

联机分析处理在门诊药品收费管理的应用^①

An OLAP Analytical Practice on Drug - use of Outpatient Department in a Hospital

戴科星 叶枫 (浙江工业大学 经贸管理学院 杭州 310014)

摘要:本文首先对联机分析处理(OLAP)和当前医院信息(HIS)的应用背景作了简要介绍。然后,笔者介绍了在实际应用中,将某医院门诊药品使用情况的原始 OLTP 数据库转化为 OLAP 数据库的过程以及该 OLAP 的架构。对中西药使用比例和抗生素药品使用频率进行了分析,并且总结了能对决策提供支持的有用信息和规律。

关键词:联机分析处理 Analysis Services 数据挖掘 门诊药品

1 引言

E. F. Codd 提出关系数据模型后,关系数据库和联机事务处理(Online Transaction Processing, OLTP)在实际应用领域得到了快速发展。随着应用在广度和深度方向上的延伸,某些企业中的数据库规模已从兆(M)、千兆(G)过渡到兆兆(T)字节,特别是某些大型金融、电信、零售企业。与此同时,用户的查询统计越来越频繁、复杂度越来越高,导致数据库服务器负荷增大、速度降低;而且所使用的特定应用程序所提供的查询统计的灵活性一般较差,用户很难根据自己需要的分析角度来对数据进行查询和统计。更为重要的问题在于,日常数据不是所有事实的叠加,其中往往隐含着不为人知的模式和规律。所以如何充分组织、利用这些海量的日常数据,使那些企事业单位能够从中找到有指导意义和实际价值的信息和知识日益成为热点。联机分析处理(Online Analytical Processing, OLAP)和数据挖掘(Data Mining, DM)在这样的背景下应运而生^[1]。

现阶段,很多经济较发达地区几乎所有的大中型医院都已经拥有医院信息系统(HIS),以此来提高整体运行效率,提升自身竞争力。一些 HIS 启用较早的医院开始向着数字化医院的目标发展。但据我们所知,多数医院的 HIS 系统仍然局限于日常数据录入、偶尔查询统计的层面上,没有充分发挥数据在决策过程中

的支持作用。在开源节流成为各级医疗单位管理主题的情况下,如何能充分利用这些大量积累的宝贵数据,为医院管理层提供辅助决策支持信息成为 HIS 的热点话题。本文就是基于此,对 HIS 的原始数据进行了 OLAP 分析。

本文所使用的数据源来自某三级甲等医院 HIS 系统的后台数据库。所考查的对象为该院 2005 年全年的门诊药品收费项目明细。

2 Analysis Service 的功能介绍

作为 Microsoft SQL Server 2000 的组件,Analysis Services 是 OLAP 和数据挖掘之间的中间层服务器。它主要包括两种决策支持机制:OLAP 和数据挖掘。其中的数据挖掘功能,是 Analysis Services 中新增的功能,弥补了之前版本 OLAP Services 只能进行 OLAP 的不足。

联机分析处理是一项能够分析大型数据库从而对其包含信息深入洞察的技术,它具有汇总、合并和聚集的功能,并且能够从不同的角度来查看信息^[2]。

Analysis Services 能够高效地检索数据和聚合数据,它可以进行的操作包括上卷、下钻、切片和切块、旋转等。Analysis Services 按照层次结构预先存储汇总的数据,允许用户以维的特性训练数据^[3]。同传统的关

① 基金项目:基于空间数据挖掘的决策支持系统体系结构研究(基金号:602096) 浙江省自然科学基金

系数据模型相比,OLAP 的优势在于:它所使用的多维概念模式和用户视图更加符合人们思维习惯;它所提供的查询统计可以灵活的满足用户的分析角度;更重要的是,它可以在一个相对固定的短时间内提交查询统计结果,受原始数据量规模的影响很小。

3 OLAP 在分析医院门诊药品使用情况中的应用

3.1 数据准备

数据准备是成功进行 OLAP 分析的前提和基础,开发人员需要将有效的、需要的数据从 OLTP 数据库转移成 OLAP 数据库,它主要包括了“校验”、“抽取”、“净化”等操作,使数据噪声尽可能小。笔者首先使用微软数据转换服务(Data Transformation Services, DTS)从该院 HIS 系统的 2005 年历史数据中将所需的事实表(门诊诊疗表)和维度表(字典药品表、字典医生表等)提取到新建的 SQL Server 2000 数据库中,建立一个名为

HIS_OLAP 数据库。然后,对各表进行必要的表结构修改,例如删除门诊诊疗表中对建立 OLAP 立方体无关的字段;删除门诊诊疗表中非药品的明细记录;创建时间维表、时间段表,这两个新建的表用于建立两个时间维度,分别从季度、月、日和每天的时间段的角度分析药品收费情况。

3.2 建立 OLAP 立方体

OLAP 是基于多维数据集的。多维数据模型围绕中心主题组织,用事实表来表示。事实表中的度量值就是我们要考察的对象,以及用维度表示各种因素是否同它有关联关系和关系的强弱程度。图 1 给出本文分析 HIS 药品收费数据所建立的名为 ClinicTreat 的多维数据集。

如上文所述,多维数据集 ClinicTreat 将 ClinicTreat 表(门诊诊疗表)设为事实表,Amount 字段(金额)设为度量值。维度的具体设置信息由表 1 给出。

表 1 多维数据集 ClinicTreat 的维度设置

维度名	维度创建方式	级别设置	构成的表
TimePeriod	星型架构	以小时划分的时间段	TimePeriod(时间段表)
TheDate	星型架构	季度、月、日	DateByDay(日期表)
TheDay	星型架构	星期	DateByDay(日期表)
DoctorDept	星型架构	科别、医生	DoctorInfo(医生信息表)
Antibiotic	星型架构	是否抗生素、单个药品	DrugInfo(药品信息表)
DrugInfo	雪花架构	药品类型、单个药品	DrugInfo(药品信息表)、 DrugType 表(药品类别表)

建立完多维数据集之后,用户就能通过 OLAP 引擎生成可以浏览的 OLAP 立方体了。接下来我们通过新生成的 OLAP 立方体进行分析,并试图寻找隐含的模式和知识。

3.3 浏览 OLAP 立方体和规律发现

3.3.1 药品分类使用情况分析

以下,我们从中西药品金额比例构成和抗生素药品使用比例两个角度浏览 OLAP 立方体。

(1) 中西药品金额比例构成。DrugInfo 维度是以雪花模型构建的,药品信息表中,每一种药都有中西药标志标识药品属于西药、中成药还是中草药。通过下面显示的图 2,我们看到,如果把中成药和中草药划归

为中药,那么西药和中药的金额比例大约在 3:1 的水平。可见,西药在该院药品使用中占据了主导地位。这同我国医院整体情况是一致的。近日由中国中药协会主办的现代中药发展战略高层论坛上一项最新数据表明:全国 12000 家西医院里,中药只占业务总收入的 16%,其他 84% 用的是西药。

(2) 抗生素药品使用比例。最近一段时间,医院抗生素的使用率的问题引起了医疗行业甚至整个社会的关注。我国某些地区的抗生素类药品使用量占医疗机构用药总量的 30% 左右,而合理使用抗生素的比例只占 50—60%。这对临床治疗带来了不利影响。这项分析我们以医生信息维度和抗生素使用维度作为旋

转表的行和列。

通过浏览多维数据集,我们得到门诊全部科室的抗生素使用率为 19.31%。可以说,门诊控制使用抗生素的总体情况还是不错的。

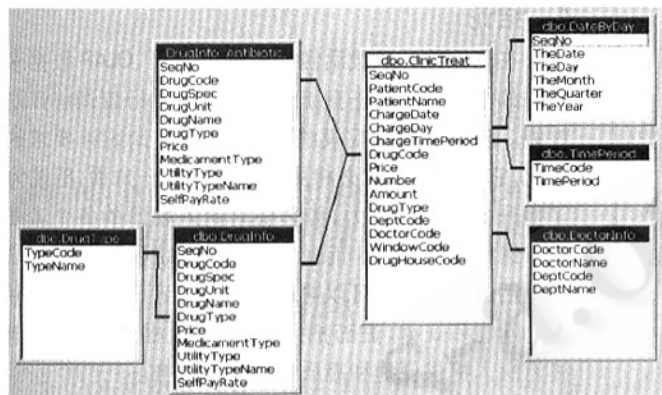


图 1 多维数据集 ClinicTreat 的结构

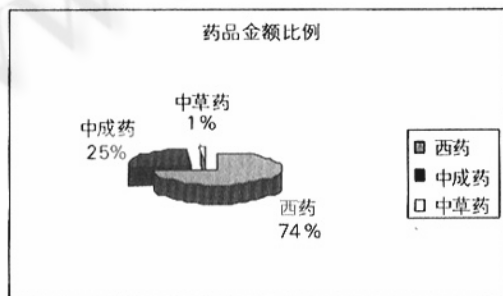


图 2 药品分类汇总统计

接下来我们对各个科室进行了横向比较。图 3 给出了药品总金额位列前 7 位科室的抗生素使用率的情况。我们知道,抗生素使用率和科室的特点是有很大关联性的。比如外科的使用率为 41%,这是符合其特点的。而儿科普通的使用率高于外科的使用率,这个信息就需要引起我们的注意了。

我们设计的多维数据集还可以进一步对科室中的各个医生的抗生素类药品的使用率进行汇总和横向比较。由于篇幅所限,不能再此进行详细的分析。但是,这个信息可以被药剂科的管理决策人员所利用。比如可以根据每个科室的特点以及历史数据制定使用率基数。每段时间进行分析,比较他们的数值是否超标,而

且可以明显地看出各个科室哪些医生所开的处方超标以及超标的程度。

3.3.2 时间和药品收费金额的关联关系

我们从浏览的结果发现,无论是季度还是月的纵向比较,药品收费金额的起伏都很小,只有日的纵向比较出现较大起伏。因此我们可以认为医院收费繁忙程度同季节之间的联系并不明显。

与此同时,我们却发现一天中的时间段和星期对医院收费繁忙程度的影响很大,我们分别用图 4 和图 5 来反映这两个因素的影响。

从图 4 我们能够总结出时间段对药品收费金额的影响作用。此图呈现类似“M”的形状。收费高峰期分别出现在每天的 9 点 - 11 点和 14 点 - 16 点这两个时间段,9 点 - 11 点达到一天的最高峰。

从图 5 我们可以发现,每周从周一到周日门诊收费的繁忙程度呈现逐步下降的情

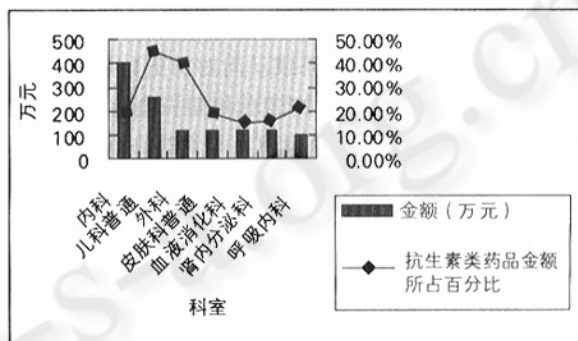


图 3 科室抗生素类药品使用率情况

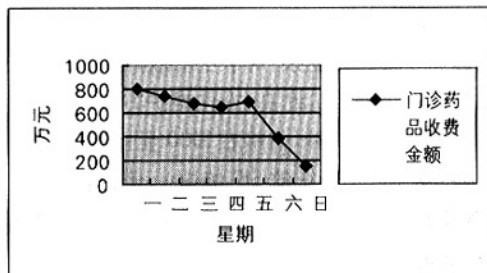


图 4 以时间段为周期的药品收费金额统计

况。抛开周六周日某些科室停诊的情况,周一到周五

的数据总体上也呈下降趋势。此图虽然反映的是按星期为周期加总后的情形,不能保证每周都复制这样的规律,但其反映的总体趋势是具有借鉴作用的。

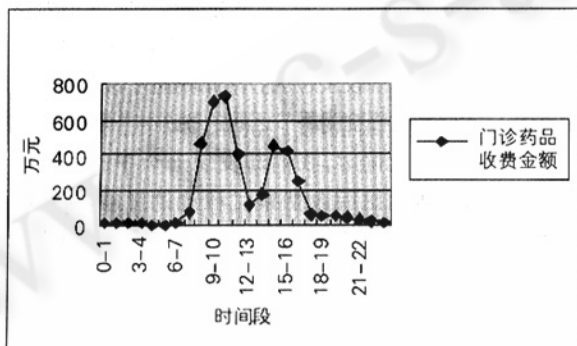


图 5 以星期为周期的日药品收费金额统计

经过对该院门诊收费的调研,我们发现,病人收费等待时间过长是长期困扰医院决策层的一个问题,甚至影响到了门诊的整体运行效率。解决这个问题一方面需要方式上的创新,比如某些城市的某些医院开始门诊预收费的尝试,从而减少收费次数;另一方面需要技术创新,比如更新收费系统。而对历史数据的充分利用,找到某些具有通性的信息和知识也是重要的。比方说,基于以上两个时间角度对门诊药品收费的分

析,能让我们找到解决该问题的一个方法,即根据不同的日期、不同的时间段合理安排收费窗口和收费人员的数量,为排班决策提供支持。

4 结论

在查询统计的灵活性和高速性方面,OLAP 数据库系统的性能大大高于直接对原始数据进行查询统计的性能。随着数据库规模的扩大、用户查询统计要求的深化,它的使用范围必将更加广泛。另一方面,医院为了提升自己的竞争力,势必要对病人就诊的各个流程环节进行优化设计。而对自身 HIS 系统积累的数据进行灵活的查询统计可以帮助相关决策者发现流程中存在的问题和规律,进而把握规律、对流程进行优化,最终为病人提供便捷和高质量的医疗服务。

参考文献

- 1 贺广生、蔡勇,基于 MS Analysis Services 的 OLAP 系统的设计与实现,信息技术,2002 年第 7 期,71-73。
- 2 Microsoft Corporation 著,分析服务[M],东方人华译,北京 清华大学出版社,2001:175。
- 3 张理强、郑兆瑞,SQL Server 2000 Analysis Services 在税务信息系统中的应用,太原理工大学学报,第 34 卷 第 4 期 2003 年 7 月。