

赵悦 周利民 傅冬梅 (北京科技大学信息学院 100083)

1 引言

数据仓库技术和联机分析处理是近年来信息领域迅速兴起的计算机技术。数据仓库经常用作决策支持系统 (DSS) 的基础。传统的 DSS 一般是直接建立在联机事务处理 (OLTP) 环境上的, 但它对分析处理的支持一直不能令人满意, 甚至有明显的冲突, 因而无法为进行联机分析的用户提供充分支持, 而数据仓库的设计解决了 DSS 的数据管理问题, 因此成为进行分析决策的基础, 决策支持还需强有力的工具进行分析和决策, 新兴的软件技术——联机分析处理 (OLAP) 就是专门设计用于支持复杂操作。在数据仓库和数据集中存储的是用于分析的数据, 而 OLAP 是使客户端应用程序实现对这些数据进行高效率访问的一种技术。

九十年代以来, 移动通信行业的运营网络管理系统、综合业务系统、计费系统、办公自动化等系统相继投入了使用, 这些计算机应用系统的运行, 积累了大量的历史数据, 但在很多情况下, 这些海量数据在原有的作业系统中无法提炼并升华为有用的信息及时提供给业务分析人员与管理决策者。因此, 管理者和决策者希望存储在系统中的数据能发挥作用, 能够直接、自由、快速地访问自己的数据, 并收集其中有价值的信息。在这种条件下, 开发移动行业的决策支持系统很有必要。本文采用数据仓库技术和 OLAP 技术来研究移动行业的决策支持系统, 主要实现网络资源、用户话务量、用户行为的分析决策, 利用数据仓库基本原理, 设计了相应的数据结构, 提出了数据仓库数据析取与联机分析处理的有效方法, 可以应分析人员要求快速、灵活地进行大数据量的复杂查询处理, 使决策者能准确掌握经营状况, 了解市场, 判断正确方案, 增加效益。

摘要: 计算机应用在移动通信行业已经有十几年的发展, 每天的业务处理都会产生大量的数据。怎样利用这些数据, 深层次地挖掘数据资源并进行分析, 使移动公司的决策者能及时掌握公司的运营情况, 并根据这些分析的结果指定长远的政策, 从而提高公司的业务发展和竞争优势, 成为移动公司目前努力的方向。

关键词: 数据仓库 联机分析处理 数据挖掘 决策支持

2 决策支持技术

决策支持的本质, 是提取收集到的数据, 进行智能化的分析, 揭示企业运作和市场情况, 帮助管理层做出正确明智的经营决定。一般现代化的业务操作会产生大量的数据, 其中一部分是决策关键数据, 但并不是所有的数据都对决策有决定意义。决策支持包括收集、清理、管理和分析这些数据, 将数据转化为有用的信息, 然后及时分发到企业各处, 用于改善业务决策。

决策支持的实现方式多种多样, 其基本体系结构往往包括三个部分: 数据仓库、多维分析和数据挖掘 (见图 1)。

2.1 数据仓库

一个企业的信息往往分布在不同的部门和分支机构, 管理者要综观全局、运筹帷幄, 必须能迅速地找到能反映真实情况的数据, 这些数据也许是当前的现实数据, 也可能是过去的历史数据, 因此, 有必要把各个区域的数据集合起来, 对其进行筛选, 将真实的、对决策有用的数据保留下来, 随时准备管理人员使用。

数据仓库中存放用户需要分析的历史数据, 它提供了丰富的工具来清洗、转换和从各地提取数据, 使得放在仓库里的数据有条有理, 易于使用。数据仓库中数据主要来自计费系统、营帐系统、客服系统、网管及其他系统, 其中包括用户资料、用户帐单、欠费信息、缴费信息、计费系统中的统计及结算信息, 另外数据挖掘阶段还用用户清单等。这些数据必须经过一定的处理, 这就是数据的抽取、转换和装载 (ELT), 这样, 将不合理的数据剔除, 保证数据仓库中的数据是干净、一致和可靠的, 然后根据不同的要求将清洗过的数据存储起来。最终用户利用各种分析工具进行分析, 把得出的结论提供给决策们。

2.2 多维分析 OLAP

管理人员往往希望从不同的角度来审视业务数值, 比如从时间、地域、功能、利润来看同一类业务。每一个分析的角度叫作一个维, 因此, 我们把多角度分析方式称为多维分析。以前, 每一个分析的角度需要制作一张报表。由此产生了在线多维分析工具, 它的主要功能, 是根据用户常用的多种分析角度, 事先计算好一些辅助结构, 以便在查询时能尽快抽取到所要的记录, 并快速地从一维转变到另一维, 将不同角度的信息以数字、直方图、饼图、曲线等等方式展现出来 [3]。

多维分析工具往往还要与一些前台分析工具配合使用, 提供简单易用的图形化界面给管理人员, 由他们自由选择要分析的数据、定义分析角度、显示分析结果。

数据仓库、多维分析是决策支持的基础, 它们完成的是对用户数据的整理和观察, 可以说, 它们的工作是总结过去。在此基础结构之上, 决策支持可以发挥更进一步的作用, 利用数据挖掘技术, 发现问题, 找出规律, 达到真正的智能效果: 预测将来。

2.3 数据挖掘

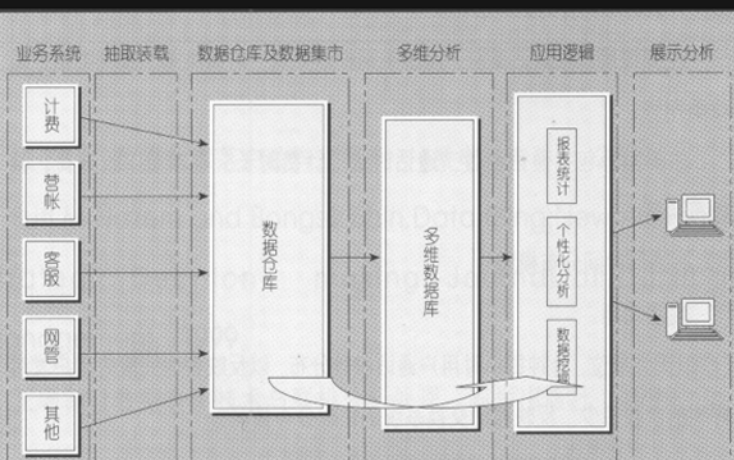
数据挖掘主要是提供隐含在数据背后的相关数据之间联系, 并对不同的因素发生变化时, 其他因素如何发生变化, 对趋势作出预测。数据挖掘使用统计、分析等数学方法, 以及机器学习和神经网络等人工智能方式 [2]。在数据仓库的数据里可以挖掘出业务人员意想不到的信息, 它比多维分析更进一步。

数据挖掘要求有数据仓库作基础, 并要求数据仓库里已经存有丰富的数据。因此, 在实施决策支持方案时, 一般分两步走: 第一步实现数据仓库和多维分析, 构造决策支持的基础, 实现分析应用; 第二步实现数据挖掘, 发挥决策支持的特色。

3 决策支持系统的分析与设计

移动通信的 DSS 组成包括日常的业务信息管理、数据仓库、OLAP 多维分析、前端展示及数据挖掘。其中日常业务信息处理包括计费、营帐、客服、网管等系统, 从这些日常业务系统中过滤、提取数据, 存放入数据仓库。本系统基于移动行业中普遍关心的话务量结构分析建立了一个决策支持系统, 其需要的数据主要有客户资料数据、号码段数据、缴费资料、全球通本地用户详单、全球通外省用户来访详单、神州行用户详单等。系统采用数据仓库作为决策支持的基础, 设计多维数据库存储, 根据性能要求及系统硬件的资源, 使用

图 1 移动通信决策支持系统逻辑框图



关系型存储方式,进行统计数据存储。

3.1 数据仓库设计

数据仓库中的数据面向主题进行组织,主题对应企业中某一宏观分析领域所涉及的分析对象[1]。经过对移动业务的分析,该系统数据仓库的主题域主要有客户资料信息、缴费资料、用户话单信息。

3.1.1 客户资料

- 客户固有信息: ID号、机主姓名、年龄、所在地县等。
- 客户业务信息: ID号、机器号码、计费代码、业务类型、信誉度等。

3.1.2 缴费信息

- 月缴费历史表: ID号、手机号、出账年、出账月、现金款等。
- 年缴费历史表: ID号、手机号、出账年、现金款、欠费款等。

3.1.3 号码段包括归属局代码(省代码)

3.1.4 欠费信息

- 月欠费表: ID号、手机号、出账年、出账月、欠费状态、月租费等。
- 年欠费表: ID号、手机号、出账年、欠费状态、月租费等。

3.1.5 用户详单历史

包括手机号码、卡号、通话时间、通话时长、呼叫费、长途费、附加费、通话类别、漫游类型、长途类型、对端类型、服务类型、归属局、漫游局等。

3.2 OLAP 需求分析

OLAP是对特定问题的联机数据访问分析。通过对信息(这些信息已经从原始数据进行了转换,以反映用户所能理解的企业的真实“维”的很多可能的观察形式)进行快速的、稳定一致和交互性的存取,允许管理决策人员对数据深入观察。在分析处理中需要对信息统计综合,所以联机分析处理除了提供直观、统一的数据访问界面外,内部还有数据综合引擎。下面对应用需求进行OLAP分析。(移动DSS的OLAP数据流图见图2)分析联机分析处理对象的“维”,即用户希望从哪些角度观察数据。根据联机分析处理的数据流图已说明系统联机分析处理对象有四个模型。

3.2.1 用户行为分析模型

主要功能:该模型主要针对需求,按主被叫分类、本地通话类型分类、按长话类型分类、按漫游类型分类、按对端类型分类、按业务类型分类、按业务品牌分类等等,并对上述各种类型组合进行交叉分析。

主要维度包括通话类别、漫游类别长话类别、对端类别、服务类别、品牌类别等。

主要指标包括通话次数、通话时长、计费时长、基本费、长途费、其他费用等。

3.2.2 通话时间分布模型

主要功能:

(1)以天为单位,分钟每小时用户通话次数分布,以反映各时间段内用户发起呼叫的次数多少,反映移动交换系统的忙时分布情况。

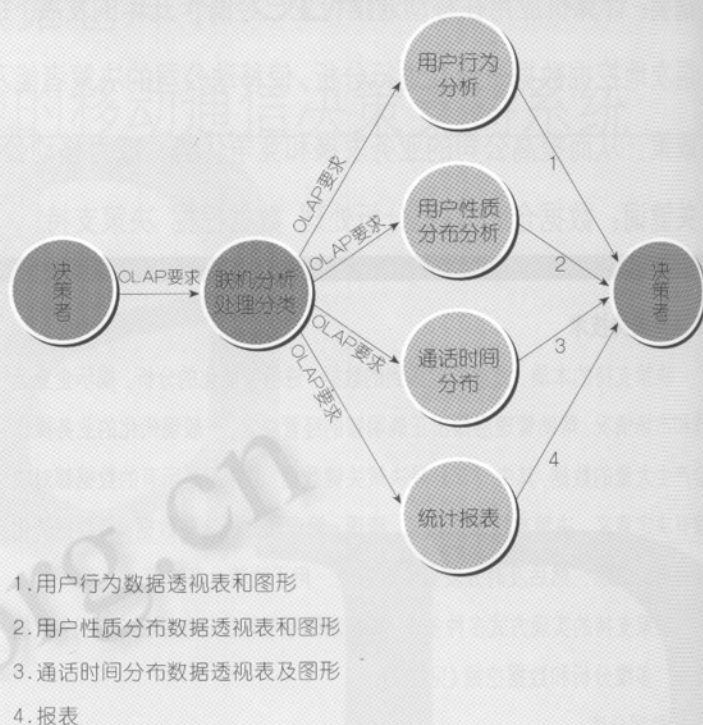


图2 决策支持系统数据流图

(2)以天为单位,分析每小时用户通话量分布。

主要维度包括时间、交换基站等。

主要指标包括通话次数、计费时长、本地呼叫费、长途费、其他费用等。

3.2.3 用户性质分布分析模型

主要功能:该模型分析不同类型用户的通话费用和通话时长、通话次数分布。

主要维度包括出生年代、性别、付款方式、用户性质、行业性质、时间段等。

主要指标包括通话次数、计费时长、总费用、用户数等。

3.2.4 统计报表模型(本系统此模型同用户行为分析模型)

主要功能:“所见即所得”地定制完成一个所需的报表。

3.3 数据挖掘应用

(1)主要目标:分析那一种因素跟短消息的使用的关系最为密切。

(2)数据字段包括:手机号、客户类型、客户所在地区、手机品牌(全球通/神州行)、行业信息、使用的短信息台个数、向该手机发送短信息的手机号码个数、短信息呼出费用、短信息接收费用等一系列统计费用。

数据挖掘方法:通过Neural Network数据分群算法,得到了7类用户。通过对7类用户的特征比较,得出结论,提供给决策者,以适当调整政策。

4 移动通信DSS的实现

该移动决策支持系统中采用基于以下软件平台的解决方案。

图 3 数据抽取

4.1 数据仓库的实现

数据仓库采用 IBM 的 Warehouse Manager, 它是 IBM 数据仓库解决方案的重要组成部分。它主要由以下几部分功能组成: 数据访问, 数据转换, 数据分布, 数据存储, 靠描述性数据查找和理解数据等。数据访问要定义原数据源及目标数据源, 转换过程通过定义过程 (process) 来完成 (见图 3), 例如选取所用的表、表的一些字段, 生成 SQL 语句, 并指定最终存放的目标表。

4.2 联机分析处理的实现

OLAP 多维分析和前端展示利用 Cognos 公司的 Cognos Impromptu Administration, Cognos PowerPlayr Transformer 及 Cognos PowerPlay 来查询数据, 构造 Cube 及观测 Cube。通过 Impromptu 与 Warehouse 连接选取所需字段, 利用 PowerPlay Transformer 确定维度、指标以建造模型 (见图 4), 观测 Cube 是通过 PowerPlay 来实现的, 它通过相关维度的不同组合分析所关心的指标, 这些观测可以从饼图、柱状图、曲线等多种形式实现。并且该软件可以把 Cube 发布到 INTERNET 提供远程查询分析功能。

4.3 数据挖掘的实现

数据挖掘一般采用 IBM 商业智能软件中的 DB2 Intelligent Miner, 其中包含了现今行业中比较流行、完善的挖掘算法。它的主要功能包括创建和定义对象、使用处理函数来准备数据、发掘数据、使用统计函数来分析数据、使用序列来使发掘过程自动化。

5 结束语

本文给出了一种基于数据仓库技术和 OLAP 的实现话务量结构分析的移动 DSS 解决方案, 将数据仓库技术和联机分析处理 (OLAP) 引入系统, 解决了传统系统忽视历史数据和 DSS 数据管理难等问题, 有效地实现了数据仓库的建立和数据抽取, 使移动公司更好地了解自己的运营情况, 为业务发展提供了良好的决策基础, 并使得用户能以一致的界面向快速地从多个角度观察, 深入分析数据。

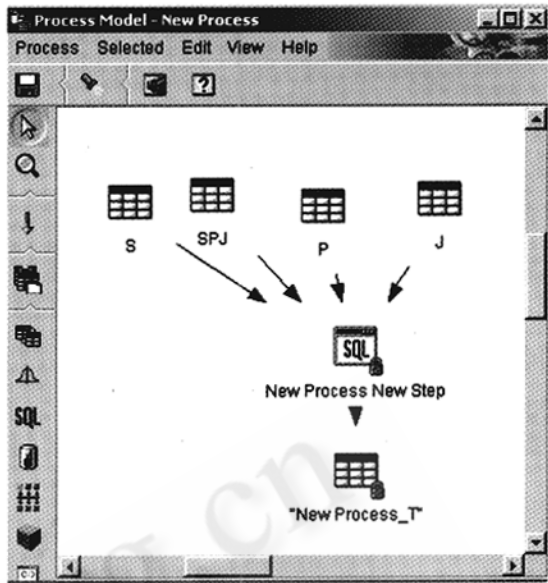
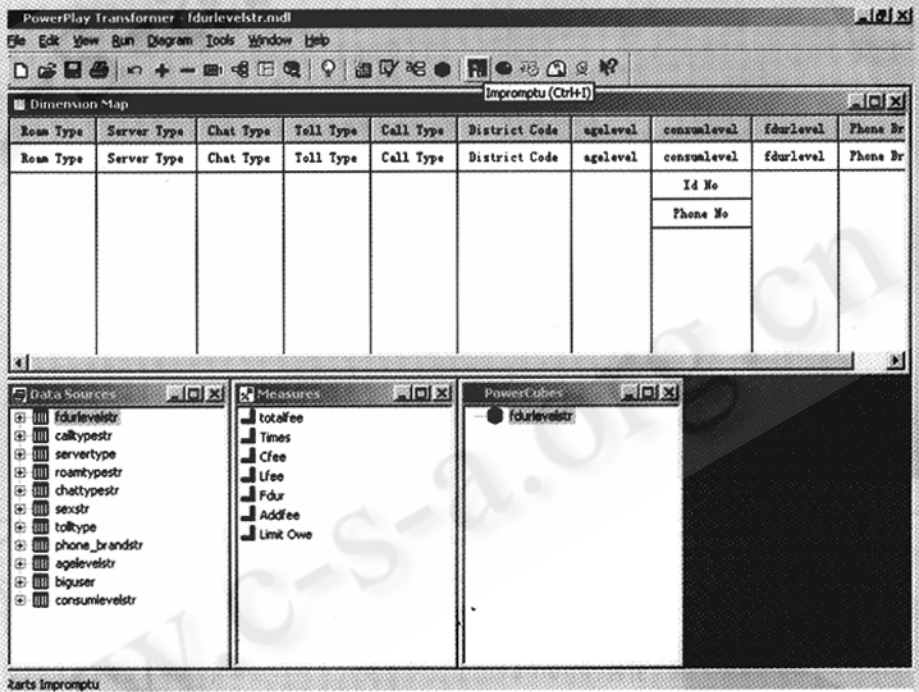


图 4 建立 cube 模型



参考文献

- 1 张冬莉, 数据仓库落户证券交易所, 《微电脑世界》, 2001.
- 2 Sufi M.Nazem and Bongsik Shin, Data Ming: New Arsenal For Strategic Decision-making. Journal of Database Management, 1999.
- 3 王珊等, 数据仓库技术与联机分析处理, 科学出版社, 1999.