

基于互表性的个性化网页研究

罗钧旻, 罗建龙

(西安工业大学 计算机科学与工程学院, 陕西 西安 710032)

摘 要: 为了适应各种用户需求, 为用户提供全面的信息, 网页往往都包容万象具有丰富庞大的信息量, 然而, 每个用户只关心自己感兴趣的信息. 本文把网页信息和用户偏好看成模糊本体, 通过求这两个模糊本体的交集得到用户感兴趣的网页信息, 为用户提供高效、便捷的个性化服务.

关键词: 网页; 用户偏好; 模糊本体; 本体求交; 个性化服务

中图分类号: TP18

文献标识码: A

文章编号: 1000-7180(2013)03-0012-04

Research on Personalized Web Based on Mutual-Description Principle

LUO Jun-min, LUO Jian-long

(School of Computer Science and Engineering, Xi'an Technological University, Xi'an 710032, China)

Abstract: To meet those various needs of the clients, websites are usually designed to contain tons of information. However, the fact is that the individual client only cares about what they are interested in. In this paper, the website information and the client's preference are considered to be fuzzy ontology and by figuring out the intersection of the both ontological modules, the author gets the information that the client is interested in. This paper introduces ontological thought to provide more efficient personalized services to the clients.

Key words: website; client preference; fuzzy ontology; intersection of ontology; personalized service

1 引言

在万维网(WWW)高速发展的今天, 其发展的方向是智能化、个性化, 为了适应各种用户需求, 提供全面的信息, 网页往往都包容万象. 然而, 每个用户只关心自己感兴趣的信息. 人们需要一种更佳准确、迅速获取到有用的信息并简单易于维护这些信息的方法, 于是语义 web 应运而生.

语义 web 并非一个独立的 web, 语义 web 是基于当前 web 的一个扩展, 其核心概念是本体. 本体是对现实世界中某个领域的抽象或概括, 是由概念及其关系构成; 网页是构成网站的基本元素, 是承载各种网站应用的平台, 我们可以把网页看成是一个本体; 用户偏好是消费心理效果中一个重要概念, 即用户对同类产品和服务中对某些品牌的嗜好, 也可把用户偏好看成是一个本体. 因此为用户提供个性

化网页可以通过求网页信息和用户偏好这两个本体的交集来完成.

本文在研究基于互表性的模糊本体^[1]概念的基础上, 提出了一种通过求网页信息和用户偏好这两个模糊本体的交集为用户提供个性化网页的思想, 给出了模糊本体求交的算法, 并用举例进行了验证.

2 基于互表性的模糊本体

2.1 互表性

不管是在哲学界还是计算机界都从不同的角度提到一个事物失去与之相关的事物或者是环境都不能存在, 这就是本文提出的互表性(互相表示性).

2.2 模糊本体

语义 web 的核心是本体, 但目前本体模型之间描述数据信息都是某种确定信息, 描述的信息之间的关系都是确定的^[2-3]. 但是, 在现实世界中概念之

间的关系是有强弱之分的. 我们将概念间关系紧密程度称为概念的模糊性. 把引入互表性和互表性的本体称为模糊本体.

模糊本体可以定义为一个三元组 $O=(C, R, E)$, C 为概念的集合, R 为关系的集合, E 为演化机制(规则). R 中的成员为一个四元组 $(r, c1, c2, f)$, 其中 r 为 $c1$ 与 $c2$ 之间的关系名, $c1, c2$ 为直接相关的两个概念, f 为关系程度, $f \in [0, 1]$, $r, c1, c2 \in C$. 在四元组的模糊本体模型之中把层次、实例、公理都纳入 R 中, 是一些特殊的关系.

例如, 在概念上的层次有 Kind-of 或 Is-a, 表示 Superclass-Subclass 关系, 我们用 (Kind-of, $c1, c2, 0.7$) 表示 $c1$ 有 70% 的可能是 $c2$ 的超类. 而对于具体的实例我们用 (Is-a, $c1, c2, 1$) 表示 $c1, c2$ 是 $c2$ 的一个实例. 对于公理集合中的每条公理代表领域知识中的永真断言, 例如, 声明关系 “Teach” 和 “Taught-by” 是互逆的, 用关系可以表示为 (Inverse, Teach, Taught-by, 1). 其中关系的程度初始值由专家的经验给出, 并不断演化.

3 网页个性化研究

3.1 个性化服务

如今, 服务商为了给用户提供更智能、更个性化的信息, 增加用户比重, 在服务端提供各种个性化服务, 网络商将收集用户信息进行整理分析后与服务器信息进行筛选和比对, 为用户整理出偏好信息进行展示, 这就是所谓的个性化服务.

目前存在着许多个性化服务系统, 它们提出了各种思路以实现个性化服务. 个性化服务系统根据其所采用的推荐技术可以分为三种: 基于规则的系统、基于内容过滤的系统和协作过滤系统.

(1) 基于规则的系统允许系统管理员根据用户的静态特征和动态属性来制定规则, 一个规则本质上是一个 If-Then 语句, 规则决定了在不同的情况下如何提供不同的服务. 基于规则的系统其优点是简单、直接, 缺点是规则质量很难保证, 而且不能动态更新, 此外, 随着规则的数量增多, 系统将变得越来越难以管理. 基于内容过滤的系统利用资源与用户兴趣的相似性来过滤信息.

(2) 基于内容过滤的系统其优点是简单、有效, 缺点是难以区分资源内容的品质和风格, 而且不能为用户发现新的感兴趣的资源, 只能发现和用户已有兴趣相似的资源.

(3) 协作过滤系统利用用户之间的相似性来过

滤信息. 基于协作过滤系统的优点是能为用户发现新的感兴趣的信息, 缺点是存在两个很难解决的问题, 一个是稀疏性, 亦即在系统使用初期, 由于系统资源还未获得足够多的评价, 系统很难利用这些评价来发现相似的用户. 另一个是可扩展性, 亦即随着系统用户和资源的增多, 系统的性能会越来越低.

针对现有个性化服务系统中存在的问题, 本文引入本体思想来提供各种个性化服务, 提出了基于互表性的模糊本体求交的方法. 采用模糊本体构建用户和网页模型, 以本体形式表示的用户偏好和网页信息进行求交运算, 获取个性化网页信息.

3.2 个性化网页的结构

基于互表性的个性化网页结构如图 1 所示. 图中, User Ontology 是保存用户偏好信息的本体, 我们称之为用户本体 O_u ; Web Ontology 是服务商所提供的门户网站, 我们称为网络本体 O_w . Interest Ontology 是根据 O_u 而从 O_w 得到的用户对门户网站的偏好信息, 我们称之为偏好本体 O_i . O_i 可以通过对 O_w 和 O_u 两个本体求交得到. 最后把 O_i 转换为网页呈献给用户, 得到了用户关心的个性化网页^[4].

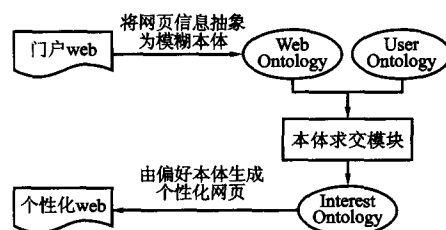


图1 个性化网页结构图

3.3 模糊本体求交算法

2个模糊本体的求交运算, 是个性化网页的核心. 由于模糊本体的求交运算不涉及本体的演化, 可以把模糊本体的定义 $O=(C, R, E)$ 简化为 $O=(C, R)$. 由于互表性, C 一定会出现在 R 中, 这样模糊本体 $O=(C, R)$ 就可以看成是 R . 两个本体的求交, 就变成了两个模糊关系集合的求交.

设 O_w, O_u 为 2 个模糊本体;

O_i 为它们的交集本体;

其对应的关系集合分别为 R_w, R_u 和 R_i ;

$(r, c1, c2, f)_i = (r_i, c1_i, c2_i, f_i) \in R_w, i=1, \dots, |R_w|$;

$(r, c1, c2, f)_j = (r_j, c1_j, c2_j, f_j) \in R_u, j=1, \dots, |R_u|$;

则 $R_i = R_w \cap R_u$ 的算法可以描述如下:

```

Ri={};
for(i=1; i≤|Rw|; i++)
for(j=1; j≤|Ru|; j++)
if(ri == rj && cli == clj && c2i ==
c2j)
Ri=Ri∪{(ri,cli,c2i,(fi+fj)/2)};

```

4 举例

图 2 所示 user ontology(用户本体)为存储用户偏好信息的模糊本体。

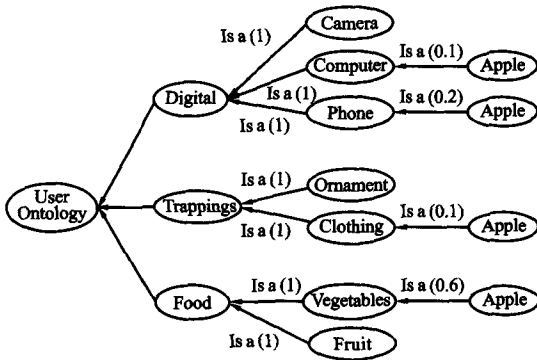


图 2 用户本体

根据图 2 的用户本体网络图得到:

$O_u = (C_u, R_u);$

$C_u = \{ \text{Digital, Camera, Computer, Phone, Trappings, Ornament, Clothing, Food, Vegetables, Fruit, Apple} \};$

$R_u = \{ R_u[1], R_u[2], R_u[3], R_u[4], R_u[5], R_u[6], R_u[7], R_u[8], R_u[9], R_u[10], R_u[11] \};$

$R_u[1] = (is-a[1], \text{Digital}[1], \text{Camera}[1], 1[1]);$

$R_u[2] = (is-a[2], \text{Digital}[2], \text{Computer}[2], 1[2]);$

$R_u[3] = (is-a[3], \text{Computer}[3], \text{Apple}[3], 0.1[3]);$

$R_u[4] = (is-a[4], \text{Digital}[4], \text{Phone}[4], 1[4]);$

$R_u[5] = (is-a[5], \text{Phone}[5], \text{Apple}[5], 0.2[5]);$

$R_u[6] = (is-a[6], \text{Trappings}[6], \text{Ornament}[6], 1[6]);$

$R_u[7] = (is-a[7], \text{Trappings}[7], \text{Clothing}[7], 1[7]);$

$R_u[8] = (is-a[8], \text{Clothing}[8], \text{Apple}[8], 0.1[8]);$

$R_u[9] = (is-a[9], \text{Food}[9], \text{Vegetables}[9], 1[9]);$

$R_u[10] = (is-a[10], \text{Vegetables}[10], \text{Apple}[10], 0.6[10]);$

$R_u[11] = (is-a[11], \text{Food}[11], \text{Fruit}[11], 1[11]).$

图 3 所示 web ontology(网络本体)是服务商多提供的门户网站信息所生成的模糊本体. 根据图 3 的网络本体网络图得到:

$O_w = (C_w, R_w);$

$C_w = \{ \text{Digital, Computer, apple, phone, Network Equipment} \};$

$R_w = \{ R_w[1], R_w[2], R_w[3], R_w[4], R_w[5], R_w[6] \};$

$R_w[1] = (is-a[1], \text{Digital}[1], \text{Camera}[1], 1[1]);$

$R_w[2] = (is-a[2], \text{Digital}[2], \text{Computer}[2], 1[2]);$

$R_w[3] = (is-a[3], \text{Computer}[3], \text{Apple}[3], 0.5[3]);$

$R_w[4] = (is-a[4], \text{Digital}[4], \text{phone}[4], 1[4]);$

$R_w[5] = (is-a[5], \text{phone}[5], \text{Apple}[5], 0.5[5]);$

$R_w[6] = (is-a[6], \text{Digital}[6], \text{Network Equipment}[6], 1[6]).$

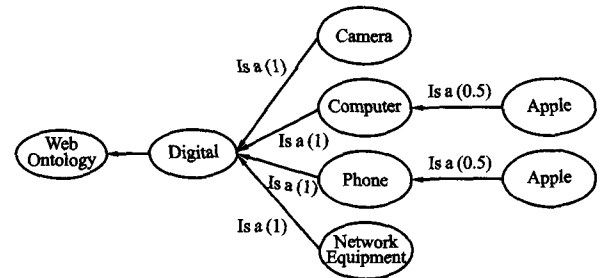


图 3 网络本体

进行用户偏好检索, 对 O_u 与 O_w 套用求交框架整理出:

$R_i = R_i \setminus \{ \} \cup \{ R_i[1], R_i[2], R_i[3], R_i[4], R_i[5] \};$

$R_i[1] = (ro = is-a[1], R_{mi} = \text{Digital}[1], R_{ni} = \text{Computer}[1], fo = F = 1/2 + 1/2 = 1[1]);$

$R_i[2] = (ro = is-a[2], R_{mi} = \text{Computer}[2], R_{ni} = \text{Apple}[2], fo = 0.1/2 + 0.5/2 = 0.3[2]);$

$R_i[3] = (ro = is-a[3], R_{mi} = \text{Digital}[3],$

$R_{ni} = \text{Phone}[3], fo = F = 1/2 + 1/2 = 1[3];$
 $Ri[4] = (ro = is - a[4], R_{mi} = \text{Phone}[4],$
 $R_{ni} = \text{Apple}[4], fo = 0.2/2 + 0.5/2 = 0.35[4];$
 $Ri[5] = (ro = is - a[5], R_{mi} = \text{Digital}[5],$
 $R_{ni} = \text{Camera}[5], fo = F = 1/2 + 1/2 = 1[5]).$

图4所示 Internet ontology(偏好本体)是 user ontology 与 web ontology 通过模糊本体交集求解生成的模糊本体. 模糊本体生成偏好本体 O_i , 偏好本体网络示意图为:

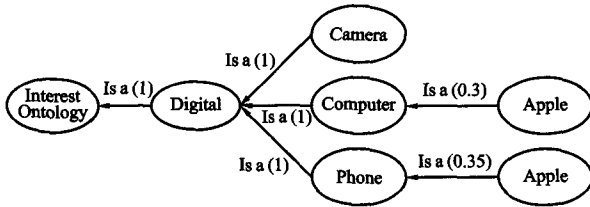


图4 偏好本体

$O_i = (C_i, R_i);$
 $C_i = \{ \text{Digital}, \text{Computer}, \text{Apple}, \text{Phone} \};$
 $R_i = \{ Ri[1], Ri[2], Ri[3], Ri[4], Ri[5] \};$
 $Ri[1] = (is - a[1], \text{Digital}[1], \text{Computer}[1], 1$
 $[1]);$
 $Ri[2] = (is - a[2], \text{Computer}[2], \text{Apple}[2],$
 $0.3[2]);$
 $Ri[3] = (is - a[3], \text{Digital}[3], \text{Phone}[3], 1$
 $[3]);$
 $Ri[4] = (is - a[4], \text{Phone}[4], \text{Apple}[4], 0.35$

$[4]);$

$Ri[5] = (is - a[5], \text{Digital}[5], \text{Camera}[5], 1$
 $[5]).$

5 结束语

本文通过引入本体思想来提供网络个性化服务, 提出了基于互表性的模糊本体求交的方法和简单的实例验证, 是对当前 web 的一个扩展和对未来语义 web 的一种展望, 为语义 web 的研究发展提供了一条新思路.

参考文献:

- [1] 罗钧旻, 武钰云, 吴彬. 模糊本体及其演化研究[J]. 微电子学与计算机, 2011, 28(5): 140-143.
- [2] 武钰云. 模糊语义网的体系结构及其推理研究[D]. 西安: 西安工业大学, 2011.
- [3] 吴斌. 模糊语义网及其演化研究[D]. 西安: 西安工业大学, 2009.
- [4] 吴彬, 罗钧旻. 基于模糊认知图的资源描述框架研究[J]. 微电子学与计算机, 2009, 26(3): 42-44.

作者简介:

罗钧旻 (1963-), 博士, 教授. 研究方向为智能软件和领域软件.

罗建龙 男, (1988-), 硕士研究生. 研究方向为计算机应用技术.