

基于 OLAM 的可视化 DM 系统的设计与实现^{*}

李献礼

(长江师范学院教育技术与信息中心 重庆涪陵 408003)

摘要 为了解决传统数据挖掘的两大缺点,结合 OLAP 和 DM 以及可视化三大技术,提出了基于 OLAM 的可视化 DM 系统模型。在该系统中,OLAP 与 DM 做到了一定程度的有机结合,用户可以参与到整个挖掘过程中,同时对该模型进行了系统的分析。

关键词 图像数据,在线数据联机挖掘,数据挖掘,关联规则

Implementation and Designment of the DM Based on OLAM

LI Xian-li

(Center of Education Technology, Yangtze Normal University, Chongqing 408003, China)

Abstract For solving the two weaknesses of traditional data mining(DM), combine the OLAP and DM and visualization, put forward the DM system model which can be visualization based on OLAM, the OLAP and the DM attains the machine of certain degree to combine in that system, user can participate to the whole analysis during the DM process, last analyzed the system Model.

Keywords Image data, OLAM, DM, Correlation rule

1 问题的提出

OLAP 建立在数据多维视图的基础上,具有与用户交互和快速响应的特性,用户可参与分析过程,动态地提出分析要求、选择分析算法,对数据进行由浅及深的分析。DM 是从大量原始数据中抽取模式的一个处理过程,它着重强调分析数据的深入和分析过程的自动化,其分析过程不需要用户的参与。CLAP 与 DM 各有所长,如果能将二者结合起来建立一种基于 OLAM 的可视化 DM 系统模型,用户就可以参与到整个挖掘过程中去。

2 系统原型结构

一个基于 OLAM 的可视化挖掘系统应具有联机选择数据挖掘功能、交互式探索性的数据分析,增加可视化工具包,可以在任何程度上都进行挖掘功能,这样才能真正地做到 OLAP 与 DM 的有机结合。充分地考虑基于 OLAM 可视化数据挖掘系统的联机性、交互性和可视化展现,并结合普通数据挖掘系统的具体实现,我们提出了如图 1 的基于 OLAM 可视化数据挖掘系统的模型^[1-3]。

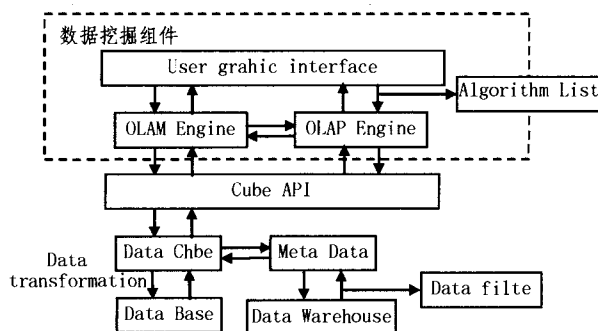


图 1 基于 OLAM 的 DM 可视化系统模型

从图中可以看到,OLAM 服务器通过用户图形接口接收用户的分析指令,在元数据的指导下,对超级立方体做一定的操作,然后将挖掘分析结果展现给用户,这个过程是动态的。从图 1 可知,该数据挖掘系统主要由数据挖掘组件、数据转换组件、过滤组件和显示组件组成,下面将分别讨论。

2.1 数据挖掘组件^[4-6]

数据挖掘组件主要由 OLAP 引擎、OLAM 引擎和用户图形接口三部分组成,完成对数据立方体的分析挖掘过程,并将挖掘出的信息输出到挖掘结果文件中保存,作为可视化展现的数据源。

(1) OLAP 引擎:其主要任务是计算用户的 OLAP 指令、对数据立方体中的数据进行快速分析处理和及时通过用户接口将分析出的信息返回给用户。它具有快速响应性、可分析性、多维性和信息性等特点。OLAP 除了具有钻取、切片、旋转等操作以外,还有创建数据立方体的能力。

(2) OLAM 引擎:OLAM 引擎是建立在 OLAP 技术基础之上的,是在 OLAP 技术中加入了数据挖掘过程。借助于 OLAP 对数据立方体进行切片、切块和钻取操作,OLAM 可以直接访问存储在底层数据库里的数据,能对任何它想要的数据进行挖掘。OLAM 引擎在数据立方体上执行的分析挖掘与 OLAP 引擎执行的在线分析处理的方式是相同的,但它执行多项数据挖掘任务,如概念描述、关联挖掘、分类、预测、聚类和时序分析等,通常是由多个集成的数据挖掘模型组成的,比 OLAP 要复杂得多,要求有比 OLAP 更强大的数据立方体构建和存取工具。

(3) 用户接口:是用户与 OLAM 引擎和 OLAP 引擎之间相互通讯的接口,主要完成向 OLAM 引擎和 OLAP 引擎发送指令、选择挖掘算法,以及接收从 OLAM 引擎和 OLAP 引擎返回的挖掘信息,然后负责将其送到挖掘结果文件中保存。通过用户接口,用户可指定相应的参数,选择挖掘算法,以加

^{*}重庆市教委科研项目(KJ071306)。李献礼 副教授,研究方向为计算机网络应用、数据挖掘。

快挖掘过程。

在数据挖掘组件中,允许多个 OLAM 引擎和 OLAP 引擎同时存在,允许多个不同的数据挖掘任务同时进行。因而用户在算法列表中选择算法时,可以同时选择多个挖掘任务,分别交给不同的 OLAM 和 OLAP 引擎去完成,这样大大地提高数据挖掘的效率。

2.2 数据转换组件

数据转换组件主要负责将这些数据转换成易于理解的直观的基本图形。从挖掘结果文件到图形文件,一般需要进行预处理和数据转换操作。

(1) 数据转换预处理

由于挖掘结果文件中的数据可能还比较多且复杂,为了减少数据转换的工作量,需要进行一些预处理工作。预处理方法比较多,如数据压缩、数据离散、聚类分析、标准化处理和概化处理等方法,依据不同的数据类型和不同的需要,可以选择不同的预处理方法。

(2) 数据转换

挖掘结果数据经过数据转换预处理后,就可通过数据转换方法将其转换成图形数据。数据转换方法放在数据转换函数库中,用户可以根据具体的需要选择数据转换方法。当然在该系统中允许用户向转换函数库添加、删除转换函数,这样就可以不断地丰富和更新转换函数库。对不同类型的数,应使用不同的转换方法,即使是同一类型的数据,在不同的应用中,使用的转换方法也可能不同。

映射是最常用的数据转换方法,是指将挖掘结果数据的属性分别与 X 坐标、Y 坐标、Z 坐标、颜色、亮度、模式等图形数据属性对应起来,从而达到将挖掘结果数据转换成图形数据的目的。需要注意的是,可视化挖掘结果数据的方式不是单一的,因而使用的数据转换方法也是多种多样的。

(3) 图形数据

图形数据是挖掘结果数据经数据转换方法而得到的,可以看作挖掘结果文件中数据的图形表示,它的属性有 X 坐标、Y 坐标、颜色、模式、大小、方向以及图的形状等。

实际上,图形数据并不是固定的,可以根据需要动态地改变。图形数据的大小可以随屏幕的尺寸而改变,图形的颜色可任意选取,其填充模式也是多种多样,用户可以从中选择。在显示图形时,图形的方向可以向上,也可以向下,可以向左,也可以向右;其形状就更是千变万化,可以是矩形、三角形、饼图、圆柱、圆锥、线、点等等。

2.3 过滤器组件

经过数据转换后得到的图形数据量可能还比较大,不便于在屏幕窗口中显示,还必须经过一个过滤过程,以便在视图区域中显示用户感兴趣数据及其属性。在过滤组件中,可视

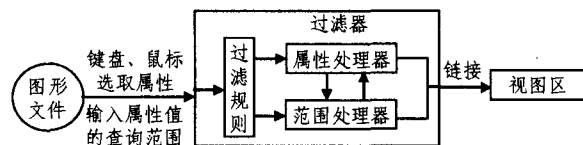


图2 可视过滤器结构示意图

化过滤器是关键组件,它定义一个基于图形数据的查询,可选择图形数据的属性和指定图形数据的某一部分或一个查询范围。当执行可视化过滤器后,将需要展现的图形数据送到视图区域中。可视化过滤器的结构如图2所示。

由图2可知,可视化过滤器主要由过滤规则、范围处理器和属性处理器三部分组成,其中过滤规则中主要存放的是对图形数据过滤时应遵守的规则,这些规则允许用户添加、修改和删除。图形数据通过过滤规则后,就到达过滤器的核心组件:范围处理器和属性处理器,其中范围处理器主要负责对输入的属性值范围进行处理,而属性处理器主要负责对鼠标、键盘所选择的图形数据属性进行处理。当然属性处理器和范围处理器可以交互地进行过滤。如果用户对过滤效果满意,就可将过滤过的数据送到视图区,否则需要重新进行过滤过程。

在整个挖掘系统中,数据过滤是必不可少的,能够确定图形文件中哪些数据最能代表数据的特征,哪些是用户最所需的,精炼了数据挖掘的结果。

2.4 显示组件

显示过程是该数据挖掘系统的最后一个过程,是指从视图区中选择图形元素显示到屏幕窗口上,允许用户在展现方法库中选取相应的展现方法。这样,通过数据挖掘系统挖掘出的信息就能最终在屏幕窗口中以直观的方式展现给用户。

展现方法库中主要存放的是一些展现方法,如二维笛卡尔坐标系、三维笛卡尔坐标系、二维墙面坐标系、三维墙面坐标系和一些特殊效果处理,如光照、雾化、融合、纹理等,还有可视化工具,如图表、曲线、决策树、规则图、boxplot图、直方图、饼图、高低区域图等等。

在整个系统中,展现部分主要负责给用户传递数据挖掘的结果信息。对于非专业用户来说,是否理解这些信息,关键还是在于此。因此,展现组件在整个系统中充当一个很重要的角色。

结束语 本文中讨论的基于 OLAM 的可视化数据挖掘系统是相对于传统的数据挖掘系统提出来的,是数据挖掘技术的具体应用,也是数据库(数据仓库)应用工具未来发展的方向。而 Internet 技术,尤其是 WWW 技术的发展,又为 OLAM 提供了一个可靠的平台,将会把基于 Web 的数据库应用推向一个新的高度。

参考文献

- [1] 林宇. 数据仓库原理与实践[M]. 北京:人民邮电出版社,2003
- [2] 段云峰. 数据仓库及其在电信领域中的应用[M]. 北京:电子工业出版社,2003
- [3] Inmon W H. 数据仓库[M]. 北京:机械工业出版社,2003
- [4] 彭木根. 数据仓库技术与实现[M]. 北京:电子工业出版社,2002
- [5] 陈京民. 数据仓库与数据挖掘技术[M]. 北京:电子工业出版社,2002
- [6] 王珊. 数据仓库技术与联机分析处理[M]. 北京:科学出版社,1999