

· 学术探讨 ·

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

陈威妮¹, 姜淼², 郑光^{2,3}, 郭洪涛^{2,4}, 周奇², 吕爱平^{2*}

(1. 广州中医药大学, 广州 510405; 2. 中国中医科学院中医临床基础医学研究所, 北京 100700;
3. 兰州大学信息学院, 兰州 730000; 4. 上海中医药大学, 上海 201203)

[摘要] 目的: 利用文本挖掘技术, 探索中成药、西药治疗 2 型糖尿病的用药规律以及中成药与西药联合应用情况。方法: 在中国生物医学文献数据库(CBM)中收集治疗 2 型糖尿病的相关文献, 建立 Access 数据库, 运用 SQL 对数据进行处理, 结合人工降噪后, 分析中成药、西药对 2 型糖尿病的治疗用药规律。结果: 消渴丸、丹参注射液、黄芪注射液、六味地黄丸、玉泉丸等中成药为治疗 2 型糖尿病文献中出现的高频药物, 滋肾养阴是中医治疗 2 型糖尿病的重要法则。二甲双胍、胰岛素泵、皮下注射胰岛素、罗格列酮、瑞格列奈、格列齐特、格列吡嗪、阿卡波糖、吡格列酮等药物或治疗方式, 为治疗 2 型糖尿病文献中出现的高频西药或治疗方式。玉泉丸与二甲双胍联用, 六味地黄丸与格列齐特联用是最常用的中西药联用组合。结论: 利用文本挖掘的方法, 从文献报道频数方面提供中西医疗 2 型糖尿病的用药规律, 为临床选药和基础研究提供参考, 中西药联合应用在药物作用机制、适应症等方面, 还值得进一步研究。

[关键词] 2 型糖尿病; 文本挖掘; 中成药; 西药; 用药规律

[中图分类号] R287 **[文献标识码]** A **[文章编号]** 1005-9903(2011)21-0299-04

Exploring Rules of Chinese Patent Medicines and Western Medicines in Treatment of Type 2 Diabetes with Text Mining

CHEN Wei-ni¹, JIANG Miao², ZHENG Guang^{2,3}, GUO Hong-tao^{2,4}, ZHOU Qi², LV Ai-ping^{2*}

(1. Guangzhou University of Chinese Medicine, Guangzhou 510405, China;

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

3. School of Information Science and Engineering Technology, Lanzhou University, Lanzhou 730000, China;

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

[Abstract] **Objective:** To explore the rules of Chinese patent medicines, western medicines, and their combination rules in the treatment of type 2 diabetes through data mining. **Method:** We queried and downloaded the data on type 2 diabetes from Sino Med on March, 2011. By transferring semi-structured data structured database of Microsoft SQL was set into, we got the raw data for data mining. By executing the algorithm of discrete derivatives, we got some simple and meaningful networks of type 2 diabetes. In cytoscape, these networks were visualized. **Result:** In the treatment of type 2 diabetes, Chinese patent medicines, i. e., Xiaoke Wan, Salvia Miltiorrhiza

[收稿日期] 20110603(009)

[基金项目] 国家自然科学基金杰出青年项目(30825047); 国家科技部创新方法专项项目(2008IM020400); 国家科技部创新方法专项项目(2008IM020900)

[第一作者] 陈威妮, 博士, 研究方向: 中西医结合防治糖尿病及其并发症, Tel:010-64014411, E-mail: whitecl@163.com

[通讯作者] * 吕爱平, 博士生导师, 研究员, 研究方向: 疾病证候分类, E-mail: lap64067611@126.com

Injection, Radix Astragali Injection, Liuweidihuang Wan, and Yuquan Wan were most frequently used. among all the western medicines, metformin, insulin pump, subcutaneous injection of insulin, rosiglitazone, repaglinide, gliclazide, glipizide, acarboseare, and pioglitazone are most concerned in the treatment of type 2 diabetes. **Conclusion:** Rules of Chinese patent medicines, west medicines, and their combinations in the treatment of type 2 diabetes are proposed in this paper.

[**Key words**] type 2 diabetes; text mining; chinese patent medicine; western medicine; drug laws

2 型糖尿病 (type 2 diabetes mellitus, T2DM) 是严重威胁人类健康的疾病之一, 主要以胰岛素抵抗伴胰岛素分泌不足, 长期慢性高血糖为主要特征的代谢性疾病, 2 型糖尿病占糖尿病患者总数的 90% 以上^[1]。2007 年我国约有糖尿病患者 3 980 万, 预计 2025 年将增至 5 930 万, 已成为仅次于印度的世界第二糖尿病大国^[2]。

糖尿病属中医学“消渴病”范畴, 在长期临床实践和中医理论指导下, 应用中医药治疗糖尿病积累了丰富的经验。中、西医治疗 2 型的文献在医学文献数据库中大量存在, 而研究中成药及西药联合用药规律的文献, 尚未见相关报道。本研究利用文本挖掘技术, 从文献报道治疗 2 型糖尿病中成药、西药用药频率方面, 探索中成药、西药治疗 2 型糖尿病的用 药 规 律, 为 临 床 用 药 和 基 础 研 究 提 供 思 路。

1 方法和材料

1.1 文本数据采集 登录中国生物医学文献数据库 (Chinese Bio Medical Literature Database, CBM) 在主题检索下检索关键词“2 型糖尿病”, 经过检索, 出现款目词、主题词、命中文献数, 合并检索主题词, 共得到文献 32 563 篇 (检索日期 2011 年 3 月 6 日)。

1.2 文本数据处理 将采集的数据按照先后顺序整合到一个平面文件 (后缀 TXT) 中, 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, 登字第 0261882 号, 登记号 2010SR073409) 对下载的非结构化文本数据进行信息提取, 保存成格式化的、便于数据库 (Access 和 Microsoft SQL Server) 处理的格式。

1.3 文本挖掘以及分析 在同一篇文章中出现的关键词, 部分反映整篇文章的信息, 相关的关键词之间存在着“共同出现”的特点。这种“共同出现”不是随机的, 而是蕴含有一定的意义^[3], 尤其对于高频共同出现的关键词对, 在一定程度上反映了科研工作者的重视程度, 因此关键词对可作为文本挖掘的主要素材^[4-6]。对于 1 篇文献中, 重复出现的关键词,

只计算 2 次。在本研究中, 首先采用图 1 的算法构造针对每 1 篇文献共同出现的关键词对, 得到 DN_pairs 数据表。然后采用图 2 的算法, 将上述数据表中相同的关键词对进行合并处理, 仅保留它们出现的频数, 得到 DN_pairs_frqcy 数据表。

```
USE Table_Initial
FOR each PMID
  k =Number_of_DescriptorName(PMID)
  j = 1
  FOR DescriptorNames(i) (i =, 1, 2, ..., k)
    DO while j ≤ k
      DescriptorNames_Pair=DescriptorNames(i)+
        DescriptorNames(j)
      j = j + 1
      OUTPUT DescriptorName_Pair INTO
        table DN_pairs
    ENDDO
  j = 1
  ENDFOR
ENDFOR
```

图 1 构建关键词对程序算法

```
USE table DN_pairs
k =max_line_number
DO while k ≥ 1
GO top
FOR DescriptorName_Pair(1) //The 1st pairs in CHD.RA
  COUNT its Frequency
EndFor
OUTPUT DescriptorName_Pair, Frequency INTO table
  DN_pairs_Frqcy
DELETE all DescriptorName_Pair(1) from table
  DN_pairs
k =max_line_number
ENDDO
```

图 2 合并筛选关键词对程序算法

1.4 应用专业知识进行数据清洗 经过医学专业知识对图 2 (DN_pairs_frqcy) 中的数据进行评估, 发现针对特定的疾病, 图 2 中仍存在噪音问题。这些噪音不是关键词的简单重复, 而是相对于专业知识来说的噪音问题。比如说, 在挖掘分析中药的过程中, 中药“白及”, 大多在“蛋白及…”中出现, 降糖西药“度和”(二甲双胍格列齐特片), 在“胰岛素抵抗程度和……”等语句中出现。这些噪音的产生, 主要是自然语言的二义性和表达方式的多样性产生的。对于这类问题, 通过对得到的药物进行原文回溯, 逐个分析, 建立规则, 根据规则, 进行数据的清洗。

1.5 数据的可视化 从 DN_pairs_frqcy 数据表中

抽出不同頻數的關鍵詞對,用 Cytoscape 2.7 進行可視化處理,得到治療 T2DM 的中成藥、西藥、及其聯合用藥和中草藥的文獻頻數圖。

2 結果

2.1 治療 2 型糖尿病中成藥文獻頻數圖(圖 3)

文本挖掘結果顯示,消渴丸、丹參注射液、黃芪注射液、六味地黃丸、玉泉丸等為治療 2 型糖尿病最常用的中成藥(頻數 > 16 次)。

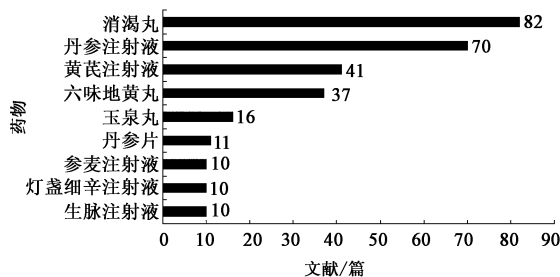


圖 3 治療 2 型糖尿病中成藥文獻頻數
(僅取頻數 > 10 次者)

2.2 治療 2 型糖尿病西藥文獻頻數圖(圖 4)

文本挖掘結果顯示,二甲雙胍、胰島素泵、羅格列酮、超短效胰島素、預混胰島素、瑞格列奈、長效胰島素、格列齊特、格列吡嗪、阿卡波糖、吡格列酮依次為治療 2 型糖尿病的高頻西藥或治療方式(頻數大於 281 次)。

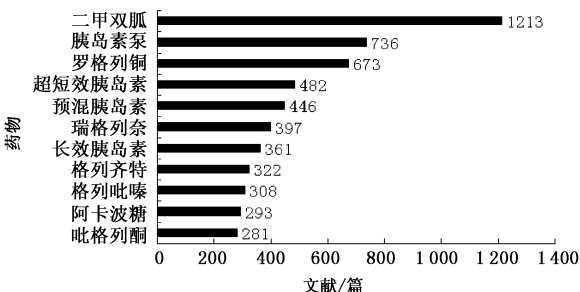


圖 4 治療 2 型糖尿病西藥文獻頻數
(僅取頻數 > 281 次者)

2.3 骨質疏松症西藥、中成藥聯合使用文獻頻數圖(圖 5)

文本挖掘結果顯示,二甲雙胍、格列齊特是口服降糖西藥與中成藥聯用時最常用的西藥,其分別與玉泉丸、六味地黃丸聯合使用最多(聯用頻數均為 4 次),這兩種中成藥也是聯合用藥中使用頻率最高的口服中成藥。丹參注射液是治療 2 型糖尿病文獻報道頻數最多的中藥注射劑,其與維生素 B₁₂ 聯用次數最多(聯用頻數為 11 次),追溯原文獻,丹參注射液與維生素 B₁₂ 聯合多用於治療糖尿病周圍神經

病變。

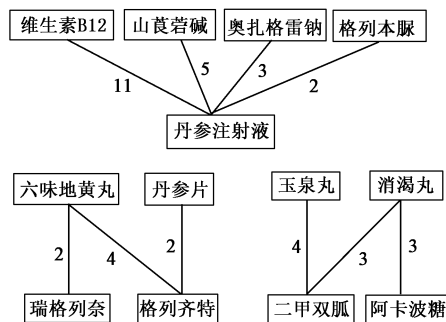


圖 5 治療 200 型糖尿病西藥、中成藥聯合使用
文獻頻數(僅取聯合使用頻數 > 2 次者)

3 討論

將數據挖掘和知識發現應用於中醫學領域的研究,是中醫學現代化研究的重要組成部分。採用文本挖掘技術從藥方、病的關係中分析治療規律和思維模式,可輔助醫生潛方用藥,以提高療效,還能為中藥新藥的臨床和實驗研究提供目標和思路,減少盲目性,縮短研究周期。通過數據挖掘尋找規律和新知識是生物學研究的熱點^[7]。

文本挖掘結果顯示,消渴丸、丹參注射液、黃芪注射液、六味地黃丸、玉泉丸等是治療 2 型糖尿病的最常用中成藥。李可建^[8]對消渴丸治療 2 型糖尿病隨機對照試驗進行系統評價,結果顯示消渴丸治療 2 型糖尿病有效,並呼喚繼續進行高質量研究予以支持。丹參注射液是丹參提取製成,臨床和實驗研究表明,丹參注射液聯合用藥或單用,多用於治療糖尿病周圍神經病變、糖尿病腎病或其他糖尿病血管病變。黃芪注射液是從黃芪中提取有效成分製成,常用於治療糖尿病腎病、糖尿病心臟病變,黃芪注射液對腎臟的保護作用,可能與促進骨形成蛋白 7 及 mRNA 的調控、抑制結締組織生長因子的表達有關,從而延緩腎臟纖維化變,緩解糖尿病腎病的發生、發展有關^[9]。倪青等^[10]觀察 1 274 例 2 型糖尿病併發冠心病治療規律,發現六味地黃丸是處方用藥中使用最多的方劑之一。玉泉丸由具有養陰生津,止渴除煩,益氣和中作用。鄧銀泉等^[11]發現玉泉丸除可改善糖尿病症狀、降低血糖外,還降低糖尿病患者促炎細胞因子水平以改善胰島素抵抗。以上 5 種中成藥主要從養陰、滋腎、健脾益氣、活血化癥等角度,針對 2 型糖尿病患者腎陰不足、陰精虧損、脾虛不運、瘀血阻絡進行改善。可見,滋腎養陰是中醫治療 2 型糖尿病的重要法則,對糖尿病併發症的治療,還可

从益气健脾、活血通络角度入手。

2 型糖尿病的西医疗疗,主要针对胰岛素抵抗和胰岛素分泌受损两方面,口服降糖药分为^[12]促胰岛素分泌剂和非促胰岛素分泌剂。磺脲类药物,主要作用为刺激胰岛 β 细胞分泌胰岛素,增加体内胰岛素水平,磺脲类药物使用不当可导致低血糖,本研究挖掘出的格列齐特、格列吡嗪属第二代磺脲类药物。格列奈类药物如瑞格列奈,通过刺激胰岛素早期分泌,以降低餐后血糖,此类药物也可引发低血糖。双胍类药物主要减少肝脏葡萄糖的输出而降低血糖,单独使用二甲双胍类药物不导致低血糖,但常引起胃肠道不适等副作用。噻唑烷二酮类主要通过促进靶细胞对胰岛素的反应而改善胰岛素敏感性,其常见副作用是体重增加和水肿,有潜在心衰的患者用这类药物可导致心衰加重,罗格列酮、吡格列酮均属本类药物。 α -糖苷酶抑制剂如阿卡波糖,主要延缓碳水化合物在肠道内的吸收,降低餐后血糖,并通过改善餐后糖负荷而改善空腹血糖,常见不良反应为胃肠道反应。胰岛素治疗也是控制高血糖的主要手段,在血糖较高的初发 2 型糖尿病患者或口服降糖药物控制不理想的患者中,应用广泛。

2 型糖尿病各种并发症错综复杂,中西药联合治疗 2 型糖尿病在临床应用中很常见。文本挖掘结果显示,玉泉丸与二甲双胍、六味地黄丸与格列齐特是口服药物中联用频率最高的;丹参注射液与维生素 B12 联用是中药注射剂与西药联用次数最多的,多用于治疗糖尿病周围神经病变。中西医结合理想的治疗方案是在循证医学指导下各种治疗方法的有机结合,应以提高临床疗效、减少不良反应和不增加医疗开支为目的^[13]。因此,对于临床常用的中成药、西药组合,应当从临床经验总结,到实验研究和临床研究证实,再回归临床推广应用,研究药物联用的科学性、合理性和必要性,阐明中成药和西药相互作用机制、有无不良反应、近期和远期疗效,规范中成药与西药联用的适应症、剂量、疗程等。因 2 型糖尿病急、慢性并发症复杂多样,本研究未对不同并发症的文献进行单独挖掘,这也是本研究的不足之处,在以后的研究中,可以更细致地对 2 型糖尿病不同并发症进行分类研究。

综上,文本挖掘是分析中成药、西药治疗 2 型糖尿病用药规律的有效方法,客观地提供了治疗 2 型糖尿病中成药与西药联用的常见组合,为合理地选

用中成药、西药提供依据,也为进一步的基础研究、临床研究和药物相互作用机制提供参考。

[参考文献]

- [1] WHO Media Center: Diabetes [EB/OL]. WHO DATABASE [2009-11]. <http://www.who.int/mediacentre/factsheets/fs312/en/index.html>.
- [2] 迟家敏.实用糖尿病学[M].3版.北京:人民卫生出版社,2010:2.
- [3] Andrea C, Rasmus P. Finding associations and computing similarity via biased pair sampling [C]. Florida; 2009 ninth IEEE international conference on data mining. 2009: 61.
- [4] Jeffrey W S. Data mining: An overview [C]. America: CRS Report RL31798,2004.
- [5] Brigitte M, Silke E. Five steps to text mining in biomedical literature [C]. Pisa; Proceedings of the second european workshop on data mining and text mining for bioinformatics, 2004: 47.
- [6] Zheng G, Jiang M, Xu Y S, et al. Discrete derivative algorithm of frequency analysis in data mining for commonly-existed biological networks [C]. Wuhan: CNMT 2010,2010: 5.
- [7] 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.
- [8] 李可建.消渴丸治疗 2 型糖尿病随机对照试验系统评价 [J]. 医药导报,2009,28(2):257.
- [9] 王收宝,田林红,孙洁,等.黄芪注射液对糖尿病大鼠肾脏 BMP-7 和 CTGF 表达的影响 [J]. 现代中西医结合杂志,2011,2(4):410.
- [10] 倪青,于丽红,陈世波.1274 例 2 型糖尿病合并冠心病证治规律数据挖掘 [J]. 中国中医药信息杂志,2010,17(9):99.
- [11] 邓银泉,范小芬,吴国琳,等.玉泉丸对 2 型糖尿病促炎细胞因子干预的影响 [J]. 中国中西医结合杂志,2006,26(8):706.
- [12] 中华医学会糖尿病分会.中国 2 型糖尿病防治指南(2007 年版) [J]. 中华医学杂志,2008,88(18):1227.
- [13] 徐浩,史大卓,刘保延,等.以临床实践数据为导向构建中西医结合临床指南的设想 [J]. 中国中西医结合杂志,2009,29(6):544.

[责任编辑 邹晓翠]