

基于融合组合赋权和最优聚类机器学习方法的五味子保胃护肝保健食品配方设计与评价

姚鉴玲¹, 刘洪宇², 程杰³, 马嘉慕¹, 宋若兰¹, 于啊香¹, 魏静¹, 魏胜利¹, 单东杰¹, 折改梅^{1*}

1. 北京中医药大学中药学院, 北京 100029

2. 国家市场监督管理总局 食品审评中心, 北京 100070

3. 东阿阿胶股份有限公司 国家胶类中药工程技术研究中心, 山东 聊城 252200

摘要: 目的 提出一种融合组合赋权、聚类、决策、评价等多种机器学习算法用于保健食品配方设计及评价的方法。方法以五味子 *Schisandre Chinensis Fructus* 为例, 构建含五味子处方数据库, 筛选与五味子配伍高频药味并挖掘其关联规则。同时结合传统中医药理论和现代科学研究建立评价指标体系, 采用兼顾主观性和客观性的层次分析 (analytic hierarchy process, AHP) -CRITIC (criteria importance though intercriteria correlation) 组合赋权法和优选的聚类算法对高频药味进行加权和聚类, 结合中医药理论设计配方并进行逼近理想解排序 (technique for order preference by similarity to an ideal solution, TOPSIS) 综合评价。结果 频次统计得到黄芪、茯苓、人参等与五味子配伍高频药味 31 个, 关联规则分析显示高频药味间更易产生强关联。构建含 3 个一级指标、7 个二级指标的评价指标体系, AHP-CRITIC 组合赋权法计算出的指标组合权重从大到小依次为药味传统功效、现代文献研究、在数据库中的出现频次, 符合主观认识和客观数据。根据与专业知识的匹配度、算法运行效率及对数据的包容度优选模糊 C 均值 (fuzzy C-means, FCM) 聚类, 将高频药味分为 5 类。结合中医药理论及上述结果设计可能的新配方共 11 个, TOPSIS 综合评价排序的结果显示, 五味子-黄芪-白术-党参是五味子保胃护肝保健食品可能的最优新配方。结论 该模型在中医药理论的指导下, 提供了既能体现传统中医药配伍理论又有足够现代科学研究成果支撑的中药类保健食品配方设计与研发的创新思路与方法。

关键词: 保健食品; 五味子; 组合赋权; AHP-CRITIC; K-均值; 自组织神经网络; 模糊 C 均值聚类; 配方设计; 综合评价
中图分类号: R283.21; TS218; TP312 文献标志码: A 文章编号: 0253-2670(2022)14-4437-10

DOI: 10.7501/j.issn.0253-2670.2022.14.023

Formula design and evaluation of *Schisandre Chinensis Fructus* health food for protecting gastric mucosa and liver based on combination of empowerment and clustering optimization machine learning algorithms

YAO Jian-ling¹, LIU Hong-yu², CHENG Jie³, MA Jia-mu¹, SONG Ruo-lan¹, YU A-xiang¹, WEI Jing¹, WEI Sheng-li¹, SHAN Dong-jie¹, SHE Gai-mei¹

1. School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing 100029, China

2. Center for Food Evaluation, State Administration for Market Regulation, Beijing 100070, China

3. National Engineering Technology Research Center for Gelatin-based Traditional Chinese Medicine, Dong'e Ejiao Co., Ltd., Liaocheng 252200, China

Abstract: Objective A formula design and evaluation model was proposed, which integrated multiple machine learning algorithms such as combination of empowerment, clustering, decision making and evaluation. **Methods** Taking Wuweizi (*Schisandre Chinensis Fructus*) as an example, constructing the prescription database containing *Schisandre Chinensis Fructus*, screening high-frequency function ingredients could be used in health food and mining its association rules. Meanwhile, a comprehensive and objective

收稿日期: 2022-01-20

基金项目: 河南省重大科技专项 (211110310100)

作者简介: 姚鉴玲 (2000—), 女, 硕士研究生, 研究方向为中药化学。E-mail: yaojianling00@126.com

*通信作者: 折改梅 (1976—), 博士生导师, 研究员, 研究方向为中 (民族) 药药效成分和新药创制研究。

Tel: (010)53912129 E-mail: shegaimei@126.com

evaluation index system was established, which combined theory of traditional Chinese medicine (TCM) and modern scientific research results. The analytic hierarchy process (AHP)-criteria importance though intercriteria correlation (CRITIC) combination empowerment was adopted, which considered both subjectivity and objectivity. K-means, self-organizing map (SOM) and fuzzy C-means (FCM) were optimized, high-frequency function ingredients were weighted and clustered by the approaches above. Then combined with TCM theory, the formulas were designed and conducted by technique for order preference by similarity to solution (TOPSIS) comprehensive evaluation. **Results** 31 Kinds of high-frequency function ingredients containing Huangqi (*Astragali Radix*), Fuling (*Poria*) and Renshen (*Ginseng Radix et Rhizoma*) were obtained by frequency statistics. The results of association rule analysis showed that strong association rules were more likely to be generated between high-frequency function ingredients. An evaluation index system containing three first-level indicators and seven second-level indicators was constructed. The combination weight of the index calculated by AHP-CRITIC combination empowerment was the traditional efficacy of medicine, modern literature research and the frequency of occurrence in the database from large to small, it is consistent with subjective knowledge and objective data. Fuzzy C-means (FCM) clustering was finally selected as the clustering algorithm in this paper according to the degree of matching with professional knowledge, the efficiency of algorithm operation and the degree of data tolerance, high-frequency function ingredients were divided into five categories. Combined with the TCM theory and results above, a total of 11 possible new formulas were designed. The results of TOPSIS method showed that *Schisandre Chinensis Fructus-Astragali Radix-Baizhu (Atractylodis Macrocephalae Rhizoma)-Dangshen (Codonopsis Radix)* was the best possible new formula of *Schisandre Chinensis Fructus* health food for protecting gastric mucosa and liver. **Conclusion** Under the guidance of the theory of TCM, this model provided innovative ideas and methods for the formula design, research and development of TCM health food, which can embody the compatibility theory of TCM and modern scientific research results.

Key words: health food; *Schisandre Chinensis Fructus*; combination of empowerment; AHP-CRITIC; K-means; self-organizing maps (SOM); fuzzy C-means (FCM); formula design; comprehensive evaluation

保健食品是声称具有特定保健功能或者以补充维生素、矿物质为目的的食品^[1]。中药类保健食品是大健康产业的主要组成部分，我国的保健食品中80%是功能保健食品，而中药类保健食品占功能保健食品的70%以上^[2]。中医养生保健理论丰富，其治未病、预防、养生、保健等理念跟保健食品定位不谋而合^[3]，中药类保健食品因其深厚的中医药理论支撑和辨证保健的理念而备受消费者青睐。但研究显示，受现有保健食品原料管理目录的限制，目前市场上的中药类保健食品配方原料组合多为简单拼凑，并未体现传统中医药的配伍思想^[4]。更重要的是保健食品在研发前期——配方设计阶段不能充分利用现有原料药功效，从而多方面制约了保健食品发展。因此，如何在有限的保健食品原料目录中，创新地设计出既能体现传统中医药配伍理论又有足够现代科学研究成果支撑的配方是中药类保健食品目前面临的迫切需要解决的问题。

数据挖掘指在大型数据存储库中自动发现有信息的过程，是传统数据分析方法与数据处理复杂算法的有机融合，包括关联规则分析、聚类分析、决策及评价等系列算法，在工业、金融、医学等领域发挥着重要的作用。有学者以保健食品可声称的功能为出发点，从已公布的保健食品配方、中成药、

方剂处方中挖掘新配方；研究方法多为借助相关平台，通过关联规则分析和聚类分析得到核心药味及可能的新组方^[5]。这种功能导向的配方设计思路有一定新颖性，方法易于操作，但存在所选算法与分析对象存在不适配的现象^[6-7]。课题组前期建立了以白及为主要原料，根据构建数据库的特点和机器学习算法原理，采用层次分析（analytic hierarchy process, AHP）-自组织神经网络（self-organizing maps, SOM）聚类-逼近理想解排序（technique for order preference by similarity to an ideal solution, TOPSIS）等深度学习进行了保肝护胃功能保健食品配方设计和评价^[8]。该方法对现有方剂、中成药、保健食品组方等数据进行深度挖掘，并对可能的新配方进行了综合评价，为保健食品的配方设计提供了创新思路。但是该研究方法未能把原料药药效学等现代科学研究成果有效的融合到保健食品配方设计中。

肝胃同治是传统中医药理论中治疗肝、胃病的一个重要思想。清代名士叶天士在《临证指南医案》中提出“肝为起病之源，胃为传病之所”。肝胃同治思想在现代临床中也得到了广泛的应用，通过疏肝理气、和胃降逆、益气健脾等治法，调畅气机，达到肝胃同治的效果^[9-10]。保健食品中对化学性肝损

伤和胃黏膜的辅助保护功能是肝、胃处于亚健康状态时的辅助治疗手段,分别对应了中医病症中的“胁痛”和“胃脘痛”,符合中医肝胃同治的基本要求。五味子为木兰科植物五味子 *Schisandra chinensis* (Turcz.) Baill. 的干燥成熟果实,味酸、甘,性温,归肺、心、肾经;具有收敛固涩、益气生津、补肾宁心的功能^[11]。现代药理研究及临床应用结果显示,五味子具有极好的保肝护肝作用,同时对胃黏膜也有一定的保护作用^[12]。

基于此,本研究以五味子为例,以传统中医药理论中的肝胃同治思想为指导,建立现有保健食品及中药相关的组方数据库,挖掘核心组合。同时结合传统中医药理论和现代科学研究成果建立全面、客观的评价指标体系,采用兼顾主观性和客观性的组合赋权法对评价指标进行赋权。最后结合最优聚类算法的筛选结果,设计可能的新配方并进行综合评价,为中药类保健食品配方的设计和研发提供新的思路和方法。本研究技术路线见图1。

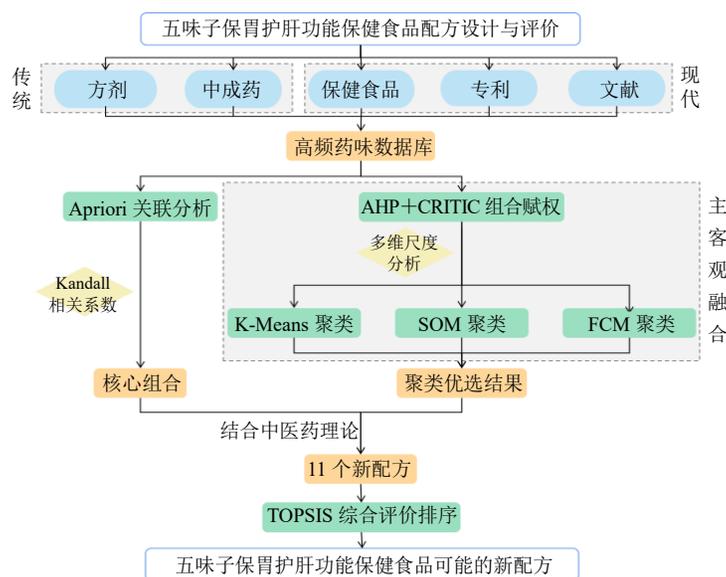


图1 五味子保胃护肝保健食品配方设计与评价技术路线

Fig. 1 Technical route of formula design and evaluation of *Schisandra Chinensis Fructus* health food for protecting gastric mucosa and liver

1 资料

1.1 数据来源、纳入及处理

数据来源包括国家市场监督管理总局 (State Administration for Market Regulation, SAMR) 特殊食品安全监督管理司网站、药智数据网 (<https://db.yaozh.com/>)、中国知网 (CNKI) 和万方数据库。检索收集组方中含五味子, 主治肝、胃疾病, 具有和胃、健脾益气、疏肝理气功能的相关保健食品批文、方剂和中成药处方及相关专利和文献。

按下列标准对收集的数据进行筛选剔除: (1) 剔除已注销、相同批准文号、配方为单味药的保健食品批文数据; (2) 剔除名称、剂型、药味组成均相同的方剂和中成药处方数据; (3) 剔除相同申请号的专利数据; (4) 相同题目及作者的同一篇文章仅录入1次。文献处方数据录入的标准: 若基础方中含有五味子, 则不同加减方分开录入; 若基础方中不含五味子, 则

仅录入加减方中含有五味子的处方。筛选后共得到相关保健食品批文89条, 方剂130首, 中成药处方181条, 相关专利166条, 文献71篇。

按以下原则进行数据清洗及规范: (1) 剔除保健食品配方中来源不明的提取物、化学药品、营养素补充剂、药用辅料及食品添加剂等; (2) 药味名称和分类参考《中国药典》2020年版^[11]和《临床中药学》第2版^[13]进行规范。涉及中药别名 (如花旗参等)、道地药材 (如川牛膝等) 或中药炮制方法 (如炒白术等), 若功效影响较小者, 则规范为同种中药名 (如麦芽与炒麦芽); 若功效影响显著者, 则分别录入 (如菊花与野菊花)。

1.2 数据库构建

将经清洗和规范后的保健食品批文、方剂、中成药、专利、文献处方数据整合为一个数据库, 使用 Microsoft Excel 2019 对数据库中药味出现频次、

性味归经及功能分类进行统计分析。依据保健食品原料目录的要求,筛选可用于保健食品的高频药味,得到高频药味原始数据库。

2 方法

2.1 关联规则分析

采用 IBM SPSS Modeler 18.0 软件中的 Apriori 模型,对高频药味进行关联规则及网络分析。

2.2 AHP-CRITIC (criteria importance though intercriteria correlation) 组合赋权

2.2.1 评价指标体系构建 中药类保健食品配方的原料药味,首先应当在保健食品原料目录中;其次,应依据拟研发保健食品的功能,根据传统中医药理论中药的性味归经、功能主治筛选。此外,随着中药物质组成和药效机制研究的深入,应综合各备选药味的现代研究结果进行筛选。评价指标体系应满足全面性、合理性、独立性的原则^[14],结合传统中医药理论及现代科学研究,设置 3 个一级指标,即药味传统功效、现代文献研究及在数据库中的出现频次;7 个二级指标,即功能分类、四气、五味、归经、单味药药理作用研究、含某药味复方研究、综述及其他。最终确定评价指标体系如表 1 所示。

表 1 评价指标体系
Table 1 Evaluation index system

一级指标	二级指标
药味传统功效 (A ₁)	功能分类 (B ₁)
	四气 (B ₂)
	五味 (B ₃)
	归经 (B ₄)
现代文献研究 (A ₂)	单味药药理作用研究 (B ₅)
	含某药味复方研究 (B ₆)
	综述及其他 (B ₇)
在数据库中的出现频次 (A ₃)	

2.2.2 AHP-CRITIC 组合赋权确定评价指标权重 主观权重的确定采用 AHP 法。根据“保胃护肝双功能”和“肝胃同治”两核心目标,基于传统中医药理论,确定各评价指标之间的相对重要关系,按 1~9 标度法对各评价指标打分赋值,构造各指标层的判断矩阵。其中 1 表示同等重要,3 表示稍微重要,5 表示明显重要,7 表示强烈重要,9 表示极端重要,2、4、6、8 为上述相邻判断的中值^[15]。对比矩阵结果通过一致性检验后,计算各一级指标主观权重系数。

客观权重的确定采用 CRITIC 法。CRITIC 法是基于评价指标的对比强度和指标之间的冲突性来综合确定指标客观权重的方法,具有同时兼顾指标变异性大小和指标之间相关性的优势^[16]。其具体计算方法如下。

(1) 评价指标量化赋值:由于评价指标体系中药味传统功效下各二级指标的原始数据为定性指标,故结合药味的性味归经及功能分类的统计分析结果,按照排序(前 5)从高到低分别赋值 5、4、3、2、1。如补虚药在统计中出现频次最高,则高频药味功能分类中凡属补虚药者,在功能分类(B₁)这一指标上均可打 5 分,构造高频药味原始矩阵。

(2) 客观权重系数计算:假设有 i 个待评价样本, j 项评价指标,则原始矩阵为:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1j} \\ \vdots & \ddots & \vdots \\ x_{i1} & \cdots & x_{ij} \end{pmatrix} \quad (1)$$

其中, x_{ij} 表示第 i 个样本第 j 项评价指标的数值。为消除因量纲不同对评价结果的影响,对原始矩阵进行无量纲化处理形成标准化矩阵。由于所选评价指标均为效益性指标,故按公式(2)进行正向化处理。

$$x'_{ij} = \frac{x_{ij} - x_{min}}{x_{max} - x_{min}} \quad (2)$$

x'_{ij} 表示无量纲处理后的数值, x_j 表示第 j 项评价指标的数值

按公式(3)计算标准差,以表示指标变异性。按公式(4)计算相关系数,以表示指标冲突性。

$$\left\{ \begin{aligned} \bar{x}_j &= \frac{1}{n} \sum_{i=1}^n x_{ij} \\ S_j &= \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}} \end{aligned} \right. \quad (3)$$

$$R_j = \sum_{i=1}^p (1 - r_{ij}) \quad (4)$$

\bar{x}_j 表示 n 个样本第 j 项评价指标的平均值, S_j 表示第 j 项评价指标的标准差; R_j 表示第 j 项评价指标的相关系数, p 表示评价指标个数, r_{ij} 表示评价指标 i 和 j 之间的相关系数

按公式(5)计算信息量,信息量数值越大,表示该评价指标在整个评价指标体系中的作用越大,应该分配更多的权重。最后按公式(6)计算第 j 个指标的客观权重。

$$C_j = S_j \times R_j \quad (5)$$

$$W_j = \frac{C_j}{\sum_{j=1}^p C_j} \quad (6)$$

C_j 表示第 j 项评价指标的信息量; W_j 表示第 j 项评价指标的

客观权重

对评价指标的主观权重和客观权重进行 Kendall 一致性检验。依据检验结果选择合适的方法计算组合权重，得到高频药味加权矩阵。

2.3 聚类算法优选

由于构建的评价指标体系数据维度较高，为避免影响聚类结果，采用 IBM SPSS Statistic 20 软件中的多维尺度分析对高频药味加权矩阵进行降维处理。

选用 K-均值 (K-means) 聚类^[17]、SOM 聚类^[18]和模糊 C 均值聚类 (fuzzy C-means, FCM)^[19-20]分别对降维后的数据进行聚类分析。以聚类结果是否符合传统中药药味分类和中医药配伍理论为核心原则，以药味聚类结果与传统中医药理论的匹配度、聚类算法运行效率及各聚类算法对本数据库的包容度为评价指标，优选适合本数据库的聚类方法。

2.4 配方设计及 TOPSIS 综合评价

根据关联规则分析结果，提取药味核心组合，综合聚类优选结果及中医药理论，以安全、有效、经济为原则，设计新配方，构造配方加权矩阵。TOPSIS 综合评价主要步骤为原始矩阵正向化→正向化矩阵标准化→计算得分并归一化。由于本研究所选指标均为效益型指标，且在确定组合权重系数时已进行无量纲化处理，故只需将配方加权矩阵进行得分计算和排序。

3 结果

3.1 频次统计分析

出现频率排名前 5 的药味为黄芪、茯苓、人参、甘草、白术，这些药味均具有补气、健脾的功效，符合肝胃同治思想中疏肝理气、益气健脾的治则。以出现频次≥40 且在保健食品原料目录中为条件，筛选出 31 个高频药味，结果见表 2。对药味功能分类及性味归经统计分析显示，温性药、甘味药、补虚药使用最多，由于肝、胃病的病因杂合，互为因果，相互转化，易形成虚实夹杂之证，故治疗中多使用性味温和且有补益作用的药物，以扶正气^[21]。使用的药味中归脾经的最多，说明在肝、胃病治疗时常以治脾为桥梁。可见，若想达到肝胃同治，补中益气、理气健脾确是一个可行的思路，结果见图 2。

3.2 关联规则分析

支持度表示数据库中频繁项集出现的比例；置信度表示频繁项集之间联系的强度；增益用于衡量关联规则前项出现提升后项出现的程度，若增益>1，则可以认为前项和后项之间有良好的相关性^[22]。

表 2 含五味子处方中高频药味

Table 2 High-frequency traditional Chinese medicines in containing *Schisandre Chinensis Fructus* prescription

序号	药味	出现频次	序号	药味	出现频次
1	黄芪	243	17	远志	75
2	茯苓	220	18	川芎	73
3	人参	217	19	肉桂	63
4	甘草	206	20	菟丝子	57
5	白术	201	21	山楂	53
6	麦冬	192	22	杜仲	52
7	当归	164	23	山茱萸	52
8	陈皮	137	24	大枣	49
9	党参	128	25	补骨脂	47
10	山药	123	26	砂仁	47
11	枸杞子	121	27	酸枣仁	45
12	白芍	116	28	牛膝	44
13	葛根	106	29	木香	43
14	丹参	100	30	肉苁蓉	43
15	地黄	99	31	乌梅	43
16	熟地	98			

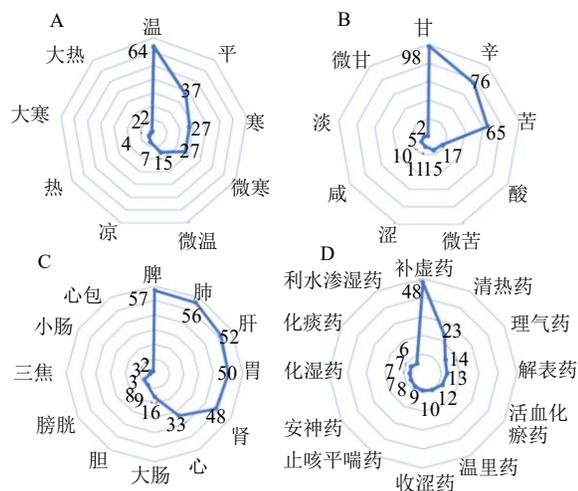


图 2 四气 (A)、五味 (B)、归经 (C) 及功能分类 (D) 分析
Fig. 2 Analysis of four qi (A), five flavors (B), meridian distribution (C) and functional classification (D)

基于数据库特性，设置支持度≥15%、置信度≥50%、最大前项数为 1，关联规则分析结果及网络图见表 3 (实例表示数据库中频繁项集出现的实际数目) 和图 3。其中，白术-茯苓、黄芪-白术、人参-麦冬、黄芪-麦冬为核心药对；关联规则网络图显示，五味子与人参、黄芪、白术、茯苓、甘草为强关联。

Stress 系数可近似理解为模型剩余的残差，其数值越接近 0 表明模型的解释度越好；RSQ 为决定系数，其数值越接近 1 证明模型的解释度越好。因此，根据模型运算结果的 Stress 系数和 RSQ 可判断降维效果。高频药味加权矩阵经降维处理后，结果见图 4。模型运行 Stress 系数为 0.0703，RSQ 为 0.9832，表示降维效果良好。

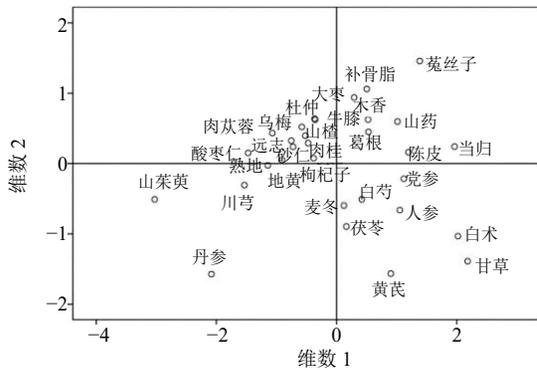
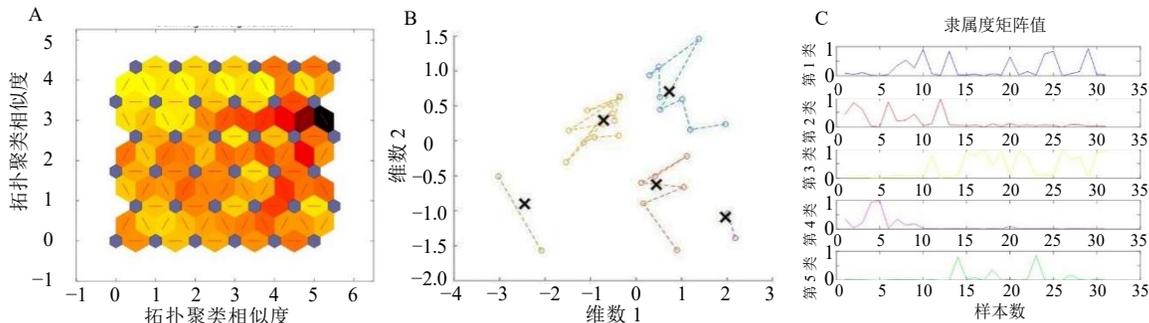


图 4 高频药味二维坐标

Fig. 4 Two-dimension coordinate of high-frequency traditional Chinese medicines

3.4.2 聚类算法对比 对高频药味加权矩阵进行 K-means 聚类，设置聚类个数为 5。在 MATLAB 软件中对高频药味二维数据进行 SOM 和 FCM 聚类，SOM 聚类中设置输出层为 8×8 矩阵，拓扑聚类相似度示意图中颜色越深，表示二者距离越远，据此将输出层 64 个拓扑节点分为 5 个类簇；FCM 聚类中设置聚类个数为 5，根据各高频药味在每一类簇中的隶属度大小，可将同一药味同时划分至不同的类簇中，如黄芪在聚类 2 中的隶属度为 0.4375，在聚类 4 中的隶属度为 0.3523，则黄芪既可归属于聚类 2，也可归属于聚类 4。K-means、SOM、FCM 聚类结果示意图见图 5。根据隶属度数值，以隶属度 ≤ 0.6 为标准，结合专业知识对聚类结果进行调整，结果见表 5。

本模型采用结合了传统中医药理论和现代科学研究结果的评价体系，以加权后的数据矩阵进行聚类分析。因此，聚类结果是否符合传统中医药药味分类和配伍理论是本模型优选聚类方法的评判标准。通过对比 K-means、SOM、FCM 3 种聚类方法的结果发现，在 K-means 中，不同分类的药味未被



A-SOM 神经网络拓扑聚类相似度 B-FCM 聚类图 C-FCM 聚类隶属度矩阵图

A-SOM neural network topology clustering similarity B-diagram of FCM clustering C-diagram of FCM clustering membership matrix

图 5 SOM、FCM 聚类示意图

Fig. 5 Schematic diagram of clustering results of SOM and FCM

表 5 K-means、SOM、FCM 聚类结果

Table 5 Clustering results of K-means, SOM and FCM

类别	K-means	SOM	FCM
聚类 1	黄芪、甘草、白术	甘草、白术	甘草、白术、黄芪、人参
聚类 2	丹参、地黄、熟地、远志、川芎、山茱萸、酸枣仁、肉苁蓉	丹参、山茱萸	丹参、山茱萸
聚类 3	肉桂、菟丝子、山楂、杜仲、大枣、补骨脂、砂仁、牛膝、木香、乌梅	地黄、肉桂、山楂、杜仲、牛膝、川芎、熟地、远志、砂仁、酸枣仁、肉苁蓉、乌梅	枸杞子、地黄、熟地、远志、川芎、肉桂、山楂、杜仲、砂仁、酸枣仁、牛膝、肉苁蓉、乌梅
聚类 4	当归、陈皮、党参、山药、枸杞子、菟丝子、山药、补骨脂、葛根、木香、白芍、葛根	熟地、远志、砂仁、酸枣仁、肉苁蓉、乌梅	当归、陈皮、山药、葛根、菟丝子、大枣、补骨脂、木香
聚类 5	茯苓、人参、麦冬	党参、人参、黄芪、白芍、茯苓、麦冬、枸杞子、陈皮	茯苓、党参、麦冬、白芍

明显区分,同一类簇中药味间的配伍关系并不明显,如聚类3中,补虚药菟丝子、温里药肉桂、理气药砂仁、木香等被划分到一起,这些药味在肝胃同治的治则下常作为佐使药辅助君药、臣药,而其相互之间的关系并不十分紧密。这可能与K-means聚类的初始点选择为随机选取,易导致聚类结果不稳定有关。

这一问题在SOM和FCM中在一定程度上得到了解决,如补虚药菟丝子、大枣与当归、山药划分到了一个类簇。就结果而言,SOM与FCM的结果比较相似。但同属补虚药且常配伍使用的人参和黄芪,在SOM中被分到了不同的类簇中,而FCM可将其划分至同一类簇,符合传统中医药理论的认识。同时,SOM计算较复杂,运行时间较长;FCM相对而言效率较高,且可根据隶属度结合实际经验与专业知识进行调整^[27],故优选FCM法作为五味子保胃护肝保健食品配方设计模型中的聚类算法。

在3种聚类方法中,清热药地黄和补虚药熟地均被划分至1个类簇,这提示聚类算法虽然能客观地挖掘样本间的深层联系,但仍然存在与专业知识不相符的情况,所以最终结果需结合专业知识进一步确定。

3.5 配方设计及TOPSIS评价结果

依据“3.2”项中关联规则分析结果,提取核心药味黄芪、茯苓、人参、甘草、白术,提取核心药对白术-茯苓、黄芪-白术,在肝胃同治理论的指导下,以五味子为核心药味,结合FCM聚类结果设计新配方,最终形成11个新配方。采用TOPSIS法对新配方进行评价排序,结果见表6。

其中,甘草及其相关制剂的不良反应近年报道越来越多,其安全性受到了临床使用及科学研究的高度重视。其不良反应涉及内分泌、心血管、神经、生殖、消化及免疫等多个系统^[28],以“假性醛固酮增多症”中所表现的高血压、低血钾、浮肿、腹水最常见,相比于那些药性剧烈或有毒性的药物而言,甘草及其制剂致不良反应的程度大多数情况下相对较低,但存在“累积效应”,即随着用药时间的增加,不良反应发生的概率随之增加。而保健食品具有适用范围较广,服用周期较长,对于原料的安全性要求高的特点。因此,尽管甘草在临床中应用广泛且在本研究中得分最高,但基于后续产品使用的安全性考虑,含甘草的配方不宜作为本研究中可选的最优配方。

表6 配方设计及TOPSIS评价结果

Table 6 Results of formula design and TOPSIS evaluation

配方	得分	排名
五味子-黄芪-白术-甘草	0.060 3	1
五味子-白术-茯苓-甘草	0.050 8	2
五味子-黄芪-白术-党参	0.049 3	3
五味子-黄芪-白术-人参	0.047 5	4
五味子-黄芪-白术-白芍	0.043 8	5
五味子-黄芪-白术-麦冬	0.038 8	6
五味子-白术-茯苓-黄芪	0.038 7	7
五味子-白术-茯苓-党参	0.036 1	8
五味子-白术-茯苓-人参	0.035 1	9
五味子-白术-茯苓-白芍	0.030 7	10
五味子-白术-茯苓-麦冬	0.026 6	11

因此,以五味子为核心药味具有护胃保肝功能可能的最优配方为五味子、黄芪、白术、党参。方中黄芪味甘微温,入脾、肺经,补中益气、升阳固表,为君药;配伍白术补气健脾,五味子益气生津、保肝为臣药,党参健脾益气协君药升提下陷之中气,与传统中医药理论相符。

4 讨论

保健食品目前的研发重点主要集中在配方设计、生产工艺及质量标准3个方面,配方设计是进行后续研究的关键和基础^[29]。对于中药类保健食品而言,配方设计最能体现其独特的优势,即在中医药理论的指导下,辨证分析,基于“君臣佐使”“七情和合”和“增效减毒”等配伍原则设计配方。如何将传统中医药理论与现代科学研究成果相结合,进行保健食品的配方设计是中药类保健食品目前面临的问题。因此,本研究提出将兼顾主客观的数据挖掘算法融入保健食品配方设计过程中,通过构建同时体现传统理论与现代研究的全面、合理的评价指标体系,将传统中医药理论与现代科学研究成果有机结合,从而进行新配方的演算。

本研究在课题组前期的研究基础上,采用赋权-聚类-综合评价的研究思路,在建模过程中,由于存在主观赋权法主观偏向性强,缺乏客观可信度,而客观赋权法有时不能反映决策者对不同指标重视程度,并且可能存在权重与实际专业知识不符合的问题。故本研究选择了能综合集成主、客观信息的组合赋权法。其中,客观赋权法并未选择目前使用较多的熵权法,而选择了CRITIC法,其优势在于

同时考虑了指标变异性和指标间的相关性, 完全利用数据自身的客观属性进行权重的计算。组合赋权方法的选用应以计算简便且全面反映信息为原则, 目前常用的方法有线性加权法^[30]、变异系数法^[31]、最小二乘法^[32]、离差最大化法^[33]等。本研究中主观权重和客观权重经一致性检验合格, 为计算简便选择主观权重和客观权重的算术平均值为组合权重。计算出的指标组合权重从大到小依次为药味传统功效(A₁)、现代文献研究(A₂)、在数据库中的出现频次(A₃)。但本研究所选指标仍有不足, 保健食品的服用周期较长, 其安全性也是生产研发中应重点关注的部分, 而目前所选指标仅关注药味的有效性, 并未关注其安全性。后续可通过检索药味相关不良反应的报道或毒性研究等文献进行评价, 以完善评价体系。

聚类算法的选择应根据聚类目的和数据特点综合确定。由于本研究建立的数据库信息较少, 为更全面、深入地挖掘药味间的联系, 综合各聚类方法的特点, 选择 K-means、SOM 和 FCM 进行对比优选。K-means 简单高效、时间和空间复杂度低, 但与专业知识存在一定偏差, 同时由于其初始点选择为随机选取, 易导致聚类结果不稳定。SOM 与 FCM 均以隶属度划分类簇, 运行结果相似, 且基本符合专业知识。但相比之下 FCM 运行效率更高, 可依据专业知识对聚类结果进行适当调整, 因此选择 FCM 为本模型的聚类方法。需要说明的是, 本研究将聚类算法用于保健食品的配方设计是为了在现有基础上挖掘数据内部的深层关联, 并不是单纯的合并同一类功效或与五味子关联密切的药味。因此, 在聚类算法的优选中, 并未选择现有聚类算法评价的外部或内部指标, 而采用了人为评价的方法, 以保证聚类结果与专业知识的一致性。

为解决现有保健食品配方单一重复、与现代科学研究联系不紧密的问题, 本研究以现有保健食品、方剂、中成药、专利和相关研究文献中的处方数据为基础, 结合多种机器学习方法进行中药保健食品配方设计, 在符合保健食品审批备案要求的前提下, 提供了既能体现传统中医药配伍理论又有足够现代科学研究成果支撑的中药类保健食品配方设计与研发的创新思路与方法。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 於洪建, 吴春福. 我国中药类保健食品的发展趋势 [J].

中草药, 2016, 47(18): 3342-3345.

- [2] 王林元, 张建军, 王淳, 等. 对中药类保健食品的认识及研究开发策略 [J]. 中国中药杂志, 2016, 41(21): 3927-3930.
- [3] 王进博, 陈广耀. 基于中医和法学对保健食品功能声称和证明标准的思考与建议 [J]. 中华中医药杂志, 2019, 34(12): 5782-5784.
- [4] 杨明, 胡彦君, 王雅琪, 等. 基于中医药理论与优势的中药保健产品设计思路 [J]. 中草药, 2017, 48(3): 419-423.
- [5] 任雪阳, 王宇, 魏胜利, 等. “保健功能-中药-中药”关联的石斛保健食品配方形规律分析及斑马鱼增强免疫力和缓解体力疲劳功能评价 [J]. 中草药, 2022, 53(8): 2435-2448.
- [6] 张睿, 王林元, 刘金莲, 等. 辅助改善记忆中药类保健食品和中成药组方规律及特点分析 [J]. 中国中药杂志, 2021, 46(13): 3222-3227.
- [7] 陈丰, 陈绍红, 柳海艳, 等. 辅助性保护化学性肝损伤保健食品配方的特点分析 [J]. 中草药, 2018, 49(7): 1703-1709.
- [8] 马嘉慕, 刘晓云, 任雪阳, 等. 基于 AHP-SOM 聚类-TOPSIS 的白及保胃护肝保健食品配方设计与评价 [J]. 中草药, 2021, 52(18): 5676-5687.
- [9] 潘晓霞, 白薛玲, 范好, 等. 肝胃同治理论及其临床应用 [J]. 陕西中医药大学学报, 2018, 41(4): 11-13.
- [10] 何伟峰, 李舒婷, 黄显伟. 叶天士“肝-胃”学术思想探析 [J]. 中国民族民间医药, 2016, 25(19): 1-3.
- [11] 中国药典 [S]. 一部. 2020: 68.
- [12] 任伟光, 张翠英. 五味子的研究进展及质量标志物(Q-marker)的预测分析 [J]. 中草药, 2020, 51(11): 3110-3116.
- [13] 王建, 张冰. 临床中药学 [M]. 第 2 版. 北京: 人民卫生出版社, 2016.
- [14] 罗宁, 贺墨琳, 高华, 等. 基于改进的 AHP-CRITIC 组合赋权与可拓评估模型的配电网综合评价方法 [J]. 电力系统保护与控制, 2021, 49(16): 86-96.
- [15] Saaty T L. Decision making with the analytic hierarchy process [J]. *Int J Serv Sci*, 2008, 1(1): 83.
- [16] Diakoulaki D, Mavrotas G, Papayannakis L. Determining objective weights in multiple criteria problems: The critic method [J]. *Comput Oper Res*, 1995, 22(7): 763-770.
- [17] Mac Queen J. Some methods for classification and analysis of multivariate observation [A] // Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability [C]. Berkeley: University of California Press, 1967: 281-297.
- [18] Kohonen T. Self-organized formation of topologically correct feature maps [J]. *Biol Cybern*, 1982, 43(1): 59-69.

- [19] Dunn J C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters [J]. *J Cybern*, 1973, 3(3): 32-57.
- [20] Bezdek J C. Numerical taxonomy with fuzzy sets [J]. *J Math Biol*, 1974, 1(1): 57-71.
- [21] 焦河玲, 唐成定, 周英武. 肝胃同病与肝胃同治 [J]. *中国实用医药*, 2007, 2(17): 93.
- [22] 张磐, 丁冷允, 姜宁, 等. 基于支持度-置信度-提升度的配网自动化系统数据挖掘算法及应用 [J]. *电测与仪表*, 2019, 56(10): 62-68.
- [23] 李艳萍, 乔琦, 柴发合, 等. 基于层次分析法的工业园区环境风险评价指标权重分析 [J]. *环境科学研究*, 2014, 27(3): 334-340.
- [24] 张玉, 魏华波. 基于 CRITIC 的多属性决策组合赋权方法 [J]. *统计与决策*, 2012(16): 75-77.
- [25] Li K, Ma Z, Robinson D, *et al*. Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering [J]. *Applied Energy*, 2018(231): 331-342.
- [26] 张美霞, 李丽, 杨秀, 等. 基于高斯混合模型聚类和多维尺度分析的负荷分类方法 [J]. *电网技术*, 2020, 44(11): 4283-4296.
- [27] 彭显刚, 赖家文, 陈奕. 基于聚类分析的客户用电模式智能识别方法 [J]. *电力系统保护与控制*, 2014, 42(19): 68-73.
- [28] 武媛媛, 顾坤, 高茜茜, 等. 甘草类制剂不良反应及其机制研究进展 [J]. *药物评价研究*, 2018, 41(7): 1363-1368.
- [29] 孙蓉, 齐晓甜, 陈广耀, 等. 中药保健食品研发、评价和产业现状及发展策略 [J]. *中国中药杂志*, 2019, 44(5): 861-864.
- [30] 马爱迪, 岳忠, 孙宝平, 等. 基于改进组合赋权法的建筑火灾保险费率研究 [J]. *中国安全科学学报*, 2020, 30(11): 134-140.
- [31] 王义保, 杨婷惠, 王世达. 基于组合赋权和灰色关联的城市公共安全感评价 [J]. *统计与决策*, 2019, 35(18): 45-50.
- [32] 高佳南, 吴奉亮, 李文福. 基于最小二乘法的优化组合权重模型在矿井环境舒适度评价中的应用 [J]. *安全与环境工程*, 2020, 27(5): 177-183.
- [33] 白丽丽, 白尚旺, 党伟超, 等. 基于离差最大化组合赋权的煤矿安全评价研究 [J]. *计算机应用与软件*, 2021, 38(4): 82-87.

[责任编辑 潘明佳]