

电子商务环境下的 Web 数据挖掘系统架构设计

黄伟建¹, 桑志超², 杜巍¹

(1 河北工程大学 信息与电气工程学院, 河北 邯郸 056038 ; 2 河北工程大学 经济管理学院, 河北 邯郸 056038)

摘要:通过对 Web 数据挖掘技术及其在电子商务中的应用进行了研究,发现传统的电子商务 Web 数据挖掘系统具有挖掘引擎封闭、难以对算法动态管理、挖掘结果难以重复利用的不足,本文提出了基于 Web 服务技术和移动 Agent 技术的适用于电子商务领域的 Web 挖掘系统的架构,此架构具有更好的跨平台性和更高的挖掘效率。

关键词:电子商务; Web 服务; Web 数据挖掘; 移动 Agent; 架构

中图分类号: TP303

文献标识码: A

The design of framework of Web data mining system under E-commerce

HUANG Wei-jian¹, SANG Zhi-chao², DU Wei¹

(1. Institute of Information and Electrical Engineering, Hebei University of Engineering, Hebei Handan 056038, China;
2. Institute of Economic Management, Hebei University of Engineering, Hebei Handan 056038, China)

Abstract: Firstly, this paper studied the Web data mining and its application in the e-commerce, and found the disadvantages of traditional e-commerce Web data mining system. The disadvantages contain closed mining engine, difficult dynamic management for algorithm and the mining result is difficult to be reused. Finally it proposes a Web data mining framework suitable for e-commerce based on Web Service and Mobile Agent and the framework has better portability and higher mining efficient.

Key words: E-commerce; Web Service; Web data mining; mobile agent; framework

随着网络的日益普及和迅速发展以及其所带来的利益,越来越多的企业加入到电子商务的行列之中。网络已经形成了一个巨大的、具有重要潜在价值的信息空间。所面临的难题是如何从这些复杂的大量数据中找到对企业有价值的信息,解决信息获取问题对电子商务的发展至关重要。传统的获取信息的方法—搜索引擎技术由于其搜索范围较小、搜索结果命中率低、不能提供多媒体搜索服务以及不能提供个性化服务,目前该技术已经不能满足电子商务对与信息和知识的需求。国内以局部挖掘算法的设计、分析、改进为研究的重点,而很少研究数据挖掘系统自身的构建、开发模式的种类等问题。本文从应用出发提出了一个综合 Web 服务和移动 Agent 技术的挖掘架构,为电子商务中数据处理提供一个实时有效、跨

平台综合利用、智能性和良好伸缩性的数据挖掘解决方案。

1 Web 挖掘系统现状分析

理论研究和应用研究是当下国内外在此方面研究的两个集中点,与国外相比,我国对此方面的研究起步略晚。我国的研究重点主要集中在算法方面,例如算法的选择、设计、分析和完善,在对 Web 挖掘系统的架构的设计和模式开发方面的研究还较弱。

现有基于电子商务的 web 数据挖掘系统的不足之处如下:

分布性和实时性问题。目前大部分已经存在的 Web 数据挖掘系统无法加载网络功能,对处于分布环境中的数据无能为力。在算法和模式方面

比较集中,要将分布在不同环境下的数据进行整合分析,要求有高速的网络,并且耗时长,系统的实时性较差。

异构数据库环境问题。网络上的任何一个站点都可以成为一个数据源,并且不同的数据源都是异构的,数据的组织不一样,由此这些数据源构成了一个海量的异构数据库环境。要想对这些数据进行数据挖掘,就要将这些不同的数据源上异构的数据集成起来,才能进行统一的挖掘。

可移植性问题。目前许多的电子商务环境下的 Web 数据挖掘系统都是为解决企业自己的商业需求,系统的体系架构比较封闭单一,没有更多的考虑到 Web 数据挖掘系统的可移植应用需要。

半结构化数据问题。电子商务系统中的数据非常的庞大,且每个数据源都有自己独特的设计,数据本身具有动态可变性,与来自数据库中的数据相比,数据源中的数据是一种半结构化的数据。处理半结构化的数据大大增加了挖掘的难度。

针对目前电子商务环境下的 Web 数据挖掘系统存在的不足,本文提出了电子商务环境下基于 Web 服务和移动 Agent 的技术的 Web 挖掘系统架构。Web 服务的工作是封装各种数据挖掘算法,移动 Agent 的工作是携带相关信息迁移执行任务组合 Web 服务应用。

2 Web 服务和移动 Agent 技术相结合

本架构利用 Web 服务是分布式计算体系的优点,每个挖掘模块都以独立的形式在 UDDI 进行注册、发布。移动 Agent 的一个优势是能够携带数据信息实现从一个节点到另外一个节点的迁移,其自主性和克隆的特点都对实现分布式的 Web 服务非常有利,将两者结合能实现优势互补。

在 Web 服务架构中,服务提供者利用标准格式来创建所拥有服务的描述,并将该服务描述在 UDDI 中心进行注册,同时也接收服务请求者有关服务调用的消息。服务请求者向注册中心提出请求,查找服务描述,通过绑定服务提供者提供的接口请求调用服务。服务注册中心担当的角色是服务请求者和提供者者的中介,将二者联系在一起。将 Web 服务和移动 Agent 技术结合,可以将一个服务组合系统的工作过程看成为一个 workflow,主要工作由移动 Agent 承担,相关信息由其携带在网络中游走,并和另外的 Agent 进行交互合作,来实现所要达到的目标。

3 Web 挖掘引擎架构设计

现有的数据挖掘引擎的架构主要分为嵌入式架构和 C/S 式的挖掘架构。但是这两种架构目前存在很多不足:嵌入式架构对电子商务系统的依赖性很强,要求的运行环境和电子商务相同,挖掘引擎没有良好的可移植性;C/S 式的挖掘架构虽然对应用环境有较好的适应性,但是采用这种形式的引擎只能以某种算法为核心,不能有效对算法进行动态管理。为满足数据挖掘的需求,本文 Web 挖掘引擎将算法和引擎相分离以便于对算法进行管理,并应用 Web 服务和移动 Agent 技术,得到较好的挖掘效果。本引擎架构中的模块主要包括两个:管理模块和算法调用模块,见图 1。

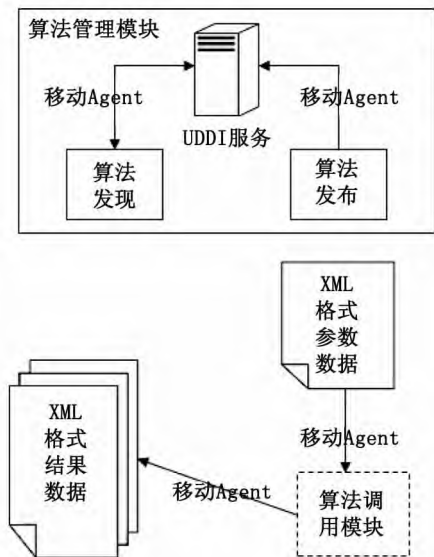


图1 Web挖掘引擎架构设计 Fig.1 Design of Web mining engine

3.1 算法管理模块

算法发布 Agent 能够传递特定的命令,这种命令由算法信息转化而来能够被 Agent 识别并被下一个 Agent 接受。下一个 Agent 被传递命令的 Agent 激活,它会移向 UDDI 注册中心,发送以一些与算法有关的请求,这些请求的内容可以是更新算法或者除去某些算法等等。注册中心会与迁移到此的 Agent 进行信息交流,接受与算法有关的请求对其进行封装存储相关的信息,这些信息包括算法的 Web 服务描述语言文档、地址等相关信息。执行至此,移动 Agent 使命终结,不再存在。

算法发现模块与算法发布模块的前期执行过

程大致相同,不同点是发送命令的内容不同,向其后一个 Agent 发送的是与查询算法有关的信息,然后激活它去执行任务。移动 Agent 也要与注册中心的服务器进行信息交流,目的是为了搜索算法。如果搜索到目标算法,依据目标算法在注册中心存储的信息和执行任务的信息迁移至服务提供者,进行信息交流,信息交流的过程可能是反复的,最终得到挖掘结果并返回。执行至此,移动 Agent 使命终结,不再存在。

3.2 算法调用模块:

该模块的功能实际上已经由移动 Agent 实现,移动 Agent 代表服务请求者与提供者进行交互,调用算法,并将挖掘结果返回。

4 数据挖掘系统的架构

依照以上的分析,本文在此将电子商务下 Web 挖掘系统的架构分为三个层次,分别为:用户层、数据层、数据挖掘层。具体细节如图 2 所示。

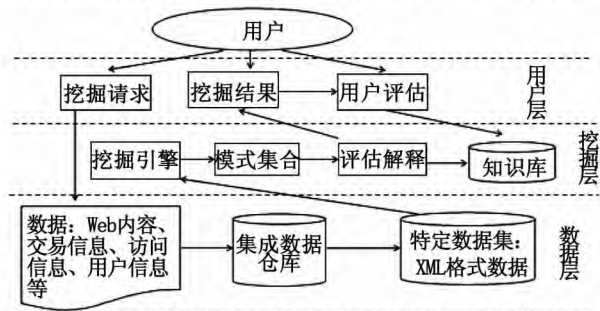


图2 Web挖掘系统总体架构

Fig.2 Framework of Web mining system

用户层:用户通过用户层提供的接口与系统进行交互,在提出挖掘请求前进行数据挖掘的参数设置,并显示挖掘之后的结果。在结果显示之后,用户可以对挖掘的结果进行评估并且可多次利用知识库中存储的知识。

数据层:数据层收集数据,并对收集到的数据进行清洗集成形成数据仓库。然后对数据仓库中的数据利用 XML 技术进行数据处理,将数据转换成方便进行挖掘的存储形式,形成特定数据集等待被挖掘。

挖掘层:利用 Web 服务和移动 Agent 技术进行数据挖掘,得到模式的集合,并对模式集合进行合理评价和解释,将结果存入知识库。评估之后将挖掘结果提供给用户使用。

5 数据挖掘系统的实现

本系统利用 Java 技术来实现移动 Agent,采用

JDBC 数据库技术和 Web 技术来实现平台的无关性;利用 JSP 技术来实现动态页面,完成与系统使用者的信息交互;系统的运行环境采用 MyEclipse;数据库采用 SQL Server;利用 Tomcat 来模拟 Web 服务器。

当网站的用户在查询商品信息时,推荐模块会根据收集到相关信息调用挖掘算法进行数据挖掘,识别是否存在一些可推荐的商品,为客户实现更个性、更全方位的服务。

6 结论

1) Web 挖掘算法以 Web 服务封装的形式且与挖掘引擎相分离单独进行注册和发布,使系统和算法实现松散耦合的目标得以实现,便于对算法库的动态管理,易于根据需求添加、删除或者修改算法。

2) 利用移动 Agent 技术来迁移数据,承担一定的通信功能,降低了网络通信的代价,加强了挖掘系统的实时响应能力。

3) 本架构不受异构平台的影响,得益于移动 Agent 技术不依赖于限定形式的主机和传输层协议的特性。应用于 Web 服务中的各种协议都是简单的、通用的,平台之间的差异不会对其造成影响,每种程序设计语言都可以实现用 Web 服务封装的算法,不受程序设计语言的制约。

参考文献:

- [1] JIAWEI HAN, MICHELINE KAMBER, JIAN PEI. Data mining: concepts and techniques[M]. Elsevier: Morgan Kaufmann, 2011.
- [2] 林芳. 基于数据挖掘建模语言的数据挖掘系统设计[J]. 制造业自动化, 2011, 33(24): 27-29.
- [3] 孙召伟, 赵建利, 朱东生. 数据结构中递归转非递归算法分析及模型设计研究[J]. 河北科技大学学报, 2011, 32(1): 43-46.
- [4] 李春梅, 李艾丹, 韩爽, 等. Web 数据挖掘中数据异构问题解决方法的研究[J]. 2012, 44(4): 85-90.
- [5] 茹蓓, 肖云鹏, 张俊鹏. 基于 Agent 的移动 Web 服务集成方案[J]. 计算机工程, 2012, 38(9): 49-50.
- [6] JIAWEI HAN, MICHELINE KAMBER, JIAN PEI. Data mining: concepts and techniques[M]. Elsevier: Morgan Kaufmann, 2011.
- [7] 莫学值, 韦必忠. Web 数据兴趣区域数据挖掘过程分析[J]. 科技通报, 2013, 54(10): 135-138.
- [8] 刘永立, 王海涛, 孙维民, 等. 基于基础数据库的煤矿应急救援指挥系统[J]. 黑龙江科技学院学报, 2010, 21(1): 44-47.
- [9] 陈远, 郑珊. 我国电子商务领域的研究热点与主体结构分析[J]. 情报科学, 2013, 31(4): 139-144.

(责任编辑 刘存英)