

doi:10.3772/j.issn.1000-0135.2010.05.009

基于加权 XML 数据模型的个性化本体研究¹⁾

李树青

(南京财经大学信息工程学院信息管理系,南京 210046)

摘要 随着近年来个性化服务技术的广泛应用和快速发展,个性化加权本体逐渐受到学者的广泛关注。虽然利用个性化加权本体可以更为有效和精确地表达用户兴趣模型,但是相关的设计实现方法仍需进一步研究。文章首先介绍加权 XML 数据模型的概念和相关特点,并利用“同构异值”的权值赋值策略设计了一种基于加权 XML 数据模型的个性化本体构造方法。然后从节点内容、层次结构和权值分配三个方面对其模式表达方法进行了详细介绍,其中重点介绍了权值分配策略及其原理。最后,给出了该种个性化加权本体的相似度比较算法,并对相关测试实验做了必要的说明。

关键词 加权 XML 本体 相似度 个性化

Study of Personalized Ontology Based on Weighted XML Data Model

Li Shuqing

(Department of Information Management, College of Information Engineering, Nanjing University of Finance & Economics, Nanjing 210046)

Abstract Along with lots of applications and rapid development in recent years, personalized weighted ontology becomes the main focus of researchers. Though personalized weighted ontology can express users' interest model efficiently and precisely, correlative methods of design and implementation needs further research. Firstly, this paper introduces the concept and characteristics of weighted XML data model, and designs a method building personalized weighted ontology based on weighted XML data model with homo-structure-hetero-value evaluation strategy. This paper also explains the methods of constructing this model from three points, such as value design, hierarchy design and weight design of which strategy and principle is explained in detail. Finally the similarity arithmetic of this personalized weighted ontology is introduced and some test experiments are discussed for comparison.

Keywords weighted XML, ontology, similarity, personalization

1 引言

随着网络技术的快速发展和 Web 信息资源的急剧增长,Web 用户越来越希望各种网络应用程序能够提供个性化的信息服务。同时,现代 Web 应用系统也广泛地采用个性化信息服务技术来解决网络信息过载问题。因此,用户个性化信息服务技术日

益受到研究者的广泛关注,其中的关键内容在于准确和有效地表达用户兴趣模型,并基于此模型来计算与推荐客体的相似度以提供推荐内容。

早期的用户兴趣模型构建方法主要采用基于关键词的方法。随着近年来本体方法的流行,开展了大量相关的研究^[1,2]。如有学者提出基于本体的用户兴趣模型可以表达成两种主要类型:一是数据视图(Data Diagram),它主要用于描述用户兴趣特征和

收稿日期:2009年7月14日

作者简介:李树青,男,1976年生,讲师,博士,研究方向:Web挖掘和信息检索。E-mail:leeshuqing@163.com。

1) 江苏省教育厅“青蓝工程”基金资助项目。

用户使用模式;二是信息视图(Information Diagram),它主要用于发现 Web 用户的信息需求主题^[3]。还有学者根据用户的信息需求和使用主题的本体模型(Topic Ontology Model),进一步获取用户兴趣中的主题信息,以此进行用户兴趣建模;并且使用模式分类模型(Pattern Taxonomy Model),通过分析用户的反馈信息和对主题本体进行评价,来进一步区分出具有特异性的用户兴趣特征,使得系统可以判断主题和用户兴趣之间的相关程度^[4,5]。

但是,目前本体的构造、映射和进化都不能自动进行,还需要人工干预,这是一个亟待解决的问题,也说明本体研究的复杂性和必要性^[6]。同时,本体间的知识重用与共享也成为一个问题,常见的处理方式就是利用本体互操作来解决,如本体映射方法等,但是相关的概念相似度计算方法往往存在严重的性能和缩放性问题^[7]。所以,我们有必要探索一种有效和新颖的本体表达方法,并且在相似度计算方法上具有更好的性能和算法优势。

随着研究的深入,与兴趣权值(Interest Score)相结合的方法也被证明有助于个性化本体的创建和进化。个性化加权本体逐渐受到研究者的关注^[8]。在很多电子商务的个性化推荐系统中也广泛采用对特定主题分配不同兴趣权值的方法,以此来创建类别驱动(Taxonomy-driven)的用户兴趣模型^[9]。还有学者计算用户访问过的 Web 网页与领域概念的相似程度,通过累加相似度权值来计算每个概念的相应权值,并利用这些具有非零权值的概念来表达用户兴趣模型^[10]。这些方法各有特色,但迄今为止,关于个性化加权本体的相似度研究还不多见^[11]。所以,本文准备在此方面做一些有益的探索。

2 加权 XML 数据模型

在各种基于个性化本体的用户兴趣模型研究中,最为常见的形式就是使用 XML 数据模型来表达本体信息。XML 的使用是实现语义 Web 网络的基本方法,它通过内在的层次结构和节点语义信息可以表达各种复杂的语义概念和它们之间的相互关系^[12]。标准的 XML 数据模型通常表现出一种带有层次结构特征的树状节点视图,利用这种层次状的数据模型可以很方便地表达一些概念较为复杂的数据内容。如在个性化本体中,使用 XML 数据模型来表达用户的兴趣特征,该 XML 数据模型的每个节点都是对应一个兴趣概念,不同的概念节点组合都可

以对应一个特定的个性化用户兴趣模型。

所谓加权 XML 数据模型,是指给每个 XML 节点赋予一个权值,以此来反映该节点对应的相关概念在整个 XML 数据实例中的重要性。为了说明一致,本文所使用的加权 XML 数据模型的结构如图 1 所示。

```

<Root weight="0">
  <Node1 weight="0">
    <Leaf11 weight="0"></Leaf11>
    <Leaf12 weight="0"></Leaf12>
  </Node1>
  <Node2 weight="0">
    <Leaf21 weight="0"></Leaf21>
    <Leaf22 weight="0"></Leaf22>
  </Node2>

```

图 1 加权 XML 数据模型的结构示例

可以注意到,图 1 中所有节点的权值都通过属性 weight 来表达,其默认值为 0。该加权 XML 数据模型可以作为个性化本体的语义知识参考数据来源,通过不同的权值属性设置来表达不同的个性化本体实例。本文将其称为领域参考本体(Domain Reference Ontology)^[13]。相应地,具体的个性化本体可以看成是领域参考本体的实例(Instance),其中的每个节点都增加了表征用户兴趣特征和访问行为为特点的兴趣权值,而且随着用户和系统交互的逐渐进行,这些兴趣权值将会被更新。

和标准的 XML 数据模型相比,这种加权 XML 数据模型除了具有层次结构和节点内容特征外,同时还增加了表达不同节点权值差异的能力,据此我们可以有效地表达一些更为复杂的概念。图 2 说明一个电子商务站点中采用加权 XML 数据模型的个性化本体,具体内容反映了用户对“播放录音”产品的购物兴趣特征。

虽然基于加权 XML 数据模型的个性化本体较为直观和简洁,但是相关的模式表达和相似度计算方法都需要进行系统的研究和设计。本文主要探讨基于加权 XML 数据模型的个性化本体在用户兴趣建模中的应用方法,包含两个主要内容,分别为模式表达和相似度比较。

```

<播放录音 ref="电池" weight="1">
  <mp3 weight="0"></mp3>
  <mp4 weight="2">
    <纽曼 weight="2"></纽曼>
    <爱国者 weight="2">
      <E5808 weight="4"></E5808>
      <F820 weight="0"></F820>
    </爱国者>
  </mp4>
</录音笔 weight="0"></录音笔>

```

图2 反映用户购物兴趣特征的基于加权 XML 数据模型的个性化本体实例

3 基于加权 XML 数据模型的个性化本体模式表达

基于加权 XML 数据模型的个性化本体具有较强的数据表达能力,可以从节点内容、层次结构和权值分配这三个方面来表达用户兴趣特征,其中权值分配是最为主要的內容。

3.1 节点内容

该个性化本体主要通过两类不同的节点来表达数据。一类节点为叶节点。在具体的应用系统中,该类节点通常都是用户所关注的最终信息内容,如选择的商品或者查询的关键词等,也是系统在信息推荐和信息展示时所需要提供给用户的信息内容。不过,这类节点本身虽然代表着详细的信息,但在语义上往往也具有一词多义或者多词一义的特点,因此,直接利用此类节点信息内容来实现个性化服务往往会产生误判。如图2中的“F820”节点,单纯以词语本身内容来看,它既有可能代表爱国者的MP4播放器,也有可能代表奥克斯的F820手机,甚至还有可能代表美利达的F820自行车。另一类节点为非叶节点,它们往往代表不同层次的语义概念,可以将不同的叶节点组织到一个完整的层次结构中。通过这种组织形式,我们可以在语义层次上有效地解决单纯使用叶节点信息内容带来的问题,并且可以更为准确地表达用户的兴趣特征模型。

3.2 层次结构

标准的 XML 数据模型的数据实例往往具有不同的结构特点,通过这些不同的结构特点,可以表达更为丰富的信息内容。这也形成了 XML 数据模型的一个显著的特点,基于 XML 数据模型的检索往往

据此区别于传统的关键词检索。不过,这在另一方面却给相应的 XML 检索带来了复杂的问题,如考虑结构层次的相似度往往在算法的复杂度和性能方面存在较为明显的缺点。为此,本文在设计基于加权 XML 数据模型的个性化本体时,采取了一个新颖的设计方案,那就是采用同一个加权 XML 数据模型结构的不同本体实例都具有相同的数据结构,不同本体实例之间的差别主要体现在权值分配的不同上。如图2中的“录音笔”节点权值为0,则说明该本体实例所反映的用户兴趣模型没有“录音笔”这个概念内容。本文将此称为“同构异值”的设计方法。所以,我们可以说,通过加权 XML 数据模型中的权值分配可以将标准 XML 数据模型中的结构问题转换为纯粹的 XML 节点权值比较问题。除此以外,基于加权 XML 数据模型的个性化本体还可以在很多其他方面充分体现结构信息的影响。如后文所提到的权值扩散方法和相似度比较方法,都大量参考和使用层次结构信息来加强计算效果。

3.3 权值分配

3.3.1 权值扩散方法

我们可以通过给不同的节点分配代表用户不同兴趣特征的权值来表示个性化的需求差异。由于用户直接访问的信息往往都是叶节点信息,因此我们可以从用户历史访问记录中获取相关信息来给权值赋值。然而,对于非叶节点而言,我们往往缺乏直接的用户访问记录。如用户没有通过和 XML 数据模型结构一致的目录来逐层定位并访问指定产品,而是通过直接点击某个叶节点对应的产品网页来选择产品,此时只能得到用户访问某个特定 XML 叶节点对应的信息。所以,我们需要对非叶节点设计一种有效的权值计算方法。

在一个基于加权 XML 数据模型的个性化本体中,存在着很多具有相互关系的节点,可以通过建立单向节点连接来表达这种扩展关系。用户在选择此处的某一产品时,权值应该在这些相关节点中自动扩散。如图2中的水平扩展主要基于节点的 ref 属性来进行,如“播放录音”节点权值可以扩散到“电池”节点中。

在权值扩散中,关键的问题在于计算扩散值。一般而言,权值扩散值都可以根据当前节点权值和扩散系数来得到,其中当前节点权值往往为反映用户兴趣的客观数值,而扩散系数的设定就较为灵活。常见的扩散系数设定方法是由系统设定一个固定的

默认值,如 0.5 等。可以看出,一个节点的最终权值应该是自己的初始权值加上所有扩散过来的权值之和。

然而,不同节点之间相关程度往往并不一致,使用一个统一的扩散系数显然不是很有有效的。我们可以通过使用节点本身的信息内容,来判断节点间连接关系的强度,从而给不同的节点连接分配一个不同的扩散系数。如有学者利用节点对应的文档向量给出一种计算方法^[13],该方法设定节点 i 和节点 s 之间的扩散系数为:

$$w_{is} = \frac{\vec{n}_i \cdot \vec{n}_s}{n_i \cdot n_s} \quad (1)$$

然而,并非所有的加权 XML 数据实例节点都附有文档向量,因此该方法的应用面不是很大。鉴于前文已经提出权值扩散存在两个方向,本文将对两个方向的扩散系数采取不同的设定方法。

3.3.2 垂直方向的扩散系数设定方法

在基于加权 XML 数据模型的个性化本体中,由于上下级节点的语义相关性是由 XML 数据的内在层次结构决定的,所以可以采用标准的双向扩散策略,即同时向上和向下进行权值扩散。很多学者都将该扩散系数设置为一个 0 到 1 区间的小数值。在这种系数设计方法下,由于用户通常选择的信息主要对应 XML 的叶节点内容,所以上层节点所获取的权值一般总会小于叶节点的权值。如图 2 所提到的本体实例就是一个扩散系数为 50% 的例子,我们会发现节点层次越高,权值往往越低。如果从节点反映用户兴趣特征这个角度来讲,这是存在问题的,因为既然用户选择了某个叶节点,那么他也就对相应的上层节点具有同样的关注度,至少不能呈现向上递减的状态。

为此,我们提出一个新颖的设计方案。在向上

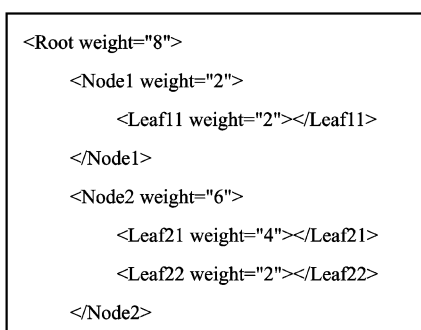
扩散过程中,扩散系数设定为 1,即所有节点都将自己的权值完整地向上扩散。为了保证这个过程的正确性,扩散过程必须从叶节点开始逐层向上进行。对于向下扩散过程而言,前文已经指出,用户往往最终选择的内容都是叶节点对应的内容,而作为语义概念层次的非叶节点往往并非用户直接关注的内容。直接根据一个上级节点而不加选择地对其所有下级节点进行扩散,可能会将很多用户其实并不感兴趣的信息包含进用户的兴趣模型中。同时这个过程所涉及的扩散节点数量非常多,相应的计算性能会产生较大的开销。极端的例子是如果用户选择了根节点,则会对所有节点进行权值扩散。因此,我们没有采用向下的权值扩散策略。

图 3(a)显示了使用该权值扩散方法的基于加权 XML 数据模型的个性化本体实例,图 3(b)显示了相应的树状节点视图。

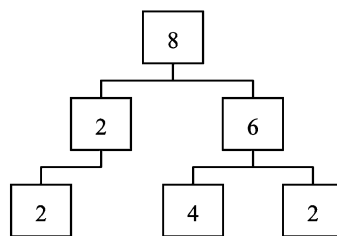
需要说明的是,由于所有节点的默认权值为 0,所以在上述实例中我们省略了权值为 0 的节点表示(下文示例依次类推)。我们可以发现,该扩散方法具有以下两个特点:① 在该数据结构中,每一层次的所有节点权值之和都为固定值,即根节点的权值;② 通常节点层次越高,越能获得更高的权值,而根节点则能够获得最高的权值。

3.3.3 水平方向的扩散系数设定方法

由于水平方向的节点连接无法直接通过连接本身来说明节点间的相关关系,一个直接的方法就是利用边数(Edge-counting)来测度,其主要思想在于寻找树状层次结构中能够连接两个节点的最小边数量^[14]。但是这个方法存在问题,特别是应用于诸如 WordNet 之类通用性较强的本体时,往往不是很准确^[15]。也有学者认为,在一个本体层次中,任意两个概念节点的相似度可以通过最近共同父节点的深



(a) 本体实例



(b) 本体实例的树状节点视图

图 3 使用权值扩散方法的基于加权 XML 数据模型的个性化本体实例

度来间接测度。如果最近共同父节点的子节点数量越多,两个节点的相似度就越可能较小,反之则有可能越大^[11]。事实上,在一个特定的应用系统中,我们认为这个假设很难一定成立。如有的产品类别下属产品很多,而有的类别下属产品较少,但是不能据此认为前者具有较高的相关性,而后者具有较低的相关性。

本文通过基于信息量(Information Content)的计算方法,得到一种因节点而异的水平方向差异化扩散系数设计方案。

在一个基于加权 XML 数据模型的个性化本体中,节点 i 的出现概率可以表示为:

$$P(i) = \text{Frequency}(i) / N \quad (2)$$

其中, $\text{Frequency}(i)$ 表示节点 i 出现的次数和节点 i 所有下级节点的出现次数总和, N 表示所有的节点总数。因此,节点 i 的信息量可以表示为:

$$IC(i) = -\log(P(i)) \quad (3)$$

据此可以得到两个节点之间的水平方向扩散系数为:

$$\text{Coefficient}_{horizontal}(n, m) = IC(\text{Super}(n, m)) \quad (4)$$

其中, $\text{Super}(n, m)$ 表示节点 n 和节点 m 的最近共同父节点。

4 基于加权 XML 数据模型的个性化本体相似度计算方法

标准 XML 数据模型的相似度主要根据节点内

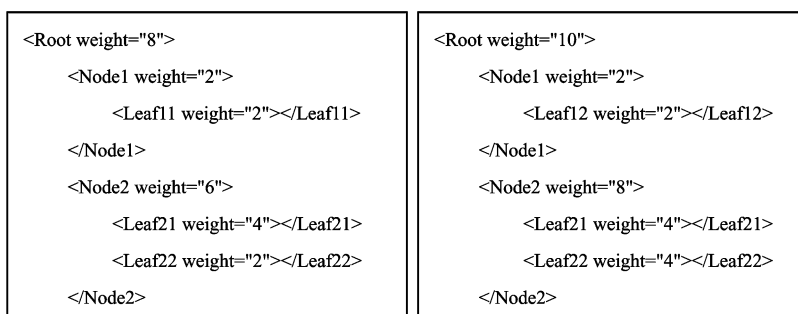
容和层次结构来进行比较,而加权 XML 数据模型通过引入节点权值,比传统方法增加了更多的量化处理能力。同时由于本文采用了“同构异值”的设计方法,所以基于加权 XML 数据模型的个性化本体所涉及的相关度计算方法具有很多新的特点。

为了便于说明,假设两个基于加权 XML 数据模型的个性化本体实例如图 4 所示。

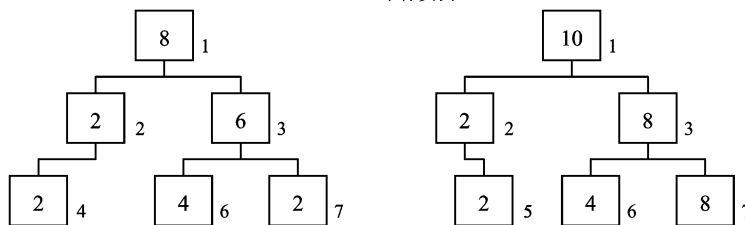
在具体解释相似度算法之前,首先说明几个基本设计原则。

1)虽然加权 XML 数据模型中具有结构层次、节点内容和节点权值三个测度依据,但是相关的相似度比较应该只侧重于节点权值的比较。这个原则的基本思路来自于“同构异值”的设计方法。不同的加权 XML 数据实例都具有相同的结构和节点内容,差异主要表现在节点权值的不同上,因此在计算相似度时,可以忽略结构层次和节点内容的影响。

退一步说,即便删除所有权值为 0 的节点,产生整体结构不一致的结果,我们也认为结构相似度可以通过权值相似度来反映。对于该加权 XML 数据模型而言,整个数据结构呈现出层次越高、差异度越小的形式,不同数据实例底层节点结构的差异往往可以通过上层节点的相似性来得以消除,所以进行相似度比较的不同数据实例之间主要的结构差异往往表现在底层的节点结构上。如图 4 两个实例中的 Node1 节点,其子节点结构的差异通过子节点的权值差异表现出来,而两个实例的 Node1 节点又都为 2,都表现出对 Node1 节点的认可,所以两个实例



(a) 本体实例



(b) 本体实例的树状节点视图 (节点方框边外的数字表示节点序号)

图 4 需要比较相似度的两个基于加权 XML 数据模型的个性化本体实例

Node1 节点的相似性又通过自身权值相等而体现出来。

2)在较低层次节点上相似的两个加权 XML 数据实例,要比在较高层次节点上相似的两个加权 XML 数据实例更为相似。层次越低的节点越能准确地反映用户的具体兴趣特征,相反层次较高的节点通常更多地用于反映用户的广义兴趣特点。所以,在相似度计算过程中,应当体现这种节点层次深度对相似度计算的影响能力。正如前文所述,在基于加权 XML 数据模型的个性化本体中,节点层次越高越能获得更高的权值,而根节点能够获得最高的权值,但是层次越高的节点对相似度计算的贡献能力却应该越低。

按照上述设计原则,我们提出了基于加权 XML 数据模型的个性化本体相似度计算方法。假设领域参考本体为 X,本体实例为 x1 和 x2,两者的相似度算法可以表示为:

```

输入 : X ,x1 ,x2
输出 : x1 和 x2 的相似度
//规范化所有节点权值
normalization (x1);
normalization (x2);

//层次衰减因子,层次越高,衰减度越大
decayFactor=1;

//X 的最大层次深度
similarityValueInLevel[n];

//倒序遍历每一层次
for each level li of X in reversed order {
//计算层次 i 相似度的中间变量
totalValueInLevel=0;

//遍历当前层次的每一个节点
for each node nj in li {
//平滑因子,为层次 i 的当前层次节点数量
总和
totalNodeNumber=gefTotalNodeNumberInLevel
(i)

//得到层次 i 中节点 j 的相似度
totalValueInLevel += |njx1 - njx2| /max (njx1,
njx2 )/totalNodeNumber

```

```

}

//得到层次 i 的相似度
similarityValueInLevel[i]= decayFactor * (1 -
totalValueInLevel);

//遍历层次越高,衰减因子衰减度逐渐增大
decrease(decayFactor);
}

//返回最终的本体实例相似度
return max(similarityValueInLevel);

```

5 系统实现与实验

为了对上述方法的有效性进行验证,笔者利用基于 JSP 的 Web 购物系统平台进行了数据测试。该应用系统的硬件平台环境为:CPU Intel Core 2 Duo P8400,内存 PC3-8500 DDR3 2.0G。软件平台为:操作系统 Windows Server 2003,SQL Server 2005, JDK1.6,Eclipse 3.3。

实验所需要的领域参考本体数据结构来自于 ODP^[16],通过解析获得一个完整的领域参考本体,所对应的 XML 数据结构节点总数为 756969,最大层次深度为 15。所抽取的“Shopping”相关实验节点总数为 5378,最大层次深度为 11。在实验系统中,所有的 XML 节点数据都采用关系表的存储方式。下面对此分别予以介绍。

1)structure-shopping 表用于存储所有节点信息,即领域参考本体信息。该表存储的顶层节点为“Top/Shopping”,二级节点有 35 个,三级节点有 541 个。表 1 展示了该表部分记录的数据内容。

2)relation-shopping 表用于存储节点间的关联信息,以便于水平方向的权值扩散。总记录数为 3067,其中链入值最大的节点只有 1 个,节点为“Top/Shopping/Sports/Apparel”,其链入值为 21;而链出值最大的节点有 4 个,链出值都为 13。表 2 展示了这些节点的情况。

表 3 展示了 relation-shopping 表部分记录的数据内容。

3)userprofile 表用于存储用户的个性化本体信息。其中,通过访问次数表达原始节点权值,默认节点权值为 0。为了节省存储空间,该表只存储权值非零的被访问节点信息,同时不存储非叶节点的相

表1 structure-shopping表的部分记录

节点号	节点名称	层次号
451605	Top/Shopping	2
451842	Top/Shopping/Children	3
451843	Top/Shopping/Children/Baby	4
451844	Top/Shopping/Children/Baby/Albums-and-Frames	5
451845	Top/Shopping/Children/Baby/Bath-and-Body	5
451846	Top/Shopping/Children/Baby/Bibs-and-Towels	5
451847	Top/Shopping/Children/Baby/Birth-Announcements	5
451848	Top/Shopping/Children/Baby/Birth-Announcements/Online	6
451849	Top/Shopping/Children/Baby/Birth-Announcements/Printed	6
451850	Top/Shopping/Children/Baby/Birth-Announcements/Printed/Photo-Cards	7
451851	Top/Shopping/Children/Baby/Diapering	5
451852	Top/Shopping/Children/Baby/Diapering/Diaper-Bags	6
451853	Top/Shopping/Children/Baby/Diapering/Disposable	6

表2 链出值最多的4个节点

节点号	节点名称
455087	Top/Shopping/Publications/Books/Business/Financial
455096	Top/Shopping/Publications/Books/Business/Opportunities
455098	Top/Shopping/Publications/Books/Business/Real-Estate
455101	Top/Shopping/Publications/Books/Business/Start-Up

表3 relation-shopping表的部分记录

序号	链出节点号	链入节点号
435696	451619	451670
435714	451636	452304
435724	451637	454193
435725	451637	454201
435735	451642	451752

关权值信息。表4展示了该表部分记录的数据内容。

表4 userprofile表的部分记录

序号	用户号	节点号	权值
4	1001	451841	7
5	1001	452302	4
6	1001	455835	13
7	1005	452303	1
8	1005	453250	2

实验提供了一个Web访问界面,不同用户可以按照自己的兴趣爱好来浏览和查询产品信息,访问的产品信息会被存储到原始用户兴趣模式表(userprofile)中。在系统后台,定期通过权值扩散方法,得到完整的用户个性化信息表(userpersonalization)。该表结构和原始用户兴趣模式表结构一致,但存储了包含非叶节点在内的所有相关节点权值信息。为了节省存储空间和提高计算速度,该表也不存储权值为0的节点信息。权值扩散算法和相似度算法都采用SQL Server存储过程来实现。

实验首先对相似度算法进行了理论数据测试。按照图4给出的两个本体实例,最终的相似度为0.43438。表5和表6给出了部分详细实验结果数据。

表5 图4本体实例的不同节点相似度

层次号	节点号	节点相似度差值	最大节点值
2	1	0	1
3	2	0.05	0.25
3	3	0.05	0.8
4	6	0.1	0.5
4	7	0.15	0.4
4	4	0.25	0.25
4	5	0.2	0.2

表6 图4本体实例的不同层次相似度

层次号	层次相似度
2	0
3	0.43438
4	0.35625

同时,实验还得出如下结论:①该相似度算法满足交换率,即 x_1 和 x_2 的相似度等于 x_2 和 x_1 的相似度。②完全不相同的两个本体实例相似度为 0,如图 5 所示的两个本体实例相似度计算结果为 0。该结果充分说明根节点权值对最终结果的相似度不起任何贡献作用,这也符合人们的一般理解。③完全相同的两个本体实例相似度为 1,如图 6 所示的两个本体实例相似度计算结果为 1。

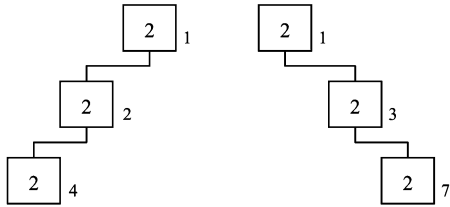


图 5 完全不同的两个本体实例

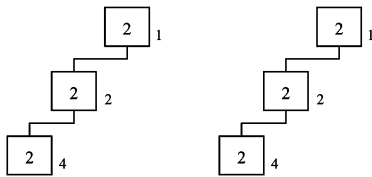


图 6 完全相同的两个本体实例

最后,应用系统采用了人工评价方法来测试相似度算法的实际有效性。系统面向 15 位测试者进行了使用测试,不同测试者独立使用,最终形成了 15 个用户兴趣模式本体实例。系统通过两两相似度比较,给每位测试者推荐 5 个相似度最高的其他用户兴趣模式本体实例,并要求测试者对 5 个结果分别给出评价,评价层次为 6 个级别,即 0 到 5。表 7 给出了最终的平均评价结果数据。

表 7 测试者的平均评价结果

序号	平均评价级别
1	3.736
2	4.005
3	3.989
4	4.041
5	3.664

6 总结与展望

本文提出了基于加权 XML 数据模型的个性化本体设计方法,并对相关模式表达和相似度计算方

法给出了详细的理论说明和实验结果分析,初步试验结果证明该算法满足要求。不过我们也发现,该方法存在很多需要进一步研究和改进的地方。主要有以下几点:①现有的权值扩散方法和相似度计算方法仍然存在可以改进的地方,我们有必要设计和探索更好的方法,并对这些方法进行比较实验分析,从而找出更优的设计方案。②文中所谈到的基于加权 XML 数据模型的个性化本体表达方法,可以看成是加权 XML 数据模型的一个应用特例,也就是采用了“同构异值”赋值方法的加权 XML 数据模型。但是如果把问题扩展开来,对于任何两个“异构异值”的加权 XML 数据实例,相关权值扩散方法,尤其是相似度计算方法都值得进一步研究,这也是我们的下一个研究目标。

参 考 文 献

- [1] Sendhilkumar S, Geetha T V. Personalized ontology for web search personalization [C] // Proceedings of the First Bangalore Annual Compute Conference (Compute'08). Bangalore, India, January 2008: 1-7.
- [2] Sieg A, Mobasher B, Burke R. Web search personalization with ontological user profiles [C] // Proceedings of the 16th ACM Conference on Information and Knowledge Management (CIKM'07). Lisbon, Portugal, November 2007: 525-534.
- [3] Li Y, Zhong N. Mining ontology for automatically acquiring Web user information needs [J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(4): 554-568.
- [4] Zhou X, Li Y, et al. Relevance assessment of topic ontology [C] // Presented at the Fourth International Conference on Active Media Technology. Brisbane, Australia, 2006.
- [5] Zhou X, Wu S-T, et al. Utilizing search intent in topic ontology-based user profile for Web mining [C] // Proceedings of 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference), 2006: 558-564.
- [6] Wong A K Y, Ray P, Waran N P. Ontology mapping for the interoperability problem in network management [J]. IEEE Journal on Selected Areas in Communication, 2005, 23(10): 2058-2068.
- [7] Mika P. Ontologies Are Us: A Unified Model of Social Networks and Semantics [C] // Yolanda Gil, Enrico Motta, et al. Proceedings of the 4th International Semantic Web Conference (ISWC 2005), volume 3729 of LNCS. Berlin Heidelberg: Springer-Verlag, 2005: 522-536.
- [8] Haase P, Sure Y, et al. Usage-driven evolution of personalized ontologies [C] // Proceedings of the 3rd

- International Conference on Universal Access in Human-Computer Interaction, HCI International 2005. Las Vegas, NV, July 2005.
- [9] Ziegler C, Lausen G, Schmidt-Thieme L. Taxonomy-driven computation of product recommendations[C]// Proceedings of ACM International Conference on Information and Knowledge Management, CIKM 2004. Washington, DC, November 2004.
- [10] Trajkova J, Gauch S. Improving ontology-based user profiles [C]// Proceedings of the Recherche d'Information Assistée par Ordinateur, RIAO 2004. University of Avignon (Vaucluse), France, April 2004: 380-389.
- [11] Zhang K, Tang J, Hong M-C, et al. Weighted ontology-based search exploiting semantic similarity [C] // Proceedings of APWeb '2006: 498-510.
- [12] Pollock J T, Hodgson R. Adaptive Information: Improving Business through Semantic Interoperability, Grid Computing, and Enterprise Integration[M](Wiley Series in Systems Engineering and Management). Wiley-Interscience, 2004.
- [13] Sieg A, Mobasher B, Burke R. Representing context in Web search with ontological user profiles[C]// Proceedings of the Sixth International and Interdisciplinary Conference on Modeling and Using Context (Context'07), Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag, 2007, 4635: 439-452.
- [14] Rada R, Mili H, Bicknell E, et al. Development and application of a metric on semantic nets [J]. IEEE Transactions on System, Man, and Cybernetics, 1989(1): 17-30.
- [15] Miller G A. WordNet: A lexical database for English[J]. Communications of the ACM, 1995, 38(11): 39-41.
- [16] DMOZ. Open Directory RDF Dump [OL]. [2009-07]. <http://rdf.dmoz.org/>.

(责任编辑 许增棋)

作者: [李树青, Li Shuqing](#)
作者单位: [南京财经大学信息工程学院信息管理系, 南京, 210046](#)
刊名: [情报学报](#) [ISTIC](#) [PKU](#) [CSSCI](#)
英文刊名: [JOURNAL OF THE CHINA SOCIETY FOR SCIENTIFIC AND TECHNICAL INFORMATION](#)
年, 卷(期): 2010, 29(5)
被引用次数: 0次

参考文献(16条)

1. [Sendhilkumar S, Geetha T V Personalized ontology for web search personalization](#) 2008
2. [Sieg A, Mobasher B, Burke R Web search personalization with ontological user profiles](#) 2007
3. [Li Y, Zhong N Mining ontology for automatically acquiring Web user information needs](#) 2006(4)
4. [Zhou X, Li Y Relevance assessment of topic ontology](#) 2006
5. [Zhou X, Wu S-T Utilizing search intent in topic ontology-based user profile for Web mining](#) 2006
6. [Wong A K Y, Ray P, Waran N P Ontology mapping for the interoperability problem in network management](#) 2005(10)
7. [Mika P Ontologies Are Us: A Unified Model of Social Networks and Semantics](#) 2005
8. [Haase P, Sure Y Usage-driven evolution of personalized ontologies](#) 2005
9. [Ziegler C, Lausen G, Schmidt-Thieme L Taxonomy-driven computation of product recommendations](#) 2004
10. [Trajkova J, Gauch S Improving ontology-based user profiles](#) 2004
11. [Zhang K, Tang J, Hong M-C Weighted ontology-based search exploiting semantic similarity](#) 2006
12. [Pollock J T, Hodgson R Adaptive Information: Improving Business through Semantic Interoperability, Grid Computing, and Enterprise Integration](#) 2004
13. [Sieg A, Mobasher B, Burke R Representing context in Web search with ontological user profiles](#) 2007
14. [Rada R, Mili H, Bicknell E Development and application of a metric on semantic nets](#) 1989(1)
15. [Miller G A WordNet: A lexical database for English](#) 1995(11)
16. [DMOZ Open Directory RDF Dump](#) 2009

本文链接: http://d.g.wanfangdata.com.cn/Periodical_qbxb201005009.aspx
授权使用: 朱馆长(wfnjcj), 授权号: dc5e7b90-62f9-49d1-b79a-9e760152e981

下载时间: 2011年1月25日