

Web 数据挖掘 在网络教育中的应用

汪全莉¹ 陈代春²

(1. 湘潭大学管理学院, 湖南湘潭 411105;

2. 湘潭大学图书馆, 湖南湘潭 411105)

摘要: 随着网络教育的发展, 将相应的网络技术及时应用到网络教学中, 提高网络教育质量, 一直是教育界探讨的重要课题。文章主要探讨了在网络教育中应用 Web 挖掘的必要性, 介绍了 Web 内容挖掘、Web 结构挖掘和 Web 使用挖掘在网络教育中的应用, 以实现网络教育的个性化, 提高网络教育质量。

关键词: Web 挖掘; 网络教育

中图分类号: G252 **文献标识码:** A **DOI:** 10.3772/j.issn.1674-1544.2008.02.009

On Application of Web Data Mining in Web - Based Education

Wang Quanli¹, Chen Daichun²

(1. Management College of Xiangtan University, Xiangtan 411105;

2. Library of Xiangtan University, Xiangtan 411105)

Abstract: With the development of Web - based education, it is the important subject to improve the quality of Web - based education by the application of Web technology in Web - based teaching . This paper mainly discusses the necessities to apply Web data mining technology in Web - based education, and the application of Web content mining, Web structure mining and Web usage mining in Web - based education to bring out the personalization of Web - based education and improve Web - based education.

Keywords: Web data mining, Web - based education

随着计算机技术、网络技术、多媒体技术、通讯技术在教育领域的广泛应用, 传统的教育形式逐步向信息化、网络化和终身化的方向发展, 不再拘泥于单一的面面对面的教育形式。一种基于 Internet 的, 与世界的学习资源实时相连, 可进行在线学习、离线学习以及在线与离线相结合的自主性、终身性教育形式——网络教育应运而生。

随着网络教育的发展, 网络上拥有的远程教育资源日益增加, 无论是教育方还是学习者都希望从日益增长的信息中发现对自己有价值的信息。这就需要借助一些先进技术, 如数据挖掘技术, 从海量信息中挖掘出我们真正需要的信息。由于网络教育主要以网络为平台, 因此, 与网络有关的 Web 挖掘在网络教育中的使用就显得尤为重要, 已成为当下较为重要的研究课

第一作者简介: 汪全莉(1974 -), 女, 湖南常德人, 讲师, 主要研究方向是信息组织、数据挖掘与信息服务。

基金项目: 湘潭大学 CSSCI 青年教师培育项目“数据挖掘在信息服务中的应用研究”(0609015)。

收稿日期: 2008年1月24日。

题。

1 关于 Web 数据挖掘

Web 数据挖掘(简称 Web 挖掘)是从数据挖掘发展而来的,集多领域技术(Web 技术、数据挖掘技术、计算机技术、信息科学技术等)为一体的一项技术,专门挖掘与万维网有关的数据,既可以是网页包含的数据,也可以是 Web 操作产生的数据。Web 数据可以分为 5 类:网页本身的内容;网页内部结构,包括 HTML 或者 XML 代码;网页之间的链接结构;描述网页被访问的使用数据;用户简档,包括与人口统计有关的信息、注册信息以及从 cookie 中获取的信息。

根据以上要挖掘的数据对象,Web 挖掘可分为 3 种:Web 内容挖掘、Web 结构挖掘及 Web 使用挖掘。其分类见图 1。

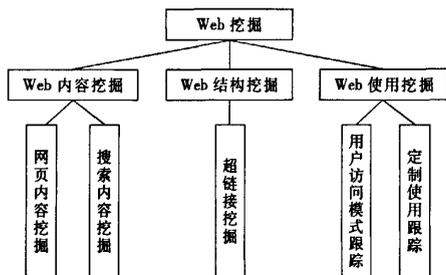


图 1 Web 挖掘分类

2 Web 挖掘在网络教育中应用的特点

与传统的教育模式相比,现代网络教育有许多独特之处,并且都与网络相联系。要更好地发挥网络教育的这些优势特性,恰当地利用 Web 挖掘技术与方法是十分必要的。

网络教育是通过网络等通讯平台来实现其教育的目的,也就是说教师在时间上、空间上、心理上与学生处于分离状态。网络教育的这一属性,带来了网络教育的开放性特征,打破了学生物理的、生理的、社会的障碍,使得“有教无类”、“按需学习”成为可能,学习者可以自由地选择时间、地点、学习内容与老师。这种分离的

属性同时也提高了可选择性,为了获得更好的选择,正确利用 Web 挖掘技术是非常必要的。

网络教育以 Web 为依托,并且网站具有与一般的 Web 页相同的特性。例如,网站上的数据不同于传统数据库中结构化的数据,而是非结构化或半结构化的数据,不能用传统的数据挖掘技术来进行处理。目前,主要是用 Web 数据挖掘技术来处理网站上的大量数据,从中发现有价值的信息。

现代网络教育是实现终身教育主体化,让个体充分、自由发展的教育,具有个别自主性。在网络教育环境中,学生是学习的中心,可以利用 Web 挖掘技术自主搜索需要的课程和感兴趣的信息,满足个性化的学习计划。

3 Web 挖掘在网络教育中的应用

网络教育主要是通过网络来实现其教与学的目的。在网络教学中,可以利用网站上的所有内容,而且网站内容的查找和使用都会留下踪迹,这些信息会自动储存在 Web 服务器的日志文件中。可以应用 Web 挖掘来分析和处理教育网站上的日志文件信息,以得到有价值的信息。例如,有多少人查找了某远程教育网站、他们来自哪里、哪些页面受欢迎、主要选择一些什么样的课程、专业与教师。利用 Web 挖掘技术所得到的这些信息,对教育方或学习方都非常重要。在这里笔者主要根据 Web 挖掘对象的不同,从 Web 内容挖掘、Web 结构挖掘及 Web 使用挖掘 3 个方面来讨论 Web 挖掘在网络教育中的应用。

3.1 Web 内容挖掘在网络教育的应用

Web 内容挖掘是从文件或描述中筛选出有价值的信息的过程。基本的 Web 内容挖掘是一种文本挖掘,可以看作是对基本搜索引擎所完成工作的一种扩展。其在网络教育中的应用主要表现在建设、检索、利用教育资源,选择教学对象及授课教师,安排教育网页本身内容等方面。

(1)网络教育资源的建设、检索、利用。网络教学资源的编辑与制作是网络教育的基础,施教方为了建立覆盖面广、功能齐全的资源库,必须编制信息资源目录,征集各类资源信息,并进行资源查重、资源信息筛选、资源信息整理与存储。而接受教育的学生,为了获得自身想要获得的教育资源,利用 Web 内容挖掘技术从教育网站或教育资源库中来获得相关的学习资源,满足学习的需要。

(2)选择教学对象及授课教师。网络教育改变了传统的面授、辅导的教学方式,实行自主式的学习。因而学生在教学对象的选择和授课教师的选择方面有很大的自主权。学生可以自由地选择授课老师,老师也可以自由地选择学生。那么,如何来了解对方,做出选择呢?这就类似于“客户关系管理”系统,了解业务对象是走向市场的第一步。要较好地完成这一步,管理者可利用数据挖掘中的关联规则技术在已建立好的老师和学生数据库中进行关联查询,找到相关的记录,提交给老师或学生,做好教师与学生的双向选择。

(3)设置个性化网页,安排个性化内容。可利用 Web 内容挖掘中的个性化挖掘技术,发现用户的访问特点及感兴趣的内容,并据此修改网页访问风格或网页内容,从而更好地适应学生的需求。这涉及为每个用户创建独特的网页。也就是说,针对用户访问特点设计网站。从这个层面来说,属于 Web 内容挖掘的应用。然而,个性化挖掘需要研究 Web 日志数据从而发现用户访问的模式,从这个层面来说,则属于 Web 使用挖掘。

3.2 Web 结构挖掘在网络教育中的应用

Web 结构挖掘可以看作万维网组织建立的一个模型。模型可以用来对网页分类或者为网页建立相似性度量。Web 结构挖掘中的 Pagerank 用于度量网页的重要性,网页的 Pagerank 值通过指向它的网页计算,这实际上是基于后向链接的一种度量。后向链接不是该网页链出去的链接,而是指向网页的链接。利用 Web 结构挖掘的相关技术,来确定网络教育的权威网站和中

心网页,并据此来选择网络教育机构。

3.3 Web 使用挖掘在网络教育中的应用

在现代网络教育中,实现个性化服务是一项重要内容。个性化服务就是以用户为中心,为每个学习者寻求适合其本身特点的发展方向,使每个学习者能够在各自特长的领域中得到充分的发展。这样不仅促进了学习者的健康成长和个性完善,同时也满足了社会发展对各种人才的需要。网络教育要彻底实现个性化服务,就必须动态地组织和呈现与学习者当前学习能力最相关的学习内容,即通过学习用户的访问模式,自动地改进站点信息的组织和显示方式。另外,系统还能预测学习者的下一步操作,将点击率极大的超链接所指向的页面提前下载到本地缓存中,从而提高学习者的访问速度。而这些功能都是需要 Web 的使用挖掘技术来实现的。

Web 挖掘在网络教育的应用主要体现在 Web 使用挖掘的应用上。Web 使用挖掘的研究对象是 Web 使用数据或者 Web 日志。Web 日志即网页访问数据。通过对 Web 日志进行分析,可以得到有助于改善网站设计的信息。Web 使用挖掘可以应用于多种目的。通过分析一个用户访问的网页序列,可以得到用户的简档信息,从而有助于实现网络教育的个性化。此外,使用网站挖掘,还可以评估网页的质量和效果,从而有助于教育网站质量的改善。

Web 使用挖掘在网络教育中的具体应用主要体现在实现网络教育个性化和改善教育网站质量这两方面。其应用过程如下。

(1)问题定义。在网络教育中实现个性化服务就是尽可能使每一个学习者在浏览网站时都有自己是该网站唯一用户的感觉,亦即尽可能地迎合与满足每一个学习者的学习兴趣和需要,不断进行自我调整来适应学习者需求的变化。具体地说,就是通过分析 Web 日志数据、学习者的学习历史记录,挖掘出用户的使用模式,从而向用户提供个性化的学习服务。

(2)建立挖掘数据库。建立挖掘数据库的过程主要是数据搜集、数据描述、数据选择、数据清理、构建元数据、构建挖掘数据库和维护挖掘

数据库。

(3) 分析数据。首先根据某个访问者的 IP 划分数据,找到每个访问者的访问记录集。然后将该访问者的每一次访问记录集为一个访问事务。最后,将所有的访问事务按时间排序,构成进行挖掘的事务集。每一个访问事务相当于访问者对站点的一条访问路径。

(4) 模式发现。利用路径分析、关联规则发现、序列模式、分类聚类等方法对 Web 日志数据进行分析,进而发现用户的访问模式。①路径分析。就是通过对 Web 服务器的日志文件中学生访问次数和路径进行分析,确定最频繁的路径访问模式或大的参引访问序列。它可以判定一个网络教育站点中最频繁访问的路径,还有一些其他的有关路径的信息可以通过路径分析得出。例如,通过分析可以得出,80%的学生存取这个站点是从/course/English 开始的;60%的学生在浏览了4个或更少的页面后就离开了。②关联规则发现。就是找到客户访问网站上的各种文件之间的相互联系,挖掘隐藏在数据间的相互关系。在网络教育中,关联规则就是找到学生访问网站上的各种知识之间的相互关系。例如,68%的学生存取这个站点是从/course/network 开始的,同时也访问了/course/java,由此可知,当学生对某一知识点感兴趣时,也可能会留意相关的知识点。这有利于学校更好地组织站点,为学生减少过滤信息的负担,从而改善网站提供信息的质量。③序列模式。在时间有序的事务集中,序列模式的发现就是指那些“一些项跟随另一个项”的内部事务模式。例如,在访问/CBecourse/chapter1 的学习者中,有60%的人曾经在过去的一个星期里用关键字“计算机辅助教育”在 yahoo 上进行过查询。发现序列模式能够便于预测学习者的访问模式,有助于针对这种模式开展有针对性的个性化参考教学。④分类和聚类。分类技术可以从个人信息或共同的访问模式中得出访问某一服务器文件的用户特征。这种描述可以用于分类学习者根

据学生的特点提供个性化服务。聚类分析可以从 Web 访问信息数据中聚类具有相似特性的学习者,便于开发和设计未来的教学模式和学习群体。

(5) 模式分析。模式分析是 Web 使用挖掘中的最后一个重要步骤。其将通过选择和观察发现的规则、模式和统计值转换为知识,再经过模式分析得到有价值的模式,即教育网站感兴趣的规则、模式。如图形化模式或为不同值赋予不同颜色的可视化模式,可使数据中的总体模式或趋势变得更加直观。

4 结 语

合理地利用 Web 挖掘技术,能够帮助教育网站及时调整教学计划,合理设置课程,改善网站设计,提供个性化教学服务,并合理利用网络教学资源。随着 Web 挖掘技术自身的发展,更多的技术与方法将会不断被引入到教育领域,提高网络教学的整体质量。

参考文献

- [1]李雄飞,李军.数据挖掘与知识发现[M].北京:高等教育出版社,2003.
- [2]刘同明,等.数据挖掘技术及其应用[M].北京:国防工业出版社,2001.
- [3]陈登科,胡翠华.数据挖掘技术在远程教育中的应用[J].情报科学,2003,21(4):445-447.
- [4]王春艳.浅谈数据挖掘与教育资源库的整合[J].教育信息化,2006(9):49-51.
- [5]范莉莎,刘刚,等.Web数据挖掘在网络教育中的应用[J].计算机技术与发展,2006,16(6):68-70.
- [6]杨永斌.数据挖掘技术在教育中的应用研究[J].计算机科学,2006,33(12):284-286.
- [7]丁琳,吴长永.数据在远程教育个性化服务中的应用[J].电化教育研究,2002(9):43-46.
- [8]BersonA,SmithS.Data Warehousing Data Mining & LAP[M].London:Mcraw-HillBook,1999:272-320.