

基于关联规则和复杂系统熵聚类的糖尿病用药规律研究

尹晓阳, 张定棋, 唐翠娟, 余方方, 胡卫*
(三峡大学医学院, 湖北宜昌 443002)

[摘要] **目的:**基于中医传承辅助平台(TCMISS)探讨糖尿病的中药处方用药规律。**方法:**从文献中收集近10余年治疗糖尿病的中药处方,采用关联规则Apriori算法,进行基于改进的互信息法的药物间关联度分析;并以关联度分析结果为基础,依据相关度与惩罚度约束,进行基于复杂系统熵聚类的药物核心组合分析;用无监督熵层次聚类算法,获得治疗糖尿病新处方;根据结果绘制出用药关联规则网络图、新方的核心组合药物网络图。上述操作基于TCMISS V1.1完成。**结果:**对筛选出的110个处方进行分析,明确了处方中药物的使用频次规律,药物之间的关联规则,挖掘出16个核心组合和8首新处方。**结论:**基于关联规则和聚类分析的数据挖掘手段,适合中医方剂用药规律的挖掘、分析和总结。以补益气阴为主并灵活配伍清热药、活血化瘀药,同时根据不同证型考究药物的配伍组合对提高糖尿病的疗效有重要意义。

[关键词] 糖尿病; 用药规律; 关联规则; 聚类算法

[中图分类号] R287.6 **[文献标识码]** A **[文章编号]** 1005-9903(2015)23-0212-04

[doi] 10.13422/j.cnki.syfjx.2015230212

Analysis on Medication Rule in Herbal Prescriptions for Diabetes Based on Apriori and Entropy Clustering Algorithm YIN Xiao-yang, ZHANG Ding-qi, TANG Cui-juan, YU Fang-fang, HU Wei* (Medical College of China Three Gorges University, Yichang 443002, China)

[Abstract] **Objective:** To analyze and excavate the medication rule in herbal prescriptions for diabetes by using Traditional Chinese Medicine in Heritance Support System (TCMISS). **Method:** Herbal prescriptions for diabetes in recent 10 years were collected from papers, and Apriori algorithm was used to analyze correlation between herbs, which was based on improved mutual information method. And according to the relevant constraints and punishment degree, core combination of herbs was analyzed by complex system entropy clustering method based on the correlation analysis between drugs. New prescriptions were acquired by using unsupervised clustering algorithm. The network diagrams of medicine association rules and core combination herbs in the new prescriptions were drawn based on the analysis results. The above operations were completed based on TCMISS V1.1. **Result:** The frequency of each herb and the association rules among the herbs were found out, 16 core combinations and 8 new prescriptions were excavated from the database about the analysis of 110 medicine prescriptions of diabetes. **Conclusion:** The data mining method based on Apriori algorithm and cluster analysis is suitable for mining, analysis and summary of the medication rules in herbal prescriptions. Herbs of nourishing *qi* and *yin*, combined with herbs for heat-clearing and activating blood circulation in accordance with different types is important to improve the curative effect on diabetes.

[Key words] diabetes; medication rules; association rules; clustering algorithm

糖尿病(diabetes mellitus, DM)是一种以高血糖为主要标志的内分泌代谢性疾病,现代研究发现,其生理病理改变

是胰岛素分泌绝对或相对不足,引发高血糖^[1]。流行病学资料表明,全球糖尿病患者已经超过4亿人,每7s就有一人

[收稿日期] 20150403(010)

[基金项目] 三峡大学科创基金项目(2015YCX047)

[第一作者] 尹晓阳,从事内分泌疾病的中药用药规律研究, Tel:13972565833, E-mail:2596585344@qq.com

[通讯作者] *胡卫,博士,副教授,从事中药治疗肿瘤与代谢性疾病的物质基础与分子机制研究, Tel:13972540162, E-mail:huwei-99@21cn.com

死于糖尿病;而我国糖尿病形势更为严峻,目前约有 1.139 亿糖尿病患者,4.934 亿糖尿病前期人群,且呈快速增加态势^[2]。糖尿病将给全球尤其是发展中国家带来非常沉重的社会和经济负担^[3-4],因此已成为中西医共同面对的重大公共卫生危机。近年来,以中医理论为指导的中药疗法,在对 2 型糖尿病的临床观察中被证实有较好疗效^[5-7]。中医药治疗糖尿病的实验研究在单味中药筛选、复方研究、中药有效成分研究以及中医药治疗糖尿病机制研究、临床辨证论治等方面都取得了较大进展^[8]。目前也有单味中草药治疗糖尿病的系统综述^[9]。然而,中药治疗糖尿病的用药和配伍规律并未见到系统研究。由于数据挖掘技术能提高方剂在疾病治疗中的用药规律分析^[10],因此本研究以 1990—2014 年公开发表文献中治疗糖尿病的中医药处方为研究对象,旨在通过关联规则 Apriori 算法和复杂系统熵聚类无监督数据挖掘手段,分析处方中药物使用频次、关联规则及其内在组方规律,以期为临床治疗糖尿病提供较客观的依据和参考。

1 资料和方法

1.1 资料来源及检索方法 中国知网数据库(CNKI, 1990—2014),维普中文科技期刊全文数据库(VIP, 1990—2014),万方中文期刊数据库(1990—2014)。中文数据库中以“中医药”并含“糖尿病”为检索词进行主题或摘要检索,再严格根据文献纳入及排除标准进行人工筛选。

1.2 纳入标准(文献质量) 1990 年至 2014 年间公开发表的相关论文,标准包括①西医诊断标准符合 WHO 1999 年修订的 2 型糖尿病诊断标准^[11];中医诊断标准符合《中药新药临床研究指导原则》^[12]中糖尿病的诊断标准及指南,符合对糖尿病进行证候分型的标准;②临床随机对照研究;③病例报告;④中医名家用药经验。

1.3 排除标准 ①单纯为 1 型糖尿病;②中药仅作为辅助治疗;③文献数据存在明显错误或重复发表的雷同文献;④动物实验;⑤综述文献。

1.4 录入与核对 将上述筛选后的处方用药录入中医传承辅助平台(TCMISS) V1.1,数据录入及统计分析过程由双人负责审核。

1.5 软件支持 本研究相关研究操作基于中国中医科学院中药研究所和中国科学院自动化研究所联合开发的 TCMISS V1.1 完成。

1.6 数据分析 通过 TCMISS V1.1 系统中“数据分析”模块中“方剂分析”功能,进行药物频次统计、组方规律分析、新方分析,统计过程使用改进的互信息法、复杂系统熵聚类、无监督的熵层次聚类算法,分析所得药物组合规则,绘制用药关联规则网络图、新方的核心组合药物网络图,实现成果的网络化、可视化展示。

2 结果

2.1 用药频次 对治疗糖尿病的 110 首方剂中 173 个药物的频次进行统计,使用频次在 10 次以上的药物有 38 味,统计结果见表 1。使用频次前 10 位的是黄芪、山药、丹参、地黄、白术、当归、天花粉、枸杞子、茯苓、山茱萸。

表 1 治疗糖尿病的处方中频次 > 10 的药物

Table 1 Frequency of medicinal herbs in prescriptions for diabetes (Frequency > 10)

| 中药名称 | 频数/次 | 中药名称 | 频数/次 | 中药名称 | 频数/次 |
|------|------|------|------|------|------|
| 黄芪 | 88 | 川芎 | 26 | 陈皮 | 14 |
| 山药 | 57 | 玄参 | 26 | 知母 | 14 |
| 丹参 | 56 | 党参 | 24 | 白芍 | 14 |
| 地黄 | 50 | 太子参 | 23 | 苍术 | 14 |
| 白术 | 35 | 牛膝 | 21 | 牡丹皮 | 13 |
| 当归 | 31 | 泽泻 | 20 | 益母草 | 13 |
| 天花粉 | 30 | 熟地黄 | 18 | 赤芍 | 13 |
| 枸杞子 | 30 | 黄连 | 16 | 半夏 | 11 |
| 茯苓 | 30 | 五味子 | 16 | 石斛 | 11 |
| 山茱萸 | 30 | 红花 | 16 | 鸡内金 | 10 |
| 甘草 | 28 | 大黄 | 15 | 地龙 | 10 |
| 葛根 | 28 | 桃仁 | 15 | 人参 | 10 |
| 麦冬 | 28 | 黄精 | 15 | | |

2.2 基于关联规则分析的组方规律分析 以支持度个数(即在所有药物中同时出现的次数)设定为 20,置信度设定为 0.9,按照药物组合出现频次由高到低排序,前 3 组分别是“山药,黄芪”,“丹参,黄芪”,“黄芪,地黄”,见表 2。分析药物关联规则,根据置信度排列,顺序结果见表 3。关联规则网络图见图 1。

表 2 处方中中药组合频次(支持度 20,置信度 0.9)

Table 2 Frequency of medicinal herb combination of prescriptions for diabetes (support ≥ 20, confidence ≥ 0.9)

| 中药组合 | 频次 | 中药组合 | 频次 |
|------------|----|------------|----|
| 山药, 黄芪 | 50 | 当归, 黄芪 | 26 |
| 丹参, 黄芪 | 44 | 天花粉, 黄芪 | 25 |
| 黄芪, 地黄 | 39 | 茯苓, 黄芪 | 24 |
| 山药, 丹参 | 32 | 山茱萸, 山药 | 23 |
| 白术, 黄芪 | 29 | 川芎, 黄芪 | 23 |
| 山药, 丹参, 黄芪 | 29 | 麦冬, 黄芪 | 22 |
| 丹参, 地黄 | 28 | 山药, 黄芪, 地黄 | 22 |
| 山茱萸, 黄芪 | 27 | 丹参, 黄芪, 地黄 | 22 |
| 枸杞子, 黄芪 | 27 | 山药, 白术 | 21 |
| 山药, 地黄 | 26 | 山药, 茯苓 | 21 |

2.3 基于熵聚类的方剂组方规律分析

2.3.1 基于改进的互信息法的药物间关联度分析并基于复杂系统熵聚类的药物核心组合分析 依据方剂数量并结合不同参数提取数据的预读,设置相关度为 8,惩罚度为 4,按照相关度与惩罚度约束,进行聚类分析,得到方剂中两两药物间的关联度。以改进的互信息法的药物间关联度分析结果为基础,按照相关度与惩罚度约束,基于复杂系统熵聚类,

表 3 处方中药物组合关联规则(置信度 > 0.9)

Table 3 Association rules of prescriptions for diabetes (support ≥ 20, confidence ≥ 0.9)

| No. | 中药组合规则 | 置信度 |
|-----|-----------------|---------|
| 1 | 山药, 当归->黄芪 | 1 |
| 2 | 山茱萸, 丹参->黄芪 | 1 |
| 3 | 天花粉, 枸杞子->黄芪 | 1 |
| 4 | 山茱萸, 山药, 丹参->黄芪 | 1 |
| 5 | 山药, 丹参, 枸杞子->黄芪 | 1 |
| 6 | 山茱萸, 丹参, 地黄->黄芪 | 1 |
| 7 | 山药, 天花粉->黄芪 | 0.95 |
| 8 | 丹参, 枸杞子->黄芪 | 0.944 4 |
| 9 | 山茱萸, 地黄->黄芪 | 0.941 2 |
| 10 | 黄精->黄芪 | 0.933 3 |
| 11 | 大黄->黄芪 | 0.933 3 |
| 12 | 川芎, 丹参->黄芪 | 0.933 3 |
| 13 | 山药, 太子参->黄芪 | 0.928 6 |
| 14 | 山茱萸, 山药, 地黄->黄芪 | 0.928 6 |
| 15 | 益母草->丹参 | 0.923 1 |
| 16 | 山药, 川芎->黄芪 | 0.923 1 |
| 17 | 山茱萸, 枸杞子->山药 | 0.923 1 |
| 18 | 山茱萸, 枸杞子->黄芪 | 0.923 1 |
| 19 | 山药, 枸杞子, 地黄->黄芪 | 0.923 1 |
| 20 | 丹参, 枸杞子, 地黄->黄芪 | 0.923 1 |

表 4 基于复杂系统熵聚类的治疗糖尿病的药物核心组合节选

Table 4 Core combinations for new prescriptions for diabetes by cluster analysis

| No. | 核心组合 | No. | 核心组合 |
|-----|---------------|-----|-----------------|
| 1 | 泽兰, 黄芩, 紫苏梗 | 9 | 泽兰, 黄芩, 虎杖 |
| 2 | 苍术, 鸡内金, 荔枝核 | 10 | 鸡内金, 川芎, 当归 |
| 3 | 甘草, 山药, 枳壳 | 11 | 甘草, 山药, 山茱萸 |
| 4 | 地龙, 川芎, 赤芍 | 12 | 地龙, 麦冬, 天花粉 |
| 5 | 熟地黄, 山茱萸, 牡丹皮 | 13 | 枸杞子, 山茱萸, 牡丹皮 |
| 6 | 白术, 地黄, 玄参 | 14 | 白术, 党参, 三七粉 |
| 7 | 知母, 大黄, 天花粉 | 15 | 知母, 麦冬, 天花粉, 玄参 |
| 8 | 太子参, 党参, 人参 | 16 | 太子参, 党参, 三七粉 |

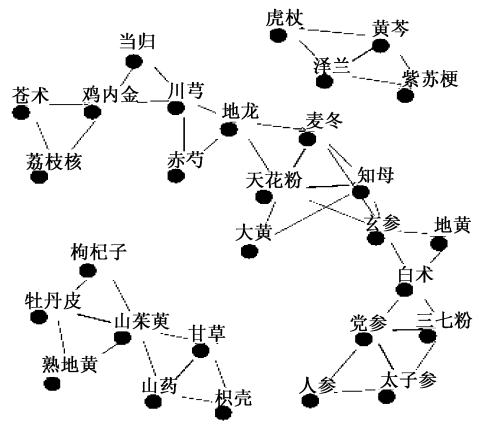


图 2 新方的药物核心组合网络展示

Fig. 2 Core combinations for new prescriptions for diabetes by cluster analysis

表 5 基于熵层次聚类的治疗糖尿病新处方

Table 5 New prescriptions for diabetes by cluster analysis

| No. | 候选新处方 |
|-----|----------------------|
| 1 | 泽兰, 黄芩, 紫苏梗, 虎杖 |
| 2 | 苍术, 鸡内金, 荔枝核, 川芎, 当归 |
| 3 | 甘草, 山药, 枳壳, 山茱萸 |
| 4 | 地龙, 川芎, 赤芍, 麦冬, 天花粉 |
| 5 | 熟地黄, 山茱萸, 牡丹皮, 枸杞子 |
| 6 | 白术, 地黄, 玄参, 党参, 三七粉 |
| 7 | 知母, 大黄, 天花粉, 麦冬, 玄参 |
| 8 | 太子参, 党参, 人参, 三七粉 |

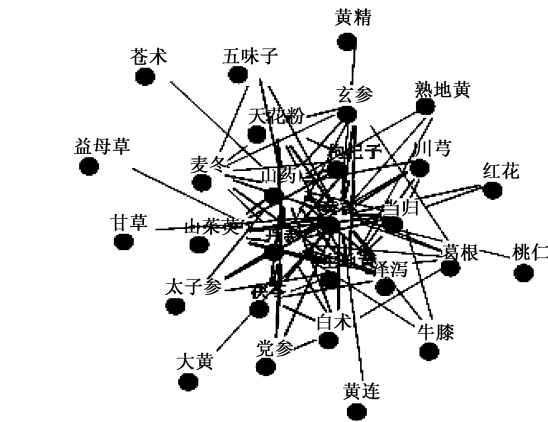


图 1 药物关联规则网络展示(支持度 11, 置信度 0.9)

Fig. 1 Association rules of medicinal herbs in prescriptions for diabetes (support ≥ 11, confidence ≥ 0.9)

演化出由 3~4 味药构成的核心组合 16 组, 节选结果见表 4。核心组合药物网络图见图 2。

2.3.2 基于无监督的熵层次聚类的新处方分析 在上述药物核心组合提取的基础上, 运用无监督的熵层次聚类算法分析获得新处方 8 个, 结果见表 5。

3 讨论

将数据挖掘如关联规则、聚类分析技术应用于中医学领域的研究, 是中医学现代化研究的重要组成部分。本研究采用数据挖掘技术分析糖尿病的治疗规律和思维模式, 挖掘出潜在而有效的用药规律不仅可辅助医生潜方用药以提高临床疗效, 还能为中医药治疗糖尿病的临床和实验研究提供目标和思路, 减少盲目性, 缩短研究周期, 提高科研效率和质量^[10]。

本研究中用药频次显示, 临床上治疗糖尿病常用的药物有补气滋阴药、清热药、活血化瘀药、利水渗湿药等(表 1)。糖尿病属于中医“消渴”范畴, 其病机主要以“气阴两虚为本, 燥热为

标,瘀血为患”,故用药频次显示以补气和养阴为主,再辅以清热、活血和利水渗湿药。譬如黄芪,味甘、微温,归脾、肺经,功能补气健脾、益卫固表,尤擅补益脾肺二气,是治疗糖尿病使用最多的一味。研究表明:黄芪多糖具有调节老年早期糖尿病肾病患者免疫功能和减少炎症反应的作用,能有效改善患者肾功能^[13]。山药益气养阴、补脾肺肾,可兼治三脏由于气阴两虚而引起的肺燥、胃热、肾虚表现出的上消、中消和下消。山药多糖通过提高糖代谢关键酶活性从而降低血糖^[14]。丹参为活血化瘀常用药,丹参多酚酸盐能显著降低早期糖尿病肾病患者的炎症反应,改善肾血管内皮功能^[15]。

基于关联规则分析的组方规律显示,常用药物组合常常发挥相须、相使的作用。如黄芪和山药,黄芪能补气健脾合山药补脾肺肾之气阴从而加强补气养阴的力度。同时显示中医治疗糖尿病的用药核心组成主要为黄芪,山药,丹参,白术,地黄,山茱萸,枸杞子,天花粉,当归,茯苓,川芎,麦冬,甘草,葛根等。该用药关联规则网络图呈密集且近似对称图形,显示了入选药物间相关性紧密。纵观此核心处方:方中黄芪主入脾肺,行补中益肺而奏润肺健脾之效;山药既补脾肺肾之气又可养脾肺肾之阴,为治疗消渴之要药;丹参既可活血祛瘀,又可凉血止痛合葛根更可清热生津止渴;地黄清热生津养阴更加天花粉善清肺胃积热、生津止渴以治内热消渴;白术、茯苓既可健脾又可化湿,常用于湿热蕴脾型糖尿病;麦冬善养肺胃二阴配伍枸杞子、山茱萸滋补肝肾二阴;当归既可补血又可活血化瘀止痛合川芎活血行气止痛,尤适用于血瘀络型糖尿病;更加炙甘草益气补中,健运脾胃,缓急止痛,兼调和药性以为使。

基于复杂系统熵聚类的药物核心组合分析结果显示:治疗糖尿病常用的药物核心组合共 16 组,以此为基础的基于无监督熵层次聚类的新处方分析结果显示:挖掘出的治疗糖尿病候选新处方 8 个,部分可能具有重要临床意义。这些新处方的发现不仅佐证了治疗糖尿病时益气与养阴并重(新方 6,8),养阴与清热共用(新方 4,7),还提示了益气与活血化瘀同施是近年来糖尿病用药的一大趋势^[16-18](新方 4)。这些潜在的规律在经过进一步临床实践的验证很可能得到广泛的应用。

综上所述,本研究采用基于关联规则算法和无监督的复杂系统熵聚类的数据挖掘手段,获得了既往传统方剂整理未获得的具有潜在应用价值的新知识、新信息,研究所获取的组方规则、用药核心组合和新处方,可为临床治疗糖尿病提供参考。但必须强调的是:数据挖掘方法亦有其局限性,应用关联规则和聚类分析研究得到的组方规则、核心组合和新处方等用药规律还需进一步临床验证,需结合中医药理论与实践综合分析、评价。希望本研究的成果能为中医药在糖尿病的基础研究和临床治疗方面提供参考。

[参考文献]

[1] Rosa G, Mingrone G, Manco M, et al. Molecular mechanisms of diabetes reversibility after bariatric surgery[J]. *Inter J Obes*, 2007, 31(9):1429-1436.

[2] Xu Y, Wang L, He J, et al. Prevalence and control of diabetes in Chinese adults[J]. *JAMA*, 2013, 310(9): 948-959.

[3] Greenfield J R, Campbell L V. Relationship between inflammation, insulin resistance and type 2 diabetes: 'cause or effect'? [J]. *Curr Diabetes Rev*, 2006, 2(2):195-211.

[4] 胡善联,刘国恩,许樟荣,等.我国糖尿病流行病学和疾病经济负担研究现状[J]. *中国卫生经济*, 2008, 27(8):5-8.

[5] 孙毅,张勉之.补肾活血法治疗糖尿病肾病临床期的疗效观察[J]. *北京中医药大学学报*, 2013, 36(5): 353-356.

[6] 张青,李琰,陈磊.黄连素对 2 型糖尿病及其并发症的治疗作用及相关机制研究进展[J]. *中国中药杂志*, 2015, 40(9):1660-1665.

[7] 魏军平,高嘉良.益气活血法治疗糖尿病肾病的临床研究进展[J]. *世界中西医结合杂志*, 2014, 9(4): 446-448.

[8] 蔡仲德,吴杰,朱惠荣.中医药治疗糖尿病研究概况[J]. *中国实验方剂学杂志*, 2002, 8(6):56-59.

[9] 李青,张玲.中草药治疗糖尿病的系统综述[J]. *中国医院药学杂志*, 2012, 32(11):882-883,893.

[10] Tari L, Anwar S, Liang S, et al. Discovering drug-drug interactions: a text-mining and reasoning approach based on properties of drug metabolism [J]. *Bioinformatics*, 2010, 26(18):1547.

[11] Mogensen C E. Management of early nephropathy in diabetic patients[J]. *Annu Rev Med*, 1995, 46(3):79-93.

[12] 中华人民共和国卫生部.中药新药临床研究指导原则[S].北京:中国医药科技出版社,2005:233-237.

[13] 邓海鸥,林凯,黎映兰,等.黄芪多糖对老年早期糖尿病肾病患者 TNF- α 、IL-6 和免疫功能的影响[J]. *中药材*, 2014, 37(4):713-716.

[14] 杨宏莉,张宏馨,李兰会,等.山药多糖对 2 型糖尿病大鼠降糖机理的研究[J]. *河北农业大学学报*, 2010, 33(3):100-103.

[15] 卢伟波,杨丕坚,李舒敏,等.丹参多酚酸盐对早期糖尿病肾病患者炎症细胞因子及肾血管内皮功能的影响[J]. *中国实验方剂学杂志*, 2014, 20(2):184-187.

[16] 殷丽平,杜联,谢春光,等.益气养阴、活血化瘀法对 2 型糖尿病血管炎症患者 TNF- α 影响的临床研究[J]. *甘肃中医*, 2010, 23(1):17-19.

[17] 黄苏萍,邱彩霞,林妙娴,等.活血化瘀药干预糖尿病心血管风险因素的系统评价[J]. *中华中医药杂志*, 2011, 26(6):1410-1415.

[18] 杨照华,夏进娥,张海燕.益气补肾化痰法治疗早期糖尿病肾病 32 例[J]. *中医杂志*, 2011, 52(3): 243-244.