

本草纲目药物警戒数据库设计与应用研究

石琳¹ 董春茹² 张冰¹

(1 北京中医药大学,北京,100029; 2 河北大学数学与信息科学学院,保定,071002)

摘要 目的:为深入分析研究本草纲目中的药物警戒思想提供新方法。方法:基于本草纲目金陵本文献著作设计了药物警戒数据库,阐述了数据库建立的要求和存储模型的选取,以及数据库的概念和逻辑设计。结果:在此数据库的基础上可建立知识发现系统,应用数据挖掘算法,对本草纲目中的药物警戒思想进行分析。结论:本草纲目药物警戒思想数据库的建立可以为深入研究药物警戒思想提供方便、准确的手段,此方法也可应用于其他文献著作研究。

关键词 药物警戒;数据挖掘;数据库;本草纲目

**Design and Application Research of Traditional Chinese Medicine Pharmacovigilance
Database based on Compendium of Materia Medica**

Shi Lin¹, Dong Chunru², Zhang Bing¹

(1 Beijing University of Traditional Chinese Medicine, Beijing 100029, China; 2 Institute of Mathematics
and Information Science, Hebei University, Baoding 071002, China)

Abstract For further analysis and research of traditional Chinese medicine pharmacovigilance in Ben Cao Gang Mu (Compendium of Materia Medica), we designed a database system based on Jinling version of the Ben Cao Gang Mu. Firstly, the paper expounds the requirements of database establishment and selection of storage model. Then it introduces the design of concept and logic of the pharmacovigilance database, and explains the structure and data flow of the knowledge discovery system based on this database. Finally the paper lists some researches on the safety of prescribing Chinese medicine that can be done with the application of data mining algorithms.

Key Words Pharmacovigilance; Data mining; database; Compendium of Materia Medica

中图分类号:R281.3 文献标识码:A doi:10.3969/j.issn.1673-7202.2017.04.051

“药物警戒”一词最早于 1974 年由法国科学家提出^[1]。2002 世界卫生组织对药物警戒正式定义为“有关发现、评估、理解和预防不良反应或其他任何与药物相关问题的科学研究和活动”^[2]。中药药物警戒思想早在几百年前已被提出,而且它的内涵比西方药物警戒更为丰富。中药药物警戒源于中医理论中“以偏纠偏”“以毒攻毒”的药物使用原则。因此,中药药物警戒思想不仅仅关注药物的不良反应,而是围绕着“识毒、用毒、解毒和防毒”多个层次展开^[3]。然而,面对有几千年的历史中医药应用实践,传统的资料搜索、数量统计等方法在数据的隐含关系发现上显得力不从心,更好的传承和发展传统中医药学需要结合飞速发展的计算机技术^[4-5]。本文主要研究内容是对《本草纲目》中的药物警戒思

想进行整理、归纳,并设计适合数据挖掘、机器学习应用的药物警戒数据库,从而为中药药物警戒研究提供丰富的数据支撑。

1 资料与方法

1.1 文献来源 中药瑰宝传世巨著《本草纲目》中积累了几百年来中医文献资料,其中记载了大量有关中药药物警戒的实践经验。这些经验作为基础数据,对于系统研究中药药物警戒体系有着不可或缺的重要作用。本研究使用的是《本草纲目》金陵本。

1.2 检索策略 《本草纲目》在药物来源、加工炮制、毒性、使用禁忌等多个方面都有药物警戒思想的体现,对安全用药有重要的指导作用。数据的检索即从上述方面进行。

基金项目:国家食品药品监督管理总局药品评价中心项目:基于中医药典籍数据挖掘的乌头属中药药物警戒思想研究,项目负责人:张冰

作者简介:石琳(1976.11—),女,在读博士研究生,副研究员,研究方向:中药药物警戒与合理用药研究,E-mail:censhizi@163.com

通信作者:张冰(1959.08—),女,博士,教授,主任医师,博士研究生导师,研究方向:中药药物警戒与合理用药研究、中药药性理论基础与应用、中药防治代谢性疾病的临床与实验研究,E-mail:zhangbing6@263.net

1.3 纳入标准 清晰理解药物警戒思想在《本草纲目》中的体现方式,才能全面、准确的针对数据源设计数据库。1) 药物来源。“释名”和“集解”中清晰描述了药物的品种、产地、形、色,可用以辨析胜品、区分伪品。药物来源是对用药安全的源头控制。2) 毒性。药物毒性的描述多见于“性味”中,以“有毒”“大毒”“小毒”“微毒”等用语表达,多指一些具有一定毒性或不良反应的药物^[6]。3) 使用禁忌。“气味”描述中以“忌恶反畏”表示药对间的配伍禁忌信息;“附方”“发明”中多以“忌”“忌食”“勿服”“勿入”“共食…”等描述表示组方与某种药物或某种食物间的使用禁忌信息;以“剧泻药”“催吐药”“活血破血”“产妇尤忌”等表示妊娠禁忌药物。4) 解毒。“主治”“发明”“附方”描述中多以“解…毒”“杀…毒”“伏”“制”表示某种药物可解毒的信息,体现了中毒解救的安全用药思想。5) 加工炮制。药物的加工炮制方法多在“修治”中进行说明。6) 用法用量。“气味”“主治”“发明”“附方”中多以“多食…”“久食…”“不可食”“…杀人”等表示某种药物在用法用量方面的安全使用事项。

1.4 排除标准 有的讨论、质疑、个例信息虽然也属于药物来源、加工炮制、毒性、使用禁忌等方面,但为保证数据的可靠性、准确性,此类数据被排除。

1.5 数据库的规范与数据库的建立

1.5.1 数据库的建立 数据库的建立需满足以下要求:1) 结合数据特征,准确表述其含义。数据完全覆盖《本草纲目》中有关药物警戒思想的内容,信息表示准确,不失原著原貌,同时对噪声数据有恰当的处理。上述各种药物警戒思想的表达均以符合各自特征的形式存储于数据库中。2) 主题特点突出,数据方便利用。数据库的建立是为了对《本草纲目》中药物警戒思想进行深入研究,所承载的核心功能是传统药物安全使用的数据分析与知识发现。3) 数据库具有一定的可扩展性。此研究的数据来源目前限定于《本草纲目》,但随着研究的深入,当增加数据源时,数据结构需不扩展或轻易扩展实现,以保护现有研究成果和实现多数据源的数据兼容。

目前常用的数据库模型主要是关系型数据库和非关系型数据库2种^[7]。虽然随着大数据和云计算应用的迅速推广,非关系型数据库正在被越来越多的使用,但结合研究主题、数据库建立要求以及2种数据库模型的特点,本设计采用的仍然是传统的关系型数据库。主要原因如下:1) 非关系型数据库的优势主要是针对大数据的特点,即数据量大、数据变

化快,以及之前可能未知的数据^[8]。此研究中的源数据明确,并且在整理录入后几乎不会再发生变化。2) 非关系型数据库对分布式计算提供更好的支持,能轻易实现“横向扩展”^[9]。此研究的数据量并不算大,也不需要提供高并发性的数据读写,不必采用分布式计算。3) 关系型数据库更适用于数据以结构化和半结构化方式存储。上述药物警戒思想中需提炼的内容均可以整理为结构化或半结构化数据。4) 关系型数据库的完整性和一致性可以对交互式应用系统提供更好的支持。这样在数据库建立后,可以在其上开发除数据分析外的其他应用。5) 关系型数据也即行数据,可轻易实现向键值存储、文档存储、列存储等非关系型数据库数据存储格式的转换。而非关系型数据转换为关系型数据却较为困难。

综上所述如果采用非关系型数据库进行设计不但无法应用其优势特点,还会增加设计的复杂性,而关系型数据库可以更好的支持数据分析和利用,并且即使以后有向大数据非关系型数据库并入的需求,也可以较容易的完成转换。本研究采用的是开源的关系型数据库 MySQL。

1.5.2 数据库的规范 概念设计采用实体关系(Entity-Relationship, ER)模型。依据《本草纲目》中药物及药物警戒思想表达方式建立实体。由于所建实体数量很多,此处只选择了几个有代表性的实体进行说明。1) 药物名称。其属性有药物代码、药物正名、部。属性只选择了有唯一取值的信息,此实体作为模型中其他实体间联系的核心。2) 药物别名。其属性有药物代码、序号、别名、出处。每种药物可以有任意多种别名,所以属性“别名”的域是无范围限制取值。别名信息的建立有助于利用此数据库产生的规则对其他文献资料进行学习。3) 性。其属性有药物代码、出处、部位、寒、热、温、凉、平。属性“出处”可用以区分《本草纲目》中记载的不同出处对药物性味的不同描述。属性“部位”用以标识药物的不同部位如实、根、叶、皮等,因为不同部位可能具有不同的性味。之所以选取“性”的取值作为属性,是因为其取值范围有限,此种设计可以减少信息的冗余存储。以属性“寒”为例,其域可以定义为“0”“1”“2”“3”分别表示“无此属性”“微寒”“寒”“大寒”。4) 配伍禁忌。其属性有药物代码1、药物代码2、出处、禁忌。记录药对间的配伍禁忌信息。属性“禁忌”的域为“忌”“恶”“反”“畏”。由于部分药物中是特定部位与其他药物有配伍禁忌关系。例如,栝楼根恶干姜、反乌头,而栝楼实则无此记载。为准确

记录此类信息可以另行设置配伍禁忌描述实体,来描述细节情况。5) 附方使用。其中有附方药物实体,其属性有附方代码、药物代码,用来记录附方中所含药物;附方描述实体,其属性有附方代码、附方内容描述。内容描述采用文本类型进行记录,不进行整理和归类,以保证其原貌并且不丢失信息;附方禁忌实体,其属性有附方代码、禁忌药物代码。6) 药物来源。其属性有药物代码、产地代码、来源描述。

还有部位、味、毒性、解毒、来源、食物、产地、炮制方法、饮食禁忌、用法、用量、注意事项等等其他实体均按照上述实体的思路进行设计。部分 E-R 图示例如图 1 所示。将 E-R 图转换为关系模型,进而即可生成 Mysql 数据库所支持的逻辑结构。

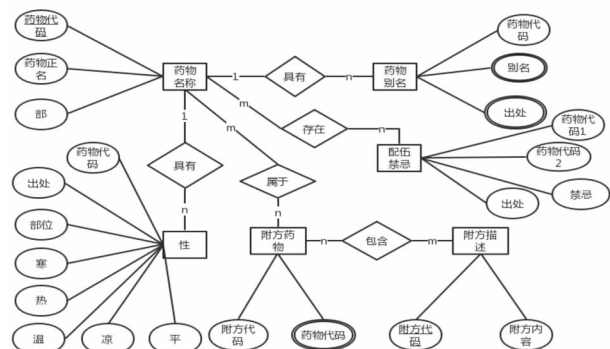


图 1 部分 E-R 图示例

1.6 数据分析 数据库建立后可应用计算机技术进行传统药物信息的深入分析和知识发现,从而更紧密的将中药实践与理论结合。在建立的药物警戒数据库上可以应用数据挖掘和机器学习技术构建知识发现系统,系统结构及数据流程如图 2 所示。数据挖掘是发现大量数据中的隐藏信息的高效手段^[10],常用的数据挖掘方法有关联分析、聚类、分类、回归分析等。而无论使用哪种数据分析方法,源数据均可以从所构建的数据库中便捷获取,并通过行列转换、数据拼接、关键词提取等数据预处理方法得到某种数据分析方法所需要的特定格式数据。数据预处理完成后,可进一步选用各类经典数据挖掘算法形成算法库,并针对研究主题选取合适的算法在数据上进行学习,最终形成知识库。下面我们列举了几种常见的数据挖掘方法在我们建立的数据集上可能的应用及相应思路。

1.6.1 关联分析 关联分析是从数据中挖掘数据项之间相互联系的常用方法之一^[11],可以通过药物的基本属性来研究药物之间安全使用的内在联系。如,可通过关联分析可研究药物性味归经、升降沉浮

属性与药物配伍禁忌之间的关系;附方药物与药物和食物间使用禁忌的关系;附方中反药组合的关系等。

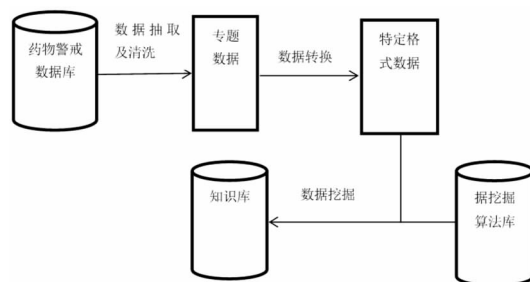


图 2 药物警戒知识发现系统结构及数据流程

1.6.2 分类及聚类 分类分析是首先从历史经验数据中选出带有类别标签的训练集,然后根据某种分类算法在该训练集上建立分类模型,最终使用所建立的分类模型对没有分类的数据进行分类^[12]。而聚类分析通过某种聚类算法对获得的数据进行分组,要求相似的记录属于同一个分组,而不相似的样本属于不同的分组。聚类和分类的区别是聚类不依赖于样本预先定义好的类标签,即不需要训练集。分类和聚类分析可用来针对不同的安全用药研究主题对药物进行分类。如可以基于分类方法从历史数据中学习关于药物相恶的知识,并用学习到的分类器判别 2 种新药物是否相恶;而基于聚类分析可以通过药物的各个属性,对不同的药物进行聚类,从而研究不同聚类的药物所具有的共性特征知识,并可进一步基于学习到的知识对新药物所属的类别进行辅助判别。

1.6.3 回归分析 回归分析是一个统计预测模型,用以描述和评估因变量与一个或多个自变量之间的关系^[13]。回归分析可用来研究药物基本属性在药物安全使用方面的主次关系;多种药物同时使用可能会存在使用禁忌的概率等。

2 结果

《本草纲目》中有丰富的药物警戒方面的思想及知识,并且这些思想和知识的表达存在其自身的特点,使人很难系统的掌握其精华。因此,为了超越以往针对该问题的简单检索、统计研究,本文在深入分析金陵版《本草纲目》的特点的基础上,设计并构建了中草药警戒数据库,从而为方便高效地利用数据挖掘和机器学习技术对中草药警戒知识进行深入分析和学习提供了数据支撑。

3 讨论

数据库的建立有助于便捷、准确、灵活地获取适

(下接第 916 页)

可见,“通阳”用于湿阻证,“温阳”用于阳虚之候。而温热药物使用的目的,皆在于纠正人体“湿盛则阳微”之状态,以利于人体气化功能的恢复,最终清除湿热之邪。

综上所述,温热药物治疗湿热病的机制渊源于《内经》《伤寒杂病论》,历代医家在继承中医经典理论的基础上结合临床不断发挥,特别是温病大家叶天士在《温热论》中的精辟论述,为我们树立了光辉的典范。温热药的合理运用,极大地丰富了湿热病的治疗内容,尤其是启迪并拓展了我们临床用药思路。在现代疾病谱湿热证逐年增多的今天,开展温热药物在湿热病治疗中应用的讨论,具有重要现实意义。

参考文献

- [1] 刘景源,王庆侠. 温病辨治汇讲第 11 讲湿热病辨治概要-湿热病的辨证纲领,治疗原则与禁忌[J]. 中国临床医生,1999,27(7):2-5.
- [2] 赵绍琴. 温热纵横[M]. 北京:人民卫生出版社,2012:13.
- [3] 曾艺鹏,李剑锋,蒋玉珍. 分消走泄方药配伍理论探讨[J]. 新中医,2006,38(3):7-9.
- [4] 刘完素. 素问玄机原病式[M]. 4 版. 北京:人民卫生出版社,1983:61.
- [5] 王士雄. 温热经纬[M]. 北京:中国中医药出版社,1996:174.
- [6] 孙思邈. 中医必读百部名著·备急千金要方[M]. 北京:华夏出版社,2008:230.
- [7] 刘景源. “分消走泄”法在湿热病治疗中的应用[J]. 中国中医药现代远程教育,2006,4(8):32-35.
- [8] 王士雄. 温热经纬[M]. 北京:中国中医药出版社,1996:69.
- [9] 赵绍琴. 温热纵横[M]. 北京:人民卫生出版社,2012:201.
- [10] 彭胜权. 温病学[M]. 北京:人民卫生出版社,2000:1012.
- [11] 赵以德. 金匱玉函经二注[M]. 北京:人民卫生出版社,1990:40.
- [12] 刘渡舟. 湿证论[J]. 北京中医药大学学报,1998,21(1):3-5.
- [13] 贾所学. 药品化义[M]. 北京:中医古籍出版社,2012:28.
- [14] 沈凤阁,王灿晖,孟树江. 叶香岩外感温热篇、薛生白湿热病篇阐释[M]. 南京:江苏科学技术出版社,1983:159.
- [15] 吴塘. 温病条辨[M]. 北京:中国书店,1994:95-133.
- [16] 叶天士. 临证指南医案[M]. 北京:华夏出版社,1995:198.
- [17] 刘景源. “分消走泄”法在湿热病治疗中的应用(二)[J]. 中国中医药现代远程教育,2006,4(9):2-3.
- [18] 刘古文. “通阳不在温,而在利小便”探析[J]. 陕西中医,1987,8(4):183-183.

(2016-06-06 收稿 责任编辑:白桦)

(上接第 911 页)

用于各种分析算法的特定格式、特定组合的数据。相比之下,关系型数据库比非关系型数据库能更好的适用此研究主题。同时可以对经典的数据挖掘、机器学习算法进行选取、整理,形成数据挖掘算法库,以方便更加科学的对中药特性进行深入理论研究。虽然此研究属于面向特定著作的主题数据库,但研究方法却可以方便地扩展到其他文献著作,从而为建立用途更为全面的中医药数据库提供了可能的设计思路。

参考文献

- [1] 刘巍,陈易新. 药物警戒的概念与起源[J]. 中国执业药师,2008,5(7):16-18.
- [2] Springer Vienna. WHO collaborating centre for international drug monitoring[M]. Dictionary of Pharmaceutical Medicine,2009:37-38.
- [3] 吴嘉瑞,梁秉中,张冰. 中药不良反应再认识[J]. 中国药物警戒,2006,3(1):40-42,49.
- [4] 邢春国,施诚. 数据库技术在中医药学文献研究方面的应用[J]. 医学信息,2008,21(11):1933-1936.
- [5] 吴春秋. 数据库技术在《本草纲目》中食物文献的研究及应用[D]. 北京:北京中医药大学,2007:17.
- [6] 王建,班炳坤,高天,等. 中药警戒思想在本草文献中的表达方式[J]. 中国药物警戒,2010,7(9):533-536.
- [7] 戴特. 数据库系统导论[M]. 北京:机械工业出版社,2007:21-22.
- [8] Kristina Chodorow, Micheal Dirolf. MongoDB 权威指南[M]. 北京:人民邮电出版社,2011:67-69.
- [9] White Tom. Hadoop 权威指南[M]. 北京:清华大学出版社,2011:102-104.
- [10] Jiawei Han, Micheline Kamber, Jian Pei. Data Mining: Concepts and Techniques-third edition[M]. Waltham USA, Morgan Kaufmann Publishers,2011:24.
- [11] Aggarwal R., Srikant R. Fast Algorithms for Mining Association Rules[J]. Proceedings of the 20th VLDB Conference Santiago, Chile,1994,15:487-499.
- [12] 孙吉贵,刘杰,赵连宇. 聚类算法研究[J]. 软件学报,2008,19(1):48-61.
- [13] 蔡章利,石为人,刁绫. 回归分析在连续型数据目标预测中的应用[J]. 重庆大学学报:自然科学版,2005,28(10):91-93.

(2016-12-02 收稿 责任编辑:王明)