

·民族药·

藏族药方剂信息库构建及知识发现的探索与实践

文成当智¹, 东改措^{1*}, 才让南加¹, 贡保东知², 格日多杰³, 泽翁拥忠⁴, 拉巴次仁⁵

(1. 青海民族大学 药学院, 青海 西宁 810007; 2. 甘肃中医药大学 藏医学院, 甘肃 甘南 747000;

3. 青海省黄南州藏医院, 青海 黄南 811300; 4. 成都中医药大学 民族医药学院, 四川 成都 611137;

5. 西藏藏医药大学, 西藏 拉萨 850000)

[摘要] 探索藏族药(以下简称藏药)方剂信息的底层框架和数据特征。基于《四部医典》等11部藏医药经典文献,收集藏药方剂信息,应用最优归类法归纳总结藏药方剂信息构成框架,梳理数据收集、规范、翻译、分析等过程中存在的关键问题和解决办法。共收集到方剂11 316首,组方药物涉及词条有139 011条,涉及效用信息共63 567条。整理发现藏药方剂信息可归纳为“编号-方源-名称-组成-效用-附录-备注”等七位一体的主体框架和18项展开层,该框架能载有方剂相关的传承、炮制、基原、剂量、语义等所有信息。基于框架提出了“历史时间轴”的方剂传承源流挖掘方法、“一体五层”的组方药物规范方法、“连-分-连”的效用信息构建方法以及适合藏药方剂知识发现研究的先进算法。藏药方剂在藏医“三因学”“五源学”,以及“味性化味”理论指导下具有明显特色和优势,该研究基于藏药方剂自身特色,提出了多层次多属性的底层数据架构,为藏药方剂信息库构建及知识发现提供了新方法和新模式,提高藏药方剂信息与各级标准一致性、互通性,以期实现“古今相连-正本清源-数据共享”的藏药方剂信息化、现代化研究路径。

[关键词] 藏族药; 方剂数据库; 知识发现; 配伍规律; 方法

Exploration and practice on construction of Tibetan medicine prescription information database and knowledge discovery

WENCHENG Dang-zhi¹, DONG Gai-cuo^{1*}, CAIRANG Nan-jia¹, GONGBAO Dong-zhi², GERI Duo-jie³,
ZEWENG Yong-zhong⁴, LABA Ci-ren⁵(1. College of Pharmacy, Qinghai Minzu University, Xining 810007, China; 2. College of Tibetan Medicine, Gansu University of Chinese Medicine, Gannan 747000, China; 3. Huangnan Tibetan Medical Hospital, Huangnan 811300, China;
4. Ethnic Medicine School, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China;
5. University of Tibetan Medicine, Lhasa 850000, China)

[Abstract] This study aimed to explore the underlying framework and data characteristics of Tibetan prescription information. The information on Tibetan medicine prescriptions was collected based on 11 Tibetan medicine classics, such as *Four Medical Canons (Si Bu Yi Dian)*. The optimal classification method was used to summarize the information structure of Tibetan medicine prescriptions and sort out the key problems and solutions in data collection, standardization, translation, and analysis. A total of 11 316 prescriptions were collected, involving 139 011 entries and 63 567 pieces of efficacy information of drugs in prescriptions. The information on Tibetan medicine prescriptions could be summarized into a "seven-in-one" framework of "serial number-source-name-composition-efficacy-appendix-remarks" and 18 expansion layers, which contained all information related to the inheritance, processing, origin, dosage, semantics, etc. of prescriptions. Based on the framework, this study proposed a "historical timeline" method for mining the origin of prescription inheritance, a "one body and five layers" method for formulating prescription drug specifications, a "link-split-link"

[收稿日期] 2022-11-05

[基金项目] 青海省科技计划项目(2022-ZJ-958Q)

[通信作者] *东改措, 硕士, 主治医师, 主要从事藏族医消化系统疾病诊治及藏族药信息化研究, E-mail: 1014976914@qq.com

[作者简介] 文成当智, 副教授, E-mail: wenchengdangzhi@aliyun.com

method for constructing efficacy information, and an advanced algorithm suitable for the research of Tibetan prescription knowledge discovery. Tibetan medicine prescriptions have obvious characteristics and advantages under the guidance of the theories of "three factors", "five sources", and "Ro-nus-zhu-rjes" of Tibetan medicine. Based on the characteristics of Tibetan medicine prescriptions, this study proposed a multi-level and multi-attribute underlying data architecture, providing new methods and models for the construction of Tibetan medicine prescription information database and knowledge discovery and improving the consistency and interoperability of Tibetan medicine prescription information with standards at all levels, which is expected to realize the "ancient and modern connection-cleaning up the source-data sharing", so as to promote the informatization and modernization research path of Tibetan medicine prescriptions.

[Key words] Tibetan medicine; prescription database; knowledge discovery; compatibility rule; method

DOI:10.19540/j.cnki.cjcm.20221126.601

随着党和国家对民族医药事业的大力支持,藏族医药(以下简称藏医药)研究在临床、教学、科研等诸多领域都取得了丰硕的成果。近年来信息学作为网络、自动化、计算机等诸多领域的基础,逐步影响着中医药的方方面面^[1],如2008年提出了“中药分子复方”^[2],2011年提出了“计算中药学”^[3],还有“网络方剂学”^[4],中药方剂网络与中药化学空间^[5]等。同样基于信息技术的藏医药研究逐年增多,在2012年有学者提出了数据挖掘技术在藏医药中的应用,本研究也先后提出了藏医药信息学^[6-7],藏族医(以下简称藏医)《四部医典》方剂数据库^[8],藏族药(以下简称藏药)“味性化味”药性理论的科学内涵、矢量结构模型等^[9-11],可见信息技术为民族医药的传承和创新带来了新的机遇和挑战。在藏医药信息化研究中,数据库作为数据结构组织、存储和管理的仓库,在知识发现研究中具有重要地位,是数据信息的直接来源,然而目前尚未发现共享式、开放式藏药方剂数据库的报道,因此,藏药方剂信息库的构建是目前待解决的科学问题,也是信息时代留给科研工作者的必要任务。

藏药方剂作为藏医药传承和发展核心对象,承载着藏医药“病-药”对接的核心理论,也蕴含着历代藏医药大师的毕生经验。藏药方剂在宏观层面涵盖了藏医药从外界自然生物到药物治疗疾病的天人合一整体思想,具体可以体现藏医药药材结构、疾病结构、药性理论、配伍方法、炮制方法等知识,具有显明特色和优势^[12],而且方剂也是藏医临床实践的主要表现形式,是目前藏医药研究聚焦点和难点。收载藏药方剂的文献丰富,方剂涉及的药材种类繁多,据统计《晶珠本草》^[13]中有2 294种药物^[14-15],《藏药晶镜本草》记载药物3 861种^[16],《中国民族药辞典》记载藏药有3 105种^[17],可见方剂信息及组方药物在藏医药信息中占据重要地位。

近年来,基于藏药方剂信息的研究,藏医药在心脑血管疾病^[18]、消化系统疾病^[19-20]、风湿类疾病^[21]、热病学^[22]中取得一定的研究成果,但根据目前研究现状来看,依然存在诸多问题,如无明确的藏药方剂底层信息结构,研究者根据需求直接截取,丢失方剂潜在的重要信息;藏药经典方剂筛选困难,尚未发现共享式藏药方剂信息库,诸多藏医药方剂研

究均立身于已翻译成中文或已编辑成结构化数据的部分现代翻译文献,数据具有一定的局限性,无法原味再现藏医古籍的学术思想;组方药物规范和统一机械化,因藏医文献多数以诗句形式记载,考虑韵味等因素,藏药“一药多名”和“一名多药”的特点突出,尚无“名-药”对接词库,统一过程中参杂很多主观因素,且统一过程过于机械化,原文记载的药物直接归纳统一,如“热制寒水石”统一为“寒水石”、“生等膏”统一为“生等”、毛诃子肉统一为毛诃子^[20]、黑沉香和白沉香统一为沉香等,缺乏原文到规范数据的具体步骤,宏观而高度概括的统一会造成文献信息与研究结果不对等或断层现象;方剂信息处理过程繁琐,无简便而高效的信息处理路径,均为学者自行拟定,方法的可重复性、可复制性较差;效用信息处理不合理,其语义涉及整个藏医学体系,目前处理方法多数对原文语义进行统一归纳,如“赤巴热”和“热赤”统一为“赤巴热症”,面对数据量多且参杂主观因素,大幅弱化时效性。总之,藏药方剂研究尚无严谨而规范的共享式数据处理平台、组方基原模糊等传统与现代脱节,且信息处理存在“主观+大量+复杂”的现象,导致研究时效性差、延展性差、可复制性低等问题。

因此,本研究基于藏药方剂信息,研究数据结构、逻辑演绎、表现形式,提出“七位一体”的藏药方剂底层数据架构,“一体五层”的组方药物统一规范路径、“连-分-连”的效用信息构建方法等,以期解决以上实际问题,实现藏药方剂数据库与国家、省部级标准相关联,以规范、高效为两翼,促进藏药新药处方筛选、加快藏医药信息化、现代化研究。

1 研究方法

1.1 文献筛选 方剂是藏医药文献中的重要内容之一,藏医药具有丰富的古籍文献,目前国内先后抢救挖掘藏医药古籍文献2 000多部(函),整理出版《藏医药大典》《中国藏医药影印古籍珍本》等。据不完全统计,藏医药古籍文献数目不下3 000种,面对浩如烟海的文献,首先需要筛选文献,本研究纳入《四部医典》、《准赛三十部教海》^[23]、《零星秘决集》^[24]、《千万舍利子》^[25]、《秀多医学集》^[26]、《孕玛冉纳医嘱选集》^[27]、《藏医临床札记》^[28]等11部文献,共收载11 316首方剂,见表1。

表 1 藏医文献收载方剂数量
Table 1 Quantity of prescriptions in Tibetan medical literature

No.	书名	作者	年代 (世纪)	方剂 数量/首
1	《度母本草》	先达杂巴	8	583
2	《妙音本草》	比如扎纳	8	332
3	《四部医典》	宇托云丹贡布	8	2 695
4	《准赛三十部教诲》	香思杰巴	11	367
5	《零星秘诀集》	贡曼贡却德勒	15	626
6	《千万舍利子》	苏喀年尼多杰	15	2 320
7	《秀多医学集》	秀多贡布	15	1 872
8	《杂玛冉纳医嘱选集》	杂玛冉纳	18	373
9	《零星方剂汇集》	杂玛冉纳等	18	743
10	《医术精选入迷》	久美丹增加措	18	292
11	《藏医临床札记》	工竹云丹嘉措	19	1 113

注:根据方剂界定不一,可能存在不同学者对同一本文献方剂数量统计不一,此数据仅表示文献中明确标注为“方剂”的方剂数量。

1.2 数据收集 根据以上文献,收集方剂数据的同时,逐步挖掘藏药方剂的数据结构,确立主要字段,并以嵌套字典模式储存,分析字段之间的关联和每一层级之间的属性关系,推导和整理藏药方剂数据结构及其之间的逻辑推理规律。

1.3 规范化处理 基原和标准药材名称以 1998 年版的《中华人民共和国卫生部药品标准藏药(第一册)》^[29]和《藏药晶镜本草》等文献为依据,以字典模式建立藏药方剂中出现的组方药物原词汇与规范词汇关系,如“诃子”:“尖嘴诃子”、“诃子去核”、“金光诃子”、“无畏诃子”等,并以字典数据为

依据,对方剂原数据逐层进行规范化处理,构建藏药方剂数据库。

2 研究结果

根据 11 部文献收集和整理方剂信息,共得到 11 316 首方剂,其中涉及组方词条有 139 011 条,效用词条有 63 567 条,整体数据结构存在多维度、多属性、多层次的特点。

2.1 “七位一体”的藏药方剂数据结构 基于 11 316 首藏药方剂,挖掘其中的关键属性字段,应用最优归类法进行归类,逐步形成属性数据架构,发现以“编号-方源-名称-组成-效用-附录-备注”等七位一体主体框架,以及 18 项展开项的结构,能够有效收纳藏药方剂相关的所有属性信息,本研究将其定为藏药方剂底层数据框架,见图 1。其中编号是方剂编号,具有唯一性,是计算机识别方剂的主要字段。方源记录方剂传承脉络,后文将重点描述。名称为方剂名称,又分为古籍明确记载的方名和未记载而作者自行命名 2 种,其中古籍记载中会出现一方多名现象,若二十一味寒水石散与木布宇杰散,故记录时需全部记录,并标注名称为原文记载,命名主要遵循 3 种模式:一是主药加药味数量命名,如七十味珍珠丸;二是功效命名,如血骚普清散;三是殊胜命名,如如意珍宝丸等,命名后要标注自行命名。组成为构成方剂的组方药物和卡擦药物,其中卡擦是指用于调节药性的加减药物。效用是指方剂功效和主治。附录包含方剂的其他属性,如剂型、用法、用时、药引、用量、辅剂、储藏等。备注为以上属性以外的信息,如方剂联用法、特殊炮制法、作者用药经验等。

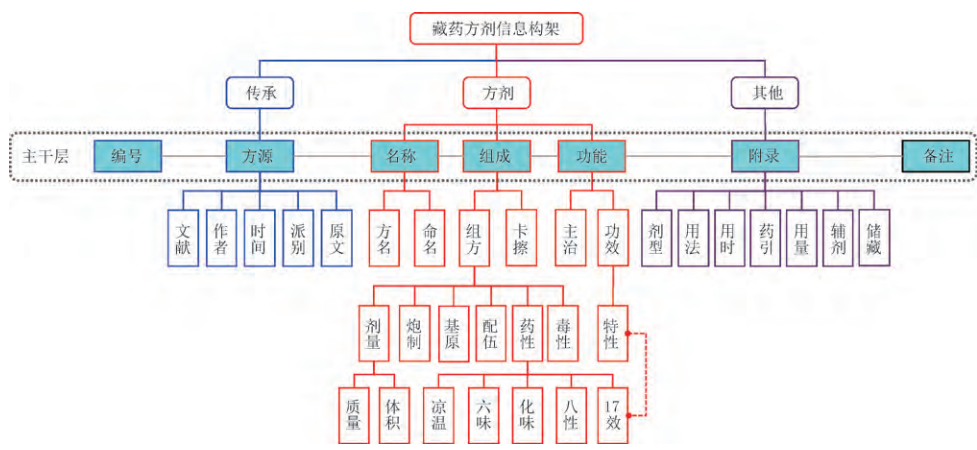


图 1 藏药方剂数据结构框架示意图
Fig. 1 Schematic diagram of data structure framework of Tibetan medicine prescriptions

在展开项中子层属性较为重要的是组方,有剂量、炮制、基原、配伍、药性、毒性等,其中配伍是藏药方剂配伍结构,描述具体组方药物在整体方剂中扮演的角色,与中药君臣佐使具有一定的相似之处。具体为君(ཏྲཱ་ཤི་མོ་)、臣(ལྷན་པོ་)、佐(འབྲས་པོ་)、妃(འབྲས་པོ་)、子(ཏྲཱ་ཤི་མོ་)、使(ཏྲཱ་ཤི་མོ་)、将(ཏྲཱ་ཤི་མོ་)、兵

(དམག་པོ་)、马(ཏྲཱ་ཤི་མོ་)、盾(ཏྲཱ་ཤི་མོ་)、矛(ཏྲཱ་ཤི་མོ་)、匠(ཏྲཱ་ཤི་མོ་)等分类,如在《四部医典》记载^[30]:消热草药剂(ཏྲཱ་ཤི་མོ་ཏྲཱ་ཤི་མོ་)中肉叶金腰为君,尼泊尔黄芩为妃,高山辣根菜为子,甘青青兰为臣,细果角茴香为佐,对治瘟疫以细果角茴香为将,同样混热以匙叶翼首草为将、血液热以圆穗兔耳草为将、赤巴热以印

度獐牙菜为将,篦齿虎耳草、川西合耳菊、乌奴龙胆等 13 味药为兵,泉水为马。《千万舍利子》记载:苏喀泻热方(བྱུང་མཁའ་གཟུང་ཚད་པའི་འཕྱུལ་བཤམས།)中冰片为君,圆穗兔耳草为妃,斑花黄堇为臣,肉叶金腰为佐,船盔乌头为将,6 味良药即肉豆蔻、豆蔻、草果、竺黄、番红花、丁香为兵,泉水为匠。泻邪金刚宝石方(གདོན་བཤམས་རྩི་རྩི་ལ་ལས།)中斑蝥、黑斑蝥、硃砂等 8 味药物共同为君,白狼毒、巴豆、大果大戟膏等 8 味药物共同为臣,6 味良药为妃,3 味黄水药即琥珀、决明子、黄葵子为子,肉托果、沉香、宽叶羌活等 13 味药物共同为兵,狗血、马血、猪血等 5 味药物为盾,佐太、阿魏、麝香等 7 味药共同为马,秃鹫粪等 24 味药物为矛。在甘露月滴方(བདུད་ཅིན་པའི་ཐིགས་པ།)中以斑花黄堇为君,船盔乌头、细果角茴香、印度獐牙菜等 3 味药物为妃,粗牡荊为子,圆穗兔耳草和匙叶翼首草为臣,打箭菊、江孜乌头苗为使,瘤果梭子芹、臭蒿为庶,臭骚草、多叶唐松草为兵,拟楼斗菜、白苞筋骨草为匠等,组方配伍结构具有详细而多元化的特点。药性是指藏医“味性化味”理论指导,具有多维多层的复杂特点,其中组方药物十七效直接和间接地决定着方剂整体功效。

用法是指方剂成方的用法,随着内服、外涂等不同,药量和炮制也发生变化,因此,对藏药剂量和炮制研究时,随用法分类而进行,不能一概而论,主要包括服、浴、涂、敷、熏、撒、滴、贴、鼻、灌、漱、含、戴等,其中“鼻”是通过鼻腔给药的治疗方法治疗锁骨以上的疾病,《四部医典》中专设一章节描述。用时为用药时间,如内服药物而言有 10 种用时,未食、食前、食中、食后、食药轮勺、食消即食物消化后、随时少量、食并、食压即食前加食后、夜间服。剂型主要有汤剂、散剂、丸剂、膏剂、药酥剂、药酒、灰剂等。药引主要有白糖、蜂蜜、红糖等 3 种。用量常以同比体积大小比喻来定,主要有豌豆、芝麻、鼠粪、羊粪、鸽蛋、拇指等,如用于瘟疫的苏玛达日方(རིམ་གྱུང་ཐོ་མ་དྲི་ཐེར་ཐུང་།)每次药量如芝麻,具体成药用量需对应形状称量来定,而组方药物用量为成药用量与组方剂量占比的积。辅剂一般指丸剂定形时所需的液体介质,主要有童子尿、生等汤、青稞酒、酥油、黄牛尿、沙棘膏、石榴汤等。

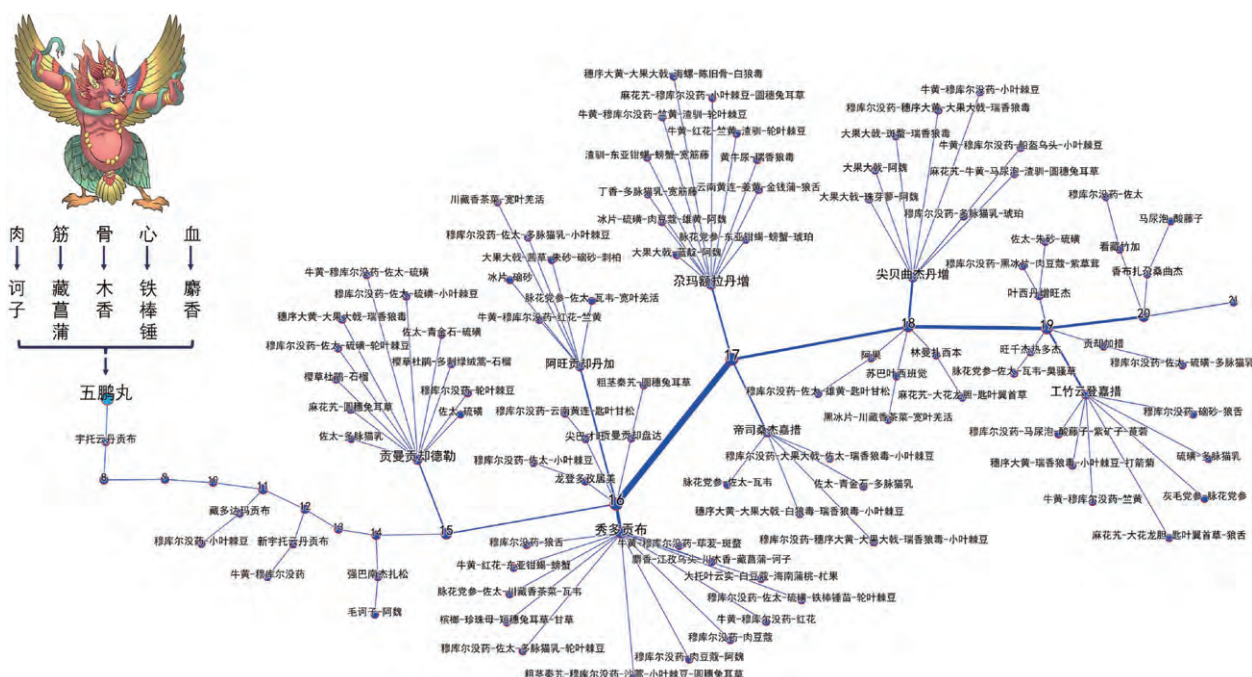
该框架整体以守正创新为宗旨,未处理数据以原文为核心、规范数据以藏药标准术语和药物名称为核心、翻译数据以药典和青海省藏药材为标准。在不丢失方剂信息、不扭曲原文语义的前提下有效衔接国家和省部级标准,横向实现“数据联用-标准共享”的基层平台、纵向实现“古今相连-正本清源”的研究模式,为藏医药数据挖掘、知识发现、基原考证等提供更加完整的底层数据结构,推动藏医药信息化研究进程。

2.2 “历史时间轴”的方剂传承源流 在“七位一体”中方源对应的是方剂的时空传承源流信息,主要包括作者、文献、时间、派别、原文,其中派别为作者的传承派系,主要有藏医南北两派,也称“强派”和“苏派”,最初有九仙神医(བོད་མཁན་མཁས་པ་མི་དགུ།)等派别,随时间又出现了宇托(ཡུ་ཐོ་ཐོག་ལུགས།)、

比吉(བི་རི་ལུགས།)、昌迪(ཅང་ཏི་ལུགས།)、唐东(ཐང་ལྷོང་ལུགས།)、贡曼(གོང་མཁན་ལུགས།)、直贡(འབྲི་གུང་ལུགས།)、司徒(སི་ཏུ་ལུགས།)、仲曼(འཛོང་མཁན་ལུགས།)等派系,因派别不同,同一种药物的基原、剂量、炮制方法都有区别,因此,派别也是方剂主要信息定位考证的主要依据。时间表示文献成书时间,若时间不清就用世纪来记载,同一方剂随着时间变迁会衍生诸多方剂,结合时间、作者、文献能够有效衔接衍生方剂之间的传承源流,如五鹏丸始载于《四部医典》,相传鹏鸟的肉、筋、骨、心、血 5 部分化成诃子、藏菖蒲、木香、铁棒锤、麝香等 5 种组方而成,不同于一般方剂,五鹏丸的文化源流和传承脉络较为浓重,历代藏医大师都有应用,上述方剂信息中检索含有五鹏丸的方剂共有 798 首,均为该方衍生方剂,选取 $7 \leq$ 组方数 ≤ 10 的方剂共有 73 首,见图 2。另外,随着时间不同,同一剂量单位实际指数不一,因此考证组方药物本原剂量时时间具有重要参考意义。原文为记录方剂整体的古籍原始文本,以直接截取或手动输入为主,不做任何分词和规范等处理,是文献到结构化数据衔接的桥梁,也是后期规范过程中遇到错误,或本原组方和本原剂量的考证时可追溯的直接依据,因藏医药古籍文献多数以七句或九句诗词形式记载,若不保留原文而主观分词难免出现错误,所以原文对收集藏医古籍方剂尤为重要。

2.3 “一体五层”的组方药物规范方法 组方药物是构成方剂的重要组成部分,若规范统一词表,将组方药物直接规范化处理,将会丢失很多重要信息,如用药部位、炮制方法等,已构建的数据也会出现原文和结构化断层现象,但如不做规范化处理,就影响数据分析,无法得到内在核心的配伍规律、高频药物等。因此,探索合理的组方药物规范化方法是藏药方剂数据构建及应用的关键技术之一。本研究根据以上方剂,分解组方药物规范的每一步骤,在原文和规范数据能够双向推导的前提下,探索最优路径,发现“一体五层”的组方药物规范方法,是藏医方剂组方信息从文献数据到信息数据升华过程中总结的 5 层路径,本质是将最大限度保留原始文献信息的同时逐步进行规范化处理,得到互通性较高的规范化处理路径,见图 3。

“一体五层”中一体指组方药物,五层依次为原文层,是将上述主干层中方源对应文本中截取组方相关的文本,侧重于原始记录。结构层是将原文在分解字符进行分词,得到结构化数据,携带者原始组方药物信息和剂量信息,该层起初以人工主观分词为主,随着分词的数量越多,构建基本词库后结合语义网络分析采用计算机自动分词,侧重于提取组方,是人机交互的核心环节,该层在五层中受主观意识影响最强,是数据库的数据准确度重要参考层面。本原层是根据“结构-本原”规范字典,将结构层的数据进一步规范化处理的结果,侧重于实际入药,记录本原组方和本原剂量,其中本原组方是指实际入药成分,如小檗中层皮、葶苈根、石榴籽、奶制寒水石、烈香杜鹃花、渣驯膏、生等膏、布谷鸟未鸣前的



中轴表示历史时间,中轴上的节点表示世纪,与中轴相连的节点为作者姓名,远端节点表示五鹏丸衍生方剂,考虑字数过多只标注五鹏丸 5 种组方药物以外的组方。

图2 基于“历史时间轴”的藏药五鹏丸及其衍生方传承源流

Fig. 2 Inheritance and origin of Tibetan medicine Wupeng Pills and the derived prescriptions based on the "historical timeline"

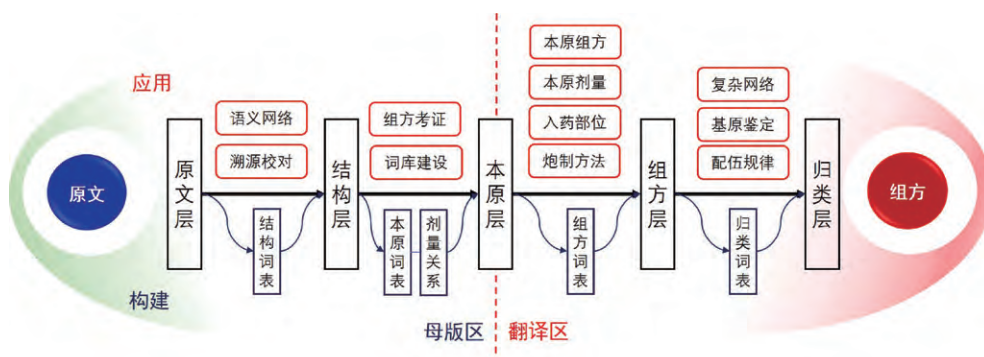


图3 “一体五层”的藏药方剂组方规范化路径示意图

Fig. 3 Schematic diagram of standardized path of Tibetan medicine prescriptions by the "one body and five layers" method

甘青青兰、能见3处(上见雪山、下见河流、前见瀑布)的大果大戟等,包含入药部位、炮制方法、道地药材、采集时间等。本原剂量是指将结构层中的剂量文本规范结果,藏医文献中剂量主要有2种:体积剂量和质量剂量,因古今剂量单位不一,关系复杂,剂量统一又是一个难以攻克的命题,但是方剂收集时发现,一般同一方剂中不会同时出现2种剂量,且研究实际入药剂量又受成药用药量的影响,如散剂1次服用量、丸剂大小等。因此,该层面只注重同一方剂中组方药物的剂量比例关系,不注重实际入药量,如若研究实际入药量,与成药用药量一同计算。明确剂量比例关系就需要古籍记载的剂量单位

转换关系,本研究经文献考证得到常用的剂量单位转换关系,见表2。组方层是本原层数据进一步归纳统一,删去炮制方法、入药部位等组方药物的属性词语,得到与植物学、动物学、矿物学能够直接对接的药材名称,侧重于药物基原,植物药归类到每一个属种,如诃子、篦齿虎耳草、线叶龙胆等;动物药归类到实际药物,如麝香、牛黄、鹿角、贝壳等;矿物药归类指实际药材,如寒水石、光明盐、硃砂等,是研究宏观用药规律、提炼方剂配伍思想的核心层次,也是组方规律可视化展示的直接依据。归类层是根据药物基原,将组方药物归纳到植物药、矿物药、动物药等3类,这层在组方规范中起不到重要作用。

但宏观分析治疗某一疾病的药物构成以及观察和判断整体方剂的用量、构成、储存、辅剂时起到关键作用,侧重于宏观归类。

表 2 基于藏医药古籍的藏药方剂剂量单位转换关系
Table 2 Conversion of dose units of Tibetan medicine prescriptions based on ancient Tibetan medical classics

No.	项目	计量单位	关系 ¹⁾
1	质量	加玛 ^{ཧུ་མ།}	12 000
2		尚 ^{ཤང་།}	1 200
3		雪 ^{ཤོ་།}	120
4		冉德 ^{རང་དེ།}	15
5		赛哇 ^{སེ་བ།}	6
6		青稞 ^{ཇུ་མ།}	1
7		楠 ^{ནུ་མ།}	0. 1
8	体积	担 ^{ཌུ་མ།}	120
9		斗 ^{ཌུ་མ།}	6
10		普 ^{ཕུ་མ།}	1

注: ¹⁾ 以青稞或普为“1”。

随着藏医药研究的深入,若发表国内和国际学术论文,藏医药数据的翻译成为重要环节,对此将上述整体路径又分为 2 个区,分别是母版区和翻译区,在母版区对象主要以原

始文本为主,在翻译区得到计量化数据后可根据研究需求进行翻译,这样会减少工作量,提高准确度。从整体来讲,该路径可根据需求进行多层次、多元化分析等特点,能有效解决单一或机械规范所带来的结果单一化和孤立化问题。

根据“一体五层”组方药物规范化处理路径,对已经收集的方剂进行处理,发现原文层等于方剂数量 11 316 条、结构层 139 011 条、本原层 2 391 条、组方层 1 290 条、归类层 5 条,见图 4。归类层中植物类有 109 813 条,占 79.0%;动物药有 16 086 条,占 11.6%;矿物药有 12 400 条,占 8.9%;除以上 3 类外含有未知类 536 条,占 0.4%,是指根据方剂原文观察是明确组方,但尚不知具体药物,排除类 176 条,占 0.1%,是指根据结构层分词时认为组方药物,但后期人工检查时视为无效数据。对 1 290 个组方药物分析发现,其中植物药 507 个,占 39.3%;动物药 618 个,占 47.9%;矿物药 163 个,占 12.6%,见表 3。本研究收集的 11 316 方剂组方药物从种类来说动物药远超过植物药,而用药次数来看植物药近占八成。在归类层至原文层,选择频次最高的词条为例,转化过程见表 4。可见应用“一体五层”路径,能够让原始数据快速收敛,得到有用信息,向前可以用于各种数据挖掘处理,向后可以对关键问题进行考证,且不同层面数据具有互通性。

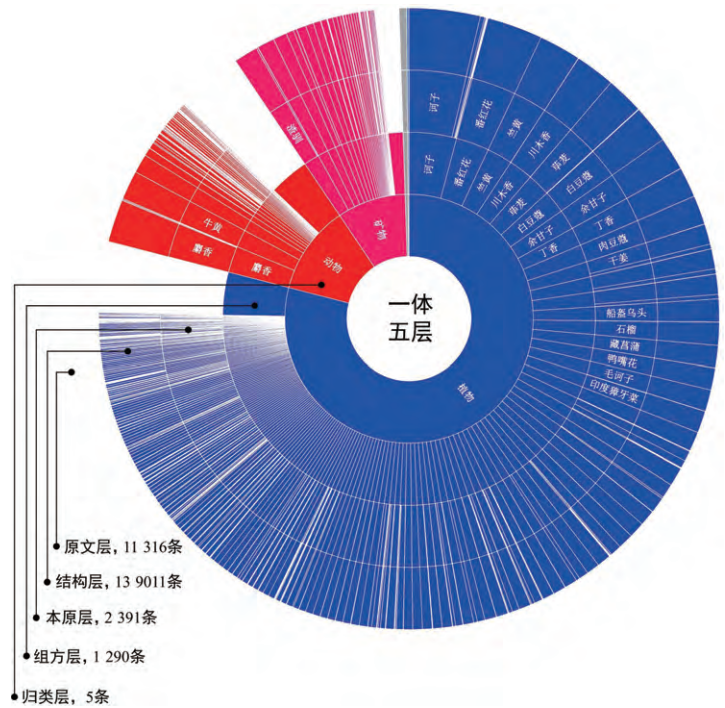


图 4 11 316 首方剂的 139 011 组方药物词条进行“一体五层”规范可视化展示
Fig. 4 Display of 139 011 entries of drug information from 11 316 prescriptions in a standardized and visual way by the "one body and five layers" method

2.4 “连-分-连”的效用信息构建 效用记载着方剂的功效与主治,是藏药方剂从文献到临床应用的依据,在以往效用

信息收集和应用步骤为一是将效用原文进行语义分词,分成有疾病名称和功效名称构成的词表,如清热解毒、治疗赤巴

表 3 11 316 首藏药方剂对应组方药物的归类占比

Table 3 Classification and proportion of drugs corresponding to 11 316 Tibetan medicine prescriptions

归类	类数/个	类别占比/%	条数/条	条数占比/%
植物	507	39.3	109 813	79.0
动物	618	47.9	16 086	11.6
矿物	163	12.6	12 400	8.9
未知	1	0.1	536	0.4
排除	1	0.1	176	0.1

表 4 基于“一体五层”的藏药方剂组方规范化路径

Table 4 Standardized path of Tibetan Chinese prescriptions based on "one body and five layers"

No.	归类层		组方层		本原层(组方层第一药物为例)		结构层(本原层第一药物为例)		原文层(结构层第一药物为例)	
	项目	数量/条	项目	数量/条	项目	数量/条	项目	数量/条	项目	数量/条
1	植物 植物 109 813	109 813	诃子 诃子 5 704	5 704	诃子 诃子 5 133	5 133	诃子 诃子 2 014	2 014		
			红花 红花 4 008		金光诃子 金光诃子 303		金光诃子 金光诃子 171			
			竺黄 竺黄 3 057		短嘴诃子 短嘴诃子 101		三果 三果 143			
			木香 木香 2 897		长嘴诃子 长嘴诃子 61		诃子 20 诃子 56			
			草茛 草茛 2 772		殊胜诃子 殊胜诃子 37		八味沉香 八味沉香 35			
2	动物 麝香 16 086	16 086	麝香 麝香 3 209	3 209	麝香 麝香 3 070	3 070	麝香 麝香 2 020	2 020		
			牛黄 牛黄 2 108		五鹏 五鹏 70		五鹏 五鹏 187			
			熊胆 熊胆 1 369		玛格达 玛格达 69		麝香没药 麝香没药 60			
			全蝎 全蝎 614		麝香 麝香 16		麝香 麝香 16			
			鹿角 鹿角 532		麝香 麝香 13		麝香 麝香 13			
3	矿物 诃子 12 400	12 400	诃子 诃子 1 856	1 856	诃子 诃子 1 729	1 729	诃子 诃子 1 298	1 298		
			诃子 诃子 1 262		诃子 诃子 116		诃子 诃子 11			
			诃子 诃子 997		诃子 诃子 8		诃子 诃子 8			
			诃子 诃子 792		诃子 诃子 2		诃子 诃子 6			
			诃子 诃子 693		诃子 诃子 1		诃子 诃子 6			

“连-分-连”三步走的藏药方剂效用信息收集及应用方法,第一步是从“连”到“分”的过程,是在方剂原文中获取效用原文,数据主要以诗或句子形式相连,根据分解字符将句子拆分成单个字节,并排除连用词“和”(དང་)、组合词“巴”(པ་)等无效字节,得到以字节构成的效用字节库,如将“培根病”拆分成“培”和“根”等。第二步是对“分”好的字节进行算法处理,采用 Apriori 关联规则算法,获取核心字节及字节关联等普遍性字节,引用 TF-IDF 关键词挖掘算法获取特异性字节,得到效用信息当中的主要字节。第三步是“分”到“连”的过程,根据已得到的字节数据,将字节相连成词、恢复语义,如字节“毒”“热”等关键度高,则恢复清热、解毒等功效,若同时具有强关联,则恢复“毒热症”等主治,以五味甘露药浴方及其衍生方剂为例,其效用收集和分析过程见图 5。“连-分-连”步骤与以往的处理步骤相比,时效性成倍提高,且能避开统一过程中的主观因素,有效解决了“主观+大量”带来的诸多问题。

3 讨论

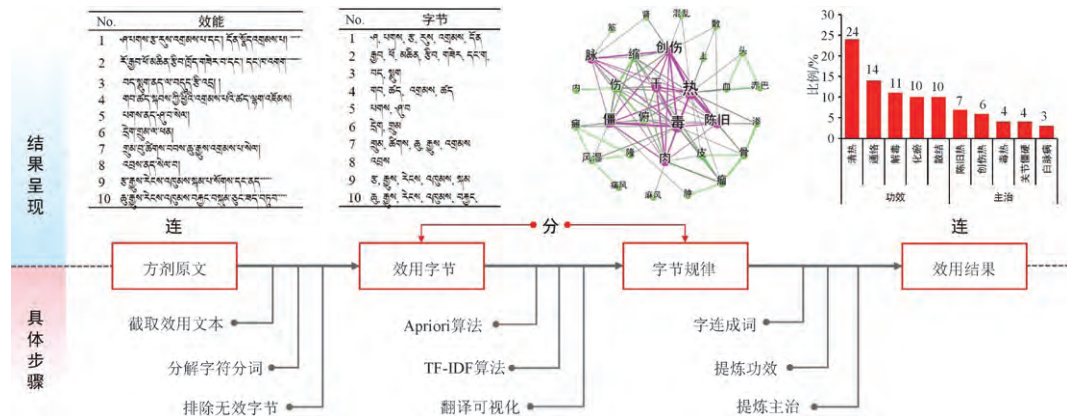
藏医药是中国传统医药的宝贵财富,与人类疾病斗争中积累了人与自然和谐共处的宝贵智慧,也为后人留下了大量

病等;二是对词表进行归类统一形成规范词表,如将消热、清热、热症统一为清热等;三是根据需求收录于数据库或数据分析应用。在这过程中分词和统一往往参与主观判断,且统一处理对象为收集的所有方剂的效用信息,故存在“主观+大量”的弊端,导致信息处理效率低、方法可复制性差、结果准确度低等问题。

本研究基于以上方剂,收集和对比所有效用信息,发现效用信息中主要由功效词语和疾病词语以及连接词语构成,且根据藏文每个字符之间由分解字符相连接的特点,提出

的理论和经验著作。当今诸多基于文献的数据挖掘方法回顾性研究,试图再现古人的学术思想,古籍文献经历史长河流传至今,每一本文献、每个字节都蕴含着古人的智慧,在筛选、规范和统一中要谨慎地判断是否排除。本研究在藏医经典文献中筛选 11 316 首方剂的相关信息为依据,探索藏药方剂的底层信息架构和逻辑规律,提出了藏药方剂信息收集和研究的新模式与新方法,得到合理、准确、高效的藏药方剂信息化处理平台。

已经构建的藏药方剂信息最终通过各种方法,提炼和总结进行知识发现研究,形成各种研究成果,这也是藏药方剂信息库的应用部分,应用部分主要涉及的是适合藏医药自身特点的先进算法,这些算法应用与否、是否合适已成为藏医药信息化、规范化研究进程滞后的重要因素,无法在数据信息处理层面获得原创性成果。本研究基于方剂数据,模拟各种经典算法,不断探索与实践,纳入有益的先进算法,如“与或非”逻辑运算,可以根据以上主干层中任意元素进行精准检索;Apriori 算法,能够获得方剂信息中的普遍性数据,在藏医治疗某疾病的方剂组方配伍规律中具有积极意义;TF-idf 算法能有效挖掘藏医治疗疾病时出现的关键信息,返回概率



图中结果呈现区域涉及数据为五味甘露药浴方数据,在 11 316 首方剂中,设置检索方法,用法:药浴,且组方药物同时含有刺柏、烈香杜鹃、大籽蒿、藏麻黄、水柏枝,得到 115 首五味甘露药浴方的衍生方剂,进行效用信息构建和分析。

图 5 基于“连-分-连”的藏药方剂效用信息构建及分析路径

Fig. 5 Construction and analysis path of efficacy information on Tibetan medicine prescriptions based on "link-split-link"

低但重要的信息。另外还有 Pagerank 算法、偏序结构算法、余弦相似度、Jaccard 相似度、蚁群算法、神经网络等,以上方法能够满足多数藏药方剂的知识发现研究,但仍需基于藏医药自身理论开发工具。本研究基于方剂数据及以上方法,构建和开发“索瓦日巴·藏药方剂数据库”(བོད་སྐད་སྐད་སྐད་མེས་པོའི་དགོངས་མཛོད། TMIP_Smansbyor_1.0)^[31],以该数据库为实践平台,设有一般检索、精准检索、数据分析等多个模块,能够有效、精准检索的同时满足诸多数据挖掘工作。中藏药新药处方筛选和发现是新药创制的首要环节^[32],该平台根据藏医药理论知识的专业判断和历代藏医大师的临床验证,结合复杂网络算法和可视化方法,能够筛选和推荐藏药新药研发的候选方剂,为藏医药知识发现研究提供基础,加快藏药新药创制提供新的思路。

本研究主要以方剂数据的信息框架为主体,以规范、高效为核心目标进行设计,构建了组方药物从原文层到归类层之间所有的对接词表以及效用信息的关键词库,如“原文-结构”词表、“结构-本原”词表、“本原-组方”词表等,在今后研究中可以借助 Python 等编程语句,可以自动在原文中截取有效字词,加快信息处理进程。总之,本研究为藏药方剂信息库的构建与应用提出了底层框架,但依然存在以下不足之处:一是目前仅以部分经典文献为依据进行探索,尚未完成所有藏医文献的方剂信息提取,无法保证数据的完整性和时间的连续性;二是方剂信息从文献到电子文本的时效性较低,藏医文献数据库共享性差,只能手动逐一录入;三是知识发现的算法纳入偏向经典算法,缺乏藏药方剂自身为依据的原创性算法等。本研究后期将以上不足点为重点,逐步解决实际问题,完善藏药方剂信息化、规范化研究路径。

[参考文献]

[1] 张滕,任明,郭利平. 数据库技术在中医药领域的应用现状[J]. 中西医结合心脑血管病杂志,2016,14(14):1628.

[2] 朱化珍,陈德兴. 分子生物技术在中药复方药理研究中的应用[J]. 中国实验方剂学杂志,2011,17(7):27.

[3] 龙伟.“计算中药学”在中药药性及复方研究中的应用[D]. 北京:北京协和医学院,2011.

[4] 范晓辉,程翼宇,张伯礼. 网络方剂学:方剂现代研究的新策略[J]. 中国中药杂志,2015,40(1):1.

[5] 龙伟,邓馨,向剑,等. 中药方剂网络与中药化学空间的构建与分析[J]. 北京中医药大学学报,2011,34(11):729.

[6] WÜNTRANG D, TAWNI T, WANG X, et al. Tibetan medical informatics; an emerging field in Sowa Rigpa pharmacological & clinical research[J]. J Ethnopharmacol, 2020, 250: 112481.

[7] WÜNTRANG D, DUNGKAR T, RIGDZIN W, et al. Dataset of illness classifications in Sowa Rigpa: compilations from the oral instructions treatise of the Tibetan medical classic (Rgyud bzhi)[J]. Data Brief, 2020, 29: 105321.

[8] 文成当智,贡保东知,夺机仁青,等. 基于 Visual FoxPro 和复杂网络分析的《四部医典》方剂数据库构建及其用药规律分析[J]. 中国实验方剂学杂志,2019,25(22):175.

[9] 文成当智,贡保东知,东改措,等. 《四部医典》用药规律:“味性化味”理论的科学内涵分析[J]. 中国实验方剂学杂志,2019,25(5):201.

[10] 文成当智,贡保东知,贡却拉姆,等. 藏族药药性量化的新方法:“味性化味”矢量结构模型构建[J]. 中国实验方剂学杂志,2019,25(19):147.

[11] 文成当智,张云森,仁真旺甲,等. 基于“味性化味-网络靶点-分子对接”的藏药五味麝香丸治疗“真布”病的作用机制研究[J]. 中国药房,2020,31(2):164.

[12] 孙春全,黎元元,谢振年,等. 藏族医药经典名方的一般筛选原则及其传承路径[J]. 中国中药杂志,2021,46(19):5130.

[13] 帝玛尔·丹增彭措. 帝玛尔医著集(藏文)[M]. 北京:民族出版社,2007.

[14] 才让南加. 《医疗瑰宝本草》的研究[D]. 北京:北京中医药

- 大学, 2017.
- [15] 罗布顿珠. 帝玛尔·格西丹增彭措药学术成就及《晶珠本草》学术价值研究[D]. 北京: 北京中医药大学, 2012.
- [16] 嘎务. 藏药晶镜本草(藏文)[M]. 北京: 民族出版社, 2018.
- [17] 贾敏如, 张艺. 中国民族药辞典[M]. 北京: 中国医药科技出版社, 2016.
- [18] 才让吉, 切尼项毛, 贡保东知, 等. 基于数据挖掘的藏医治疗中风(萨滞布病)的用药规律分析[J]. 世界科学技术(中医药现代化), 2020, 22(4): 1361.
- [19] 格日多杰, 文成当智, 久先, 等. 基于数据挖掘和“方-药-性-效”的藏医治疗慢性胃炎用药规律研究[J]. 世界科学技术(中医药现代化), 2020, 22(1): 108.
- [20] 贡保东知, 罗晴方, 余羊羊, 等. 基于数据挖掘和整合药理学平台的藏医治疗脾胃病用药规律及作用机制[J]. 中国中药杂志, 2018, 43(16): 3368.
- [21] 文成当智. 基于方-药-性-效信息化平台的藏医治疗真布病临床用药规律及作用机制研究[D]. 成都: 成都中医药大学, 2020.
- [22] 武佳杰, 任小巧, 李佳林, 等. 基于藏医热病学思想的藏医肺炎诊疗规律探析[J]. 世界科学技术(中医药现代化), 2020, 22(6): 2063.
- [23] 享·斯基巴. 享·斯基巴医学著作考源(藏文)[M]. 成都: 四川民族出版社, 2021.
- [24] 贡曼贡却德勒. 零星秘诀集(藏文)[M]. 北京: 民族出版社, 2004.
- [25] 苏喀年尼多杰. 千万舍利子(藏文)[M]. 北京: 民族出版社, 2005.
- [26] 秀多贡布. 秀多医学集(藏文)[M]. 北京: 民族出版社, 2006.
- [27] 尕玛冉纳. 尕玛冉纳医嘱选集(藏文)[M]. 北京: 民族出版社, 2008.
- [28] 工竹云丹嘉措. 藏医临床札记(藏文)[M]. 北京: 民族出版社, 2006.
- [29] 中华人民共和国卫生部药典委员会. 中华人民共和国卫生部药品标准·藏药[M]. 北京: 人民卫生出版社, 1998.
- [30] 玉多·云登贡布. 四部医典(藏文)[M]. 拉萨: 西藏人民出版社, 1982.
- [31] 文成当智, 东改措. 索瓦日巴藏药方剂库系统(TMIP_Smansbyor_1.0); 2021SR1252511[P]. 2021-08-24.
- [32] 杨洪军, 唐仕欢, 申丹. 源于中医传统知识与临床实践的中药新药发现研究策略[J]. 中国实验方剂学杂志, 2014, 20(14): 1.

[责任编辑 丁广治]