

模糊本体及其演化研究

罗钧旻, 武钰云, 吴 彬

(西安工业大学, 陕西 西安 710032)

摘 要: 本体是共享概念模型的规范说明, 但以往的描述方法不能表示资源间关系的模糊性及其演化. 针对这一问题, 引入了模糊本体的概念并给出了形式化的描述, 用模糊资源描述框架(FRDF)对模糊本体进行表示, 提出了一种模糊本体演化算法, 并用实验进行了验证.

关键词: 语义网; 模糊本体; 模糊资源描述框架; 模糊语义网

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2011)05-0140-04

Research on the Fuzzy Ontology and it's Evolution

LUO Jun-min, WU Yu-yun, WU Bin

(Xi'an Technological University, Xi'an 710023, China)

Abstract: The ontology is the formal criteria of the shared conceptual model, but the former method can not describe their vague relationship between resources and its evolution. To this issue, we introduced the concept of fuzzy ontology and give a formal description, fuzzy ontology is described by fuzzy resource description framework (FRDF), and an evolution algorithm and experiment are given.

Key words: semantic web; fuzzy ontology; fuzzy resource description framework; fuzzy semantic web

1 引言

目前, 语义网理论把资源间的关系看成是确定不变的. 然而, 资源的语义在不同的领域、不同的文化、不同的社会环境中有所不同; 同时, 资源的语义还会随着外界因素的变化而发生演变^[1]. 所以就需要把语义网的理论进行扩充, 引入描述资源的模糊性、资源间关系的模糊性及其发展变化的成分. 因此, 文中在研究语义网现有理论的基础上, 提出了模糊本体的概念并给出其形式化表示, 同时对模糊本体的演化作了研究, 并以实验进行了验证.

2 模糊本体及其形式化

模糊语义网的核心是模糊本体, 模糊本体是将模糊概念引入到传统的本体模型中, 以解决某一领域的不确定问题^[2]. 为了描述概念及概念间关系的不确定性, 文中引入模糊关系, 得到模糊本体模型. 下面给出本体的形式化定义.

模糊本体可以定义为一个四元组 $O = (C, R, E, I)$, C 为概念的集合, R 为关系的集合, E 为演化机制(规则), I 为推理机制(应用或解释规则). R 中的成员为一个四元组 (r, c_1, c_2, f) , 其中 r 为 c_1 与 c_2 之间的关系名, c_1, c_2 为直接相关的两个概念, f 为关系程度, $f \in [0, 1]$, $r, c_1, c_2 \in C$. 层次、实例、公理都可纳入 R 中, 是特殊的关系, 例如, 在概念上的层次有 Kind-of 或 Is-a, 表示 Super class—Subclass 关系, 用 $(\text{Kind-of}, c_1, c_2, 0.7)$ 表示 c_1 有 70% 的可能是 c_2 的超类. 而对于具体的实例用 $(\text{Is-a}, c_1, c_2, 1)$ 表示 c_1 是 c_2 的一个实例. 对于公理集合 X 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 $(\text{Inverse}, \text{Teach}, \text{Taught-by}, 1)$. 其中关系的程度初始值由专家的经验给出, 并不断演化.

3 用模糊资源描述框架表示模糊本体

本体可以用 RDF 表示, 那么模糊本体用模糊资

源描述框架(Fuzzy Resource Description Framework, FRDF)表示. FRDF模型用三元组(主体、(谓词、权值)、客体)来定义Web上的模糊资源和描述资源之间的模糊关系. 模糊本体定义中的 R 是四元组 (r, c_1, c_2, f) ,它可以用FRDF表示, r, c_1, c_2, f 分别对应FRDF中的谓词、主体、客体以及权值. 例如:在图1中,本体模型中的关系 $R=(\text{name}, \text{person123}, \text{tom}, 0.9)$ 用FRDF可以描述为 $(\text{person123}, \text{name}(0.9), \text{Tom})$ 三元组的形式^[3].

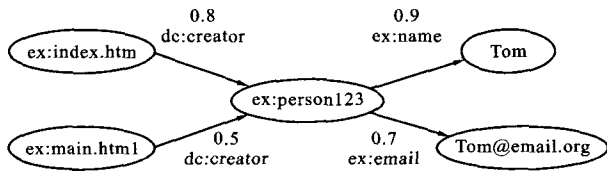


图1 模糊认知图对FRDF的表示

4 模糊本体的演化算法

在创建本体的初期,领域专家根据以前的经验及对当前事实的考察定义了领域范围内的本体.但是,随着社会的进步和科技的发展,使得本体也发生了改变.可能是添加了某些新的概念,也可能是旧的概念被新的概念所取代,还有可能原来关系很微小的两个概念之间的关联明朗化.这就需要领域专家能够及时地修改本体内容,还要能够根据历史的走向预知未来的发展趋势.

根据反馈神经网络的思想,可将本体中的概念看作是网络上的节点,演化的过程可看成是根据节点之间连接的变化情况确定节点的语义.任意两个节点之间都有连接权值,两个节点之间的连接权值表示了节点之间的连接强度^[4].如图2所示,为基于反馈神经网络的演化示意图.

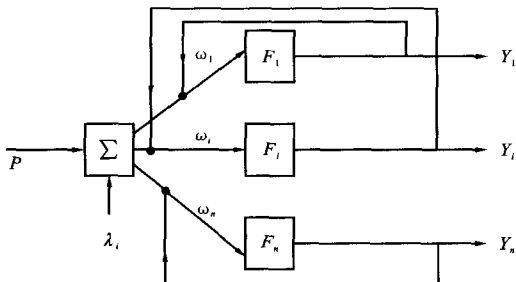


图2 演化示意图

图2中的 P 是反馈神经网络的输入,代表用户当前的关注概念; Σ 是神经网络的神经元,作用是将用户层次的概念转化为本体层次的概念; F 也是神经网络的神经元,也是中间结点,它代表的是输入概念相关联的概念; Y 是神经网络的输出,也是用户

概念语义的输出; ω_i 是输入 P 与和中间结点 F 的连接权值,代表了用户当前关注的概念和与其相关联的概念之间的关联程度; λ_i 是连接系数.

根据对反馈型神经网络的理解研究,提出本体的演化算法:

$$\omega_t(i) = (1 + \lambda_{t-1}(i))\omega_{t-1}(i) \quad (1)$$

其中, ω_t 为第 t 次迭代时的权值; $\lambda(i)$ 为主体 A 与第 i 个客体 B_i 之间连接的常数.

λ_i 的初值为相同的较小正数,其大小根据网络的规模来定.网络规模大, λ_i 的取值较小;网络规模小,则 λ_i 取值较大些.并且在演化的过程中, λ_i 的取值不变.演化的过程中,根据用户的偏好选择,选中的输出与用户输入节点之间的连接权值按式(1)演变,其他与用户输入节点相关的连接权值则不发生变化.一次演化过程结束时,权值要归一化,归一化公式如下:

$$\omega_t(i) = \omega_t(i) / \sum_{j=0}^n \omega_t(j) \quad (2)$$

归一化的目的是使得每次演化后相关权值的和总是等于1,权值的大小就表示了关联强度的相对大小.这样,虽然演化算法每次只改变用户选择的偏好,但是,经过归一化后,与用户当前输入有关的连接权值都发生了演变.

5 模糊本体的演化实验

5.1 系统实现步骤

(1) 创建本体

本体的创建分为通用本体和用户本体的创建,它们都是创建在服务器端的本体库.这里主要研究的是用户本体的创建.

① 定义类及继承关系

图3定义了类及继承关系.其中有网站、数码、服装及水果四个基本类.

基本类与子类之间的关系在某种意义上存在模糊性,图中的权值代表了基本类与子类之间的关联程度,权值的大小表示子类在父类中所占的比重的大小.模糊度的初始值的取值受到应用领域的限制,可以用连接权值来表示这种关系^[5-6].

② 定义类之间的关系

类之间关系的定义如图4所示.图中定义了苹果类与水果类、数码类及服装类之间的关系,苹果与数码、服装之间的关系是品牌关系,与水果之间的关系是继承关系.这里的关系同样具有模糊性,图中的权值表示了类之间关系的模糊性,即关联程度的大

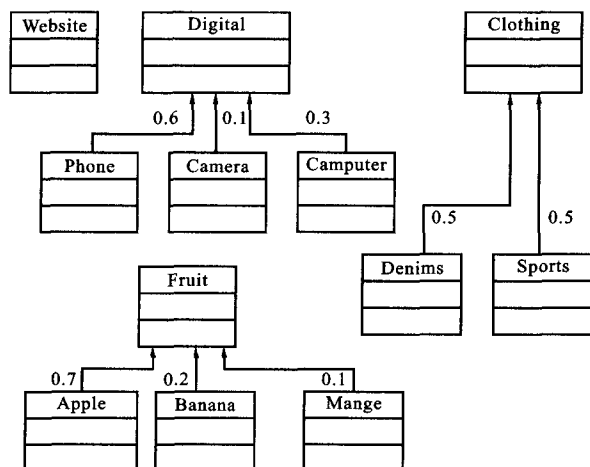


图 3 定义的类及层次关系

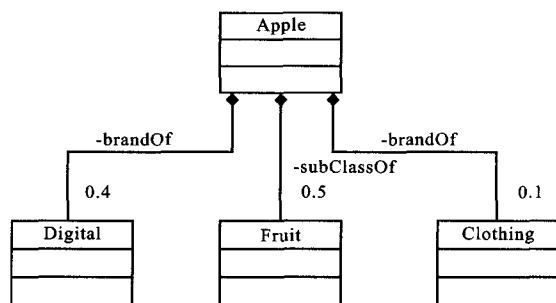


图 4 类之间的关系

小。同样,这里权值的设定也要考虑应用领域。

与类及继承关系的定义一样,对于类之间关系的定义中,一个类和所有与其相关联的类之间的权值的和要等于 1。图 4 同样只列举了几个常见的与苹果类有关联的类。

③ 定义类的属性

类的属性如图 5 所示。

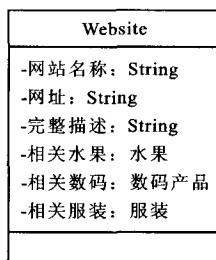


图 5 类的属性

(2) 对信息资源进行标注

对信息资源进行语义标注并生成实例文件,由于语义检索是针对格式统一的文档,该文档是基于 XML 语法并且以 FRDF 三元组的形式进行描述信息资源,并结合元数据库中的元数据如网站名称、网站描述、网址等。用这些元数据来生成实例文件,这些元数据和本体中的概念和属性是一致的。由于对信息资源添加了模糊性的描述,因此,采用手工标注

技术对这些信息资源进行标注。

5.2 实现结果分析

(1) 信息检索

举例,输入“苹果”一词。系统经过语义分析推理,判断“苹果”的语义是“水果”、“数码”、“服装”还是“电影”的模糊度即权值的大小,然后访问资源描述库,最后将查询结果按权值大小顺序返回到客户端。查询结果如图 6 所示。

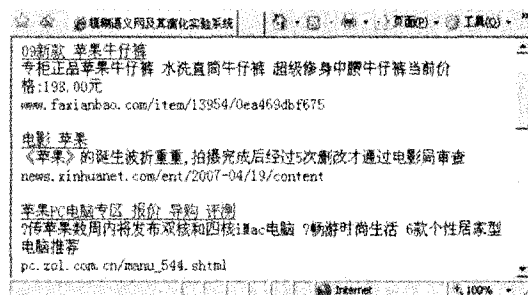


图 6 查询结果

此外,系统还可以根据同一用户前一次的用户偏好来判断当前的用户层次概念的语义。这一智能化的检索加快了检索的速度,提高了检索的效率^[7]。

(2) 演化演示

演化的演示通过演化曲线图来实现,系统根据用户输入的主体查询数据库,读取与此主体相关的客体及主客体间的连接权值,绘制演化曲线。曲线的走势代表了本体演化的方向,如图 7 所示。

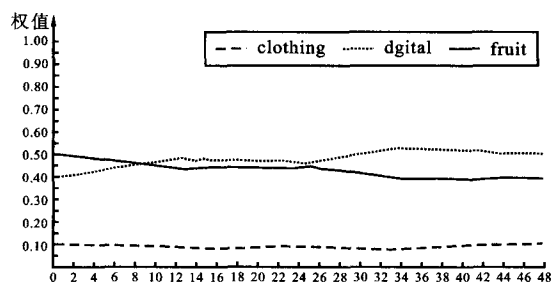


图 7 演化曲线图

本体的建立要在特定的领域、特定的自然环境及特定的社会环境内进行。文中主要研究个体本体的演化情况,并且将本体按演化层次分别存储在不同的数据表中。即将前面 5.1 节中的类与其子类之间的关系以 FRDF 三元组的形式构成本体库,存储在关系数据库中。而将类之间的关系存储在另一个本体库中,即存储在不同的数据表中。这样就形成了上层本体与下层本体的结构。

模糊语义网的演化具有多重性,即演化分多个层次,从底层开始逐层向上演化,并且底层的演化会

对上层的演化起到一定的影响作用.例如,用户的偏好是苹果牌的牛仔裤,那“服装类”与“牛仔”的连接权值就会变大,传递到上一层级,通过推理就可得出“苹果类”与“服装类”之间的关联程度相对较大,那么二者之间的连接权值就会变大^[8].

演化是一个漫长的、不确定性的循环过程,在经过了长期的演化后,就可清晰地看到权值演化的趋势,即本体语义演化的趋势.演化的部分数据如图8所示.

main	object	weight
2 digital	computer	0.25
3 digital	phone	0.3
3 digital	mp3	0.3
4 digital	camera	0.15
5 apple	digital	0.4
5 apple	clothing	0.1
7 apple	fruit	0.5
8 clothing	sport	0.3
9 clothing	leisure	0.3
10 clothing	occupation	0.1
11 clothing	death	0.3
12 fruit	real	0.6
13 fruit	tropical	0.4
14 clothing	sport	0.298211
15 clothing	leisure	0.298211
16 clothing	occupation	0.0994036
17 clothing	death	0.304175
18 apple	fruit	0.499002
19 apple	digital	0.399202
20 apple	clothing	0.101796
21 clothing	sport	0.296406
22 clothing	leisure	0.296406
23 clothing	occupation	0.0948925
24 clothing	death	0.308382

图8 演化部分数据

模糊语义网的演化使得基于语义网的智能检索更加快速、准确.对收集的所有信息资源进行标注后的结果以 FRDF 三元组的形式存储在基于 XML 语法的实例文件中. FRDF 三元组中的权值会随着用户的选择偏好而改变.例如,输入“苹果”一词,会返回多个与此相关的网站或网页,用户会在其中选择一个或多个结果.在经过漫长的交互后,用户选择次数多的网站或网页与“苹果”这一概念之间的连接权值就会有变大的趋势.这里所指的连接权值的改变依据,除了用户的选择偏好,还要考虑用户浏览此网页的时间,可选取一个时间临界值,小于此临界值,则可以认为此网页的内容并不是用户想要的,在这种情况下,连接权值不会变大,而会变小.每次返回的结果都会以与“苹果”之间连接权值的大小来排序.另一种情况,系统在对当前用户输入进行语义分析推理时,会根据此用户先前的输入记录来推理当

前用户输入的语义.这样会使得检索效率大大提高.

6 结束语

文中重点探讨语义网 RDFS 层和本体层,给出了模糊本体的概念及其形式描述,用模糊资源描述框架(FRDF)对模糊本体进行表示,提出了一种模糊本体演化算法并给出了实验验证,为语义网的研究发展提供了一条新的思路.

参考文献:

[1] 陆建江,张亚非,苗壮,等. 语义网原理与技术[M]. 北京:科学出版社,2007. 3.

[2] 刘植惠. 本体(Ontology)与语义网(Semantic Web)[J]. 重庆图情研究,2006(3):1-4.

[3] 吴彬,罗钧旻. 基于模糊认知图的资源描述框架(RDF)研究[J]. 微电子学与计算机,2009(3):42-44.

[4] 刘萍萍,罗钧旻,乔奎贤. 基于神经网络的模糊认知图的演化研究[J]. 微电子学与计算机,2007,24(7):69-71.

[5] 韩亚洪,许卓明. 怎样在关系数据库中存储 RDF[J]. 计算机与现代化,2002(11):70-72.

[6] 张桂元,贾燕枫. Eclipse 开发入门与项目实践[M]. 北京:人民邮电出版社,2006.

[7] Yu Wei, Cao Jiaheng, Chen Junpeng. Novel method for searching ontologies on semantic web[J]. Journal of Southeast University(English Edition), 2006,22(3):11-17.

[8] Kang Dazhou, Xu Baowen, Lu Jianjiang, et al. Description logics for fuzzy ontologies on semantic web [J]. Journal of Southeast University (English Edition). 2006,22(3):343-347.

作者简介:

罗钧旻 男,(1963—),博士,教授.研究方向为智能软件和领域软件.

武钰云 男,(1984—),硕士研究生.研究方向为智能软件与领域软件研究.

吴彬 女,(1983—),硕士研究生.研究方向为智能软件与领域软件研究.