

模糊本体生成系统的研究

何爱娜, 罗钧旻

(西安工业大学 计算机科学与工程学院, 陕西 西安 710032)

摘 要: 用 protégé 建立的本体中概念及其关系都是确定的, 但是, 现实世界中事物及其关系往往具有模糊性且这种模糊性会随着事物的发展而变化. 文中以模糊资源描述框架(FRDF)为基础, 把其引入本体的建立工具 protégé, 对其进行相应的扩充与修改以适应模糊本体的建立, 着重讨论了模糊本体的存储方式、模糊 protégé(FProtégé)的功能、结构和实现方法.

关键词: 本体; protégé; 模糊本体; 模糊 protégé

中图分类号: TP18

文献标识码: A

文章编号: 1000-7180(2011)05-0105-04

Fuzzy Ontology Generation System

HE Ai-na, LUO Jun-min

(Department of Computer Science and Engineering, Xi'an Institute of Technology, Xi'an 710032, China)

Abstract: The concept and relation of the ontology information with protégé is definite. However, things in the real world and their relations are often fuzzy and this ambiguity may change as things develop. Based on the FRDF (Fuzzy Resource Description Framework) in this paper, the ontology creation tools protégé is introduced and supplemented and modified to fit for the Fuzzy ontology establishment, and impressed the storage methods of fuzzy ontology, the function of fuzzy protégé (FProtégé), structure and realization methods.

Key words: ontology; protégé; fuzzy ontology; fuzzy protégé

1 引言

本体是语义网基础结构的主要组成部分. 目前, 国内外对本体的研究比较多, 但对本体编辑工具 protégé 的研究比较少, 而且 protégé 建立的本体信息都是确定的信息. 文中研究基于模糊资源描述框架 FRDF 的模糊 protégé, 以弥补 protégé 的不足, 增强其语义描述的功能.

2 模糊本体的意义

本体是对客观世界的存在进行系统化描述, 方便知识的重用集成和交互, 但是人们描述客观世界是从大量的、不完全的、模糊的和随机的数据中提取隐含在其中的, 人们事先不知道的但又是潜在有用信息和知识的过程. 模糊本体设想的提出, 使得本体的概念具有模糊性, 概念间的关系也同样具有模糊

性. 并且, 这种模糊性的大小会因为外界因素的影响而改变, 对事物的认知理解会随着种种因素的改变而改变^[1].

在创建本体的初期, 领域专家根据以前的事实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. 模糊本体是对实际问题进行模糊判断、模糊决策、模糊识别等. 由于模糊性是客观存在的, 而且系统的复杂性越高, 精确化能力越低, 这就意味着模糊性越强. 所以, 模糊本体就应运而生^[2].

3 模糊本体的概念及其形式化描述

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.为了描述概念及概念间的不确定关系,文中引入模糊系数,得到模糊本体模型.

(1) 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).

其中 R 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 Find-of 或 Is-a, 表示 Superclass—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. 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 中的每一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 $(\text{Inverse-of}, \text{Teach}, \text{Taught-by}, 1)$. 其中关系的程度由专家的经验给出, 并不断演化^[3].

(2) 模糊资源描述框架

模糊 protégé 是在 FRDF 的基础上产生的. RDF 数据模型的核心是用(主体、谓词、客体)三元组来定义 Web 上的资源, 如果三元组中的客体是资源的话, 那么这个三元组就描述了两个资源间的关系. 但是, 这种情况下所描述的关系是静态的. 不能够表示出关系的程度以及关系程度随时间、空间的改变而改变的情况. 所以, 在(主体、谓词、客体)中的谓词部分加入一个表示谓词的模糊性, 即主体与客体间关系的模糊度的权值. 提出一个新的概念——模糊资源描述框架(Fuzzy Resource Description Framework, FRDF), 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^[4].

4 模糊 protégé 的功能

(1) 模糊本体的创建. 它将 Web 上的资源赋予模糊性, 描述了资源间的模糊关系. 向本体中的属性加入隶属度并设置隶属度的隶属区间来创建模糊本体.

(2) 模糊本体的存储. 除内存中存储数据模型之外, 对数据量较大的 Data 或者 Ontology 数据可以通过 FRDF 提供的数据库将其存储在数据库

表中.

(3) 模糊本体更新. 本体演化是一个本体面对不断产生的变化要能够具备随时间变化的适应能力, 随着时间的推移, 由于知识具有内在的变化性质, 具有高度的动态性质, 领域的变化, 对不同任务的适应性, 以及本体在概念化和关系上的改变均要求给出的知识表示方法要具备很好的动态适应性.

(4) 模糊本体的使用(查寻与推理). 在模糊逻辑中, 命题都是由模糊集表述的模糊命题. 模糊逻辑的最终目的是将模糊理论作为一种主要工具, 为不精确命题的近似推理提供理论基础. 为实现这一目的, 引入模糊推理^[5].

5 模糊 protégé 的结构

Protégé 是基于框架的体系是可以与 OKBC (Open Knowledge Base Connectivity) 协议兼容. Protégé 的体系结构是单一的体系结构. 一个本体由一个在一个包容层次结构组织的类的集合来表示一个域的突出概念, 一个相关的类槽集来描述他们的属性和关系, 以及对这些类的实例设置. 框架特性包括:

(1) 一个用户界面, 可定制模型知识, 让用户和输入的内容广泛的友好形式的数据集.

(2) 一个插件架构, 可与定制设计元素, 如图形组件(如图表和表格), 媒体(声音、图像、视频), 不同的存储格式(如 RDF, XML 和数据库后端), 和更多的支持工具(如本体可视化、推理等).

(3) 一个基于 Java 的应用程序编程接口 (API), 它是为插件和其他应用程序访问, 使用, 并与 protégé 框架创建显示本体.

其中它的扩展性由插件来完成. Protégé 知识库由框架表示类、槽、面(facets)、实例和约束组成. 类 Classes 是论域领域中的概念. 它们是领域中的抽象概念的条目或命名的实例的集合. 槽 Slots 是两值的关系, 它描述了类的属性或性质. Facets 描述槽的性质用来说明槽值许用范围的约束, 如槽值(每个槽可以允许有几个值), 类的实例 Instances 有明确的槽值. 约束 Constraints 指定实例间的附加的关系. Protégé 系统构成复杂, 约有 600 多个类(.java 组成)组成. 分为 11 个大类, 但是系统架构比较清晰, 有利于系统的扩展和分析. 系统分为界面层、模型层和存储层, 还有大量辅助性的类包^[6]. Protégé 的体系结构如图 1 所示.

Protégé 体系结构的最主要的优势就是它的开

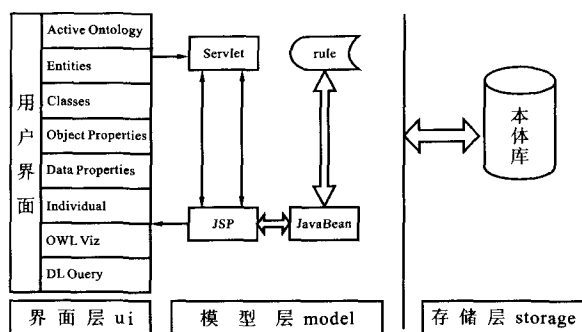


图1 Protégé的体系结构图

放的模块化的风格. 基于组件的体系结构使系统开发者可以通过生成恰当的插件增加新的功能. 插件可以分为三类:一类是后端插件,使用户可以以多种格式来存储和输入知识库;一类是 slot widgets 类插件,用来为特定的域或特定的任务合并槽或显示和边界槽值;第三类是 tab 插件,通常与 Protégé 知识库一样,提供基于知识的应用. 后端插件支持在 RDF Schema、带 DTD 的 XML 文件、XML Schema 文件中存储和导入知识本体. slot widgets 插件包括显示 GIF 图片和音频视频的用户界面组件. tab 插件非常普及,提供高级可视化、知识本体合并、版本管理、推理等功能^[7].

模糊 protégé 的设计: 模糊 protégé 是在 protégé 的基础上制作一个模糊插件来实现模糊 protégé 的功能. 步骤如下:

(1) 在 protégé 的 Tabs 插件的下拉菜单 Create new tab 建一个权值按钮 (Weight value).

(2) 后端插件用来保存建立好的模糊本体. fowl 文件. Protégé 建的本体是用 owl 存储的. 其中在本体资源部分(主体、谓词、客体)中谓词部分加入一个表示权值的元素,即谓词的模糊性,即主体与客体间关系的模糊度(主体、(谓词,隶属度)、客体). 隶属度记为 μ_A , 定义 $\mu_A(x): x \in X \rightarrow [0, 1]$, 表示元素 x 属于 A 的程度. 模糊集合 A 可以表示成 $A = \{ (x, \mu_A(x)), x \in X \}$. 当隶属函数的值域由区间 $[0, 1]$ 蜕化为两个数 $\{0, 1\}$ 时,相应的模糊子集就成为普通的集合. 在 FOWL 中,隶属度的概念可视作对公理描述客观事务的真实程度. 如可以指派给 $\mu_A(x)$ 在 $[0, 1]$ 之间的一个值,以表达断言 “ x is A ” 的程度.

例1 一个模糊本体 Tom 很年轻的隶属度为 0.8, 他喜欢篮球隶属度为 0.9. 即 Tom. fowl 的存储后如下:

```
<owl:Class rdf:ID="Person"/>
<owl:Class rdf:ID="YoungPerson">
```

```
<owl:subClassOf rdf:resource="# Person"/>
</owl:Class>
<owl:Class rdf:ID="Minor">
<owl:subClassOf rdf:resource="# Person"/>
<f:membershipOf>
<f:subClassOf rdf:resource="# YoungPerson"/>
<f:moreOrequivalent f:weigh=0.8>
</f:membershipOf>
</owl:Class>
<owl:EnumeratedClass rdf:ID="Hobby">
<owl:oneOf parseType="Collection">
<owl:Thing rdf:about="# Baseball">
<owl:Thing rdf:about="# Basketball">
<owl:oneof>
<owl:EnumeratedClass>
<owl:ObjectProperty rdf:resource="like">
<rdfs:domain rdf:resource="# Person">
<rdfs:range rdf:resource="# Hobby">
</owl:ObjectProperty>
```

(3) slot widgets 类插件显示出例 1 的模糊本体图形. 如图 2 所示.

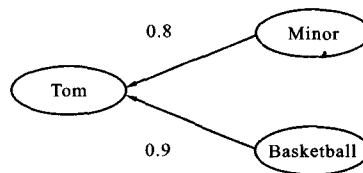


图2 模糊本体图

6 关系权值的选取问题

Protégé 中主要有类 (Classes)、属性 (Object Properties 和 Data Properties)、个体 (Individuals). 下面讨论 FRDF 中谓词权值的确定原则:

(1) 当创建的本体是来自不同的类即来自同一层次的类时,要申请 Disjoint Classes,这时取权值为 $u=0$,即两个个体之间没有关系.

(2) 当创建的本体表示的是一个特殊的意思时比如 protégé 中的函数属性 (Functional Properties),通过这个属性只能连接一个个体. 函数属性图如图 3 所示.

图 3 中 Jean, Peggy, Margaret 是人名, Peggy 和 Margaret 是同一个人. 取隶属度值为 $u=1$,即表示的是同一个本体.

以上两种情况是相当于模糊数学中的当 u_A 仅取值 $[0, 1]$ 的两个端点 $\{0, 1\}$ 时,模糊集合 A 就蜕化成普通集合 A .

(3) 除了上述两种情况之外创建的本体表示不

确定的概念时,取 $\{0,1\}$ 之间的值.

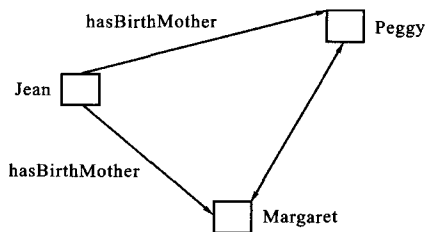


图3 函数属性图

属性中有 Domains(intersection)表示定义域和 Ranges(intersection)表示值域. protégé 中使用属性对类进行描述和定义,这要通过对属性加以限制,这样 protégé 中就引入了 Property Restriction. 在 Property Restriction 中有一个为量词限制(Quantifier Restriction). 量词限制由三部分构成,分别为量词(quantifier)、属性和 filler. 量词分为存在量词(some 等)和全称量词(only),即 owl 中的 some value form.

现在,protégé 中只是说明一个本体由一些 some(existential)组成,而没有说明它们所占的比重. 用模糊 protégé 建出来的本体是一个四元组 $O = (C, R, E, I)$. 其中, R 就是模糊关系. 权值的初值是根据专家经验来确定^[8].

(4)一个物体本身不止和一个属性有关,它可以和好多属性有关,比如,一个 pizza 它可以由好多味道组成,即可以有好多属性. 模糊数的大小由各自占的比例决定. 如图 4 所示.

7 结束语

Protégé 是一个本体编辑工具,它建立的本体是确定的信息,文中是在模糊资源描述框架 FRDF 的基础上提出了一个模糊 protégé,用它建立的信息具有模糊性,更符合人们的思维,也提高复杂知识及其语义的表示处理能力.

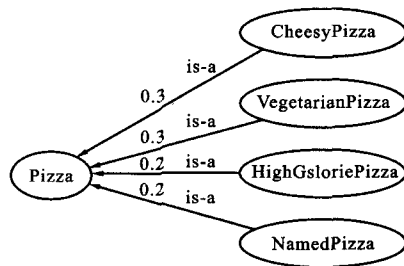


图4 模糊本体多重属性图

参考文献:

- [1] 吴彬,罗钧旻. 基于模糊认知图的资源描述框架(RDF)研究[J]. 微电子与计算机,2009,26(3):42-44.
- [2] 陆建江,张亚非,苗壮,等. 语义 web 原理与技术[M]. 北京:科学出版社,2007.
- [3] 洪娜,张智雄. protégé 在科研本体构建与推理中的实践研究[M]. 北京:中国科学院研究生院,2009.
- [4] 宋炜,张铭. 语义网简明教程[M]. 北京:高等教育出版社,2004.
- [5] 王晓露. 一种面向对象的模糊知识表示及其推理研究[D]. 北京:中国科学院研究生院:计算技术研究所,1996.
- [6] protégé[EB/OL]. (2009-03-01)[2010-03-13]. <http://protégé.stanford.edu>
- [7] A practical guide to building OWL ontologies using protégé 4[EB/OL]. (2010-03-19)[2010-06-13]. [http://owl.cs.manchester.ac.uk/tutorials/protegeowl-tutorial/resources/protegeOWL TutorialP4_v1_2.pdf](http://owl.cs.manchester.ac.uk/tutorials/protegeowl-tutorial/resources/protegeOWL%20TutorialP4_v1_2.pdf).
- [8] BernersLee T, Hendler J, Lassila O. The semantic web[J]. Scientific American, 2001,284(5):34-43.

作者简介:

何爱娜 女,(1979-),硕士研究生. 研究方向为智能软件和领域软件.

罗钧旻 男,(1963-),博士,教授. 研究方向为智能软件和领域软件.