80% 的训练集和 20% 的测试集,然后在训练集上训练了一个随机森林回归模型,并在测试集上对其进行了测试。在模型训练完成后,就得到了每个特征的重要性(以百分比表示),并按重要性对特征进行了排序。

本研究选择了前 20 个重要性最高的波长进行后续模型构建。所选波长如表 2 所示。

表 2 木材单板经过特征提取后的部分样本光谱反射率数据

Table 2 Partial sample spectral reflectance data of wood veneer after feature extraction

波长/nm	光谱反射率		
	单板 1	单板 2	单板 3
455.43	0.104	0.092	0.100
447.52	0.108	0.097	0.107
468.62	0.089	0.078	0.081
546.26	0.052	0.043	0.041
561.59	0.051	0.042	0.040
473.90	0.084	0.074	0.070
450.16	0.106	0.095	0.103
463.34	0.097	0.085	0.090
444.89	0.109	0.098	0.109
397.66	0.362	0.350	0.436
542.91	0.051	0.043	0.042
400.28	0.321	0.308	0.374
460.70	0.099	0.087	0.091
452.79	0.105	0.094	0.100
556.25	0.050	0.041	0.039
402.90	0.282	0.269	0.327
545.57	0.051	0.042	0.040
550.91	0.050	0.042	0.040
532.25	0.054	0.046	0.045
569.61	0.054	0.044	0.042

3 模型描述

3.1 基因表达编程

基因表达编程 (gene expression programming, GEP) 是一种遗传算法,它通过采用线性染色体结构和基因表达机制,基于遗传规划 (genetic programming, GP),提高了遗传算法的性能和效率。与传统的遗传规划不同,GEP 将基因型和表型分离开来,从而更有效地搜索解空间^[23]。在本研究中,将 GEP 应用于木材染色领域,因为它有潜力生成高度准确和可解释的模型。

3.1.1 GEP 编码原理

基因表达编程的核心是个体的概念,个体由一个或多个基因组成。每个基因由一个包含函数和终端的头部以及仅包含终端的尾部组成。尾部的长度是根据头部中函数的最大数量计算的,这使得表型表示具有更大的灵活性。头部基因包括函数和函数符号集 (例如 +、-、*、√、sin 等) 和终端符号集 (输入),而尾部基因仅包含终端符号集。基因的头部和尾部之间存在固定的关系,如果头部的长度为 h ,并且基因的函数集中包含的操作的最大数量为 n ,那么尾部基因的长度 e ^[24] 为

$$e = h(n - 1) + 1 \tag{2}$$

为了增强算法的搜索能力和适应性,在基因的尾部附加了一个多样性贡献 (diversity contribution, DC) 域。DC 域在基因编码中引入了额外的信息。在算法中,本研究使用一个随机的 DC 域。DC 域的长度等于尾部基因的长度,即

$$L_{DC} = e \tag{3}$$

基因结构如图 2 所示。按顺序分为头基因、尾基因和额外的 DC 域。每个 GEP 基因被解码成一个表达树,终端“?”代表一个随机常数,从左到右用 DC 域的符号替换。假设随机 DC 域生成的数组是 $[a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}]$,则将其转换为表达树如图 3 所示。

0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2
+	?	*	+	?	*	*	a	a	a	?	?	a	a	a	6	8	0	8	3	2	9	5

图 2 基因结构示例

Fig. 2 Example of gene structure

3.1.2 GEP 遗传操作

(1) 突变操作:父染色体被用作突变的对象。根据突变率随机选择基因片段进行突变。如果突变发生在头部,相应的基因片段可以突变为任何函数符号集和终端符号集。如果突变发生在尾部,只能突变为终端符号集^[25-27]。

(2) 重组操作:重组可以分为多点重组、两点重

组和单点重组。单点重组是指在两个父染色体上选择一个点,然后交换这个点后面的所有基因。两点重组是指在每个父染色体上选择两个点,然后交换这两个点之间的所有基因。在多点重组中,首先随机选择 N 个交叉点。然后,从第一个基因开始,子代交替地从两个父代中复制基因。每次到达一个交叉点时,就改变从哪个父代复制基因。多点重组的示例如图 4 所示,其中选择了 3 个交叉点,具体位置分别为 2、5 和 7。

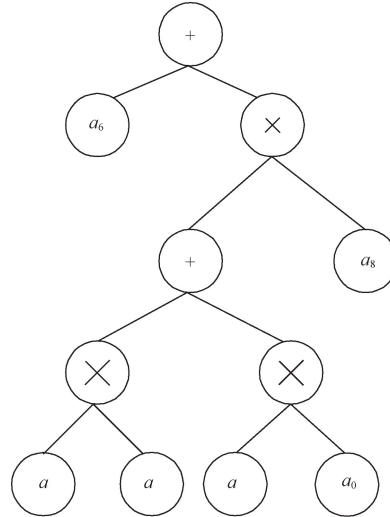


图 3 相应的表达树

Fig. 3 Corresponding expression tree

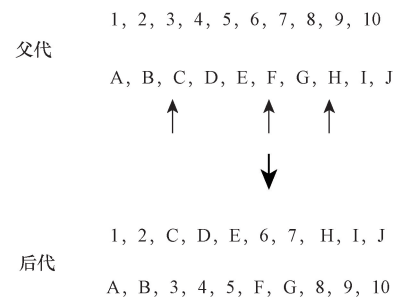


图 4 多点重组示例

Fig. 4 Multi-point recombination example

(3) 插串:包括插入序列 (insertion sequence, IS) 插串和根插入序列 (root insertion sequence, RIS) 插串,在 IS 插串中,随机选取一小段序列 (通常是尾部的一部分),并插入到头部的某个位置。头部通常对应于表达式树的更高层次的节点,有助于改变头部的结构和行为;RIS 插入与 IS 插入类似,只是被选取的序列被插入到头部的开始位置 (也就是根位置)。这是一种更激进的插入,因为它可以彻底改变表达式树的根,从而可能产生完全不同的行为。如图 5 示例,在 IS 中将 D、E 插入头部第二个位置,在 RIS 中将 D、E 插入跟位置。

3.2 多基因表达编程算法

为了更好地模拟光谱反射的复杂过程和3种颜色浓度函数的挖掘,引入了多基因染色体的概念。在遗传算法中,染色体是解的表示,基因是解的一部分。在这种情况下,每个染色体代表一个可能的模型,其中每个基因对应于模型中一种颜色的浓度。因此,得到的染色体由3个基因组成,每个基因对应于3种颜色中的一种的浓度,如图6所示。

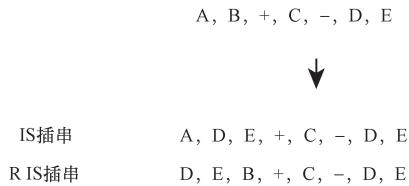


图5 插串和跟插串实例

Fig. 5 Insertion and root insertion examples

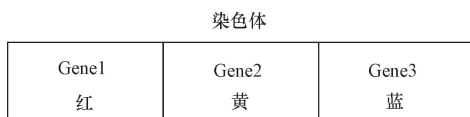


图6 多基因染色体

Fig. 6 Multi-gene chromosome

使用多基因染色体的优势在于可以更好地捕捉各种颜色浓度之间的相互关系,同时保持模型的适应性。

3.3 多表达式编程算法

在传统的GEP算法中,基因表达的长度是固定的,这可能会限制算法在处理复杂问题时的表达能力。这种固定长度的限制可能会导致性能下降。为了解决GEP中表现出的低效率和缺乏精确度的问题,在引入多基因表达编程算法的基础上,本研究进一步引入了多表达式编程算法(MEP)。MEP的最大优势是引入了多表达的概念,允许每个个体具有多个基因表达,从而更好地适应不同问题的复杂性和特征。

3.3.1 多表达式原理

基于先前构建的多基因表达编程算法,本研究将每个染色体表示为3个独立的线性基因串。MEP中的基因串由一系列基因组成,每个基因可以是一个函数(带有一个或多个参数)或一个终结符(不带参数)。为了从基因串生成表达式,需要根据基因的类型(函数或终结符)决定如何处理它们。如果一个基因是一个函数,则需要获取它的参数,这些参数可能在后续的基因中。如果一个基因是一个终结符,可以直接将其视为一个表达式。

具体过程如下:

- (1) 从基因串的开头开始,取出第一个基因。
- (2) 检查该基因的类型。如果它是一个函数,

根据函数的参数数量,从其后面的基因中取出相应数量的基因作为参数,然后将这个函数和参数组合成一个表达式。如果它是一个终结符,则直接将其视为一个表达式。

(3) 保存这个表达式。

(4) 将当前位置移动到下一个未处理的基因,并返回步骤2)。当整个基因串已经遍历完毕时,所有表达式都将被获取到。

3.3.2 自适应概率改进

在传统的GEP中,诸如突变等操作完全是随机的,没有考虑个体的适应度。这可能导致高适应度个体关键特征的破坏,而适应度低的个体则没有足够的改进机会。因此,优化遗传算子操作策略可以提高算法的性能。为了在预测性能和效率方面实现更多的改进,进一步引入了两种优化和改进方法。

(1) 自适应突变概率策略:根据个体的适应度动态调整突变概率。适应度高的个体将具有较低的突变概率,以保留其优秀特征,而适应度较低的个体将具有较高的突变概率,以增加多样性和可探索性。这种适应性可以提高算法的收敛性和搜索能力。

在改进自适应突变概率策略时,本研究首先设置了突变率的范围,即 $[R_{\min}, R_{\max}]$ 对于一个个体染色体,如果其适应度为 f ,那么这个染色体的突变概率为

$$P = R_{\min} + (1 - a)(R_{\max} - R_{\min}) \quad (4)$$

$$a = \frac{f_{\text{chro}} - f_{\min}}{f_{\max} - f_{\min}} \quad (5)$$

式中: f_{\max} 和 f_{\min} 分别为当前种群中的最大和最小适应度; f_{chro} 为当前染色体的适应度值。从式(5)可以看出,突变概率与当前个体的适应度呈线性负相关。因此,适应度较低的个体将被分配更高的突变率。

(2) 自适应重组概率策略:根据种群中所有个体适应度的标准差,自适应调整单点重组和双点重组的概率。标准差 d 用于评估种群的多样性。适应度标准差的范围被设置为 $[T_{\text{upper}}, T_{\text{lower}}]$,那么重组概率为

$$P = \begin{cases} 0.9P_c, & d > T_{\text{upper}} \\ P_c, & T_{\text{lower}} < d < T_{\text{upper}} \\ 1.1P_c, & d < T_{\text{lower}} \end{cases} \quad (6)$$

式(6)中: P_c 为初始给定的重组概率。由式(6)可知,当多样性较大或较小时,重组概率可以自适应调整。

3.4 改进算法描述

MEP算法的流程图如下(图7)。

步骤1 根据预定义的种群大小,随机初始化染色体以创建初始种群。将每个染色体中的基因数设置为3,分别对应不同的染料浓度。设置种群大小

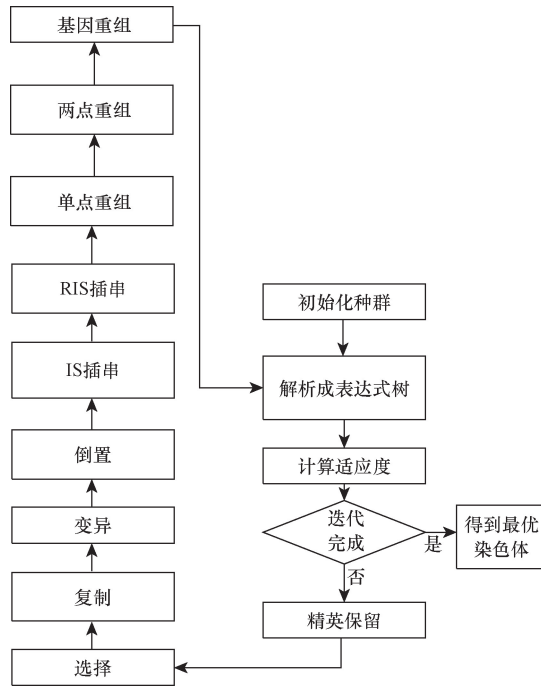


图7 MEP算法流程图
Fig. 7 MEP algorithm flowchart

和迭代次数。

步骤2 将每个染色体中的每个基因解码为表达式树(expression tree, ET), 并使用特定的适应度函数对其进行评估。本研究设计了基于均方误差(mean squared error, MSE)方法的适应度函数。第*i*个个体的均方误差定义为

$$E_i = \sum_{j=1}^n (P_{i,j} - T_j)^2 \quad (7)$$

式(7)中: $P_{i,j}$ 为变量 y 关于求得函数 f 的预测值; T_j 为变量 y 的真实值。那么第 i 个个体的适应度函数为

$$f_i = 1000 \times \frac{1}{1 + E_i} \quad (8)$$

由均方差的含义可知, f_i 的范围为 (0, 1 000)。

步骤3 对每个染色体进行遍历, 计算每个染色体中3个基因的适应度, 并求其平均值。找到具有最高适应度的染色体, 然后对种群进行遗传操作。

(1)选择: 采用轮盘赌选择方法。它根据个体(染色体)的适应度值来分配下一代的复制份额。适应度较高的染色体被选中并复制到下一代的机会更大。设第 i 个个体的适应度为 f_i , 在轮盘赌方法下染色体被保留的概率为

$$P = \frac{f_i}{\sum_i f_i} \quad (9)$$

(2)突变: 在修改后采用自适应突变作为突变方法。

(3)倒置: 倒置操作也需要在基因级别执行。在基因的头部和尾部区域随机选择两个位置, 然后反转这两个位置之间的基因序列, 即将这部分基因序列颠倒。

(4)插串。

(5)重组: 根据改进后的自适应重组概率实现单点重组和双点重组。

(6)基因重组: 基因重组运算符专门针对多基因染色体。对于多基因染色体, 随机选择一个基因, 然后交换两个染色体中相同位置的基因。

步骤4 为了确保进化结束时种群的性能比前一代更好, 或至少与前一代相等, 采用精英保留策略, 将当前代中最不适应的个体替换为上一代中最适应的个体。重复这个过程直到迭代结束, 然后输出最优染色体。

3.5 并行多表达式编程(MEP)方法

并行计算在处理大规模遗传规划问题时可以极大地提高效率和可扩展性。本文中介绍了一种并行多表达式编程(MEP)方法, 可以有效地利用多个处理器进行遗传规划计算。

并行MEP首先将总体种群分成几个子种群, 以利用多个处理器。这些子种群分配给不同的处理器, 然后在并行中执行遗传操作和适应度评估。这种方法的关键优势在于它大大减少了单个处理器需要处理的数据量, 从而提高了计算速度。

本研究使用Python的多进程库实现并行处理, 其能够在每个子种群上并行执行计算密集型的遗传操作。子种群的适应度评估也是并行执行的, 进一步加快了计算速度。

3.6 模型参数设置

首先, 对GEP和MEP的参数进行了设置。考虑到最终选择的输入为20维, 输出为3维, 并考虑到大量的输入数据, 头基因长度被精心选择, 以平衡计算效率与数据探索之间的关系; 设置过长会显著降低GEP的运行效率, 而设置过短可能无法充分利用函数关系。对于改进后的MEP, 引入了多表达式编程的概念, 该算法能够彻底地利用每个基因的表达树, 但这需要大量的计算。因此, 在广泛验证后, 本研究决定相对于GEP减小MEP的种群大小, 以增强算法的效率。在设置自适应重组适应度标准差间隔时, 程序运行了多次以计算种群的平均标准差。然后将该间隔设置为平均标准差的75%~125%。GEP和MEP的参数如表3所示。

数据被分为训练集和测试集, 其中在测试集中选择了10个随机配方进行验证。

表3 GEP和MEP参数设置

参数	GEP	MEP/MEP
种群规模	500	100
单染色体基因数目	3	3
头部基因长度	20	20
迭代次数	100	100
函数集	+、-、√、×、sin、cos、ln、exp、 x^2 、sqrt、 $\frac{1}{x}$	
变异概率	0.3	自适应
倒置概率	0.1	0.1
IS插串概率	0.2	0.2
RIS插串概率	0.2	0.2
单点重组概率	0.3	Adaptive
两点重组概率	0.3	Adaptive
自适应变异概率区间		[0.2, 0.6]
自适应重组初始概率		0.3
自适应重组适应度标准区间		[1.35, 2.25]

4 模型性能和结果分析

在Python环境中,根据设定的程序,分别应用了GEP和MEP进行数据函数挖掘。在实验中,无论是GEP还是MEP,主要都依赖于适应度函数值来进行染色体的选择和进化。从图8和图9的结果观察到,相比于GEP,MEP方法在相同迭代次数和更小的种群规模下,得到的染色体具有更高的适应度和更低的均方误差。这表明,MEP方法生成的模型能够更好地拟合训练数据。

GEP和MEP寻优结果对比如表4所示。

GEP和MEP获得的表达式在测试集上进行验证,使用的评价指标是配方的相对偏差。GEP和MEP获得的表达式在测试集上进行验证,使用的评价指标是配方的相对偏差。平均相对偏差的公式为

$$r = \sum_{i=1}^3 \frac{|y_i - p_i|}{y_i} \quad (10)$$

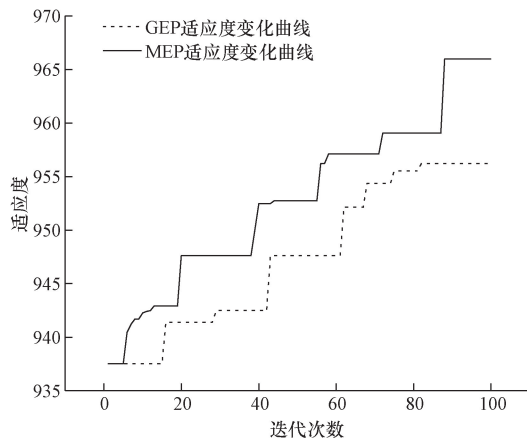


图8 适应度寻优曲线

Fig. 8 Fitness optimization curve

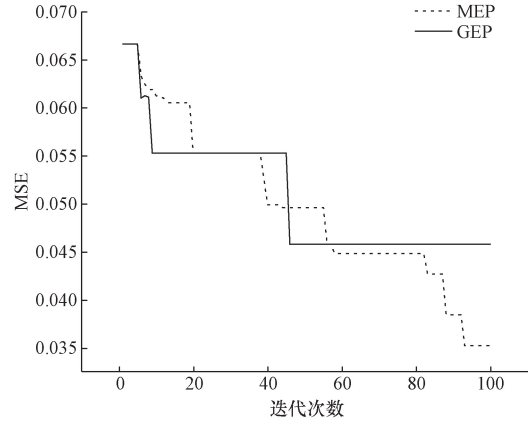


图9 均方误差寻优曲线

Fig. 9 Mean squared error optimization curve

表4 GEP和MEP寻优最佳染色体

Table 4 Optimal chromosome results for GEP and MEP optimization

模型	最优染色体
GEP	+ - + (sqrt) + + K * * (sin) (inv) (sqrt) (sin) (ln) (inv) * (inv) + (sin)/IARSQDSBP NHGOGKNTETJQ101152198142151219412614103116193 (inv) (exp) + / (sin) O (ln) (ln) B (exp) E (inv) * / (inv) (sin) + RTOOTROGTD
MEP	CFFCQAEK? N? T11206162011375031861613116131114 (sqrt) + KC + / (X2) (sin) / / (inv) * (sqrt) K (exp) (ln) / (inv) (inv) (inv) BTCDMJB JMCPQPBKAEKBMF118111721623200861151341814817 (X2) + G - - S + (sqrt) (sqrt) (inv) (inv) (ln) (ln) (inv) (sqrt) (X2) + G - - SHDHTDD ARLJKSAILDEFID 12131413191614114709831622010161119/ (sqrt) (inv) (ln) (sin) (inv) D (X2) + - (inv) (ln) (ln) TK (exp) (sin) + + (sqrt) HLO LGHFFA? GJJPGO? RBRO31612188144082023741761287916 (X2) (sin) - (sqrt) C (sin) * (X2) (X2) - + (X2) - (inv) - (X2) + (ln) (sqrt) OKBMD? H? EHNIC? LPDCF? QO31015112019811716231434201811617

注:三种染料浓度的函数关系是通过解码染料浓度的染料来获得的。

式(10)中: $y_i (i = 1, 2, 3)$ 为测试样本的真实比例; $p_i (i = 1, 2, 3)$ 为测试样本的预测染料比例,预测结果如表5、表6所示。

根据GEP模型和MEP模型的配方预测结果,绘制配方相对偏差曲线图,如图10所示。由曲线可以看出,与GEP预测结果相比,MEP预测效果更好,对大部分测试样本得到的配方相对偏差更小,表7为测试样本的平均相对偏差,即MEP模型预测结果与实际配方更为接近。这表明MEP算法在处理复杂的配方预测问题上具有优势。相对于GEP模型,

表5 GEP 预测结果

Table 5 Predicted results for GEP

样本 编号	真实配方			GEP 预测配方			相对偏差
	活性红	活性黄	活性蓝	活性红	活性黄	活性蓝	
1	0.14	0.38	0.05	0.158	0.417	0.046	0.101
2	0.27	0.23	0.1	0.330	0.272	0.073	0.224
3	0.25	0.65	0.05	0.176	0.567	0.054	0.167
4	0.19	0.52	0.03	0.196	0.557	0.028	0.054
5	0.04	0.21	0.04	0.032	0.197	0.047	0.147
6	0.13	0.11	0.02	0.170	0.143	0.014	0.295
7	0.3	0.6	0.1	0.328	0.694	0.073	0.174
8	0.14	0.38	0.03	0.184	0.252	0.026	0.269
9	0.23	0.1	0.12	0.246	0.113	0.113	0.089
10	0.03	0.06	0.01	0.032	0.062	0.010	0.035

表6 MEP 预测结果

Table 6 Predicted results for MEP

样本 编号	真实配方			MEP 预测配方			相对偏差
	活性红	活性黄	活性蓝	活性红	活性黄	活性蓝	
1	0.14	0.38	0.05	0.161	0.452	0.051	0.121
2	0.27	0.23	0.1	0.285	0.270	0.118	0.139
3	0.25	0.65	0.05	0.261	0.663	0.051	0.029
4	0.19	0.52	0.03	0.216	0.668	0.036	0.202
5	0.04	0.21	0.04	0.040	0.212	0.041	0.013
6	0.13	0.11	0.02	0.109	0.126	0.025	0.179
7	0.3	0.6	0.1	0.287	0.625	0.118	0.088
8	0.14	0.38	0.03	0.167	0.374	0.035	0.128
9	0.23	0.1	0.12	0.284	0.105	0.141	0.153
10	0.03	0.06	0.01	0.032	0.063	0.011	0.076

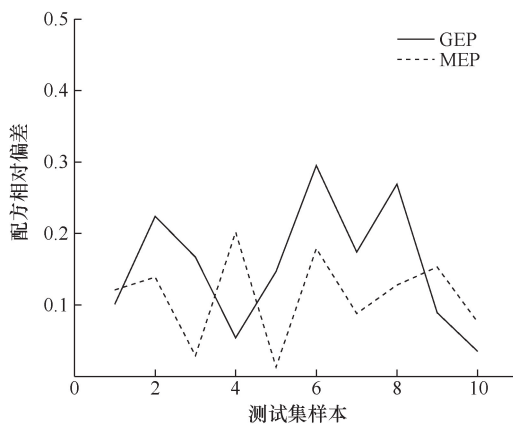


图10 GEP、MEP 模型预测配方相对偏差

Fig. 10 Relative deviation of recipe prediction by GEP and MEP models

表7 GEP 模型和 MEP 模型配方相对偏差对比

Table 7 Comparison of mean relative deviation for GEP and MEP models

模型	GEP	MEP
配方相对偏差均值	0.156	0.113

MEP 模型能够更准确地挖掘出配方中各个成分之间的复杂交互关系,从而提高了预测的精度和准确性。

5 结论

从本次研究结果可以看出,在探索与木材染色浓度相关的函数模型方面,多表达式编程(MEP)技术展示了其优越性。MEP 通过将多个简单表达式整合成一个复杂模型,有效地描述了染色浓度与影响因素之间的复杂关系。这种方法不仅提高了模型的描述能力,也增强了预测精度,从而优化了染色工艺流程,提升了生产效率和产品质量。

MEP 的成功也依赖于准确的输入数据和模型。所以,准确收集和测量染色参数、变量和目标是获得可靠的模型所必要的。此外,模型中的误差和不确定性可能会影响优化结果的准确性和可靠性,并且所得到的函数模型在一般化方面可能存在弱点。

通过关注以上因素,MEP 在木材染色浓度预测和优化方面的应用将更加稳健和有效。

参 考 文 献

- [1] Raj A, Jhariya M K, Yadav D K, et al. Environmental and sustainable development through forestry and other resources [M]. New York: Apple Academic Press, 2020: 1-24.
- [2] Smith P H, Lore-Chavez I. The matter of mimesis [M]. Leiden: Brill, 2023: 27-53.
- [3] Li J, Lindenmayer D B, Yang W, et al. Diversity and density patterns of large old trees in China [J]. Science of the Total Environment, 2019, 655: 255-262.
- [4] Jons G M, Keane J J, Gutiérrez R J, et al. Declining old-forest species as a legacy of large trees lost [J]. Diversity and Distributions, 2018, 24(3): 341-351.
- [5] Chen M, Tsang H S, Tsang K T, et al. An hybrid model CMR-color of automatic color matching prediction for textiles dyeing and printing [C]//Neural Computing for Advanced Applications; Second International Conference. Guangzhou: Springer, 2021: 603-618.
- [6] Furferi R, Governi L, Volpe Y. Color matching of fabric blends: hybrid Kubelka-Munk + artificial neural network based method [J]. Journal of Electronic Imaging, 2016, 25(6): 061402.
- [7] Wang Q, Shuai L, Dai X, et al. PSO-BP neural network-based gelatin hollow capsule color matching model [C]//2020 IEEE International Conference on Mechatronics and Automation (ICMA). New York: IEEE, 2020: 94-99.
- [8] 李文峰. 基于极限学习机的杉木单板染色智能配色模型研究 [D]. 哈尔滨: 东北林业大学, 2022.
Li Wenfeng. Intelligent color matching model for wood dyeing using genetic algorithm and extreme learning machine [D]. Harbin: Northeast Forestry University, 2022.
- [9] 黄焯. 基于 GA-BP 神经网络对原液着色纤维素色纱线的颜色预测 [D]. 天津: 天津工业大学, 2019.
Huang Ye. Color prediction of solution-dyed cellulose color yarn based on GA-BP neural network [D]. Tianjin: Tianjin Polytechnic University, 2019.
- [10] Chaoch S, Moussa A, Ben M I, et al. Application of genetic algorithm to color recipe formulation using reactive and direct dyestuffs

- mixtures[J]. *Color Research & Application*, 2020, 45(5): 896-910.
- [11] Zhan J, Zhang X, Wu J, et al. Dyeing recipe prediction of cotton fabric based on hyperspectral colour measurement and an improved recurrent neural network[J]. *Coloration Technology*, 2021, 137(2): 166-180.
- [12] Li F, Chen C, Mao Z. A novel approach for recipe prediction of fabric dyeing based on feature-weighted support vector regression and particle swarm optimisation [J]. *Coloration Technology*, 2022, 138(5): 495-508.
- [13] 李博文, 宋文广, 李浩源, 等. 基于遗传算法优化单类支持向量机的油田离心泵注水站异常检测[J]. *科学技术与工程*, 2023, 23(1): 283-289.
- Li Bowen, Song Wenguang, Li Haoyuan, et al. Anomaly detection for oilfield centrifugal pump water injection stations based on one-class support vector machine optimized by genetic algorithm [J]. *Science Technology and Engineering*, 2023, 23(1): 283-289.
- [14] Ferreira C. Gene expression programming: a new adaptive algorithm for solving problems[J]. *Complex Systems*, 2001, 13(2): 87-129.
- [15] Peng Y Z, Yuan C A, Qin X, et al. An improved gene expression programming approach for symbolic regression problems[J]. *Neurocomputing*, 2014, 137: 293-301.
- [16] Zhang H, Zhou A. RL-GEP: symbolic regression *via* gene expression programming and reinforcement learning [C]//2021 International Joint Conference on Neural Networks (IJCNN). New York: IEEE, 2021: 1-8.
- [17] Wan Y, Wagner N, Rondinelli J M. Symbolic regression in materials science[J]. *MRS Communications*, 2019, 9(3): 793-805.
- [18] Liu G, Li X, Xu H, et al. Preliminary application research of regression analysis in camouflage color matching [C]// Pacific Rim Laser Damage 2019: Optical Materials for High-Power Lasers. Washington: Society of Photo-Optical Instrumentation Engineers (SPIE), 2019: SPIE 11063.
- [19] Alav A H, Gandomi A H, Sahab M G, et al. Multi expression programming: a new approach to formulation of soil classification [J]. *Engineering with Computers*, 2010, 26: 111-118.
- [20] Awan H H, Hussain A, Javed M F, et al. Predicting marshall flow and marshall stability of asphalt pavements using multi expression programming[J]. *Buildings*, 2022, 12(3): 314.
- [21] Fallhpour A, Wong K Y, Rajoo S, et al. An evolutionary-based predictive soft computing model for the prediction of electricity consumption using multi expression programming [J]. *Journal of Cleaner Production*, 2021, 283: 125287.
- [22] Wag J, Sun X, Cheng Q, et al. An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting [J]. *Science of the Total Environment*, 2021, 762: 143099.
- [23] Laksmanaprabu S K, Shankar K, Ilayaraja M, et al. Random forest for big data classification in the internet of things using optimal features [J]. *International Journal of Machine Learning and Cybernetics*, 2019, 10: 2609-2618.
- [24] Iqba M F, Liu Q, Azim I, et al. Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming [J]. *Journal of Hazardous Materials*, 2020, 384: 121322.
- [25] Khn M A, Memon S A, Farooq F, et al. Compressive strength of fly-ash-based geopolymer concrete by gene expression programming and random forest [J]. *Advances in Civil Engineering*, 2021, 2021(3): 6618407.