

.NET 平台下电子商务数据挖掘服务设计与实现

丁小宝,唐跃武,陈国栋
(台州学院,浙江 台州 317000)

摘要: 根据电子商务环境中分布式和异构性数据挖掘服务需求,设计了基于移动 Agent 和 Web Service 的五层分布式数据挖掘服务框架,实现数据挖掘服务与电子商务系统的松散耦合。从整体框架给出数据挖掘服务质量评价体系,包括支撑性服务质量评价和算法服务质量评价。具体分析了 .NET 平台下移动 Agent 的迁移和 WCF 技术创建数据挖掘服务过程组件的实现。

关键词: 数据挖掘服务;电子商务;移动 Agent;Web Service

中图分类号: TP302.1

文献标识码: A

文章编号: 1674-7720(2012)07-0017-03

Design and realization of E-commerce data mining service based on .NET framework

Ding Xiaobao, Tang Yuewu, Chen Guodong
(Taizhou University, Taizhou 317000, China)

Abstract: According to the data mining service needs of a distributed and heterogeneous E-commerce environment, this paper designs a five-layer distributed data mining service framework, realizes the loose-coupling between data mining service and E-commerce systems, provides the data mining service quality evaluation system from overall framework, including the supporting service quality evaluation and algorithm service quality evaluation. The implementation of mobile Agent migration under the .NET platform and how to construct data mining service components is analyzed.

Key words: data mining service; E-commerce; mobile Agent; Web Service

企业在实施电子商务或者全程电子商务过程中,由于采用不同的信息系统提供资源和服务,会产生异构的、不确定性的和非结构化的海量数据,其复杂程度已经远超出人类现有的信息分析和理解能力。企业如何利用这些海量数据吸引更多的客户,判定有价值的客户和保持客户等问题已成为企业战略决策的关键问题。数据挖掘技术成为企业解决这些问题的一个重要手段。数据挖掘工具多建立在数据仓库基础上,各大数据库厂商提供的数据库产品及相应分析工具、数据挖掘工具(如 IBM 公司的 Intelligent Miner、SPSS 公司的 Clementine、SAS 公司的 Enterprise Miner 等^[1]),给用户带来了便利。但各厂商产品不兼容、难以耦合,尽管部分数据库产品的数据挖掘工具支持二次开发,而大部分只是支持自身的数据仓库环境,缺乏普适性和灵活性;企业容易形成对单一厂商数据挖掘产品和服务的依赖性,针对不同环境需要重复构建相同功能的数据挖掘软件。

另一方面进行电子商务系统数据挖掘,所需要的数据主要是 Web 内容、Web 结构、Web 日志记录、客户信息、交易数据、查询信息等,这类数据具有分布性、异构性及海量性,而且目前的挖掘工具主要采用单机运行、集中处理结构化的同质数据的模式。而对于分布式存储的数据需要集中才能处理,这个过程对网络速度、稳定性要求比较高,数据挖掘结果响应时间比较长、数据的私有性和安全性遭到威胁。另外,一般的数据挖掘引擎是基于封闭体系构建的,难以管理和维护多个挖掘算法,结果格式不统一,许多数据挖掘结果无法直接给其他系统提供服务。针对以上问题,本文提出了面向电子商务应用的分布式数据挖掘服务的体系,集成分布式技术应用,提供给企业统一的挖掘结果格式的服务。

1 技术概要

1.1 移动 Agent 技术

在分布式计算领域,把在分布式系统中持续自主发

挥作用的,具有自主性、交互性、反应性、主动性特征的可移动计算实体称为移动 MA (Mobile Agent),目的是尽量把 Agent 靠近数据源,降低网络通信负载。一个完整的 Agent 系统包括两个部分:移动 Agent 和 Agent 服务环境。在 Agent 服务环境中,Agent 不断迁移和运行,通过 Agent 通信语言 ACL (Agent Communication Language) 与其他 Agent 协作通信。移动 Agent 的体系结构模型如图 1 所示。

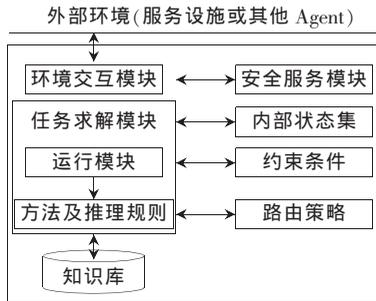


图 1 移动 Agent 体系结构模型

图 1 中,Agent 利用环境交互模块感知和作用外部环境;安全服务模块执行安全策略,负责 Agent 与外部环境的安全交互并阻止外部环境对 Agent 的非法访问;任务求解模块由运行模块、相关推理方法和规则组成。Agent 把移动过程中所获取的知识及任务求解模块保存在知识库中;Agent 在任务求解过程中,由内部状态集、约束条件及路由策略控制执行。内部状态集控制执行过程中的状态;约束条件约束 Agent 的行为和性能,路由策略用于满足复杂和非确定任务的求解,既可以是静态服务列表也可以是基于规则动态路由。

移动 Agent 的自主性和移动性是解决分布性、海量性数据处理的一种有效方法,能有效提高数据挖掘服务的智能性、并行性和稳定性^[2]。

1.2 Web Service

Web Service 是一种面向服务的体系结构和基于标准的应用集成方式,其流程包括:定义抽象的服务、实现具体的服务、发布并查找相关的服务、实现服务实例选择,最终实现可操作服务的使用;能将分布在 Intranet/Internet 分布式服务器上的应用集成到一起。Web Service 体系结构包括三种角色:服务提供者、服务注册用户及服务注册中心。服务提供者既是 Web Service 的提供者也是服务的拥有者,负责服务的发布、更新和回收。具体来说,定义 Web Service 的服务描述,将服务直接提供给服务用户或者提交到服务注册中心,服务注册中心是连接服务提供者和服务注册用户的第三方平台,汇集大量在线 Web 服务。Web Service 具体实现是基于以往的 Web 技术,通过定义相关协议和标准来实现,主要包括 XML、SOAP (Simple Object Access Protocol)、WSDL (Web Services Description Language)、UDDI (Universal Description Discover and Integration)。

通过 Web Service 将异构数据源处理服务标准化,便于将服务寄宿在不同的计算机系统环境中,避免数据挖掘时有大数据量的传输。

2 基于移动 Agent 和 Web Service 的数据挖掘

2.1 数据挖掘服务技术框架设计

在电子商务应用中,数据主要以分布和异构的形式存在,数据挖掘服务需要能提供一种有效、可信、安全及扩展性比较好的数据挖掘服务解决方案。本文采用一种基于 Web Service 标准封装数据挖掘算法组件、由移动 Agent 负责智能组合数据挖掘服务、实现数据挖掘服务与电子商务系统的松散耦合、便于数据挖掘服务的统一管理和维护的框架。这种结构能动态减轻分布式数据迁移量,满足各种异构系统的数据挖掘服务请求。

该架构可以抽象成由组合服务入口、服务组合代理 (S_{Agent}) 和 Web 服务三个实体组成。用户在组合服务入口输入服务组合规范和初始参数,系统将这些规范转化为处理模式,并生成一个 Agent 来完成这个计划;行为匹配算法将这个计划转化为 S_{Agent} 的物理行为,Agent 间通过 ACL 语言完成如克隆、复制、执行、归并等操作协作完成用户服务请求,如图 2 所示。具体可以划分为 5 层:(1)底层是电子商务系统数据层,包含电子商务环境中不同的数据源,由不同的组件提供访问服务;(2)数据挖掘服务组件层,主要在数据层上提供整个数据挖掘流程所必须的数据处理算法的组件;(3)数据挖掘服务组合层,根据不同的数据源组合和封装相关的处理组件,并提供 XML 结构化的数据给语义推理层调用;(4)语义推理层,主要协调用户请求和机器处理,服务推理逻辑解析用户请求,根据本体知识库已有的知识,标准化用户请求,设置约束条件,本体知识库中包含面向电子商务数据挖掘服务的统一的概念和关系的描述;(5)应用表示层主要是面向用户的 Web 应用请求和应用程序请求。

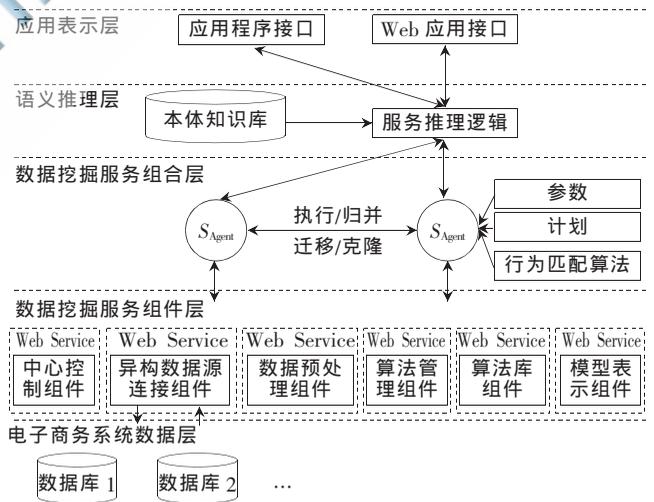


图 2 基于移动 Agent 和 Web Service 的电子商务分布式数据挖掘服务框架

2.2 数据挖掘服务评价体系设计

为电子商务企业提供分布式数据挖掘服务,服务质量是一个重要的因素,服务质量从框架的服务推理逻辑

辑、S_{Agent} 服务组合以及数据挖掘 Web 服务三个方面评价,服务质量概括为包括支撑服务质量和算法服务质量。支撑性服务质量指的是服务框架整体效果,包括可用度、响应时间、成功率、声誉以及价格;算法服务质量主要指数据挖掘 Web 服务效果,是整个数据挖掘服务评价体系重要方面,包括吞吐量、可伸缩性、存储空间、运行时间、鲁棒性。服务质量可以从这两个方面评价^[3],如表 1 所示。

表 1 数据挖掘服务评价体系

类别	评价指标	说明
支撑服务质量	可用度	最近一段时间内,服务的可访问时间
	响应时间	整个框架响应服务请求的时间
	成功率	单位时间内服务请求正确响应次数
	声誉	用户的体验评价
	价格	服务获得价格
算法服务质量	吞吐量	单位时间内处理的数据量
	可伸缩性	算法对海量数据处理能力
	存储空间	数据处理对数据存储的要求
	运行时间	处理一定量数据的时间
	鲁棒性	整个数据挖掘 Web 服务健壮性

3 .NET 平台下框架技术实现分析

电子商务分布式数据挖掘服务框架主要结合移动 Agent 和 Web Service 技术,Agent 的迁移和服务的封装是两个关键问题,具体实现采用 .NET 技术。目前移动 Agent 开发主要在通信结构相同的两端,采用 Java 语言的序列化和类装载机制实现。但电子商务数据系统是分散的,无法保证同构性,而且对于常见协议支持较弱,难以在分布式环境中提供完整的服务。.NET 是 Microsoft XML Web services 平台,XML Web services 支持应用程序在 Internet 上通信和数据共享,对于不同操作系统、设备或编程语言是透明的,能有效支持 Internet 协议和相关标准。

3.1 基于 .NET 移动 Agent 迁移

移动 Agent 迁移主要是解决 Agent 程序体和数据在异构节点间的迁移和恢复问题。完整地移动 Agent 的数据状态由代码态、数据态和执行态构成^[4]。代码态是指能完成 Agent 功能并做控制的任务程序代码;数据态指的是 Agent 对象的一些全局变量和属性,是对 Agent 的一些元数据的描述,如 Agent 标识、目的节点、迁移路径等;执行态指的是 Agent 执行过程中需要保存的上下文信息。在实际迁移过程中,Agent 代码态和数据态分别可以在 .NET 平台生成 dll 文件和结构化 XML 文件实现迁移;接收端可以利用 .NET 反射技术和 XML 类解析,还原 Agent 代码态和数据态。因此,技术实现上,构建 Agent 代码态类,管理这些类的管理类以及负责 Agent 代码态类和数据态文件迁移的通信类。针对分布式环境文件传输,.NET 平台通过 System.Net 和 System.Net.Socket 命名空间提供网络通信的类和方法;而对于 XML 文件

的传输则通过 System.XML 命名空间提供的 XmlWriter 类和 XmlReader 类实现 Agent 数据态 XML 文件生成和解析。

3.2 基于 WCF 构建数据挖掘组件 Web 服务

移动 Agent 要完成分布式数据挖掘服务,需要动态调用针对不同数据源的数据挖掘服务组件,Web 服务组件实现采用 SOA (Service Oriented Architecture) 设计的 WCF 技术。WCF 是微软基于 SOA 而设计的技术架构,WCF 技术能很好地开发分布式应用程序,支持 HTTP、TCP、Named Pipe、MSMQ、Peer-To-Peer TCP 等协议,能寄宿在 IIS、Windows Service、Console、Windows Application 等多种宿主(Host)中。技术实现上,一个具体的 WCF 服务由宿主、服务类和端点构成。宿主是指服务运行的运行环境。服务类是指具体的服务方法,在 .NET 平台中采用 [ServiceContract] 特性声明接口,或者直接声明类。端点是 WCF 服务的一个重要组成,由地址 (Address)、绑定 (Binding)、契约 (Contract) 构成。地址即标识消息发送的终点,绑定是描述消息的发送方式,契约是指消息包含的内容、消息的组织 and 操作方式^[5]。

对于移动 Agent 所调用的 Web 服务利用 WCF 技术实现,构建的服务能寄宿在多种环境中,支持多个通信协议。利用 WCF 创建 Web Service 分成 4 步:(1) 引入命名空间 System.ServiceModel 和 System.ServiceModel.Description;(2) 定义服务类,将服务接口添加 [ServiceContract] 特性,将服务算法的方法添加 [OperationContract] 特性;(3) 创建 Host 宿主。WCF 的服务可以运行在程序、Windows 服务或者 asp.net 上面,由于数据挖掘过程提供数据挖掘 Web Service,WCF 配置节点 <system.serviceModel> 终端,绑定协议选择 <endpoint address = " … " binding = " wsHttpBinding " contract = " … ">, 将数据挖掘 Web Service 寄宿在 IIS 上;(4) 创建 WCF 运行时间。

本文针对电子商务系统提出了分布式数据挖掘服务框架和相应服务评价体系,利用分布式技术移动 Agent 解决电子商务环境中数据异构和分布的问题,将数据挖掘过程封装成一个个服务寄宿在数据源端向移动 Agent 提供服务,为独立的电子商务数据处理中心建设提供依据。

参考文献

- [1] ABDULLAH M. GHOSON A L. Decision tree induction & clustering techniques in SAS enterprise miner, SPSS clementine, and IBM intelligent miner—a comparative analysis[J]. International Journal of Management & Information Systems, 2010, 14(3): 57–58.
- [2] 陈珂, 庞景中. 基于移动 Agent 的分布式数据挖掘平台的设计与实现 [J]. 计算机应用与软件, 2011, 28(7): 183–184.
- [3] 陈增科, 肖基毅, 陈灵娜, 等. 基于物元和 QoS 约束的数据挖掘服务选择 [J]. 计算机工程, 2009, 35(24): 90–91.

- [4] JAISANKAR N, SARAVANAN R, SWAMY K D. Intelligent intrusion detection system framework using mobile agents[J]. International Journal of Network Security & Its Applications (IJNSA), 2009, 1(2): 73-74.
- [5] 韩旭, 王海波, 柳克俊. 基于 .NET Framework WCF 的面向服务 SOA 中间件设计[J]. 小型微型计算机系统, 2010, 31(12): 2359-2361.

(收稿日期: 2011-11-10)

作者简介:

丁小宝, 男, 1985 年生, 硕士, 主要研究方向: 中小企业信息化与异构信息系统集成。

唐跃武, 男, 1982 年生, 讲师, 硕士, 主要研究方向: 电子商务应用, 物流与供应链管理。

陈国栋, 男, 1980 年生, 讲师, 在读博士, 主要研究方向: 知识管理。

