

数据仓库技术在配送决策支持中的应用

欧广宇 王红安 (郑州大学工学院电信学院 450002)

摘要: 文章介绍了数据仓库和数据挖掘的概念以及 OLAP 和 ADO MD 技术, 结合连锁超市配送决策支持系统, 阐述数据仓库的建立过程、建立立方体的方法, 并举例说明了在立方体上运用 MDX 语言进行数据挖掘的方法。

关键词: 数据挖掘 联机分析处理 MDX

1 引言

在大规模连锁超市的经营中, 商品配送过程是关系到经营业绩的重要环节。通过对配送需求的预测可以保证合理的商品配置和降低配送费用。以往, 配送需求由人工制订, 主要是依据个人经验, 有着种种不足之处。与此同时, 企业在经营活动中, 积累了大量的数据, 如销售数据、历史配送数据、库存数据、计划数据、商品结构数据、地区数据、市场数据等。这些数据蕴涵着宝贵的信息, 如能充分利用, 对企业的发展极为有利。

随着数据仓库理论、数据挖掘和 OLAP 技术的发展成熟, 使得数据挖掘辅助企业管理者进行决策成为可能。作者开发的连锁超市配送决策支持系统, 在海量数据的基础上, 构建全新结构的数据仓库, 以 OLAP 为手段, 运用数据挖掘方法, 为商品配送需求参数的确定提供决策支持。

2 数据仓库及相关技术

2.1 数据仓库

W.H.Inmon 在其所著的 "Building the Data Warehouse" 一节中给出了数据仓库的定义: 数据仓库是面向主题的、整合的、稳定的, 并且时变的收集数据以支持管理决策的一种数据决策形式。^[1]

数据仓库系统是多种技术的综合体, 由数据仓库、数据仓库管理系统 (DBMS)、数据

仓库工具三部分组成。数据仓库管理系统负责管理整个系统的运行, 是整个系统的引擎, 数据仓库包括远期基本数据、近期基本数据、轻度综合数据和高度综合数据, 是整个系统的核心; 数据仓库工具则是通过联机分析处理 (OLAP, On-line Analytical Processing) 工具、数据挖掘 (DM, Data Mining) 工具等高效的工具来发挥数据仓库的真正作用。

2.2 数据挖掘

数据挖掘是从大量的数据中挖掘稳含的、有用的、潜在的信息和知识。^[2] 它在数据仓库的基础上利用人工智能、机器学习、统计学等技术, 自动的分析数据, 进行归纳性的推理和联想, 从中挖掘出潜在的模式和规则, 帮助企业决策者调整市场策略, 规避风险, 做出正确的判断和决策。

2.3 在线分析处理 OLAP (online analytical processing)

联机分析处理 (OLAP) 是一种针对特定问题的联机数据访问和分析处理技术。^[2] 它从企业的数据集中收集信息, 并运用数学运算和数据处理技术, 灵活、交互式的提供统计、趋势分析和预测报告。通过多种 OLAP 工具对数据仓库中的数据进行多维分析、汇总, 形成图表或报表, 使决策者可以清晰、直观的看到分析结果。数据密集型行业 (data rich industries) 是 OLAP 软件的主要需要者。

3 配送决策支持系统

3.1 系统功能简介

连锁超市配送中心对库存商品入库、出库、退货、配送、盘点、报损等相关操作进行管理。系统主要功能是配送计划的制订和调拨计划的生成。此外, 还可以进行商品信息和业务信息的查询。

在大规模连锁超市的管理中, 各个门店的商品大部分是由配送中心来统一配送的, 配送中心根据门店所需商品的品种和数量, 均衡考虑各门店的请求制定配送计划, 再经过调拨指令处理, 使期更为合理, 然后按配送计划将商品送往各门店。

3.2 数据仓库的建立

连锁超市配送决策支持系统由后台数据仓库和前端分析工具两部分组成。数据仓库是运用 OLAP 工具进行分析的基础, 它的建立包括三步:

- (1) 确定用户需求;
- (2) 设计和构建数据仓库;
- (3) 提取、净化和加载数据。

3.2.1 确定用户需求

在大规模连锁超市的管理中, 各个门店的商品大部分是由配送中心来统一配送的, 配送的依据有两种, 一种是各门店根据销售情况和店内存放货品的能力向配送中心发出配货请求, 另一种是总部根据各个门店的动态库存情况确定需要配送的品种和数量。配

送中心根据门店所需商品的品种和数量, 均衡考虑各门店的请求制定配送计划, 将商品送往各门店。

门店保存有恰当数量的店存商品(以下简称店存), 既不影响销售又可以降低存储和运输的成本。对于连锁超市是至关重要的, 我们称这个数量为最佳店存。确定某种商品的最佳店存要考虑近期的销售情况和该类商品的季节性销售因素, 同时还要排除一些非正常的销售数据。但是由于商品分类和品种太多, 人工非常难于直观判定。

建立数据仓库的目的是通过分析门店的商品目录、销售数据, 并综合门店动态店存、配送中心库存等因素来决策最佳店存, 并在此基础上对连锁店和配送中心的存储、配送等活动进行分析和预测。

从不同的方向分析销售数据, 可以得到许多信息。例如分析销售时间, 可以发现特定商品在一年中某个特定时间段销售情况较好; 分析销售的商品, 可以发现哪种商品卖得好, 哪种商品卖得不好; 分析销售发生的门店, 可以比较各门店的经营状况等等。

由于商品品牌推广等销售活动的影响, 单独某一品种商品的销售情况并不说明什

么问题, 考察某一类商品的销售才有意义。因此在分析季节性因素对商品销售的影响时, 依据的数据应当是该商品分类的销售数据; 而在确定近期销售模型中则应着重考虑单品的销售。

根据以上的分析, 我们通过数据挖掘的方法建立一个按照商品类在时间维上的季节性模型和商品品种的近销售对照模型来作为确定最佳店存的决策支持数据。

3.2.2 设计和构建数据仓库

(1) 数据分析。建立数据仓库需要以下信息:

事实表—如销售量、销售额、交易量、店存等。

维—是事实信息的属性。如销售发生的时间、门店、销售了何种商品等。

粒度—是维度划分的单位。如时间单位可以是日, 也可以是月、季度或年。

(2) 设计数据仓库结构。数据仓库的数据模型有星型和雪花模型两种。星型模型由事实表和维表组成, 多个维表之间形成多维数据结构, 体现了空间的多维立方体, 较为直观, 便于用户安排不同的查询, 并可提高查询的性能, 所以通常采用星型模型表示立方体。

(3) 创建数据表和索引。创建事实表和维表后, 还要创建索引以加快信息检索和更新立方体的速度, 特别要针对经常进行的查询创建相应的索引。

3.2.3 抽取、净化和加载数据

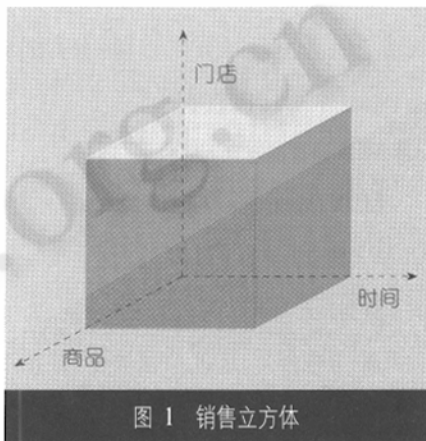
数据仓库的结构建立以后, 需要将事务处理系统中的数据经过数据验证、迁移、净化和转换, 加载到数据仓库中。

经过上述步骤, 数据仓库就建立起来了。

3.3 OLAP 立方体的建立

数据仓库中的 OLAP 立方体存储预先计算好的统计数据, 能够以极短的响应时间支持用户的动态分析, 是决定系统性能的关键因素。

下面以销售立方体 (Sales) 为例, 创建立





方体如图 1 所示。

销售立方体以商品日销售表作为唯一的事实表, 并三个维度分别是时间维、商品维和门店维。

事实表是存储真实数据即事实的地方, 企业的销售信息主要以商品日销售表来记录, 它记录了每一次交易发生的时间、地点、商品销售的品种、数量、金额以及其他基本销售信息, 时间维以年、季度、月、日作为维的粒度级别, 商品维以商品大分类、商品细分类和商品(单品)作为维的粒度级别。

3.4 MDX 语言实现立方体查询

MDX (Multidimensional Expressions) 是 OLAP 多维立方体查询语言, 在功能上类似于关系数据库查询语言 SQL, 它语法丰富, 执行效率高, 功能强大, 是 OLAP 服务器与外界交互的专用语言。

下面, 结合实例, 用 MDX 语句实现对立方体的多维分析。

(1) 查询某商品的销售走势, 其 MDX

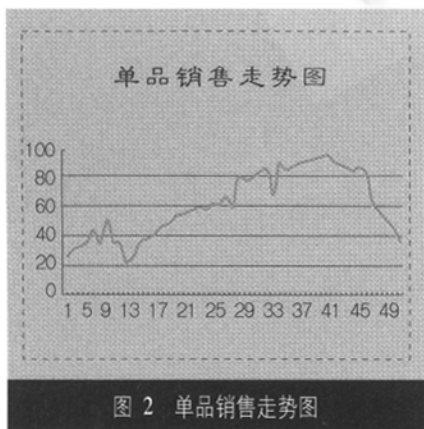


图 2 单品销售走势图

语句为:

```
WITH
Member [Measures].[Xssl per week] AS
SUM(WTD(),[Measures].[Xssl])
SELECT
{[Time].[Week].members} on columns,
{[Measures].[Xssl per week]} on rows
FROM Sales
```

WHERE {[Product] [003].[0030202].[1356006]}其中, [Measures].[Xssl] 表示销售数量, [Measures].[Xssl per week] 表示一周总的销售数量, 查询结果如图 2 所示。

(2) 查询不同门店某商品库存走势, 其

```
MDX 语句为:
SELECT
{[Time].[All Time].[2001].[Jan]:[Time].[All Time].[2001] [Dec]} on columns,
{[Stores].[All Stores].[001]:[Stores].[All Stores].[010]} on rows,
{[Measures].[mdkc]} on pages
```

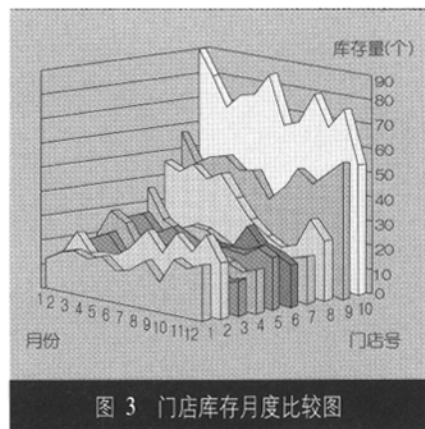


图 3 门店库存月度比较图

FROM Salse

```
WHERE ([Product].[AllProduct].[002].
[0020201].[6300786])
```

在客户端就可将查询结果展开为三维网络, 如图 3 所示。

4 总结

连锁超市配送决策支持系统充分利用企业已有数据资源, 运用数据挖掘思想, 结合 OLAP、ADO MD 等手段, 实现配送计划的科学制定与自动维护。

利用 OLAP 手段对商品的销售和店存进行分析, 其结果是建立在海量数据的基础上的, 与使用传统的数据库查询相比, 效率得到大大提高, 同时这种基于海量数据的查询, 从很大程度上排除了单个非法数据的干扰, 可以有效的体现出商品销售和店存的季节因素、节日因素以及消费趋向, 使决策更加切合实际, 更具有实用性。 ■



参考文献

- 1 W.H. Inmon, 数据仓库 [M], 机械工业出版社, 2000.
- 2 (美) Erik Thomsen, George Spofford, Microsoft OLAP 解决方案 [M], 潇湘工作室译, 人民邮电出版社出版, 2000.
- 3 Microsoft SQL Server 7.0 联机帮助.
- 4 王珊等, 数据仓库技术与联机分析处理, 科学出版社, 1998.5.