

# 基于文本挖掘探索溃疡性结肠炎药物的治疗规律

刘元艳<sup>1</sup>, 刘学文<sup>2</sup>, 谭勇<sup>3</sup>, 赵思宇<sup>1</sup>, 郭洪涛<sup>3</sup>, 郑光<sup>3</sup>, 姜森<sup>3</sup>, 吕爱平<sup>3\*</sup>

- (1. 北京中医药大学中药学院, 北京 100102;
2. 首都医科大学潞河教学医院中医科, 北京 101100;
3. 中国中医科学院中医临床基础医学研究所, 北京 100700)

**[摘要]** **目的:**探索溃疡性结肠炎的中西药用药规律。**方法:**利用数据挖掘技术,结合可视化,在中国生物医学文献数据库中收集治疗溃疡性结肠炎的文献数据,采用基于敏感关键词频数统计的数据分层算法,结合文献回溯、人工阅读降噪,挖掘溃疡性结肠炎的中西药治疗规律。规律通过一维频次及二维的网络图进行显示。**结果:**治疗溃疡性结肠炎使用频数排在前五的西药依次为柳氮磺吡啶、甲硝唑、地塞米松、泼尼松和氢化可的松,地塞米松最常与其他西药联用。治疗溃疡性结肠炎使用频数排在前五的中成药依次为云南白药、乌梅丸、补脾益肠丸、理中丸和西瓜霜,中草药依次为黄连、白术、白头翁、黄芪和乌梅。补中益气丸和补脾益肠丸常与其他中成药联用,黄芪、黄连、白术常与其他中草药联用,补脾益肠丸、云南白药和乌梅丸常与地塞米松、甲硝唑、柳氮磺吡啶联用。**结论:**文本挖掘技术能较客观地总结溃疡性结肠炎中西药的用药规律,为临床经验总结提供有益的探索与参考。

**[关键词]** 文本挖掘; 溃疡性结肠炎; 药物治疗规律; 中西药; 配伍规律

**[中图分类号]** R283.6 **[文献标识码]** A **[文章编号]** 1005-9903(2013)15-0329-04

**[doi]** 10.11653/syjf2013150329

## Exploration of Drug Treatment Law for Ulcerative Colitis Based on Text Mining Technology

LIU Yuan-yan<sup>1</sup>, LIU Xue-wen<sup>2</sup>, TAN Yong<sup>3</sup>, ZHAO Si-yu<sup>1</sup>,  
GUO Hong-tao<sup>3</sup>, ZHENG Guang<sup>3</sup>, JIANG Miao<sup>3</sup>, LV Ai-ping<sup>3\*</sup>

- (1. School of Material Medica, Beijing University of Chinese Medicine, Beijing 100102, China;
2. Department of Traditional Chinese Medicine, Luhe Teaching Hospital, Capital Medical University, Beijing 101100, China;
3. Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China)

**[Abstract]** **Objective:** To explore treatment law of ulcerative colitis with traditional Chinese medicine and western medicine. **Method:** Taking data mining technology, combined with visualization, literarure data for treating ulcerative colitis were collected in Chinese biomedical literature database. Data slicing algorithm based on sensitive keywords frequency statistics was adopted, combined with literature review and artificial reading noise reduction, then treatment law of traditional Chinese medicine and western medicine for ulcerative colitis was investigated. Law showed of one-dimensional and two-dimensional network diagram. **Result:** Using frequency of western drugs for treatment of ulcerative colitis was in order of sulfasalazine > metronidazole > dexamethasone > prednisone > hydrocortisone, dexamethasone was commonly associated with other western medicine; Using frequency of Chinese patent drugs was in order of Yunnan Baiyao > Wumei pills > Bupi Yichang pills > Lizhong

**[收稿日期]** 20130131(011)

**[基金项目]** 科技部“重大新药创制”项目(2009ZX09502-019);科技部创新方法学专项(2008IM020900);国家自然科学基金青年科学基金项目(30902000,81001623);北京中医药大学课题(2009JYBZZ-JS037)

**[第一作者]** 刘元艳,副教授,博士,从事中药化学研究, Tel:010-84738658, E-mail:yyliu\_1980@hotmail.com

**[通讯作者]** \*吕爱平,研究员,博士生导师,从事病证关联研究, Tel:010-64067611, E-mail:lap64067611@126.com

pills > mirabilitum praeparatum, Buzhong Yiqi pills and Bupi Yichang pills was always associated with other Chinese patent drugs; Chinese materia medica was in order of Rhizoma Coptidis > Rhizoma Atractylodis Macrocephalea > Radix Pulsatillae > Radix Astragali > Fructus Mume, Rhizoma Coptidis, Rhizoma Atractylodis Macrocephalea and Radix Astragali often combined with other herbs; Bupi Yichang pills, Yunnan Baiyao and Wumei pills were always associated with dexamethasone, metronidazole and sulfasalazine. **Conclusion:** Text mining technology could be adopted for objective summing up treatment law of traditional Chinese medicine and western medicine for ulcerative colitis, then provide useful exploration and reference to clinical experience.

[ **Key words** ] text mining; ulcerative colitis; medication treatment law; traditional Chinese medicine and western medicine; compatibility regularity

溃疡性结肠炎(UC)是一组慢性、炎症性、特发性肠道疾病状态<sup>[1]</sup>,其最佳药物治疗方法呈多方面和个性化<sup>[2]</sup>。UC多被归于中医学“泄泻”、“痢疾”、“腹痛”、“肠风”、“脏毒”、“肠癖”等范畴。既往研究表明,中医和中西医结合治疗溃疡性结肠炎疗法优化、疗效可靠。本实验采用数据挖掘技术<sup>[3-5]</sup>,结合原文献回溯,人工阅读分析等方法,对大量中文文献进行挖掘分析,探索中西药治疗 UC 的用药规律,为 UC 的临床用药提供参考。

### 1 材料

**1.1 数据采集** 登录中国生物医学文献数据库(CBM),检索关键词“溃疡性结肠炎”,共检索到 17 078 篇文献(检索日期截止至 2012 年 4 月 13 日),选择“详细”和“显示全部”的显示格式,以获得每篇文献的流水号、标题、摘要、主题词等信息,备用。

**1.2 文本数据处理** 将采集的数据按先后顺序整合至一个平面文件(后缀 TXT)中,以 ANSI 编码格式保存。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,相关关键词间存在着“共同出现”的特点,且该特点不是随机的,蕴含一定意义<sup>[6]</sup>,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。这些关键词对可作为文本挖掘的主要素材<sup>[6-8]</sup>。采用图 1 的算法构造针对每一篇文献共同出现的关键词对,得到 DN\_pairs 数据表;通过图 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。

**1.4 数据的可视化** 从 DN\_pairs\_frqcy 数据表中抽出不同频数的关键词对,用 Cytoscape 2.7 进行可视化处理,获得反映中西药单独及联用治疗溃疡性结肠炎用药情况的柱状图和网络图。

```
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 UC
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

## 2 方法与结果

**2.1 西药文本挖掘** 西药文本挖掘一维频数结果见图 3,显示结肠炎相关西药共 130 种,使用频数前九名由高到低依次为柳氮磺吡啶、甲硝唑、地塞米松、泼尼松、氢化可的松、磺胺嘧啶、硫唑嘌呤、西咪替丁、胱氨酸。西药文本挖掘二维网络结果显示结肠炎相关西药组合共 171 种,选取 ≥2 的文献频数构建网络图(图 4),发现网络中药物的显示度和 node 大小呈正相关,地塞米松、甲硝唑、泼尼松、硫唑嘌呤、柳氮磺吡啶、甲氨蝶呤、氧氟沙星、胱氨酸等具有较高显示度,提示这些西药常被联合使用。

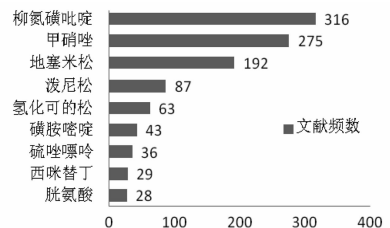


图 3 治疗结肠炎的西药使用频数

**2.2 中成药及中药文本挖掘** 中成药及中药文本挖掘一维频数结果分别见图 5,6,显示结肠炎相关

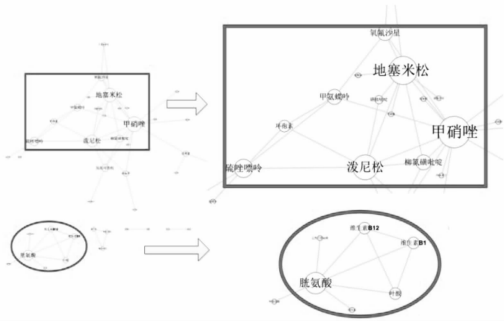


图4 治疗结肠炎的西药组合网络(组合频数≥2)

中成药共33种,中草药共219种,使用频数前10名的中成药由高到低依次为云南白药、乌梅丸<sup>[14]</sup>、补脾益肠丸、理中丸、西瓜霜、珍珠散、逍遥丸、香连丸<sup>[15]</sup>、健脾丸、驻车丸。使用前十名的中草药由高到低依次为黄连、白术、白头翁、黄芪、乌梅、苦参、白及、黄柏、黄芩、丹参。

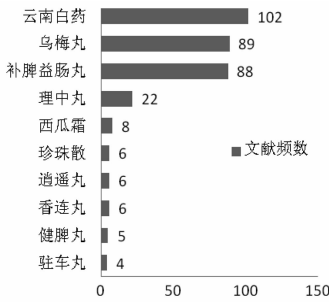


图5 治疗结肠炎的中成药使用频数

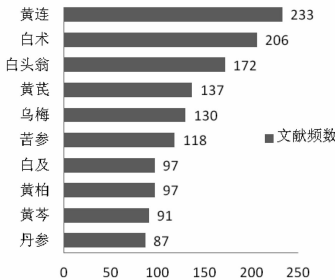


图6 治疗结肠炎的中药使用频数

中成药文本挖掘二维网络结果显示,结肠炎相关中成药组合共5种,中药组总共2602种,分别选取≥1,3的文献频数构建网络图,结果见图7,8。网络中药物的显示度和node大小呈正相关,补中益气丸、补脾益肠丸等具有较高显示度,提示这2个中成药常与其他中成药联合使用。黄芪、黄连、白术、党参、茯苓、当归、白芍、柴胡、大黄、黄柏、黄芩、白头翁、乌梅、诃子、苦参、地榆、白及、苍术、补骨脂、三七、赤芍、蒲公英等具有较高显示度,提示这些中草药常被联合使用。

**2.3 西药、中成药联用文本挖掘** 文本挖掘显示,结肠炎相关西药、中成药联用组合共25种,选取≥1和≥4的文献频数分别构建网络图9A,9B,发现网

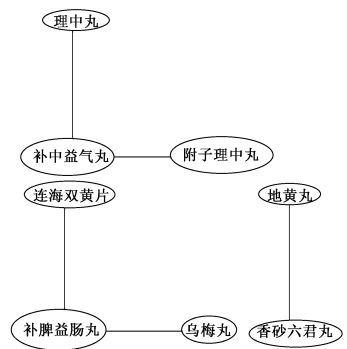


图7 治疗结肠炎的中成药组合网络

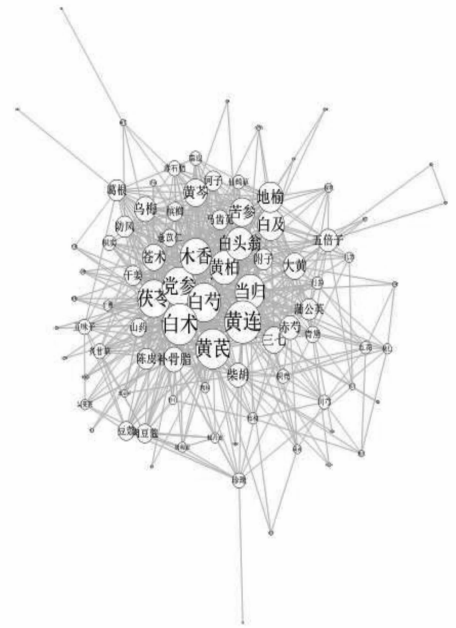


图8 治疗结肠炎的中草药组合网络(组合频数≥3)

络中药物显示度和node大小呈正相关,补脾益肠丸、云南白药、乌梅丸常与地塞米松、甲硝唑、柳氮磺吡啶联用。

### 3 讨论

文本挖掘技术是以统计数理分析、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的交叉学科<sup>[9]</sup>。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方面。根据中医理论或专业知识,计算机挖掘中医药文献库与生物医学信息库的非关联知识可为中西医结合研究提供新的思路和途径,且结果更客观,可重复性强<sup>[10]</sup>。

西药文本挖掘结果表明,5种常用药物为柳氮磺吡啶、甲硝唑、地塞米松、泼尼松和氢化可的松,符合临床实践中氨基水杨酸制剂、糖皮质激素为主要治疗药物的实际,而甲硝唑临床中多用于灌肠;地塞米松是最常与其他西药联用治疗UC的糖皮质激素类药物,与其具有更强烈抗炎作用有关。

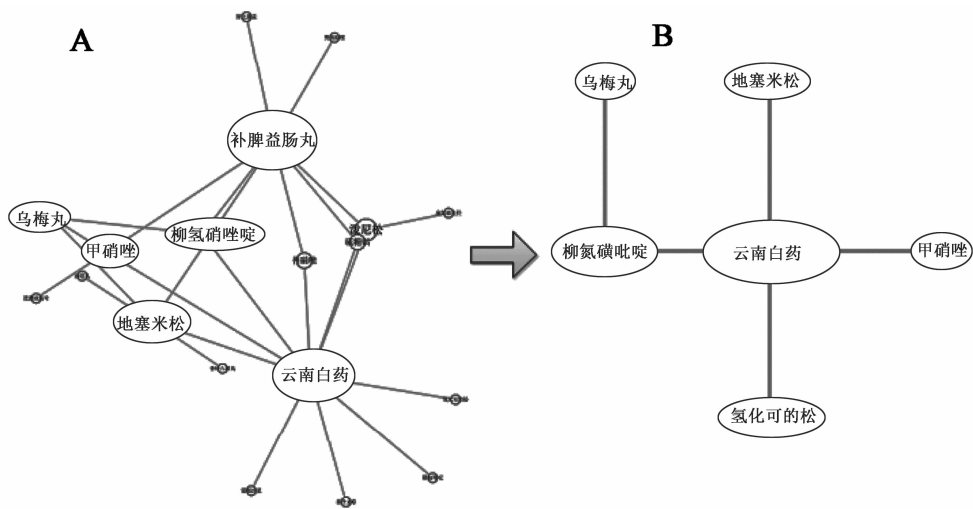


图 9 治疗结肠炎的西药与中成药联用网络

中医治疗溃疡性结肠炎急性发作期以清热化湿为主,缓解期以健脾益气为本<sup>[11]</sup>。挖掘所得中成药中,常用补脾益肠丸、补中益气丸、理中丸、健脾丸、逍遥丸、驻车丸等,均体现了健脾益气、治病求本的治则。云南白药具有养血、活血、止血、解毒功效,用于溃疡性结肠炎的治疗,攻补兼施、标本兼治,是最常用的中成药。西瓜霜清热解毒,用其保留灌肠有利于发挥药效而达到消除病邪、修复损伤组织的作用,对溃疡性结肠炎有较好疗效<sup>[12]</sup>。由此可见,将甲硝唑及西瓜霜一同用于灌肠,可作为治疗溃疡性结肠炎的新思路。

挖掘所得的常用中草药及其联用特点,体现了黄连解毒汤、白头翁汤和乌梅丸的组方配伍特征,反映了清热化湿解毒的中医治疗思路。许多其他功能类中药应用时,如补气用的黄芪、丹参等,虽使用频率不高,多为经验型用药或参考现代药理研究结果,但也体现了现代中医人在治疗溃疡性结肠炎时的多样性和灵活性。中西药联用治疗溃疡性结肠炎,能减轻西药的毒副作用,提高综合疗效,是目前治疗慢性溃疡性结肠炎的重要方法<sup>[13]</sup>。

[参考文献]

[ 1 ] Charles N B, Michael F, Krabshuis J H, et al. 2010 年世界胃肠病学组织关于炎症性肠病诊断和治疗的实践指南[J]. 胃肠病学, 2010, 15(9): 548.  
[ 2 ] Sohrabpour A A, Malekzadeh R, Keshavarzian A. Current therapeutic approaches in inflammatory bowel disease [J]. Curr Pharm Des, 2010, 16(33): 3668.  
[ 3 ] 郭洪涛, 郑光, 张弛, 等. 利用数据挖掘技术探索类风湿关节炎与糖尿病“同证”的科学基础[J]. 世界科学技术—中医药现代化, 2010, 12(5): 818.  
[ 4 ] 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: 5.  
[ 5 ] 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 Mining, 2011, 4 (1): 18.  
[ 6 ] Andrea C, Rasmus P. Finding associations and computing similarity via biased pair sampling [ C ]. Miami: ICDM, 2009: 61.  
[ 7 ] Jeffrey W S. Data mining: An overview [ R ]. CRS, 2004: RL31798.  
[ 8 ] Brigitte M, Silke E. Five steps to text mining in biomedical literature [ C ]. Pisa: In Proceedings of the Second European Workshop on Data Mining and Text Mining for Bioinformatics, 2004: 47.  
[ 9 ] 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.  
[ 10 ] LI S, ZHANG Z Q, WU L J, et al. Understanding ZHENG in traditional Chinese medicine in the context of neuro-endocrine-immune network [ J ]. IET Syst Biol, 2007, 1(1): 51.  
[ 11 ] 王新月, 刘果, 盛益华. 溃疡性结肠炎流行病学特点与中医病因病机探讨 [ J ]. 中国中医基础医学杂志, 2011, 17(5): 481.  
[ 12 ] 刘增刚, 冷秋萍. 西瓜霜喷剂灌肠治疗放射性直肠炎 32 例疗效观察 [ J ]. 山东医药, 2009, 49(37): 36.  
[ 13 ] 崔丽君, 孙洪莲. 中西药的配伍应用 [ J ]. 实用中医内科杂志, 2012, 26(1): 88.  
[ 14 ] 李燕, 史成和. 高忠英教授应用乌梅丸加减治疗慢性溃疡性结肠炎的临床经验 [ J ]. 中国实验方剂学杂志, 2012, 18(14): 322.  
[ 15 ] 王欣, 王伟明, 宋亚娟, 等. 香连胶囊对大鼠急性肠炎的作用 [ J ]. 中国实验方剂学杂志, 2013, 19(3): 212.

[责任编辑 全燕]