

DOI: 10.13288/j.11-2166/r.2021.02.007

基于方剂相似度的核心方药及其适应症挖掘方法研究 ——以失眠症为例

李新龙¹, 刘岩², 周莉¹, 王宁³, 杜元⁴, 周雪忠³, 刘保延², 何丽云^{2✉}

1. 北京中医药大学东直门医院, 北京市东城区海运仓胡同 5 号, 100700; 2. 中国中医科学院中医临床基础医学研究所; 3. 北京交通大学计算机与信息技术学院; 4. 辽宁中医药大学第一临床学院

摘要 目的 探索建立中医师辨证论治核心方药的数据挖掘方法。方法 以中医师辨证治疗失眠症的临床诊疗数据为例, 基于方剂相似度对目标人群进行模块划分, 构建核心方药及其有效人群特征的复杂网络, 采用子网筛选、分级核心子网挖掘算法, 挖掘中医师辨证治疗失眠症有效人群的核心方药及其适应症。结果 共采集 101 例失眠症患者 245 次诊疗数据, 其中有 78 例就诊 1 次以上, 主要疗效结局显示有效患者 68 例 (87.2%), 部分有效患者 2 例, 无效患者 8 例。78 例 (144 次) 患者诊疗数据进行复杂网络挖掘, 根据各次方剂相似度可划分为 2 个模块, 模块度为 0.184。其中, 模块一核心方由枳实、半夏、青礞石、竹茹、茯苓、黄连、橘红、大黄、龙齿、珍珠母组成, 其有效人群以舌红、苔黄腻、入睡困难、早醒、多梦为主要表现, 兼见大便干、口干口苦、情志不宁、头晕、脉弦实等。结论 基于方剂相似度的复杂网络挖掘方法, 能够初步揭示中医师辨证论治用药特点, 为辅助名老中医经验挖掘和传承提供方法。

关键词 核心方药; 方剂相似度; 复杂网络; 辨证论治; 数据挖掘; 失眠症

中医辨证论治的基本思想起源于《黄帝内经》, 形成于张仲景的《伤寒杂病论》, 后世医家对其不断丰富和完善, “辨证论治”一词于 1957 年被明确提出^[1]。辨证论治是一个医患交互、以医生为主体、个体化、动态的诊疗过程^[2], 是医生、特定患者群、治疗方案间不断完善达到和谐成熟的过程^[3]。作为中医学的基本特点^[4], 中医师个体化辨证论治及其理法方药是一种自治的知识体系, 是一类具有复杂概念层次和网状结构的独特知识体系。方剂是“辨证”认识的主要体现, 同时也是“论治”实践的主要途径, 作为辨证论治过程的枢纽, 方剂数据是研究辨证论治的重要着力点和突破口。本研究以中医师辨证论治失眠症临床数据为例, 基于方剂相似度, 探索构建挖掘分析中医师核心方药及其适应症的流程, 以期辅助名老中医经验的挖掘和继承提供方法学探索。

1 资料来源

1.1 数据来源

依托国家自然科学基金重点项目“阶梯递进

的辨证论治疗效评价方法研究”^[5-7]中河南省中医院队列数据, 在真实世界诊疗环境下, 采用观察性研究方法, 前瞻性连续采集中医师一段时间内辨证论治失眠症的临床诊疗数据, 并通过预处理策略^[8]完成数据预处理。

1.2 研究对象

1.2.1 患者纳入标准 1) 以失眠为主诉, 满足失眠症诊断标准 (DSM-V)^[9]; 2) 年龄 18 ~ 65 岁; 3) 同意参加本项临床观察。

1.2.2 患者排除标准 1) 合并尚未得到有效控制或者极为严重的疾病, 如各种原因引起的持续疼痛、恶性肿瘤、心脑血管疾病、神经内分泌疾病、神经肌肉疾病和消化道疾病等导致睡眠障碍者; 2) 严重抑郁、精神病患者; 3) 妊娠及哺乳期患者。

1.2.3 中医师纳入标准 1) 具备 20 年以上中医药治疗失眠症相关疾病临床经验, 已形成稳定的辨证论治失眠的经验; 2) 诊疗的失眠患者人群较为固定; 3) 有专门的科研助手协助课题开展; 4) 愿意参加本研究, 依从性较好。

1.3 治疗方案

针对患者临床表现, 中医师按照自身辨证论治知识体系进行诊治, 治疗药物及非药物治疗措施均

基金项目: 国家自然科学基金 (81673964, 81230086); 中央高校基本科研业务费专项 (2019-JYB-XJSJJ-018)

✉ 通讯作者: hely3699@163.com

由自医师所在医疗机构提供, 治疗频次及疗程由中医师根据患者病情变化动态调整。

1.4 结局指标

1.4.1 主要疗效指标及判定标准 参照国际研究^[8], 主要疗效指标由匹斯堡睡眠质量指数 (PSQI)^[9] 和睡眠效率 (SE)^[10] 判定: 有效: PSQI 和 SE 均改善 (PSQI 评分降低 ≥ 0 分, SE 提高 ≥ 0) 的前提下, PSQI 评分降低 ≥ 3 分或 SE 提高 $\geq 10\%$; 部分有效: PSQI 和 SE 其中之一有所改善 (PSQI 评分降低 ≥ 3 分或 SE 提高 $\geq 10\%$), 但另一个指标恶化; 无效: PSQI 评分降低 < 3 分且 SE 提高 $< 10\%$ 。

1.4.2 次要结局指标 1) 医师疗效总体评分 (CGIS); 2) 失眠严重程度 (ISI); 3) 睡眠时间: 总睡眠时间 (TST)、入睡潜伏时间 (TSO)、夜间觉醒次数 (TADS)、夜间觉醒时间 (WTDS)、早醒时间 (EMAIS); 4) 汉密尔顿焦虑量表 (HAMA)、汉密尔顿抑郁量表 (HAMD)。

1.4.3 安全性指标 记录试验期间发生的不良事件。

2 数据挖掘方法

数据挖掘过程包括: 目标人群的界定、模块划分、核心方药挖掘、核心方有效人群特征挖掘等, 具体流程见图 1。

2.1 目标人群界定

对纳入受试者, 首先, 剔除无法进行疗效判定患者, 包括只就诊 1 次患者、未报告主要结局指标患者等。其次, 因患者末次就诊后未进行随访, 末次就诊的疗效缺失, 因此末次就诊时的辨证论治诊疗数据并未纳入数据挖掘分析。

2.2 模块划分

采用 Jaccard 相似系数衡量两个诊次方剂相似

度, 分别对各诊次方剂两两比较, 计算其 Jaccard 系数, 公式如下:

$$J(F_1, F_2) = \frac{|F_1 \cap F_2|}{|F_1 \cup F_2|} = \frac{|F_1 \cap F_2|}{|F_1| + |F_2| - |F_1 \cap F_2|}$$

其中 F_1 、 F_2 代表两个诊次所用方剂, “ $F_1 \cap F_2$ ” 示两方剂中相同的中药种数, “ $F_1 \cup F_2$ ” 表示两方剂所有中药的种数, $J(F_1, F_2)$ 取值在 0 到 1 之间, $J(F_1, F_2)$ 越大, 表示两方剂相同中药的种数占二者所含中药种数越大, 即两方剂越相似。以患者不同诊次方剂的 Jaccard 相似度为划分依据, 采用 Gephi 软件 (Version 0.9.2) 快速模块性优化方法^[12] 对其进行模块划分。快速模块优化算法是一种基于模块度的迭代算法, 通过不断划分模块使得最终网络的模块度值不断增大, 该算法结合了局部优化与多层次聚类技术, 对稀疏网络具有线性时间复杂度, 同时可获得较高的聚类质量^[13]。该方法采用模块度来衡量模块划分的质量, 其绝对值在 0 ~ 1 之间, 绝对值越大, 模块划分质量越高^[14]。

2.3 核心方药挖掘

基于复杂网络挖掘方法对上述模块进行核心方挖掘, 挖掘算法通过 Liqueurice 软件 (Version 2012) 实现, 该软件基于 JUNG 可视化程包^[15] 进行网络的可视化及网络分析, 主要采用以下算法: 1) 子网筛选器: 依据边、节点度和间距等因素进行网络分析, 其中节点度的有效性已在多项中医药数据挖掘研究中得到证实^[16]。2) 分级核心子网挖掘算法^[17]: 从整体网络中挖掘多层次核心子网, 用于从数据库中挖掘共性知识 (核心处方、核心症状等) 及知识的主要调整 (核心处方加减、主要兼症) 等。

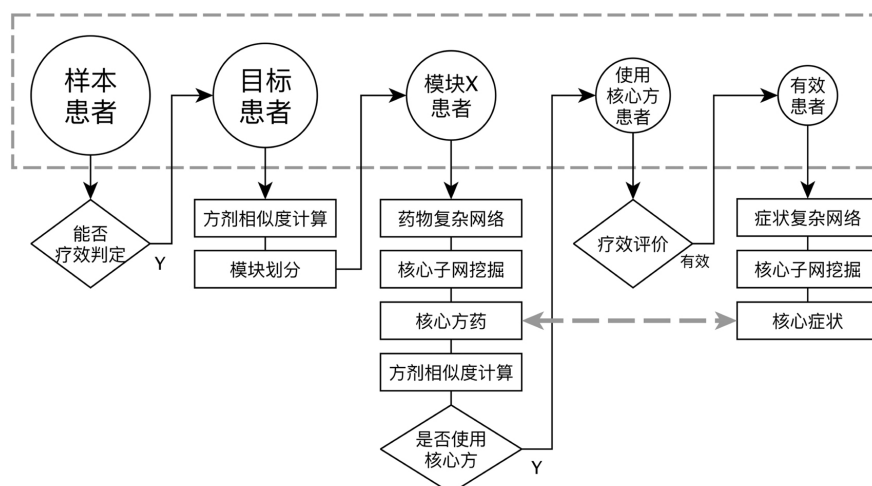


图 1 核心方药及其适应症数据挖掘流程图

2.4 核心方有效人群特征挖掘

1) 使用核心方人群筛选: 将模块中患者各诊次方剂与核心方药进行比较, 采用余弦相似度衡量其相似性, 余弦相似度计算公式如下:

$$\text{Cos}(F_1, F_2) = \frac{\sum_{i=1}^n (H_{1i} + H_{2i})}{\sqrt{\sum_{i=1}^n (H_{1i})^2} \times \sqrt{\sum_{i=1}^n (H_{2i})^2}}$$

其中 F_1 、 F_2 代表两个诊次所用方剂, H_{1i} 和 H_{2i} 分别代表两方剂中某味中药的有无, 当方剂中具有某一味中药时, $H_i = 1$; 不含有该中药时, $H_i = 0$ 。 $\text{Cos}(F_1, F_2)$ 取值范围在 0~1 之间, 值越大, 代表两个方剂越相似。对目标人群各诊次实际使用方剂进行两两比较并计算余弦相似度, 以其余弦相似度中位数作为判定标准。模块中与核心方药相似度大于等于上述中位数时, 判定该诊次使用了核心方药, 反之则没有使用核心方药。

2) 使用核心方有效人群判定: 上述使用核心方的人群中, 以主要疗效指标“有效”患者作为使用核心方的有效人群。

3) 核心方适应症挖掘: 采用复杂网络挖掘方法对使用核心方的有效人群的临床表现进行分析, 方法同“2.3 核心方药挖掘”。

3 结果

3.1 样本患者一般情况描述

自 2015 年 3 月 1 日至 2016 年 12 月 31 日, 共

采集 101 例失眠患者 245 诊次诊疗数据。101 例患者平均年龄 (43.76 ± 12.24) 岁, 其中女性 71 例 (70.3%); 就诊 1 次 23 例 (22.8%), 就诊 2 次 23 例 (22.8%), 就诊 3 次 44 例 (43.6%), 就诊 4 次及以上 11 例 (10.9%)。

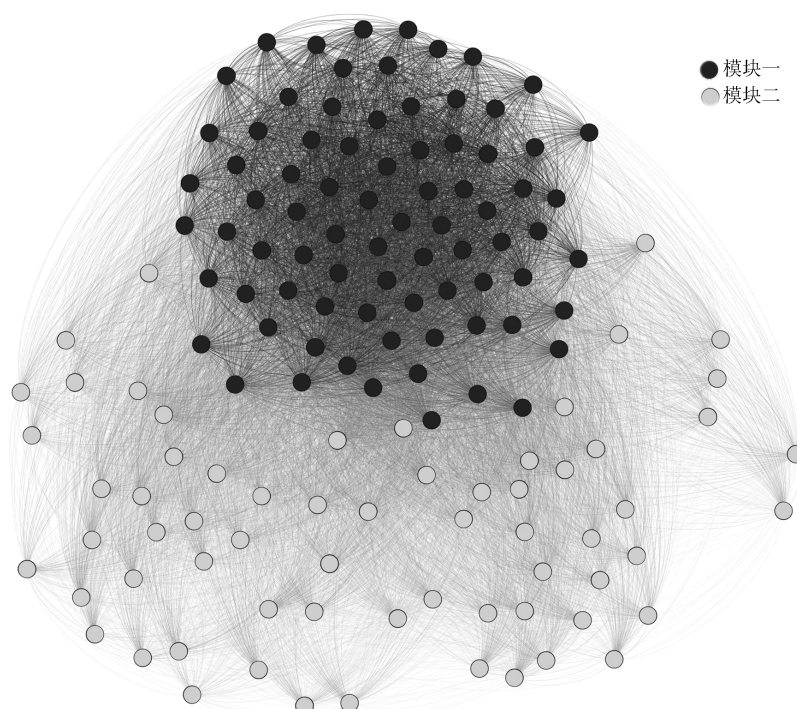
3.2 模块划分

根据目标人群界定标准, 共 78 例 (144 诊次) 患者诊疗数据可进行复杂网络挖掘, 采用 Jaccard 相似度的计算时采用了两两比较, 对 144 诊次所用中药方剂进行两两比较共得出 10296 条结果; 根据各诊次方剂相似度可划分为 2 个模块, 模块度为 0.184, 详见图 2, 可以看出模块一各方剂的紧密程度大于模块二, 反映出模块一方剂间相似性更好。其中, 模块一包括 42 例 (79 诊次) 患者, 各模块处方名称分布见表 1。

表 1 78 例失眠症患者治疗处方诊次及分布情况

模块	处方名称*	诊次数	比重/%	例数	比重/%
模块一	失眠 2 号方	79	100	42	100
模块二	失眠 1 号方	24	36.9	12	32.4
	失眠 4 号方	22	33.8	13	35.1
	失眠 5 号方	7	10.8	4	10.8
	失眠 3 号方	5	7.7	2	5.4
	失眠 2 号方	4	6.2	4	10.8
	柴胡疏肝散	2	3.1	1	2.7
	半夏白术天麻汤	1	1.5	1	2.7

注: * 处方名称由医师命名。



注: 图中节点代表某诊次的处方, 节点间的连线代表两处方的相似度

图 2 78 例失眠症患者治疗处方模块划分结果

表 2 78 例失眠症患者治疗前后结局指标比较

(分, $\bar{x} \pm s$)

结局指标	例数	治疗前	治疗后	治疗后 - 治疗前	t 值	P 值
WTDS	78	2.597 ± 0.165	1.390 ± 0.113	-1.208 ± 0.145	-8.342	<0.001
TST	78	3.776 ± 0.092	6.083 ± 0.124	2.308 ± 0.146	15.831	<0.001
TSO	78	99.615 ± 5.815	48.013 ± 4.285	-51.603 ± 4.905	-10.521	<0.001
TADS	78	101.389 ± 9.149	38.542 ± 5.300	-62.847 ± 7.484	-8.397	<0.001
SE	78	50.346 ± 1.554	76.423 ± 1.229	26.077 ± 1.807	14.429	<0.001
PSQI	78	16.269 ± 0.198	11.218 ± 0.395	-5.051 ± 0.415	-12.163	<0.001
ISI	78	15.948 ± 0.428	7.414 ± 0.333	-8.534 ± 0.538	-15.867	<0.001
HAMD	78	13.141 ± 0.697	6.769 ± 0.646	-6.372 ± 0.543	-11.743	<0.001
HAMA	78	16.654 ± 0.748	7.282 ± 0.732	-9.372 ± 0.727	-12.895	<0.001
EMAIS	78	5.654 ± 0.125	5.936 ± 0.103	0.282 ± 0.052	5.415	<0.001
CGIS	78	4.474 ± 0.068	3.051 ± 0.145	-1.423 ± 0.137	-10.373	<0.001

注: WTDS, 夜间觉醒时间; TST, 总睡眠时间; TSO, 入睡潜伏时间; TADS, 夜间觉醒次数; SE, 睡眠效率; PSQI, 匹斯堡睡眠质量指数; ISI, 失眠严重程度; HAMD, 汉密尔顿抑郁量表; HAMA, 汉密尔顿焦虑量表; EMAIS, 早醒时间; CGIS, 医师疗效总体评分。

3.3 目标患者疗效评价

共有 78 例就诊 1 次以上, 主要疗效结局显示: 有效患者 68 例 (87.2%), 部分有效患者 2 例 (2.6%), 无效患者 8 例 (10.3%), 患者治疗前后各结局指标变化详见表 2。

3.4 核心方挖掘

根据患者诊次创建中药的同质网络, 以模块一为例, 对 79 诊次中药方剂分别与挖掘出的核心方进行比较共得出 79 条结果, 挖掘其核心方及加减规律如图 3-4、表 3-4。图 3-4 中节点代表药物, 药物出现频次越高, 节点直径越大; 节点之间的连线表示药物的共现 (药对), 两药物共现次数越多, 连线越粗。

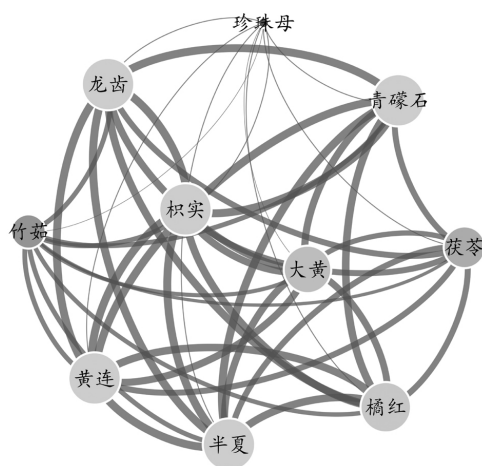


图 3 失眠症处方模块一核心方

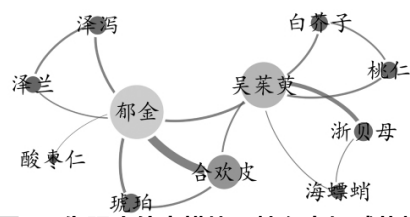


图 4 失眠症处方模块一核心方加减药物

表 3 失眠症处方模块一核心方药物用量及节点度

序号	中药	平均用量 /g	用量范围 /g	权重	度密度	节点度
1	枳实	10.2 ± 1.0	10 ~ 15	677	1.00	9
2	半夏	10.3 ± 1.7	10 ~ 20	677	1.00	9
3	青礞石	29.7 ± 1.6	20 ~ 30	677	1.00	9
4	竹茹	19.9 ± 1.2	10 ~ 20	561	0.89	8
5	茯苓	20.0 ± 0.0	20 ~ 20	583	0.89	8
6	黄连	12.0 ± 0.0	12 ~ 12	616	0.89	8
7	橘红	12.0 ± 0.0	12 ~ 12	610	0.89	8
8	大黄	5.9 ± 1.0	4 ~ 12	602	0.89	8
9	龙齿	20.6 ± 2.8	15 ~ 30	616	0.89	8
10	珍珠母	20.0 ± 0.0	20 ~ 20	183	0.33	3

注: 度密度, 节点与网络中其他节点的连接密度, 定义为网络中与该点连接的点的数量与所有可能与该点连接的点的数量的比值。

表 4 失眠症处方模块一核心方加减药物用量及节点度

序号	中药	平均用量 /g	用量范围 /g	权重	度密度	节点度
1	郁金	12.8 ± 1.3	12 ~ 15	81	0.60	6
2	吴茱萸	2.0 ± 0.0	2 ~ 2	68	0.60	6
3	合欢皮	15.8 ± 4.2	10 ~ 20	49	0.30	3
5	琥珀	2.4 ± 0.5	2 ~ 3	24	0.20	2
6	浙贝母	12.1 ± 2.6	8 ~ 15	24	0.20	2
7	泽泻	27.2 ± 5.5	12 ~ 30	23	0.20	2
8	白芥子	6.0 ± 0.0	6 ~ 6	21	0.20	2
9	桃仁	10.7 ± 3.0	4 ~ 15	21	0.20	2
10	泽兰	30.0 ± 0.0	30 ~ 30	20	0.20	2
11	海螵蛸	28.8 ± 3.5	20 ~ 30	16	0.20	2
12	酸枣仁	29.3 ± 2.6	20 ~ 30	7	0.10	1

注: 度密度, 节点与网络中其他节点的连接密度, 定义为网络中与该点连接的点的数量与所有可能与该点连接的点的数量的比值。

3.5 核心方有效人群特征挖掘

目标人群各诊次方药余弦相似度中位数为 0.65, 模块一共 40 例, 76 诊次与上述核心方余弦相似度 ≥ 0.65 , 其中有效患者 38 例, 73 诊次; 部分有效 1 例, 2 诊次; 无效 1 例, 1 诊次。根据患者诊次, 创建核心方有效人群症状体征的同质网

络, 目标人群模块一核心方人群主症及兼症如图 5-6、表 5-6。

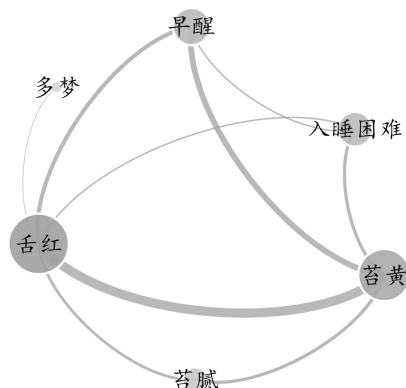


图 5 失眠症疾病模块一核心方主症

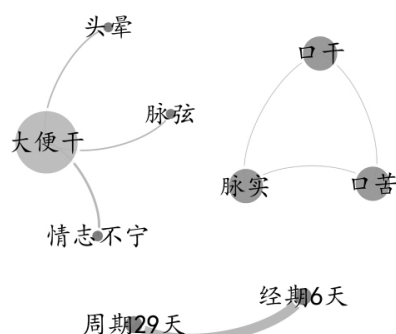


图 6 失眠症疾病模块一核心方兼症

表 5 失眠症核心方有效人群主症节点度

序号	主症	权重	度密度	节点度
1	舌红	275	1.00	5
2	苔黄	236	0.80	4
3	早醒	166	0.60	3
4	入睡困难	154	0.60	3
5	苔腻	107	0.40	2
6	多梦	48	0.20	1

注: 度密度, 节点与网络中其他节点的连接密度, 定义为网络中与该点连接的点的数量与所有可能与该点连接的点的数量的比值。

表 6 失眠症核心方有效人群兼症节点度

序号	兼症	权重	度密度	节点度
1	大便干	109	0.38	3
2	口干	70	0.25	2
3	口苦	70	0.25	2
4	脉实	70	0.25	2
5	经期 6 天	46	0.13	1
6	周期 29 天	46	0.13	1
7	情志不宁	37	0.13	1
8	脉弦	36	0.13	1
9	头晕	36	0.13	1

注: 度密度, 节点与网络中其他节点的连接密度, 定义为网络中与该点连接的点的数量与所有可能与该点连接的点的数量的比值。

4 讨论

本研究纳入中医师对失眠症的辨证论治, 已初步形成了其治疗失眠症的 5 个基本方, 即失眠 1 号方至失眠 5 号方。本研究期间纳入患者以使用失眠 2 号方为主, 有效患者比例达到 87.2%, 模块划分结果与医师类方分类具有较好的一致性。核心方药挖掘结果提示, 模块一核心方药由枳实、半夏、青礞石、竹茹、茯苓、黄连、橘红、大黄、龙齿、珍珠母为主组成, 即由礞石滚痰丸、泻心汤合黄连温胆汤化裁而来。方中青礞石长于坠痰下气, 定惊安神。大黄、黄连清胃之火, 助礞石导痰下行; 半夏、竹茹、橘红、茯苓、枳实健脾理气化痰, 以消痰源; 龙齿、珍珠母镇惊安神。诸药合用, 共奏化痰消痰、清热安神之用, 主治痰热扰神证。基于该方剂有效人群症状体征分析提示, 舌红、苔黄腻, 入睡困难、早醒、多梦是其主要表现, 兼见大便干, 口干口苦, 情志不宁, 头晕, 脉实、弦等症。以此方为基础, 易半夏、橘红为胆南星针对痰热病机, 调茯苓为茯神增强安神之效, 增生白术、薏苡仁, 减珍珠母, 祛邪而不伤正, 进一步形成了其治疗痰热扰神失眠的系列方剂^[13]。

本研究所采用的核心方挖掘方法, 在研究患者群上共分五步进行: 第一步, 根据能否进行疗效判定从样本人群中筛选出可进行疗效判定的目标患者群。第二步, 根据患者所使用的方剂相似度进将目标患者划分为不同模块。第三步, 分别构建各模块患者的药物复杂网络, 采用分级子网挖掘方法对各模块患者的核心方药子网进行分析。某一模块患者并不一定全部使用了所挖掘出的核心方, 因此对核心方和相应模块患者各诊次所采用的实际处方进行了两两比较, 进一步筛选出模块中使用核心方的患者。此外, 不同中医师临证处方过程中方剂的离散程度存在较大差异, 在判定患者某诊次是否采用核心方时, 将该中医师所有目标人群处方之间方剂相似度的中位数作为标准, 在一定程度上既形成了相对统一的标准, 也保证了符合医师自身的处方规律(本研究因目标人群样本量较小, 故采用中位数代表医师处方间相似度的平均水平, 当样本量足够大时, 也可采用相似度的平均数)。第四步, 依据主要疗效指标是否有效, 筛选使用核心方的有效患者。第五步, 构建使用核心方的有效患者的症状复杂网络, 挖掘其核心症状子网。对核心方药物与适应症之间的关系, 本研究仅依据二者在同一有效患

者同一诊次的共现作为依据,有待进一步研究具体分析核心方药症对应关系。

中药方剂的配伍过程具有无尺度复杂网络现象,无尺度网络的特点就是存在共性的核心网络结构,具体体现为中医医师在处方过程中对药物的使用具有比较集中的趋势。以无尺度网络的幂值为基础,周雪忠团队开发了挖掘其中共性或核心的药物配伍子网的具体算法^[19-20]。针对中医医师治疗某一病证的方剂数据进行药物配伍网络挖掘时,核心子网络则表达了其处方思路或首选处方结构,为了保证所挖掘核心处方临床有效,本研究预先界定数据挖掘人群,剔除无法进行疗效判定诊次的数据,只选择治疗有效患者进行核心方适应症的挖掘。为了提高数据挖掘效率,本研究首先对所挖掘处方进行了模块划分(聚类),考虑到中医医师辨证论治过程中的个体化和动态性特征,并未将中医证型诊断作为方剂模块划分的依据,而是采用了方剂相似度。所谓方剂相似度,即方剂之间的相似程度,可以用相似系数来表达,其计算依据和方法目前尚无统一标准,本研究采用了文本挖掘中常用的余弦相似度、Jaccard 相似度来衡量两个处方的相似度,从所划分模块结果与中医医师对处方的命名比较来看,已经满足了现有需求,但二者在实际应用中的具体差别以及方剂相似度的标准有待进一步研究。

综上,本研究初步形成了一套基于方剂相似度和复杂网络的中医医师核心方药挖掘流程,能够初步揭示和梳理中医医师的辨证论治用药特点,在一定程度上提高了对医师经验的传承和研究效率,能够辅助名老中医经验的挖掘和传承。

参考文献

- [1] 秦伯未. 中医“辨证论治”概说 [J]. 江苏中医, 1957 (1): 2-6.
- [2] JIANG L, LIU B, XIE Q, et al. Investigation into the influence of physician for treatment based on syndrome differentiation [J]. Evid Based Complement Alternat Med, 2013, 2013: 1-16. doi: 10.1155/2013/587234.
- [3] 何丽云, 李新龙, 刘岩, 等. 中医医师辨证论治知识图谱构建的思路与方法 [J]. 中医杂志, 2017, 58 (19): 1650-1653.
- [4] 孙广仁, 郑洪新. 中医基础理论 [M]. 北京: 中国中医药出版社, 2012.
- [5] 李新龙. 中医医师辨证论治失眠症知识图谱构建方法研究 [D]. 北京: 中国中医科学院, 2018.
- [6] LIU B. Individualized treatment with traditional Chinese medicine for the patients with insomnia disorder: a cohort study [EB/OL]. [2017-06-14] (2019-08-24).

https://www.clinicaltrials.gov/ct2/show/NCT02402959?term=sleep&rcv_d=14&show_rss=Y.

- [7] 李新龙, 刘岩, 王宁, 等. 中医师个体化辨证论治失眠诊疗数据预处理方法研究 [J]. 中国中医药信息杂志, 2017, 24 (12): 92-96.
- [8] American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5) [M]. Washington DC: American Psychiatric Association, 2013.
- [9] BUYSE DJ, GERMAIN A, MOUL DE, et al. Efficacy of brief behavioral treatment for chronic insomnia in older adults [J]. Arch Intern Med, 2011, 171 (10): 887-895.
- [10] BUYSE DJ, REYNOLDS CF, MONK TH, et al. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research [J]. Psychiatry Res, 1989, 28 (2): 193-213.
- [11] BUYSE DJ, ANCOLI-ISRAEL S, EDINGER JD, et al. Recommendations for a standard research assessment of insomnia [J]. Sleep, 2006, 29 (9): 1155-1173.
- [12] BLONDEL VD, GUILLAUME JL, LAMBIOTTE R, et al. Fast unfolding of communities in large networks [J/OI]. J Stat Mechanic Theor Experi. [2008-10-09] (2019-08-24). <https://doi.org/10.1088/1742-5468/2008/10/P10008>.
- [13] 孔维莲, 徐丽丽, 薛燕星, 等. 基于复杂网络的薛伯寿教授临床处方用药规律分析研究 [J]. 世界科学技术-中医药现代化, 2017, 19 (1): 55-62.
- [14] NEWMAN ME, GIRVAN M. Finding and evaluating community structure in networks [J]. Phys Rev E Stat Nonlin Soft Matter Phys, 2004, 69 (2 Pt 2): 026113. doi: 10.1103/PhysRevE.69.026113.
- [15] O'MADADHAI N, JOSHUA A, FISHER R, et al. JUNG-Java universal network/graph framework [EB/OL]. (2010-01-01) [2017-06-29]. <http://jung.sourceforge.net/index.html>.
- [16] ZHOU X, LIU B. Network analysis system for traditional Chinese medicine clinical data [C] // International Conference on Biomedical Engineering & Informatics. Tan-jing: IEEE, 2009.
- [17] ZHOU X, ZHANG X. Extraction of hierarchical core structures from traditional Chinese medicine herb combination network [M]. Beijing: Posts & Telecom Press, 2008: 262-267.
- [18] 赵景州, 候园园, 王松龄. 王松龄治疗痰热扰神失眠的经验 [J]. 中国中医药现代远程教育, 2018, 16 (8): 75-77.
- [19] 周雪忠. 中医临床数据仓库构建与临床数据挖掘方法研究 [D]. 北京: 中国中医科学院, 2007.
- [20] 周雪忠, 刘保延, 王映辉, 等. 复方药物配伍的复杂网络方法研究 [J]. 中国中医药信息杂志, 2008, 15 (11): 98-100.

Prescription Similarity-Based Analysis of Core Formulas and Medicinals and Related Indications through Data Mining: Taking Insomnia as an Example

LI Xinlong¹, LIU Yan², ZHOU Li¹, WANG Ning³, DU Yuan⁴, ZHOU Xuezhong³, LIU Baoyan², HE Liyun²

1. Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing, 100700; 2. Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences; 3. School of Computer and Information Technology, Beijing Jiaotong University; 4. The first Affiliated Hospital of Liaoning University of Traditional Chinese Medicine

ABSTRACT Objective To explore the data mining method for establishing the core formulas and medicinal herbs for traditional Chinese medicine (TCM) doctors to give treatment based on syndrome differentiation. **Methods** By taking insomnia as an example, the data mining method was explored to analyze the clinical data in TCM practice. Based on the similarity of prescriptions, the target population was identified through classification models; the complex network of core formulas and medicinals, and the characteristics of people benefited from the prescription was constructed; subnetwork screening as well as a hierarchical core subnetwork mining algorithm were adopted to explore the core formulas and medicinals for insomnia patients who were benefited from the TCM treatment. **Results** Data from 101 insomnia patients involving 245 hospital visits were collected, among which 78 patients visited the hospital more than once. Sixty-eight patients (87.2%) showed effective result, while two as partially effective, and eight as non-effective. A complex network of clinical data of 78 patients (144 visits) were constructed for data mining. Target population was divided into two modules according to the similarity of prescriptions, and the module degree was 0.184. The core formula of module one was composed of Zhishi (*Fructus Aurantii Immaturus*), Banxia (*Rhizoma Pinelliae*), Qingmengshi (*Lapis Chloriti*), Zhuru (*Caulis Bambusae in Taenia*), Fuling (*Poria*), Huanglian (*Rhizoma Coptidis*), Juhong (*Exocarpium Citri Rubrum*), Dahuang (*Radix et Rhizoma Rhei*), Longchi (*Dens Draconis*), and Zhenzhumu (*Concha Margaritiferae Usta*); it was effective for patients who mainly had a red tongue, and yellow and greasy coating; had difficulty falling asleep; waked up early; and had many dreams; certain patients may present dry stool, dry mouth, bitter taste, uneasy feelings, dizziness, wiry and excess pulse. **Conclusion** The prescription similarity-based complex network data mining method can be used to initially reveal the characteristics of using drugs based on syndrome differentiation by TCM doctors, the result of which may further facilitate the data mining and inheritance of TCM famous doctors' experience.

Keywords core formulas and medicinals; prescription similarity; complex networks; treatment according to syndrome differentiation; data mining; insomnia

(收稿日期: 2019-08-24; 修回日期: 2020-02-12)

编辑: 柴倩云