

基于本体的电子商务个性化推荐的点击流信息资源开发利用研究

易 明, 张金隆

(华中科技大学 管理学院 武汉, 430074)

摘 要:首先提出了点击流信息资源开发利用的方法体系,并以基于本体的电子商务个性化推荐为例,分析了语义层次的点击流信息资源开发利用的方法。该方法保留了概念属性之间以及概念之间的关系,支持使用文档与用户当前会话的柔性匹配,能有效提高个性化推荐集的准确度和覆盖度。

关键词:点击流;信息资源开发;电子商务

中图分类号:F724.6

文献标识码:A

文章编号:1001-7348(2007)03-0143-06

0 前 言

与传统的基于内联网的信息系统相比,基于 Web 的信息系统的一个重要特征就是 Web 用户的高度自治性,这种高度自治性使得 Web 用户具有很强的心理优势,从而真正成为“上帝”。因此,对 Web 用户点击行为的分析与应用就成为当前研究的热点问题。事实上,对用户点击行为进行分析与应用,关键在于如何开发利用反映用户点击行为的点击流信息资源。本文首先应用我国著名信息学家钟义信提出的信息运动过程模型,构建了点击流信息运动过程模型,以此为基础,架构点击流信息资源开发利用的方法体系,并以基于本体的电子商务个性化推荐为例,分析语义层次的点击流信息资源开发利用方法。

1 点击流信息的层次

依据万维网委员会的定义^[1],页面视图就是指一次显示在用户浏览器的所有文件,包括多个页面文件、图片、插入的组件以及脚本文件等,而点击流就是指用户的一系列页面视图请求。本文所分析的点击流是指用户访问特定 Web 站点的一系列页面视图请求,也就是交易事务文件。而点击流信息资源是一种狭义的信息资源,特指点击流信息。

1.1 本体论层次的点击流信息

从本体论层次来看,点击流信息就是 Web 用户(本体论信源)点击行为所产生的本体论信息,它是用户点击行为状态及其变化方式的自我表述。本体论层次的点击流信

息不会因为认识主体(Web 管理者)的存在与否而转移。事实上,Web 管理者往往不能现场观察用户的点击行为,所以无法直接获取本体论层次的点击流信息,也正因为如此,点击流信息一直都没有得到 Web 管理者足够的重视。

1.2 认识论层次的点击流信息

虽然,Web 管理者无法直接观察用户的点击行为,但是可以通过配置站点服务器、代理服务器以及客户机(认识论信源)来产生反映用户点击行为的记录(如 Web 服务器日志),也就是所谓的信息变换。根据 Web 管理者认识深度的不同,可以将认识论层次的点击流信息进一步划分为语法层次、语义层次和语用层次 3 种。其中,语法层次的点击流信息是 Web 管理者基于各种认识论信源而感知(表述)的反映用户点击行为的各种形式上的信息,也就是反映用户点击行为状态及其变化方式的各种符号。如 Web 管理者利用 Web 服务器日志可以观察到特定用户在某个时间访问了某个(或几个)页面。语义层次的点击流信息是 Web 管理者进一步感知(表述)的反映用户点击行为含义的各种信息。如 Web 管理者可以进一步认识到特定用户访问的某个(或几个)页面的具体内容。语用层次的点击流信息则是 Web 管理者所感知(表述)的语法层次或语义层次的点击流信息的效用信息。如 Web 管理者可以认识到特定用户访问的这几个页面对于个性化推荐所具有的价值。

2 点击流信息资源开发利用的方法体系

点击流信息资源开发利用的实质就是点击流信息运

收稿日期:2006-07-25

作者简介:易明(1978-),男,华中师范大学信息管理系电子商务专业教师,华中科技大学管理学院管理科学与工程专业博士研究生,研究方向为电子商务站点数据挖掘。张金隆(1952-),男,教授,博士生导师,华中科技大学管理学院院长,研究方向为电子商务、信息资源管理等。

动的过程,可以应用信息运动过程模型进行分析^[2],如图1所示。为了便于研究,图1只描述了最为关键的4个环节。第一个环节是点击流信息获取,即Web管理者通过一定的信息变换手段将本体论信息脱离用户以实现认识论信息的获取——交易事务文件集。第二个环节是点击流信息认知,即Web管理者借用一定的信息处理手段,对点击流信息(交易事务文件集)进行加工,从中提炼出能够帮助Web管理者解决实际问题的知识——用户点击行为模式,也就是点击流信息资源开发。根据认识主体的认知能力不同,信息认知可以细分为语法层次的信息认知、语义层次的信息认知和语用层次的信息认知。如Web管理者开发点击流信息资源的直接目的就是解决用户点击行为的优化问题,而个性化推荐就是解决这些问题的重要方法。由此,Web管理者可以从3个层次分别提炼出实现个性化推荐所需要的知识。第三个环节是点击流信息再生,即在信息认知基础上形成解决实际问题的各种策略,这一环节属于点击流信息资源利用的范畴。由于信息再生是以信息认知为基础,而且在实际应用过程中这两个环节往往是集成在一起的,所以信息认知的方法层次将直接决定信息再生的方法层次,也就是说,Web管理者可以从以上3个层次产生电子商务个性化推荐策略(个性化推荐页面集)。第四个环节是点击流信息施效,即施效执行主体通过实施这些策略以优化用户点击行为。这一环节也属于点击流信息资源利用的范畴。基于以上分析,点击流信息资源开发利用方法可以分为语法、语义和语用3个层次。

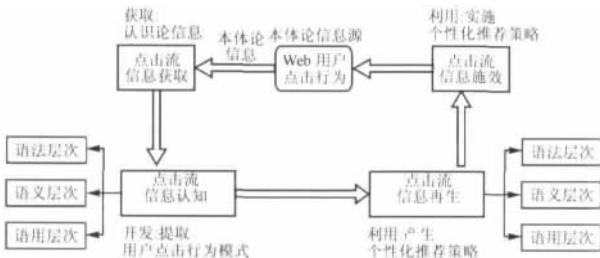


图1 点击流信息运动过程模型

2.1 语法层次

语法层次的点击流信息资源开发利用方法,也就是基于Web使用挖掘的点击流信息资源开发利用方法,是当前研究的重点。目前,这种层次的方法已经形成了通用的模式,即以Web使用数据(主要是Web服务器日志)为挖掘对象,利用各种Web使用挖掘方法(如聚类分析、关联规则、序列模式等)提炼出各种用户点击行为模式,然后根据用户当前会话与这些用户点击行为模式的匹配情况,实现面向用户的各种Web应用。

由于这种层次的方法仅利用了语法层次的点击流信息,只能在纯粹形式化的水平上进行识别、推理和判断,因此,这种点击流信息资源开发利用方法实现起来虽然较为简单,但最终形成的各种Web应用系统的准确度较低,用户兴趣满足程度也低。另外,如果Web的内容是在有规律的变化中,或者用户当前所点击的页面比较少,那么这种层次的方法将会面临很多问题。

2.2 语义层次

语义层次的点击流信息资源开发利用方法是当前研究的前沿。这种方法需要首先将语法层次的点击流信息转化为语义层次的点击流信息,如可以利用站点领域本体,将语法层次的点击流信息表示为站点领域本体中各个概念的实例及其权重的形式,然后再利用相关的数据挖掘算法,从中提取知识,进而实现面向用户的各种Web应用。由于这种层次的方法不仅利用了语法层次的点击流信息,而且整合了语义层次的点击流信息,可以进行语义层次的逻辑推理,所以这种方法实现起来虽然较为复杂,但最终所形成的Web应用系统的准确度相对较高,能够有效弥补语法层次方法的诸多不足。

2.3 语用层次

目前,关于语用层次的点击流信息资源开发利用方法的研究成果很少。这种方法主要依据Web页面对满足用户需求的效用度来实现各种面向用户的Web应用,准确度与效用度最高,其关键在于如何构建面向此层次Web访问的用户效用函数。在当前的技术条件下,用户效用函数的构建需要Web应用系统与用户不断地进行交互,所以成本相对较高,而且还会干扰用户的正常点击。所以,从经济的角度考虑,在当前的技术条件下,利用语用层次的点击流信息还需适度进行;从发展的眼光来看,随着技术的不断进步,为了获得语用层次点击流信息所需要付出的代价会越来越来少,而由此带来的收益则会越来越多。

3 基于本体的电子商务个性化推荐模型

如图2所示,基于本体的电子商务个性化推荐模型分为离线处理和在线推荐两个模块。

离线处理阶段的核心任务,就是在构建站点领域本体的基础上生成语义层次的使用文档。语义层次使用文档的生成有两种方法。第一种方法就是首先利用站点结构数据预处理的结果和站点服务器日志,完成站点使用数据预处理工作,即从大量的Web日志中识别用户与会话并进行路径补充,以形成语法层次的交易事务文件集;其次,利用聚类算法对语法层次的交易事务文件进行聚类,并利用每个聚类的质心点矢量作为代表,以形成语法层次的使用文档;最后,利用站点领域本体,将语法层次的使用文档转换为语义层次的使用文档。第二种方法的原理与第一种基本类似,它是利用站点领域本体产生语义层次的交易事务文件集,然后直接生成语义层次的使用文档。

在线处理阶段,首先需要利用站点结构数据预处理的结果,分析当前服务器日志文件,以形成语法层次的用户当前会话;其次,利用站点领域本体,将语法层次的用户当前会话转化为语义层次的用户当前会话;最后,将语义层次的使用文档或同语义层次的用户当前会话进行匹配,形成能够反映用户兴趣的匹配文档或,并将匹配文档转化为实际的Web页面,即产生个性化推荐集或;或者直接依据用户当前会话的语义特征,在站点领域

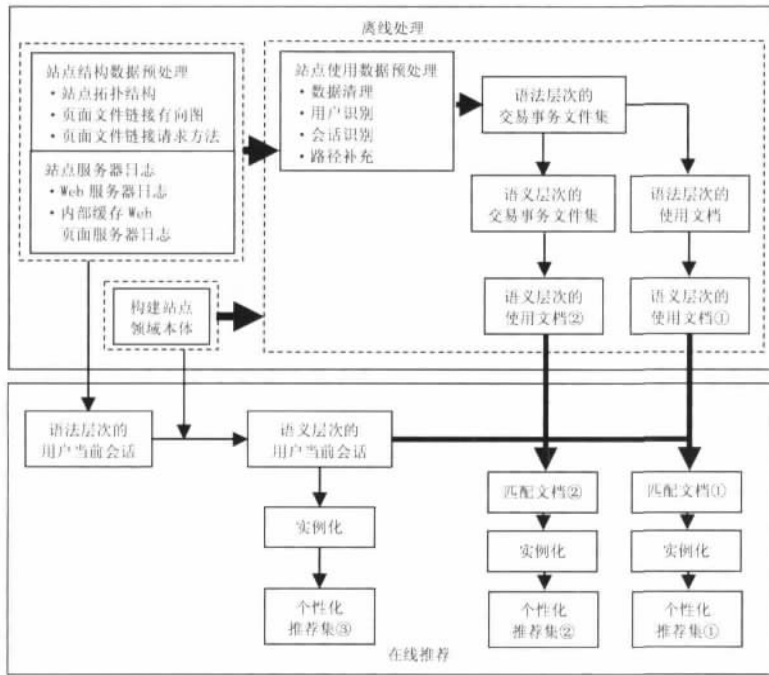


图 2 基于本体的电子商务个性化推荐模型

本体中寻找相似的对象,并将这些对象转化为实际的 Web 页面,以个性化推荐集。

考虑到个性化推荐集、的生成方法在本质上是一致的,同时为了便于研究,本文重点分析个性化推荐集的形成过程。

4 站点领域本体的构建

目前,领域本体的构建方法主要有 IDEF5 方法、骨架法、TOVE 法、KACTUS 工程法、METHONTOLOGY 法、SENSUS 法以及七步法等。这些方法都是在具体开发过程中总结出来的,应用领域很有限,方法细节也比较粗略,而且相关技术较少,因此有着一定的局限性。汪方胜、侯立文等通过借鉴以上 7 种本体构建方法,在遵循本体建立准则的基础上,抽象总结出了一套领域本体的知识工程构建方法^[9]。这种本体构建方法比较符合人类的思维认知,逻辑性强,具有很强的可操作性和可拓展性,并且在实践中得到了检验。本文将以此种方法为依据,说明站点领域本体构建的一般过程。

4.1 确定站点领域本体的范围与领域

首先,应明确站点领域本体将覆盖的专业领域,本体的目的、作用及其系统开发、维护和应用对象。这些对于领域本体的建立过程有着很大的关系。以当当网为例,它是全球最大的中文网上图书音像商城,经营 30 万种图书、CD/VCD 及众多的游戏与应用软件等商品,因此,当当网领域本体将覆盖其零售商品领域。构建本体的目的是揭示与当当网相关的商品、用户所固有的属性以及商品之间、商品与用户之间的复杂关系,进而与电子商务个性化推荐过程进行有机整合。

4.2 列举站点领域本体中重要的术语和概念

在站点领域本体创建的初始阶段,应尽可能地列举出系统想要陈述的重要术语和概念,如图书、电影、音乐、作者、导演等。在此阶段,不需要关注这些术语和概念之间的意思是否重叠,也不需要考虑这些概念到底用何种方式来表达。

4.3 构建站点领域本体框架

上一步骤中已经产生了与当当网涉及领域相关的无序的词汇表,这时需要按照一定的逻辑规则把它们进行分组,形成不同的工作领域,以使得同一工作领域的概念具有较强的相关性。另外,对其中每一个概念的重要性进行评估,选出关键性术语,摒弃那些不必要或者超出领域范围的概念,从而形成一个领域知识的框架体系,得到站点领域本体的框架结构。

4.4 设计站点领域元本体

元本体是指本体的本体,其术语用于定义本体中的概念,如实体、关系、角色等。它可以说是更高层次的本体,是领域内概念的抽象。在设计元本体时,应尽量做到领域无关性,并且包含的元概念数目要尽可能的少。除了概念,还要定义概念之间的关系,这些关系不仅仅涉及同一工作领域的概念,不同工作领域的概念也可以相关,只是这些关系总是属于某一个工作领域。

(1) 定义类及其层次体系。创建的概念中,很大一部分属于类,而对类的层次的定义主要有自顶向下法、自底向上法和综合法 3 种。这 3 种方法各有利弊,采用哪种方法取决于开发人员对领域认识的角度。由于当当网所销售的都是一些常用的零售商品,人们对这些商品都有自顶向下的认识,所以建议采用自顶向下法。

(2) 定义类的属性。仅仅通过类还不足以确切地描述一个领域,一旦定义好了类就还要进一步描述这个类的内部结构。在从前面列举的站点领域本体中的重要术语、概念中选择了类之后,大部分剩下的概念就可能成为这些类的属性。此外,还需要描述类中的个体成员与其它类之间的关系。如“图书”这个类就包括书名、作者、出版社、出版时间、类型等属性,而属性“作者”实质上也是一个类,它包含姓名、年龄、国籍、代表作等属性。

(3) 定义属性值。定义属性值即说明属性的取值类型、取值个数及有关值的其它特征。

(4) 创建实例。创建实例即选择概念类并创建该概念类的实例。

4.5 对站点领域本体编码和形式化

选用合适的本体描述语言对上述建立的领域本体进行编码和形式化。目前大多数已经建立的本体模型都是基于一阶逻辑,也有的是基于描述逻辑。考虑到由于描述逻辑的语法容易转换成 XML/RDF 形式,因此基于描述逻辑的本体模型更适合 Web 环境下的概念建模与信息查询。

目前,本体的逻辑表示主要有四元素表示法和六元组表示法。前者源于 Gruber 博士的观点,在世界范围内得到了比较高的认同,但是形式过于灵活,不易掌握^[6];后者则是 2002 年由新加坡南洋理工大学的 Myo- Myo Naing 博士在一篇国际会议论文中提出的,其可操作性强,被国内研究者接受的程度比较高^[9]。站点领域本体模型构建完成以后,应选用合适的本体描述语言对上述建立的领域本体进行编码和形式化,如 XML(eXtensible Markup Language,可扩展标识语言)、RDF(Resource Description Framework,资源描述框架)、RDFS(RDF Schema)、OWL(Web Ontology Language,网络本体语言)等。

4.6 站点领域本体的检验评价

站点领域本体被形式化以后是否满足了需求?是否满足本体的建立准则?本体中的术语是否被清晰地定义?本体中的概念及其关系是否完整?这些问题还需要进行检验和评估。由此,站点领域本体建立后,需要反复地将其应用到实际系统中或与领域专家不断交流,出现问题及时反馈,并利用以上方法加以改进和完善。

5 导出语义层次的使用文档

5.1 生成语法层次的使用文档

站点使用数据预处理会产生 n 个语法层次的交易事务文件和 m 个属于交易事务文件的页面记录,每一个页面记录都由相应的 url 唯一标识。

定义 1: $T=\{t_1, t_2, \dots, t_n\}$, T 为站点使用数据预处理后产生的语法层次交易事务文件集。

定义 2: $P=\{p_1, p_2, \dots, p_m\}$, P 为站点使用数据预处理后产生的页面记录集。

定义 3: 语法层次的交易事务文件 $t_i=\langle w(p_1, t_i), w(p_2, t_i), \dots, w(p_j, t_i), \dots, w(p_m, t_i) \rangle$ 为 m 维向量, $w(p_i, t_i)$ 是 p_i P 所代表的页面在 t_i 中的权重, $i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, m\}$ 。

可以利用标准的聚类算法将具有相似使用模式的交易事务文件聚集在一起,以形成 k 个聚类, $T=\{u_1, u_2, \dots, u_k\}$ 。然而,这些经过初次聚类而形成的聚类集合还不能有效地反映站点用户的使用模式,因为,对于大型电子商务站点而言,每个交易事务文件聚类都包含了成千上万的交易事务文件,而每个交易事务文件又是由上百个页面文件构成的。由此,可以将 k 个语法层次的交易事务文件聚类变成赋有权重的页面集合,以生成语法层次的使用文档。

定义 4: $Tpr=\{pr_1, pr_2, \dots, pr_k\}$, Tpr 为语法层次的使用文档集, pr 为语法层次的使用文档。

对每一个交易事务文件聚类 u, 利用 k 个聚类的质心点矢量生成使用文档 $pr^{[6]}$, 以页面和权重二维形式表示为: $pr=\langle p, weight(p, pr) \rangle | p \in P$

其中:

(1) $weight(p, pr)$ 是页 p 面在形成的语法层次的使用文档 pr 中的权重,其计算方法可以表示为:

$$weight(p, pr) = \frac{1}{|u|} \times \sum_t w(p, t)$$

(2) $w(p, t)$ 是页面 p 在每个交易事务文件 t u 中的权重。

5.2 导出语义层次的使用文档

给定语法层次的使用文档 pr, 利用站点领域本体, 通过在每个页面 p 中抽取实例, 可将它转化为语义层次的使用文档 $spr=\langle o_1, w_1 \rangle, \langle o_2, w_2 \rangle, \dots, \langle o_x, w_x \rangle$, o 是站点领域本体中各个概念的实例, w 为权重, 可依据 $weight(p, pr)$ 和站点领域本体中各概念的重要性转化而来。对于大型电子商务站点而言, spr 可能包含成千上万个实例, 同样也需要采取一定的整合方法来约简 spr。由此, 可以考虑将语义层次使用文档中那些属于同一概念的实例进行整合, 以形成简化的语义层次使用文档, 这样也会大大减轻推荐引擎的计算负担^[7]。

定义 5: $C=\{c_1, c_2, \dots, c_f\}$, C 为站点领域本体的概念集, f 为概念总数。

如果将 spr 中属于同一概念的实例归为一类, 将形成 f 个子类, 即: $spr=\{g_1, g_2, \dots, g_f\}$, $g_i=\langle o_1^i, w_{o_1^i}^i \rangle, \langle o_2^i, w_{o_2^i}^i \rangle, \dots, \langle o_y^i, w_{o_y^i}^i \rangle$, $i \in \{1, 2, \dots, f\}$ 。其中, o^i 为概念 c_i 的实例, y 为概念 c_i 的实例总数, $w_{o^i}^i$ 为概念实例对应的权重 ($i=1, 2, \dots, y$)。表 1 给出了语义层次的使用文档 spr 中概念 'Book' 的一个例子, 后文将以此表中的数据为例进行分析。

表 1 概念 'Book' 示例 1

o^{Book}	w_o^{Book}	name	author	publisher	year	genre	...
Book1	1	{name1}	{A 1}	{publisher1}	{2001}	图书 科技 计算机/网络 网络与数据通信 电子商务、电子政务	...
Book2	0.8	{name2}	{B 0.6; A 0.4}	{publisher2}	{2002}	图书 科技 计算机/网络 网络与数据通信 网络配置与管理	...
Book3	0.6	{name3}	{C 0.5; B 0.3; D 0.2}	{publisher3}	{2003}	图书 科技 计算机/网络 数据库 数据仓库与数据挖掘	...
Book4	0.3	{name4}	{D 0.6; A 0.4}	{publisher2}	{2004}	图书 科技 计算机/网络 网络与数据通信 网络协议	...

将 spr 分割为 f 个子类后, 问题就转化为如何整合每个 g_i 。此时需要为概念 c_i 的每个属性 a 赋予一个整合函数 φ_{∞}

对于属性 'name', 其整合函数 φ_{name} 就是一个联合操作, 所有的 'name' 属性值将被联合在一起, 并通过累加产生每个属性值的权重。比如将 φ_{name} 应用到 $\{\langle \{name1\}, 1 \rangle, \langle$

{name2}, 0.8>, <{name3}, 0.6>, <{name4}, 0.3>}, 将会产生一个关于属性 “name” 的整合实例 {<name1, 1>, <name2, 0.8>, <name3, 0.6>, <name4, 0.3>}。

对于属性 “author”, 其属性值包含了一个赋有权重的对象集, 这些对象都是 “Author” 这个概念中的实例。在这种情形中, φ_{author} 就是一个基于矢量的加权操作, 每个对象的权重的计算方法为^[9]:

$$ww_o = \frac{\sum l w_d^i \times ww_o}{\sum l w_d^i}$$

其中, $l=\{1, 2, \dots, y\}$, y 为概念 c_i 的实例总数, w_d^i 为每个概念实例对应的权重, ww_o 是原始属性实例的属性值中每个对象 (author) 的权重, 而 ww_o 也就是整合后每个对象的权重。比如将 φ_{author} 应用到 {<{A, 1}, 1>, <{B, 0.6; A, 0.4}, 0.8>, <{C, 0.5; B, 0.3; D, 0.2}, 0.6>, <{D, 0.6; A, 0.4}, 0.3>} 之中, 将会产生一个关于属性 “author” 的整合实例 {<A, 0.53>, <B, 0.24>, <C, 0.11>, <D, 0.11>}。

对于属性 “publisher” 来说, 其整合函数 $\varphi_{publisher}$ 就是一个联合操作, 所有的 “publisher” 属性值将被联合在一起, 并通过累加产生每个属性值的权重。比如将 $\varphi_{publisher}$ 应用到 {<{publisher1}, 1>, <{publisher2}, 0.8>, <{publisher3}, 0.6>, <{publisher2}, 0.3>}, 将会产生一个关于属性 “publisher” 的整合实例 {<publisher1, 1>, <publisher2, 1.1>, <publisher3, 0.6>}。

对于属性 “year” 来说, 其整合函数 φ_{year} 就是一个联合操作, 所有的 “year” 属性值将被联合在一起, 并通过累加产生每个属性值的权重。比如将 φ_{year} 应用到 {<{2001}, 1>, <{2002}, 0.8>, <{2002}, 0.6>, <{2003}, 0.3>} 之中, 将会产生一个关于属性 “year” 的整合实例 {<2001, 1>, <2002, 1.4>, <2003, 0.3>}。

对于属性 “genre”, 整合函数 φ_{genre} 直接执行基于图的算法来抽取概念层次的部分, 不必赋予权重。比如将 φ_{genre} 应用到表 1 之中, 将会产生一个关于属性 “genre” 的整合实例 {图书 科技 计算机/网络}。

利用整合函数 φ_{σ} 约简后的语义层次 Web 使用文档 nspr 将会形成, 可以表示为: $nspr=\{<o_1, nw_1>, <o_2, nw_2>, \dots, <o_i, nw_i>\}$ 。此时, o 表示经过整合函数处理后形成的每个概念的整合实例, nw 为对应的权重, 可以依据概念在站点

领域本体中的重要性来确定。同时, 表 1 也可以转化为表 2。

6 在线推荐

在线推荐阶段, 同样需要首先利用站点领域本体, 将语法层次的用户当前会话转化为语义层次的用户当前会话 $S=\{<o_{s1}, w_{s1}>, <o_{s2}, w_{s2}>, \dots, <o_{sf}, w_{sf}>\}$ 。其中, o_s 表示经过整合函数处理后形成的每个概念的整合实例, w_s 为对应的权重, 可以依据概念在站点领域本体中的重要性来确定。然后, 利用语义相似性测度方法来匹配语义层次的用户当前会话和语义层次的使用文档。对于任何一个 $nspr=\{<o_1, nw_1>, <o_2, nw_2>, \dots, <o_i, nw_i>\}$, 语义向量 S 和 $nspr$ 的相似性 $Sim(S, nspr)$ 取决于整合后各概念实例对应的语义相似性 $SemSim(s, nspr_i)$:

$$Sim(S, nspr) = \sum_{i=1}^f \alpha_i \times SemSim(o_s, o_i)$$

其中, α_i 为概念 c_i 在站点领域本体中的重要性。对于 $SemSim(o_s, o_i)$, 其计算方法可以表示为:

$$SemSim(o_s, o_i) = \sum_j \beta_j \times Similarity(o_s, a, o_i, a)$$

其中, o_s, a 表示 o_s 的属性 a 的实例, o_i, a 表示 o_i 的属性 a 的实例, β_j 为属性 a 在概念中的权重。当然, $Similarity(o_s, a, o_i, a)$ 的计算需要根据各概念属性的不同而采取不同的方法。

基于以上分析, 可以完成 $SemSim(o_s, o_i)$ 和 $Sim(S, nspr)$ 的计算。由此, 对于每一个语义层次的使用文档 $nspr$ 和语义层次的用户当前会话, 每个整合后的概念实例 o^i 的推荐值 $Rec(o^i, S)$ 的计算方法可以表示为:

$$Rec(o^i, S) = \begin{cases} 0, & \text{if } o^i \in S \\ nw \times Sim(S, nspr), & \text{otherwise} \end{cases}$$

对于所有的语义层次使用文档和语义层次用户当前会话 S , 可以得出语义层次的个性化推荐概念实例集 $URec(S)$, 以整合后的概念实例和推荐值的二维形式表示为:

$$URec(S) = \{<o^i, Rec(o^i, S)> | Rec(o^i, S) \geq \varepsilon\}$$

考虑到同一概念在不同的语义层次 Web 使用文档中, 其包含的 (整合后的) 概念实例一般是不同的, 但可能相同, 所以对于不同语义层次使用文档中

相同的 (整合后的) 概念实例 o^i 的推荐值取最大值。那些推荐值低于一定阈值 ε 的 (整合后的) 概念将被过滤掉。由此, 前文所提及的匹配文档 也就形成。最后, 推荐引擎将匹配文档 与电子商务站点页面再次进行语义层次的匹配, 以寻找合适的 Web 页面推荐给用户, 实现个性化推荐。

表 2 概率 “Book” 示例 2

o^{Book}	name	author	publisher	year	genre	...
Book	{<name1, 1>, <name2, 0.8>, <name3, 0.6>, <name4, 0.3>}	{<A, 0.53>, <B, 0.24>, <C, 0.11>, <D, 0.11>}	{<publisher1, 1>, <publisher2, 1.1>, <publisher3, 0.6>}	{<2001, 1>, <2002, 1.4>, <2003, 0.3>}	图书 科技 计算机/网络	...

参考文献:

[1] R. Cooley, B. Mobasher, and J. Srivastava. Data preparation for mining World Wide Web browsing patterns [J]. *Journal of Knowledge and Information Systems*, 1999(1):5- 32

[2] 钟义信.信息科学原理(第三版)[M].北京:北京邮电大学出版社, 2002

[3] 汪方胜, 侯立文, 蒋馥.领域本体建立的方法研究[J].*情报科学* 2005, 23(2): 241- 244

[4] Stevens R., Goble C.A., Bechhofer S. Ontology- based knowledge representation for bioinformatics [J]. *Briefings in Bioinformatics* 2000(4):398- 414

[5] Naing M. M., Lim E. P and Goh D. H. On extracting link information of relationship instances from a web site[C]. In Proc. of ICWS- Europe, Erfurt, 2003:213- 226

[6] B. Mobasher, Honghua Dai, Tao Luo, et al. Discovery of Aggregate Usage Profiles for Web Personalization [C]. In Proc. of the Web Mining for E - Commerce Workshop (WEBKDD'00), Boston, 2000

[7] H. K. Dai and B. Mobasher. Using ontologies to discover domain- level web usage profiles [C]. In *Second Workshop on Semantic Web Mining, at the 6th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'02)*, Helsinki, 2002:61- 82

[8] B. Mobasher and Honghua Dai. A road map to more effective web personalization: Integrating domain knowledge with web usage mining [C]. In *International Conference on Internet Computing, Las Vegas* 2003

(责任编辑: 来 扬)

A Study on the Exploitation and Utilization of “Click Stream” Information Resource: Personalized Recommendation in Electronic Commerce Based on Ontology

Abstract: This article firstly proposes the exploitation & utilization method system of click stream information resource, and then analyzed the personalized recommendation in e-commerce based on ontology, which is an example of the semantic exploitation & utilization method. This method retains the relationships both among attributes of concepts and among concepts, providing more flexibility in matching usage profiles with current user session, which can improve the precision and coverage of the personalized recommendation sets.

Key Words: Click Stream; information resource; personalized recommendation; ontology

《软件导刊》杂志征稿函

《软件导刊》杂志是由湖北省科技厅主管、湖北省信息学会主办的关于软件开发与管理的学术期刊。自创刊以来,得到了广大读者、作者的广泛关注和大力支持,为进一步加强刊物与作者的互动,提高办刊质量,特向广大作者征稿。本刊国际标准刊号:ISSN1672- 7800,国内统一刊号:CN42- 1671/TP。

1 征稿范围

1.1 管理论坛

介绍最新的软件企业管理理念与项目管理模式,交流与探讨行之有效的管理经验,降低项目风险,提高企业效益。

1.2 软件开发实务

发布最新的软件开发工具和务实的开发技术,增强企业及个人软件开发能力,提高软件产品的开发效率和质量。

1.3 软件技术评述

提供相关软件技术的权威研究报告,对软件行业的前沿技术进行新视角、宽领域的探索,提供软件技术交流平台。

1.4 电子商务与政务

介绍相关政策法规,探讨电子商务发展赖以支撑的信用、认证、支付和现代物流等方面的建设,提升电子商务技术和服务水平,促进企业及公民的电子商务应用意识;推进电子政务建设,

提高行政办公效率和行政管理水平,促进政务公开和廉政建设。

1.5 网络与信息安全

探讨网络开发与应用技术,介绍最新网络安全设备,促进网络技术的发展,提高网络服务能力,为企事业单位信息化提供借鉴。

1.6 数字校园

介绍国内外先进软件产业人才培养模式,探讨如何利用高等教育、网络教育和在职培训体系,培养高素质、复合型软件管理与开发人才。

2 征稿要求

2.1 稿件请发至邮箱: softwareguide@163.com。本刊地址:武汉市洪山路2号湖北科教大厦D座5楼; 邮编:430071; 电话:027-87891823

2.2 来稿要求按出版规范写出论文的中英文标题、作者单位、摘要及关键词,并附上作者简介、通信地址、电话。属各级、各类科研课题的论文请注明项目名称和项目编号。“作者简介”请写明作者出生年、性别、籍贯、工作单位、学位学历、职务职称、研究方向和成果。其它个人资料也可附上,供本刊发表时选用。