

基于 J2EE 的 Web 挖掘客户数据的价值

邢国春

(长春师范学院 信息学院, 吉林 长春 130032)

摘 要: 简要介绍了 Web 挖掘的概念以及功能, 阐述数据挖掘系统结构, 并对其体系框架及实现方法进行了探讨, 最后给出基于分布计算 J2EE 平台的 Web 挖掘系统逻辑结构。从而系统将打破制约信息系统开发的瓶颈, 提高客户数据挖掘的效率。

关键词: 挖掘; J2EE; Web; 系统架构

中图分类号: TP393

文献标识码: A

文章编号: 1007-7634 (2006) 11-1758-03

User Data Value of Web Mining Based on J2EE

XING Guo - chun

(Information College , Changchun Normal University , Changcun 130032 , China)

Abstract: It briefly introduces the concept and the Function of Web Mining , elaborates the data mining system structure , and has carried on the discussion to its body sub - frame and the achievement method , finally produces the Web Mining system logical organization based on the Distributed Computing as the J2EE platform. Thus the bottleneck of the system will be broken. The system of information will be developed , and the Data Mining efficiency will be improved.

Key words: excavation ; J2EE ; Web ; system overhead construction

1 引 言

数据挖掘技术目前国内发展非常迅速, 面对 Web 上海量、分布、动态、异质、复杂、非结构化的丰富资源, 客户如何从中查找, 抽取自己想要得到的有用信息, 由此产生了挖掘技术。Web 挖掘就是从大量的 Web 文档和 Web 活动中发现, 抽取感兴趣的、潜在的有用模式和隐含的、事先未知的潜在信息。

综合运用计算机网络、数据库、数据仓库、人工智能、信息检索、信息提取、机器学习、统计学、可视化、自然语言理解等多领域的技术, 将传

统数据挖掘与 Web 结合起来^[1]。为了成功的应用数据挖掘技术, 围绕数据挖掘过程需要涉及问题的理解, 数据的理解、收集和准备, 建立数据挖掘模型, 评价机制, 同时系统应提供支持所有任务的必要手段和功能, 并最大限度地为用户使用这些功能提供方便的接口、选择和操作。

应用较普遍的数据在 Web 上的应用, 即 Web Mining 技术, 利用有效的客户信息, 大大降低运营成本, Web 上的电子商务是交互式的, 它的发展方向是顾客可以制定和指定产品服务, 交换信息。Web Mining 工具可以用来对 Web 上的商业模式建模、预测、了解影响销售的各种因素, 以便迅速调整市场价格、库存等。还可以发现顾客和访问者的

收稿日期: 2005 - 12 - 13

作者简介: 邢国春 (1963 -), 男, 吉林农安人, 长春师范学院信息学院副教授, 从事数据开发研究。

爱好、生活模式等，并且可以充分利用这些信息来发展新的客户，发掘新的商业机会。可以将 Web 挖掘定义为：从与 WWW 相关资源和行为中抽取感兴趣有用的模式和隐含信息^[2]，Web 挖掘的分类如图 1 所示。

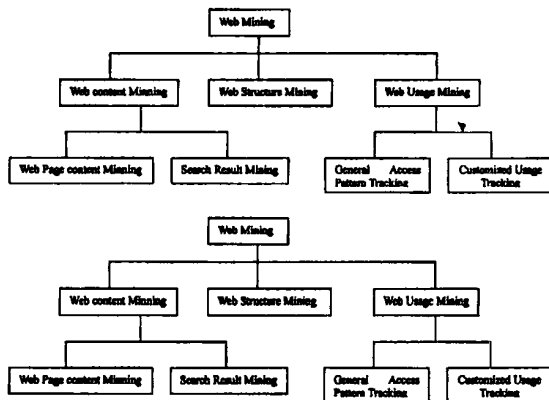


图 1 Web 挖掘的分类

2 基于 J2EE 的 Web 挖掘系统逻辑结构

系统中采取客户/服务器发展结构。在客户端基于 JAVA Applet 实现用户界面部分，open suite 放在应用服务器中实现，底层数据库/数据仓库放在数据服务器端实现。

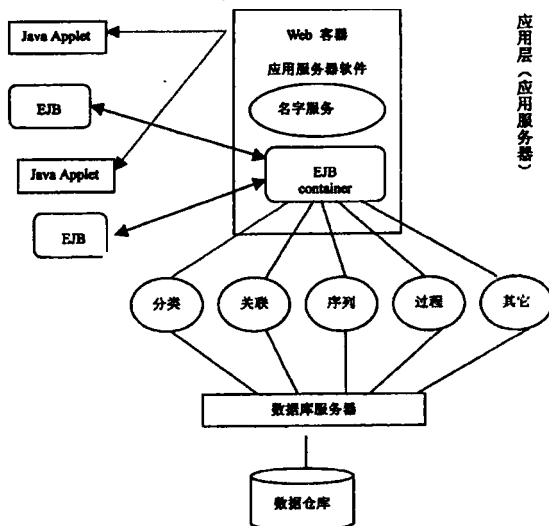


图 2 基于分布计算 J2EE 平台的 web 挖掘系统

将数据挖掘过程放在应用服务器端的实现方式使网络上的数据传输量较小，只需传输 JAVA Applet，DMQL 请求及挖掘结果。在下载至客户机后，JAVA Applet 通过 RMI/IIOP 协议与应用服务器的名

字服务建立连接，获取在其上注册的应用服务对象的对象引用，将 DMQL 查询封装为该次请求参数；数据挖掘工具完成指定算法求解过程，并将计算结果返回客户端，在此过程中，任一工具对象都可请求与数据库服务器通信，实现对数据存取操作^[3]。

采用基于 J2EE 技术最大优点是使系统有良好的可扩展性。当有新的数据挖掘工具加入系统时，无须改动和重新编译客户端应用和 open suite 注册，同时更新 Web 服务器的相应 Web 页面即可，见图 2 绘出一般框架。

用户界面分为三部分：图形界面 (GUI)、命令行界面及 Web 界面。通过 GUI 或 Web 界面提高数据挖掘任务请求，当发送 open suite 之前，任务请求都被转化成 DMQL 语言形式。

3 掘系统架构

挖掘系统包括规则生成子系统和应用评估子系统两部分，规则生成子系统主要完成数据仓库提供的历史数据，统计并产生相关规律，并输出相关结果。具体包括数据抽取转换、挖掘数据建立、建模、模型评估、结果的分布^[4]。应用评估子系统可以理解为生产系统中的挖掘代理程序，根据子系统产生的规则按照一定的策略对数据非类预测。通过系统的任务计划对生产数据产生评估指标。包括核心业务系统数据自动转入数据平台，规则实时评估、评估结果动态显示，实际效果评估 (见图 3)。

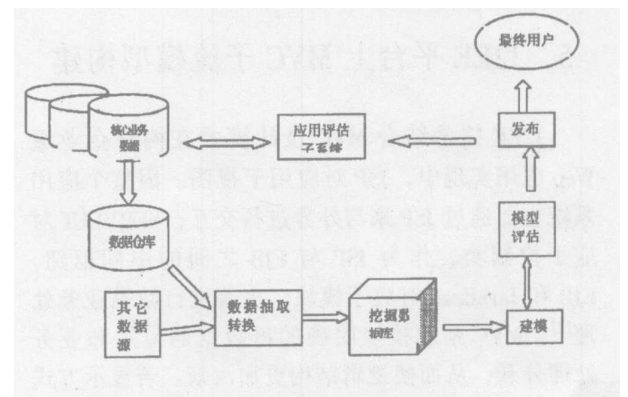


图 3 数据挖掘系统架构

4 Web 的集成价值

Web 集成 (Web Integration, WI) 是一套系统方法。它选择分析 Web 数据，然后把它们转换、加载数据仓库。WI 是一门包含数据仓库、因特网、

数据采集和搜索引擎的复杂技术。经过 WI 处理后, 外部数据在数量和质量上都应该等同甚至是超过内部数据^[5]。关键是怎样快速浏览 Web, 发现和获取有商业价值的信息。

Web 工具用来自动浏览 Web 并把经过选择的网站信息载入到数据服务器。Web 工具有搜索引擎功能, 植入的自动机制使 Web 工具自动细心地在因特网上查询网站和页面, 选择网站的标准和规则是定义在 WI 目录中。WI 目录放映了组织机构对 Web 数据商业兴趣和价值取向, 促成了 Web 域名组成的层次结构, 如 .com, .net, .edu 或 .org 的域在各个组织机构中各占多大比例。它还维持了需要定期查询及时信息的网站和 Web 工具列表。对于数据仓库, 更多有价值的数据输入将提升数据仓库的功效, WI 的首要任务就是从 Web 上提取有价值的信息。从某种程度看, 带给组织机构的 Web 数据的真实价格决定了 Web 集成的价值^[6]。与别的集成信息相比, 因特网数据更容易纳入数据仓库。过去人们不得不去图书馆或政府部门收集外部数据, 这些数据格式多样, 而在因特网上, 只需花几次鼠标点击, 人们就能在世界各地找到所需的数据。因此, Web 数据的基于比别的媒体格式的数据更容易载入数据仓库。

目前, 市场上已有了从 Web 上获取数据并把数据直接载入数据仓库的商业软件产品, 这些软件工具从网页上找回数据, 把相关数据并入新的 HTML 格式文档或商业应用。

5 J2EE 平台上 MVC 子流模型构建

J2EE 技术结合 MVC 设计模式在构建企业级 Web 应用实现中, JSP 对应用于视图。因整个应用系统主要通过 JSP 来与外界进行交互; Servlet 对应于控制类, 作为 JSP 与 EJB 之间的中间枢纽, EJB 和 JavaBean 对应于模块, 主要进行数据业务处理^[7]。MVC 系统模块明确的将数据的显示和业务处理分开, 从而使逻辑结构更加清晰。若显示方式改变只需要更改 JSP 视图页面, 并不要求更改相应的处理数据模块, 因而系统可以很容易加入新的业务, 可以灵活适应各种需求的变化。在 Web 客户端的 HTML 或 JSP 网页向服务器提交时, 控制器 Servlet 统一处理这些请求, 这个控制器 Servlet 根据提交业务不同, 将请求传递相应的业务 Bean 操作处理, 然后将业务处理 Bean 的结果再传递给视图

JSP。视图 JSP 在服务器上处理之后的 HTML 的方式回显给客户端, 见 MVC 模式构筑, Web 应用框架如图 4 所示。

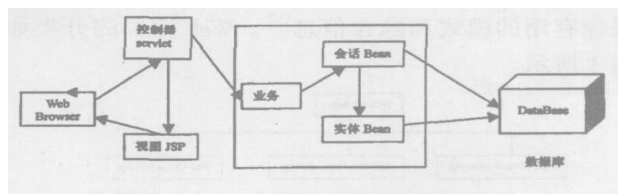


图 4 MVC 模式构筑的 J2EE 平台

MVC 模式中最关键的部分在于控制器的实现, 控制器能够很好的处理视图与模式的交互。

6 结 论

基于 J2EE 的 Web 数据挖掘技术发展已经到了信息系统开发的新阶段。目前先进的 J2EE 的层体系子框架, 提出了一种能够有效集成信息流和工作流平台, 给出了基模型结构和功能。但还有一些潜在的问题有待解决。其一是怎样把各种各样的 Web 数据集成到一个同构的书库框架中, 第二自动数据收集将依赖 Web 集成的效率。而多媒体数据的挖掘方式和算法还需要投入大量精力去开发。还有个别问题就是如何提高系统的灵活性, 缩短开发周期, 提高开发效率, 降低开发过程的复杂性。使系统的设计准确地反映系统的需求, 从而数据挖掘具有更强的适应性。

参考文献

- 1 袁梅冷, 等. 基于 J2EE 构建 BB 电子商务平台[J]. 计算机应用研究, 2003, (5): 140 - 141.
- 2 陆荣幸, 等. J2EE 平台上 MVC 设计模式研究与实现[J]. 计算机应用研究, 2003, (5): 144 - 145.
- 3 岳 昆. Web 服务核心支撑技术[J]. 软件学报, 2004, (3): 429 - 440.
- 4 宋善德, 刘 晋. 把 Web 数据并入数据仓库的研究[J]. 计算机应用研究, 2003, (4): 43 - 45.
- 5 Java Servlet API Specification 2. 2 [EB/OL]. <http://java.sun.com/products/servlet>, 2002 - 04 - 05.
- 6 都艺兵. 信息系统中如何实现数据库系统与 Web 的集成[J]. 计算机系统应用, 1998, (9): 33 - 35.
- 7 张志伟, 等. Paul Perrone, et al. J2EE 构建企业系统[M]. 北京: 清华大学出版社, 2001. 25 - 30.

(责任编辑: 滕代娣)