

**摘要：**伴随中医医案文献的逐渐增加,怎样对大量的医案予以整理及挖掘,因此找出其中潜藏的知识及规律,即为现阶段医案研究的侧重点。而 Data mining (数据挖掘)的广泛应用,为中医医案的发掘铺平了道路。Data mining 即为从海量、残缺、模糊的数据内提炼其内部隐性信息及知识的一个过程,也是发现知识的一个基本过程。文章将以数据挖掘技术在中医医案研究中的应用作为切入点,在此基础上予以深入的探究,相关内容如下所述。

**关键词：**数据挖掘技术；中医；医案研究；应用

## 1. 中医医案研究中 Data mining 的形式

### 1.1. Frequency 统计

Frequency 统计方法相对简单,且对医案的发掘意义深远,能够找到大量的有价值的结论。张明等统计了四千例当代名医医案的舌象与脉象分布 Frequency 状态,通过研究发现很多以往不为人们注意的新知识,如嫩舌亦为主实证等。梁国栋等人对高血压及冠心病医案予以统计分析发现,Data mining 结果可以从根本体现医案对上述疾病的诊治特色及经验。邓明华统计了五百个老中医痹证医案,进而构建了痹证医案参数库,通过 Frequency 分析对证候与成因以及发病时间等予以统计。分析结果和传统中医理论相统一,匹配于过去中医对痹证的研究结果。一些研究人员构建了古代情志医案中心身疾病关联式数据库管理系统,通过 Frequency 分析措施,对 Frequency 超过五十次的五十四种药物予以分析,总结了心身疾病用药的基本趋势。

### 1.2. 逻辑斯蒂回归分析

逻辑斯蒂回归分析方法是明确超过两种变数间相互依存定量关系的一种统计分析形式。刘玉民构建了历代医案参数库,择取非条件逻辑斯蒂多元递进回归筛选外感病因的症状,在此基础上提供各症状对诊断外感病因的有效性及特异性。分析表明,外感病因的证候成因主要为寒邪、风邪、热邪、燥邪、湿邪、外毒、疠气以及疔邪。同时依附于此数据库,择取非条件逻辑斯蒂多元逐步回归方法,构建中医证候与其临床特性的知识库,同时设计“中医临床辅助诊断系统”软件,此软件仅需输入病例的四诊信息,即可提供辨证结论。李志华通过流行病学调查有关胆石病的临床资料,对三百例胆石病病例临床症状及临床诊断总结的中医病因病机构建了逻辑斯蒂回归模型,对中医病因病机予以了统计研究,同时将研究结果与临床诊断结果予以对比显示,流行病学调查所择取的中医病因病机诊断指标具有较强的合理性,所获取的三百例胆石病病例的主要中医病因病机分布也是拥有较高可信度的分布。

### 1.3. Association Rule 挖掘

在数据挖掘方面, Association Rule 即为应用相对广泛的研究方向。多个变量取值间工程特定的规律性, 例如“同时发生”及“从一个目标能够推出另一个目标”即称为 Association Rule。Association Rule 挖掘是利用关联分析寻求数据库内隐性的知识, 通过这些隐性知识能够依附于已知情况对未知问题予以有效的推测。张玉梅等择取 Association Rule 分析名中医四百五十例哮喘医案的成因、证候以及四诊信息之间的关系, 病因、证候、四诊信息和用药的内质联系, 通过研究发现中医医案中的四诊信息、用药、病因以及证候间存在较强的正相关, 能够利用 Association Rule 找到内在的规律。姜建国以古代心悸医案为基础构建数据库, 予以 Frequency 及 Association Rule 分析, 进而挖掘出治疗心悸的有效方药为陈皮、半夏、人参、白术、茯苓、甘草、白芍、当归、远志、酸枣仁以及茯神。刘明等出版发行的名中医牙痛医案, 择取 Association Rule 与链接分析举措予以统计, 研究表明牙痛治疗包括下述基础组方, 以赤芍、浙贝母、连翘、蒲公英、金银花以及菊花为第一基础组方, 知母、石膏、生地黄、牡丹皮以及牛膝为第二基础组方, 上述两组方和牙痛的中医治疗理论相契合。王红丽以六百例名医病案为目标, 依附于频繁模式树的算法, 同时对证型-症状、症状-药物, 证型-药物间的 Association Rule 予以了挖掘。经研究显示, 所挖掘出的规则都可以得到科学的解释, 同时具有较强的价值。闫丕庚等人通过 Frequency 分析与 Association Rule 挖掘技术, 提出了一个萃取古代各家名医治疗各种病症的有效方法, 在此基础上通过中风病症的验证, 证明此模式具有较强的有效性。通过分析结果挖掘出古代名医治疗中风较为多见且有效的几类中药。

而本人认为:通过古代中医医案的 Data mining, 能够将古代名医的治疗经验予以整体全面的提炼。提炼出的知识不仅具有较强的有效性, 同时还具备一定的临床价值。

#### 1.4. Cluster analysis 挖掘

Cluster analysis 即为通过物以类聚的基本原则, 将无序的参数分门别类, 这在一定程度上利于对大量数据内的规则进行认识, Cluster analysis 隶属数理统计, 能够把观察目标依附于相关特征进行分类, 目前已被应用于医学分类问题。在中医医案研究中, Cluster analysis 可以规避分类环节中的主观因素, 进而客观的表现研究目标, 同时从中发现潜在的规律。宋晓晨等统计了古代二十六部现存医案中的津液亏损病案五百例, 择取 R 型系统 Cluster analysis, 总结津液亏损中各亚型的基本用药, 在此基础上明确了我国古代津液理论的核心特性。张培超等搜集了古代医家各学派的六十例医案资料, 在此基础上通过系统 Cluster analysis 对变量予以研究, 进而挖掘治疗消渴病的用药有效性, 结果显示知母、生地、甘草、麦冬、佩兰叶、肉桂、茯苓、山药以及熟地即为治疗消渴病的有效中药, 对临床实践具有一定的影响。

#### 1.5. ANNs 挖掘

ANNs 即为模拟人脑的一种计算形式, ANNs 是通过非处理模块所构建的非线性大规模自适应机制, 通过近似于人脑神经网络的并行处理体系予以信息的整体计算。李旭东等人把 ANNs 和粗糙集理论互溶, 通过粗糙集理论对 ANNs 予以预处理, 进而计算主体属性, 在此基础上应用至中医类风湿病的诊断建模。仿真结果显示,

通过粗糙集理论设计 ANNs, 从根本深化了 ANNs 的学习能力, 同时在实践中获取了良好的效果。

## 2. 现阶段中医医案研究中数据挖掘技术所存在的问题及发展

通过上述研究我们可以发现, 大量的研究人员都逐渐侧重于 Data mining 对中医医案的发掘及利用, 目前已积累了医案 Data mining 的有效方法, 同时发现 Data mining 可以从根本找到医案中潜藏的基本规律, 因此指导及深化中医理论及实践所存在的个别问题, 不过我们现阶段还没有系统的予以医案的挖掘。Data mining 在医案研究中应用还具有下述问题: (1) 研究的主要内容已然以疾病的诊治用药分析为主, 而基础研究则少之又少, 不过近年来有逐渐增加之势; (2) 一些研究结果虽然和临床数据相契合, 不过依然有大量结论和临床实践结果不符; (3) 医案挖掘方法已然以统计学方法为基点, 繁琐的挖掘方法在医案中应用较少; (4) 从事医案挖掘的研究工作者较少, 而且具有较强的集中性。

在未来医案挖掘环节, 我们需侧重于下述两点内容: (1) 对医案的特性予以更为整体的研究, 择取有针对性的挖掘方法, 进而促使挖掘结果可以更为直接的体现医案中潜藏的规律, 为深化中医临床实践水平奠定良好的基础; (2) 侧重于开发计算机技术中的文本挖掘, 更为全面的使用 Cluster analysis、ANNs 等挖掘方法。

## 3. 总结

综上所述, Frequency 统计方法相对简单, 且对医案的发掘意义深远, 能够找到大量的有价值的结论。逻辑斯蒂回归分析方法是明确超过两种变数间相互依存定量关系的一种统计分析形式。多个变量取值间工程特定的规律性, 例如“同时发生”及“从一个目标能够推出另一个目标”即称为 Association Rule。而 Cluster analysis 即为通过物以类聚的基本原则, 将无序的参数分门别类, 这在一定程度上利于对大量数据内的规则进行认识, Cluster analysis 隶属数理统计, 能够把观察目标依附于相关特征进行分类, 目前已被应用于医学分类问题。ANNs 则是模拟人脑的一种计算形式, ANNs 是通过非处理模块所构建的非线性大规模自适应机制, 通过近似于人脑神经网络的并行处理体系予以信息的整体计算。上述几类数据挖掘技术目前均被应用于中医医案研究中, 而我们需要在此基础上深化数据挖掘的形式及有效性, 进而为中医医案研究铺平道路。

总而言之, 数据挖掘是相对前沿的技术, 虽然现阶段在中医医案研究应用中处于发展阶段, 不过随着技术的完善及方法的改进, 其应用面势必更为广泛。

### 参考文献:

- [1]. 马莉; 章玲; 数据挖掘技术在图书馆领域的应用 [A]; 新环境下图书馆建设与发展——第六届中国社区和乡镇图书馆发展战略研讨会征文集 (下册) [C]; 2016 年
- [2]. 郭云峰; 张集祥; 一种基于位向量的关联规则挖掘算法 [A]; 2015' 中国信息技术与应用学术论坛论文集 (一) [C]; 2015 年

- [3]. 周雪忠;刘保延;姚乃礼;李平;王映辉;王春山;于彤;穆钊;徐兴良;中医临床数据仓库的研究及构建[J];国际中医中药杂志;2014年06期
- [4]. 姚小磊;彭清华;数据挖掘技术在中医眼科应用的设想[A];中华中医药学会第七次眼科学术交流会论文汇编[C];2015年
- [5]. 杨钧;刘建平;张颖;吴磊;基于数据挖掘技术的中医药科研方法的研究[A];中华中医药学会中医药传承创新与发展研讨会专辑[C];2016年
- [6]. 刘明;韩娜娜;杨建业;马力;基于数据挖掘技术的下肢深静脉血栓形成常用中药分析[A];中华中医药学会周围血管病分会2014年学术大会论文集[C];2014年
- [7]. 王雪峰;董丹;梁茂新;张斌;张明卫;数据挖掘技术在小儿肺炎中医临床疗效评价指标研究中应用的思路与方法[A];第24届全国中医儿科学术研讨会、中医药高等教育儿科教学研讨会、儿科名中医讲习班论文汇编[C];2014年
- [8]. 沈波;李国春;吴勉华;王均琴;周仲瑛医案瘀热病机理论的数据挖掘方法及思路[J];山东中医药大学学报;2013年05期
- [9]. 郭超峰;施学丽;黄克南;邓巍;关联规则数据挖掘技术在中医药研究中的应用概况[J];广西中医学院学报;2015年04期
- [10]. 张天嵩;张素;李秀娟;张伟伟;潘宝峰;杨克敏;治疗肺纤维化中药复方用药规律的数据挖掘[A];第十一次全国中西医结合防治呼吸系统疾病学术研讨会论文集[C];2014年