

数据挖掘技术及其在医疗质量管理中的应用^{*}

郑娟 许建强

(徐州医科大学 徐州 221004)

〔摘要〕 介绍数据挖掘概念、基本流程及任务, 阐述数据挖掘在医疗质量管理中的应用, 包括医疗质量指标变化趋势分析、基于多维数据模型的数据挖掘分析等方面, 指出将数据挖掘技术应用于医疗质量管理具有广阔的发展前景和较高的应用价值。

〔关键词〕 数据挖掘; 医疗质量; 单病种医疗质量

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2018.03.015

Data Mining Technology and Its Application in Medical Quality Management ZHENG Juan, XU Jian-qiang, Xuzhou Medical University, Xuzhou 221004, China

〔Abstract〕 The paper introduces the conception, basic process and tasks of data mining technology, expounds on the application of data mining in medical quality management, including analysis of the variation trend of medical quality indicators, data mining based on multidimensional data model and other aspects, and points out that the application of data mining technology in medical quality management boasts broad prospects for development and high application value.

〔Keywords〕 Data mining; Medical quality; Medical quality of single disease

1 引言

我国新医保政策的实施、医院之间优劣评比方式的改变等使医院之间的竞争压力不断增加。医疗质量问题受到医院各界的广泛关注, 如何提高医疗质量成为各医疗机构亟待解决的重点问题。提高医疗质量可以通过提高疾病诊疗和医院管理两个方面

实现。目前国内多数医院都使用各自的医院信息系统进行管理工作, 其收集的病案数据非常适合运用计算机技术进行分析和研究。近年来随着医疗数据挖掘的兴起, 使用适当的数据挖掘技术开展医疗质量研究的时机已经成熟, 在理论和技术上均具有极高的可行性^[1]。

2 数据挖掘概述

2.1 概念

关于数据挖掘 (Data Mining, DM) 的概念, 不同学者给出不同的解释, 目前应用最广泛的是 Fayyad^[2] 等人在 1996 年提出的: 数据挖掘是从大量的、不完全的、有噪声的、模糊的、随机的数据

〔修回日期〕 2018-01-18

〔作者简介〕 郑娟, 讲师, 发表论文 8 篇。

〔基金项目〕 江苏高校哲学社会科学基金项目 (项目编号: 2017SJB0985); 国家自然科学基金项目 (项目编号 71673170)。

中，提取出隐含在其中的、有效的、新颖的、潜在有用的且最终可理解的模式的非平凡过程。作为一门新兴的边缘学科，数据挖掘涉及多个专门领域，包括信息科学、数据库技术、可视化、统计学、机器学习以及其他学科等。数据挖掘涉及领域^[3]，见图 1。

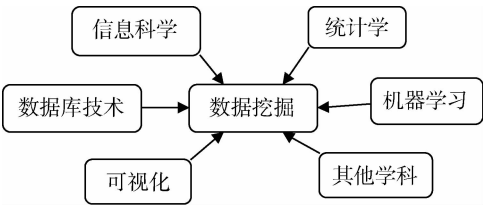


图 1 数据挖掘涉及领域

2.2 基本流程及任务

一个完整的数据挖掘流程主要包括数据收集、数据整理（清理和转换）、数据挖掘、挖掘结果评价和分析决策^[4]。数据挖掘的基本流程，见图 2。数据挖掘的任务根据可解决问题的性质归纳为：概念描述、分类、预测、划分和链接分析等。数据挖掘方法及相应软件，见表 1。

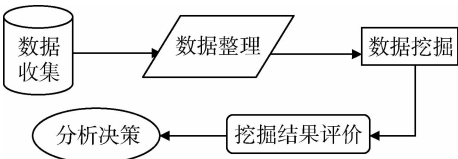


图 2 数据挖掘基本流程

表 1 数据挖掘方法及相应软件^[5]

功能	数据挖掘技术	数据挖掘软件
概念描述：描述变量以及变量之间的关系	查询技术、可视化技术、简易交叉报表等	SGL（SQL）Enterprise Miner
分类：根据预测变量，将一个新实例分类给已经类别	判别分析、逻辑回归、关联规则、神经网络等	Clementine Darwin
预测：用一个或多个自变量预测一个未知的因变量	多元线性回归、逻辑回归、关联规则、决策树、神经网络、遗传算法等	Intelligent Miner Business Miner
划分：在依据研究对象相关的变量集合基础上对目标分组	决策树、聚类分析、神经网络等	-
链接分析：依据隐含信息来确认相关的购买模式	如关联规则法	-

3 数据挖掘在医疗质量管理中的应用

3.1 应用概述

医疗数据挖掘技术随着计算机技术的普遍应用应运而生^[6]。数据挖掘在医学卫生领域主要应用于遗传学、疾病辅助诊断、疾病预测预后分析、药物开发、医学影像、医院信息系统等。如有研究者将数据挖掘方法运用于 DNA 序列分类以及基因组测序数据分析，发现非线性相关统计法（如神经网络、AMI 等）有重大的应用价值^[7]；Young-MoonChee 等学者使用韩国医疗保险的数据库，发现决策树技术可以为高血压管理以及人群健康评估提供相应的决策信息^[8]。数据挖掘技术从产生至今仅有 10 年的历史，其在各行业和领域如商业、工程、

电信等都应用广泛并成就显著，但在医学领域尚处于起步阶段。医疗系统及医学技术的特性使医学数据挖掘具有明显的独特性，未来数据挖掘技术在医疗科学领域将有广阔的应用前景。

3.2 医院医疗质量管理难点

目前我国各医疗机构在医疗质量的管理上还存在很多困难。首先在效率方面，各医疗机构往往安排专门的统计人员进行医疗质量指标分析，人力、物力等资源消耗严重，且专业人员在进行分析时由于力度有限，往往无法找出问题根源，从而找不到解决问题的方法。很多情况下数据分析人员在提交报告时经常对庞大的数据进行再加工，导致效率低下。其次在数据质量方面，医院有很多医疗质量的数据是记录在不同的系统中，仅根据病案首页无法

得知患者具体的病情信息，并且医院中很多的原始数据在记录过程中填写不全，甚至填写有误，数据质量问题严重。

3.3 医疗质量管理中的数据挖掘分析

3.3.1 医疗质量指标变化趋势 在我国大多数医院采用的医疗质量指标是 ISO9000 体系下的评价指标，如平均住院天数、住院人数、住院患者死亡率等。应用数据挖掘技术可以分析医疗质量指标的变化趋势，预测未来的发展走势，如付学宇^[1]采用数据挖掘技术中的贝叶斯网络模型分析医院医疗质量指标（住院患者死亡率、围手术期死亡率、非计划再返手术室率、出院 31 天内再入院率等）在 2004 -2007 年间的变化趋势。有研究^[9]指出可利用数据挖掘技术中的环基比和定基比来分析医疗质量指标的增长和发展速度，并且可用曲线拟合来预测未来的发展趋势。

3.3.2 医疗质量指标影响因素 有助于医院管理者更好地保证和提高医疗质量，如平均住院天数可以全面衡量医院医疗质量以及管理水平。有研究指出利用数据挖掘技术中的灰色关联分析方法从患者住院日的时间长短、医疗质量及床位利用等方面综合评价其对平均住院日的影响，发现影响住院天数的主要因素有超长住院平均住院日、2 日内住院人次、手术人次、床位利用指数、转科人次等^[10]。

3.3.3 医疗效果 医疗质量管理的一项重要内容是对疾病进行预后情况的评价。可以利用患者主记录数据仓库对其进行预后情况的分类，如可通过分析死亡率、治愈率、治愈好转率等对出院患者、死亡患者和转科患者情况等进行分析并制定相关治疗方案。郑西川等^[11]使用决策树分析方法对某医院 8 405 例出院患者进行数据挖掘分析，得出影响患者死亡率的几个重要因素，提出医疗质量持续改进的指导原则和方法。

3.3.4 医疗质量综合评价 应用数据挖掘技术综合评价医疗质量可以促进医疗质量的提高。有研究指出医疗质量的综合评价可应用同一评价指标、模式、权重系数、标化方法、分类方法等对医院业务报表以及医疗和护理工作资料进行分析和评价^[11]。

3.4 利用数据挖掘技术加强医疗质量管理

3.4.1 数据收集和整理（清理、转换） 在进行数据挖掘前，首先要将所需要的数据收集到某一特定类型的文件或数据库中。多个系统数据进行整合时，可根据定义标准化数据进行整合。对于无法归类或无效的数据，可根据标准格式剔除样本数据集中存在的冗余和不相关信息。

3.4.2 基于多维模型的数据挖掘 多维模型主要包括“摘要”、“维度”和“信息”。医疗质量管理过程中常将相关的人、事、物等字段纳入到多维模型中。数据挖掘的多维分析（包括横向的交叉分析和纵向的潜入分析）体现其强大的分析功能。要明确问题，可以在宏观层面发现问题，也可通过横向交叉分析以及纵向潜入分析来寻找问题的根源所在^[12]。

3.4.3 单病种医疗质量分析举例 应用数据挖掘技术和多维模型分析可以了解并分析单病种医疗质量问题。从医院前 10 位多发病种中找出死亡率最高的几类病种，应用数据挖掘技术分析死亡率高的原因，如患者自身原因、诊断和治疗方面原因以及医生原因等。从医疗行为所涉及的人、事、物等关系中找出与死亡率指标相关方面，从而分析出问题的根源，规范医疗机构的医疗行为，提高医院运行效率。单病种死亡率等医疗质量指标分析，见图 3。

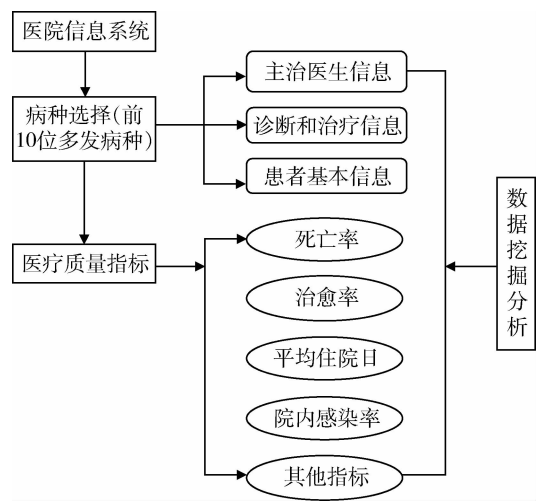


图 3 单病种死亡率等医疗质量指标分析

3.5 应用前景

目前各医疗机构对相关数据的统计只局限于采用普通的统计学分析方法,很少应用数据挖掘技术。数据挖掘技术与传统的统计学方法相比,其具有较好的前沿性,可更全面、系统、准确地分析医院医疗质量问题,更快地找出存在的问题,从而帮助医院管理者找出问题,分析根源,找到相应解决策略。在医疗质量管理领域应用数据挖掘技术可以扩大该技术的应用范围,也间接说明相关技术在医疗质量管理问题上的解决能力较高,具有较高的理论和实践应用价值。

4 结语

尽管在医疗质量领域应用数据挖掘技术具有很好的发展前景和应用价值,但在实际操作中非常复杂和困难。对医疗数据的挖掘需要相关学科、技术及医学相融合,是从全部的医学信息数据库中不断提炼和提取相关知识的过程,对于医疗服务的整体决策及其科学化等具有重要意义。对医疗数据的挖掘是一项复杂的工程,其特点主要包括以下几点:首先,挖掘对象具有广泛性,主要是因为医学数据具有复杂性。其次,数据挖掘要求算法具有高效性,因为医学数据具有繁杂冗长的特点。此外,对于医学知识提取所用的决策分析需要有更高的准确性。目前各机构现有的医学信息数据库并不完善,很大程度上要求计算机、统计学分析以及数学算法等各部门学科的相互融合,那么在相关知识提取、算法及信息提取关键技术可能取得更大的突破。随着数据挖掘方法的日益发展和不断改进,未来医学相关领域该技术的应用会更加深入和广泛,其发展

必将会为人们带来更大的经济效益和社会效益。

参考文献

- 1 付学宇. 基于贝叶斯网络模型的医疗质量挖掘研究 [D]. 天津:天津大学,2009.
- 2 Fayyad U M, Piatetsky - Shapiro G, Smyth P. The KDD Process for Extracting Useful Knowledge from Volumes of Data [J]. Communication of the ACM, 1996, 39 (11): 27 - 34.
- 3 王红影. 数据挖掘在医疗费用分析中的应用研究 [D]. 大连:大连海事大学,2008.
- 4 李鲁宁. 基于数据挖掘的中医胃炎辨证论治规律研究 [D]. 沈阳:沈阳建筑大学,2012.
- 5 Sandra S. Liu, Jie Chen. Using Data Mining to Segment Healthcare Markets From Patients' Preference Perspectives [J]. International Journal of Health Care Quality Assurance, 2009, 22 (2): 117 - 134.
- 6 Prather J C, Lobach D F, Goodwin L K, et al. Medical data mining: knowledge discovery in a clinical data warehouse [J]. A Conference of the American Medical Informatics Association, 1997, 4 (1): 101 - 105.
- 7 张吉英. 数据挖掘技术在针灸处方配伍中的应用研究 [D]. 济南:山东中医药大学,2014.
- 8 张继. 基于数据挖掘技术的妇科肿瘤病人住院费用研究 [D]. 郑州:郑州大学,2011.
- 9 陈郁韩. 数据挖掘技术在医院管理中的应用 [J]. 中国医学装备, 2014, 11 (1): 62 - 65.
- 10 赵莉丽. 基于灰色关联法分析平均住院日影响因素 [J]. 中国卫生统计, 2010, 27 (2): 194 - 195.
- 11 郑西川, 秦环龙, 厉永灏. 基于数据挖掘和决策支持的医疗质量分析 [J]. 中国医院管理, 2006, 26 (4): 22 - 24.
- 12 余元龙, 郭茜. 数据分析与挖掘技术在医疗质量管理中的应用 [J]. 医学信息学杂志, 2011, 32 (1): 34 - 37.