

基于文本挖掘方法探索寒性热性中药的病证方药相应规律

梁非¹, 展俊平^{2,3}, 李立², 郑光^{2,4}, 吕爱平², 姜淼^{2*}, 喻长远^{1*}

(1 北京化工大学, 北京 100029; 2 中国中医科学院中医临床基础医学研究所, 北京 100700;
3 上海中医药大学, 上海 201203; 4 兰州大学信息学院, 兰州 730000)

[摘要] **目的:**探索寒性中药和热性中药的病证方药相应规律。**方法:**在 CBM 数据库中收集三味寒药(苦参、龙胆、大青叶)和三味热药(肉桂、吴茱萸、胡椒)文献数据,采用基于敏感关键词频数统计的数据分层算法,挖掘与之相关的疾病、证候、症状、汤药规律,这些规律通过一维频次的解释以及二维网络图进行展示。**结果:**疾病方面:与寒性药和热性药相关的疾病涉及各个系统多种疾病,并无一定规律可循。证候方面:热性药的中医证候以阳虚、寒凝、气滞、血瘀为主;寒性药则以湿热、阴虚、热毒、风热为主。症状方面:热性药相关的症状是以消化系统和神经系统症状为主;寒性药除了消化系统症状外,尚且有皮肤疾病表现。中药方剂方面:中药方剂方面,热性药相关的主要是温里剂和温阳剂,寒性药相关的主要是清热剂和和缓剂。**结论:**文本挖掘技术,结合文献回溯,人工阅读降噪,能够比较客观的总结寒热药性的病证方药相应规律,为进一步阐明中药寒热药性差异的客观性提供探索与参考。

[关键词] 文本挖掘; 寒性中药; 热性中药; 中医证候

[中图分类号] R287 **[文献标识码]** A **[文章编号]** 1005-9903(2013)15-0333-05

[doi] 10.11653/syfy2013150333

Relationships between Prescription and Syndrome on Chinese Herbal Medicines with Properties of Cold or Hot: A Text Mining Analysis

LIANG Fei¹, ZHAN Jun-ping^{2,3}, LI Li², ZHENG Guang^{2,4}, LV Ai-ping², JIANG Miao^{2*}, YU Chang-yuan^{1*}

(1. Beijing University of Chemical Technology College, Beijing 100029, China;

2. Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Science, Beijing 100700, China;

3. Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China;

4. School of Information Lanzhou University, Lanzhou 730000, China)

[Abstract] **Objective:** To explore the relationships between prescription and syndrome on Chinese herbal medicines with properties of cold or hot. **Method:** Data of *Sophorae Flavescentis Radix*, *Gentianae Radix et Rhizoma*, *Isatidis folium*, *Cinnamoni Cortex*, *Eudiae Fructus*, *Piperis Fructus* were downloaded from CBM database. The rules of disease, pattern/syndrome, symptom, and decoction were mined out with data slicing algorithm. The results are visually demonstrated with cytoscape. **Result:** For diseases, there was no clear separation between cold herbs and hot herbs, because they covered disease of several system. For traditional Chinese Medicine (TCM) pattern/syndrome, the hot herbs were mainly focused on yang deficiency, coagulated cold, qi stagnation, blood stasis. For cold herbs, they were mainly focused on dampness-heat, yin deficiency,

[收稿日期] 20121128(003)

[基金项目] 国家自然科学基金杰出青年项目(30825047);国家自然科学基金面上项目(30973975,81072982);国家自然科学基金青年基金项目(30902003);中国博士后基金面上项目(20110940553);中国中医科学院自主选题项目(Z0172);国家自然科学基金项目(81273631)

[通讯作者] *姜淼,副研究员,Tel:010-64014411-2397, E-mail:miao_jm@vip.126.com; *喻长远,教授,E-mail:yuey@mail.buct.edu.cn

heat-toxicity, wind-heat. For symptoms, hot herbs were mainly focused on diseases of digestive and nervous systems. But for cold herbs, not only the digestive system, but also the skin diseases, i. e. , pruritus and papule were included. For decoctions, formulae of Wen-Yang and Wen-Li were associated with hot herbs, on the country, cold herbs were mainly focused on formulae of Qing-Re and He-Jie. Take digestive disease for example, these relationships between prescription and syndrome are existed. **Conclusion:** Text mining, together with artificial reading for anti-noising, is an important approach in exploring the rules among disease, TCM pattern/syndrome, symptom, and prescription associated with cold and hot herbs.

[**Key words**] text mining; cold herbs; hot herbs; pattern/syndrome

药性理论是中药理论的基本组成部分,包括四气、五味、升降浮沉,归经、有无毒性等,其中的四气理论(寒、热、温、凉)是中药药性理论的重要组成部分^[1]。然而,如何阐明中药寒热药性差异的客观性仍然是中医药基础研究的难点和热点。本文尝试以四气中寒热这两个对立的特性为出发点,从寒热药中各选出 3 味较常用的中药作为目标药物,借助课题组不断成熟的文本挖掘技术^[2],结合原文献回溯,人工阅读分析等方法,对与这 6 味中药相关的中医文献进行挖掘,分析以此 6 味药为代表的寒性和热性中药,在临床应用中的病证方药规律。

1 材料与方法

1.1 寒热性药物的选择 为了能很好的研究寒热药性之间的具体差异,笔者选择大家所熟识、常用,且其寒热药属性记载始终一致的中药作为代表药物。鉴于本课题组一直在进行中药寒热药性的相关研究,并且分别构建了寒性药与热性药的实验数据库^[3],笔者从中各选 3 味中药,结合前期数据挖掘结果,寒性中药以苦参、龙胆、大青叶为目标药物;热性中药以肉桂、吴茱萸、胡椒为目标药物。

1.2 文本数据收集 方法概述如下:在中国生物医学文献数据库(Chinese BioMedical Literature Database, CBM, <http://sinomed. cintcm. ac. cn/index. jsp>)中,“ 缺省(智能)”状态下以“苦参”为检索词进行检索,共得到文献 7 672 篇(检索日期:2012 年 6 月 1 日),下载后作为文本数据挖掘的原始数据。依照同法,其他的 5 味中药文本数据收集结果见表 1。

1.3 文本数据处理 将收集来的数据,按照下载的顺序,整合到一个平面文件(后缀 txt)里面,以 ANSI 编码格式保存。然后,利用专有的文本提取工具(软件著作权,软著登字第 0261882 号,登记号 2010SR073409),对下载的非结构化的 txt 文本数据进行信息提取,保存成格式化的、便于大型关系型数据库(Microsoft SQL Server,以下简称 SQL)处理的格

表 1 寒性及热性中药文本数据收集

药性	药名	Latin Name	文本收集结果/篇
寒性药	苦参	<i>Sophora flavescens</i>	7 672
	龙胆	<i>Gentia nascabra</i> Bunge	3 557
	大青叶	<i>Isatis indigotica</i> Fort	672
热性药	肉桂	<i>Cinnamomi cassiae</i>	2 141
	吴茱萸	<i>Evodia rutaecarpa</i>	1 622
	胡椒	<i>Piper nigrum</i> L.	769

式,然后导入 SQL 中进行下一步的挖掘分析。假设每一篇文献的贡献度是相同的,一篇文献中重复出现的关键词,只需要计算一次,据此构建算法进行数据清洗工作^[2,4-5]。清洗完毕后的数据,既可以提取挖掘对象的一维频次,也可以得到挖掘对象的二维关系,进行可视化呈现。抽出不同频次的关键词对,用 Cytoscape 2. 8 软件进行可视化处理,形成可视化的网络图,然后结合专业知识进行解析,一旦发现不合理的结果,即回溯原文献数据集,如果是噪音,仍按算法进行噪音清洗,直至噪音降到满意为止。最后的结果可视化成图,结合专业知识进行解析。

2 寒热药文本挖掘结果的评价和分析

2.1 疾病文本挖掘 一维频次文本挖掘结果:热性中药相关疾病,肉桂 49 种,吴茱萸 38 种,胡椒 19 种;寒性中药相关疾病,苦参 61 种,龙胆 48 种,大青叶 16 种。

二维网络图构建:分别将 3 味热药和 3 味寒药中频次最高的前 3 种疾病整合到一个数据表中,剔除重复出现的疾病数据,保证该表中所有的疾病只出现一次,然后,构建寒性药和热性药与相关疾病构成的网络图(图 1)。图中正方形代表热性药对应的疾病,八边形代表寒性药对应的疾病,圆形代表寒性药与热性药共有的疾病。

由图 1 可知,热性药相关的疾病,主要包括胃癌、溃疡性结肠炎、休克、多发性骨髓瘤、糖尿病、胃炎,涉及消化系统、内分泌系统、血液系统;寒性药相

关的疾病主要包括肝硬化、过敏性紫癜、痛风、流行性感、白血病、心肌炎、心律失常,涉及心血管系统、血液系统、内分泌系统、消化系统、呼吸系统。从结果中可以发现,与寒性药和热性药相关的疾病,涉及各个系统多种疾病,并无一定规律可循。

2.2 中医证候文本挖掘结果 一维频次文本挖掘结果:热性中药相关证候,肉桂 111 种,吴茱萸 80 种,胡椒 14 种;寒性中药相关证候,苦参 72 种,龙胆 158 种,大青叶 22 种。

二维网络图构建结果:分别将 3 味热药和 3 味寒药中频次最高的前 5 种证候整合到一个数据表中,剔除重复出现的证候数据,保证该表中所有的证候只出现一次,然后构建寒性药和热性药与相关证候构成的网络图(图 2)。图中正方形代表热性药对应的证候,八边形代表寒性药对应的证候,圆形代表寒性药与热性药共有的证候。

由图可知,热性药相关的中医证候,以阳虚(肾

阳虚、脾肾阳虚)、寒凝(寒凝血瘀、寒凝气滞、胃虚寒、风寒湿痹、寒滞肝脉)、气滞(寒凝气滞、气滞血瘀)为主;寒性药相关的中医证候,以湿热(肝胆湿热、湿热蕴结、湿热下注、湿热内蕴、湿热证)、阴虚(肝肾阴虚、阴虚火旺)、热毒(热毒炽盛)、风热(肝经风热)为主。

2.3 症状文本挖掘结果 一维频次文本挖掘结果:热性中药相关症状,肉桂 118 种,吴茱萸 99 种,胡椒 43 种;寒性中药相关症状,苦参 114 种,龙胆 148 种,大青叶 60 种。

二维网络图构建结果:分别将 3 味热药和 3 味寒药中频次最高的前 8 种症状整合到 1 个数据表中,剔除重复出现的症状数据,保证该表中所有的症状只出现一次,然后构建寒性药和热性药与相关症状构成的网络图(图 3)。图中正方形代表热性药对应的症状,八边形代表寒性药对应的症状,圆形代表寒性药与热性药共有的症状。

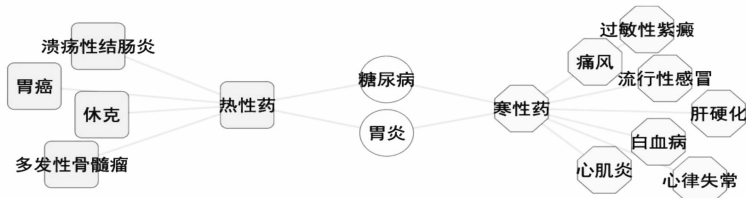


图 1 寒性药和热性药与相关疾病构成的网络

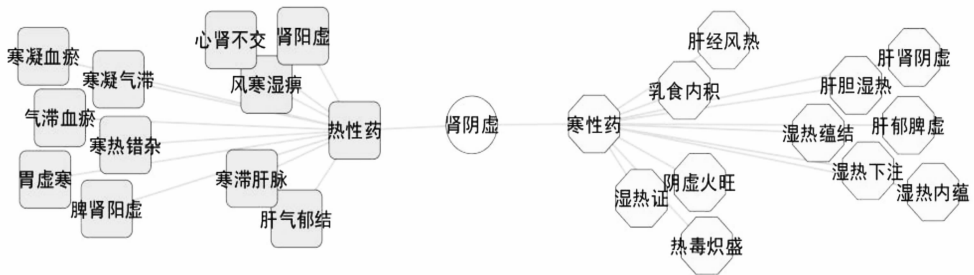


图 2 寒性药和热性药与相关中医证候构成的网络

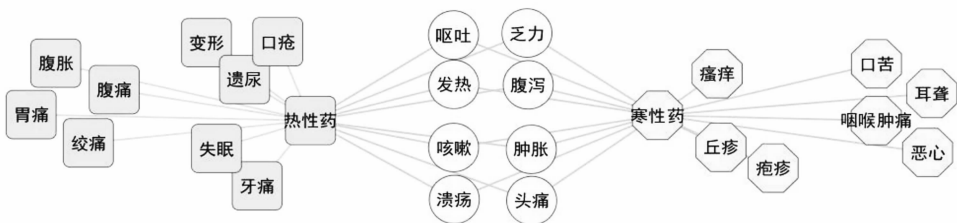


图 3 寒性药和热性药与相关症状构成的网络

由图可知,热性药相关的症状多与消化系统有关,表现为腹胀、腹痛、胃痛、呕吐、腹泻、溃疡,而发热、乏力、头痛、失眠多是神经系统的症状;寒性药除消化系统症状外,还有皮肤疾病的表现,如瘙痒、丘

疹、疱疹等。
2.4 汤药文本挖掘 一维频次文本挖掘结果:热性中药相关汤药,肉桂 76 种,吴茱萸 79 种,胡椒 5 种;寒性中药相关汤药,苦参 71 种,龙胆 126 种,大青叶

18 种。

二维网络图构建:分别将 3 味热药和 3 味寒药中频次最高的前 5 种汤药整合到一个数据表中,剔除重复出现的症状数据,保证该表中所有的汤药只

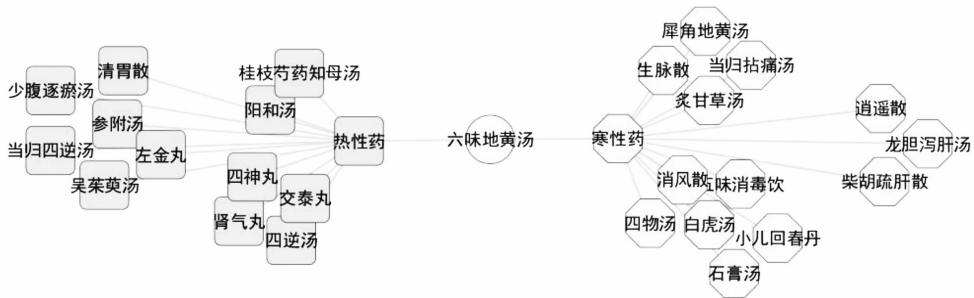


图 4 寒性药和热性药与相关汤药构成的网络

由 4 图可知,热性药相关的主要汤剂是温里剂(当归四逆汤、吴茱萸汤、参附汤、四逆汤、阳和汤),温阳剂(肾气丸、四神丸);寒性药相关的主要汤剂是清热剂,包括清气分热(白虎汤、石膏汤)、清热解毒(五味消毒饮)和清脏腑热(龙胆泻肝汤),寒性药中还有和解剂(逍遥散、柴胡疏肝散)。

3 讨论

文本挖掘技术是以计算语言学、统计数理分析为理论基础,服务于生物、医药、文献研究等学科的新兴的交叉学科^[6]。应用于中医药领域,文本挖掘能从海量的中医药文献中发现知识以促进中医临床研究和中药复方研发等多个方面^[7]。

本研究尝试从中药寒热药性出发,结合文本挖掘技术,探讨与其的病证方药规律。挖掘结果显示:疾病方面,与寒性药和热性药相关的疾病涉及各个系统多种疾病,并无一定规律可循,这可能与每一种疾病都有不同的寒热证候分类,甚至随着疾病的发展演变,证候也可能出现寒热转变有关,所以并不能单从中药的寒热属性来确定相关的疾病;证候方面,热性药的中医证候以阳虚、寒凝、气滞、血瘀为主;寒性药则以湿热、阴虚、热毒、风热为主。《神农本草经》云:“疗寒以热药,疗热以寒药”,《素问·至真要大论》云:“寒者热之,热者寒之。”本研究对寒性及热性中药相关证候文本挖掘结果,与中医理论基本一致;症状方面,从结果中可以看出,很多消化系统症状与热性药物和寒性药物都相关,这可能与消化系统疾病中寒证和热证都比较常见有关,所以,寒药和热药都出现了消化系统症状,另外,中医皮肤科多以湿热毒蕴证候为主,故寒性药物可并见皮肤病症状;中药方剂方面,文本挖掘的结果显示,热性药相

出现一次,然后构建寒性药和热性药与相关汤药构成的网络图(图 4)。图中正方形代表热性药对应的汤药,八边形代表寒性药对应的汤药,圆形代表寒性药与热性药共有的汤药。

关的主要是温里剂和温阳剂,寒性药相关的主要是清热剂、和解剂,这一结果与寒性热性中药在临床应用时组方用药、治则治法原则基本相符。

基于以上文本挖掘的结果,选取高频数的利用 Cytoscape 2.8 软件形成可视化的网络图,展示临床上的常见证候、常用药物的使用,能够直接客观的反应寒热药物与临床应用的现状。值得一提的是,消化系统相关的疾病、证候、症状、方药在本研究中多次出现,提示在未来的研究中,可以尝试单独分析寒性热性药物与消化系统疾病的病证方药规律。目前,基于文本挖掘技术探索中医药在治疗疾病中的应用规律研究,已经有广泛的开展,并且结果可靠^[8,10],随着该方法逐步在中医学研究中的推广,希望能对中医药规范化研究提供有益的帮助。

【参考文献】

[1] 金锐,张冰. 中药药性理论复杂性特征分析[J]. 中国中药杂志,2012,37(12):3340.

[2] ZHENG G, JIANG M, HE X J, et al. Discrete derivative: a data slicing algorithm for exploration of sharing biological networks between rheumatoid arthritis and coronary heart disease [J]. Bio Data Min, 2011, 4:18

[3] Liang F, Zhan J P, Zheng G, et al. Exploring associated networks of cold herb and hot herb within the frame work of traditional Chinese medicine through text mining [C]. Taipei: International Conference on New Trends in Information Science, Service Science and Data Mining, 2012.

[4] Andrea Campagna, Rasmus Pagh. Finding associations and computing similarity via biased pair sampling [C]. Miami, Florida: 2009 Ninth IEEE International Conference on Data Mining, 2009:61.

中药药性物质基础研究中数学建模思路探讨

王晓燕^{1,2*}

(1. 山东中医药大学药学院, 济南 250355; 2. 山东中医药大学理工学院, 济南 250355)

[摘要] **目的:**探讨数学建模策略在中药药性物质基础研究中的应用,并尝试建立一套客观可重复的建模思路与方法。**方法:**从数据库建立、数据预处理、建模策略选择和评价标准制定 4 个关键步骤,讨论中药药性特征标记识别模型的建立。**结果:**从 64 种建模策略中选出 17 种判别良好的模型,继而考察其可视化表达能力选出 4 种适宜模型,最后经统计模拟,选出 PLS 作为目前最适宜模型。**结论:**中药药性与其发挥作用的物质基础之间难以找到现成的公式或规律来表达二者关联,可用穷举策略,经数据库建立、数据预处理、建模策略选择和评价 4 步筛选适宜的中药药性特征标记识别模型。

[关键词] 数学建模; 药性; 物质基础; 穷举建模策略

[中图分类号] R287 **[文献标识码]** A **[文章编号]** 1005-9903(2013)15-0337-03

[doi] 10.11653/syfy2013150337

Study on Mathematical Modeling in Material Foundation of Traditional Chinese Medicine

WANG Xiao-yan^{1,2*}

(1. College of Pharmaceutical, Shandong University of Traditional Chinese Medicine (TCM), Ji'nan 250355, China;
2. College of Science and Technology, Shandong University of TCM, Ji'nan 250355, China)

[Abstract] **Objective:** To discuss the application of mathematical modeling strategies in material foundation of traditional Chinese medicine (TCM), and try to establish a set of objective modeling methods. **Method:** From the database establishment, data preprocessing, modeling strategy choosing and model's evaluations the most appropriate recognition model for CHMP-markers was choose. **Result:** Seventeen kind of modeling strategies from 64 was selected, and 4 kind was chosen in view of the visual expression ability, and PLS was selected as the most appropriate one via statistical simulation. **Conclusion:** It is difficult to find ready-made formula or law express the relation between the relation of Chinese medicine medicinal properties and material base, exhaustive method can be used to choose the most suitable model, database establishment, data preprocessing, modeling strategy choosing and model's evaluations are the 4 key steps.

[收稿日期] 20121126(007)

[基金项目] 国家重点基础研究发展计划项目(2007CB512601)

[通讯作者] * 王晓燕, 博士生, 讲师, 从事中药质量控制与资源研究, E-mail: sdnwxy@126.com

[5] Jeffrey W Seifert. Data mining: An overview [R]. CRS Report RL31798, 2004.
[6] 薛为民, 陆玉昌. 文本挖掘技术研究[J]. 北京联合大学学报: 自然科学版, 2005, 19(4): 59.
[7] 吕婷, 姜友好. 文本挖掘在生物医学领域中的应用及系统工具[J]. 中华医学图书情报杂志, 2010, 19(4): 56.
[8] 李立, 周奇, 郑光, 等. 基于文本挖掘技术分析中成药、西药对慢性胃炎的治疗规律[J]. 中国实验方剂

学杂志, 2011, 17(24): 228.
[9] 郭洪涛, 郑光, 赵静, 等. 基于文本挖掘分析甲型 H1N1 流感的中医药治疗特色[J]. 世界科学技术——中医药现代化, 2011(5): 772.
[10] 谭勇, 杨静, 赵宁, 等. 利用文本挖掘技术探索中西药治疗慢性乙型肝炎的用药规律[J]. 中国实验方剂学杂志, 2011, 17(24): 232.

[责任编辑 邹晓翠]