

利用文本挖掘技术探索 中医药治疗疾病的用药规律*

□谭 勇** (中国中医科学院中医临床基础医学研究所 北京 100700)

郭洪涛 (上海中医药大学 上海 201203)

郑 光 张 弛 杨 静 吕 诚 查青林 姜 森 吕爱平**
(中国中医科学院中医临床基础医学研究所 北京 100700)

摘 要:用药规律研究是中医临床研究的核心。在当今中西医融合、高度信息化的时代,利用文本挖掘技术对中医用药规律进行挖掘分析,具有现实可行性和积极意义。中医用药规律包括“证-药”、“药-药”和“药-效”三方面的关系,其中“药-药”关系是中心和纽带。以现代医学疾病为研究对象,以探索中医治疗疾病的核心药物及其组合关系为研究先导,在不预设中医学推导规则的前提下,对海量文献数据从不同层次和角度进行挖掘分析,并对不同疾病的挖掘结果进行对比,从而探寻到能用现代科学语言解释、便于国际交流的用药规律。既往的研究表明,利用文本挖掘技术进行中医用药规律研究具有客观性、可行性和先进性。

关键词: 中医药 用药规律 文本挖掘

doi: 10.3969/j.issn.1674-3849.2010.05.030

自《黄帝内经》以来,中医学理论和实践在不断归纳、总结、升华中完善,历代医家在临床实践中为提高疗效,孜孜不倦地探索。中医用药规律是探索的核心。中医临床经典著作《伤寒杂病论》、《温热论》、《医宗金鉴》、《临证指南医案》等都是探索用药规律的结晶。这些著作对指导临床实践、提高疗效

具有重要意义。当今时代,全世界累以千万计的科研和医务人员在研究、使用中医,借助现代传媒技术,大量的中医药信息得以交流共享。如何充分利用这样的信息平台探索用药规律,是值得我们关注并亟待解决的问题。文本挖掘是新兴的交叉学科技术,能够处理海量的文本信息,用其探索中医用药规律有望取得一定的突破。

收稿日期: 2010-10-13

修回日期: 2010-10-15

* 科学技术部“十一五”国家科技支撑计划项目(2006BAI04A10):基于二次临床研究的中医药治疗类风湿性关节炎的临床评价,负责人:吕爱平;国家科学技术部创新方法学专项项目(2008IM020900):中医药科学方法研究——中医药科学方法总论研究,负责人:吕爱平;国家自然科学基金杰出青年项目(30825047):中西医结合基础——疾病证候分类研究,负责人:吕爱平;国家自然科学基金项目(30902003):基于生物网络构建技术的中药寒热药性科学基础,负责人:姜森。

** 联系人:谭勇,助理研究员,主要研究方向:中药毒性评价研究,Tel:010-64014411-3402,E-mail:tcmt@yaho.com.cn;吕爱平,博士生导师,研究员,主要研究方向:疾病证候分类,E-mail:lap64067611@126.com。

[World Science and Technology/Modernization of Traditional Chinese Medicine and Materia Medica] 823

一、既往中医用药规律研究的启示

近代以降,由于各种信息传媒和通量的不断加大,中医用药规律研究呈现多元化的趋势。在临床实践中,师徒口传相授、笔录总结是最基本的研究方法。归纳总结前人著作,结合临床验证,也是探索中医用药规律途径之一。随着信息仓储和检索技术的不断进步,用人工方法集成与某个疾病相关的中医诊治文献,对其进行分析和总结,可以找到一定的中医用药规律。计算机技术的兴起,使以往的纸质文本得以电子化和集成化,这为探索中医用药规律提供了更为便捷的条件。随着大量医学文献不断涌现,充分利于文献数据,建立数据库,用现代数理统计方法探索中医用药规律是目前研究的主要方法。

利用文献数据进行名老中医用药规律研究及针对症状的用药规律研究是目前中医用药规律研究的热点^[1-2]。前者对于继承和弘扬老一辈中医学术思想具有重要意义,后者为症状的治疗提供了用药方面的指导。与累以千万计的文献比较,名老中医以一人之力诊治疾病的数量有限,取得良好疗效并有完整记载的文献较少,更何况治疗单个病种的有效文献就更少。对少量的文献进行统计分析,过程中很难有效地降低各种干扰和偏倚,从而所得到的用药规律缺乏普遍性。另外,名老中医用药规律研究过程中,预设的一些用药规则,也限制了隐含的有用知识的发现。症状具有主观性,轻重程度难于量化,不同疾病可能出现同一个症状,而同一症状的产生本质不同,因此,针对症状的用药差别必然加大,加之症状疗效又难于评估,所以,针对症状的用药规律研究值得进一步商榷。总之,目前中医用药规律研究存在一定的问题,亟需研究思路和方法的更新。

二、中医用药规律研究的新思路

1. 以西医疾病为研究对象

西医疾病具有诊断明确、治疗技术统一规范、国际公认的特点。以其为对象开展中医用药规律研究有利于研究成果与现代医学接轨,有利于推动和深化中医用药理论和技术走向世界。而且能为开发国际化的中医新药研究提供线索和目标,有利于减少盲目性,缩短研究周期^[3]。

2. 以探索核心药物及其配伍关系作为研究的先导

中医诊治疾病遵行“证→药→效”的思路,临床用药以“证”为依据,以“效”为评价标准。因此,中医用药规律研究应该把“药及配伍关系”作为研究中心和先导,并兼顾“证-药”和“药-效”的研究。

3. 要重视不同疾病用药规律的对比研究

用药规律研究不仅要以单一疾病为对象,也要多种疾病互参对比研究。对比研究不仅有利于对比各方用药规律的再认识和深入探索,还有助于发现相同用药下的不同疾病共同生物学机制,从而为中医“异病同治”思想提供客观依据,也为开发兼具精确和广泛适应症的现代中药提供具体数据支持。

4. 应在不预设中医学规则的前提下对海量文献进行挖掘分析

为了客观真实地反映某个疾病的中医用药规律,对相关信息整个挖掘分析过程应该避免中医知识、规则的介入。小样本、不完整的信息对整体缺乏代表性,不但使挖掘分析偏倚增大,而且有可能影响信息分析的架构,从而使结果的真实可靠性降低。因此,保证采集信息的全面性和完整性,是探索中医用药规律的必要条件。

三、文本挖掘是中医用药规律研究的重要工具

针对疾病和海量数据进行深层次用药规律研究,技术问题是关键。在生物学领域,由于生物学数据和生物医学文献数量的急骤增长,通过数据挖掘寻找规律和新知成了生物学研究的一个新热点和重要分支^[4]。文本挖掘是从文本集中发现潜在、隐藏的归纳性的知识,如有用的模式、模型、趋势、规则等,它是人工智能、机器学习、自然语言处理、相关自动文本处理如信息抽取、信息检索、文本分类等理论和技术相结合的产物^[5-6]。由于大量文本数据库的存在,文本挖掘成为数据挖掘领域的一个热点研究方向。文本挖掘与传统数据挖掘的本质区别在于其分析对象的非结构化或半结构化。由于直接面向文本型数据的算法的复杂性很高,所以目前文本挖掘是将文本型数据转换为结构化数据,再利用数据挖掘成熟的算法进行分析^[7]。

中医用药是中医辨证论治的完整体现,中医辨证充满非线性思维,“证-药-效”间的多层关联、序列组合、集群对应形成了整体论的思维方式和原则。文本挖掘能以线性和非线性方式解析数据,且能进行高层次的知识整合,善于处理模糊和非量化数据^[6]。

文本挖掘的可视化可用图形化方法描述数据间的相互关系,脉络清晰,可理解性强。可见,文本挖掘具有探索中医用药规律的技术特征和条件,可作为研究的主要方法。

四、采用文本挖掘进行中医用药规律研究 具有可行性和先进性

基于上述研究思路,我们以西医疾病为研究对象,从中国生物医学文献数据库(Chinese BioMedical Literature Database, CBM)中收集数据,利用专有的文本提取工具抽出不同频数的关键词对,进行结构化处理,然后存入 Access 数据库和 SQL 数据库进行挖掘和分析,并对结果进行可视化。以下举例说明我们文本挖掘的结果及可视化分析讨论情况,以此展示文本挖掘的可行性和先进性。

1. 单一疾病分析

通过对有关“心律失常”57447 篇文献的分析,得到用药关系网络图 1~2。药物之间连线上的数字是两个药物共同出现的频数。频数大小体现了两个药物在疾病治疗中的重要程度。与某个药物连接的线条越多,其他药物与这个药物配伍的几率越高。图 1 展示的是两药同时出现频次在 16~31 次之间的疾病用药情况,丹参与 10 个药物有配伍关系,体现了丹参的核心配伍功能。其与图中其他药物构成了治疗心律失常的主要中药用药范畴。图 2 展示的是两药同时出现频次在 32~99 次之间的疾病用药情况,图中展现的药物组合有:人参、麦冬、五味子是生脉饮的组成药物,心肺同治,主补益心气心阴;甘草、桂枝、阿胶、人参是炙甘草汤的主要组分,能够通经脉、助心阳、补心气、养心血;麻黄、附子、细辛是治疗寒凝胸肺、心阳不振的主要药物;黄芪、当归组成当归补血汤,补气生血,有助心主血脉功能正常发挥。本研究结果体现了中医治疗心律失常的主要思想:心、肺、脾三脏同调,三位一体,以温阳、补气、活血为主,配以养血、解毒、宣肺等;并发现了一些特殊配伍现象和模式,如甘松、黄精形成的药对,以及丹参、麦冬、甘草、桂枝、人参、黄芪组成的环状药物组合结构,值得深入研究。

2. 不同疾病的对比分析

按照现代疾病分类,哮喘属于呼吸系统疾病,肾小球肾炎属于泌尿系统疾病,对这两个疾病的文献进行文本挖掘并对比研究,是探索两者共同用药

情况,并进一步探索共同生物学基础的重要方法。通过挖掘有关“哮喘”的文献 34248 篇和有关“肾小球肾炎”的文献 11017 篇,形成高频可视化图 3。图 3-A 显示,麻黄、半夏、甘草、杏仁、细辛、射干和地龙是治疗哮喘的主要药物,其中麻黄是配伍的核心药物,这些药物多有宣肃肺气、化痰活血的功效。图 3-B 显示,黄芪、益母草、丹参、茯苓、泽泻、雷公藤、山药、当归是治疗肾小球肾炎的主要药物,其中黄芪是配伍的核心药物,这些药物具有补气养血、活血利水、清热解毒的功效。合并挖掘两者文献,图 3-C 显示,两种疾病高频共同出现的药物有黄芪、白术、党参、丹参和地龙,其中黄芪是核心药物。分析药物组合的中医学意义,补益中气兼活血养血是两者共同的治疗方法,与临床实际一致。两个不相

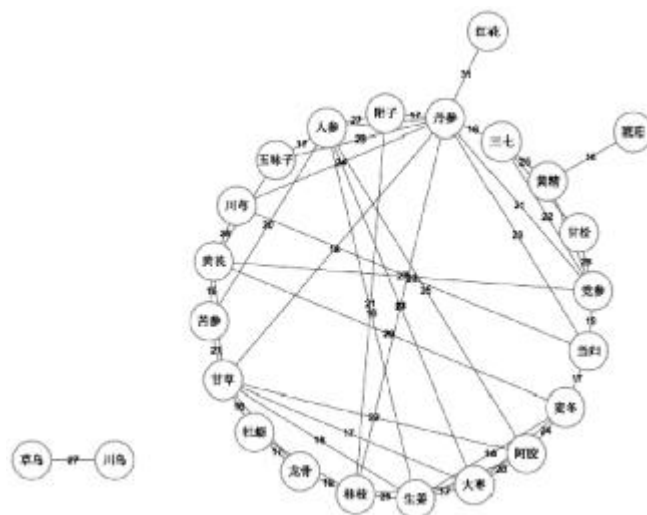


图 1 用药关系网络图(频次 16~31)

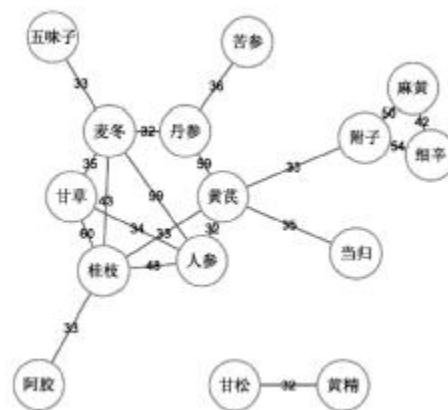


图 2 用药关系网络图(频次 32~99)

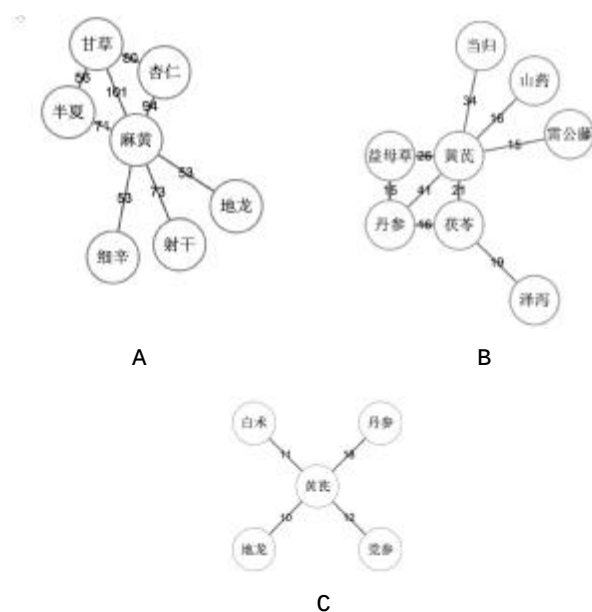


图3 高频可视图

干的疾病,有共同的用药,机制何在? 这些共同药物不仅为中医异病同治思想提供了客观依据,也为临床及实验研究两者的生物学基础提供了进一步探索的线索和目标。

五、总结与展望

“21 世纪是系统生物学时代”,在系统论指导下,利用多学科技术研究中西医结合理论实质及其内在联

系,有利于促进西医学找到共同语言^[8]。利用文本挖掘技术对中医疾病用药历史数据进行智能分析,不仅能更客观地印证、重现既有的疾病用药规律,还能发现新的知识和线索。此类研究对于发掘中医科学内涵、带动中医学术水平提高、拓展中医生存空间具有促进作用。

参考文献

- 1 王丽平,张其慧,李萌,等.许彭龄教授治疗慢性胃炎用药规律研究.中华中医药杂志,2010,25(3):475-478.
- 2 包祖晓,田青,高新彦,等.235 首抑郁情绪治疗方剂的用药组方规律分析.浙江中医药大学学报,2010,34(5):763-766.
- 3 吕爱平,姜森,赵静.立足中医理论探索药物研发新途径.中国中西医结合杂志,2008,26(5):163-166.
- 4 Tari L, Anwar S, Liang S, et al. Discovering drug-drug interactions: a text-mining and reasoning approach based on properties of drug metabolism. Bioinformatics, 2010, 26(18): 1547-1553.
- 5 Al-Mubaid H, Singh RK. A text-mining technique for extracting gene-disease associations from the biomedical literature. Int J Bioinform Res Appl, 2010, 6(3):270-286.
- 6 Rodriguez-Esteban R. Biomedical text mining and its applications. PLoS Comput Biol, 2009, 5(12):e1000597.
- 7 Tsai RT, Dai HJ, Lai PT, et al. PubMed-EX: a web browser extension to enhance PubMed search with text mining features. Bioinformatics, 2009, 25(22):3031-3032.
- 8 刘永琦. 系统生物学对中西医结合研究的影响. 中医杂志,2007,48(10):869-875.

Based on Text Mining Techniques to Explore Medication Regularities of Diseases Treatment with Traditional Chinese Medicine

Tan Yong¹, Guo Hongtao², Zheng Guang¹, Zhang Chi¹, Yang Jing¹,
Lv Cheng¹, Zha Qingling¹, Jiang Miao¹, Lv Aiping¹

(1. Institute of Basic Research In Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China;

2. Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China)

Abstract: Exploring medication regularities of treating diseases are one of the most important tasks in traditional Chinese medicine clinical research. Nowadays, it is possible and important to explore medication regularities of treating diseases with traditional Chinese medicine by text mining techniques. Medication regularities include "syndrome-herb medicine", "herb medicine-herb medicine" and "herb medicine-therapeutic effect" three relationships. The relationship among herb medicines is the center and link of the above-mentioned three relationships. Thus, in the absence of traditional Chinese medicine principles, taken modern medicine diseases as study object, the most frequently used single herb and herbs combination relationship from extensive literature should be studied in advance. Besides, the research of medication regularities should attach importance to comparative studies of different

diseases. Medication regularities, which can both be explained in modern science and is internationally recognized, are expected to be obtained through the above research ideas and methods. Previous studies indicated that exploring medication regularities of treating diseases with traditional Chinese medicine by text mining techniques is feasible and advanced.

Keywords: Traditional Chinese medicine, Medication regularity, Text mining

(责任编辑:李沙沙,责任译审:吕爱平)