

基于文本挖掘探讨中医外治法和食疗在糖尿病中应用规律

沈姗姗¹, 姜淼¹, 郑光², 杨静¹, 宋梦梦², 史欢欢², 吕爱平^{1,3*}

(1. 中国中医科学院 中医临床基础医学研究所, 北京 100700;

2. 兰州大学 计算机信息学院, 兰州 730000; 3. 香港浸会大学 中药学院, 香港 999077)

[摘要] **目的:**利用文本挖掘技术探索糖尿病的中医外治法和食疗规律特点。**方法:**在中国生物医学文献数据库(CBM)中收集有关糖尿病所有文献资料,建立 Access 数据库,采用基于敏感关键词频数统计数据分层算法,运用 SQL 软件对数据进行专业处理,并根据医学专业知识进行降噪,最终挖掘出糖尿病中医外治法以及食疗规律,并对挖掘出的数据进行一维频次解释及二维频次网络图展示,全面解析挖掘数据情况。**结果:**通过文本挖掘得出当前文献已有报道中糖尿病最常用中医外治法依次为针灸(频数为 2 079),光疗(频数为 1 237),中药外敷(频数为 348),穴位注射(频数为 327),按摩(频数为 207),最常用的食疗对象依次是:三七(频数为 446),山药(频数为 262),柠檬(频数为 234),枸杞子(频数为 143)。**结论:**文本挖掘结果,选取高频数参与网络关系图构建,直接客观的反应临床应用现状。本研究通过文本挖掘以及回溯文献,并结合人工阅读降噪,从文献报道频数方面比较客观的总结糖尿病中医外治法和食疗规律,为中医临床医师诊治方案制定提供有益的探索与依据,并为临床其他疾病施治中如何选择最佳中医适宜技术提供了方法学参考,而对于糖尿病应用中医外治法和食疗具体作用机制、适应症状等方面,还值得进一步研究。

[关键词] 文本挖掘;糖尿病;中医外治法;食疗

[中图分类号] R287.6 **[文献标识码]** A **[文章编号]** 1005-9903(2015)01-0209-04

[doi] 10.13422/j.cnki.syfjx.2015010209

[网络出版地址] <http://www.cnki.net/kcms/detail/11.3495.R.20141113.1453.002.html>

[网络出版时间] 2014-11-13 14:53

Application of Traditional Chinese Medicine External Therapy and Food Therapy on Diabetes Treatment

SHEN Shan-shan¹, JIANG Miao¹, ZHENG Guang², YANG Jing¹, SONG Meng-meng², SHI Huan-huan², LYU Ai-ping^{1,3*} (1. *Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China*; 2. *School of Information Science and Engineering Technology, Lanzhou University, Lanzhou 730000, China*; 3. *School of Chinese Medicine, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hongkong 999077, China*)

[Abstract] **Objective:** To explore the rules among external therapy and food therapy of traditional Chinese medicine (TCM) in treating diabetes by applying text mining technology. **Method:** The relative literatures were collected from Sinomed database (CBM), and an Access database was established. The data was processed by SQL using frequency statistical method based on sensitive keywords hierarchical algorithm, and noise was manually removed according to the medical professional knowledge. The rules among external therapy and food therapy of TCM were deeply analysed and comprehensively displayed by the frequency of one-dimensional and two-dimensional network diagram. **Result:** The results showed that the most common external therapies of diabetes were acupuncture (frequency 2 079), phototherapy (frequency 1 237), external application with TCM (frequency 348), acupoint injection (frequency 327) and massage (frequency 207) in turn. The most common food therapies were followed by panax notoginseng (frequency 446), Chinese yam (frequency 262), lemon (frequency 234) and wolfberry (frequency 143). **Conclusion:** Text mining, together with artificial reading for anti-noising, is an useful

[收稿日期] 20140429(013)

[基金项目] 国家自然科学基金青年基金项目(30902003)

[第一作者] 沈姗姗,博士,从事中医内科学研究,Tel:15901495477,E-mail:ras0929@163.com

[通讯作者] *吕爱平,博士,研究员,从事中西医结合基础研究,Tel:010-64067611,E-mail:lap64067611@126.com

approach in exploring the clinical rules of TCM in treating diabetes. However, its exact mechanism and certain indication in treating diabetes is worthy of further study.

[Key words] text mining; diabetes; external therapy of traditional Chinese medicine; food therapy of traditional Chinese medicine

糖尿病临床上是以血糖升高为基本特征的一种内分泌代谢性疾病,其病理生理改变是胰岛素分泌的绝对或相对不足,引起碳水化合物、脂肪、蛋白质等的代谢紊乱。目前,糖尿病仅次于心脑血管疾病和肿瘤,已发展成为发达国家致死率第三高的慢性病,而我国的糖尿病患者多达 4 000 多万,已发展成为严重的公共卫生问题^[1]。糖尿病属于祖国医学“脾瘕”、“消渴”范畴,中医治疗方法在改善患者各种临床症状方面,凸显了它的优势。现代中医在糖尿病理论和临床研究方面也取得了较大进展,通过临床辨证分型施治,使患者获得了较好疗效^[2],并且在降糖、消除症状、改善各种生化指标方面积累了一些临床经验^[3]。文本挖掘是数据挖掘技术的一个分支,其在中医药领域的应用已日渐成熟,通过检索相关文献,发现已有将文本挖掘技术应用于糖尿病中成药及西药用药规律的分析^[4],并且取得了一些研究成果,但关于数据挖掘糖尿病中医外治法和食疗的应用规律,尚未有文献报道。本文应用文本挖掘方法^[5],系统地探索并总结中医外治法和食疗在糖尿病中的应用规律,为临床诊疗提供理论参考依据。

1 材料与方法

1.1 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) 中以“缺省[智能]:”状态下检索“糖尿病”,共得到文献 187 529 篇(检索日期:2014 年 2 月 9 日),依次下载所有文献并保存。

1.2 文本数据处理 将收集来的数据,按照下载先后顺序,整合到一个平面文件(后缀 txt)里面,以 ANSI 编码格式保存。然后,利用专用的文本提取工具(软件著作权,软著登字第 0261882 号,登记号 2010SR073409),对下载的非结构化 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) 处理的格式,然后导入 SQL 中进行下一步的挖掘分析。算一次,据此构建算法进行数据清洗工作^[5]。清洗完毕后的数据,既可以挖掘对象的一维频次,也可以得到挖掘对象二维关系,进行可视化呈现。抽出高频次的主要关键词对,运用 Cytoscape 2. 8 软件进行可视化处理,形成可视化网络图,然后结合专业知识进行解析,一旦发现不合理结果,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。最后结果再一次可视化成图,结合专业知识进行最终解析。

2 挖掘结果

2.1 中医外治法挖掘结果 文本挖掘共提取到 50 种中医外治法,选择前 6 个依次列出(图 1),分别是:针灸、光疗法、穴位注射、中药外敷、按摩、推拿。通过一维频次挖掘结果可

以看出,在糖尿病的中医外治法中,针灸在临床上是使用频率最高的,其次是光疗法和穴位注射、中药外敷。此外,按摩、推拿方法,也在临床得到应用推广。

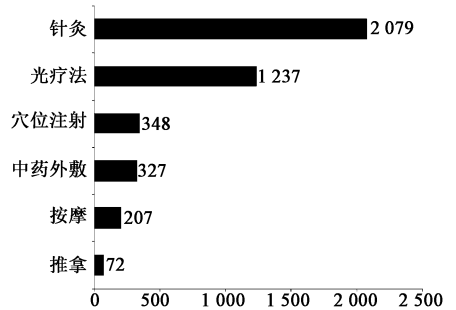


图 1 一维频次中医外治法挖掘结果

Fig. 1 External treatment of traditional Chinese medicine mining results based on one-dimensional frequency

基于二维频次,构建糖尿病中医外治法两两之间网络关系图(图 2),图中圆圈内为中医外治法名称,连线代表外治法两两之间联系。治法连线愈多,代表该治法与疾病关联程度越高,越处于核心位置;圆圈越大,代表该治法在文献中出现频次越高,其权重越大。

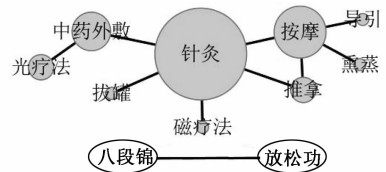


图 2 二维频次中医外治法挖掘结果

Fig. 2 External treatment of traditional Chinese medicine mining results based on two-dimensional frequency

文本挖掘共提取到 39 项外治法组合,图 2 为频次 > 2 的治法组合参与构建的网络图。由图可知,针灸、按摩、中药外敷圆圈较大,其相应代表的挖掘文献频次也较高。其中针灸是与其他治法联用最多的,处于核心位置,权重最大。左侧圆圈中的中药外敷治法,除与针灸联用次数比较多外,还常与光疗法配合应用于糖尿病的临床治疗。右侧圆圈中的按摩疗法,除与针灸配合应用外,还常与推拿、导引、熏蒸疗法分别联合应用,其他外治法频次较低。基于文本挖掘归纳出的这些外治法可以得出:糖尿病的中医外治法联用中,主要以针灸、按摩、中药外敷为主,光疗、推拿、拔罐法等为辅的治疗特点。

2.2 中医食疗挖掘结果 文本挖掘共提取到 152 种食疗法,选择前 14 个依次列出(图 3),分别是:三七、山药、柠檬、枸杞子、花生、猪胰、苦瓜、地黄、玉米、灵芝、海藻、黄豆、荞麦、南瓜。通过一维频次挖掘结果可以看出,在糖尿病患者

的食疗方案中,三七、山药、柠檬是频数最高的,最常用于作为降糖食材。此外,枸杞子、花生、猪胰、苦瓜等,也是作为糖尿病患者食疗的较好选择。同时,由挖掘结果可归纳出食疗方案的一个特点是:频次较高的三七、山药、柠檬、枸杞子、花生、苦瓜等性味皆属于甘温、甘平、甘苦寒之类,可以作为糖尿病患者选择食物时性味取向的参考。

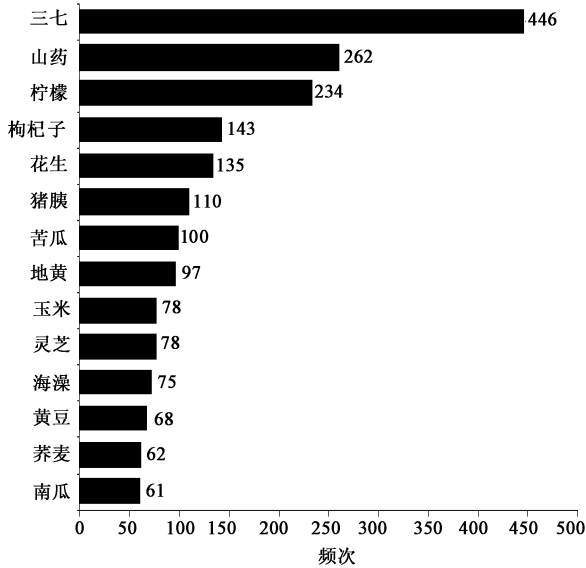


图 3 一维频次食疗挖掘结果

Fig. 3 Food treatment results based on one-dimensional frequency

基于二维频次,构建糖尿病食疗两两之间网络关系图(图 4):文本挖掘共提取到 435 项食疗组合,图 4 为频次 > 3 的组合参与构建的网络图。由图 3 可知,山药、枸杞子、三七、大蒜、南瓜与其他食材比较,相对较大,相应挖掘文献频次也较高,说明在食疗方案中是较常选择作为联用的食材对象,而与其他食材联合应用频次最高的是山药,权重也最大,玉竹、山楂、萝卜、猪胰等频次相对较低。由图 4 结果可以得出:糖尿病的食疗联用中,主要以性味甘温平的山药、枸杞子为主要食材,配合三七、山楂、玉竹等其他性味相似的食物,用于糖尿病患者日常饮食的推荐方案,大蒜、南瓜与洋葱、荞麦、萝卜等其他食材的配合应用也值得推广。

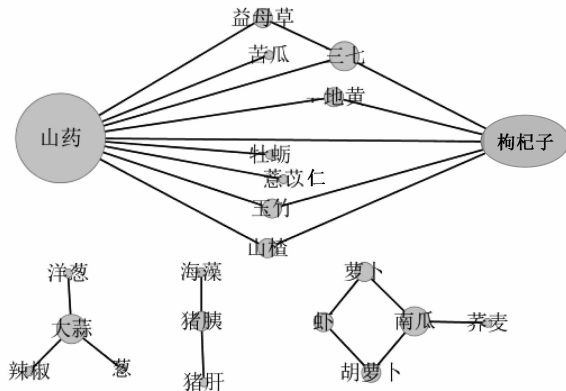


图 4 二维频次食疗挖掘结果

Fig. 4 Food treatment results based on two-dimensional frequency

3 讨论

文本挖掘(Text Mining)技术是以数理分析、计算语言学为理论基础,服务于医药、生物、文献研究等学科的新兴交叉学科^[6]。通过在大量文献中进行信息抽取、语词识别、发现知识间的关联等^[7],能够有效地找到所需信息,并发现隐藏的知识信息,解决了由于文档的数量庞大、缺乏组织整理以及格式多种多样,不能充分利用这些数据的问题,它可对海量数据进行整合、分析,获得的结果更具有代表性、可信度更高^[8]。在中医药领域,此方法从某种层次上促进中医临床研究和中药复方研发等多个方面,为中西医结合研究提供新的思路和途径,并且结果更加客观,可重复性强^[9]。本研究针对从 CBM 收集到 187 529 篇糖尿病相关文献,采用文本挖掘技术^[5],结合回溯原文献,人工阅读降噪,对糖尿病在中医外治法和食疗方面的应用规律进行挖掘,挖掘的结果既看到它们在文献中的频次,又可以发现它们各自在关联网络中的分布规律。

本次挖掘的中医外治法在糖尿病中的应用特点是:一维频次结果中针灸疗法的频次最高,其次是光疗法、穴位注射、中药外敷、按摩、推拿治法。二维频次结果中以针灸和中药外敷、针灸和按摩联合应用为主,针灸和推拿、针灸和拔罐联合应用为辅的特点。中医针灸对糖尿病的治疗积累了很多经验,疗法安全,疗效肯定、无毒副作用,一定程度上可弥补西医的不足,是目前比较理想的治疗方法。Head^[10]通过文献荟萃研究发现针灸等替代医学疗法能够通过刺激腧穴达到益气养阴、活血通络之效,从而控制血糖,给糖尿病患者带来极大好处。Zhao 等^[11]临床研究也证实,温针灸肾俞、脾俞等穴位,可明显改善糖尿病患者临床症状,降低患者血脂和血液黏度,改善微循环状况。而其他外治法,如光疗法能够通过光波照射起血管扩张,局部循环改善,增强代谢达到改善糖尿病患者周围血管病变的作用^[12-13]。通过对糖尿病患者背部腧穴以及足底脏腑反射区的推拿、按摩,可以明显降低患者空腹血糖(FBG),糖化血红蛋白(HbA1c)水平,有效改善患者临床症状体征^[14]。同样关于穴位注射、中药外敷等临床研究发现,局部应用活血化瘀、舒筋通络药物,可以改变血流量和血流速度,改善微循环,进而改善糖尿病周围血管病变出现的周围组织或器官缺血缺氧症状。文本挖掘获得的中医外治法结果,和临床实践方案相一致,并能得到临床验证。

本次挖掘的食疗在糖尿病中的应用规律是:一维频次较高的是三七、山药、柠檬,其次是枸杞子、花生、猪胰、苦瓜、地黄、玉米、灵芝、海藻、黄豆、荞麦、南瓜。二维频次结果中以山药、枸杞配合其他食材联合应用频次较高,三七、大蒜、南瓜联合其他食材应用频次相对较低。祖国医学通常将饮食调理与医疗联系在一起,早在《备急千金要方》中就有关于饮食治疗疾病的经文记载,并指出“食能排邪而安脏腑,如其所犯,以食治之”的食疗作用。同时中医学认为食物同药物一样,具有“寒、凉、温、热”四气和“酸、苦、甘、辛、咸”五味特性。根据文本挖掘出的糖尿病食疗结果,按照性味大致

可以分为三组:一组是性味甘平或甘温的食物,此类食物多以滋补功效为主,如山药、枸杞子、柠檬、花生、山楂、猪胰、南瓜、荞麦、黄豆、虾;一组是性味甘苦或寒的食物,此类食物除了补益之外,还兼具清热润燥功效,如地黄、玉竹、苦瓜,一组是性味辛温或咸寒的食物,如大蒜、益母草、海藻之类,此类食物具有辛散、活血化痰之功效。糖尿病的基本病机是阴虚为本,燥热为标,治疗原则当清热润燥、养阴生津。本次挖掘出的食物性味具有滋补、清热润燥、活血散结的功效,符合糖尿病患者治疗原则要求。同时,文本挖掘出的食疗对象多具有降糖的活性成分。现代药理学研究,山药中所含有的山药多糖可降低四氧嘧啶致糖尿病模型大鼠血糖,升高 C 肽水平^[15]。三七中所含有的主要活性成分三七总皂苷,对 2 型 DMKK-Ay 鼠具有抗高血糖、抗肥胖作用^[16]。柠檬中的活性成分柠檬酸能够降低血糖,可作为糖尿病患者日常饮食对象的选择。苦瓜中含有多种降糖活性成分,如三萜类、甾类、苷类和多肽等^[17-18],经常食用苦瓜,可以有效控制糖尿病患者血糖水平。文本挖掘的食疗结果,有充分的药理学理论支持其具有良好的降糖效应。

总之,糖尿病中医外治法和食疗文本挖掘的结果,在一定程度上反映了当前临床上治疗糖尿病过程中,中医外治法以及食疗的应用规律特点。文本挖掘的结果,选取高频数参与网络关系图的构建,展示出糖尿病患者临床上的常见中医外治法、常用食疗素材,直接客观的反应临床应用的现状。回溯原文献,人工阅读评价,使得结果更加客观,一维频次和二维网络构图,从不同角度显示挖掘对象在文献数据集中的分布,均有不同的意义,可以为临床医师治疗糖尿病时,优化治疗方案提供有益的参考。

[参考文献]

[1] 潘长玉. 中国糖尿病控制现状—指南与实践的差距[J]. 国外医学:内分泌分册, 2005, 25(3):174-178.

[2] 黄晓华. 中医辨证治疗糖尿病性胃轻瘫 86 例[J]. 四川中医, 2010, 28(10):71-72.

[3] 任昶, 耿福太, 张蓉洲. 中医辨证治疗口服降糖西药控制不佳 2 型糖尿病患者 46 临床观察[J]. 中医杂志, 2012, 53(14):1213-1216.

[4] 陈威妮, 姜森, 郑光, 等. 利用文本挖掘技术探索中成药和西药治疗 2 型糖尿病的用药规律[J]. 中国实验方剂学杂志, 2011, 17(21):299-302.

[5] Zheng G, Jiang M, He X, 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(1):1-18.

[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):59-63.

[7] Lamirel J C, Shehabi A J, Francois C, et al. Intelligent patent analysis through the use of a neural network: experiment of multi-viewpoint analysis with the multi SOM model [C]. Japan: The ACL 2003 Workshop on Patent Corpus Processing, Sapporo, 2003.

[8] Rodriguez-Esteban R. Biomedical text mining and its applications [J]. PLOS Compute Biol, 2009, 5(12):e1000597.

[9] Li S, Zhang Z Q, Wu L J, et al. Understanding ZHENG in traditional Chinese medicine in the context of neuroendocrine-immune network[J]. IET Syst Biol, 2007, 1(1):51-60.

[10] Head K A. Peripheral neuropathy: pathogenic mechanisms and alternative therapies[J]. Altern Med Rev, 2006, 11(4):294-329.

[11] 赵建玲, 李忠仁. 温和灸治疗糖尿病周围神经病变临床观察[J]. 中国针灸, 2008, 28(1):13-16.

[12] 潘志伟, 何春阳, 周凯, 等. 半导体激光对创伤愈合的影响作用机制研究[J]. 中国激光医学杂志, 2002, 9(4):263-264.

[13] 成侃, 王懿, 占世坤, 等. 人体半导体激光外照射的疗效及其作用机理研究[J]. 激光生物学报, 2002, 9(2):133-136.

[14] 张继红, 焦晓明. 慢跑配合手足按摩对 2 型糖尿病的疗效观察[J]. 中国康复医学杂志, 2009, 24(3):275-276.

[15] 张忠泉, 陈百泉, 许启泰. 山药多糖对大鼠血糖及胰岛释放影响的研究[J]. 上海中医药杂志, 2003, 37(10):52-53.

[16] Chen Z H, Li J, Liu J, et al. Saponins isolated from the root of panax notoginseng showed significant anti-diabetic effects in KK-Ay mice [J]. Am J Chin Med, 2008, 36(5):939-951.

[17] Okabe H, Miyahara Y, Tatsuo Y, et al. Studies on the constituents of *Momordica charantia* L. isolation and characterization of momordicosides A and B, glycosides of a pentahydroxyl-cucurbitane triterpene [J]. Chem Pharm Bull, 1980, 28(2):2753-2762.

[18] 张瑜, 关键, 崔炯漠, 等. HPLC 测定不同产地苦瓜中降糖活性成分苦瓜皂苷元 L 含量[J]. 中国中药杂志, 2010, 35(5):620-622.

[责任编辑 邹晓翠]