

基于信息熵理论的中药提取工艺优选

吴璐, 杨华生*

(江西中医学院, 南昌 330004)

[摘要] 目的: 探讨信息熵在中药提取工艺评价中的应用。方法: 以菝葜为模型药物, 以总黄酮、白藜芦醇、落新妇苷和黄杞苷含量为评价指标, 计算多指标的信息熵及权重系数, 确定综合评价指标。结果: 利用基于信息熵的数据处理方法处理正交试验数据, 筛选出影响菝葜提取效果的因素, 优选菝葜提取的最佳方案为加 10, 8 倍量水分别提取 2, 1.5 h。结论: 基于信息熵理论的数据处理方法可应用于中药提取工艺综合评价。

[关键词] 信息熵; 菝葜; 提取工艺; 综合评价

[中图分类号] R283.6 **[文献标识码]** A **[文章编号]** 1005-9903(2012)09-0029-03

Optimization of Extraction Technology of Traditional Chinese Medicine Based on Information Entropy Theory

WU Lu, YANG Hua-sheng*

(Jiangxi University of Traditional Chinese Medicine, Nanchang 330004, China)

[Abstract] **Objective:** To explore application of information entropy theory in evaluation of extraction technology for traditional Chinese medicine. **Method:** With *Smilax china* as a model drug, with the content of total flavonoids, resveratrol, astilbin and engeletin as indexes, comprehensive evaluation index was determined by calculating information entropy and weight coefficient of multi-indexes. **Result:** Orthogonal test data was handled by data processing method basing on information entropy, factors were selected which affected extraction effect of *S. china*, optimized optimum program of extraction technology for *S. china* was: added 10, 8 times the amount of water to extract 2, 1.5 h, respectively. **Conclusion:** Data processing method could be used in comprehensive evaluation of extraction technology for traditional Chinese medicine which basing on information entropy theory.

[Key words] information entropy; *Smilax china*; extraction technology; comprehensive evaluation

由于中药成分复杂, 指标成分多, 采用单一指标考察具有一定的片面性, 目前多根据研究基础或文献资料确定中药多个指标成分, 再采用主观权重算法或客观赋权法得到综合评价指标, 优化中药提取工艺^[1-2]。主观赋权法是基于研究者直接给出偏好

信息的方法, 受主体主观影响大; 客观赋权法是基于决策矩阵信息的方法, 经过对实际发生的资料进行整理、计算和分析, 从而得出的权重系数, 例如标准离差法、多目标最优化法、主成分分析法等。

熵(Entropy)的概念源于热力学, 是热能变化量除以温度所得的商, 目前已在工程技术、经济社会中广泛应用; 依据热力学中熵的概念, 把信号集的平均信息量称为信息熵^[3-4], 并利用概率统计方法给出了信息熵的定义 $H(X) = -k \sum p_i \ln p_i$ ^[5]。在信息论中, 信息熵 $H(X)$ 是系统不确定性的一个度量, 是系统无序性的度量。近年来, 信息熵在多属性决策问题中得到广泛的应用。本文提出了一种基于信息熵的中药提取工艺评价方法, 用熵值法计算指标权重, 将中药多项评价指标综合成单一度量指标, 用

[收稿日期] 20111011(003)

[基金项目] 江西省“十一五”重点学科青年教师培养计划项目(200327); 江西省高等学校教学改革研究课题(JXJG-10-9-21)

[第一作者] 吴璐, 硕士, Tel: 0791-87118628, E-mail: wulu456@126.com

[通讯作者] * 杨华生, 博士, 副教授, 从事中药新剂型与新技术研究, Tel: 0791-87118645, E-mail: yanghuasheng456@126.com

以客观评价提取工艺的新方法。

菝葜含有黄酮、白藜芦醇、皂苷及等多种成分。黄酮类化合物具有抗肿瘤、抗炎等生物活性^[8]；白藜芦醇具有抗炎抗肿瘤、调节脂肪代谢、抑制血小板聚集和保护心血管损伤等多种作用。故本研究采用水为提取溶媒，以总黄酮、白藜芦醇、落新妇苷和黄杞苷含量为评价指标，优选菝葜的提取工艺。

1 材料

752 型紫外-可见分光光度计(上海欣茂仪器有限公司), HP1200 型液相色谱仪(美国安捷伦), 菝葜购自湖南三湘饮片厂, 产地湖南, 经江西中医学院付小梅副教授鉴定为百合科植物菝葜 *Smilax china* L. 的根茎。芦丁、白藜芦醇对照品(中国药品生物制品检定所, 批号分别为 10080-20707, 111535-200502), 落新妇苷、黄杞苷对照品(自制, 纯度分别为 99.12%, 98.89%), 水为双蒸水, 其他试剂均为分析纯。

2 方法与结果

2.1 信息熵计算步骤 在一个评价指标系统中, 设有 m 个评价对象, n 个评价指标, 其形成的原始指标数据矩阵为 $Y = (y_{ij})_{mn}$ 。对于某项指标 i , 指标 y_{ij} 的差距越大, 则表明该项指标处于“无序”状态, 则该指标在评价中所起的作用越大; 如果某项指标值相等(可等价的认为 p_i 的值相等), 则说明该项指标处于“有序”状态, 指标的信息熵大, 该项指标在综合评价中不起作用。故试验可利用信息熵计算各指标的权重, 为中药提取多指标的综合评价提供依据^[6]。其计算步骤为:

- (1) 建立原始评价指标矩阵 $(X_{ij})_{mn}$;
- (2) 将原始数据阵 $(X_{ij})_{mn}$ 转为“指标概率”矩阵 $(P_{ij})_{mn}$;

在信息熵公式中 P_i 为某个信息的概率, 满足 $0 \leq p_i \leq 1$, 故必须对矩阵 $(X_{ij})_{mn}$ 做归一化处理, 处理后的矩阵可视为指标的“概率”矩阵。其中 P_{ij} 表示第 j 个试验在第 i 个评价指标下的“概率”。

$$P = \begin{pmatrix} P_{11} & P_{11} & \cdots & P_{1n} \\ P_{21} & P_{21} & \cdots & P_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ P_{m1} & P_{m2} & \cdots & P_{mn} \end{pmatrix} = (P_{ij})_{mn}$$

$$X = \begin{pmatrix} 157.77 & 192.96 & 144.42 & 275.15 & 166.67 & 158.17 & 195.79 & 191.34 & 121.76 \\ 2.01 & 2.83 & 1.73 & 3.332 & 0.76 & 1.32 & 5.00 & 2.52 & 0.94 \\ 32.18 & 11.58 & 42.09 & 71.23 & 28.15 & 5.83 & 27.99 & 19.97 & 28.17 \end{pmatrix}$$

计算 P_{ij} , 将原始评价矩阵转为“概率”矩阵。

其中 $P_{ij} = X_{ij} / (\sum_{j=1}^n X_{ij})$, P_{ij} 表示第 j 次实验在 i 指标下的概率。

(3) 计算属性输出的熵值, 确定第 i 个评价指标信息熵; 信息熵 $H_i = -k \sum_{j=1}^n P_{ij} \ln P_{ij}$, $k = 1/\ln n$ 。

(4) 计算第 i 项指标的权系数 $W_i = (1 - H_i / \sum_{i=1}^m (1 - H_i))$ 。

由上述公式可知, 当 H_i 越小, W_i 越大, 即当 X_{ij} 值相差越大, 表明该指标传递的信息量越多, 作用越大, 其权重值越大。

(5) 根据各指标的权重, 计算多属性综合评价指标, 对正交试验数据进行综合分析。

2.2 基于信息熵理论的菝葜提取工艺评价

2.2.1 菝葜药材提取工艺正交试验^[7-8] 称取菝葜 30 g, 按照正交试验因素水平表进行试验, 加入水回流提取 2 次, 滤过, 合并提取液, 定容至 500 mL。水提取正交试验选取加入溶剂倍数、提取时间、浸泡时间 3 个因素为考察因素, 各因素的水平设计见表 1。各供试样品中总黄酮提取量(FVN)、白藜芦醇含量(RST)、落新妇苷和黄杞苷提取量(ASAEN), 测定结果见表 2。

表 1 菝葜药材提取工艺正交试验因素水平

水平	A	B	C
	浸泡时间/h	加溶剂量/倍	提取时间/h
1	1.0	10,8	2.0,1.5
2	0.5	8,6	1.5,1.0
3	0	6,4	1.0,0.5

表 2 菝葜药材提取工艺正交试验含量测定 mg

No.	FVN	RST	ASAEN
1	157.77	2.01	32.18
2	192.96	2.83	11.58
3	144.42	1.73	42.09
4	275.15	3.332	71.23
5	166.67	0.76	28.15
6	158.17	1.32	5.83
7	195.79	5.00	27.99
8	191.34	2.52	19.97
9	121.76	0.94	28.17

2.2.2 基于信息熵的菝葜提取工艺研究的数据处理 按照前面所给的步骤, 建立原始评价指标矩阵, 原始评价矩阵如下:

$$P = \begin{vmatrix} 0.0984 & 0.1203 & 0.0900 & 0.1715 & 0.1039 & 0.0986 & 0.1221 & 0.1193 & 0.0759 \\ 0.0983 & 0.1384 & 0.0846 & 0.1630 & 0.0372 & 0.0646 & 0.2446 & 0.1233 & 0.0460 \\ 0.1204 & 0.0433 & 0.1575 & 0.2666 & 0.1054 & 0.0218 & 0.1048 & 0.0747 & 0.1054 \end{vmatrix}$$

计算每项指标的信息熵,得评价指标的信息熵。

$$H_i = |0.9884 \quad 0.9329 \quad 0.9204|$$

计算第 i 项指标的权系数 W_i 。

$$W_i = |0.0731 \quad 0.4240 \quad 0.5028|$$

从 9 次试验可知,落新妇苷、黄杞苷提取量的信息熵最小,在不同试验条件下,落新妇苷、黄杞苷提取量变化显著,表明在菝葜提取工艺研究中,落新妇苷、黄杞苷提取量是要重点考察的指标,总黄酮提取量信息熵最大,说明在不同的浸泡时间、不同的溶媒量及不同提取时间,总黄酮提取量数值变化不大。

对概率矩阵的数据进行加权处理,得综合评价指标 M ,并进行方差分析,确定最优实验方案。正交试验结果见表 3,方差分析见表 4。

表 3 菝葜药材提取工艺正交试验概率矩阵
数据加权处理综合评价

No.	A	B	C	D	M
1	1	1	1	1	0.1094
2	1	2	2	2	0.0893
3	1	3	3	3	0.1217
4	2	1	2	3	0.2157
5	2	2	3	1	0.0763
6	2	3	1	2	0.0456
7	3	1	3	2	0.1653
8	3	2	1	3	0.0986
9	3	3	2	1	0.0781
K_1	0.3204	0.4904	0.2536	0.2638	
K_2	0.3376	0.2642	0.3831	0.3002	
K_3	0.3420	0.2454	0.3633	0.4360	
R	0.0216	0.2450	0.1295	0.1722	

表 4 综合评价方差分析

方差来源	SS	f	F	P
A	0.0001	2	0.0158	>0.05
B	0.0062	2	2.2571	>0.05
C	0.0016	2	0.5908	>0.05
D(误差)	0.0027	2		

注: $F_{0.05}(2,2) = 19.00, F_{0.01}(2,2) = 99.00$ 。

由极差分析结果可知,各因素对菝葜综合指标

的影响主次顺序为 $B > C > A$,即溶媒用量 > 提取时间 > 浸泡时间,方差分析各因素水平间均无显著性影响,故选 $A_3B_1C_2$ 为菝葜药材的提取的最佳方案。

2.3 验证试验 为确证该工艺的优劣和稳定性,按最佳条件提取菝葜 3 批进行验证试验,测得提取物中各指标成分含量,计算综合评分, M 平均值为 0.2279, RSD 1.85%,表明基于信息熵的中药提取数据处理优化的试验方案稳定可行。

3 讨论

信息熵近年来逐渐被引入生物医学领域,因其在处理复杂数据方面的优势而备受关注。本研究在正交试验的基础上,引入信息熵对试验数据进行分析,探索影响中药提取效果的主要条件对评价指标信息熵的影响;在实际研究工作中,可考察评价指标在单因素试验中信息熵的变化,计算熵权,对于在不同条件下信息熵没有明显变化的指标,可以舍弃,不仅可节省人力、物力,还可使实验设计更加科学合理。该方法有其合理性和应用价值,在实际应用时,可编制相应的计算软件,使计算简单易行。

[参考文献]

[1] 付克,张坤,闫广利.多成分评价优化柴芩清肝方提取工艺研究[J].中国实验方剂学杂志,2010,16(7):10.

[2] 杨华生,张坤,尹小英,等.多指标正交试验优选罗布麻定时脉冲片提取工艺[J].中国实验方剂学杂志,2010,16(12):14.

[3] 贾中裕.经济与管理数学模型[M].北京:冶金工业出版社,2000:146.

[4] 张东,张宁.物理学中的熵理论及其应用研究[J].北京联合大学学报:自然科学版,2007(3):4.

[5] 邱苑华.管理决策与应用熵学[M].北京:机械工业出版社,2001:135.

[6] 张少艳.信息熵在教学质量分析中的应用[J].红河学院学报,2007,5(2):78.

[7] 肖晶,魏峰,林瑞超.菝葜属植物的研究进展[J].中国药学杂志,2001,36(6):370.

[8] 王学江,文朝阳,丰平,等.复方菝葜的抑瘤机理[J].中国病理生理杂志,1996,12(4):614.

[责任编辑 全燕]