

基于 Agent 的商务智能模型研究

徐 娜¹ 赵文耘² 彭 鑫²

¹ (复旦大学软件学院 上海 200433) ² (复旦大学计算机科学与工程系 上海 200433)

摘 要 针对传统的商务智能系统存在的被动性和集成困难的缺点,提出一个基于 Agent 的商务智能模型,可利用 Agent 建立和维护数据仓库,并进行数据挖掘。给出了其主动监控商务环境原理,以及对推理规则的定义,从中找到决策支持信息,最后结合了一个出版管理系统的应用进行了实践。

关键词 商务智能 智能代理 数据挖掘

RESEARCH ON BUSINESS INTELLIGENCE MODEL BASED ON AGENT

Xu Na¹ Zhao Wenyun² Pen Xin²

¹ (Software School, Fudan University, Shanghai 200433, China)

² (Computer Science and Engineering Department, Fudan University, Shanghai 200433, China)

Abstract To improve the disadvantages of traditional business intelligence systems such as passivism and hard to integrate, an Agent-based business intelligence model is given. Agents are used for data warehouse establishing and maintaining and data mining as well. The definition of reasoning rules is introduced. How the proposed model monitors the business environment actively based on these rules, and how to find out decision supported information are analyzed. At last the model is carried out through a publication management system.

Keywords Business Intelligence Agent Data Mining

0 引 言

20 世纪 80 年代随着企业信息化的深入和企业数据的增加,商务智能技术应运而生。1989 年 Gartner Group 提出了商务智能 (business intelligence, 简称 BI) 概念,认为“BI 是将数据转换成信息的过程,然后通过发现将信息转化为知识”^[1],随后 IBM 在 2000 年正式推出了一项商务智能的企业级启动计划。如今 BI 技术被广泛应用到电信、保险、医疗等行业。很多财富 500 强公司意识到商务智能的重要性并建立了商务智能数据库。

目前学术界对于商务智能没有统一的定义。笔者认为,商务智能通过运用数据仓库 (data warehouse, 简称 DW)、联机分析处理 (Online Analytical Processing, 简称 OLAP) 和数据挖掘 (data mining, 简称 DM) 技术来对业务数据进行分析,从中得到支持业务决策的有用信息,亦即用以对业务决策进行支撑。

典型的 BI 系统有 Oracle BI, Brio, Cognos, BO (Business Objects) 等,学术界一些学者包括 Simon Fong^[2], Jun-Jang Jeng^[3] 等人也提出了商务智能模型。这些系统主要特点是:

- 以数据仓库为核心,对商业信息进行提取、转换、加工 (ETL) 后放入数据仓库中。
- 利用数据挖掘和 OLAP 等技术对数据仓库中的信息进行分析处理,并根据用户不同的分析要求表现出来,形成图表或表格。

但以上系统在主动性和系统集成方面存在以下缺陷:

a) 被动地分析报表,不能积极主动地监督商业环境变化并作出反应,而飞速增长的信息量和日趋激烈的竞争要求企业对信息作出反应的时间必须尽可能缩短^[4]。

b) 需要与已有的商务系统集成,才能对数据进行采集和分析,而实际商务环境中现有系统的复杂多样性使得集成难以实施。

为了弥补以上现有成果的不足,本文提出在商务智能模型中利用 Agent 建立、维护数据仓库,当商务环境发生变化时,根据规则进行推理分析,采取合适的响应措施。

近几年兴起的 Agent 技术,由于其自治性、响应性、主动性等特点^[5],被广泛应用于网络管理、动态路由、电子商务、远程教育系统、工业以及医药卫生等方面^[6]。简单说来,Agent 是一个存在于特定环境的软件对象。但是不同于一般软件对象,Agent 具有智能性,能够代表用户进行决策。Agent 的基本特性包括^[7]:自治性、响应性、主动性、社交性。此外,Agent 还有类似于人类的精神化概念,如知识、信念、意图、义务。Agent 的智能主要依靠这些概念实现。文献 [8] 的研究认为,Agent 的上述特性可以帮助进行数据挖掘,以及发现、定位、向用户报告最有价值信息。

数据仓库是 BI 的基础,是“面向主题的、集成的、时变的、非易失的数据集合,用于管理决策指定过程”^[9]。

收稿日期:2005-09-20。国家 863 计划 (2004AA112070, 2004AA113030),国家自然科学基金 (60473061)。徐娜,硕士,主研领域:软件工程,Agent。

1 基于 Agent的商务智能模型

基于 Agent的商务智能模型如图 1所示。

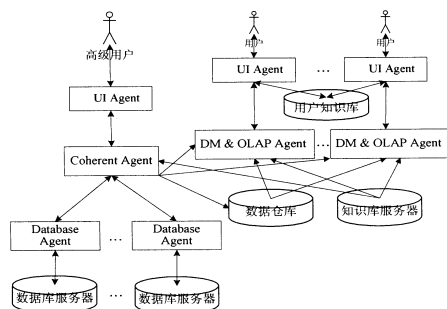


图 1 基于多 Agent的商务智能模型

a) 用户

普通的决策用户,主要是指那些需要利用商务智能系统对数据进行分析以获得决策支持的人员。

b) 高级用户

高级用户对整体各数据库结构有一定的了解,熟知商务环境。高级用户收集和分析业务需求,据此建立数据模型,完成数据仓库的物理和结构的设计,并负责建立规则库。

c) 用户知识库

存放用户描述需求的语言和本行业领域内专业术语对应关系等相关知识,同时用户知识库维护一个用户特征表,用以记录用户的喜恶主题,以及用户对该主题的访问频度和预订信息。

d) 数据库服务器

以一定结构存储具体的、未经处理的商业数据。各数据库服务器可能是异质异构的。

e) 数据仓库服务器

由高级用户利用 Coherent Agent建立,供 DM&OLAP Agent查询分析,由 Coherent Agent负责其更新和维护。

f) 知识库服务器

知识库服务器存储规则库以及领域知识。其中领域知识由领域专家建立,描述该领域中的相关概念和这些概念之间关系的概念模型,以解决该行业领域中的术语和词汇的二义性,实现知识的理解和共享。规则库是由高级用户设定的 ECA(事件-条件-动作 event-condition-action)规则库^[10],Coherent Agent利用规则库进行匹配或者推理决定商业环境发生变化后采取的行为。

g) 用户 Agent

驻留在数据仓库服务器上的 UI Agent负责用户与 Coherent Agent、DM&OLAP Agent的交互工作,它可以以友好的界面与用户交互,也可以让用户预订偏好的主题以定期自动发出该主题的分析请求并向用户报告结果。

普通用户用特定语言描述所分析的高层次目标,UI Agent获得用户需求后访问用户知识库,将目标转化成用专业术语描述的分析目标,并告知 DM&OLAP Agent。当后者返回分析结果后,UI Agent将这些分析结果可视化地呈现给用户。高级用户用特定的数据库脚本语言详细描述设计要求,UI Agent将设计要求告知 Coherent Agent,后者直接执行该脚本完成数据仓库的设计。

h) 数据挖掘 & 联机分析 Agent

驻留在数据仓库服务器上。它了解数据仓库的结构信息,通过对知识库的查询,将用户的需求描述转化成低层次的挖掘任务,进行数据分析和挖掘。该 Agent可以自己将用户的需求转化成低层次的特定查询语言的合法语句来完成数据分析和挖掘功能,也可以调用环境中的 DM工具或者 OLAP工具。

i) 协同 Agent

这是整个系统的关键 Agent,驻留在数据仓库服务器上,它负责建立和维护数据仓库,协调各 Database Agent的行为。

j) 数据库 Agent

驻留在数据库服务器上,了解本地数据库的存储和结构信息。负责本地数据库的查询和监控,如果本地数据库有变化负责向 Coherent Agent报告。

2 商务智能模型实现的关键技术

2.1 ECA 规则

本模型中的规则用四元组 $R = (D, E, C, A)$ 描述存储,其中 D 为规则的唯一标记, E 为发生的事件, C 表示当事件 E 发生时触发 A 必须满足的条件, A 为执行的操作。为了便于解析执行, ECA 规则的事件 E 以库名 + ' _ ' + 表名 + ' _ ' + 键值 + ' _ ' + 字段名 + ' _ ' + 变化类型 (New/Delete/Increase/Decrease) 的格式定义,条件 C 以库名 + ' _ ' + 表名 + ' _ ' + 字段名及其数学运算符来组合定义。事件 E 或者条件 C 中凡是参数部分均以 % 表示,如某规则的事件 E 中键值为参数,则该规则的事件 E 为库名_表名_%键值_%字段名_变化类型。事件 E 中的参数在规则查询时可以忽略,而条件 C 中的参数在验证条件时必须进行赋值。涉及到数据仓库的执行操作 A 以数据库脚本语言的形式描述,只要在规则中指定该脚本的路径即可被 Agent 读取并执行。

规则库支持按照 D 或者 E 检索。当事件发生时, Agent 将首先根据事件检索规则库,查询结果可能有多个规则, Agent 读入这些规则并进行解析执行。

算法 1 规则执行算法

```
ArrayRule = QueryRuleByEvent(E);
for(int i=0; i<ArrayRule.length; i++)
{
    if(Verify(ArrayRule[i].C))
    {
        Do(ArrayRule[i].A);
    }
}
```

2.2 协同 Agent 的处理逻辑

协同 Agent 的功能描述如下:

a) 它维护描述各数据库服务器信息以及各 Database Agent 信息的数据库表。当系统中有新的数据库服务器加入时,为该服务器分配一个 Database Agent 驻留在该数据库服务器上。

b) 由于 Coherent Agent 和 Database Agent 知道对方所处的站点位置,所以各 Database Agent 可以直接与 Coherent Agent 通信,同时各 Database Agent 之间也可以通过 Coherent Agent 相互通信。

c) 在高级用户提供必备的数据仓库的设计信息后,与各数

数据库服务器上的 Database Agent 交互,通过 Database Agent 完成该地数据库的数据采集工作,以建立数据仓库。

d) 当数据库有变化时 Coherent Agent 将评估变化,并根据高级用户设定的规则采取响应措施。

e) 必要时,Coherent Agent 要协调 Database Agent 的行为,解决数据定义和表示的冲突。

为了实现 Coherent Agent 对商务环境变化的响应,本文设计其结构如图 2 所示。

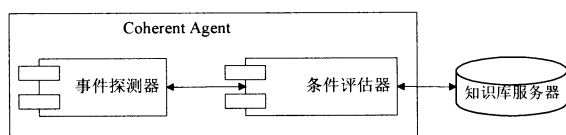


图 2 Coherent Agent 结构图

驻留在数据服务器上的 Database Agent 监控本地数据库的数据变化,并向 Coherent Agent 报告变化,报告内容包括发生数据变化的库名、表名、变化的记录、字段名、原值以及变化后的值,变化类型。事件探测器主要负责判断从 Database Agent 接收到的消息是否为数据库变化的报告,如果是的话,事件探测器将调用条件评估器。后者从报告的内容中根据规则库中事件的格式提取出事件,根据事件查询位于知识服务器上的 ECA 规则库得出数条规则,推理分析后决定采取何种措施响应变化。

2.3 数据仓库

在本文提出的模型中,建立数据仓库过程如下:

a) 了解应用数据的高级用户对数据仓库进行结构设计,用数据库脚本语言描述设计要求,要求中必须指定数据仓库的多维模型存在形式,比如星形模式、雪花模式或事实模式^[5]。同时高级用户可以给出多维模型中每个维表的概念分层。例如存在一个“时间”维,那么可以把这个维按照年-月-日层次分层,也可以按照年-季-月-日层次分层^[11]。UI Agent 获得用户要求后,与 Coherent Agent 交互,将设计要求告知 Coherent Agent。

b) Coherent Agent 与所有的 Database Agent 交互,由 Database Agent 自身或者调用环境中的 ETL (Extract Transform Load) 工具,从所在的数据库服务器中转换、清除和提取数据。

在 Coherent Agent 的协调下,各个 Database Agent 向数据仓库中输入数据,完成数据输送过程。

2.4 商务智能的实现

本模型在实现传统 BI 系统被动进行数据分析的基础上,进一步实现了对商业环境的主动监控。

当用户连接到系统时,系统中的 Coherent Agent 将为用户分配一个 UI Agent 和一个 DM&OLAP Agent。用户提出的高层次目标经由 UI Agent 传送给 DM&OLAP Agent。接到目标后,DM&OLAP Agent 运用领域知识,将其转化成一系列的低层次的数据分析和挖掘任务,据此选择实施数据挖掘方法和算法(聚类分析、关联分析、时间序列预测、决策树)或者调用环境中的分析挖掘工具,然后生成分析结果返回给 UI Agent,由 UI Agent 以报表、图形等方式可视化地呈现给用户。

模型的智能决策框架如图 3 所示。当商业环境有变化时,将引起商业度量数据的改变,具体体现在系统中某个数据库的数据变化上。监控到数据变化之后,当地的 Database Agent 与 Coherent Agent 通信,报告变化情况。当 Coherent Agent 的事件探测器收到数据变化的报告时,根据 ECA 规则库结合数据仓库

中的数据进行推理和分析,并采取响应措施,如更新数据仓库或者向用户报告,从而实现了商业环境的主动监控。同时通知所有的 DM&OLAP Agent 重新进行数据分析和挖掘。

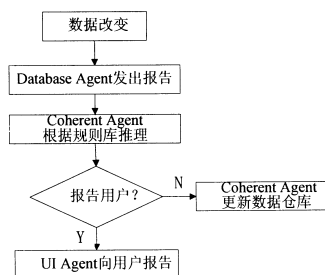


图 3 智能决策框架

3 应用实例

目前,上海和江苏等地出版业广泛应用的一个管理信息系统——复旦天翼出版管理信息系统,由 3 大部分组成,各部分分别由数个模块构成,各模块可以根据需要动态增加或者减少。每个模块有自己独立的数据库,同时也可能需要从其他模块的数据库中读取相关数据。

3.1 系统架构

该系统主要模块如图 3 所示。出版部分的模块主要负责出版流程管理、纸张材料管理、作者稿酬管理以及出版印刷工程管理等功能;发行部分的模块主要负责产品信息管理、客户信息管理和业务管理等功能;财务部分的模块主要负责出版社的财务管理、产品的成本管理和固定资产管理等功能。

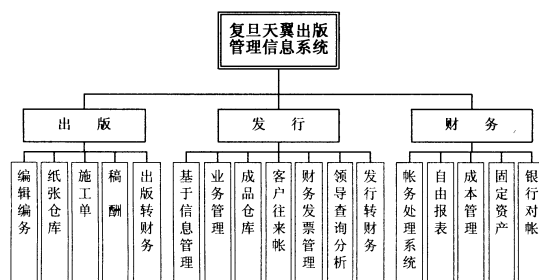


图 4 复旦天翼出版管理系统功能模块图

各个模块之间是相互关联的,并且在业务上相互影响。每当有一笔销售业务发生,销售人员在管理系统中录入业务数据。由于销售信息的改变直接影响成品库存的调整、利润增减和出版计划的制定,所以用户希望根据销售数据对库存和盈利状况等信息进行实时分析。

3.2 商务智能实现

基于以上提取的各个数据库,我们建立了一个数据仓库。该仓库采用星形模式多维模型,如图 5 所示。该多维模型以销售表作为事实表,维度表包括:作者信息表、书籍类别表、库存表、客户信息表等等。

为了实现 Agent 的智能推理,我们建立了一个 ECA 规则库,其中有两个销售数据改变的推导规则:

(1, SaleBase_SaleBill_% Book D % _Sale_Increase, StorageBase_Storage_% Book D % _Totalstorage <= StorageBase_Storage_% Book D % _baseline, Notifypanner)

(下转第 64 页)

3 结束语

在医学诊断及研究中,从 CT机上得到 DICOM 图像文件,要在目前计算机操作平台上直接进行处理与操作,就必须把它转换成计算机能处理的位图文件(*.BMP)。但是,在实际的应用中,如果涉及到网络使用问题,还需考虑 DICOM 文件的网络传输协议,本文未涉及到该内容。

参 考 文 献

- [1] Zhao Mingchang, Tian Jie, He Huiguang, Li Guangning A Surface Reconstruction Algorithm for High Resolution Medical Images, Proceedings of International Conferences on CAD&CG2001, International Academic Publishers, Kunming, China, 2001, 1: 286 - 290.
- [2] 杰·亚历山大,等. CT技术及其临床应用. 汪臻,顾英,译. 1988.
- [3] 侯自强. 医学诊断数字影像技术. 1994.

(上接第 15 页)

(2, SaleBase_SaleBill_% BookD%_Sale_Increase, StorageBase_Storage_% BookD%_TotalStorage > StorageBase_Storage_% BookD%_baseline,
D:\datawarehouse\script\Redeploystorage.txt)

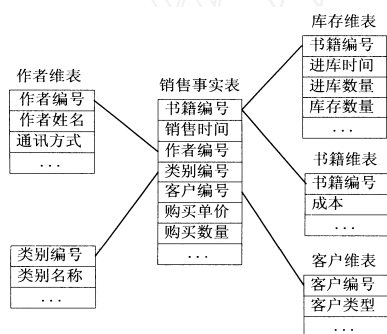


图 5 星形模式多维模型

规则 1 的含义是当销售业务发生即某本书销售量增加后,系统检查库存判断该成品在各个仓库中的总库存是否低于警戒线,如果低的话,通知计划者制定该书出版计划;规则 2 的含义是如果其他仓库有多余库存的话,执行 Redeploystorage.txt 脚本,调配数量。

建立数据仓库和本体知识库后,利用商务智能对该数据仓库的多维实时分析,我们可以分析挖掘出有效的决策支持信息。商务智能在出版管理系统中的应用主要有以下几个方面:

a) 库存分析 分析成品库存情况,及时调整库存或制定出版计划。销售数据变动后,系统将分析目前的各产品的库存,对照用户预先设定的警戒线,并根据本体知识库服务器上的规则库进行推理,以决定在各个成品仓库之间调整成品数量或者向出版计划制定者建议出版品种和数量。

b) 利润分析 分析销售和成本数据。比如通过分析作者与所写书籍的销量、价格和成本等信息的数量关系,可以得到创造最多利润的作者,从而采取邀稿形式主动与作者合作。

c) 客户分类和特点分析 通过分析客户的信息、购买的书籍数量和类别,可以得出各类型客户偏好采购的书籍类别,从而按照结果有针对性地进行新产品介绍和推广。

d) 市场营销策略制定 采用商务流通领域回归分析预测

系统的回归模型^[12]对销售数量与外界因素如价格、稳定客户数量等相关因素进行分析,对销售情况进行预测。比如通过分析往年季度与各类型书籍的销量的数量关系,可以预测下季度销售较多的书籍,并根据这些书籍的当前库存量制定本季度的生产计划。

4 结 论

本文提出的商务智能架构利用 Agent 对各个数据库的实时监控来感知商务环境的变化,由于企业中商务信息大部分是以数据库的方式存储,所以该架构易于和商务环境中的现有系统集成,具备良好的通用性;该架构除了能满足用户分析数据的需求之外还能主动维护数据仓库,实时监控数据的变化,并根据规则进行推理,以对数据变化采取相应措施,从而实现了主动性。本文在出版管理信息系统中应用了提出的商务模型,证明了该模型的合理性。相对于传统的 B 系统,本模型可以在各种商务环境中快速实施,提高决策的及时性和准确性,从而提高了企业的效率。随着 Agent 技术的成熟,本文提出的框架有很强的理论和现实意义。

未来的工作将致力于如何实现商务智能系统对文档和图像等非结构化商务信息的分析,以实现对企业信息资产的充分利用。

参 考 文 献

- [1] BM. 2001 年 BM BI & e-Commerce 暑期研讲班讲义汇编 (1 - 2) [R]. 上海:复旦大学计算机科学与工程系, 2001.
- [2] IEEE International Conference on Data Mining, 2002. ICDM 2002. Proceedings 9 - 12 Dec, 2002: 759 - 762.
- [3] Jun Jang Jeng, Schiefer J, Cheang H. An agent-based architecture for analyzing business processes of real-time enterprises. Enterprise Distributed Object Computing Conference. Proceedings Seventh IEEE International 16 - 19 Sept, 2003: 86 - 97.
- [4] 姚卫新,黄丽华. 事件驱动式商务智能——商业竞争的新利器 [J]. 2004, 2: 59 - 61.
- [5] Wooldridge M, Jennings N R. Agent Theories, Architectures, and Languages: A Survey. In: Wooldridge M eds Proc of the ECAI-94 Workshop on Agent Theories, Architectures, and Languages Amsterdam, 1994. Berlin: Springer-Verlag, 1995: 1 - 22.
- [6] Wooldridge M, Jennings N R. Application of intelligent Agents. Agent Technology: Foundation, Application, and Markets, 1998.
- [7] Wooldridge M, Jennings N R. Intelligent Agents: Theory and Practice [J]. Knowledge Engineering Review, 1995, 10(2): 115 - 152.
- [8] Chi R, Turban E. Distributed Intelligent Executive System. Decision Support System, 1995, 14: 117 - 130.
- [9] JiaWeHan, Micheline Kambr. Data Mining Concepts and Techniques [M]. 北京:高等教育出版社, 2001.
- [10] Simon E, Dittrich A K. Promises and realities of active database systems. Proceedings of the 21th International Conference on Very Large Data Bases, 1995: 642 - 653.
- [11] 李泽海,赵吉贵,赵君. 商业智能技术及行业应用分析. 2003, 21 (1): 55 - 60.
- [12] 周春光,邢辉,徐振龙,王哲. 商业数据的预测模型及其算法研究. 2002, 20(3): 53 - 60.