

本文引用: 潘国兴, 易 钢, 卢彦杰. 基于改进双聚类算法的中医治疗过敏性紫癜性肾炎用药剂量规律的研究[J]. 湖南中医药大学学报, 2023, 43(10): 1872-1879.

基于改进双聚类算法的中医治疗过敏性紫癜性肾炎 用药剂量规律的研究

潘国兴, 易 钢*, 卢彦杰

湖南中医药大学信息科学与工程学院, 湖南 长沙 410208

[摘要] **目的** 采用模糊 C 均值聚类算法和改进的 CC 双聚类算法, 对中医治疗过敏性紫癜性肾炎 (Henoch-Schönlein purpura nephritis, HSPN) 用药剂量进行数据挖掘, 并结合中医理论, 分析数据挖掘结果。**方法** 检索 CNKI、万方数据库、维普中文期刊服务平台、中国生物医学文献数据库, 自建库至 2022 年 8 月 26 日关于中医治疗 HSPN 的文献资料, 提取其中的药物方剂。首先, 使用模糊 C 均值聚类算法创建药物的大剂量、中剂量、小剂量隶属矩阵, 将所有药物剂量进行分类, 再结合改进的 CC 双聚类算法挖掘中医治疗 HSPN 关键药物的组分信息。**结果** 共纳入中医治疗 HSPN 处方 387 则, 其中包括药物 174 味, 高频药物 (频次 ≥ 30 次) 38 味, 高频药物组合 10 组, 常见药物剂量组合有大剂量水牛角、生地黄配小剂量小蓟 (2:2:1); 小剂量连翘配伍小剂量金银花 (1:1); 大剂量当归配伍中剂量黄芪 (1:2) 等, 以清热解毒、疏散风热、益气活血药为主。**结论** 经改进的 CC 双聚类算法联合模糊 C 均值聚类算法, 可以从众多的方剂数据集中挖掘出关键的药物组分, 并根据药物的剂量大小进行区间划分, 具有高效率数据挖掘的特点, 可为 HSPN 中医临床治疗提供一定参考。

[关键词] 数据挖掘; 过敏性紫癜性肾炎; CC 双聚类算法; 模糊 C 均值聚类算法; 用药规律

[中图分类号] R255.7

[文献标志码] A

[文章编号] doi:10.3969/j.issn.1674-070X.2023.10.019

Dosage rules of Chinese medicines treating Henoch-Schönlein purpura nephritis based on the improved biclustering algorithm

PAN Guoxing, YI Gang*, LU Yanjie

School of Information Science and Engineering, Hunan University of Chinese Medicine, Changsha, Hunan 410208, China

[Abstract] **Objective** To mine the dosage of Chinese medicines in treating Henoch-Schönlein purpura nephritis (HSPN) by Fuzzy C-means algorithm and improved Cheng-Church biclustering algorithm, and to analyze its data mining results combined with Chinese medicine theory. **Methods** CNKI, Wanfang, VIP, and China Biomedical Literature Database were searched, the literature on the treatment of HSPN by Chinese medicine was screened from the established time to August 26, 2022, and the medicines and formulas were identified. Fuzzy C-means clustering algorithm was used to set the high-, medium-, and low-dose membership matrices of medicines, and all the doses were classified, and finally the component information of key Chinese medicines for HSPN were mined with the improved Cheng-Church biclustering algorithm. **Results** A total of 387

[收稿日期] 2023-04-15

[基金项目] 湖南省中医药管理局项目 (A2023048)。

[第一作者] 潘国兴, 男, 硕士研究生, 研究方向: 电子信息。

[通信作者] * 易 钢, 男, 硕士, 副教授, 硕士研究生导师, E-mail: e_gang@163.com。

formulas were included in the treatment of HSPN, with 174 medicines. Thirty eight high-frequency medicines were taken more than 30 times. Ten groups of high-frequency combinations include: high-dose Shuiniujiao (Bubali Cornu)+high-dose Shengdihuang (Rehmanniae Radix)+low-dose Xiaoji (Cirsii Herba) (2:2:1); low-dose Lianqiao (Forsythiae Fructus)+low-dose Jinyinhua (Lonicerae Japonicae Flos) (1:1); high-dose Danggui (Angelicae Sinensis Radix)+medium-dose Huangqi (Astragali Radix) (1:2), etc. The medicines mainly function is to clear heat and remove toxins, release the exterior with pungent-cool, benefit qi and circulate blood. **Conclusion** The improved Cheng-Church biclustering algorithm combined with fuzzy C-means algorithm can mine the key medicinal components from massive formula data, and the intervals can be divided according to the doses, which is highly efficient and can provide certain reference for TCM clinical treatment of HSPN.

[**Keywords**] data mining; Henoch-Schönlein purpura nephritis; Cheng-Church biclustering algorithm; fuzzy C-means clustering algorithm; dosage rules

过敏性紫癜性肾炎(Henoch-Schönlein purpura nephritis, HSPN)患者通常表现为蛋白尿、血尿、腹痛、皮肤紫癜等症状^[1]。中医学将其归为“血证”“紫癜风”“水肿”等范畴^[1],认为是由于患者先天阴虚质燥,营血中有伏火,复又外受风热、温热或药毒之邪,导致两热相搏,血热炽燔,从而灼伤肤络,血溢肤表而发为紫癜^[2]。临床研究表明,采用中药治疗 HSPN 效果明显,且不良反应小^[3]。但是,在对当前中医治疗 HSPN 相关文献进行研究分析后发现,在用药剂量、药物配伍等问题上缺乏系统性归纳,存在局限性。近些年,大量学者采用现代统计分析和数据挖掘技术对中医用药规律进行研究,有效推动了中医的发展,但采用的算法大体相同且较为原始,挖掘的结果也较为浅显。因此,本研究采用聚类算法的数据挖掘技术探讨中医治疗 HSPN 用药剂量规律,以期为临床治疗和新药研发提供参考。

人工智能技术的发展与应用,对中医药的发展起到促进作用。中医知识体系庞大,内容极为丰富,且具有经验性、辨证性的特点,仅仅依靠人力难以对中医基础理论、中药方剂学和中医文献学等进行规律总结和隐藏关系有效发掘。数据挖掘技术具备从大量的数据中挖掘出具有特殊联系的隐藏信息的能力,因此被广泛应用。聚类算法作为数据挖掘技术中的一个重要的分支,在中医用药规律的发掘中具有举足轻重的作用。例如,倪瑛等^[4]采用 K-means 聚类算法探讨治疗新型冠状病毒肺炎处方的组方配伍规律,聚类分析获得 6 类药物社团,具有清肺解毒、芳香化湿、辟秽化浊、通腑解毒祛瘀、回阳救逆、补气养阴等配伍特征;邹锦等^[5]将熵聚类应用于发掘中医经方治疗抑郁症的用药规律,通过聚类分析得到

新方 3 个,为临床辨治抑郁症提供参考。

针对中医治疗 HSPN 的用药特点,本文采用改进的 CC 双聚类算法挖掘其中的隐藏规律信息,并采用模糊 C 均值(fuzzy C-means, FCM)聚类算法对挖掘出的药物剂量大小进行区分。旨在得出中医治疗 HSPN 用药剂量的规律,验证算法模型的有效性,为中医治疗 HSPN 研究提供参考。

1 资料与方法

1.1 资料来源

以“过敏性紫癜性肾炎”作为关键词,检索项选择“主题”,文献类型选择“期刊论文”,对 CNKI、万方数据库、维普中文期刊服务平台、中国生物医学文献数据库进行检索,筛选年限为建库至 2022 年 8 月 26 日,共收集到中医治疗 HSPN 相关研究文献 508 篇。

1.2 纳入及排除标准

纳入标准:(1)符合 HSPN 的中西医诊断标准;(2)中医内服方剂治疗 HSPN 的临床研究文献;(3)治疗结局判定有效;(4)中药处方药物完整,剂量记载明确;(5)患者年龄区间为 20~30 岁。

排除标准:(1)综述、Meta 分析、个案;(2)针灸、熏蒸、西医结合等联合治疗文献;(3)动物实验;(4)重复处方。

1.3 数据提取及录入

将文献中的方剂按照方剂名、症状、药物(剂量)的格式录入 Microsoft Excel 2019 软件进行存储。

中药名称规范化:参照《中医诊断学》^[6]《中华人民共和国药典》^[7],对录入的症状名称和中药名称进行规范化处理,如“血尿”“尿血”等规范为“尿血”,“银花”“双花”规范为“金银花”,“丹皮”规范为“牡丹

皮”等。

药物剂量规范化:为方便后续数据挖掘工作,将所有药物剂量单位统一为克,按照一枚标准大枣 10 g、一片生姜 3 g 等转化为“大枣 10 g”“生姜 3 g”。

数据录入:录入工作由两人合作完成,采用资料双人背对背录入方式,以此保证数据的准确可靠性。

1.4 数据预处理

对数据进行整理过程中发现,个别处方中有 1~2 味其他中医专家不曾使用的药物。因为不同的中医专家有着不同的临床经验,对药物的配伍、用量、服用顺序等各不相同。此外,患者体质与病情不可能完全一致。因此,会选择性加入一些药物,以达到更好的疗效。然而,低频药物仅占整体数据集中的一小部分,如果将它们都加入数据集则会产生大量噪声数据,导致算法运行效率降低、结果解释性降低、过拟合等问题。因此,为寻求普适性用药规律,需要设置界定阈值。在对药物出现频次的不同值作为阈值进行了多组实验后,结果显示,频次=5 次作为阈值时可以得到较好的聚类效果,并且频次 ≥ 5 次的药物占整体数据集的比例适中。故将药物界定阈值设为 5,得到出现频次 ≥ 5 次的药物 174 味,建立药物-剂量矩阵,矩阵行代表方剂,矩阵列代表药物名,矩阵内数值代表剂量,缺失值填 0。详见表 1。

1.5 FCM 聚类算法

表 1 药物-剂量矩阵

| 方剂 | 生地黄 | 牡丹皮 | 白茅根 | 赤芍 | ... |
|-----|-----|-----|-----|-----|-----|
| 1 | 15 | 10 | 30 | 10 | ... |
| 2 | 0 | 15 | 0 | 0 | ... |
| 3 | 9 | 6 | 0 | 9 | ... |
| 4 | 0 | 10 | 30 | 0 | ... |
| 5 | 12 | 0 | 10 | 0 | ... |
| 6 | 10 | 10 | 0 | 10 | ... |
| ... | ... | ... | ... | ... | ... |

模糊算法源于模糊数学理论,属于智能算法,常见的模糊算法有均值模糊和高斯模糊。FCM 是一种基于划分的聚类算法,其目的在于使划分到同一簇的对象之间的相似度最大,而不同簇之间对象的相似度最小。FCM 是一种模糊划分,相较于其前身 C 均值聚类算法的硬性划分,FCM 的划分更为柔性^[8]。

在中药剂量区间划分研究中,刘丹^[9]采用现代统计学的传统划分方法,易于理解和实现、适应性强、更容易操作,划分模式比较直观和易于操作,适用于处理药物剂量比较稳定以及算力比较弱的情况。但传统划分方法还存在一些局限性,例如划分模式非此即彼,不能全面反映药物剂量效应和区间,而且对于中医药领域中药物剂量复杂且效应不确定的情况有限制。因此,需要考虑应用更细致化和更符合实际情况的剂量划分方法。如模糊聚类算法,其可以很好地处理和利用非精确性数据,在中医药研究中有很好的应用前景。与传统方法相比,模糊聚类算法可以更好地适应不同病情和患者的个体差异,实现药物剂量的个性化调整。本次研究采用 FCM 算法对药物剂量区间进行划分,FCM 算法由 RUSPINI 首先提出,之后经过 DUNN 与 BEZDEK 的改进^[10],成为现在人们熟知的 FCM 算法。FCM 算法大致流程如下:

第一步:随机初始化聚类中心;

第二步:对于每个数据点,计算其与每个聚类中心的相似度(距离);

第三步:根据计算得到的相似度值,将每个数据点分配到最相似的聚类中心;

第四步:更新每个聚类中心,计算每个聚类中心的向量值,使其尽可能接近同一类别中的所有数据点;

第五步:重复第二步~第四步,直到聚类中心不再发生变化或达到预设的最大迭代次数^[11]。

1.6 CC 双聚类算法

1.6.1 原始算法 CC 双聚类算法由 CHENG 和 CHURCH 于 2000 年提出^[12]。算法的核心思想在于利用平均平方残基来寻找矩阵中具有相似特征的子矩阵^[13],例如,一个 X 行 Y 列的矩阵 A,其平均平方残基可以表示为:

$$H(X, Y) = \frac{1}{|X| |Y|} \sum_{x \in X, y \in Y} (\alpha_{xy} - \alpha_{xY} - \alpha_{Xy} + \alpha_{XY})^2 \quad (1)$$

$$\alpha_{xY} = \frac{1}{|X| |Y|} \sum_{x \in X, y \in Y} \alpha_{xy} \quad (2)$$

$$\alpha_{Xy} = \frac{1}{|X|} \sum_{x \in X} \alpha_{xy} \quad (3)$$

$$\alpha_{xy} = \frac{1}{|X|} \sum_{x \in X} \alpha_{xy} \quad (4)$$

$$RS_{xy} = \alpha_{xy} - \alpha_{xy} - \alpha_{xy} + \alpha_{xy} \quad (5)$$

上述公式中,公式(5)计算结果即为矩阵中每个元素的残基,将该元素依次减去所在行的平均值和所在列的平均值,最后加上矩阵所有元素的平均值;计算出矩阵中每个元素的残基后就可以得出该矩阵的平均平方残基 $H(X,Y)$,设置阈值 ε ,当矩阵的 $H(X,Y) \leq \varepsilon$ 时,即该矩阵具有较强的内部一致性。因此,对于 $H(X,Y) > \varepsilon$ 的矩阵,若想达到较强内部一致性的目标,就需要依次删除 $H(X,Y) > \varepsilon$ 的行或列 R ,删除条件公式为:

$$R = \{x \in X; \frac{1}{|Y|} \sum_{y \in Y} RS_{xy}(x,y)^2 > H(X,Y)\} \quad (6)$$

为避免错误删除平均平方残基小于阈值的行列,加入填补行列条件公式:

$$R = \{x \notin X; \frac{1}{|Y|} \sum_{y \in Y} RS_{xy}(x,y)^2 \leq H(X,Y)\} \quad (7)$$

最后输出目标子矩阵;经过一轮的删除与输出,为了下次迭代能够输出新的子矩阵,需要对删除的行或列进行随机值的填补,经过反复迭代,最终输出多个不同的具有较强内部一致性的子矩阵。CC 双聚类算法的伪代码形式如下:

输入:矩阵 A

输出:簇 C1,C2,⋯, Ck

- 1.初始化:将所有行作为一个簇
- 2.hile (有多于一个簇) do
- 3.计算任意两个簇之间的相似度
- 4.合并相似度最大的两个簇
- 5.endwhile

返回聚类结果

1.6.2 改进算法 利用药物-剂量矩阵进行 CC 双聚类算法的剂量挖掘。由于中医处方所涉及的药物种类十分庞大,因此药物-剂量矩阵中存在大量的0值;经过上述对原始 CC 双聚类算法内容的描述,可以预料到算法会将矩阵中大量的0值作为具有较强内部一致性的子矩阵输出,这样的挖掘结果没有任何意义。除此之外,算法迭代过程中还会将删除的行或列进行随机数的填补,这必然会影响到原始的剂量信息,导致输出结果不准确。

针对上述出现的两个问题,提出“惩罚策略”^[14],具体算法步骤如下:

第一步:创建与药物-剂量矩阵 $\rho(X,Y)$ 列) 维度相同的 0 矩阵 θ ,遍历矩阵 ρ ,当 $\rho(x,y)=0$ 时,设 $\theta(x,y)=1$;

第二步:创建惩罚系数 $\omega=|X| \times |Y| \times \mu$,其中 μ 为初设阈值,可以通过调节其大小来控制惩罚力度;

第三步:对原始 CC 双聚类算法中平均平方残基公式(1)进行改进,即为:

$$H(X,Y)' = H(X,Y) + (\theta_{xy} \times \omega) \quad (8)$$

其中 θ_{xy} 代表矩阵 θ 中 X 行, Y 列的取值。

第四步:对原始 CC 双聚类算法中删除行列条件公式(6)进行改进:

$$R = \{x \in X; \frac{1}{|Y|} \sum_{y \in Y} RS_{xy}(x,y)^2 + (\theta_{xy} \times \omega) > H(X,Y)\} \quad (9)$$

通过在(6)式中添加“惩罚项”可以保证优先删除矩阵 ρ 中的 0 值;

第五步:对原始算法中添加随机值条件公式(7)进行改进:

$$R = \{x \notin X; \frac{1}{|Y|} \sum_{y \in Y} RS_{xy}(x,y)^2 + (\theta_{xy} \times \omega) \leq H(X,Y)\} \quad (10)$$

改进后的随机值填补条件公式可以使算法优先填补不含 0 值的行列;

第六步:对于一次迭代输出的目标子矩阵,记录元素所在位置 (x,y) ,并将矩阵 $\theta(x,y)=1$,算法在进行下一次迭代时,因为惩罚项的因素删除第一次迭代输出的子矩阵相同位置的元素,即使该位置被赋予满足输出条件的随机值。

将改进思路应用到 CC 双聚类原始算法中,改进后的算法伪代码如下:

输入:矩阵 A

输出:簇 C1,C2,⋯, Ck

- 1.初始化:将所有行作为一个簇
- 2.while(有多于一个簇) do
- 3.计算相似度矩阵 D
- 4.对 A 进行标记,初始化 $flag0=[0,0,\dots,0]$, $flag1=[0,0,\dots,0]$
- 5.初始化 $aaa=A \times flag0 \times flag1.T$
- 6.初始化权重矩阵 w
7. for i from 1 to n do


```

8. for j from 1 to m do
9.a_ij=(a_ij-mean(a_i)-mean(a_j))^2
10.a_ij=a_ij+w_ij×惩罚因子×权重矩阵中元素
11.endfor
12.endwhile
13.选取 D 中值最大的部分,保证形成的子矩阵
符合聚类条件
14.剔除保留之后不需要的行和列
15.使用 add 函数动态添加行和列
16.endwhile
返回聚类结果

```

该改进算法主要通过引入惩罚因子和动态调整方剂数据矩阵来优化聚类结果。在初始化阶段,通过设置权重矩阵和惩罚因子,用于调整元素距离的影响力。然后,使用 CC 双聚类算法对方剂数据矩阵进行聚类,得到聚类的平均值。接下来,根据惩罚因子对方剂数据矩阵进行调整。将某些行和列置零,并计算调整后的行和列的均值。然后,对方剂数据矩阵进行旋转操作,并减去旋转后的均值。将调整后的方剂数据矩阵加上聚类的平均值,并进行平方操作。在计算惩罚项时,加上惩罚因子,并计算所有非零元素的平均值。通过判断惩罚值,评估聚类结果的稳定性是否满足要求。根据聚类结果的稳定性,动态地添加或删除方剂数据矩阵的某些行或列。通过重复以上步骤,直至聚类结果稳定且满足要求。该改进算法通过引入惩罚因子和动态调整方剂数据矩阵,以及重复迭代的方式,优化了聚类结果,避免了过于紧密或离散的情况,从而实现更好的聚类效果。

2 结果

2.1 药物频次分析

共纳入中医治疗 HSPN 处方 387 则。在 Pycharm 2021(Python 3.7)平台编写频数统计程序,截取出现频次 ≥ 30 次的药物 38 味。详见表 2。使用频次由高到低的中药依次是生地黄、牡丹皮、白茅根、赤芍、甘草等。生地黄具有清热生津、凉血止血的作用,常用于热病伤阴、发斑发疹、衄血等,因此,被广泛应用于 HSPN 的治疗^[15]。

表 2 治疗 HSPN 的高频药物(频次 ≥ 30 次)分布表

| 序号 | 名称 | 频次/次 | 序号 | 名称 | 频次/次 |
|----|-----|------|----|-------|------|
| 1 | 生地黄 | 229 | 20 | 白花蛇舌草 | 59 |
| 2 | 牡丹皮 | 197 | 21 | 女贞子 | 57 |
| 3 | 白茅根 | 152 | 22 | 三七 | 56 |
| 4 | 赤芍 | 137 | 23 | 金银花 | 54 |
| 5 | 甘草 | 133 | 24 | 白术 | 49 |
| 6 | 黄芪 | 132 | 25 | 黄柏 | 47 |
| 7 | 小蓟 | 122 | 26 | 川芎 | 45 |
| 8 | 紫草 | 117 | 27 | 泽泻 | 43 |
| 9 | 当归 | 104 | 28 | 党参 | 42 |
| 10 | 丹参 | 100 | 29 | 益母草 | 41 |
| 11 | 茯苓 | 88 | 30 | 黄芩 | 39 |
| 12 | 连翘 | 87 | 31 | 太子参 | 38 |
| 13 | 茜草 | 83 | 32 | 白芍 | 38 |
| 14 | 防风 | 75 | 33 | 薏苡仁 | 34 |
| 15 | 仙鹤草 | 70 | 34 | 蒲公英 | 34 |
| 16 | 山药 | 64 | 35 | 知母 | 33 |
| 17 | 蝉蜕 | 61 | 36 | 芡实 | 33 |
| 18 | 水牛角 | 60 | 37 | 山茱萸 | 31 |
| 19 | 墨旱莲 | 60 | 38 | 土茯苓 | 30 |

2.2 药物剂量 FCM 聚类分析

选用 MATLAB R2021b,利用 FCM 聚类算法分别对表 1 的每列进行去 0 值聚类分析。

算法初始化内容:模糊加权指数 m 取值范围为 $[1,2.5]$,通常取 $m=2$ 可达到较好的效果;聚类中心点个数 $cluster_n=3$;最大迭代次数为 300。

可以得到 174 味药物大剂量、中剂量、小剂量的中心点信息。详见表 3。

经过上述的 FCM 聚类,可以得到每味药物的隶属度矩阵以及分布曲线。以“茯苓”的部分隶属度为例进行分析。详见表 4、图 1。

在 88 首含茯苓的方剂中,茯苓剂量有 6 种,分别为 6、10、12、15、20、30 g,表 4 展示这 6 种剂量的隶属度,图 1 展示 88 首方剂中茯苓剂量的隶属度形成的隶属度曲线。从以上结果可以得出,茯苓小剂量中心点约为 10 g,中剂量中心点约为 15 g,大剂量中心点约为 20 g。有研究根据张仲景方剂,对茯苓的剂量疗效关系进行分析,以 10 g 与 20 g 作为边界,将茯苓的剂量分为小剂量、中剂量和大剂量^[16],与上述的模糊聚类结果相同。因此,采用模糊聚类算法可

表3 药物剂量中心

| 药物 | 小剂量/g | 中剂量/g | 大剂量/g |
|-----|-------|-------|-------|
| 生地黄 | 10.00 | 15.00 | 20.02 |
| 牡丹皮 | 10.00 | 15.00 | 20.00 |
| 白茅根 | 15.00 | 20.00 | 30.00 |
| 赤芍 | 10.00 | 12.00 | 15.00 |
| 甘草 | 5.00 | 6.00 | 10.00 |
| 黄芪 | 15.00 | 30.00 | 50.04 |
| 小蓟 | 10.00 | 15.00 | 30.00 |
| 紫草 | 10.00 | 15.00 | 30.00 |
| 当归 | 10.00 | 12.00 | 15.00 |
| 丹参 | 10.00 | 15.00 | 30.00 |
| 茯苓 | 10.00 | 15.00 | 20.00 |
| 连翘 | 10.00 | 15.00 | 20.00 |
| 茜草 | 10.00 | 15.00 | 30.00 |
| 防风 | 6.00 | 10.00 | 15.00 |
| 仙鹤草 | 10.00 | 15.00 | 30.00 |
| 山药 | 15.00 | 20.00 | 30.00 |
| 蝉蜕 | 6.00 | 10.00 | 15.00 |
| 水牛角 | 10.00 | 15.00 | 30.00 |
| 墨旱莲 | 10.00 | 15.00 | 20.00 |
| 金银花 | 10.00 | 15.00 | 30.00 |
| ... | ... | ... | ... |

表4 茯苓剂量隶属度矩阵

| 药物 | 剂量/g | 小剂量隶属度 | 中剂量隶属度 | 大剂量隶属度 |
|-----|------|---------|---------|---------|
| 茯苓 | 6 | 0.385 8 | 0.322 2 | 0.292 0 |
| 茯苓 | 10 | 0.917 7 | 0.044 3 | 0.038 0 |
| 茯苓 | 12 | 0.377 5 | 0.345 0 | 0.277 4 |
| 茯苓 | 15 | 0.052 5 | 0.895 0 | 0.052 5 |
| 茯苓 | 20 | 0.772 0 | 0.122 8 | 0.105 3 |
| 茯苓 | 30 | 0.309 4 | 0.329 8 | 0.360 9 |
| ... | ... | ... | ... | ... |

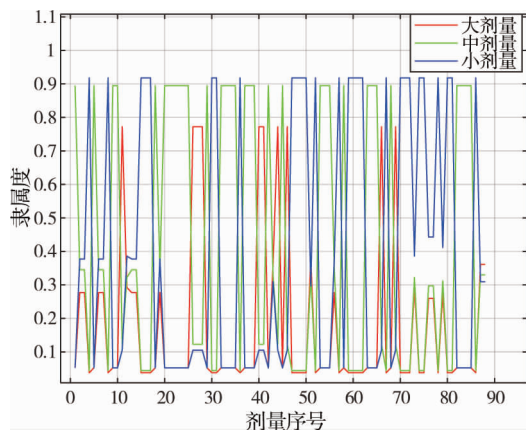


图1 茯苓隶属度曲线

以有效地将药物的剂量关系进行划分。

2.3 药物剂量双聚类分析

实验中,首先利用原始的CC双聚类算法对药

物-剂量矩阵进行聚类分析,聚类结果详见表5。可以直观地看到,原始的CC双聚类算法在面对稀疏矩阵时并不能发挥出对非0数据的聚类作用,显然无法达到本次研究的目的。

表5 CC双聚类算法部分结果

| 方剂 | 乌梢蛇 | 瓜蒌 | 独活 | 没药 | ... |
|---------|-----|----|----|----|-----|
| 芪芍茜根散 | 0 | 0 | 0 | 0 | 0 |
| 清热健肾汤II | 0 | 0 | 0 | 0 | 0 |
| 养阴健肾汤 | 0 | 0 | 0 | 0 | 0 |
| 益气健肾汤 | 0 | 0 | 0 | 0 | 0 |
| ... | 0 | 0 | 0 | 0 | 0 |

将药物-剂量矩阵导入改进CC双聚类算法中。初始设定惩罚系数中初设阈值 μ 为1,进行双聚类,并观察聚类结果。如果结果过分承认噪声点,则将 μ 值的大小适当增加,重复这个过程直到结果合理。设定一组 μ 的可能值,分别进行多次双聚类,选出噪声点最少的一组,取得该数值。面对不同类型的数据集,可以根据数据的特征进行设定,当数据包含噪声较多时,可以适当增加 μ 的大小,当数据中噪声很小时,则 μ 可以设定为较小值。经过多次重复实验,最终设定参数 $\mu=10$;迭代次数 $n=120$,进行双聚类分析,挖掘出10组药物剂量组分子矩阵,部分内容展示见表6—8。根据“2.2”得到的药物剂量隶属度,分别用low(小剂量)、mid(中剂量)、high(大剂量)来标注药物的剂量大小。

表6所列举的3首自拟方中均用到水牛角、生地黄和小蓟,该类型的自拟方大多是犀角地黄汤合小蓟饮子加减。犀角地黄汤联合常规西医干预治疗儿童过敏性紫癜具有良好的疗效,能显著改善症状与免疫功能。方中君药水牛角起到祛风解毒、凉血解毒、清热定惊的作用^[7]。小蓟饮子为理血剂,具有凉血止血、利水通淋之功效,方中君药小蓟凉血止血、祛瘀消肿^[18-19]。两首方剂中均使用到的生地黄味甘、苦,性寒,归心、肝、肺经,能清热凉血、养阴生津。

从生成的药物剂量矩阵来看,水牛角与生地黄大多使用大剂量和中剂量,配合小剂量的小蓟。水牛角与生地黄用量宜大,二者配伍剂量比例约为1:1,且常见于30g左右;小蓟常规用量为5~12g,属小剂量,与生地黄、水牛角剂量比例约为1:2;水牛角、

生地黄、小蓟剂量比例大约控制在 2:2:1 时,清热凉血之效益彰,治疗热迫血行之衄血、崩中下血、尿血等症具有显著的疗效^[20]。

表 6 药物组分子矩阵 I

| 方剂 | 水牛角 | 生地黄 | 小蓟 |
|---------|----------|----------|---------|
| 自拟方 I | 30(high) | 25(high) | 12(low) |
| 自拟方 II | 20(mid) | 20(high) | 10(low) |
| 自拟方 III | 15(mid) | 15(mid) | 10(low) |

表 7 显示有连翘、金银花。连翘与金银花均具备清热解毒、疏散风热之功效,两药配伍,升浮宣散、清热解毒之力倍增^[21]。丁影^[22]研究不同配伍比例对金银花-连翘药对中 9 个成分提取率的影响,发现当金银花-连翘药对比例为 1:1 时,其中的绿原酸、咖啡酸、獐牙菜苦苷等 9 种主要成分含量最高。此外,金银花与连翘均具有解热抗炎作用,吴嘉瑞等^[23]采用网络药理学方法研究两种药物对炎症性疾病的作用机制,当金银花与连翘比例为 1:1 配伍使用时,抗炎作用最强,且解热作用最为明显。因此,在众多治疗 HSPN 的方剂中,使用金银花-连翘药对时,会将药对的剂量比例控制在 1:1 左右,使其发挥出最大的疗效^[24]。

表 7 药物组分子矩阵 II

| 方剂 | 连翘 | 金银花 |
|-------|----------|---------|
| 解毒凉血方 | 20(high) | 20(mid) |
| 连翘败毒散 | 10(low) | 9(low) |
| 清热活血方 | 9(low) | 9(low) |

表 8 显示 4 则方剂均用到黄芪、当归和茯苓。黄芪具有补气升阳、利水消肿的功效;当归具有补血活血、调经止痛的作用^[25];茯苓能够利水渗湿、健脾宁心。

黄芪-当归具有改善血液循环、抗脏器纤维化、抗炎等作用^[26]。临床研究发现,二者在治疗糖尿病、肾病综合征等疾病中具有良好的疗效。尚伟庆等^[27]研究了黄芪-当归合剂对肾病综合征患者肾功能、炎症机制及疗效的影响,在采用常规治疗的基础上对患者使用黄芪 50 g,当归 25 g(黄芪:当归=2:1),结果显示黄芪当归合剂可显著改善患者肾功能,减轻患者微炎症状态。从聚类结果上看,诸多中医专家

在用到黄芪-当归药物配伍时,习惯将二者的剂量比例控制在 2:1,以达到较好的治疗效果。

表 8 药物组分子矩阵 III

| 方剂 | 黄芪 | 当归 | 茯苓 |
|-----------|---------|----------|----------|
| 归脾汤 | 15(low) | 10(low) | 12(low) |
| 健脾脱敏清络汤 | 15(low) | 9(low) | 10(low) |
| 清热凉血化瘀汤 | 30(mid) | 15(high) | 30(high) |
| 益气汤合桃红四物汤 | 30(mid) | 15(high) | 20(high) |

3 讨论

将 FCM 聚类算法和 CC 双聚类算法相结合,目的是为了更全面、更深入地研究药物的剂量规律和关键药物组合。在 CC 双聚类算法中,药物的剂量已经被明确地归为某一具体数值,但是仅仅一个具体的数值并不能准确描绘出药物的剂量和药效之间的关系。因此,使用 FCM 聚类算法将药物剂量分为大、中、小剂量,可以在一定程度上涵盖药物的剂量效应和区间,从而更好地把握药物剂量对治疗的影响。

同时,药物的剂量大小并不是唯一决定药物疗效的因素,与药物组合、剂量比例密切相关。通过 CC 双聚类算法,可以找到出现频率较高的药物组合和对应的药物剂量比例。但是药物组合的适当剂量并不是根据出现频率得出的,而是需要考虑具体的治疗目标和当前患者病情。因此,在实际临床应用中,算法得出的“最佳剂量比例”并非每个患者的最佳剂量比例。而将 FCM 聚类算法得出的大、中、小剂量区间和 CC 双聚类算法得出的药物组合、剂量比例结合在一起,能够较好地综合考虑药物剂量和药物组合的综合效应,以更准确地指导临床实践,呈现出更好的治病效果。

在这项关于中药治疗 HSPN 的研究中,本研究团队使用改进的 CC 双聚类算法和 FCM 聚类算法对众多中医专家的药物剂量进行分析和处理,得出了有效的药物剂量区间和药物组合,并总结了相应的实验结论。这种研究方法和结论可为个性化药物治疗提供参考,具有一定的研究价值。此项研究仍需进一步深入,例如如何根据患者特征提出更加个性化的治疗方案;如何进一步优化药物的剂量区间和药物组合,以更好地满足患者的临床需求,等等。未

来,本研究团队希望能够借助更加人性化和高效的算法,探究药物的潜在作用及其剂量的潜在效应,并从中找到隐性的治疗策略,以更好地为患者提供个性化的治疗方案,为发展中药治疗及个性化医疗提供支持 and 实践依据。

参考文献

- [1] 牛静,周雪,陈团营. 小儿过敏性紫癜的中医研究进展[J]. 中国医学创新, 2023, 20(19): 184-188.
- [2] 赵旭晖. 过敏性紫癜性肾炎中医证型及方药的文献分析研究[D]. 北京: 北京中医药大学, 2013.
- [3] 王龙,张霞,任献青,等. 中医阶梯治疗方案治疗血尿和蛋白尿型儿童紫癜性肾炎的疗效观察[J]. 时珍国医国药, 2020, 31(5): 1167-1169.
- [4] 倪瑛,张一鸣,范欣生,等. 基于复杂网络与PCA降维和K-Means聚类探讨治疗COVID-19组方配伍特点[J]. 南京中医药大学学报, 2020, 36(6): 897-901.
- [5] 邹锦,杨学琴,张小勇,等. 基于关联规则和熵聚类的中医经方治疗抑郁症用药规律[J]. 中国实验方剂学杂志, 2017, 23(10): 211-215.
- [6] 任健. 中医诊断学[M]. 济南: 山东科学技术出版社, 2020.
- [7] 国家药典委员会. 中华人民共和国药典: 2020年版[M]. 北京: 中国医药科技出版社, 2020.
- [8] 赵琦,唐晨,徐敏,等. 基于改进模糊c均值聚类算法的条纹投影背景去除[J]. 激光与光电子学进展, 2022, 59(24): 118-125.
- [9] 刘丹. 基于我校国医堂中医专家临床用药的剂量研究[D]. 武汉: 湖北中医药大学, 2018.
- [10] 蔡静颖. 模糊聚类算法及应用[M]. 北京: 冶金工业出版社, 2015.
- [11] 何婷,赵春兰,李屹,等. 基于FCM聚类的模糊综合评价方法[J]. 陕西师范大学学报(自然科学版), 2023, 51(1): 111-119.
- [12] 王贝伦. 机器学习[M]. 南京: 东南大学出版社, 2021.
- [13] 薛源. 基于智能优化算法的双聚类方法研究[D]. 南京: 南京邮电大学, 2022.
- [14] 周骏. 基于高维数据的双聚类算法研究与应用[D]. 南京: 南京理工大学, 2009.
- [15] 程龙,朱翠玲,闫奎坡,等. 张仲景运用生地黄特点[J]. 河南中医, 2021, 41(7): 977-980.
- [16] 李敏,何庆勇,陈胤峰,等. 张仲景方中茯苓量效关系的研究[J]. 中华中医药杂志, 2015, 30(12): 4311-4313.
- [17] 关艳楠,张晓霞. 犀角地黄汤治疗儿童过敏性紫癜血热妄行证临床疗效及对免疫功能影响[J]. 辽宁中医药大学学报, 2021, 23(5): 96-99.
- [18] 李蓝轩. 泻黄散合小蓟饮子治疗紫癜性肾炎血热伤络证的临床疗效观察[D]. 成都: 成都中医药大学, 2021.
- [19] 林玲. 小蓟饮子及其配伍对关木通毒性的减毒作用及机理研究[D]. 武汉: 湖北中医学院, 2007.
- [20] 郑植彬. 基于数据挖掘的国医大师应用清热类药物的剂量及相关研究[D]. 武汉: 湖北中医药大学, 2021.
- [21] 刘天洋,谈珍瑜,尤昭玲. “以通促养法”在宫腔粘连术后的应用[J]. 河南中医, 2022, 42(3): 392-396.
- [22] 丁影. 不同配伍比例对金银花-连翘药对中9个成分提取率的影响[J]. 中国药师, 2020, 23(11): 2142-2146.
- [23] 吴嘉瑞,金燕萍,王凯欢,等. 基于网络药理学的“金银花-连翘”药对作用机制分析[J]. 中国实验方剂学杂志, 2017, 23(5): 179-183.
- [24] 丁晓彦,林志军,王岱. 金银花-连翘药对的成分和药理作用研究进展[J]. 山东科学, 2019, 32(3): 36-41.
- [25] 高玉伟,王兴华,杨洪娟,等. 扶阳络法治疗脾肾阳虚型慢性肾脏病的疗效观察[J]. 中国中西医结合肾病杂志, 2020, 21(12): 1101-1103.
- [26] 向璐,张巧艳,赵琦明,等. 黄芪-当归化学成分、药理作用及临床应用的研究进展[J]. 中草药, 2022, 53(7): 2196-2213.
- [27] 尚伟庆. 黄芪当归合剂对肾病综合征患者肾功能、炎症机制及疗效的影响[J]. 中华中医药学刊, 2017, 35(5): 1339-1341.

(本文编辑 贺慧娥)