

## • 文献研究 •

利用文本挖掘技术分析中成药和西药治疗  
脑梗死的基本用药规律赵 宁<sup>1</sup> 郑 光<sup>1,2</sup> 谭 勇<sup>1</sup> 郭洪涛<sup>1,3</sup> 姜 淼<sup>1</sup> 吕爱平<sup>1\*</sup>

(1. 中国中医科学院中医临床基础医学研究所,北京市东直门内南小街 16 号,100700;

2. 兰州大学信息学院;3. 上海中医药大学)

基金项目:国家自然科学基金青年基金资助项目(30825047);国家自然科学基金资助项目(30902003);

科技部创新方法工作专项项目(2008IM020900)

\* 通讯作者:lap64067611@126.com, (010)64067611

**[摘要]** 目的 利用文本挖掘技术总结中成药和西药及其联合用药治疗脑梗死的基本用药规律。方法 收集中国生物医学文献数据库(CBM)中治疗脑梗死的相关文献,建立 Access 数据库,运用结构化查询语言对数据进行处理,并运用数据分层算法分析中成药和西药及其联合用药治疗脑梗死的基本用药规律。结果 共检索出相关文献 50 286 篇,其中治疗脑梗死所用西药频数最高的是尿激酶(1 241),其次是低分子肝素(980)、依达拉奉(867);所用中成药频数最高的是丹参注射液(1 063),其次是川芎嗪(431)、葛根素(294);西药和中成药联合应用频数最高的是奥扎格雷联合丹参注射液(84),其次是阿司匹林联合丹参注射液(57)、依达拉奉联合丹参注射液(55)。结论 利用文本挖掘的方法,可从文献报告频数方面分析中成药、西药及其联合应用治疗脑梗死的用药规律,并为研究疾病用药规律提供方法。

**[关键词]** 脑梗死;文本挖掘;西药;中成药

The Medication Regularity of Chinese Patent Medicine and Western Medicine for Cerebral Infarction by Text Mining

ZHAO Ning<sup>1</sup>, ZHENG Guang<sup>1,2</sup>, TAN Yong<sup>1</sup>, GUO Hongtao<sup>1,3</sup>, JIANG Miao<sup>1</sup>, LV Aiping<sup>1</sup>

(1. China Academy of Chinese Medical Sciences, Beijing 100700; 2. School of Information Science and Engineering Technology, Lanzhou University; Shanghai University of Traditional Chinese Medicine)

**ABSTRACT Objective** To summarize the medication regularity of Chinese patent medicine, western medicine and their combination for cerebral infarction by text mining. **Methods** The pertinent articles on cerebral infarction were collected from Chinese Biomedical Literature Database (CBM). Access database was established and data was processed with structured query language. Slicing algorithm was used to analyze the medication regularity of Chinese patent medicine, western medicine and their combination for cerebral infarction. **Results** Totally 50 286 articles were collected from CBM. The commonly used western medicines for cerebral infarction were urokinase (1 241), low molecular heparin (980) and edaravone (867) in the order of frequency. The commonly used Chinese patent medicines were Danshen Injection [Injection of Danshen (*Radix et Rhizoma Salviae Miltiorrhizae*)] (1 063), ligustrazine (431) and puerarin (294). The commonly used combinations were Danshen Injection with ozagrel (84), Danshen Injection with aspirin (57) and Danshen Injection with edaravone (55). **Conclusion** The text mining could analyze the medication regularity of Chinese patent medicine, western medicine and their combination for cerebral infarction. It is an important method to research the medication regularity of diseases for popularization and application.

**Keywords** cerebral infarction; text mining; western medicine; Chinese patent medicine

脑梗死,又称缺血性脑卒中,占全部脑卒中的 80%,具有高发病率、高病死率、高致残率的特征,已成为严重危害人类健康的疾病<sup>[1]</sup>,在中医学属于“中风”范畴。在生物学领域,由于生物学数据和生物医学文献数量的急剧增长,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<sup>[2]</sup>。文本挖掘是从文本集中发现潜在、隐藏的归纳性知

识的一门技术<sup>[3]</sup>,它能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<sup>[4]</sup>。既往研究表明<sup>[5]</sup>,文本挖掘具有探索治疗疾病用药规律的技术特征和条件,可操作性强。本研究利用文本挖掘技术,从文献报告的频数方面进行西药、中成药以及西药和中成药联合应用治疗脑梗死用药规律,以期临床治疗脑梗死用药提供参考。

## 1 材料与方法

### 1.1 文本数据收集

数据库:中国生物医学文献数据库(CBM);检索方式:主题检索;检索词:脑梗死、脑梗塞等;检索日期:1985—2011年。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50286篇。

### 1.2 文本数据处理

将数据整合至平面文件(TXT)里,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、便于数据库(Access)和大型数据库处理的格式。提取出来的信息主要是关键词(包括核心和非核心两种类型)。提取出来的数据,首先存入Access数据库,作为下一步数据处理的材料,然后导入结构化查询语言(SQL)中进行下一步的挖掘分析。

### 1.3 数据一次清洗

根据生成的Access数据库,将“结果”数据表导入SQL中,以“Table\_Initial”为表名称,针对“序号”和“关键词”进行处理。为了方便处理,将“序号”和“关键词”两个字段分别用PMID和Descriptor Name来表示。

经过对原文献的分析,发现相同的关键词在1篇文献的标题和摘要中存在着重复出现的问题。对于文本挖掘来说,假设每一篇文献的贡献度是相同的,按照这个假设,对于1篇文献中重复出现的,只需要计算1次。据此进行数据清洗工作。

### 1.4 数据挖掘及分析

针对得到的数据,应用文献[4]中的基于敏感关键词频数统计的数据分层算法,计算治疗脑梗死西药、中成药及其联合应用数据。

### 1.5 数据的可视化

根据得到的数据表DN\_pairs\_frqcy,抽出不同频数的关键词对,分别用Excel、Cytoscape 2.8等软件进行可视化处理,得到治疗脑梗死的西药、中成药及其联合用药的文献频数图。

## 2 结果

### 2.1 治疗脑梗死所用西药名称及文献频数

治疗脑梗死所用西药有359种,依据频数高低排序,97以上的文献频数如表1所示,所涉及的西药包括以下几类,溶栓药:尿激酶、降纤酶、巴曲酶、

蚓激酶;抗凝剂:低分子肝素、阿司匹林;抗血小板药:奥扎格雷、噻氯匹定;降纤剂:东菱克栓酶;神经保护剂:依达拉奉、纳洛酮、尼莫地平、胞二磷胆碱;对症治疗药:甘露醇;具有综合疗效的药物:前列地尔、丁咯地尔等。其中溶栓药和抗凝药的频数最高,说明临床工作者在对脑梗死的治疗中对这两类药物的应用最为广泛,使用频率最多。

表1 治疗脑梗死所用西药名称及文献频数(频数≥97)

| 药物名称  | 文献频数 |
|-------|------|
| 尿激酶   | 1241 |
| 低分子肝素 | 980  |
| 依达拉奉  | 867  |
| 阿司匹林  | 830  |
| 奥扎格雷  | 815  |
| 降纤酶   | 667  |
| 巴曲酶   | 658  |
| 纳洛酮   | 553  |
| 尼莫地平  | 543  |
| 胞二磷胆碱 | 509  |
| 甘露醇   | 485  |
| 前列地尔  | 198  |
| 东菱克栓酶 | 189  |
| 丁咯地尔  | 179  |
| 噻氯匹定  | 114  |
| 蚓激酶   | 97   |

### 2.2 治疗脑梗死所用中成药名称及文献频数

治疗脑梗死所用中成药有86种,6以上的文献频数如表2所示,其中丹参注射液、川芎嗪、葛根素和血塞通注射液使用频数最高,说明在治疗脑梗死的过程中,活血化瘀药的应用最为广泛。

### 2.3 治疗脑梗死西药、中成药联合使用情况及文献频数

治疗脑梗死西药和中成药联合应用的组合有191种,10以上文献频数如图1所示(图中横线代表药物两两之间的联系,数字代表两种药物联合使用在文献出现的篇数)。从图中可以看出,中成药应用都是单独出现,两两之间没有发生交联,而其联合的西药可以是一种或多种。联合应用的中成药的种类比较集中,以丹参注射液为最多,其次是川芎嗪、葛根素、血塞通注射液、血塞通注射液和舒血宁注射液等。其中,丹参注射液的联合用药囊括了几乎所有的西药,可以分别联合溶栓药、抗凝药、抗血小板聚集药、降纤药、神经保护剂等药物。

## 3 讨论

文本挖掘是从非结构化的文本数据中抽取有意

表 2 治疗脑梗死所用中成药名称及文献频数(频数 $\geq 6$ )

| 药物名称    | 文献频数  |
|---------|-------|
| 丹参注射液   | 1 063 |
| 川芎嗪     | 431   |
| 葛根素     | 294   |
| 血塞通注射液  | 210   |
| 舒血宁注射液  | 146   |
| 黄芪注射液   | 124   |
| 灯盏细辛注射液 | 121   |
| 血栓通注射液  | 115   |
| 灯盏花素注射液 | 97    |
| 红花注射液   | 78    |
| 脉络宁注射液  | 75    |
| 刺五加注射液  | 74    |
| 清开灵注射液  | 62    |
| 安宫牛黄丸   | 56    |
| 醒脑静注射液  | 55    |
| 香丹注射液   | 44    |
| 参麦注射液   | 40    |
| 丹参片     | 36    |
| 生脉注射液   | 35    |
| 复方丹参片   | 22    |
| 参附注射液   | 11    |
| 注射用灯盏花素 | 10    |
| 心脑舒通胶囊  | 8     |
| 脑血栓片    | 8     |
| 冠心病宁注射液 | 7     |
| 生脉胶囊    | 6     |

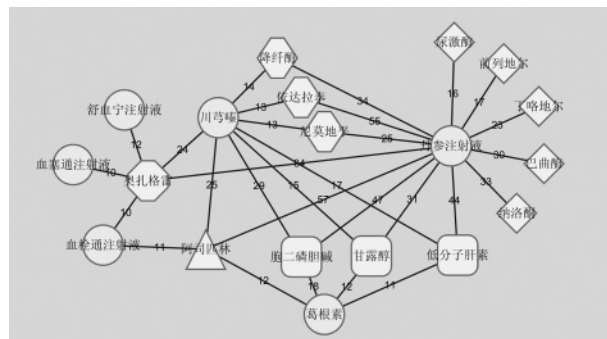


图1 治疗脑梗死西药、中成药联合使用情况及文献频数  
(频数 $\geq 10$ )

义的数据。具体说,文本挖掘应用到生物、医学上,可以分为文本数据收集、处理、结构化分析、可视化以及评价 5 个步骤。本研究挖掘了西药和中成药治疗脑梗死的用药规律,基本能够反映临床西药、中成药及其联合用药的情况。

治疗脑梗死的西药频数由高到低依次是尿激酶、低分子肝素、依达拉奉、阿司匹林、奥扎格雷、降纤酶、巴曲酶、纳洛酮等,涉及到的药物类型包括溶栓药、抗凝剂、抗血小板聚集药、降纤剂、神经保护剂、对症治疗药以及其他具有综合疗效的药物等,并以溶栓药和

抗凝药为最多,说明目前脑梗死的西药治疗是以溶栓、抗凝、抗血小板、降纤以及脑神经保护为主,兼具改善脑循环以及对症治疗,并且以溶栓为首选的治疗方法,这也和国际上对于治疗脑梗死最有效的途径是“溶解血栓”和“恢复血供”的认识相一致。

中成药的应用中,活血化瘀药的应用频数最高,其中丹参注射液、川芎嗪、血塞通注射液、舒血宁注射液、血栓通注射液、红花注射液 6 种排名靠前的中成药,所涉及的文章共有 2043 篇,已经占到中成药所涉及的 3338 篇文献的 61.20%,可见治疗脑梗死所用的中成药是以活血化瘀为主,这和中医对于中风病脑脉痹阻或血溢脉外,引起昏仆不遂,发为中风的病机相一致。

中成药和西药的联用中,丹参注射液居首位,其联合用药的范围也非常广泛,囊括了几乎所有的西药,说明丹参注射液仍然是西医医院里治疗脑梗死使用频率最高的中成药,同时也是研究最多的中成药。这虽然说明了西医医院对于中成药的使用已经非常广泛,同时也说明,西医对于中成药的使用还有盲目性和片面性,缺乏辨证论治;应当增强西医医生对中医药的认识以及规范其正确合理使用中成药。加强中医药知识的宣传和普及也是中医药工作者的重要任务。

采用文本挖掘技术分析脑梗死的西药、中成药用药情况,为临床医生研究疾病用药规律提供参考,并通过这些结果的分析,以及回溯查询,能够从整体上梳理出西药、中成药治疗脑梗死的用药规律,为认识疾病,更新知识,制定疾病的临床指南等方面提供有价值的参考,也为脑梗死的科学研究提供思路。

## 参考文献

- [1] 王维治. 神经病学[M]. 北京: 人民卫生出版社, 2006: 1.
- [2] 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): i547 - i553.
- [3] Al-Mubaid H, Singh RK. A text-mining technique for extracting gene-disease associations from the biomedical literature[J]. Int J Bioinform Res Appl, 2010, 6(3): 270 - 286.
- [4] Rodriguez-Esteban R. Biomedical text mining and its applications[J]. PLoS Comput Biol, 2009, 5(12): e1000597.
- [5] 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 - 822.

(收稿日期:2011-07-15;修回日期:2011-11-28)

疗药以及其他具有综合疗效的药物等,并以汤剂药和

编辑:崔京艳