

基于点击流技术的个性化信息服务研究

Research on the Clickstream - based Personalized Information Service

杜文华

(中南民族大学管理学院 武汉 430074)

摘 要 分析了目前个性化信息服务的机制及缺点,提出了基于点击流技术的个性化信息服务方案和体系结构。

关键词 个性化信息服务 点击流技术 数据挖掘

个性化信息服务就是将用户感兴趣的信息主动提交给用户。个性化信息服务既是一种个性化服务,又是一种信息服务。个性化信息服务首先应该是能够满足用户的个体信息需求的一种服务,即根据用户提出的明确要求提供信息服务,或通过对用户个性、使用习惯的分析而主动地向用户提供其可能需要的信息服务。其次,个性化信息服务应能够根据用户的知识结构、心理倾向、信息需求和行为方式等来充分激励用户需求,促进用户有效检索和获取信息,促进用户对信息的有效利用并在此基础上进行知识创新。本文所倡导的个性化服务是为用户创立和管理自己的信息或兴趣群组,并提供分析统计服务,跟踪用户以浏览信息的历史记录,帮助用户及时获取最需要的信息。

1 目前个性化信息服务的机制及缺点

目前网站提供的各类搜索引擎往往适用于短暂的随机性查询,它们一般是利用用户提供的关键词进行搜索,返回系统认为相关的文档。然而用户的个性化信息需求是相对稳定的、时间相对长久的信息需求,与那些随机的、临时的查询有所区别,因为它还是比较稳定的需求,所以需要保存这种需求,并可以根据用户对于所推荐信息的兴趣程度加以修改,以便获得更好的效果。所以在互联网上出现海量信息后,通过搜索引擎获得的信息仍然不能满足用户高效获取信息的需要。

早期的个性化 Web 信息服务主要由新闻剪裁、股票报价和目录推荐等内容组成,这些内容都是通过浏览器的 Cookie 文件和一些信息服务者提供的智能后端工作实现的。目前较流行的个性化信息服务,主要是网站为用户通过关键字进行界面定制建造创立和管理自己的信息或兴趣群组,一些网站主要是让用户键入所需信息的关键词进行查找,还有就是通过词表导航,旨在满足用户的检索需求而提供的一种检索帮助,在线用户实施检索信息的过程中,系统自动地显示与其相关的词,并按出现词频高低排序。虽然,目前一些网站的个性化信息服务的做法日趋流行,但都有致命弱点,即局限性太强。当前信息服务局限性主要表现在:a.简单的关键词匹配,往往输出大量的信息,而真正相关的信息却很少,更不用说快速准确地检索出用户需求信息,没有用户兴趣的保存和维护的功能,这使得用户耗费更多的时间和精力处理一些毫不相关的信息。b.基于关键字的需求模型不能全面地、准确地反映出用户的兴趣。原因是系统内部通用性分类并不一定和用户个性化需要信息分类一致。信息用户如果不能很好地理解网

站信息分类和信息关键字的抽取方法,可能自己填写的关键字与网站的信息特征有偏差,同样网站提供信息分类和关键字可能不能准确体现用户信息需求,造成信息推荐产生偏差。c.定制方式不能满足信息用户的动态信息需求。当用户的信息需求转移时,根据经验用户一般情况是不去修改自己的资料,或者是不能及时更改自己信息需求资料,这样缺乏用户信息兴趣跟踪的信息推荐就不能准确体现信息用户当前的信息需求。

随着 Web 领域内人工智能技术、数据库技术、知识发现技术的不断完善和发展,专家学者及一些实践者不断对个性化信息服务功能进行改进。数据挖掘工具的发展和点击流技术的出现,能够对个性化信息服务功能进行极大的提升,应用的方式主要是对用户浏览网页和信息的 Web 记录进行数据挖掘、分析从而得出用户对信息的兴趣和偏好知识,不断地进行提供信息内容的调整,动态地为用户提供所需的信息。根据用户的兴趣类型和信息需求,为用户提供更为细致有效的帮助是很有必要的,只有这样信息服务者才能为用户提供真正意义上的个性化信息服务。点击流技术的出现能够很好地跟踪用户历史行为,进而分析用户访问信息的兴趣,使个性化信息服务更精确。将点击流技术应用到个性化信息服务势必带来全新的信息服务模式,有效提高服务的质量和满意度,拓展信息服务的发展空间,并有可能带来信息服务产业的变革。

2 基于点击流技术的个性化信息服务研究

2.1 点击流 点击流是指用户在网上冲浪结束后所留下的实际的踪迹。一个点击流就是一个用户在网络上的行为记录,这个记录中包括用户浏览的每一个站点及站点内的用户浏览的每个页面,用户在网站或页面上停留了多长的时间,用户按照什么次序浏览的页面,用户参与的任何一个新闻组和用户发送的 E-mail 地址。通过采集用户在站点上运动情况的方法,可用于跟踪记录访问过的链结点,包括用户的来源地点、浏览站点的路线和最终到达的目标。链接分析包括对点击过的链接的观察、它们在屏幕上的相关位置、用户在网页上停留的时间以及点击过的链接间的关系和最终结果(用户是否发生了电子交易等)。通过对这些数据的有效分析,能够对网站的建设起到指导作用,增强网站的粘着度。当网络的用户查询一些信息或无意间随意地点击网页时,就为商家留下了可利用的分析资料。

无论是 ISP 服务商们还是网站都有能力捕获用户的点击流。

在电子商务网站环境中,点击流的分析越来越有价值,已经远远超出点击流的范围,对这些数据深层次分析已经成为电子商务网站了解经营状况、了解用户行为的有效工具。点击流分析以 Web 上的点击流数据为基础,利用 OLAP、数据挖掘等技术满足电子商务企业的所有人员(市场、销售、工程与管理)的需求。不同的部门有不同的需求,通过对点击流数据不同角度的分析来达到不同的目的。本文中引入点击流技术主要是解决用户信息兴趣特征集合的分析。

2.2 基于点击流技术的个性化信息服务方案 基于点击流的个性化信息服务方案主要是对电子商务网站信息用户浏览信息的行为采用点击流技术进行记录、采用数据挖掘的方法,分析得出用户在不同时期内的准确而全面的信息兴趣。同时为了弥补方案初期用户浏览信息记录数据的不足、分析不准确的缺点,在用户成为网站用户的初期采用用户定制信息的方式。具体实现机制是:a. 在用户最初注册时根据用户填写的基本信息,为用户提供个性化信息推荐服务;b. 利用点击流技术对用户信息行为记录(如搜索关键字记录,信息浏览记录,信息回复记录等等)进行跟踪,通过建立点击流数据仓库、点击流数据集市抽取用户信息行为的相关数据,再利用商务智能工具进行用户当前信息需求偏好的分析,对用户需求的类型加以修正、补充;c. 再通过点击流技术对用户推荐的信息行为跟踪,对分析模型进行评测、改造,直至提出一个完善的用户信息行为分析模型。

基于点击流技术的个性化信息服务方式和目前的个性化信息服务方式相比的优点主要体现在以下三个方面:a. 动态地跟踪用户访问信息兴趣进而为用户提供所需信息。点击流技术对用户信息行为的跟踪,根据用户最新浏览信息记录及时反映出用户信息需求的变化,及时调整推荐信息,更加贴合用户的需求,不会因为用户没有及时更新信息定制内容,而总是给用户发布垃圾信息。b. 准确地为用户提供所需信息。根据用户信息行为记录分析出的信息兴趣特征值组合实际上是提供了高级的信息查询方式,能够为用户提准准确全面的信息。在信息定制服务中用户可能对网站信息的分类不是太了解,关键字需求定制模型不能全面地、准确地反映出用户的兴趣,而基于点击流技术的个性化信息服务方式对用户信息行为的跟踪更能反映出用户对信息的实际需求。c. 无需用户定制或更新信息需求。传统的信息定制服务方式要求信息用户及时提供信息改变需求才能发送正确信息,基于点击流技术的个性化信息服务方式能够自动分析用户的信息需求变化,发送信息。

进行点击流分析可以根据 Web 日志记录中用户浏览的信息内容进行信息类别挖掘,虽然短期来说这种信息需求分析是杂乱无章的,但长期来说这种需求是稳定的。而且这种根据用户信息浏览行为的分析会和网站的信息分类、关键字含义一致,能更准确地、全面地反映出用户的信息需求。同时根据分析的时间维度设计,这种信息需求分析能够得出目前这段时期内用户的信息需求。

2.3 基于点击流技术的个性化信息服务结构体系 如前所述,个性化信息服务就是将用户感兴趣的信息主动提供给用户。要实现这项服务,服务系统必须具备两方面的能力:一是构建用户兴趣信息分析模型,即跟踪用户信息浏览行为,从而得出用户的信息兴趣集合;二是实施信息推荐,即网站根据分析得出用户

信息兴趣集合向信息用户推荐信息。所以基于点击流技术的个性化信息服务的实现过程主要包括两个部分,用户信息兴趣分析部分和信息发布部分。本文所描述的用户信息兴趣分析主要使用 Web 内容挖掘和 Web 使用记录挖掘对 Web 日志文件和网站信息页内容进行挖掘,得出加权的用户曾访问过信息类和信息关键字,然后通过聚类的方法得出用户的信息兴趣集合来判定用户感兴趣的信息。然后通过信息推荐平台把网站的信息页和分析得出的信息兴趣集合进行比对,符合一定条件的信息将由系统自动推荐给用户。系统的体系结构见图 1,图中分为两个部分,用户信息访问兴趣分析部分和信息发布部分。由于每个用户都有不同的登录时间和兴趣类型,信息发布是根据用户访问行为的不同实时地进行信息发布的,因此信息发布部分为在线处理部分,而用户访问信息兴趣分析部分则为离线处理部分。

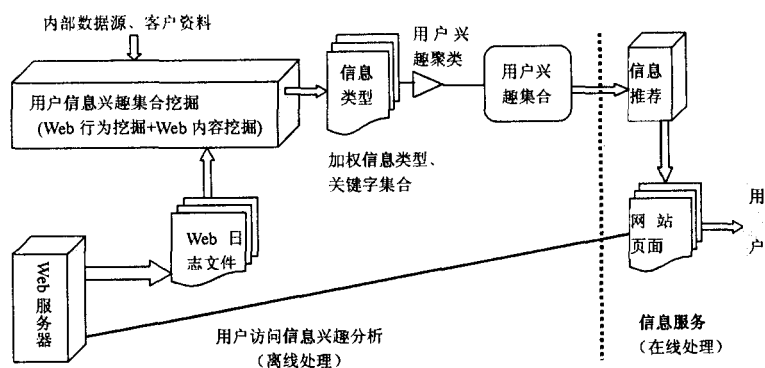


图 1 基于点击流技术的个性化信息服务体系结构

3 结 论

可以看出,用户个性化推荐系统就是通过点击流技术来收集和统计用户访问过的历史数据,挖掘当前用户感兴趣的页面,获取用户兴趣模型,以便在用户以后的访问过程中根据兴趣模型自动向用户推荐内容,指导用户的浏览行为,提高浏览信息效率。可见,推荐系统的首要任务是推荐结果的准确性,如果总是向用户推荐不感兴趣的信息将会丧失用户的忠诚度;其次推荐的信息应尽可能多地覆盖用户的实际兴趣范围,以最大程度地提高推荐的效果。这些都对分析算法的准确性和分析统计的实时性提出了较高的要求,在分析过程中也需要与用户实时交互,及时调整分析和推荐方案,提升推荐信息的效果。

参 考 文 献

- 1 张玉峰,晏创业. 基于 Agent 的个性化信息服务模型研究. 情报学报, 2001;(5)
- 2 饶增阳. 网络环境下的个性化信息服务. 情报探索, 2004;(1)
- 3 What is Personalization. <http://www.personalization.org/faqsl/html#Questionone/>, 2003
- 4 Eileen Colkin, Personalization Tools dig Deeper. <http://www.informationweek.com/story/IWK20010823S0003>, 2003
- 5 Yoshiyuki Koseki. Developing Personalized Information Services on the Internet. <http://www.computer.org/proceedings/mse/8925/89250086abshtml/2003>
- 6 Jakob Nielsen. Personalization is Over-Rated. <http://www.useit.com/alertbox/981004.html>, 2003
- 7 毛 军. 网络环境下的个性化信息服务. 计算机周刊, 2001;(4)
- 8 Clickstream, www.webopedia.com/TERM/C/clickstream.html

(责编:阳桦钧)