

Web 挖掘妙用

浩如烟海的 Web 数据，如果有挖掘技术能充分利用其价值，其意义不可估量。

编译 周姝嫣

客户的交互活动能反映其重要倾向和模式，可以帮助公司设计网站，有效地表达和销售其产品和服务。

各种规模和种类的公司都殷切期望成为下一个亚马逊，因而争相建立他们的电子商务网站。公司通常关注交易处理方法，设置动态目录和购物系统，但是往往不能规划网站产生的大量客户数据。大多数公司没有意识到电子商务能否取得成功取决于如何利用 Web 数据将访客转化为顾客。销售产生的 Web 数据比销售本身更具价值，因为 Web 数据可以与顾客建立长期的利益关系。

不管最终是否售出产品，每次有人访问零售网站都可以产生重要的客户行为数据。每位访客的行为都是一份数字签名，可以反应其习惯、偏好和倾向。这些交互活动反应的重要倾向和模式能够帮助公司设计网站，有效表达和销售其产品和服务。公司可以聚集、增强和挖掘 Web 数据，从而了解哪些产品卖得好、哪些措施行得通，而哪些不行，以及哪些人在买、哪些人没买。

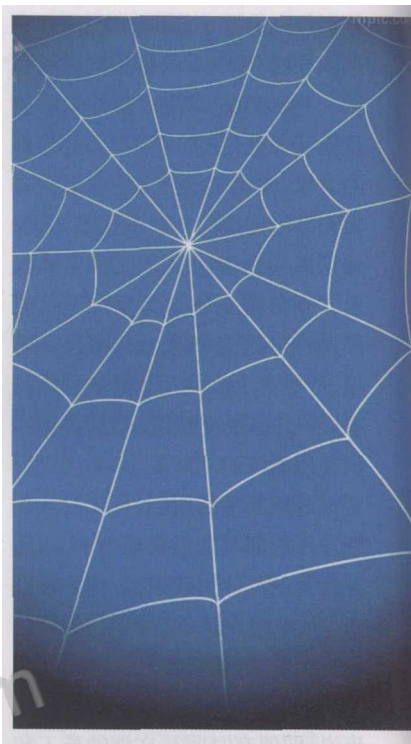
创建 Web 数据

游客每次访问你的网站都表示对产品和服务感兴趣，因此你必须仔细观察每次交互活动。但是，Web 数据种类繁多、数不胜数。游

客产生的服务器日志文件、表单数据库和电子邮件可以捕捉到许多数据，要分析电子商务数据，你必须将这些分散的数据集中为连贯、综合而全面的视图。如果你事先计划如何捕捉重要的客户信息，你就能轻松地集成和挖掘 Web 数据。

在建设电子商务网站之前进行战略规划，你就可以捕捉到访客的偏好、在线行为等重要信息。花点时间考虑网站的总体设计，如在主页上设置怎样的提示和链接，你才可以跟踪访客的行动。另外，在访客点入网站、查询或购买产品时，提示进行快速、简短的注册，同样可以捕捉到重要的个人信息，你可以利用这些信息增强和挖掘数据。

汇总和捕捉顾客信息的关键在于获取唯一的标识符：访客 ID 号。这种策略已经过验证：通过提供特殊服务或鼓励措施，让用户在首次访问网站时就注册。允许他们访问网站的某个特殊部分（举办竞赛或门票抽奖）。关键是你只有在他们注册后，才能设置 Cookies，而 Cookies 可作为唯一的 ID 号。从这个角度而言，零售商能够利用唯一的 ID 号追踪每位客户的交互活动。网站可以利用唯一的 ID 号，将日志文件和表单数据库与公司的数据仓库、其它流动信息和本地信息、活动目录服务器网络或者协同过滤引擎等相连。



增强 Web 数据

服务器日志文件提供访客使用的域种类、访问时间、关键词和搜索引擎，并指导观察访客如何来到网站，以及使用什么样的关键词才锁定网站。服务器分配的 Cookies 可以追踪看过的浏览器记录和网页，观察访客打开网站的频率，以及进入的区域。表格可以提供重要的访客私人信息，如性别、年龄、邮编等。这可能是最重要的客户视图，因为所包含的信息可以用于增添附加数据，如来自数据仓库的数据。你可以添加客户的流动信息和家庭信息，包括客户可能有多少收入、开什么车、以及有几个孩子。

外部信息可以与网站数据相连，还可以识别顾客身份、特征、生活方式及行为。许多供应商提供这类信息，如 Acxiom、Equifax、Experian、MetroMail、Polk 等公

司。目前出现了一个行业专门从事客户行为信息的分离、分类工作,并转卖给公司,当然包括那些拥有网站的公司。

另外,最近出现了全新的供应商——“网络使用者统计商”,他们出售软件或服务,有时候兼售软件和服务,实现协同过滤和关联销售,了解顾客概况。这些新的数据供应商代表着一类全新的网络公司,都在设法捕捉和形成有关互联网用户行为和偏好的信息。包括DoubleClick、Engage Technologies、Firefly、Net.Perceptions等公司。这些供应商利用大量的解决方案追踪和描述客户信息,包括借助专利软件和数据库,或者通过服务器网络混合Cookies。

所有的内部信息和外部信息都可以写入Oracle表格或一般文件,然后连接或输出到数据挖掘工具。挖掘工具包括自动工具,主要在数据仓库中使用,用于抽取顾客种类、倾向和关系;还包括新颖、简单易用的数据挖掘工具,具有专门为业务人员和销售人员设计的GUI接口。这些数据挖掘分析工具可以提供许多格式的可行方案,可以与负责设计、维护和销售电子商务网站的人员分享。

挖掘Web数据

迄今为止,Web数据的分析大多涉及日志流量报告,而大部分报告只能提供服务器活动的累积量,而不能真正从商业角度观察顾客动向和在线行为。目前,NetIntellect、Bazaar Analyzer Pro、HitList、NetTracker、Surf Report、WebTrends等大多数流量分析软件都是在分析日志文件的基础上,给

出服务器活动的预定义报告。因此,这些软件的使用范围就局限于统计域名、IP地址、Cookies、浏览器以及其它专门针对TCP/IP的机对机活动。

另外,挖掘电子商务网站的Web数据可以分析顾客行为和概况,但不能分析服务器统计信息。电子商务网站需要了解访客的习惯爱好和生活方式。这种情况下的数据挖掘工作需要解决如下商业问题:谁在购买什么商品,购买什么价位的商品。你还应该了解哪些产品比较畅销,从而调整动态目录,规划订单和发货方式。你需要知道哪些销售方式、激励措施、产品和广告切实有效,以及如何设计网站才能获取更大利润。

在这个竞争激烈、人群拥挤的市场中,要了解和熟悉客户,关键是挖掘数据。在竞争异常激烈的市场中,有策略地利用客户信息才是生存的重点。

商人从网站和数据库收集的信息可以反映哪些产品具有交叉销售的机会,或者根据访客性别、年龄、统计和生活方式,应该提供哪些相应的信息和措施。这个过程如下:从服务器日志、Cookies和表单捕捉重要的客户信息,添加到家庭信息和动态信息中;然后,利用神经网络、技术学习和遗传算法等强大的模式识别技术,预测客户的购买倾向。

利用数据挖掘工具和机器学习算法,可以将网站数据库分割为多个顾客群,每个顾客群具有自己特定的行为。同时还能从统计学角度对数据进行测试,在没有分析人员和销售人员参与的情况下,将市场分为多个区域。这些数据挖掘工具

可以自动识别数据间隔和范围,区分好坏。这些工具通常以决策树或“如果/那么”的形式输出结果。商人可以利用这种“Web”挖掘方式,以商业规则的形式预测客户的盈利情况;而商业规则可以直接从Web数据抽取获得。

如果搜索关键词为“PC 软件”、性别“男”、年龄“24-29”,那么可以预测平均销售额 ≤ 267.26 美元。

或者,

如果搜索关键词为“数学软件”、搜索引擎“雅虎”、子域AOL,那么可以平均销售额 ≤ 379.95 美元。

另外,也可利用数据挖掘工具和反向传播神经网络预测顾客的购买倾向。神经网络可以用于构建顾客行为模型,预测哪些人会购买商品、或者可能投入多少资金。学习能力是神经网络的一大特征。神经网络并非事先设计好的,而是需要驯化。神经网络通过驯化样本,构建预测模型,为顾客在购买倾向“打分”。

通常,神经网络会观察数据关系,然后再“驯化”。

据Web挖掘方案采取措施

你很可能需要在独立的分析专用服务器中执行Web挖掘。分析之后,你需要通过某些生产系统验证结果,如利用电子邮件完成销售测试运动。请注意,比较电子邮件、物理邮件或电话销售的成本,可以快速评估Web挖掘和销售效果。当

然,评估Web挖掘项目非常简单,给潜在客户发送邮件只需花费5美分,而直接寄信需要5美元,电话销售则需要8到12美元。以前,规划和执行传统的销售运动需要几个月的时间;今天,利用网站、电子邮件只需几小时即可。由于网站可以快速评价预测模型,从而加速了一对一销售的发展和Web挖掘结果的论证。

评估 Web 挖掘效果及其投资收益率并不复杂。只需考虑Web挖掘前后的广告或横幅点击率。考虑产品信息的售出率和回应率,以及分析Web挖掘带来的购买量。考虑执行数据挖掘前后的购买率。

如果你刚开始根据数据挖掘分析结果发起电子邮件销售运动,请考虑回应比例,将邮件分离为两类:一类是你的分析所针对的人员,另一类则是排除的人员。查看Web挖掘分析针对的人员比排除的人员在回应率和销售率上具有多大优势。

你所处行业和市场的情况决定你应该多久挖掘一次网站数据。数据挖掘间隔取决于客户修改属性的频率。例如,银行可能拥有交叉销售模型,供调用站点使用,而且在几个月内始终有效。可能需要按季或按月创建银行模型,同时还与银行试图回答的业务问题相关,如CD、银行卡、贷款等金融产品的交叉销售机会。

对搜索引擎等门户而言,模型可能需要每周更新,因为内容、客户及其特征等可能会比银行客户变化更快。这些门户预测的终端产品同样容易频繁更改——对银行而言,终端产品是指贷款;对门户而言,则是广告。

集中完成Web挖掘

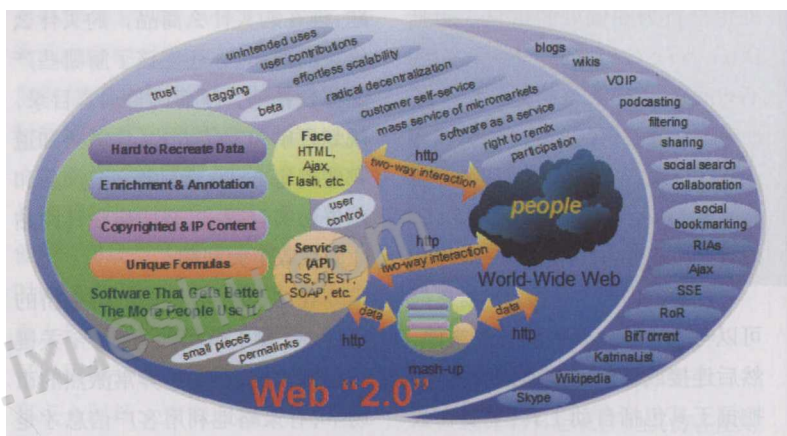
利用软件分析邮件分离情况,

将邮件产品分层,提供给之前通过挖掘分析得到确认的潜在客户。利用目标邮件,为那些有可能对你的产品和服务感兴趣的访客提供鼓励措施。请记住,将邮件发送给对你的产品了解甚少的人员就无异于垃圾邮件。自动回复、路由、管理和分离电子邮件,你就可以通过邮件有效地回应顾客,实现直接销售。

佳潜在客户的概况,包括动向、品位以及邮箱。通过供应链有效管理后台物流。

大量的商业交互活动最终汇总为收入,因此,在这个充满活力的市场中,顾客和商家之间存在十分微妙的关系转换。神经网络、计算学习、遗传算法等人工智能工具在设计时模仿了人类的技能,利用这

Elements of the Web's Next Generation



Web 2.0 原理图

通过自动或分离的邮件,为顾客提供便捷服务。

利用你对Web挖掘的分析了解顾客的动向、偏好、价值观和生活方式。将你对顾客的了解合并在一起,并采用合并后的基调、方式和方法与他们交流。找出当前顾客和潜在顾客的共同属性。无论是在线联系还是离线联系,都应管理好顾客的联系方式。在顾客购买产品和服务时,你可以通过销售要求、会议、电话或邮件询问与顾客联系,都应该集中客户和交易资料。

追踪销售广告,了解广告效果及其原因。监测哪些广告起到效果,哪些确实售出了产品。明确最

些工具挖掘网站数据,可以在全新的、不断进化和突变的商业环境下,尽力识别、预期和了解顾客的购买习惯和偏好。

在网络经济时代,零售商必须适应和接纳客户的需求,这一点最为重要。在当前不断扩张、竞争激烈且反复无常的环境下,Web挖掘已成为一项非常重要的流程,影响到零售商能否取得长远成功;如果不能快速反应、适应和发展,就会使顾客在点击鼠标时陷入“搅拌状态”。电子销售反映了顾客和商家之间的数据交换越来越频繁,并且不断进化和改变——就像生物体逐步发展神经系统一样。^[5]



论文写作，论文降重，
论文格式排版，论文发表，
专业硕博团队，十年论文服务经验



SCI期刊发表，论文润色，
英文翻译，提供全流程发表支持
全程美籍资深编辑顾问贴心服务

免费论文查重：<http://free.paperyy.com>

3亿免费文献下载：<http://www.ixueshu.com>

超值论文自动降重：http://www.paperyy.com/reduce_repetition

PPT免费模版下载：<http://ppt.ixueshu.com>

阅读此文的还阅读了：

- [1. 基于fuzzing的Web漏洞检测工具设计与实现](#)
- [2. 基于Web的机械设备设计、制造及管理系统信息化控制软件的设计与实现](#)
- [3. 基于制造业CAX/PDM/ERP集成研究](#)
- [4. 优化网络应用性能的电量采集与分析系统设计](#)
- [5. Framework of Web Data Integrated LBS Middleware](#)
- [6. Web数据挖掘](#)
- [7. Web的数据挖掘](#)
- [8. 互联网3D技术的实现方式](#)
- [9. 软件测试在Web开发中的应用](#)
- [10. Data Migration for Ontology Evolution](#)
- [11. 基于Web的在线复杂公式编辑器的实现](#)
- [12. 基于web的九江学院二手物品交易平台的设计与实现](#)
- [13. 如何在工业自动化中使用Web服务](#)
- [14. 基于web的面向对象模型库系统](#)
- [15. Web结构挖掘](#)
- [16. 构建基于Web的企业级电子邮件系统](#)

17. 一种wap与web交互式博客的实现方案

18. 基于ASP.NET+ORACLE的节水管理系统

19. 基于房产信息导航系统在高校房屋管理中的应用

20. 基于web的科技规划决策支持系统研

21. 浏览器中applet的访问数据库的方法探讨

22. Web环境下高校教学任务分配系统的开发

23. JavaScript调试方法以及常见错误

24. 基于web的实验报告提交系统

25. 基于web的肺结核病人规范管理转诊系统的设计与实现

26. 基于XML的Web挖掘

27. 基于职业仿真性网络实验室建设方案探讨

28. Web挖掘研究

29. 机械产品设计开发web方式的PDM集成

30. 基于Web的天然气计量远程实时管理系统

31. 挖掘Web潜力

32. 数据挖掘、Web挖掘与Web日志挖掘之研究

33. 基于java的web应用程序的性能测试

34. 基于web的智能教学系统研究

35. Web挖掘技术

36. 数据挖掘的妙用

37. Delphi下Web数据库发布

38. 惠普正式宣布放弃web OS平板，举世震惊

39. Web挖掘研究

40. 巧用预览把关迅雷下载

41. 基于 μ Clinux的嵌入式Web服务器的设计

42. 挖掘Web数据

43. Web挖掘研究

44. 我院教学评价系统的开发与应用

45. 数据库中间件技术及在三层客户机/服务器模型中的实现

46. Web数据挖掘

47. 基于web和关系数据库的super-Domain系统的实现

48. Web挖掘研究

49. 利用Java Servlet实现在Web服务器平台上的开发

50. NET在远程测试仪器WEB中的应用