

基于数据仓库的医院数据分析平台建设与应用^{*}

马少锋 温明锋 陈健超

(广东省江门市五邑中医院 江门 529000)

〔摘要〕 分析传统医院信息系统存在的不足, 建立医院数据挖掘分析平台, 介绍平台架构, 搭建院领导分析、抗菌药物监管、医院质量分析、单病种分析、医保控费等主题的数据挖掘模型, 在此基础上提出提高医院服务水平的智能解决方案。

〔关键词〕 数据挖掘; 数据仓库; 数据分析; 分析平台

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2017.07.004

Construction and Application of Data Analysis Platform of the Hospital Based on Data Warehouse MA Shao-feng, WEN Ming-feng, CHEN Jian-chao, Wuyi Traditional Chinese Medicine Hospital in Jiangmen, Guangdong Province, Jiangmen 529000, China

〔Abstract〕 The paper analyzes the deficiencies of the traditional Hospital Information System (HIS), establishes the hospital data mining and analysis platform, introduces the platform architecture, builds the data mining models themed by analysis of hospital leaders, supervision and management of antibacterial drugs, analysis of hospital quality, analysis of single disease, expenditure control by medical insurance, etc., and puts forward the smart solutions to improve the service level of hospitals on the basis of above.

〔Keywords〕 Data mining; Data warehouse; Data analysis; Analysis platform

1 引言

目前, 我国大部分医院已建立了基本的医院信息系统来辅助事务管理, 且随着计算机信息技术特别是联机事务处理 (On-line Transaction Processing, OLTP)、数据库技术等快速发展, 医院信息

系统对信息的生成、存储、传输和管理等数据的操作处理已十分成熟有效, 成为医院科学管理和提高医疗服务水平的有力手段。但医院业务系统及其数据保存格式主要是面向流程设计的事务处理, 而不是面向分析管理, 因此很难根据现有数据和信息进行深度挖掘, 不能从海量数据中提取潜在知识, 辅助医院管理者的决策, 加强医院的管理水平。在此情况下, 使用数据仓库技术建立医院数据分析平台来帮助医院科学地管理, 综合统一地使用各业务系统中分散的数据, 以及分析和利用已积累的每日不断增加的大量数据, 成为当前信息化建设的一个新的发展方向^[1-4]。

〔收稿日期〕 2017-03-06

〔作者简介〕 马少锋, 主任医师; 通讯作者: 陈健超。

〔基金项目〕 2014 年广东省江门市科技计划项目“数据挖掘在医院管理中的应用研究” (项目编号: 20140200054620)。

2 传统医院信息系统存在的不足

2.1 大量的业务数据难以充分利用

传统医院信息系统产生的业务明细数据的组织方式是面向业务应用而不是面向管理，从中抽取有价值信息的难度非常大，花费的人力、财力、时间等代价都很高。这些业务数据主要表现出异构的数据结构、不同的数据库、分布存放、备份到各种介质、数据量极大等特点。

2.2 查询系统开发成本高、灵活性差、效率低

传统联机处理业务系统是帮助企业运行业务的，是一种结构化体系，由于其静态性和可调整的困难性，对数据的访问受到很大限制，面对分析人员随时发生变化的需求，查询系统开发周期较长，灵活性差，在执行复杂、大量计算的查询时效率低下，查询系统和基础业务系统会相互干扰，影响正常业务系统的稳定运行。

2.3 基础分析数据不一致

多个业务系统可能导致多个查询系统，使用起来非常不方便，多个查询系统又会导致统计口径存在差异，从而引起数据的不一致性，甚至错误数据，可能误导决策。

2.4 缺乏对数据进行深度挖掘分析的能力

查询系统缺乏对数据的分析和挖掘功能，部分查询还需要人工汇总加工，很难发现隐藏在数据中的有价值的信息，针对业务分析方面往往不够全面，无法寻找到规律性内容。

3 数据仓库与数据挖掘技术概述

3.1 数据仓库

数据仓库是一个面向主题的、集成的、相对稳定的、反映历史变化的数据集合，用于支持管理决策，其是一个过程而不是一个项目。数据仓库将决策支持型数据处理从事务型数据处理中分离出来，

数据按照一定的周期，从事务型数据库导入到决策支持型数据库（即“数据仓库”）中^[3]。数据仓库是一个用于存储企业经营的关键知识及信息的仓库，是按业务的“主题”组织数据的。

3.2 数据挖掘

数据挖掘（Data Mining，DM）是一个从大量数据中抽取挖掘出未知的、有价值的模式或规律等知识的复杂过程。其主要步骤包括：（1）数据清洗（Data Cleaning），清除数据噪声和与挖掘主题明显无关的数据。（2）数据集成（Data Integration），将来自多数据源中的相关数据组合到一起。（3）数据转换（Data Transformation），将数据转换为易于进行数据挖掘的存储形式。（4）数据挖掘（Data Mining），利用智能方法挖掘数据模式或规律知识。（5）模式评估（Pattern Evaluation），根据一定评估标准从挖掘结果中筛选出有意义的模式知识。（6）知识表示（Knowledge Presentation），利用可视化和知识表达技术，向用户展示所挖掘出的相关知识^[5]。

4 医院数据挖掘分析平台设计

4.1 平台架构

医院数据挖掘分析平台采用商业智能架构实现，主要由数据抽取 - 转换 - 装载（Extract - Transform - Load，ETL）、数据集成、数据分析和数

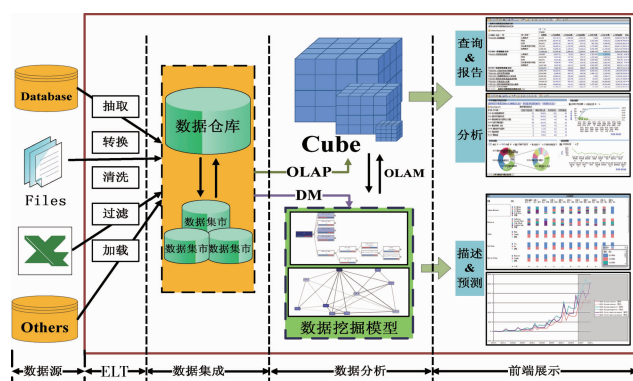


图 1 医院数据挖掘分析平台架构

4.2 数据抽取 - 转换 - 装载

负责完成信息集成工作,从各个原有的业务系统中获取数据,解决异构数据整合问题,完成数据质量的控制、数据清洗、转换、加载等,将数据按照业务要求整理为统一元数据并加载到数据仓库中。该平台采用的 ETL 工具为微软公司企业级数据整合平台 SQL Server Integration Services (SSIS),其提供了构建企业级 ETL 应用程序所需的功能及性能,是一个灵活、快速和可扩展的体系结构,能够对当前业务环境数据进行有效集成^[6]。使用 SSIS 可以在作业过程将需要数据库处理的部分直接处理完毕,可将所有的作业包括文字采矿、汇总、合并、资料比对等整合在同一个服务器中,分离新增数据库与 ETL 工具的作业范围,强化数据库的高可用度,可直接通过 SSIS 进行问题的警示与回报。

4.3 数据集成

即数据仓库层,有效的数据集成是制定有效决策的基础。数据仓库设计是该平台最基础、最重要的过程,良好的数据仓库设计是后续工作顺利进行的保证。医院的数据仓库按不同业务主题分类,将医院需要从各类业务系统中获取的信息通过标准的集成渠道收集并存储起来,实现医院范围内跨科室、跨系统的数据集中和整理。

4.4 数据分析

通过元数据管理、多维分析等技术,实现对数据的多种分析,满足用户的多层次要求。该平台设计多维数据集采用微软公司的 SQL Server Analysis Services。其多维数据集是一种多维结构,可以快速访问大量的预聚合数据,使终端用户能够以思维的速度洞察相关的业务数据,管理人员(用户)可以从不同角度(维度)、通过不同的度量值来观察分析所关心的事实数据,逐步摆脱对固定报表的依赖。SQL Server Analysis Services 作为用于联机分析处理和数据挖掘的中层服务器,可以构造用于分析的多维数据集,提供对多维数据集信息的快速客户端访问;可将数据仓库中的数据组织成包含预先计

算聚合数据的多维数据集,以便为复杂的分析查询提供快速解答;其用户可从多维数据集和关系数据源中创建数据挖掘模型,还可对这两类数据应用数据挖掘模型。

4.5 数据应用

该层为用户访问层,用户通过前端展示工具访问数据分析平台,以多种灵活的格式展现给用户用以分析研究,如图表、报表、嵌入式展示等。该平台采用的前端展示工具为先特计公司的 Analyzer,利用 Analyzer 提供的报表、实时查询、联机分析处理等强大分析功能,让使用者在客户端轻松地存取、分析置放于 SQL Server Analysis Services 多维结构数据库中的数据,在数以秒计的时间内,将图文并茂、易于理解、编排美观的分析结果呈现在管理者眼前。Analyzer 提供的常用数据分析功能主要有同比分析、环比分析、定比分析、趋势分析、相关性分析、差异分析、结构分析、ABC 分析、类比分析、预警分析、80/20 分析、协同报告和个性化分析等。

5 医院数据挖掘分析平台主要应用模块

5.1 院领导分析平台

提供一个院领导分析平台,将院领导关心的指标与数据以仪表盘、图形、报表等形式展现出来,使其可以实时掌握全院运作管理情况及发展趋势,对异常情况及时、有效地采取相关措施。

5.2 运营管理分析平台

主要是针对门诊业务和住院业务的运营管理分析。通过门诊患者(年龄、性别、地区等属性)、患者类型、医生、科室、日期等多个维度结合分析门诊的常用指标,如门诊挂号人次、门诊初诊人次、门诊接诊人次等,轻松组合多样的实用报表,分析医院管理不足之处,以便做出及时调整,方便患者,减少误诊率,减轻患者的经济负担,提高医院的管理水平。对出入院人次、平均住院日、药比、住院费用等进行分析,充分挖掘数据,使其转

化为医院进行宏观管理和科学决策的依据, 提高患者的治愈率及宏观经济效益。

5.3 科主任管理平台

覆盖各个项目工作, 基本实现工作处处有统计、时时有监控, 为科主任的管理提供更好的手段与方式。科主任能快速查询其关心的数据, 分析科室及其医生的工作情况, 从中找出相关规律, 为科室的决策提供辅助。

5.4 护理工作分析

主要从高工作量、高技术含量、高危药物护理、高风险人群护理、特殊科室统计等方面来统计衡量护理人员的工作量情况。

5.5 阳光用药分析平台

通过用药监测智能分析系统, 协助医院管理层从多角度分析和监控药品的流向、用量、金额等情况。从药品的入库、出库, 到各个部门、科室、医生的用药情况, 严格监控药品的使用, 从类型、剂型、抗菌、国家基本药物、类别等多个角度, 分析药品在全院、科室、医生的使用数量、金额、排名、占比等, 为医院严格掌握药品使用提供了极大便利, 也为医院分析用药趋势及规律提供了科学依据。

5.6 抗菌药品分析

实现对抗菌药品采购、医生开方用药、药品使用的规范化、数据化和网络化, 对医生抗菌药品进行信息分析及在线追踪监控, 从而实现对医疗机构非常态化抗菌药品使用现象的及时发现、预警提示、评价分析、纠正改进的全过程。通过分析抗菌药品在临床的使用情况, 有效控制抗菌药品的合理使用, 从而提高抗菌药物的治疗效果、减少细菌性耐药、提升临床药物治疗水平、保证医疗质量和医疗安全。

5.7 处方点评

系统定时将医院信息系统中的用药相关数据导入处方点评库 (区别于医院信息系统数据库), 自

动按照《处方管理办法》、配伍禁忌等要求进行分析, 筛选出用药不当问题的记录, 然后由资深药师在软件辅助下对问题记录再行审核与评价, 最终结果可以按照原卫生部的要求自动形成每月上报报表, 同时形成医院管理所需的系列统计表。该系统主要实现功能有人工随机抽样点评、计算机辅助点评和点评工作表生成。

5.8 医疗质量指标分派

为全面推进深化医药卫生体制改革, 逐步建立医院评审评价体系, 促进医疗机构加强自身建设和管理, 不断提高医疗质量, 保证医疗安全, 改善医疗服务, 通过对医院的质量如院感、护理、不良事件反应、平均住院日、疗效等指标进行监控, 同时对质量指标进行分派监控实行计划 - 执行 - 检查 - 行动 (Plan - Do - Check - Action, PDCA) 管理流程。

5.9 医院质控指标分析

医院运行、医疗质量与安全监测指标反映医疗质量在一定时间和条件下的结构、过程、结果等概念和数值, 建立一套科学的医疗质量评价指标, 实施持续性的医疗质量评价监测, 依此对医疗机构进行追踪评价, 促进医疗质量持续改进。

5.10 医保分析

伴随新医改的逐步推行, 医保覆盖的患者人群日益增多, 而医保中心给予医院的通常是按照定额结算、人次付费, 医院经常面临医保费用超标的问题。医保中心开展的门诊统筹、儿童医保, 对于医院经费的拨付条件更加严格。在此背景下, 利用医保控费系统对在医院就诊的医保患者费用进行适当控制, 对费用信息进行分析, 可达到不超或者少超医保的目的, 实现科学管理。

5.11 病案统计分析

病案系统包含医院很多数据, 医院管理趋于细化, 对数据的要求越来越精细, 医院信息部门或统计部门日常面对多样的数据需求, 需要灵活地报

(下转第 50 页)

工作的紧迫感和责任感,提高安全意识,重视网络安全和信息化建设,重点推进信息安全等级保护、电子认证、安全制度建设和信息安全检查。建立健全本地、本单位网络与信息安全规划和保护策略,进一步完善网络与信息安全管理,明确、落实安全管理责任,加强疾病预防控制信息安全管理。

6 结语

总体来说,近年来我国疾病预防控制信息化建设取得了一定的成绩,但仍存在不少问题,应结合“十三五”卫生信息化规划,完善建设规划,强化标准应用,加大资金投入,重视人才培养,提高信息安全意识,加强安全管理。

参考文献

- 1 国务院办公厅. 国务院办公厅关于印发深化医药卫生体

制改革 2016 年重点工作任务的通知 [Z]. 2016.

- 2 葛辉,李言飞,鲍一丹,等. 基于互联网的流行病学调查数据采集定制系统模型建立 [J]. 中国数字医学, 2011, 6 (7): 37-40.
- 3 孟郁洁,戚晓鹏,马睿,等. 流行病学动态数据采集平台在公共卫生调查中的应用 [J]. 医学信息学杂志, 2013, 34 (6): 18-22.
- 4 张业武,刘冬云,赵自雄,等. 二〇一一年度省级疾控机构信息化建设和应用现状分析 [J]. 中国数字医学, 2012, 7 (6): 2-4.
- 5 夏兰芳,戚晓鹏,孟郁洁,等. 省市级疾控中心卫生信息化人才队伍现状调查 [J]. 中国公共卫生, 2013, 29 (9): 1306-1310.
- 6 中国疾病预防控制中心公共卫生监测与信息服务中心. 中国公共卫生信息化人才队伍现状与发展策略研究报告 [Z]. 2012.
- 7 刘军. 省级疾控机构信息化建设中的信息安全问题与对策 [J]. 医疗卫生装备, 2015, 36 (4): 120-122.

(上接第 21 页)

送,日常工作量非常大。随着信息系统的发展,统计部门应当由数据采集转变为数据分析。

5.12 病案单病种分析

以医院病案统计管理系统为数据源,以出院患者费用及 ICD 诊断为中心,分析 ICD 病种费用、住院天数、药品、院感情况等,对医院控制病种的费用、病种相关性提供有价值的信息,有利于提高医院整体医疗质量和管理水平。

6 结语

运用数据仓库和数据挖掘技术,搭建基于数据仓库的商业智能平台,对医院医疗活动过程中产生的海量数据进行深度加工和多方位分析,从中得到长期、系统、综合、丰富的辅助决策信息,可协助医院管理者及时掌握医院运营状况,为其提供医院的日常运营管理,制定长、中、短期发展规划时所

需要的信息与功能;同时也可对医院的运作情况及趋势发展进行深度分析,结合先进的管理思想与理念,实现更为深度的分析与管理,为医院管理从传统经验型向科学量化方向发展,特别是为重大决策提供科学数据分析基础上的依据。

参考文献

- 1 汤宇松,刘相峰,黄亚楼,等. 数据挖掘系统设计 [J]. 系统工程理论与实践, 2000, 20 (9): 56-63.
- 2 胡蓉,郭芳,张燕. 数据仓库技术在医院信息管理及决策中的应用 [J]. 现代医院管理, 2005, (3): 28-29.
- 3 杨海清. 数据挖掘技术在医院管理中的应用 [J]. 中华医院管理杂志, 2005, 21 (7): 497-499.
- 4 刁琰. 数据仓库与数据挖掘技术在医院信息管理中的应用 [J]. 华南国防医学杂志, 2008, 22 (4): 63-64.
- 5 史忠植. 知识发现 [M]. 北京:清华大学出版社, 2002: 13-17.
- 6 张亚男,董亮. 基于数据挖掘技术的医院决策支持系统建设 [J]. 中国数字医学, 2014, (1): 101-103.