

文章编号: 1004-4353(2022)04-0358-07

## 基于机器学习的马尾松松花粉 提取工艺研究

罗文莉, 刘赛虎, 金莉莉  
( 延边大学 药学院, 吉林 延吉 133002 )

**摘要:** 为提高马尾松松花粉提取率,对采用正交实验法和机器学习法提取马尾松松花粉的效果进行了对比研究. 研究显示:以马尾松松花粉浸膏为指标时,机器学习方法获得的马尾松松花粉浸膏的优化验证值为 21.06%,正交试验的验证值为 20.98%;以松花粉多糖为指标时,机器学习方法获得的多糖提取率的优化验证值为 2.74%,正交试验的验证值为 2.61%;以松花粉综合评分为指标时,机器学习方法获得的综合评分优化验证值为 92.01%,正交试验的验证值为 89.78%. 该结果表明,采用机器学习法提取马尾松松花粉浸膏和多糖的效果显著优于正交实验法,同时该方法可为同类提取工艺提供参考.

**关键词:** 机器学习; 松花粉; 提取工艺; 多糖; 综合评分

**中图分类号:** R931.6

**文献标识码:** A

## Research on optimization of *Pinus massoniana* Lamb extraction process based on machine learning

LUO Wenli, LIU Saihu, JIN Lili  
( College of Pharmacy, Yanbian University, Yanji 133002, China )

**Abstract:** To improve the extraction rate of *Pinus massoniana* Lamb, the orthogonal experiment method and machine learning method were adopted and compared. According to the results, when the *Pinus massoniana* Lamb extract was taken as the index, the optimized validation value of the *Pinus massoniana* Lamb extract obtained by the machine learning method reached 21.06%, and the validation value of the orthogonal test reached 20.98%. In addition, taking polysaccharide of the *Pinus massoniana* Lamb as an index, the optimized validation value of the extraction rate of polysaccharide obtained through the machine learning method was 2.74%, and the validation value of orthogonal test reached 2.61%. Furthermore, taking the comprehensive score of the *Pinus massoniana* Lamb as the index, the optimized verification value of the comprehensive score obtained by the machine learning method was 92.01%, and the validation value of the orthogonal test reached 89.78%, which shows that the machine learning optimization results are better than the orthogonal experiment results in extraction rate of extract and polysaccharides. At the same time, this method can provide reference for the extraction of similar processes.

**Keywords:** machine learning; *Pinus massoniana* Lamb; extraction technology; polysaccharides; comprehensive score

收稿日期: 2022-06-17

基金项目: 吉林省教育厅“十三五”科学技术项目(JJKH20191156KJ)

第一作者: 罗文莉(1998—),女,硕士生,研究方向为中药现代化.

通信作者: 金莉莉(1978—),女,博士,副教授,研究方向为中药现代化.

## 0 引言

松花粉是马尾松(*Pinus massoniana* Lamb)、油松(*Pinus tabulaeformis* Carr)、赤松(*Pinus densiflora* Siebet Zucc)、黑松(*Pinus thunbergii* Parl)等松科植物的花粉<sup>[1]</sup>.松花粉含有油脂、多糖、黄酮类、胆碱、甾醇等多种成分<sup>[2-5]</sup>,其中多糖是松花粉的主要活性物质之一.研究表明,松花粉具有抗炎、抗肿瘤、抗氧化、抗疲劳等作用<sup>[6-9]</sup>.目前,对松花粉的提取及相关研究仍停留在初级阶段<sup>[10]</sup>.2016年,袁桂香等利用正交试验对马尾松松花粉中的多糖进行了工艺优化筛选,但该方法仅仅能对已有的试验水平进行筛选<sup>[11]</sup>,并不能对工艺范围内的最优工艺进行全局求解.机器学习方法因能够有效解决非线性问题以及进行全局寻优,因此近年来被应用于提取工艺的优化中,并取得了良好效果<sup>[12]</sup>.基于此,本文利用正交试验和机器学习两种方法,以松花粉浸膏、多糖提取率和综合评分为指标,对松花粉提取工艺进行优化和对比,以期对松花粉提取工艺的择优选取提供参考.

## 1 试验材料与仪器

### 1.1 材料

马尾松松花粉(产于长白山),经延边大学药学院安仁波教授鉴定;无水乙醇、葡萄糖、苯酚均为分析纯(天津市科密欧化学试剂有限公司);浓硫酸(吉林省辽源市西安区化学试剂厂).

### 1.2 仪器

分析天平,上海精天电子仪器有限公司;多功能超纯水系统(Unique-R20),厦门锐思捷水纯化技术有限公司;WZ-180SP 旋转蒸发仪,上海申生科技有限公司;HH-W 型恒温水浴箱,江苏省常州市金坛区恒丰仪器厂;紫外分光光度计(UV-2450 型),岛津公司;鼓风干燥箱,金坛市恒丰仪器厂;L-500 台式低速自动平衡离心机,上海四科仪器设备有限公司.

## 2 试验方法

### 2.1 松花粉浸膏的制备

将 180 g 干燥马尾松松花粉等质量地分为 9 份后,分别置于 9 个烧瓶中.采用乙醇回流方法提

取马尾松松花粉,并将得到的提取液趁热过滤后再用旋转蒸发仪进行减压浓缩.浓缩后所得的膏体即为浸膏.

### 2.2 松花粉多糖的提取率测定

#### 2.2.1 对照溶液的配制

称取 20 mg 干燥的葡萄糖置于 250 mL 容量瓶中,加入超纯水将其定容.取葡萄糖对照溶液 2 mL 置于试管中,加入 2 mL 苯酚溶液(质量分数为 6%)和 5 mL 浓硫酸(体积分数为 98%)后将试管置于 40 ℃ 的恒温水浴锅中.40 min 后取出试管,室温下保存备用.

#### 2.2.2 供试溶液的配制

取 500 mg 松花粉浸膏置于 500 mL 容量瓶中,加入超纯水定容后超声、离心(3 000 r/min, 15 min).将上清液减压浓缩至 5 mL 后加入 15 mL 无水乙醇,放置 24 h(4 ℃)后离心 15 min(3 000 r/min).在离心得到的沉淀中加入少量蒸馏水,沉淀溶解后加入无水乙醇(加入量为溶液体积的 3 倍)使溶液再次出现沉淀.试验平行 3 次,并将沉淀物置于干燥箱中烘干(60 ℃, 60 min).取 5 mg 烘干品,研磨后加少量水溶解后转入 50 mL 容量瓶中定容.取 2 mL 样品液置于试管中,加入 1 mL 的苯酚(质量分数为 6%)和 5 mL 浓硫酸(体积分数为 98%),静置 20 min 后摇匀.将上述试管放入恒温水浴锅中水浴 15 min(40 ℃),取出后即得马尾松松花粉的多糖供试液<sup>[13]</sup>.

#### 2.2.3 标准曲线的绘制

在 250 mL 容量瓶中,用葡萄糖标准液和超纯水分别配制 8.0、12.0、16.0、20.0、24.0、28.0、32.0  $\mu\text{g/mL}$  的标准葡萄糖溶液<sup>[14]</sup>.取 1 mL 苯酚溶液(质量分数为 6%)置于 7 个试管中,然后加入 2 mL 配制的不同质量浓度的葡萄糖标准液;充分混匀后加入 5 mL 浓硫酸(体积分数为 98%),静置 10 min 后摇匀;在恒温水浴锅(40 ℃)中继续静置 15 min,然后在 490 nm 处测定 7 组试管的吸光度值<sup>[15]</sup>.测定平行 3 次,取平均值.利用 Origin 软件绘制标准曲线(以葡萄糖标准液的浓度为横坐标,以该浓度下测定的葡萄糖吸光度为纵坐标),如图 1 所示.由图 1 可知,葡萄糖标准液的浓度(X)和吸光度(Y)之间存在良好的线性关系: $Y = 0.0149X - 0.0656$ ,  $R^2 = 0.9962$ .

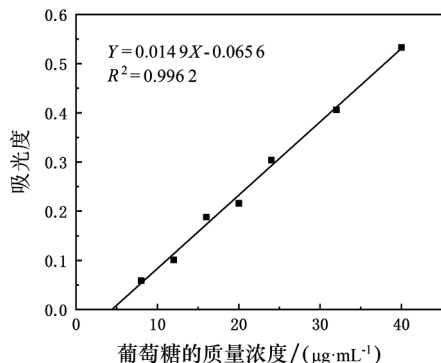


图 1 苯酚-硫酸法测定葡萄糖含量的标准曲线

#### 2.2.4 松花粉多糖的含量测定

取马尾松松花粉多糖供试品,平行测定 5 次吸光度值,并根据公式( $P = CV \times \frac{W_1}{W_2} \times 100\%$ )计算多糖提取率.公式中  $P$  为松花粉的多糖提取率,  $C$  为多糖的质量浓度( $\text{mg/mL}$ ),  $V$  为溶解体积( $\text{mL}$ ),  $W_1$  为浸膏质量( $\text{g}$ ),  $W_2$  为生药质量( $\text{g}$ ).

#### 2.2.5 加样回收率测定

取 3 份马尾松松花粉浸膏各 50 mg 置于试管中,并在浸膏中分别加入不同量的对照品(供试品已知量的 80%、100%、120%),然后在 490 nm 处测定溶液的吸光度和计算其 RSD 值.经计算,松花粉多糖的平均加样回收率为 98.41%,  $RSD = 0.949\%$ .

#### 2.3 工艺数据的处理

为同时考察浸膏和多糖的提取率,本文按公式( $X_{\text{nor}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$ )对数据进行归一化处理(因浸膏提取率与松花粉中多糖含量存在数量级差别),其中  $X_{\text{nor}}$  为归一化后的数据,  $X$  为原始数据,  $X_{\min}$  为数据最小值,  $X_{\max}$  为数据最大值.

本文根据文献[16]中的方法对松花粉的多目标提取工艺进行综合评分,评分公式为  $S = \gamma \times D + (1 - \gamma) \times P$ ,其中  $\gamma$  为松花粉浸膏的提取权重系数(取 0.4),  $1 - \gamma$  为松花粉多糖提取率的权重系数(取 0.6).

#### 2.4 正交试验的设计

基于文献[17],本文选取乙醇体积分数、提取温度和提取时间作为影响因素,每个因素均选取 3 个水平(如表 1 所示).正交试验  $L_9(3^3)$  利用 Minitab 软件设计.

表 1 正交试验的各因素水平

水平	乙醇体积分数/%	提取温度/℃	提取时间/h
1	70	65	1.5
2	80	75	2.5
3	90	85	3.0

#### 2.5 机器学习的优化算法

图 2 为利用机器学习方法对松花粉提取工艺进行优化的流程图.优化时,将正交试验所得的数据作为人工神经网络(artificial neural network, ANN)的训练数据集,并对训练后的数据进行皮尔逊相关性分析(Pearson correlation analysis),以此得到各个工艺参数的影响权重.为使神经网络模型能够快速收敛,对数据进行归一化处理,并在训练过程中采用  $K$  折交叉验证( $K$ -fold cross validation)的方式对模型进行评估,以此确保模型的泛化能力.

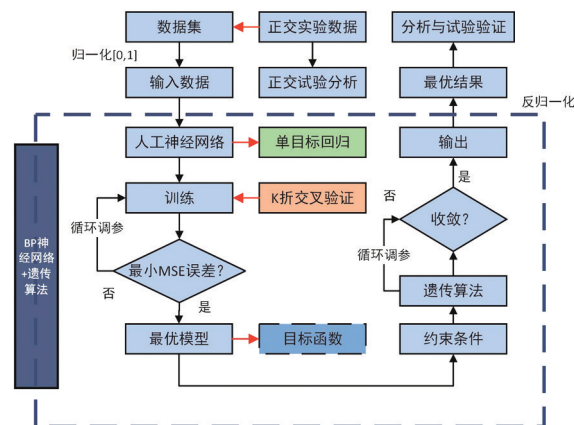


图 2 利用机器学习的松花粉提取工艺优化流程

利用 Python 语言中的 Pytorch 框架构建 BP (back propagation) 神经网络<sup>[18]</sup>,并利用该网络对马尾松松花粉正交试验提取的数据进行训练和预测.训练时,交叉验证的  $K$  折数设定为 6,遍历次数取 2000,优化器选用 Adam,学习率设定为 0.01.在不同目标下对松花粉浸膏提取、多糖提取和综合评分进行网络结构分析,所得的结果见表 2.

表 2 不同目标下的网络结构

工艺目标	第 1 层网络	第 2 层网络
浸膏提取	6(sigmoid)	6(sigmoid)
多糖提取	10(sigmoid)	8(sigmoid)
综合评分	8(tanh)	6(sigmoid)

由于遗传算法在优化数据时需给定目标函数和约束条件,因此本文根据浸膏提取、多糖提取及综合评分的正交试验结果给出如下约束条件: $60\% \leq \text{乙醇体积分数} \leq 90\%$ ,  $40\text{ }^{\circ}\text{C} \leq \text{提取温度} \leq 90\text{ }^{\circ}\text{C}$ ,  $0.75\text{ h} \leq \text{提取时间} \leq 3.2\text{ h}$ . 迭代数据,并将种群规模设置为 1 000,最大进化代数设置为 500,重组概率设置为 0.7.

3 结果与分析

3.1 正交试验方法的相关性分析

正交试验结果如表 3 所示. 由表 3 可以看出,影响浸膏提取、多糖提取及综合评分的因素其大小顺序依次为乙醇体积分数、提取时间、提取温度.

图 3 为浸膏、多糖提取率及综合评分的主效应图. 由图 3 可知:当乙醇体积分数为 90%、提取温度为 85  $^{\circ}\text{C}$ ,提取时间为 3 h 时,松花粉浸膏的提取率最高;当乙醇体积分数为 80%、提取温度

为 75  $^{\circ}\text{C}$ ,提取时间为 3 h 时,松花粉多糖的提取率最高;当乙醇体积分数为 80%、提取温度为 85  $^{\circ}\text{C}$ ,提取时间为 3 h 时,综合评分最高.

表 3 正交试验结果

试验序号	乙醇体积分数/%	提取温度/ $^{\circ}\text{C}$	提取时间/h	浸膏提取率/%	多糖提取率/%	综合评分/%
1	70	65	1.5	15.76	0.908	0
2	70	75	2.5	18.54	1.396	39.86
3	70	85	3	18.67	1.487	44.25
4	80	65	2.5	19.71	2.223	79.59
5	80	75	3	19.24	2.533	87.35
6	80	85	1.5	19.96	2.178	79.89
7	90	65	3	20.85	2.229	88.78
8	90	75	1.5	18.83	1.792	56.77
9	90	85	2.5	20.23	2.038	76.85
R(浸膏提取)		2.31	0.85	1.40		
R(多糖提取)		0.91	0.16	0.64		
R(综合评分)		0.54	0.11	0.28		

注: R 为正交试验的极差值(由 Minitab 软件计算得出).

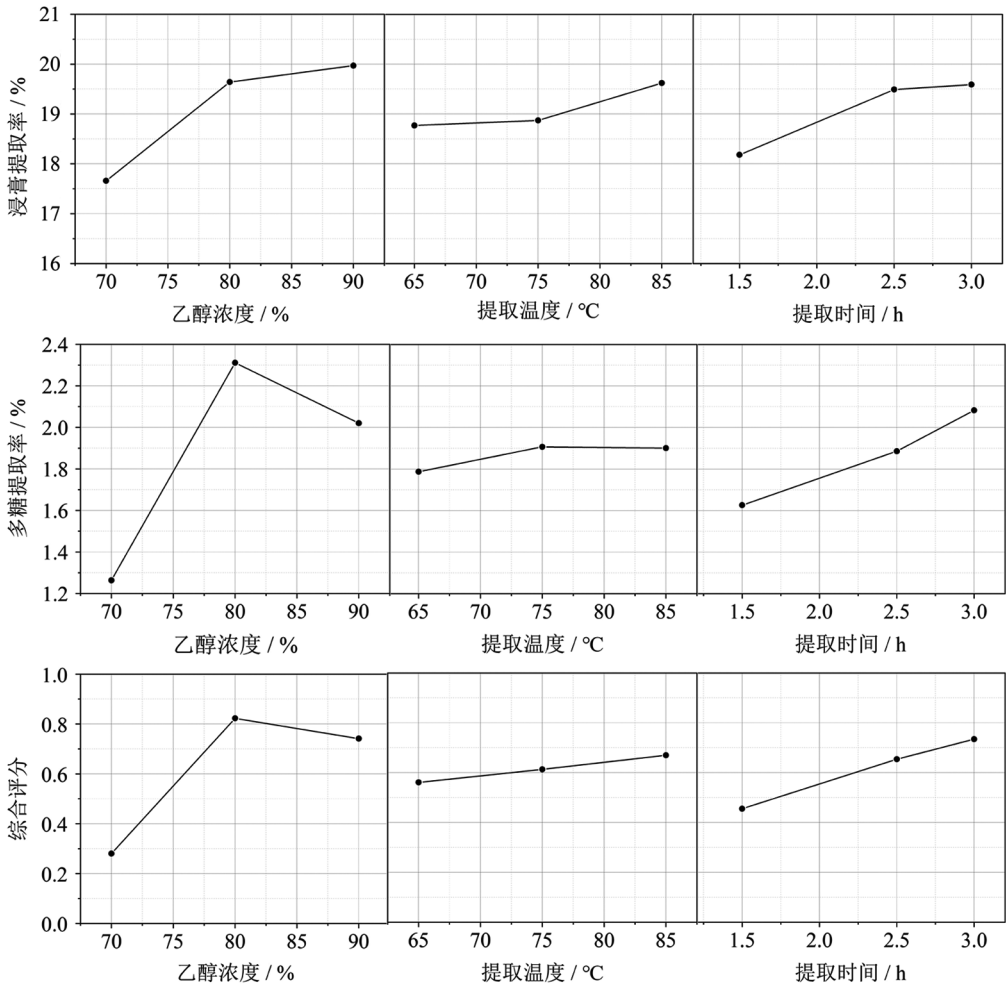


图 3 浸膏、多糖提取率及综合评分的主效应图



### 3.2 机器学习方法的预测结果

图 4 为松花粉提取工艺的皮尔逊相关性分析图. 由图 4 可知, 影响两种方法(正交试验法和机器学习方法)的权重(浸膏提取率、多糖提取率和综合评分的权重)其顺序均相同, 即由大到小的顺序均为乙醇浓度、提取时间、提取温度.

图 5 为基于 BP 网络的马尾松松花粉浸膏、多糖提取率和综合评分的响应图. 由图 5(a)和图 5(b)可以看出, 乙醇体积分数为 80% 时, 浸膏提取率随提取温度或提取时间的增加而增加; 由图 5(c)可以看出, 当提取温度为 75 °C 时, 浸膏提取率随提取时间的增加呈先上升后下降的趋势; 由图 5(d)和图 5(e)可以看出, 提取温度为 75 °C 或提取时间为 2.5 h 时, 多糖提取率随着乙醇体积分数的增加呈先上升后下降的趋势; 由图 5(f)可以看出, 提取温度为 75 °C 时, 多糖提取率随提取

时间的增加而增加; 由图 5(g)和图 5(h)可以看出, 提取温度为 75 °C 或提取时间为 2.5 h 时, 多糖提取率随乙醇体积分数的增加呈先上升后下降的趋势; 由图 5(i)可以看出, 提取温度为 75 °C 时, 综合评分随提取时间的增加而增加.

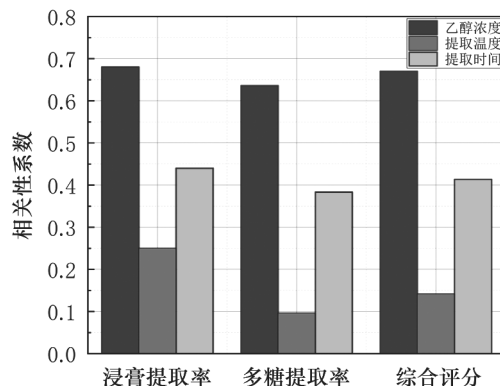


图 4 松花粉提取工艺的皮尔逊相关性分析图

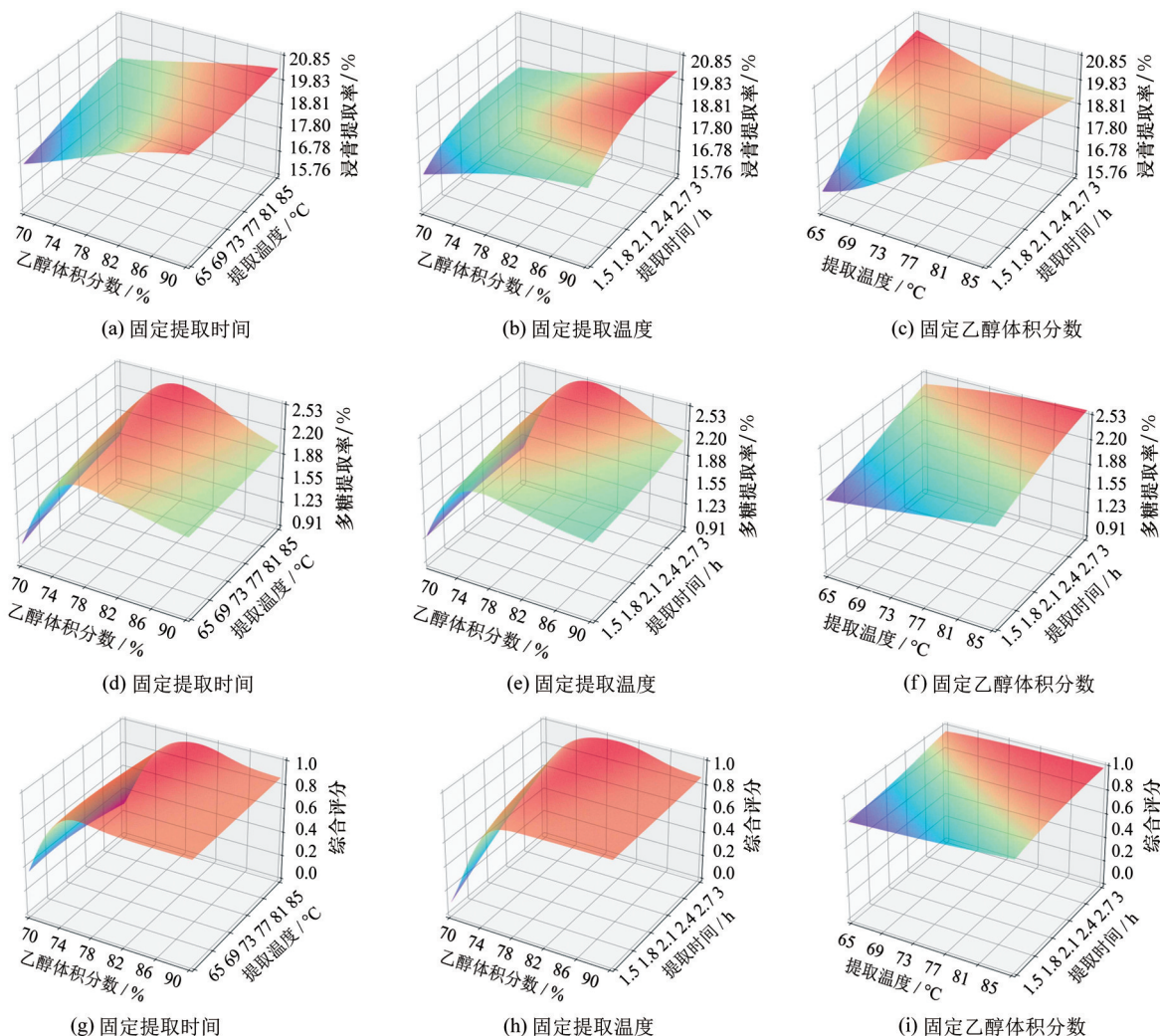


图 5 基于 BP 网络的马尾松松花粉浸膏、多糖提取率和综合评分的响应图

3.3 试验工艺验证

图 6 为种群个体目标函数值随代数变化的曲线。由图 6 可知,经过 20 次迭代后,遗传算法已找到最优工艺值。表 4 为正交试验和机器学习的优化结果及验证值。由表 4 可知,以松花粉浸膏为指标时,机器学习方法获得的松花粉浸膏的优化验证值为 21.06% (提取工艺为乙醇体积分数 90%、提取温度 65℃、提取时间 3.2 h),正交试验的验证值为 20.98%;以松花粉多糖为指标时,机器学

习方法获得的多糖提取率的优化验证值为 2.74% (提取工艺为乙醇体积分数 81%、提取温度 89℃、提取时间 3.2 h),正交试验的验证值为 2.61%;以松花粉综合评分为指标时,机器学习方法获得的综合评分优化验证值为 92.01% (提取工艺为乙醇体积分数 83%、提取温度 89℃、提取时间 3.2 h),正交试验的验证值为 89.78%。上述结果表明,机器学习法的提取率显著优于正交试验法,且所得的提取率相对稳定。

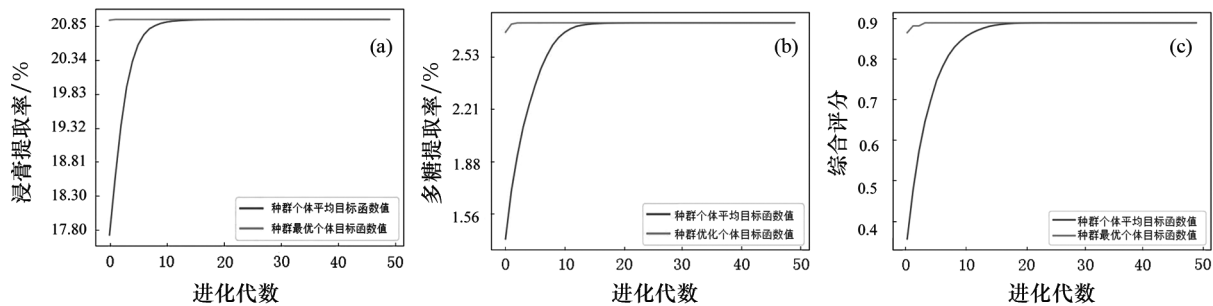


图 6 种群个体目标函数值随代数变化的曲线

表 4 正交试验和机器学习的优化结果及验证值

方法	考察工艺	乙醇体积分数/%	提取温度/℃	提取时间/h	试验验证值/%	优化预测值/%
正交试验	浸膏提取	90	85	3	20.98	-
	多糖提取	80	75	3	2.61	-
	综合评分	80	85	3	89.78	-
机器学习	浸膏提取	90	65	3.2	21.06	21.01
	多糖提取	81	89	3.2	2.74	2.65
	综合评分	83	89	3.2	92.01	88.96

4 结论

本文利用正交试验设计和机器学习两种方法,对马尾松松花粉提取工艺进行对比研究显示:以松花粉的浸膏、多糖提取率及综合评分为评价指标时,采用机器学习方法获得的松花粉浸膏、多糖提取率和综合评分显著优于正交试验设计方法,这表明机器学习可用于非线性的复杂问题的拟合,且使用该方法求解得到的极值通常为工艺的最优解。该方法同时具有良好的稳定性,并可降低试验费成本和时间,因此可为松花粉的工业化提取提供参考。在今后的研究中,我们将对马尾松松花粉多糖的生物活性进行研究,以促进马尾松松花粉的开发和利用。

参考文献:

[1] 于巧宁,仲米存,王锐莹,等. 马尾松和云南松花粉的特异性分子鉴定[J]. 食品科技,2020,45(8):56-60.

[2] 朱巧莎,侯占群,段盛林,等. 松花粉的主要活性成分及其功能的研究进展[J]. 食品研究与开发,2019,40(9):194-198.

[3] ZHOU S, SHANG H Q, MIAO Y Q, et al. Polysaccharides from *Pinus massoniana* pollen improve intestinal mucosal immunity in chickens[J]. Poultry Science, 2020,100(2):507-516.

[4] 刘其耸,石丽花,赵巧丽,等. 远红外处理延缓松花粉脂质的氧化作用[J]. 现代食品科技,2020,36(4):63-72.

[5] LIANG S B, LIANG N, BU F L, et al. The potential effects and use of Chinese herbal medicine pine pollen (*Pinus* pollen): abibliometric analysis of pharmacological and clinical studies[J]. World Journal of Traditional Chinese Medicine, 2020,2(6):163-170.

[6] WEI K, SUN Z H, YAN Z G, et al. Effects of Taishan *Pinus massoniana* pollen polysaccharide on immune response of rabbit haemorrhagic disease tissue inactivated vaccine and on production performance of rex rabbits[J]. Vaccine, 2011,29(14):2530-2536.

[7] CHENG Y, QUAN W, HE Y J, et al. Effects of

- postharvest irradiation and superfine grinding wall disruption treatment on the bioactive compounds, endogenous enzyme activities, and antioxidant properties of pine (*Pinus yunnanensis*) pollen during accelerated storage [J]. LWT-Food Science and Technology, 2021, 144:111249.
- [8] 梁海燕, 马平川, 朱文凯, 等. 泰山松花粉抗病毒作用的研究进展[J]. 山东化工, 2019, 48(8):78-79.
- [9] SU F C, SUN M M, GENG Y.  $^1\text{H}$ -NMR metabolomics analysis of the effects of sulfated polysaccharides from masson pine pollen in RAW264.7 macrophage cells[J]. Molecules, 2019, 24(9):1841.
- [10] 薛雨. 基于遗传算法和伴随方法的人居环境空气质量逆向问题研究[D]. 天津:天津大学, 2016.
- [11] 袁桂香, 罗小菊, 刘映良. 马尾松松花粉多糖提取工艺[J]. 林业工程学报, 2016, 1(4):80-84.
- [12] 尹涛, 何昭璇, 孙睿睿, 等. 机器学习在针灸研究领域的应用现状与展望[J]. 中国针灸, 2020, 40(12):1383-1386.
- [13] 郝晓华, 曹丽蓉, 罗淑政. 响应面优化超声波法提取薄荷多糖的研究[J]. 中国饲料, 2021(21):26-31.
- [14] 李瑞雪, 王钰婷, 夏家凤, 等. 桑黄菌丝体中多糖提取工艺优化及其体外抗氧化活性分析[J]. 中国农学通报, 2019, 35(29):143-150.
- [15] 王淑娟, 魏凯, 黄河, 等. 泰山松花粉多糖体外抗鸡传染性法氏囊病病毒的效果[J]. 中国兽医学报, 2018, 38(9):1670-1676.
- [16] 金小越, 帕提古丽·玉苏普, 王华洋, 等. 莲威阿那其处方药材水提工艺优化[J]. 中国药房, 2015, 26(25):3551-3553.
- [17] 卢奕, 马力. 响应面法优化松花粉总酚的超声提取工艺[J]. 中国实验方剂学杂志, 2012, 18(16):53-56.
- [18] 张锐, 顾志荣, 吕鑫, 等. 基于球面对称设计和 GA-BP 神经网络模型结合 AHP-CRITIC 优化清咽止痛方提取工艺[J]. 中药材, 2022(7):1692-1699.

~~~~~  
(上接第 326 页)

- [3] CHAMBOLLE A, POCK T. A first-order primal-dual algorithms for convex problem with applications to imaging [J]. Journal of Mathematical Imaging and Vision, 2011, 40:120-145.
- [4] GU G, HE B S, YUAN X M. Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: A unified approach[J]. Computational Optimization and Applications, 2014, 59:135-161.
- [5] HE B S, YUAN X M, ZHANG W X. A customized proximal point algorithm for convex minimization with linear constraints[J]. Computational Optimization and Applications, 2013, 56:559-572.
- [6] HE B S, YOU Y F, YUAN X M. On the convergence of primal-dual hybrid gradient algorithms[J]. SIAM Journal on Imaging Sciences, 2014, 7:2526-2537.
- [7] MA F, NI M F. A class of customized proximal point algorithms for linearly constrained convex optimization[J]. Computational and Applied Mathematics, 2018, 37:896-911.
- [8] ZHU Y, WU J, YU G H. A fast proximal point algorithm for  $l_1$ -minimization problem in compressed sensing[J]. Applied Mathematics and Computation, 2015, 270:777-784.
- [9] BAI J C. A new insight on augmented lagrangian method and its extensions[OL/J]. arXiv:2108.11125v2[2021-06-21]. <https://arxiv.org/abs/2108.11125v2>.
- [10] BONETTINI S, RUGGIERO V. On the convergence of primal-dual Hybrid Gradient algorithms for total variation image restoration[J]. Journal of Mathematical Imaging and Vision, 2012, 44:236-253.