

# 基于数据仓库数据挖掘的医院档案 信息化管理

边荣云

(山东莱芜市人民医院, 莱芜 271100)

[摘要] 医院档案管理信息化存在着海量信息数据无法有效利用的问题, 对医院医疗质量提高, 管理水平提升的辅助决策支持功能不足。本文详细探讨了医院决策支持系统中疾病转归统计分析数据仓库(DW)的设计方法及过程, 并在此基础上构建联机分析处理(OLAP)系统, 根据数据分析的需求选取OLAP主题, 进行多维数据模型的设计, 运用MDX多维查询语言实现疾病转归统计数据分析, 并应用决策树对影响因素进行深度数据挖掘, 实现高效管理和决策的支持作用。

[关键词] 数据仓库; 疾病转归统计分析; 决策支持

中图分类号: R194 文献标识码: A 文章编号: 2055-5200(2014)02-009-03

DOI: 10.11876/mimt201402003

**Information Management of Medical Records Based on Data Warehouse & Data Mining** BIAN Rong-yun. (Shandong Laiwu People's Hospital, Laiwu, 271100)

[Abstract] The information management of medical records may result in the inefficient use of Big Data, thus failing to adequately support in the aid decision-making on the improvement of medical quality and management level. This paper will discuss in detail the design methodology and its process of the Data Warehouse concerning the statistical analysis of disease outcome within the hospitals' aid decision-making system, and try to build an OLAP(On-Line Analytical Processing) system. By selecting a relevant OLAP subject according to the need of data analysis and designing a multidimensional data model, it's possible to complete the statistical analysis of disease outcome with the MDX multidimensional query language and make In-depth data mining of effect factors with the application of Decision Tree. In this way, efficient management and decision-making can be achieved.

[Key Words] Data Warehouse; statistical analysis of disease outcome; decision support

医院档案包括医院诊断、治疗、护理等医疗活动中形成的文字、图表、影像等诸多材料, 是医院医疗业务信息的主要载体。经过系统归纳整理后可以作为医院医疗、教学、科研提供丰富基础资料, 通过病案分析和统计处理可以为医院管理提供一手的详实资料。

随着信息技术的发展, 围绕着医院信息系统(Hospital Information System, HIS)及配套接入开发的诸多系统, 医院的管理模式实现了由经验管理向信

息管理的转变。HIS系统以病人医疗信息、医院经济信息和医院管理信息3条信息主线来提供技术支持平台, 使大量的医院日常医疗、经济和管理事务电子化, 实现了高效的数据录入、查询、统计等功能, 对提高医院工作效率、质量和服务水平等方面起到重要作用<sup>[1]</sup>。档案管理的信息化也在逐步完善, 但是病案资料积累越多, 信息化后的内容越丰富, 信息流的作用越强, 反馈出来的数据处理要求就越高, 如果没有一定的技术手段方法是实现不了这些要求的。

面对医院信息数据量的迅速膨胀，传统数据库陷入“数据爆炸，知识贫乏”的境地。数据仓库(Data Warehouse, DW)的出现有效地弥补了传统数据库在分析型应用上的缺陷。利用数据仓库技术将传统病案系统中所积累的海量数据背后隐藏的有价值的信息挖掘出来，从而获取二次相关信息为医院的决策管理、医疗、科研和教学服务提供了一种科学的解决方案<sup>[2]</sup>。本研究以患者疾病转归统计分析相关信息为数据源，建立了医院医疗质量分析的多维数据模型及多维数据集的数据仓库，通过联机分析处理对疾病转归进行数据挖掘从而提供医疗决策的相关信息，将对医院的管理产生积极作用。下文将对疾病转归统计分析决策支持系统的构建进行阐述。

## 1 数据仓库的构建

### 1.1 疾病转归统计分析数据仓库的设计

以住院病人疗效为仓库主题，建立星型仓库模型，以一个事实表和4个维表为主要内容，其中维表分别为科室维，时间维，病种维和病人基本情况维，如图所示。事实表包含联机分析处理(OLAP)多维数据集中所有的度量值，包括住院天数、各项费用；维度表中的不同属性应用文字，离散值等表示，如年龄段分为5段(0~20岁、21~30岁、30~50岁、50~60岁、60岁以上)；疾病名称以ICD-10主类、ICD-10亚类、ICD-10子类表示，维度关键字是与事实表关联的桥梁，是用于查询事实表数据的唯一标识符<sup>[3]</sup>。通过这种构架，可进行多维度查询。

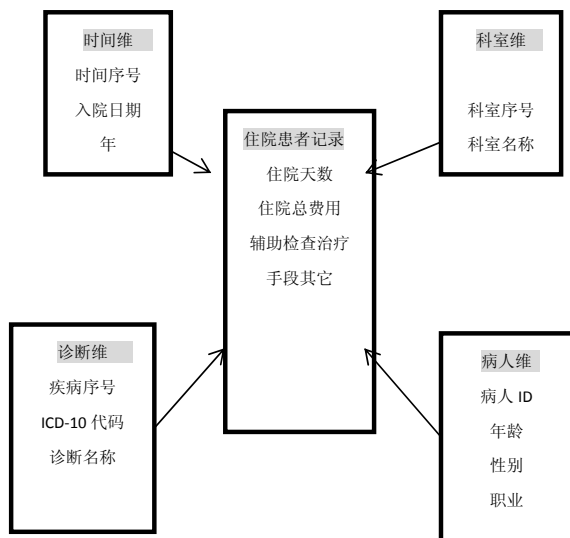


图1 疾病转归统计分析数据仓库星型模型

### 1.2 数据预处理

HIS系统中大量的数据信息中有一部分为空缺数

据、不一致数据或包含有错误或存在偏倚的噪声数据，在进行数据挖掘前必须对数据进行预处理，清洗后的数据才能保证数据的一致性和有效性。数据预处理主要包括数据集成和数据清理两部分，其中数据集成是将所有的病人的基本信息和医疗信息数据进行整合，将从病人主索引表中抽取的所有信息集成到统一的数据表中，对不一致的字段长度，不一致的赋值等进行数据转换，整合为统一的格式，如日期形式2012-04-10转换为时间为上定义的年，月，日格式：2012,4,10；数据清理是将缺陷数据且无法进行转换的数据屏蔽，如筛查2012年的数据资料，经信息提取后发现存在部分2011年的数据资料时，可通过定义语句进行清洗。

### 1.3 数据仓库的加载和转换

在SQL Server 2000中使用Transact—SQL、数据转换服务(DTS)和BSP将数据加载到数据仓库中。DTS是以OLE DB或ODBC等接口技术连接数据源与数据目的<sup>[4]</sup>，其驱动数据进行导入、导出等转换操作，实现数据源与数据目的间的交换。转换结构图见图2。通过上述操作过程，医院数据仓库建立完成，为后续多维数据的分析和数据挖掘进行了铺垫。

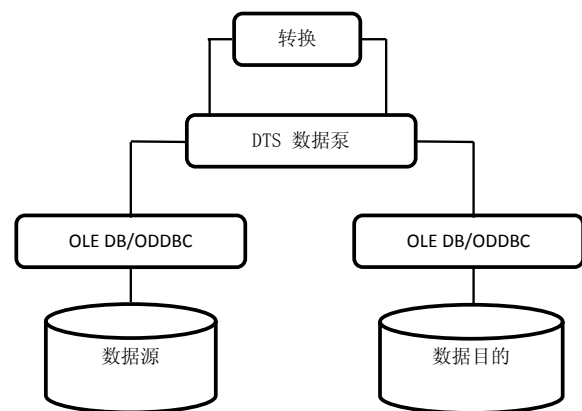


图2 DTS数据转换流程图

## 2 基于数据仓库的疾病转归统计分析 OLAP 建模分析

多维数据集的构建完成后可进行以病人、疾病和费用等多个主题开展的OLAP分析。如通过医疗付款方式维度进行住院费用构成分析；通过月份或年份来分析总住院患者的变化趋势，或在展现层上定义一些附加度量值，如“增减率”，以反映科室病种的变化趋势。在进行OLAP分析中，可以运用切片、切块，旋转，钻过和钻透等操作程序，或通过编写多维查询语言MDX (Multi-Dimensional Expressions, 多维

表达式)程序进行查询,可以获得不同汇总结果上的观察数据,也可以获得住院患者相关因素间的关系,我们也可以选择2个以上的维度的数据实现多维数据集的分析<sup>[5-6]</sup>,如以“科室”和“费用类型”类型为关键词。查询各科室医疗费用占比的MDX代码及结果如下图3:

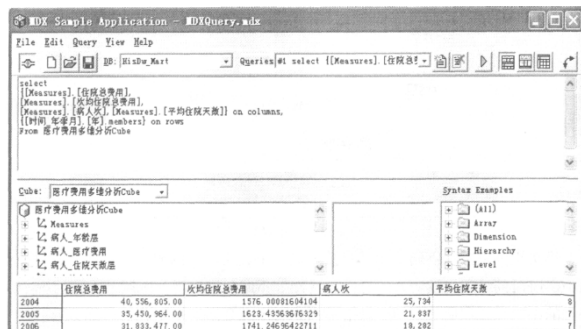


图3 查询结果示意图

### 3 分析决策支持的实现

数据挖掘(Data Mining,DM)技术是决策分析的重要工具,目前应用比较广泛的DM算法主要有基于规则系统的决策树、聚类分析、神经网络、支持向量机方法等<sup>[7]</sup>。在实际应用中,数据挖掘可为医院管理决策提供多层次帮助。例如利用决策树方法提取数据库中的患者信息,包括性别、年龄、入院情况、药品费用等各个属性,构建相关性网络,应用相应的运算方法分析患者信息,得到总医疗费用的强关联项住院天数与药品费,最终得到医疗费用的层决策树,如图4所示。决策树由“Generate decision tree”给定的训练数据产生,训练样本有表示住院费用的每个属性以离散值的形式表示,运算过程中产生的各个节点为不

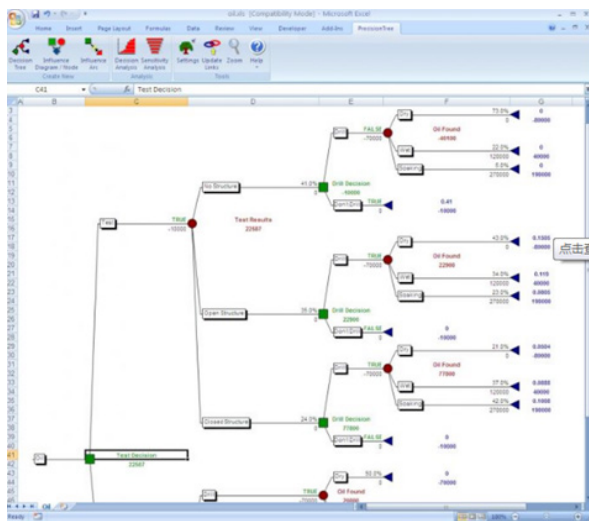


图4 医疗费用决策树示意图

同的分支点,显示费用的不同影响属性,而每一个节点内的所有样本均为同一类别。从数据挖掘的最终结果看,药品费用和住院天数是医疗费用的强影响点,与医疗费用呈正相关,住院天数越长,药品费用越高则患者的总住院费越高。在模型判断的基础上进行有针对性的费用控制决策将是一种目的明确、针对性强,高效的决策支持方法<sup>[8-9]</sup>。

### 4 结论

针对医院档案利用效率低,海量数据对信息管理系统决策支持不足的现状,尝试应用数据仓库和基于多维数据集的OLAP技术及数据挖掘技术,对医学信息数据进行分析,挖掘出更多深层次的、有价值的信息,为医院管理者提供了辅助决策的依据。同时,随着医院信息技术的不断发展,还可以开发预算警示模块等多个实用的,有针对性的模块,建立面向医院各级管理层的决策支持系统,切实地、真正地实现医院的精细化、循证化管理<sup>[10]</sup>。系统的开发使用为提高医院管理水平提供了高效、可行的技术手段。

### 参 考 文 献

- [1] 马刚,刘天时,程国建.基于数据仓库技术的医院信息系统应用研究[J].西安石油大学学报(自然科学版),2010,25(4):99-102.
- [2] 叶明全,宋念东.基于HIS的医院数据仓库设计与实现[J].计算机工程与设计,2008,29(3):779-781.
- [3] 李慧宗,张学森.面向主题的医疗费用分析数据仓库建模[J].中小企业管理与科技,2009(31):274.
- [4] 周治宇,郑昊.基于医院信息系统数据仓库的设计[J].中国数字医学,2010,05(9):81-83.
- [5] 郑丹青.基于OLAP技术的医疗信息多维数据集设计与分析[J].吉林师范大学学报:自然科学版,2011,32(3):56-58.
- [6] 孔琳.基于HIS的数据仓库构建及多维分析[J].医学信息学杂志,2011,32(11):6-9.
- [7] 胡志东.数据挖掘及数据仓库综述[J].黑龙江科技信息,2010(26):100.
- [8] 朱秋丽,张涛,丁云芳,等.呼吸道合胞病毒感染儿童住院费用影响因素的回归树分析[J].复旦学报(医学版),2011,38(4):294-299.
- [9] Lemon SC,Roy J,Clark MA,et al.Classification and regression tree analysis in public health:methodological review and comparison with logistic regression[J].Ann Behav Met,2003,26(3):172-181.
- [10] 唐熊,张巨发,段昌奉,等.基于移动智能终端的医院管理决策支持系统开发及应用[J].现代仪器与医疗,2013,19(1):47-49.