

· 论著 · 方法学研究 ·

基于粗糙集理论的正交设计提取工艺因素方法的研究

崔红新^{1, 2}, 杨帅朋¹, 张泞杰¹, 王子翰¹, 黄琼¹, 赵迪^{1, 2}, 冯素香^{1, 2, 3}

1. 河南中医药大学药学院（郑州 450046）
2. 河南中医药大学呼吸疾病中医药防治省部共建协同创新中心（郑州 450046）
3. 河南中医药大学河南省现代中药研发与应用工程研究中心（郑州 450046）

【摘要】目的 研究药物制剂生产过程中因素之间的内在规律，探讨应用粗糙集理论寻求最优的工艺方案方法。**方法** 以经典名方四物汤为例，利用正交设计试验预处理数据，建立决策表，利用粗糙集规则模型进行分析。**结果** 当四物汤的提取时间、加水量和提取次数的因素水平逐渐增高时，多糖成分含量水平也呈递增趋势。当提取时间为 1.5 小时、加水量为 1:20、提取次数为 2 次时，试验结果最好，为四物汤优化提取工艺的三因素三水平最佳组合。**结论** 应用粗糙集理论可以探究药物制剂生产工艺因素之间以及因素与试验结果之间的相关关系，更全面地揭示其内在动态规律，并提供一种新颖的研究方法，进一步提高药物制剂的工艺质量。

【关键词】 粗糙集理论；正交设计；因素；决策规则模型

Study on extracting process factor method by orthogonal design based on rough set theory

Hong-Xin CUI^{1,2}, Shuai-Peng YANG¹, Ning-Jie ZHANG¹, Zi-Han WANG¹, Qiong HUANG¹, Di ZHAO^{1,2}, Su-Xiang FENG^{1,2,3}

1. School of Pharmacy, Henan University of Chinese Medicine, Zhengzhou 450046, China
 2. Collaborative Innovation Center for Chinese Medicine and Respiratory Diseases Co-constructed by Henan province & Education Ministry of P.R.China, Henan University of Chinese Medicine, Zhengzhou 450046, China
 3. Henan Engineering Research Center of Modern Chinese Medicine Research, Development and Application, Henan University of Chinese Medicine, Zhengzhou 450046, China
- Corresponding author: Su-Xiang FENG, Email: fengsx221@163.com

【Abstract】Objective To study the inherent law between factors in the pharmaceutical preparation production process, and explore the application of rough set theory to seek the optimal process plan method. **Methods** Take the classic recipe Siwu decoction as an example. Based on preprocessing data of the orthogonal design experiment, the decision table was established, and the rough set rule model was used to analyze. **Results** When the factor levels of the extraction time, the amount of water and the extraction times of Siwu decoction

DOI: 10.12173/j.issn.1004-4337.202310040

基金项目：河南省科技攻关项目（212102311095）

通信作者：冯素香，博士，教授，博士研究生导师，Email: fengsx221@163.com

were gradually increased, the content of polysaccharide showed an increasing trend. When the extraction time was 1.5 hours, the amount of water was 1:20 and the extraction times were 2, the experimental results were the best and the best combination of three factors and three levels to optimize the extraction process of Siwu decoction. Conclusion Rough set theory can be applied to explore the correlation among factors of pharmaceutical production process and the correlation between factors and test results, to reveal its inherent dynamic law more comprehensively and provide a novel research method to further improve the quality of pharmaceutical preparations.

【Keywords】Rough set theory; Orthogonal design; Factor; Decision rule model

在药物制剂生产工艺过程中,为了达到最佳工艺条件,技术人员常把因素对工艺的影响作为重要课题研究^[1-2]。传统研究因素间关系的常用分析方法包括正交设计、均匀设计及回归分析等,这些方法不仅需要统计的相关性判断、假设检验等,还需要进行大量的试验或足够大的样本量,同时传统统计方法的结果形式比较单一,无法动态分析因素之间、因素和试验结果指标间的相关关系^[3-7]。粗糙集理论是一种能有效分析不精确、不完整和不一致信息数据的数学方法,不需要提供所处理数据之外的其他先验知识,而是通过对数据进行分析和推理,挖掘隐含的关系,进而揭示因素之间动态变化的潜在规律。

粗糙集理论目前已广泛应用于数据挖掘、故障检测和工程控制等领域,但在医药领域中的应用较少^[8-12]。本研究以《中医药统计学》中“优化经典名方‘四物汤’提取工艺,提高药材的有效成分提取率”为例^[13],以该案例中正交设计试验预处理数据为数据来源,采用粗糙集理论决策方法,探讨四物汤提取生产工艺过程中因素之间的相互作用,提高有效成分提取率,寻求最佳组合搭配。

1 粗糙集理论简介

粗糙集理论由波兰数学家 Z.Pawlak 于 1982 年提出^[14]。该理论的基本思路是首先确立考察工

艺的因素,对数据资料离散化处理,建立信息表,继而形成条件属性与决策属性的决策表,进行决策表属性约简,利用粗糙集理论的决策分析系统,处理决策表的数据,得出条件属性和决策属性之间的决策规则模型。通过进一步分析决策规则模型,探寻各因素之间以及因素和结果指标之间的动态内在规律。

2 粗糙集模型的建立

2.1 模型因素的确定

本研究旨在探究四物汤提取生产工艺过程中提取时间、加水量和提取次数对试验指标药材多糖成分含量的影响,利用正交设计的正交表 L₉(3⁴)安排试验,进行预处理^[13],试验因素水平及试验结果分别见表 1 和表 2。

2.2 决策表及决策规则模型的建立

利用正交设计试验方案的数据作为论域,提取时间、加水量和提取次数作为条件属性,多糖成分含量作为决策属性,条件属性按其数值从小到大进行离散,决策属性按其数值从小到大进行离散,建立因素与试验指标多糖成分含量的相关关系决策表,见表 3。

建立决策表后,进行属性约简,建立决策规则模型: Rx: desc([X]C) → desc([X]D), 其中 desc([X]C) 表示因素提取时间、加水量和提取次数的特定取

表1 四物汤提取因素水平

Table 1. The extraction factor level of Siwu decoction

水平	因素		
	提取时间A(h)	加水量B(倍数)	提取次数C(次)
1	0.5	1 : 10	1
2	1.0	1 : 15	2
3	1.5	1 : 20	3

表2 四物汤提取因素正交试验结果
Table 2. Orthogonal test results of Siwu decoction extraction factors

试验号	列号				试验结果 多糖含量 (mg/g)
	A1	B2	C3	空白列4	
1	1 (0.5)	1 (1:10)	1 (1次)	1	13.71
2	1	2 (1:15)	2 (2次)	2	17.39
3	1	3 (1:20)	3 (3次)	3	17.65
4	2 (1.0)	1	2	3	25.07
5	2	2	3	1	24.95
6	2	3	1	2	19.03
7	3 (1.5)	1	3	2	25.43
8	3	2	1	3	19.24
9	3	3	2	1	25.56

表3 四物汤提取因素水平与多糖成分含量水平相关关系的决策表
Table 3. The decision table of the correlation between the extraction factor level and the content level of polysaccharide in Siwu decoction

组别	条件属性			决策属性
	提取时间A (h)	加水量B (倍数)	提取次数C (次)	
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	7
5	2	2	3	6
6	2	3	1	4
7	3	1	3	8
8	3	2	2	5
9	3	3	1	9

值, $\text{desc}([\text{X}]D)$ 表示试验指标药材多糖成分含量的特定取值。决策规则 $Rx: \text{desc}([\text{X}]C) \rightarrow \text{desc}([\text{X}]D)$ 表示因素提取时间、加水量和提取次数取不同水平时, 指标多糖成分含量水平的变化。通过此决策规则模型可以探究因素对多糖成分含量的影响, 以及因素之间相互作用的内在规律。

2.3 决策规则模型的分析

采用粗糙集软件 Rosetta 进行数据分析, 将因素水平与多糖成分含量水平相关关系的决策表载入 Rosetta 系统后, 软件基于粗糙集理论的 KDD 决策分析系统, 根据粗糙集理论对决策表数据进行处理、逐步分析, 得到决策规则结果。按照决策属性的离散原则, 决策属性水平越高表示药材多糖成分含量越高, 提取效果越好, 根据此规则

对四物汤提取工艺规则表模型进行分析。利用 Excel 软件绘制三维曲面图, 横轴分别表示提取时间、加水量或提取次数的水平, 纵轴表示多糖成分含量的水平, 曲面的不同颜色表示多糖成分含量的不同水平, 水平颜色从高到低依次为浅灰色、粉色、浅蓝色、绿色、深蓝、黄色、深灰色、深橘色、橘色。

3 结果

根据 Rosetta 软件的决策模型, 按照选取支持度均达到最小值 0.1 和置信度达到 1 的标准, 获得 Ga (genetic algorithm) 规则表, 见表 4。

由表 4 的决策规则 1~9 条可以看出, 提取时间水平为 1、加水量水平为 1~3 时, 多糖成分含

表4 四物汤提取工艺模型规则表

Table 4. Rule table of the extraction process model of Siwu decoction

序号	Ga Rule	LHS	RHS	RHS	LHS	RHS	RHS	LHS	RHS
		Support	Support	Accuracy	Coverage	Coverage	Stability	Length	Length
1	提取时间A(1) AND 加水量B(1) => 决策属性(1)	1	1	1.0	0.111 111	1.0	1.0	2	1
2	提取时间A(1) AND 加水量B(2) => 决策属性(2)	1	1	1.0	0.111 111	1.0	1.0	2	1
3	提取时间A(1) AND 加水量B(3) => 决策属性(3)	1	1	1.0	0.111 111	1.0	1.0	2	1
4	提取时间A(2) AND 加水量B(1) => 决策属性(7)	1	1	1.0	0.111 111	1.0	1.0	2	1
5	提取时间A(2) AND 加水量B(2) => 决策属性(6)	1	1	1.0	0.111 111	1.0	1.0	2	1
6	提取时间A(2) AND 加水量B(3) => 决策属性(4)	1	1	1.0	0.111 111	1.0	1.0	2	1
7	提取时间A(3) AND 加水量B(1) => 决策属性(8)	1	1	1.0	0.111 111	1.0	1.0	2	1
8	提取时间A(3) AND 加水量B(2) => 决策属性(5)	1	1	1.0	0.111 111	1.0	1.0	2	1
9	提取时间A(3) AND 加水量B(3) => 决策属性(9)	1	1	1.0	0.111 111	1.0	1.0	2	1
10	提取时间A(1) AND 提取次数C(1) => 决策属性(1)	1	1	1.0	0.111 111	1.0	1.0	2	1
11	提取时间A(1) AND 提取次数C(2) => 决策属性(2)	1	1	1.0	0.111 111	1.0	1.0	2	1
12	提取时间A(1) AND 提取次数C(3) => 决策属性(3)	1	1	1.0	0.111 111	1.0	1.0	2	1
13	提取时间A(2) AND 提取次数C(2) => 决策属性(7)	1	1	1.0	0.111 111	1.0	1.0	2	1
14	提取时间A(2) AND 提取次数C(3) => 决策属性(6)	1	1	1.0	0.111 111	1.0	1.0	2	1
15	提取时间A(2) AND 提取次数C(1) => 决策属性(4)	1	1	1.0	0.111 111	1.0	1.0	2	1
16	提取时间A(3) AND 提取次数C(3) => 决策属性(8)	1	1	1.0	0.111 111	1.0	1.0	2	1
17	提取时间A(3) AND 提取次数C(1) => 决策属性(5)	1	1	1.0	0.111 111	1.0	1.0	2	1
18	提取时间A(3) AND 提取次数C(2) => 决策属性(9)	1	1	1.0	0.111 111	1.0	1.0	2	1
19	加水量B(1) AND 提取次数C(1) => 决策属性(1)	1	1	1.0	0.111 111	1.0	1.0	2	1

续表4

序号	Ga Rule	LHS	RHS	RHS	LHS	RHS	RHS	LHS	RHS
		Support	Support	Accuracy	Coverage	Coverage	Stability	Length	Length
20	加水量B(2) AND 提取次 数C(2) => 决策属性(2)	1	1	1.0	0.111 111	1.0	1.0	2	1
21	加水量B(3) AND 提取次 数C(3) => 决策属性(3)	1	1	1.0	0.111 111	1.0	1.0	2	1
22	加水量B(1) AND 提取次 数C(2) => 决策属性(7)	1	1	1.0	0.111 111	1.0	1.0	2	1
23	加水量B(2) AND 提取次 数C(3) => 决策属性(6)	1	1	1.0	0.111 111	1.0	1.0	2	1
24	加水量B(3) AND 提取次 数C(1) => 决策属性(4)	1	1	1.0	0.111 111	1.0	1.0	2	1
25	加水量B(1) AND 提取次 数C(3) => 决策属性(8)	1	1	1.0	0.111 111	1.0	1.0	2	1
26	加水量B(2) AND 提取次 数C(1) => 决策属性(5)	1	1	1.0	0.111 111	1.0	1.0	2	1
27	加水量B(3) AND 提取次 数C(2) => 决策属性(9)	1	1	1.0	0.111 111	1.0	1.0	2	1

注: LHS: left hand side, 左边; RHS: right hand side, 右边; Support: 支持度; Accuracy: 精确度; Coverage: 覆盖度; Stability: 稳定度; Length: 长度。

量总水平平均值为 $6/3$; 提取时间水平为 2、加水量水平为 1~3 时, 多糖成分含量总水平平均值为 $17/3$; 提取时间水平为 3、加水量水平为 1~3 时, 多糖成分含量总水平平均值为 $22/3$ 。说明当提取时间水平、加水量水平增高时, 多糖成分含量水平呈递增趋势, 见图 1。

由表 4 的决策规则 10~18 条可以看出, 提取

时间水平为 1、提取次数水平为 1~3 时, 多糖成分含量总水平平均值为 $6/3$; 提取时间水平为 2、提取次数水平为 1~3 时, 多糖成分含量总水平平均值为 $17/3$; 提取时间水平为 3、提取次数水平为 1~3 时, 多糖成分含量总水平平均值为 $22/3$ 。说明当提取时间水平、提取次数水平增高时, 多糖成分含量水平呈递增趋势, 见图 2。

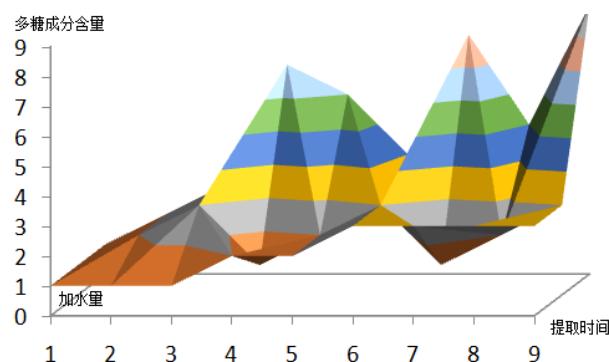


图1 提取时间、加水量与多糖成分含量水平的
三维曲面图

Figure 1. Three-dimensional curved surface diagram of extraction time, water content and polysaccharide content level

注: 横坐标1~9分别对应Ga规则表4的1~9条。

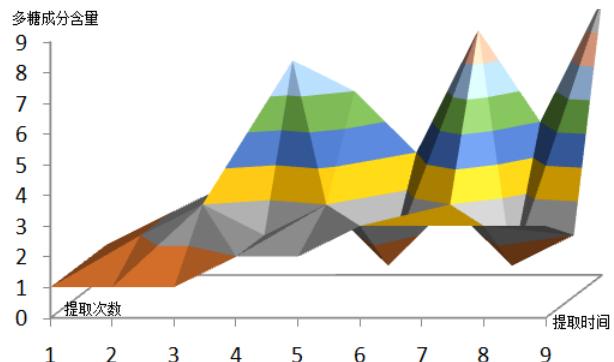


图2 提取时间、提取次数与多糖成分含量水平的
三维曲面图

Figure 2. Three-dimensional curved surface diagram of extraction time, extraction times and polysaccharide content level

注: 横坐标1~9分别对应Ga规则表4的10~18条。

由表4的决策规则19~27条可以看出,加水量水平为1、提取次数水平为1~3时,多糖成分含量总水平平均值为 $6/3$;加水量水平为2、提取次数水平为1~3时,多糖成分含量总水平平均值为 $17/3$;加水量水平为3、提取次数水平为1~3时,多糖成分含量总水平平均值为 $22/3$ 。说明当加水量水平、提取次数水平增高时,多糖成分含量水平呈递增趋势,见图3。

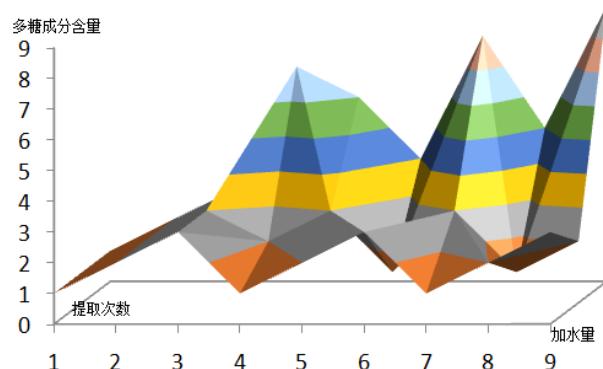


图3 加水量、提取次数与多糖成分含量水平的三维曲面图

Figure 3. Three-dimensional curved surface diagram of water addition, extraction times and polysaccharide content level

注: 横坐标1~9分别对应G₀规则表4的19~27条。

根据决策规则1~9可知,当提取时间为3水平、加水量为3水平时,多糖成分含量为9水平;由规则10~18可知,当提取时间为3水平、提取次数水平为3或2时,多糖成分含量水平为8或9水平;由规则19~27可知,加水量为3水平、提取次数水平为3或2时,多糖成分含量水平为8或9水平。综合平衡以上三因素与试验指标的内在关系,四物汤优化提取工艺的三因素三水平最佳搭配为提取时间为3水平、加水量为3水平、提取次数为2水平,即提取时间为1.5小时、加水量为1:20、提取次数为2次。

4 讨论

基于三维曲面图,结合决策规则中提取时间、加水量、提取次数水平与多糖成分含量水平的动态变化规律可以发现,当提取时间、加水量、提取次数水平均增高时,多糖成分含量水平呈递增趋势。

本研究结果显示,当提取时间为1.5小时、

加水量为1:20、提取次数为2次时,试验结果最好,此结果与参考资料中采用正交设计分析的结果一致^[13]。正交设计分析方法是利用方差分析的原理,多次对因素不同水平进行计算,得出最佳搭配,但该方法计算量大,且不能充分反映因素之间的内在联系。本研究采用粗糙集理论方法能够以动态变化形式,以结局指标与三因素不同水平变化影响的过程为研究重点,充分反映各因素之间以及因素和试验结果之间的相关关系,从更深层次揭示因素之间的内在规律。

本研究结果表明,粗糙集理论方法能够有效分析和推理不完整的数据信息,挖掘因素之间的相关关系,它不同于基于概率论为基础的数据挖掘,不需要进行相关性的系列假设检验。与模糊数学理论及证据理论方法的数据挖掘相比,粗糙集理论方法不需要提供所处理信息之外的任何预知知识。因此,该方法可作为医药科学领域解决实际问题的一种实用方法。

综上所述,本研究以四物汤为例,探究工艺因素之间以及因素与结果指标之间的相关关系,应用粗糙集理论决策规则模型的方法更全面地揭示事物之间的内在动态规律,并提供一种新颖的研究方法,从而进一步提高药物制剂的工艺质量。

参考文献

- 张爱丽,刘德军,苟琼心.银翘解毒丸制备工艺中绿原酸转移率的影响因素研究[J].药物评价研究,2016,39(1): 92–96. [Zhang AL, Liu DJ, Gou QX. Study on factors affecting the transfer rate of chlorogenic acid in process of preparing Yinqiaojiedu pills[J]. Drug Evaluation Research, 2016, 39(1): 92–96.] DOI: 10.7501/j.issn.1674-6376.2016.01.017.
- 王璐瑶,何淑蹇,冉锐敏,等.基于Plackett-Burman试验设计筛选大豆分离蛋白膜制备的主要工艺因素[J].中国油脂,2021,46(11): 73–77. [Wang LY, He SJ, Ran RM, et al. Screening of main process factors impacting the comprehensive performance of soy protein isolate film using Plackett-Burman design[J]. China Oils and Fats, 2021, 46(11): 73–77.] DOI: 10.19902/j.cnki.zgyz.1003-7969.210039.
- 袁雨冰,张登山.正交设计法优化富马酸喹硫平口服混悬液处方工艺[J].中国药物应用与监测,2023,20(1): 19–23. [Yuan YB, Zhang DS. Optimization of

- formulation process of quetiapine fumarate oral suspension by orthogonal test[J]. Chinese Journal of Drug Application and Monitoring, 2023, 20(1): 19–23.] DOI: 10.3969/j.issn.1672-8157.2023.01.005.
- 4 李军, 丁娟, 马志杰. 早期胃癌患者 ESD 术后迟发性出血危险因素的 Logistic 回归方程构建及预测价值 [J]. 实用癌症杂志, 2023, 38(5): 815–819. [Li J, Ding J, Ma ZJ. Logistic regression equation construction and predictive value of risk factors for delayed bleeding after ESD surgery in patients with early gastric cancer[J]. The Practical Journal of Cancer, 2023, 38(5): 815–819.] DOI: 10.3969/j.issn.1001-5930.2023.05.031.
- 5 杨传浩, 于钦辉, 杨勇, 等. 基于正交设计优化桂枝加葛根汤的提取工艺 [J]. 中药材, 2022, 45(7): 1696–1698. [Yang CH, Yu QH, Yang Y, et al. Orthogonal design was used to optimize the extraction process of cinnamon twig and kudzu root decoction[J]. Journal of Chinese Medicinal Materials, 2022, 45(7): 1696–1698.] DOI: 10.13863/j.issn1001-4454.2022.07.031.
- 6 刘会芳, 王振, 张学顺. 基于均匀设计法优化活血止痛凝胶膏剂处方工艺研究 [J]. 成都中医药大学学报, 2022, 45(3): 43–48. [Liu HF, Wang Z, Zhang XS. Optimization of formulation process of huoxue zhitong gel paste based on uniform design method[J]. Journal of Chengdu University of Traditional Chinese Medicine, 2022, 45(3): 43–48.] DOI: 10.13593/j.cnki.51-1501/r.2022.03.043.
- 7 陈家兰, 林芳花, 曹泽粼, 等. 羊栖菜多糖提取工艺的均匀设计法优选及含量测定 [J]. 时珍国医国药, 2021, 32(1): 54–57. [Chen JL, Lin FH, Cao ZL, et al. The sargassum fusiforme polysaccharides extraction process was optimized by uniform design method and the content was determined[J]. Lishizhen Medicine and Material Media Research, 2021, 32(1): 54–57.] DOI: 10.3969/j.issn.1008-0805.2021.01.14.
- 8 褚旭, 胡宗玉, 王珂清, 等. 粗糙集在烤烟化学质量综合评价中的应用 [J]. 中国烟草学报, 2021, 27(4): 20–27. [Chu X, Hu ZY, Wang KQ, et al. Application of rough set in comprehensive evaluation of flue-cured tobacco chemical quality[J]. Acta Tabacaria Sinica, 2021, 27(4): 20–27.] DOI: 10.16472/j.chinatobacco.2020.323.
- 9 赵婷, 王鹏莉, 张鲁玉, 等. 基于粗糙集理论的 MRI 设备扫描序列指标体系的优化 [J]. 郑州大学学报(医学版), 2020, 55(1): 92–96. [Zhao T, Wang PL, Zhang LY, et al. Optimization of scan sequence indicator system of MRI equipment based on rough set theory[J]. Journal of Zhengzhou University (Medical Sciences), 2020, 55(1): 92–96.] DOI: 10.13705/j.issn.1671-6825.2019.02.088.
- 10 侯文婷. 基于粗糙集理论的建筑工程成本分析 [J]. 现代电子技术, 2018, 41(19): 83–88. [Hou WT. Cost analysis of construction project based on rough set theory[J]. Modern Electronics Technique, 2018, 41(19): 83–88.] DOI: 10.16652/j.issn.1004-73x.2018.19.020.
- 11 周凯, 孙宏悦, 姜辉, 等. 基于粗糙集理论和 TOPSIS 的绿色包装评价方法 [J]. 包装工程, 2018, 39(17): 142–146. [Zhou K, Sun HY, Jiang H, et al. The evaluation method of green packaging based on rough set theory and TOPSIS[J]. Packaging Engineering, 2018, 39(17): 142–146.] DOI: 10.19554/j.cnki.1001-3563.2018.17.023.
- 12 马周明, 林梦雷, 张海洋. 《粗糙集理论及其应用》研究生课程教学改革与实践 [J]. 闽南师范大学学报(自然科学版), 2020, 33(1): 93–98. [Ma ZM, Lin ML, Zhang HY. Reform and practice of graduate education course of rough set theory and its application[J]. Journal of Minnan Normal University (Natural Science), 2020, 33(1): 93–98.] DOI: 10.16007/j.cnki.issn2095-7122.2020.01.016.
- 13 何雁. 中医药统计学 [M]. 北京: 中国中医药出版社, 2021: 147–150. [He Y. Statistics of traditional Chinese medicine[M]. Beijing: China Traditional Chinese Medicine Press, 2021: 147–150.]
- 14 刘秀峰, 陈振虎, 詹秀菊. 粗糙集理论在医学数据挖掘中的研究与应用 [J]. 计算机与现代化, 2011, 7: 176–178. [Liu XF, Chen ZH, Zhan XJ. Research and application of rough sets theory in medical data mining[J]. Computer and Modernization, 2011, 7: 176–178.] DOI: 10.3969/j.issn.1006-2475.2011.07.048.

收稿日期: 2023 年 10 月 10 日 修回日期: 2023 年 11 月 22 日

本文编辑: 张苗 黄笛

引用本文: 崔红新, 杨帅朋, 张泞杰, 等. 基于粗糙集理论的正交设计提取工艺因素方法的研究[J]. 数理医药学杂志, 2023, 36(12): 881–887. DOI: 10.12173/j.issn.1004-4337.202310040
 Cui HX, Yang SP, Zhang NJ, et al. Study on extracting process factor method by orthogonal design based on rough set theory[J]. Journal of Mathematical Medicine, 2023, 36(12): 881–887. DOI: 10.12173/j.issn.1004-4337.202310040