

数据挖掘在医疗中的应用研究

李雄伟

(武汉科技大学 信息科学与工程学院, 湖北 武汉 430081)

摘要: 随着医院信息化的进程加快, 医院信息的积累越来越丰富, 为医学研究提供了方便和研究价值。以出院患者的随访记录建立数据库, 依靠数据挖掘技术, 筛选出有价值的信息, 建立一个能够为以后的治疗提供有价值的参考作用的医院医疗方案库。从而形成一个对患者更负责任的体系, 完善医院服务, 提高医院诊断就医效率。

关键词: 数据挖掘; 随访; 医疗方案库

中图分类号: TP31

文献标识码: B

Application research of data mining in medical treatment

LI Xiong Wei

(Department of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan430081,China)

Abstract: Currently, with the rapid development of informational construction of hospital and medical treatment in our country, while mostly hospitals haven't built more perfect and effective information database in medical treatment. The paper demonstrates the significance and effect of data mining in large number of information, meanwhile bettering medical treatment solution database, discussing the problems in data mining of medical treatment. Applying data mining technology in medical treatment can standardize and better the construction of hospital service. In this way, it also improves the effect and efficiency.

Key words: data mining; medical treatment; MTSD

1 数据挖掘

1.1 数据挖掘概念

数据挖掘^[1-2]是指从数据库中提取出隐含的人们事先未知的、潜在的有用信息和知识, 所提取的知识可以表示为概念规则、规律、模式等形式。它具备有效性、新颖性、潜在有用性及最终可理解性, 即发现的知识可接受、可理解、可运用, 最好能用自然语言表达发现的结果。处理的原始数据可以是结构化的, 如关系数据库中的数据; 也可以是半结构化的, 如文本、图形、图像数据; 甚至是分布在网络上的 Web 数据和异构型数据。获取的知识可用于信息管理、查询优化、决策支持、过程控制等, 还可以用于数据自身的维护。因此, 数据挖掘是一个介于统计学、模式识别、人工智能、机器学习、数据库技术以及高性能并行计算等领域的交叉新学科, 已在经济、商业、金融、天文等行业得到了成功的应用,

在国际上掀起了一股空前的研究热潮即所谓的“数据淘金潮”。我国学者在这个领域也已开展了很多研究, 主要来自计算机科学及相关领域^[1-3]。

1.2 数据挖掘过程

数据挖掘过程主要由 3 个阶段组成: 数据准备、数据开采、结果表达和解释。

(1) 数据准备

这个阶段又可以进一步分成 3 个子步骤: 数据集成、数据选择、数据预处理^[2-4]。数据集成将多文件或多个数据库运行环境中的数据进行合并处理, 解决语义模糊性、处理数据中的遗漏和清洗脏数据等; 数据选择的目的是辨别出需要分析的数据集合, 缩小处理范围, 提高数据挖掘的质量; 数据预处理是为了克服目前数据挖掘工具的局限性。

(2) 数据开采

在这个阶段进行实际的数据开采操作，主要包括：

① 如何产生假设，是让数据挖掘系统为用户产生假设，还是用户自己对于数据仓库中可能包含的知识提出假设。前一种称为发现型(Discovery-Driven)的数据挖掘，后一种称为验证型(Verification)的数据挖掘；

② 选择合适的工具；

③ 挖掘知识的操作；

④ 发现知识。

(3) 结果的表达和解释

根据最终用户的决策目的对提取的信息进行分析，把最有价值的信息区分出来，并且通过决策支持工具提交给决策者。因此，这一步骤的意义不仅是把结果表达出来(例如采用信息可视化方法)，还要对信息进行过滤处理。如果不能令决策者满意，则需要重复以上数据挖掘的过程。数据挖掘过程如图1所示。

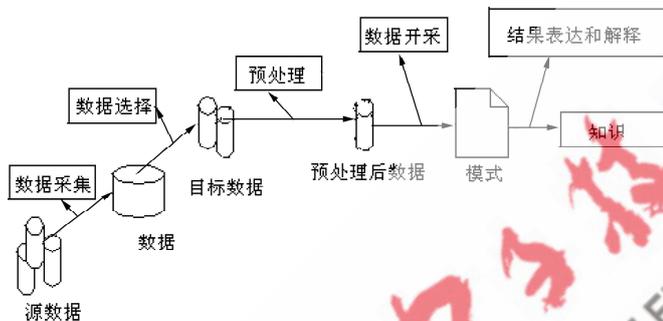


图1 数据挖掘过程

2 医院随访现状

随访是医疗卫生部门为定期或不定期了解门诊或出院病人在院期间医疗处理的愈后、康复情况及远期疗效通常采用的一种工作手段。随着医疗技术的日新月异，随访工作的开展与否已成为衡量一家医院医疗管理是否先进的重要指标之一。随访，作为医疗及医学信息应用的主要内容，为医院的医疗、教学、科研提供了不可多得的宝贵资料。患者随访服务将成为医院在今后市场竞争中的重要因素，通过为患者提供有效的服务，设法吸引患者成为医院经营的焦点之一。然而，由于各种原因，患者随访依然是当前医院管理中的薄弱环节之一。本文建立的随访数据库，提供全面的患者随访管理，完成以患者为中心的服务转变，以为“完善医院服务，提高病人满意度，降低病人流失率”最终目标，全面提高医院的服务效果。

3 数据挖掘在医院医疗中的应用

3.1 数据挖掘在随访信息库中的应用

医院作为一个庞大的社会服务机构，在过去若干年的时间里积累了海量的、以不同形式存储的数据资料，但是如果对其中的每个数据都仔细处理是不可能的，也没有实效和用处。本文以电子病历系统的数据为原始

库，接收电子病历系统数据，通过数据清理，建立一个随访信息库，用于对出院病人的跟踪调查。在随访信息库的基础之上，进一步进行有价值的信息挖掘，建立一个医院治疗方案库，以便总结治疗经验，对以后医疗方案的实施有重要的指导意义^[6-8]。

3.2 原始电子病例资料库

电子病历包含病人的基本资料、医生信息、检查情况、诊断情况、医嘱、手术情况、护理信息等数据信息，随访病人资料主要有病人的基本资料、病人出院情况记录、出院后的用药记录、病人的复诊情况等。

病人的基本资料包括姓名、年龄、性别、身高、体重、血型、医保费用类别等；医生信息包括科室、姓名、职称、诊疗时间等；检查情况包括主要症状的描述；检查报告书包括放射、超声波、心电图、内窥镜等；医技检查诊断情况包括诊断依据、门诊诊断、入院诊断、出院诊断等；医嘱包括日期、主治医师姓名、临时医嘱、长期医嘱等；手术情况包括手术时间、参与人员、手术方案、手术记录等；护理信息包括护士姓名、护理日期、护理计划、床位设置、护理情况等。

将以上信息纳入随访信息数据库中，是对病案信息资源的二次开发。随着医院医疗工作的开展，随访工作量日益加大，设计完善的随访系统是科学、准确地做好随访工作的前提和保障。根据原始电子病例资料库，运用适当的算法进行信息的清理，把对病患随访的无效的数据清理掉，然后通过各种随访手段进行随访，建立起一个随访病人资料库。随访库建立流程如图2所示。

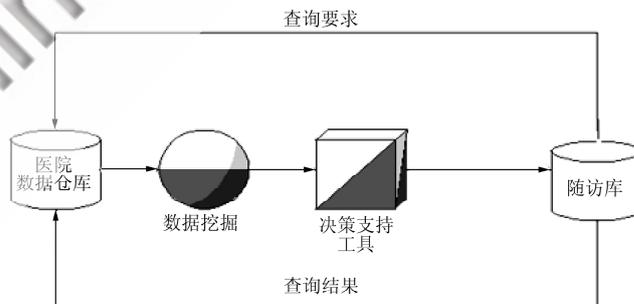


图2 随访库建立流程

3.3 随访病人资料

随访病人的基本资料应该与医院中的电子病例数据库中病人的基本信息一致。病人的出院情况记录疾病的种类、治疗时间、出院时间、主治医师以及随访人员编号。出院后的用药记录包括用药种类、时间以及用药效果等，病人的复诊记录包括复诊时间、复诊主诊医生、检查情况、回复情况等。当然，具体到每一个病例，不一定会包括所有的内容，可能会涉及到新的内容，这就需要具体情况具体分析。

医院信息系统蕴藏着海量数据，全部调用既不可

(下转第82页)

(上接第79页)

能,也无功效。所以通过现代的数据挖掘技术将随访库和原始数据库结合,进行数据分析、清理和挖掘,得出具有代表性和指导意义的病例,并将其添加到治疗方案当中,逐渐形成一个大规模的、有价值的医院治疗指导信息库,使得以后其他病例有所借鉴,最终形成信息的共享。基于医院数据仓库的数据挖掘功能如图3所示。

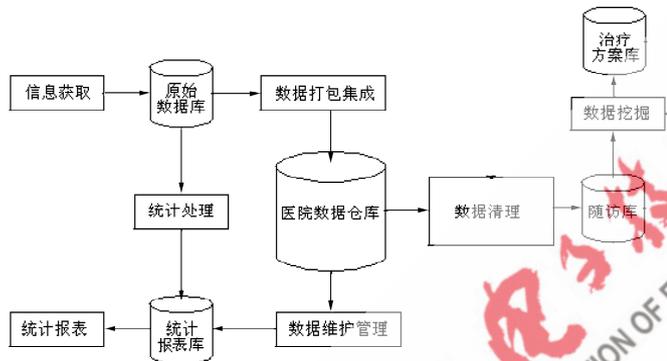


图3 基于医院数据仓库的数据挖掘

通过医院原始的信息库建立一个随访库,对患者进行跟踪随访,记录出院后的数据,收集病人的随访数据,然后通过随访数据库进行数据挖掘,建立起一个有价值和指导意义的医院医疗库。医生通过查阅医院医疗库中信息资料,可以了解某种疾病或某种手术方

案以及以往病人的治疗情况,并结合自己的临床经验,对治疗方案进行筛选,最终提高医疗疗效,达到提高医疗质量的目的。

参考文献

- [1] 范明,孟小峰.数据挖掘概念与技术[M].北京:机械工业出版社,2001.
- [2] 康晓东.基于数据仓库的数据挖掘技术[M].北京:机械工业出版社,2004.
- [3] 普通高等学校本科专业目录和专业介绍[M].北京:高等教育出版社,1997.
- [4] 范明,范宏建.数据挖掘导论[M].北京:人民邮电出版社,2006.
- [5] JIA Wei Han, MICHELINE K. Data mining concepts and techniques [M]. China Machine Press, 2005.
- [6] LECH P, SHUSAKU T, TSAU Y. Rough set method and applications: new developments in knowledge discovery in information systems [M]. Physica-Verlag, 2000.
- [7] 谷建军,王洪国,丁艳辉.粗糙理论及其在数据归约中的应用[J].信息技术与信息,2006(3).
- [8] 伞冶,叶玉玲.粗糙理论及其在智能系统中的应用[J].智能系统学报,2007,4(2).

(收稿日期:2009-03-17)