

基于 LDA 主题模型与关联规则的岭南骨伤用药规律

武凤震¹, 胡劲红², 刘子晴¹, 陈凯佳^{2*}

(1. 广州中医药大学医学信息工程学院, 广州 510006;
2. 广州中医药大学基础医学院, 广州 510006)

[摘要] 目的:挖掘岭南骨伤名家用药规律。方法:收集整理岭南骨科名家著作、手稿、论文等文献,建立岭南骨伤方药数据库,采用频数分析,潜在狄利克雷分布(LDA)主题模型,关联规则等方法对其用药规律进行分析。结果:岭南骨伤文献中共涉及方剂802首、中药811味,所使用药物以温性最多,占28.3%,药物类别以活血化瘀药、补虚药、理气药、清热药、祛风湿药、温里药为主,LDA主题模型共得5个核心方,通过关联规则分析,共得出21条药对与药组关联规则,核心药物有当归、川芎、白芍、赤芍等。结论:岭南骨科名家基本遵从骨伤科用药的常规,重视活血化瘀法的治疗,用药呈现比较明显的骨伤早、中、晚3期用药规律,早期或兼清热解毒,或强调止血,或开窍醒神;中期攻补兼施,兼以补益肝肾;晚期以补为主,益气健脾养血,兼以活血、补益肝肾。在此基础上,用药也体现了重视健脾化湿清热的岭南骨伤特色治法。

[关键词] 岭南骨伤; LDA主题模型; 关联规则

[中图分类号] R287 **[文献标识码]** A **[文章编号]** 1005-9903(2017)11-0214-06

[doi] 10.13422/j.cnki.syfjx.2017110214

[网络出版地址] <http://kns.cnki.net/kcms/detail/11.3495.R.20170309.1014.052.html>

[网络出版时间] 2017-03-09 10:14

Medication Rules of Famous Doctors in Lingnan Bone Injury Based on LDA Thematic Model and Association Rules

WU Feng-zhen¹, HU Jing-hong², LIU Zi-qing¹, CHEN Kai-jia^{2*}

(1. College of Medical Information Engineering, Guangzhou University of Chinese Medicine, Guangzhou 510006, China; 2. Basic Medicine College, Guangzhou University of Chinese Medicine, Guangzhou 510006, China)

[Abstract] **Objective:** To dig up the medication rules used by famous doctors in Lingnan bone injury.

Method: The database of Lingnan bone injury prescriptions were established through collecting and analyzing famous doctors' works, manuscripts, papers and other literature. Frequency analysis, Latent Dirichlet Allocation (LDA) thematic model and association rules were used to analyze their medication rules. **Result:** The literature of lingnan bone injury involved 802 prescriptions and 811 kinds of traditional Chinese medicine (TCM) herbs. The warm drugs were used most, accounting for 28.3%. The drugs categories were dominated by dissipating blood stasis drugs, tonifying deficiency drugs, Qi-regulating drugs, heat-clearing drugs, removing wind-dampness drugs, and interior-warming drugs. Five core prescriptions were obtained from the LDA thematic model, and the through association rules analysis, a total of 21 herbal pairs and herbal group association rules. The core herbs included *Angelicae Sinensis Radix*, *Chuanxiong Rhizoma*, *Paeoniae Radix Alba*, *Paeoniae Radix Rubra* and so on.

Conclusion: The famous doctors of Lingnan bone injury emphasized the treatment of activating blood and resolving

[收稿日期] 20161216(007)

[基金项目] 广东省科学技术厅—广东省中医药科学院联合科研项目(2014A020221086);广东省科技计划项目(2014A020221039)

[第一作者] 武凤震,在读硕士,从事中医药信息学研究,Tel:13226651902,E-mail:26006968080@qq.com

[通讯作者] *陈凯佳,博士,副教授,从事岭南医学研究、中医文献与信息技术交叉研究,Tel:13660182826,E-mail:chenkaijia@163.com

stasis. On this basis, the drugs showed obvious early, middle and late stages of drugs use for bone injury: in early stage, clearing heat and detoxicating, or emphasizing hemostasis, or opening orifice and waking up; in middle stage, both attacking and tonifying, or both tonifying liver and kidney; in late stage, mostly tonifying (tonifying Qi, blood and spleen), combined with activating Qi, blood, and nourishing liver and kidney. On this basis, the medication also reflected the importance of invigorating spleen and eliminating dampness in characteristic treatment in Lingnan bone injury.

[Key words] Lingnan bone injury; LDA thematic model; association rules

岭南中医骨伤科学术流派以精确的理伤手法及独特的固定方法与有效伤科药剂著称于世,有蔡忠、管镇乾、李才干、梁财信、何竹林五大分支^[1],流派囊括何竹林、蔡荣、李广海、陈渭良、黄耀燊、岑泽波、陈基长等广东省名中医,名家创制的多种骨科方药,依然广泛应用于临床,如何竹林生肌膏、驳骨散、田七跌打风湿霜、蔡忠跌打万花油、黄耀燊骨仙片、双柏散、李广海跌打酒、陈渭良伤科油、三龙驳骨散、白药膏、还魂汤、理伤消肿口服液、疗筋膏等。对岭南骨伤名家所用方药进行研究,有利于探索发现岭南骨科常用药物、处方、药对,为骨科临床提供借鉴。课题组在前期岭南骨科名家著作、手稿、论文等文献收集整理的基础上,建立岭南骨伤方药数据库,此次研究利用 LDA 主题模型算法与关联规则算法对其进行建模分析,以更好地总结和归纳岭南骨伤名家用药规律及特色。

潜在狄利克雷分布 (Latent Dirichlet Allocation, LDA) 算法最早用于文本分类,是一种基于无监督学习的潜在语义生成模型^[2]。模型假设待分类文本集存在 T 个潜在主题,每个文本都以一定概率对应一个或多个主题;由于文本的基本组成单元为词,通过识别不同的词分布即可确定文本主题;对于一个特定主题,与之对应的出现概率最高的 N 个词即为其主题词。同理,若给定一组中药方剂,将一首方看做一个文本,则每首方剂会以一定的概率对应一类或多类功效主治,可以将其视为组方主题;方剂的基本组成单元为单味药,通过分析单味药的分布即可确定这组中药方剂所包含的潜在组方主题,亦即功效主治;每类组方主题所对应的出现概率较高的中药,即为达到该功效主治的方剂最核心的用药。此次研究将与特定组方主题最相关的前 N 味药定义为对应功效主治的核心组方,简称核心方。关联规则 (association rule) 最早由 Agrawal 等^[3]于 1993 年首先提出,用于从大量数据中发现满足一定条件的项集之间隐藏的关联。关联规则分析首先根据支持度从数据集中提取出所有的频繁项集,然后

产生关联规则并根据置信度保留符合条件的规则^[4]。近年来,关联规则被引入中医药领域,在中医药知识发现与规律总结方面发挥了重要的作用。

1 资料与方法

1.1 数据来源与规范化 收集整理目前可见岭南骨伤著作文献《何竹林正骨医粹》,管炎威《伤科讲义》,《外科讲义》,《骨折与脱位的治疗》,《岭南骨伤科名家何竹林》,《广东广汉中医专门学校伤科讲义》,《陈渭良骨伤科临证精要》,《西关正骨·李氏临床经验》,《岭南正骨精要》等骨伤著述及中国知网,万方数据知识服务平台、维普中文期刊服务平台等在线数据库中检索到的相关论文中所涉及的岭南骨伤方药,共得有效方剂 802 首。建立岭南骨伤方药数据库,录入包括编号、方名、所属流派、药物组成、药物用量、制作用法、主治功效等在内的方剂信息。以《中国药典》^[5]以及《中药大辞典》^[6]中记载的中药正名为标准规范方剂中药物名称。

1.2 数据分析软件与算法 “中医药古籍文献知识服务平台”是由广东省中医院大数据团队研发,该平台是基于本体语言 OWL 的知识表示及推理算法构建,利用现有大规模古籍文献全文数据库与知识发现技术,根据专科、专病、流派等知识特点形成一套标准化的中医药知识元构建工具和方法,系统整理中医药领域知识并构建知识地图,通过对数据分析建模进而对中医理法方药进行科学的阐释的知识服务平台^[7]。

描述性分析是中医药数据挖掘基本上都会使用的方法,随着近年来数据挖掘技术的发展,该方法多作为其他算法的基础,很少单独使用^[8]。近年来,专家学者们把 LDA 主题模型引入中医药数据挖掘领域并取得了一定的成果。如张小平^[9]利用 LDA 模型来分析中医诊疗数据,客观、系统的分析总结中医辨证论治规律;霍蕊莉等^[10]利用主题模型算法研究了中医治疗消渴病痹痿证的症状与用药的关系;王建华等^[11]收集 982 例心脑血管合病患者的四诊信息

并建立数据库,采用频数统计及主题模型等数据挖掘方法分析了心脑合病的证候特征。关联规则被引入中医药数据挖掘领域后在中医方剂研究领域、医案研究领域、证候研究等领域有着广泛的应用。如刘根等^[12]利用中医传承辅助平台数据挖掘软件中关联规则算法对治疗老年性痴呆方剂核心药物组合的筛选研究。刘广等^[13]对中医胃炎原始病案数据预处理和数据结构规范化,采用关联规则挖掘方法进行中医胃炎疾病的医案数据库挖掘,得出了症状之间、处方之间、症状与处方之间的关联。艾军等^[14]480 例小儿肺炎病例为数据源,进行症状关联规则运算,分析了小儿肺炎中医证候与病机的关系。

1.3 资料处理与数据分析 通过 Excel 建立岭南骨伤方药数据库,并将数据导入“中医药古籍文献知识服务平台”数据挖掘软件中,利用软件中描述性分析,LDA 主题模型分析、关联规则分析等数据分析建模功能对岭南骨伤数据进行用药规律挖掘。

2 结果

2.1 描述性分析结果

2.1.1 药物频数、频率分析 纳入分析的 802 首岭南骨伤方剂中共涉及中药 811 味,药物出现频数总计 6 959 次,使用频率最高的是当归,共使用 620 次,频率为 77.31%。其次是甘草,其使用频数为 435 次,使用频率为 54.24%;川芎位居第 3 位,其使用频数为 252 次,使用频率为 31.42%;白芍位居第 4 位,其使用频数为 234 次,使用频率为 29.18%;赤芍共使用 233 次,使用频率为 29.05%。其中,使用频数排列在前 8 位的中药分别是当归、甘草、川芎、白芍、赤芍、红花、茯苓、桃仁,以上 8 味药使用频数均超过 200 次。使用频数在 80 次以上、使用频率超过 10% 的药物见表 1。

2.1.2 药类使用频数、频率分析 根据《中国药典》与《中药大辞典》中药物分类标准,将 811 味中药进行分类。具体药物类别、频数、频率分布见表 2。由表 2 可见,岭南骨伤方剂中药物类别主要为活血化瘀药、补虚药、理气药、清热药、祛风湿药、温里药,累计频率可达 43.06%,其中活血化瘀药又以使用 1 013 次,占总使用次数的 14.56% 高居第 1 位;补虚药使用频数仅次于活血化瘀药为 733 次,占总是用频率的 10.53% 位居第 2 位;理气药使用频数为 366 次,占总使用频率的 5.26%,位居第 3 位;清热药使用频数为 358 次,占总使用频率的 5.14%,位居第 4 位;祛风湿药使用 284 次,占总使用次数的 4.08%,位居第 5 位;温里药使用频数为 143 次,占

表 1 使用频数超过 80 次的药物

Table 1 Herbs used over 80 times in prescriptions

药物	频数/次	频率/%	药物	频数/次	频率/%
当归	620	77.31	三七	137	17.08
甘草	435	54.24	黄连	132	16.46
川芎	252	31.42	黄芩	131	16.33
白芍	234	29.18	防风	109	13.59
赤芍	233	29.05	牡丹皮	106	13.22
红花	223	27.81	肉桂	105	13.09
茯苓	217	27.06	牛膝	100	12.47
桃仁	212	26.43	柴胡	98	12.22
人参	195	24.31	黄芪	96	11.97
没药	188	23.44	骨碎补	94	11.72
地黄	186	23.19	木香	93	11.60
乳香	183	22.82	桔梗	90	11.22
白术	173	21.57	栀子	86	10.72
熟地黄	171	21.32	黄柏	85	10.60
陈皮	163	20.32	续断	84	10.47
大黄	163	20.32	枳壳	84	10.47
白芷	150	18.70	羌活	81	10.10
半夏	150	18.70	香附	81	10.10

表 2 每类药物使用频数和频率分布

Table 2 Use frequency and frequency distribution for each type of herb

药物类别	频数/次	频率/%	累积频率/%
活血化瘀药	1 013	14.56	14.56
补虚药	733	10.53	20.09
理气药	366	5.26	30.35
清热药	358	5.14	35.49
祛风湿药	284	4.08	39.57
温里药	143	3.49	43.06

总使用次数的 3.49%,位居第 6 位。

2.1.3 药物性味归经分析 对上述 811 味药进行性味归经分析,药物性味归经情况见图 1。岭南骨伤所使用药物以温性最多,占 28.3%,其次是平性所占比例为 22.6%,寒性药物所占比例为 22.5%,位居第 3。药味以甘、辛、苦为主,所占比例分别为 26.8%,25.1%,23.3%。岭南骨伤所使用药物中归肝经的药物占 21.9%,归肺经的药物占 13.8%,归脾经的药物占 13.4%,归胃经的药物占 12.4%。

2.2 LDA 模型核心方建模分析结果 运用 LDA 主题模型^[15]对岭南骨伤数据进行主题分析建模。经

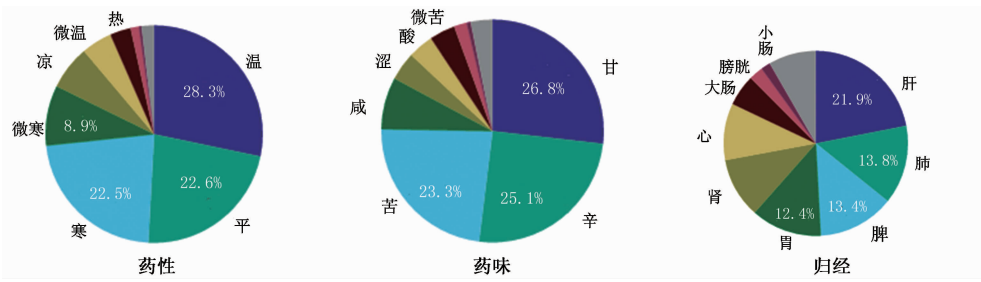


图 1 药性味归经分布

Fig.1 Properties, flavours and meridian distributions of herbs

建模测试并结合相关中医理论,最终设置主题数为 5,主题词数为 10,迭代次数为 1 000 次,一维数组为 10%,二维数组为 10% 进行主题分析共得 5 个核心方。5 个核心方中核心药物为主题 1,当归、甘草、地黄、红花、桃仁、赤芍、大黄、栀子、黄芩、黄连;主题 2,当归、赤芍、三七、骨碎补、牛膝、白芷、乳香、没药、红花、续断;主题 3,甘草、当归、茯苓、人参、白术、川芎、熟地黄、陈皮、白芍、半夏;主题 4,乳香、没药、血竭、麝香、朱砂、三七、冰片、大黄、儿茶、龙骨;主题 5,三七、白及、白芷、艾叶、藕节、桑白皮、血余、桑白皮、百草霜、海螵蛸、香附。

2.3 关联规则分析结果 运用软件中分析建模功能中的关联规则对上述药物进行分析,设置参数:置信度 $\geq 60\%$,支持度 $\geq 5\%$ [置信度表示前项(LHS)中后项(RHS)出现概率的度量值,是对关联规则准确性的测量,支持度表示 LHS 与 RHS 同时出现概率的度量值,是对关联规则普遍性的测量^[16]]。共得出 21 条药对与药组关联规则,详细内容见表 3。由表 3 可见,川芎、白芍->当归置信度最高,其置信度为 93.38%;川芎->当归支持度最高,其支持度为 14.27%;当归、乳香->没药的提升度最高,其提升度为 7.25%。关联规则可视化内容见图 2。核心药物分布见图 3,由图 3 可知当归是所有关联规则中的最核心药物,纳入核心药物的有当归、川芎、白芍、赤芍、熟地黄、红花、桃仁、地黄、甘草、牛膝、白芷、茯苓、乳香、没药、肉桂、黄芪,基本上是桃红四物汤的加减方。

3 讨论

骨伤科是研究防治人体皮肉、筋骨、气血、脏腑经络损伤与疾患的一门科学。在古代属于“折疡”、“金镞”等的范畴。历史上有“金疡”、“接骨”、“正骨”、“伤科”等不同的称谓。骨伤疾患由于外感六淫、邪毒感染、外力损伤,体质、先天禀赋不足、七情内伤加重气血内耗等原因,导致气血瘀滞,经络阻塞,出现出血、疼痛、功能障碍等症状。血瘀既是伤

表 3 岭南骨伤方剂数据关联规则分析

Table 3 Association rules of herb combinations for Lingnan bone injury

关联规则	置信度/%	支持度/%	提升度/%
川芎、白芍->当归	93.48	5.58	2.45
川芎->当归	85.94	14.27	2.25
川芎、甘草->当归	85.19	5.97	2.23
乳香->没药	84.27	9.73	7.22
乳香、当归->没药	83.67	5.32	7.17
当归、没药->乳香	83.67	5.32	7.25
没药->乳香	83.34	9.73	7.22
熟地黄->当归	81.25	8.43	2.13
黄芪->当归	79.37	6.49	1.08
牛膝->当归	72.73	5.19	1.91
白芍->当归	70.54	10.25	1.84
桃仁、红花->当归	68.75	5.71	1.80
当归、桃仁->红花	68.75	5.71	4.49
赤芍->当归	67.57	9.73	1.77
地黄->当归	64.52	7.78	1.69
红花->当归	64.41	9.86	1.69
肉桂->当归	64.18	5.58	1.68
茯苓->甘草	63.16	8.43	2.26
桃仁->红花	62.75	8.30	4.10
桃仁->当归	62.75	8.30	1.65

损之后的病理产物,又是引起骨伤患者功能障碍的致病因素,因而历代骨伤医家均重视活血化瘀的治疗。岭南骨伤是全国中医骨伤的一个重要组成部分,也是岭南医学的重要组成部分,其用药规律既有骨伤科共性又有岭南地域特色。本研究结果表明,岭南骨科名家的用药基本遵从骨伤科用药的常规,重视活血化瘀的治疗,用药偏温性,归肝经最多,桃红四物汤是其核心处方,当归是其最核心的用药。

用药呈现比较明显的骨伤早、中、晚 3 期用药规律,在活血化瘀的基础上,早期或兼清热解毒,或强

不散,局部肿痛较明显,根据受伤部位、受伤程度的不同,出现不同的临床表现,开放性骨折或伤筋可能会出现以出血为主,或创口感染出现毒邪炽盛,或蒙蔽清窍出现神志不清。闭合性骨折或伤筋以及脱位则可能更多表现为肿痛。但不管怎样,中医认为,“瘀不去,则骨不能生”、“瘀去新骨生”,所以治疗以活血化瘀、行气消散为主,不可过早滋补,否则加重瘀血积滞。

中期骨伤部瘀肿大部分吸收,疼痛减轻,一般的软组织损伤已修复,骨折断也初步稳定,纳食正常,原始骨痂也开始逐步形成。中医认为肝主筋,肾主骨、生髓,肝肾同源,补肝肾可以达到壮筋强骨的作用。所以治疗以祛瘀生新、接骨续筋、舒筋活络为主,在活血化瘀的基础上加用补肝肾的药物。

晚期骨伤部瘀肿基本吸收,常常已有骨痂生长、骨折断端也比较稳定,而日久不愈,常常因为气血不足,“气为血之帅,血为气之母”,气血濡养充足,筋骨功能就会强壮,正气亏虚,则难以托邪外出,迁延不愈。所以治疗以补益气血、濡养筋脉、滋补肝肾为主,强筋健骨,同时疾患日久,损伤脾胃,在补益的基础上,常常要加用健脾益气之品,以促进吸收^[17]。

岭南地处大庾岭,骑田岭,都庞岭,萌渚岭,越城岭五岭以南,濒临海洋,大部分属亚热带湿润季风气候,以高温多雨为主要气候特点。故病人体质亦多湿热,骨折伤损之后,湿热与瘀血互结,致瘀湿热同时为患。故岭南骨伤疾患的治疗早期在活血化瘀基础上,多用大黄、栀子、黄连、黄芩、藕节等清热利湿之品进行加减。岭南地区地卑土薄,人与天地相应,南方人身材大多较北方人矮小,体质上多脾胃虚弱,脾虚易湿热内生,湿热的气候又阻碍脾胃的运化,加重了脾虚,这使得岭南骨科名家在骨伤疾患治疗后期更重视健脾益气,行气化湿,并且注意少用滋腻碍脾等补益品。

在前期的文献整理研究中,笔者发现,部分岭南骨科名家,如何竹林,喜欢使用岭南地方草药,如毛藿香、透骨消、徐长卿、千斤拔、过岗龙等;蔡荣对岭南藤类中药材亦青睐有加,其献方之一《骨科外洗一方》,主药即含宽筋藤;蔡荣的传承人陈基长对于病势位于下肢为重者,多喜配用宽筋藤煎汤内服治之^[18]。在本次数据建模分析中,这些岭南地方草药并没有进入核心药物范畴,这可能是因为此次收集是岭南骨伤医家总体用药情况的数据,反映的是

这些骨伤医家用药的共性规律所致。

[参考文献]

- [1] 刘小斌,郑洪. 岭南医学史(中)[M]. 广州:广东科技出版社,2012:364.
- [2] Griffiths T L, Steyvers M. Finding scientific topics[J]. Pro Nat Acad Sci, 2004, 101 (S1):5228-5235.
- [3] Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases[J]. Acm Sigmod Record, 1999, 22(2):207-216.
- [4] 熊平. 数据挖掘算法与Clementine实践[M]. 北京:清华大学出版社,2011:76-77.
- [5] 国家药典委员会. 中华人民共和国药典[M]. 北京:中国医药科技出版社,2015:1-1268.
- [6] 南京中医药大学. 中药大辞典[M]. 上海:上海科学技术出版社,2006:1-3875.
- [7] 郑玮琳. 痛经文献分析及益气化瘀法治疗子宫内膜异位症痛经临床研究[D]. 广州:广州中医药大学,2016.
- [8] 李强. 基于方剂数据挖掘的痹证证治规律研究[D]. 北京:中国中医科学院,2014.
- [9] 张小平. 主题模型及其在中医临床诊疗中的应用研究[D]. 北京:北京交通大学,2011.
- [10] 霍蕊莉,刘保延,何丽云,等. 基于主题模型的消渴病痹痿症药关系研究[J]. 北京中医药, 2014, 33(3):163-166.
- [11] 王建华,张哲,肖蕾,等. 基于主题模型982例心脑血管合病患者证候特征研究[J]. 世界科学技术—中医药现代化, 2015, 17(12):2527-2531.
- [12] 刘根,贺文彬,赵子强,等. 基于中医传承辅助平台对老年性痴呆防治方剂核心药物组合的筛选研究[J]. 中国实验方剂学杂志, 2016, 22(7):223-228.
- [13] 刘广,吴磊. 基于关联规则的中医胃炎诊疗与处方间的挖掘研究[J]. 江苏中医药, 2011, 43(11):72-74.
- [14] 艾军,汪受传,戴铭,等. 基于症状关联规则的小儿肺炎中医证候病机学研究[J]. 北京中医药大学学报, 2016, 39(2):140-143.
- [15] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. JMLR, 2003, 3:993-1022.
- [16] 薛薇,陈欢歌. SPSS Modeler数据挖掘方法及应用[M]. 北京:电子工业出版社,2014:209.
- [17] 张朝驹,何川,刘伟,等. 徐昌伟分期论治骨伤科疾患经验[J]. 湖北中医杂志, 2015, 37(5):31-32.
- [18] 徐险峰,黄枫,郑晓辉. 岭南骨伤科应用岭南藤类中药材举隅[J]. 河南中医, 2014, 34(7):1410-1411.

[责任编辑 邹晓翠]