

文章编号: 1000-582X(2002)09-0146-04

汽车销售分析与预测系统的设计与开发*

李勇, 肖智, 陈玲

(重庆大学工商管理学院, 重庆 400044)

摘要: 市场分析和预测现在已成为了企业重要的决策依据和手段, 而有效地存储和利用已有的销售数据是进行分析和预测的基础。数据仓库和联机分析技术提供了强大的数据存储和数据集成的功能, 分析预测等数学方法提供了一系列的数据统计分析工具。将信息技术和数学方法结合起来应用于销售数据的分析与预测领域, 提出了一个基于数据仓库决策支持应用子系统的设计和实施方案, 展示出构建数据仓库技术平台和 OLAP 多维数据的过程, 并介绍了分析和预测中所采用的统计预测方法, 体现了先进的信息技术与决策科学的良好结合。

关键词: 数据仓库; OLAP; 决策支持; 分析预测方法

中图分类号: F270.7

文献标识码: A

在当今日益激烈的竞争环境下, 企业的决策者已不仅仅满足与对数据的简单维护和查询, 而是更希望能够有效地对变化的商业环境进行分析, 既要求有反映市场当前状态的信息, 还要求有反映历史及趋势的信息。在这种外在需求的牵引下, 一方面, 建立一个企业的决策支持系统成为了企业信息化的一种必然趋势; 而另一方面, 随着数据库技术的不断发展, 数据仓库的出现为决策支持工具的充分利用提供了基础平台, 也为企业的管理信息化、决策科学化提供了可行的技术基础。

W. H. Inmon^[1]对数据仓库的定义为: 数据仓库是支持管理决策过程的、面向主题的、集成的、时变的、持久的数据集合。从企业管理角度, 数据仓库是一种管理技术, 它与决策支持系统的目标用户都是企业的中高层领导, 执行的都是决策和趋势分析类的应用。从功能角度, 数据仓库着重于数据的集成和存储, 它将分散的商业数据集成到一起, 把支持决策分析的数据事先收集、归纳、处理, 使企业的业务操作环境和信息分析环境分离开来; 决策支持系统着重于运用统计分析方法进行数据的分析和信息的抽取。而决策支持系统中一些传统的统计分析方法及一些智能决策技术是可以集成到数据仓库中去的, 使数据仓库的决策分析能力得以大大地提高, 形成了基于数据仓库的决策支持系统。

以建立一个决策支持应用为线索, 从后台数据仓库以及中间 OLAP 多维数据的建立, 前端用户操作界面和分析预测方法的使用 3 个方面对汽车销售管理系统中分析与预测子系统的设计与开发实例进行描述。

1 数据仓库和 OLAP 多维数据的创建

子系统采用的是 OLAP 应用的典型模式, 在总体上可以分为 3 个组成部分: 后台数据仓库、中间 OLAP 多维数据、前端用户操作界面。其结构如图 1 所示:

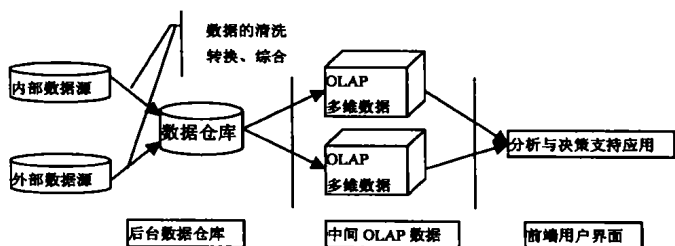


图 1 OLAP 应用的典型模式结构图

1.1 数据仓库

数据仓库是决策支持应用建立的基础, 它决定和限制着该应用的实现。它的建立一般为如下几步:

1) 确定信息需求

汽车销售的分析和预测的信息需求主要来自于: 汽车信息(汽车固有属性、汽车销售信息)、客户信息(客户固有属性、客户购买信息)。此外, 还有一些辅助

* 收稿日期: 2002-01-09

作者简介: 李勇(1967-), 男, 四川德阳人, 讲师, 重庆大学博士生。研究方向为商务智能、信息系统与决策支持系统。

信息,如:供应商的信息、保险公司的信息、宏观经济信息等^[2]。

2) 数据仓库结构

采用星形模式,将上述两方面的信息需求作为数据仓库中的两个主题,分别对其设计和建立事实表和多个维度表^[3]。

主题的星形模式如图 2 所示。

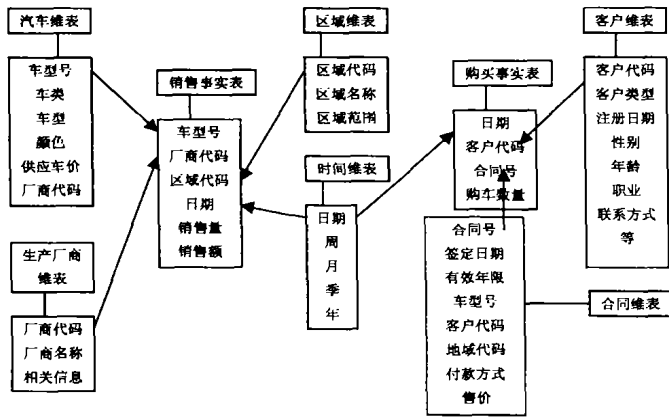


图 2 数据仓库的星形构架

3) 数据加载

当数据仓库结构建立以后就可以从事务处理数据库中加载数据了,通常有简单堆积、轮转综合、快照简化 3 种方法。针对汽车销售信息的分析和预测的需求频率不象证券行业要求那么高,采用以周、月、季、年为单位的轮转综合的方法来处理和加载数据。在应用中相应还提供了预处理提醒和积累处理的功能。

经过了上述过程,就建立了一个可用于汽车销售分析与预测的数据仓库。

1.2 OLAP 多维数据

OLAP 多维数据是针对具体应用的需要,将数据仓库中的细节数据按照具体应用的要求进行综合,并存储起来,用来快速响应和支持用户的动态分析;它体现

了 OLAP 应用的性能好坏。OLAP 多维数据的事实信息和维度信息是与数据仓库中的事实表和维度表相对应的,即 OLAP 多维数据是综合后的数据仓库中多个表的镜像^[4]。

1) 构造 OLAP 多维数据

OLAP 多维数据面向 OLAP 应用反映 OLAP 应用的数据组织模式。在子系统中 OLAP 应用体现为支持销售数据的分析与预测,在大多数的多维数据中都包含有时间这一维度。所使用的变量是销售量、销售额、合同份数。在比较分析中,对销售量和销售额这两个变量使用了[车类 + 时间]、[车类 + 车型 + 时间]、[区域 + 车类 + 时间]3 个多维数据;合同份数变量使用了[付款状态 + 时间]、[结帐状态 + 时间]两个多维数据。在相关分析中,对销售量和销售额这两个变量使用了[车价 + 车类 + 时间]、[车类 + 车类 + 时间]多维数据。

2) OLAP 多维数据的维护

多维数据的维护一部分体现于数据仓库的数据维护中,另一部分则是 OLAP 应用独立的维护方式。常用有 3 种维护方式:完全更新方式、完全刷新方式、增量刷新方式。增量刷新的方式在计算量和有效性上为最佳方式,但是其实现难度较大,基本思路为对应数据仓库中改变的记录来改变原有的多维数据中与之相关的结果,怎样来标识改变的数据与多维数据之间的对应关系是这种方式的关键^[5]。

2 前端用户界面

数据仓库和中间的 OLAP 多维数据建立好以后,在 SQL SERVER 2000 环境下利用 Powerbuilder 进行前端用户界面的开发,构造了一个直观、方便的操作环境。分析界面见图 3、图 4。

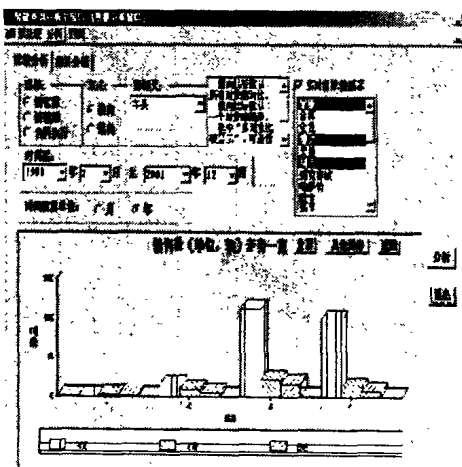


图 3 分析—比较分析窗口

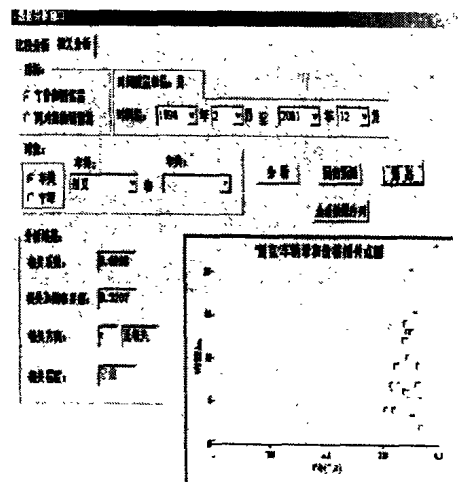


图 4 分析—相关分析窗

3 分析和预测的方法

数据仓库的建立使得传统的统计预测方法得以良好应用,在分析模块中使用了比较分析、相关分析、回归分析3种方法;在预测模块中使用了线性回归预测、二次指数平滑预测、季节预测3种方法。这几种分析和预测的方法都能直观地反映出销售数据的历史和趋势,为决策的制定提供了客观的数据依据,提高了决策的科学性。现以上述的预测方法为例来说明。

1) 方法的选择及特点

销售数据的预测是以时间序列预测为核心,因此以时间序列预测方法为核心,将影响预测目标变化的一切因素由‘时间’综合起来描述。选择指数平滑和季节预测这两种时间序列预测的方法,代表性地反映长期趋势和季节变动这两类因素的影响。而且在具体实践过程中发现指数平滑方法的长期误差较小,具有较高的可靠性。因此,使用二次指数平滑方法和季节预测方法来进行预测。而线性回归主要用于确定相关因素之间的函数关系,从而通过给定自变量的值来得到预测目标的值,由于它对于时间来说仅仅体现的是随时间的直线形的上升或下降趋势,因此在这里我们将此方法作为一般性的趋势观察来使用^[6]。

2) 方法的逻辑模型^[7]

◆ 线性回归

- 建立一元线性回归模型: $y_t = a + b * T$
- 检验相关系数的显著性 $R = (\sum xy - nxy) / (\sqrt{\sum x^2 - nx^2} * \sqrt{\sum y^2 - ny^2})$
- 标准误差: $s_y = \sqrt{\frac{\sum y^2 - a\sum y - b\sum xy}{n - 2}}$
- 点估计值: $y_t = a + b * T$
- 预测区间值: $[\hat{y}_t \pm t_{(\alpha/2), (n-2)} * s_y * \sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{n\sum x^2 - (\sum x)^2}}]$

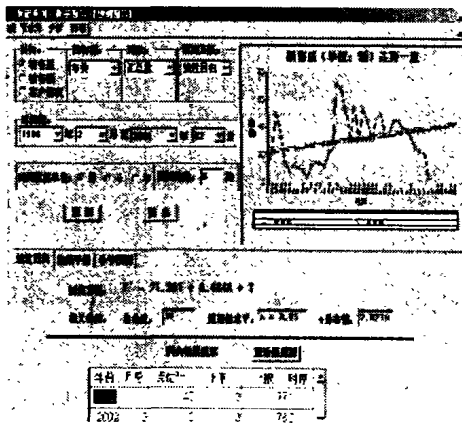


图5 线性回归预测窗口图

◆ 二次指数平滑

- 一次平滑: $s_t^{(1)} = \alpha y_t + (1 - \alpha) s_{t-1}^{(1)}$
- 二次平滑: $s_t^{(2)} = \alpha s_t^{(1)} + (1 - \alpha) s_{t-1}^{(2)}$

$$\cdot \text{直线趋势模型: } \begin{cases} \hat{y}_{t+T} = a_t + b_t T \\ a_t = 2s_t^{(1)} - s_t^{(2)} \\ b_t = \frac{\alpha}{1 - \alpha} (s_t^{(1)} - s_t^{(2)}) \end{cases}$$

$$\cdot \text{均方误差: } MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2$$

◆ 季节预测

$$\cdot \text{季节系数: } \begin{cases} \bar{x} = \frac{1}{t * 4} \sum_{i=1}^t \sum_{j=1}^4 x_{ij} \\ \bar{x}_j = \frac{1}{t} \sum_{i=1}^t x_{ij} \\ \beta_j = \frac{\bar{x}_j}{\bar{x}} \end{cases}$$

$$\cdot \text{预测年值: } \hat{y}_{t+1} = \frac{y_1 + 2y_2 + \dots + ny_t}{\sum_{i=1}^t i}$$

$$\cdot \text{平均年值: } \hat{y}_{t+1,j} = \frac{\hat{y}_{t+1,j}}{4}$$

$$\cdot \text{修正平均年值得到不同的季节值: } \hat{y}_{t+1,j} = \beta_j \hat{y}_{t+1,j}$$

3) 结果比较

在分别通过线性回归、二次指数平滑、季节预测3种方法的预测后,对比3个预测结果以及每种方法的误差,发现对于汽车销售数据的预测,反映长期趋势的指数平滑方法取得了很好的效果,并且季节预测方法也反映出本实例的销售呈一、二季高、三、四季低的季节变动趋势;而线性回归由于模型本身的限制只能观察到一般的趋势,预测结果的误差较大。预测的结果如图5、图6、图7所示。

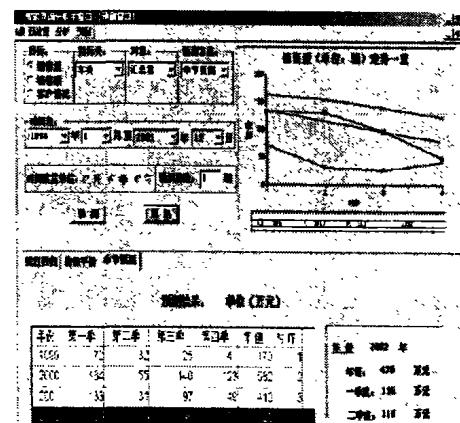


图6 季节预测图

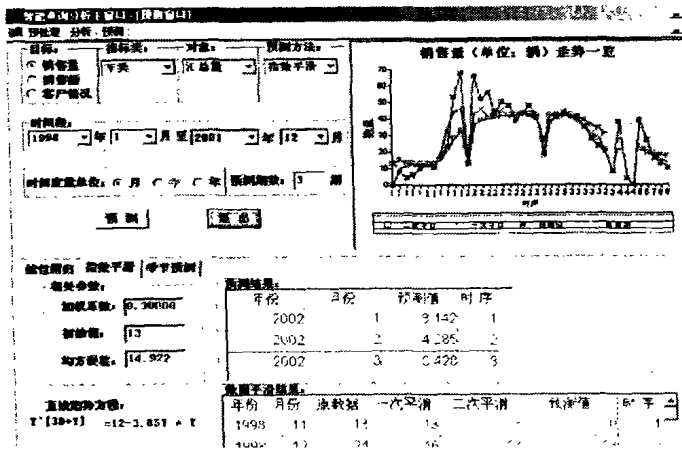


图 7 二次指数平滑预测图

4 结束语

在实践开发过程中由数据仓库途径改造信息系统并开发 DSS 的过程从设计思想到实现细节上都充满了挑战性的问题。但是应该看到的是先进的信息技术与决策科学相结合来开发 DSS 是一种必然的趋势,对比过去‘拍脑袋’的主观决策来说无疑是更加的客观和

科学,必将受到企业决策者们越来越多的重视,这也是企业管理信息化、决策科学化的必然要求;并且这种结合将随着信息技术的不断发展而更加紧密。

参考文献:

- [1] INMON W H. Building the Data Warehouse[M]. 北京:机械工业出版社,1993.
- [2] 张基温. 信息系统开发案例[M]. 北京:清华大学出版社,1999.
- [3] INMON W H. Building the Data Warehouse, Second Edition [M]. 北京:机械工业出版社,1996.
- [4] 王珊. 数据仓库技术与联机分析处理[M]. 北京:科学出版社,1998.
- [5] 沈兆阳. Microsoft SQL Server 2000 OLAP 解决方案——数据仓库与 Analysis Services[M]. 北京:清华大学出版社,2001.
- [6] 倪加勋,袁卫,易丹辉等. 应用统计学[M]. 北京:中国人民大学出版社,1999.
- [7] 暴奉贤,陈宏立. 经济预测与决策方法[M]. 广州:暨南大学出版社,2000.

Design and Development of Analysis and Forecast System for the Distribution of Car

LI Yong, XIAO Zhi, CHEN Ling

(College of Business Administration, Chongqing University, Chongqing 400044, China)

Abstract: Analysis and forecast for market have become important methods of making decision in most enterprises. It is the base of analysis and forecast to effectively store and make full use of the data in the distribution. Technology of data warehouse and OLAP provides the strong function of storing and integrating data, and mathematics methods provides a series of statistics and analysis tools. Combine IT and mathematics methods in the field of analysis and forecast for the data of distribution to put forward a design and implementation plan of application system in supporting decision based on data warehouse. The design exhibits the processes how to build data warehouse and multi - dimensional data, recommends the methods of analyzing and forecasting, and shows a good integration of advanced information technology and decision science.

Key words: data warehouse; OLAP; decision support

(责任编辑 刘道芬)