

基于电子病历结构化的名老中医经验数据挖掘研究

刘鸿燕 胡红濮

张 越

(中国医学科学院医学信息研究所 北京 100020)

(北京中医药大学管理学院 北京 100029)

〔摘要〕 数据挖掘在中医药信息化领域的重要地位日益突显,在名老中医经验传承方面尤为突出。名老中医经验数据挖掘的基础为结构化的中医电子病历,详细阐述中医电子病历的表达架构,数据采集技术,名老中医经验数据挖掘中对显性知识和隐性知识的常用挖掘方法,提出存在的问题及相应建议。

〔关键词〕 电子病历;名老中医经验;数据挖掘

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2015.12.003

Data Mining Research on Experience of Famous Traditional Chinese Medicine Doctors Based on Structured Electronic Medical Records LIU Hong-yan, HU Hong-pu, Institute of Medical Information, Chinese Academy of Medical Science, Beijing100020, China; ZHANG Yue, School of Management Administration, Beijing University of Chinese Medicine, Beijing100029, China

〔Abstract〕 Data mining acquires an increasingly greater significance in Traditional Chinese Medicine (TCM) field, it plays a more important role in inheriting experience of famous doctors of TCM. Data mining of their experience is based on structured TCM Electronic Medical Records (EMR). By explaining the expression structure of TCM EMR and data collection technology, the common used methods for both explicit knowledge and tacit knowledge in data mining of experiences of famous TCM doctors, the paper proposes existing problems and relevant suggestions.

〔Keywords〕 Electronic Medical Records (EMR); Experience from the famous doctors of TCM; Data mining

1 引言

中医学是中华民族的文化瑰宝,传统中医学是经验医学,依靠医生在长期行医过程中经验的积累。中医学发展现代化、国际化的基础就是对现已形成的中医药理论体系科学地传承与发展。采用数据挖掘技术对中医电子病历中结构化与非结构化的

数据进行分析,提取总结名老中医行医规律,精炼个性化诊疗特征,提取更为科学、易于推广的模型与模版。模版可存储大量的结构化的基础方和古方,再根据患者个性化情况酌情加减,适当修改治疗方案,这样既提高诊疗效率,又减少了漏诊、误诊、配伍禁忌等问题的发生,也有利于青年医生从中学习名老中医经验,加深对临床知识的理解^[1]。

2 名老中医经验数据挖掘意义及作用

2.1 名老中医经验的重要地位

名老中医是具有丰富临床诊疗经验的专家,医术在长期临床诊疗过程中不断精进,为中医药创新

〔修回日期〕 2015-11-30

〔作者简介〕 刘鸿燕,在读硕士研究生;通讯作者:胡红濮,副研究员,研究方向为卫生信息管理,发表论文 50 余篇。

发展提供动力。因此有效利用名老中医的经验、用药规律和学术思想,并进行总结和传承工作十分必要。现已成型的中医药体系和经验主要通过文献古籍和师承两种形式保存并发展,师承是继承名老中医宝贵诊疗经验的最佳方式,但由于受众较少,成才缓慢等方面的局限,无法缓解名老中医减少,宝贵的学术经验流失的问题。因此,1990 年国家原人事部、原卫生部、国家中医药管理局联合颁布的《关于采取紧急措施做好老中医药专家学术经验继承工作的决定》取得显著成效。《医药卫生中长期人才发展规划(2011-2020 年)》^[2]中提到:开展全国老中医药专家学术经验和基层老中医药专家临床经验继承工作。国家的大力支持可见名老中医经验的优势与地位已日渐凸显。

2.2 中医电子病历的发展辅助名老中医经验数据挖掘

2.2.1 概念与特点 中医电子病历,即在中医医院使用,符合中医临床记录特点,满足所有的医疗、法律和管理需求的计算机化的病历^[3]。中医电子病历除具备一般电子病历的特征外,还具有自身的特点,主要表现在内容、诊断、处方的特殊性等方面。(1) 内容。中医学有许多特有的术语,如病名、方剂名、穴位名等。此外,古代文学对中医药有深远的影响,在阐述中医观点时,引用中医经典著作中的原文十分普遍。(2) 诊断。中医有自成体系的诊疗术语,包括四诊(望、闻、问、切)、诊断、辨证要点、治法治则等等。这些显著的中医药特征体现了中医辨证论治和“个性化诊疗”的思想。(3) 处方。中医处方大多为复方。复方通过配伍增强治疗作用,同时降低甚至消除某些副作用。君臣佐使是方剂配伍组成的基本原则,临症时根据患者的个体情况对基础方进行加减调配,具有一定的随机性,因此需要依靠电子病历结构化来降低名老中医经验数据挖掘、传承的难度。

2.2.2 中医电子病历在名老中医经验整理工作中的作用 20 世纪 70 年代开始,国内外专家利用现代计算机技术,研究中医临床诊断系统和专家辅助诊断系统^[4],但将现代数学与计算机技术相融合的

名老中医经验数据分析与挖掘方面,仍然存在信息采集的结构化问题、数据库问题、分析挖掘中解析的问题、无法保留中医数据的动态特征等问题。名老中医经验存在于临床过程形成的医案中,在医院信息化与电子病历发展充分之前,医案以纸质版存在居多,或者纯文本电子文档形式存储。名老中医经验数据挖掘要求医案信息数字化、结构化和标准化。中医电子病历的结构化成为了名老中医经验信息挖掘的数据基础,中医电子病历的质量影响着数据挖掘的程度。

3 中医电子病历结构化

3.1 概况

结构化是指将自然语言表达的信息分解为计算机可识别的一系列规范化元素^[5]。电子病历结构化才能实现进一步加工,从简单的检索到复杂的决策支持都依赖于电子病历的结构化。电子病历信息的结构化程度往往决定了服务临床及科研系统能力的强弱,体现了电子病历的发展程度。在电子病历中,检验报告、药疗医嘱等内容都是结构化的,检查报告、病情症状是半结构化的;病程记录等内容则基本上是自由化的。中医电子病历的结构化既需要满足法律、医疗、管理方面的要求;还要满足中医医院各信息系统采集的要求^[6]。

3.2 表达架构

3.2.1 语法层面 既包括病历书写规范,也包括病历元素规范。标准规范的中医临床术语为结构化电子病历提供支撑服务。如由国家中医药管理局制定发布的《中医电子病历基本规范(试行)》对规范医疗机构中医电子病历的使用管理,保证医患双方合法权益,病历质量监控、医疗卫生服务信息及数据统计分析和医疗保险费用审核提供技术支持^[7]。

3.2.2 语义层面 强调对病历内容的编码化,对照受控医学术语集,将病历以编码化的形式组织起来,这方面的标准如 HL7 临床文档结构、DICOM SR 协议等。国家质量监督检验检疫总局、国家标准

化管理委员会批准发布《中药方剂编码规则及编码》、《中药编码规则及编码》、《中药在供应链管理中的编码与表示》等系列中医药国家标准并于 2015 年 12 月 1 日开始实施^[8]。

3.2.3 语用层面 对文档内容进行理解应用,例如在验证患者性别为“男”时,电子病历不应该出现月经史这样的选项。目前语用层面还处于探索阶段,已出现资源描述框架(Resource Description Framework, RDF)等标准,但并没有针对中医电子病历合适的标准规范。

3.3 数据采集技术

3.3.1 采集要素 所有与名老中医临床经验有关的因素,都是需要采集的临床要素,包括:患者的基本信息、病史、舌脉诊、理化检查、诊断(包括中西医诊断)、辨证要点、治则治法(治疗原则及中药治法、针灸治法)、汤药处方、成药及西药处方、针灸处方、其他治疗记录、预防与调摄、依从性评价、指标追踪、结局评估等方面。以上采集要素需要结构化的标准化临床术语支持,包括临床信息相关的各种国际标准、国家标准(如中医疾病诊断、中医证候诊断术语、症状类术语、治则治法术语、方剂术语、中药术语、西医疾病 ICD-10 字典、辅助检查术语等),同时还需要现代医学 SNOMED 等标准规范。除了结构化的标准化临床术语外,中医电子病历中半结构化的数据,或文本电子病历系统的数据、使用文档记录的临床医案信息,需要通过结构化的方式进行数据加工,提高利用率和准确率,为后续的数据挖掘提供基础。

3.3.2 采集技术 既要留存名老中医经验信息的原始文本,又要保证信息动态、多维、结构化的特性,生成有利于直接分析的结构化数据^[9]。采集技术主要包括两种,自然语言处理和机构化数据录入。自然语言处理的目的是从自由文本中自动地提取代码化的医疗数据^[10],其优点在于直接采集现有的自由文本,临床医生在表达临床发现和做出决策等方面不必改变他们习惯的记录方式^[11],不足之处是数据存在大量噪声,有解释错误的可能,需核查得出的结论,进行人为修正。结构化数据录入是

数据直接输入到结构化模块。在电子病历模板中设置结构标志,允许进行自由文本录入,用户可根据需要调用不同的模版并提取结构,从而应对复杂的电子病历录入。虽然结构化数据录入的结果满足计算机对信息的处理要求,但缺点是用户无法自由描述,较难表达真实意思^[12]。

4 名老中医经验数据挖掘

4.1 概况

在第一届中国中医药信息大会上,指出“十二五”以来中医医院信息化建设成效显著,全国 82.25% 的中医医院建立了医院信息系统,94.98% 建立了药品管理信息系统,42.41% 建立了医生工作站,53.7% 建立了电子病历。在国家的大力支持下,信息化助力中医药事业发展,为中医电子病历的发展提供广阔空间,也有利于利用中医电子病历实现名老中医临床诊疗经验数据采集、挖掘。目前上海、广东等省市实现了区域内中医医院系统共享以及跨区域业务协调。北京、重庆、浙江等省市纷纷建立了名老中医临床诊疗经验研究等特色系统。

4.2 数据挖掘

根据分析挖掘知识的特点,将中医临床经验分为显性及隐性两大类,显性经验是指通过频次统计等一般的统计分析方法即可获得的经验知识,而隐性经验是蕴含在其内部的,通过复杂的关系分析才能获得的知识。后者常用方法包括复杂网络分析、点式互信息、隐树模型^[13]等方法,以人为主,人机结合,辅助专家从大量的数据中获取简洁的框架知识和模型,提高研究的效率及准确性。

4.2.1 显性知识 显性知识的快速挖掘通常采用联机分析处理方法(On-line Transaction Processing, OLAP),是指针对某个特定的主题进行联机数据访问、处理和分析,通过直观的方式从多个维度、多种数据综合程度将系统的运营情况展现给使用者^[14]。最流行的支持 OLAP 的数据仓库模型是(Multi-dimensional OLAP)^[15],实现名老中医临床诊疗数据的多维存储,实时在线多维检索查询分

析,形成主题分析数据库。适用于大规模结构化数据的即时快速查询分析,提供丰富的分析图表。OLAP 是医疗领域知识发现的重要工具,将医疗数据挖掘模型及其结果与电子病历相结合,使电子病历中的数据转化成其本身可以利用的信息,实现系统的闭环控制,使电子病历具有自动学习、自我完善的能力^[16]。

4.2.2 隐性知识 常见的方法有:关联规则、判别分析、决策树、支持向量机、神经网络等^[17]。在名老中医经验数据方面应用较为成熟的方法有复杂网络分析、点式互信息、多因子降维分析、马尔科夫决策过程模型等。(1) 复杂网络 (Complex Network) 将领域科学问题建模为以节点与边表示的网络模型,在此基础上进行网络拓扑分析。针对疾病和人体生命系统的复杂性,基于复杂网络视角的网络医学已经成为研究热点^[18]。罗静等^[19]运用复杂网络探索挖掘了中医药治疗冠心病不稳定型心绞痛有效病例的用药配伍规律,针对单个有效症状分析核心用药。运用复杂网络分析方法将中药处方构建成为以药物为节点、以配伍联系为边的中药网络,能够发现有效处方和总结用药规律,具有推广价值,但是在分析研究结果时需结合临床实际情况考虑。王丹妮等^[20]利用复杂网络分析方法对李发枝教授在诊治艾滋病方面的宝贵经验进行数据挖掘与分析,王映辉^[21-22]等利用复杂网络分析方法与点式互信息法分析挖掘出某位名老中医在临床复方形成的药物配伍网络,反映了在某些病症条件下,临床处方的配伍经验知识。利用复杂网络分析方法得到的结果具有明确而直接的临床含义,可靠性高。第 1 层核心子药物子网一般解释为共性的核心处方;第 2 层解释为主要药物配伍;第 3 层解释为次要药物配伍。这些药物配伍分别对应样本的核心病机如主要疾病和主要症候、兼证等。同时利用点式互信息法发现有效的药物配伍关系,最终得出治疗某病的有效处方。(2) 多因子降维法 (Multifactor Dimensionality Reduction, MDR) 是近年统计学中发展起来的一种新的分析方法。其中,“因子”即交互作用研究中的变量,“维”是指研究中多因子组合的个数^[23]。该方法主要应用于卫生统计学、流行病学及

遗传学中。MDR 针对诸如疾病状态等分类变量,采用数据降维的思想,从而在有限样本量条件下,分析高维数据之间的交互作用^[24]。传统多因子降维分析算法的基础上,根据中医药数据的特点及分析挖掘需求,开发了新的算法,分析药物-药物,药物-症候,药物-症状与疗效的相关强度,在药物配伍时,如有效率高于只用其中的一味或几味,同时也高于不用这些药物配伍时的疗效,则提示为疗效较好的配伍,由此可分析药物配伍对疗效的影响,从而发现有效的药物配伍关系,最终得出治疗某种疾病的有效处方。(3) 马尔科夫模型可以看作是一种随机过程,模型首先设定一个状态数 N ,连续的 N 个状态数可以决定下一状态^[25]。中医诊疗过程中,是医生根据患者状态,确定干预措施并不断调整的过程。在人工智能领域,部分可观察马尔可夫决策过程 (Partially Observable Markov Decision Process, POMDP) 描述的是当前世界模型部分可行的情况下,智能体如何采取有效的策略来决策的行动序列,部分可感知模型中,要同时考虑状态和动作的不确定性,这种模型能够更加客观地描述真实世界,因此应用十分广泛。结合系统的感知、行动和反馈,是不确定环境下进行序贯决策的一种强大的数学框架,通过每一次决策后状态的变化,得出较优的解决方案,等同于医生根据治疗后患者的反应,不断完善治理方案,形成最终的治疗方案。

4.3 存在问题及建议

4.3.1 电子病历结构化 (1) 术语本身。对名老中医经验进行数据挖掘的基础是临床术语结构化,目前大多数研究者应用的术语是在中医基础研究中获得的,与临床术语有很大的差别。随着中医药信息化、标准化与结构化的发展,术语发展程度的提高,名老中医经验的数据挖掘工作会更加准确。(2) 结构化方法。常用的结构化方法有自然语言识别、数据结构化录入等,由于临床描述的个体差异性、多样性,而又缺乏规范的临床术语,自然语言识别过程中存在错误率高、噪声大等问题,使数据失真。临床采用较多的是根据预先设计的表格,直接将相应的内容录入数据库中,这种方法可

以确保数据录入的准确性,但灵活性差,遇到新的信息后,较难结构化。建议在实践中将自然语言识别法和结构化录入方法结合使用,设计电子病历结构化描述模型集,提供大量模板进行电子病历数据采集,允许医生在模块内自由描述表达临床事件,应用自然语言处理技术对文档进行计算机处理分析。这样既提高了描述性数据的复用性,又减少了临床医生在数据录入方面的时间,这样有利于提高电子病历录入的质量与效率。

4.3.2 名老中医经验数据挖掘 (1) 人才问题。数据挖掘是一个多学科交叉研究领域,且技术开发更新迅速,需要有计算机、统计学等多学科知识,对中医药临床人员来说难度较大。名老中医经验传承最终目的是老专家临床经验和学术思想的总结和传播,数据挖掘方法的应用是有力的辅助手段。建议加强人才队伍建设,大力发展中医药信息学专业,理论与实践相结合,此外还可以通过培训的方法来提高人员的综合素质。(2) 结果解读分析缺乏验证。名老中医诊疗经验主要以隐性知识的形式存在,通过数据挖掘技术,可把部分隐性知识中的内在规律呈现出来。但在某些情况下,有出现解释错误的可能,如挖掘结论与临床实践有明显的出入,或者出现不符合名老中医辨治思想的结果,这样的情况下要和名老中医进行沟通,对数据挖掘流程进行优化和改进。如在运用关联规则对名老中医经验方和药方的挖掘过程中,研究者往往只注重从数据关联的角度提取有效的药物关联,但这不够,还有必要进一步从中医专业角度审视数据内容(即药物)之间的配伍关系^[26]。建议加强与名老中医或有经验的医生的沟通,复核数据挖掘得出的结果,提高可靠性与真实性。

5 结语

在中医电子病历结构化的基础上,采集有效的名老中医经验数据,利用多种分析与展示方法,可以实现对名老中医辨证、辨病、辨症、选方、用药等方面经验的数据挖掘及比较研究,使名老中医经验传承工作成为可能^[27]。但目前中医电子病历仍存

在一定的局限,如中医电子病历只是提供了采集信息的方法和手段,还需临床人员实时采集,严格执行质量控制措施,形成高质量的临床数据,为挖掘分析研究提供基础。

参考文献

- 1 丁琦,郝琦. 中医电子病历在数字化医院的应用与探讨 [J]. 内蒙古中医药, 2013, (23): 51.
- 2 中华人民共和国卫生部. 卫生部关于印发《医药卫生中长期人才发展规划(2011-2020年)》的通知 [Z]. 2011.
- 3 陈硕臻. 基于XML的中医门诊电子病历的研究与应用 [D]. 南昌: 南昌大学, 2012.
- 4 王映辉,姜在阳,闫英杰,等. 基于信息和数据挖掘技术的名老中医临床诊疗经验研究思路 [J]. 世界科学技术, 2005, 7 (1): 98-105.
- 5 李包罗,傅征. 医院管理学(信息管理分册) [M]. 北京: 人民卫生出版社, 2003: 468-493.
- 6 梁兆晖,杨小波,印鉴,等. 中医专科专病临床科研一体化电子病历的构建 [J]. 中国中医药信息杂志, 2007, 14 (8): 101-102.
- 7 高新军. 中医电子病历基本规范5月1日起施行 [J]. 中医药管理杂志, 2010, (5): 468-468.
- 8 廖利平,曾庆明,吴培凯,等. 中药编码系列国家标准发布 [J]. 中医药管理杂志, 2015, 23 (12): 50.
- 9 王映辉,刘保延,姚乃礼,等. 动态结构化名老中医临床诊疗信息采集系统研究开发与应用 [J]. 中国中医药信息杂志, 2008, (2): 106-107.
- 10 R S Dick, E B Steen, D E Dether. The Computer-based Patient Record. An Essential Technology for Health Care [M]. Second ed. Washington DC: National Academy Press, 1997.
- 11 张润顺,王映辉,姚乃礼,等. 名老中医电子病历中病史动态结构化数据录入规范 [J]. 中国中医药信息杂志, 2007, (3): 100-101.
- 12 金新政. 现代医院信息系统 [M]. 北京: 人民卫生出版社, 2009: 517-521.
- 13 黄源,杨铭,陈佳蕾,等. 基于复杂网络的中医药治疗慢性阻塞性肺病的用药配伍特点 [J]. 中国试验方剂学杂志, 2012, 18 (15): 7-11.
- 14 黄薇. 基于数据仓库的数学模型库架构的研究 [D]. 武汉: 武汉理工大学, 2006.

- 15 孙亚男, 宁士勇, 鲁明羽, 等. 中医诊疗决策支持系统中的 OLAP 与 DM 融合 [J]. 计算机工程, 2006, 32 (9): 251-252.
- 16 陶敏. 电子病历系统中处方用药推荐模块的设计与实现 [D]. 杭州: 浙江大学, 2015.
- 17 吴丽, 周良. 数据挖掘方法在中医药领域的应用浅析 [J]. 医学信息, 2009, (5): 576-578.
- 18 Albert-László Barabási, Natali Gulbahce, Joseph Loscalzo. Network Medicine: a network-based approach to human disease [J]. Nat Rev Genet, 2011, 12 (1): 56-68.
- 19 罗静, 徐浩, 周雪忠, 等. 基于复杂网络的不稳定型心绞痛中药配伍应用规律研究 [J]. 中国中西医结合杂志, 2014, (12): 1420-1424.
- 20 王丹妮, 李真, 徐立然, 等. 李发枝治疗艾滋病带状疱疹及其后遗神经痛的配伍精要 [J]. 中国中药杂志, 2013, (15): 2497-2500.
- 21 王映辉, 周雪忠, 张润顺, 等. 利用复杂网络与点式互信息法分析挖掘名老中医用药经验研究 [J]. 中国数字医学, 2011, (4): 76-80.
- 22 周雪忠, 刘保延, 王映辉, 等. 复方药物配伍的复杂网络方法研究 [J]. 中国中医药信息杂志, 2008, (11): 98-100.
- 23 彭艳. 调神法在功能性胃肠病治疗中的应用临床研究 [D]. 北京: 北京中医药大学, 2015.
- 24 王承龙, 张大武, 王培利, 等. 急性冠脉综合征中医证候要素的多因子降维分析 [C]. 乌鲁木齐: 世界中医药学会联合会心血管病专业委员会第五届学术大会, 2011: 165-171.
- 25 程显毅, 朱倩, 王进. 中文信息抽取原理及应用 [M]. 北京: 科学出版社, 2010: 41-42.
- 26 古求知, 老膺荣, 范宇鹏, 等. 名老中医临床医案数据挖掘研究思路探讨 [J]. 中国中医药信息杂志, 2013, (4): 10-11.
- 27 王映辉, 张润顺, 周雪忠, 等. 名老中医经验共性规律及个性差异比较研究 [J]. 世界科学技术 (中医药现代化), 2009, 11 (6): 793-799.

2016年《医学信息学杂志》征订启事

《医学信息学杂志》是国内医学信息领域创刊最早的医学信息学方面的国家级期刊。主管: 国家卫生和计划生育委员会; 主办: 中国医学科学院; 承办: 中国医学科学院医学信息研究所。中国科技核心期刊 (中国科技论文统计源期刊), RCCSE 中国核心学术期刊 (武汉大学中国科学评价研究中心, Research Center for Chinese Science Evaluation), 美国《化学文摘》、《乌利希期刊指南》及 WHO 西太区医学索引 (WPRIM) 收录, 并收录于国内 3 大数据库。主要栏目: 专论, 医学信息技术, 医学信息研究, 医学信息组织与利用, 医学信息教育, 动态等。读者对象: 医学信息领域专家学者、管理者、实践者, 高等院校相关专业的师生及广大医教研人员。

2016 年《医学信息学杂志》国内外公开发行, 每册定价: 15 元 (月刊), 全年 180 元。邮发代号: 2-664, 全国各地邮局均可订阅。也可到编辑部订购: 北京市朝阳区雅宝路 3 号 (100020) 医科院信息所《医学信息学杂志》编辑部; 电话: 010-52328673, 52328674, 52328671。

《医学信息学杂志》编辑部