维普资讯 http://www.cqvip.com



# RDF可信度扩展在领域本体构建中的应用

陈 坚,何洁月
(东南大学 计算机科学与工程系,江苏 南京 210096)

**摘 要**:仅由少数权威领域专家的参与难以实现领域本体构建的繁重任务。文中对RDF进行可信度扩展,标记每个声明被领域专家认可的程度,协助多用户共同参与领域本体的构建,在一定程度上解决了目前仅由少数权威领域专家构建领域本体工作量繁重的困难,为知识动态发展和更新迅速的前沿领域进行本体构建提供了方便,同时也为领域本体的表示引入了量化的机制。

**关键词**:本体构建;领域本体;可信度
**中图分类号**:TP301 **文献标识码**:A **文章编号**:1005-3751(2006)01-0120-03

# A Credibility Extension to RDF and Its Application for Domain-Ontology Building

CHEN Jian, HE Jie-yue
(Dept. of Computer Science & Engineering, Southeast University, Nanjing 210096, China)

**Abstract**: Domain ontology building is a great burden with few authorities. So extend credibility to RDF, marking the acceptability of every statement among domain experts, which presents a credibility-based method for multi-user collaborative building of domain ontology and mitigates the burden of domain ontology building. Gives convenience to ontology building among fast developing domains and introduces a quantitative measurement to the ontology representation.

**Key words**: ontology building; domain ontology; credibility

本体作为对"共享概念模型的明确的规范化说明"[1],在解决数据异构问题和实现计算机对数据语义理解的巨大前景,引起各个领域尤其是生物医学领域的关注。目前大部分领域本体都由专门机构组织少数权威领域专家来构建,如GO[2]。由于领域知识的海量性、关联的复杂性和发展的动态性,若仅由少数权威领域专家手工构建领域本体及对本体的调整和修改则工作量庞大。目前各个组织往往针对领域的特定小范围构建本体,如生物医学领域有:Sequence Ontology, MGED Ontology,其负面效果是割裂了领域内各知识的关联。

文中提出一种RDF的可信度扩展,模拟共享概念形成的过程,协助领域内多个专家共同参与领域本体的构建,在一定程度上克服了仅由少数权威领域专家构建领域本体的困难,并为领域本体的表示引入了量化的机制。

## 1 RDF的可信度扩展

传统模式下RDF声明(statement)表示成三元组形式<subject, predicate, object>[3],一个声明只有两个状态:存在或不存在。但在领域知识的发展过程中,概念模型间的声明从提出到被大多数领域专家认可有一定过程,而且并不一定被所有领域专家认可。这里对传统的RDF表示方法进行扩展,为每个声明设置可信度(credibility),标记该声明被领域专家认可的程度,表示成<subject, predicate, object, credibility>。扩展后的本体表示见图1。

```
<onke:药物_C rdf:ID="多巴胺_I">
    <onke:cause_P rdf:resource="#肾小动脉扩张_I"/>
    <onke:cause_P rdf:resource="#心率加快_I"/>
    <onke:cause_P rdf:resource="#血管扩张_I"/>
    <onke:cause_P rdf:resource="#血压增高_I"/>
</onke:药物_C>
<onke:药物_C rdf:ID="多巴胺_I" onke:credibility=1>
    <onke:cause_P rdf:resource="#肾小动脉扩张_I" onke:credibility=0.95/>
    <onke:cause_P rdf:resource="#心率加快_I" onke:credibility=0.9/>
    <onke:cause_P rdf:resource="#血管扩张_I" onke:credibility=0.8/>
    <onke:cause_P rdf:resource="#血压增高_I" onke:credibility=0.85/>
</onke:药物_C>
```

图1 基于RDF可信度扩展的本体表示

这种扩展建立在传统本体表示方法的基础上,可以通过损失部分信息,如不考虑credibility或将所有的credibility值都设为1,弱化为传统的本体,并可以利用jena, racer等相关工具进行处理。

## 2 RDF的可信度扩展在多用户协作领域本体构建中的应用

### 2.1 可信度的范围及相关阈值

credibility标记每个statement被领域专家认可的程

收稿日期:2005-04-24
作者简介:陈 坚(1979—),男,福建龙岩人,硕士研究生,研究方向为语义Web、本体集成、生物信息;何洁月,副教授,研究方向为数据库技术、数据挖掘、生物信息学。

维普资讯 http://www.cqvip.com



度,取值范围为(0,1)。领域本体构建过程中多用户的协作通过 credibility 体现出来,如图 2 所示,statement 刚创建时 credibility 赋予创建初值,随着领域专家对概念模型间的声明提出各自的观点,credibility 值发生变动,大于有效阈值时表示这个 statement 可以有效参与本体的查询推理等各种操作,大于最大阈值时,这个 statement 将具有相对稳定性(具体见后面对 credibility 值的调整),反之,如果小于删除阈值则这个 statement 将被删除。这几个阈值根据具体的应用设定。

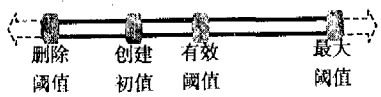


图 2　credibility 的取值范围及 4 个阈值

在此基础上推理规则演变成:

$(s_{i3}, p_{k3}, o_{j3}, \text{credibility}_3) \leftarrow (s_{i1}, p_{k1}, o_{j1}, \text{credibility}_1), (s_{i2}, p_{k2}, o_{j2}, \text{credibility}_2)$

iff:

i. 在传统的本体表示中满足:$(s_{i3}, p_{k3}, o_{j3}) \leftarrow (s_{i1}, p_{k1}, o_{j1}), (s_{i2}, p_{k2}, o_{j2})$;

ii. $\text{credibility}_3 = \text{credibility}_1 * \text{credibility}_2 \geqslant$有效阈值。

### 2.2　多用户协作领域本体构建中可信度的调整

statement 可信度的调整决定于多个领域专家的观点,一个领域专家对某个声明的认可程度用 viewpoint 表示,取值范围为(-1,1)。领域专家可以对概念模型提出新的声明,也可以对已有声明发表或修改自己的观点。为了反映不同领域专家观点的重要性,笔者对领域专家的经验值进行管理,用 experience 表示,取值范围为(0,1),随着领域本体的调整,若领域专家的观点与领域本体的变动一致,则其经验值提高,反之降低。针对某个领域专家的操作,按如下的算法调整相应声明的可信度(ViewNum 是当前状态下对这个声明发表观点的领域专家人数):

(1)领域专家查询两概念模型间的 statement,如果没有与自己观点相对应的 statement,转(2),否则转(3)。

(2)新建 statement,credibility = viewpoint * experience,转(5)。

(3)对指定概念模型提出新的声明并转(4),或者修改自己的 viewpoint,转(6)。

$$(4)\begin{cases}\text{credibility}' = (\text{credibility} \times \text{ViewNum} + \text{viewpoint} \times \text{experience}) / (\text{ViewNum} + 1) \\ \text{ViewNum}' = \text{ViewNum} + 1\end{cases}$$

并对从本节点到根节点的类层次路径上的每一个 statement 进行调整:

$$\begin{cases}\text{credibility}' = (\text{credibility} \times \text{VewNum} + 1) / (\text{ViewNum} + 1) \\ \text{ViewNum}' = \text{ViewNum}\end{cases}$$ 转(5)

(5)若 credibility < 删除阈值,通知对该 statement 发表过观点的经验值最高的 10 个领域专家,若一定时间段内超过 5 人赞同删除,则删除该 statement,否则放弃删除计划转(7)。

(6)修改领域专家观点:

$\text{credibility}' = \text{credibility} + (\text{viewpoint}' - \text{viewpoint}) \times \text{experience} / \text{ViewNum}$,转(7)。

(7)完成。

$\alpha_1$ 是系数,取值范围为(0,1),它尚在调整中,暂设为 0.995。模型中的 4 个阈值也尚在调整中,暂定为:删除阈值 = 0.10,创建初值如算法所述,有效阈值 = 0.3,最大阈值 = 0.9。

本体中概念模型应该具有相对的稳定性,虽然没有对这种稳定性额外加以控制,但在上面的设计中,每个声明的 credibility 值的变化都导致节点到根的类路径上概念模型间声明的可信度增大,客观上维持了上层节点相对的稳定性。

### 2.3　基于可信度扩展的多用户协作的领域本体构建过程

在上面讨论的基础上,给出图 3 的单个 statement 调整的模式图。随着领域专家都对概念模型间的声明提出自己的观点,声明的可信度发生调整。新的概念模型因声明的需要而创建,当某个声明的可信度低于指定阈值时,该声明将被删除。若一个概念模型和其它概念模型间的声明都被删除后,该概念模型成为孤立的节点也将被删除。从而将概念模型的调整转换为声明的调整,在图上表现为节点间边的变动,从而转移了本体构建过程中本体调整和修改的困难。领域本体构建的整个过程表现为多用户协作的以可信度变化引起的概念模型间声明的添加、删除以及由此导致的概念模型的变动。

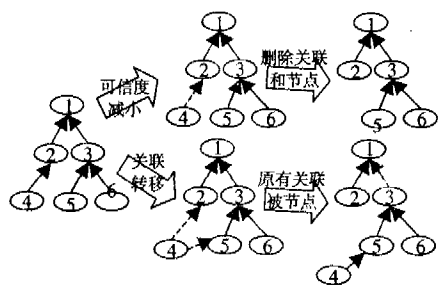


图 3　单个 statement 调整模式图

图中,随着 $\text{credibility}_{2,4}$ 的减小,当 < 删除阈值时,$\text{Statement}_{2,4}$ 将被删除,如果没有其他节点与节点 4 关联,将删除孤立节点 4。

## 3　原型系统的实现

在此基础上,实现了多用户协作领域本体构建工具,以 java applet 形式发布,可直接用 IE 访问。领域本体的图形化显示工作是在 karlsruhe 大学 KAON 相关工作基础上修改的,关于 KAON 请查看文献[4],而领域本体局部信息的显示包括某个类的父类、子类、属性。被选中的 statement 在系统中列出所有领域专家的观点,其中包括该领域专家对自己观点的解释,该解释一方面可以说服其他领域专家认可自己的观点,同时也是别的领域专家发表

维普资讯 http://www.cqvip.com



观点的参考。每个用户可以修改自己的解释。

每个 statement 有 3 个参数,图 4 是放大的显示,如标记 2 所示,第一个参数(1.0)本文没介绍,第二个参数(0.55197155)是本文介绍的 credibility,第三个参数(1.0)表示当前登录用户对这个 statement 的认可程度。双击某条边可以对当前用户的观点进行修改,如标记 1 所示。

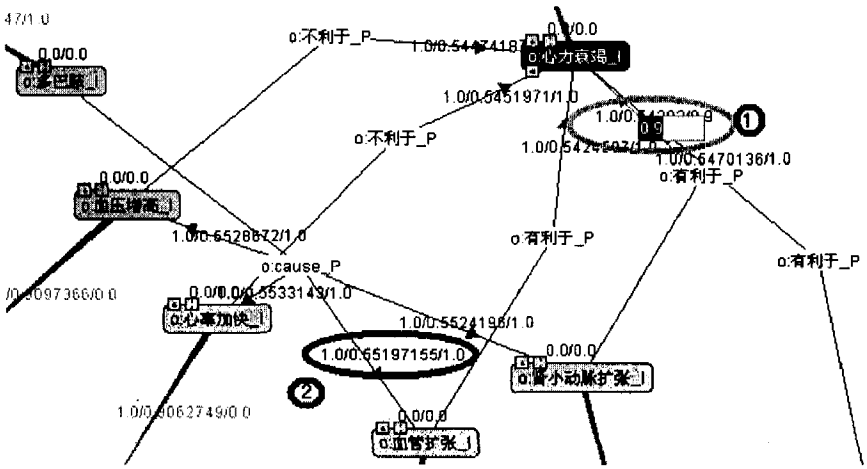


图 4 多用户驱动的领域本体演化系统参数的图形化显示和修改

## 4 和现有的本体构建工具的比较

目前支持本体构建的工具很多,主要有 Protégé[5],OntoEdit[6],KAON[4]等。Protégé 已成为广泛应用的本体构建和修改工具,提供强大的本体构建的功能,但侧重于单用户环境下的本体构建,没有提供对多领域专家协同构建本体的支持。OntoEdit 主要从本体工程的角度对本体构建过程中各个周期进行规划和建模。KAON 是德国 Karlsruhe 大学 AIFB 开发的本体工具,针对现有本体编辑工具的不足提供了全局的 log 功能和 undo/redo 功能,尤其支持本体构建后的演化,提出演化策略的概念,允许用户进行演化策略的选择,支持多个本体以 include 关系互相联系时的一致性维护,但是这些工具都没有对本体构建中的多用户协作提供特别的支持,本质上是一种单用户的本体构建。

文中对 RDF 进行可信度扩展,协助多个领域专家共同参与领域本体的构建,领域专家根据自己的领域知识和经验对已有概念模型间的声明发表自己的观点或提出新的声明,甚至不需深入理解领域本体构建的细节。系统根据领域专家的观点调整声明的可信度,相对客观地模拟共享概念间声明的形成过程,较好地解决了目前仅有少数权威领域专家构建领域本体的困难,也有利于全面及时地收集领域内各个专家的观点,在知识动态发展和更新迅速的前沿领域如生物医学领域具有比较明显的优势,也为领域本体的表示引入了量化的机制。

## 5 总 结

对 RDF 进行可信度扩展,在此基础上讨论了基于可信度的多用户协作的领域本体构建,通过多个领域专家的协作,模拟共享概念的形成过程,一定程度上解决了目前由少数权威领域专家构建领域本体的困难,为知识动态发展和更新迅速的前沿领域进行本体构建提供了方便,同时也为领域本体的表示引入了量化的机制。下一步将完善原型系统,并利用可信度所引入的量化机制,进一步探讨生物医学领域基于本体的语义关联的发现和排序。

**参考文献:**


[1] Studer R, Benjamins V R, Fensel D. Knowledge Engineering, Principles and Methods[J]. Data and Knowledge Engineering, 1998, 25(122): 161-197.

[2] Harris M A, Clark J, Ireland A, et al. The Gene Ontology (GO) database and informatics resource[J]. Nucleic Acid Res, 2004, 32(D2): 58-61.

[3] Manola F, Miller E. RDF primer. W3C Working Draft[EB/OL]. http://www.w3.org/TR/rdf-primer/. 2004.

[4] Haase P, Sure Y, Vrandecic D. Ontology Management and Evolution-Survey, Methods and Prototypes. Project Report D3.1.1, Institute AIFB, University of Karlsruhe[EB/OL]. http://www.aifb.uni-karlsruhe.de/WBS/pha/publications/main_d311.pdf. 2004-12-20.

[5] Noy N, Fergerson R, Musen M. The knowledge model of Protege-2000: Combining interoperability and flexibility[A]. In: Dieng R, Corby O. Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management: Methods, Models, and Tools (EKAW 2000), volume 1937 of Lecture Notes in Artificial Intelligence (LNAI)[C]. Juan-les-Pins, France: Springer, 2000. 17-32.

[6] Sure Y, Angele J, Staab S. OntoEdit: Multifaceted Inferencing for Ontology Engineering[A]. In: Spaccapietra S, March S, Aberer K. Journal on Data Semantics, LNCS, Volume 2800/2003[C]. Heidelberg: Springer-Verlag, 2003. 128-152.


(上接第 119 页)


[6] 王 实,高 文.数据挖掘中的聚类算法[J].计算机科学,2000,27(4):42-45.

[7] 周永权,焦李成.基于属性稀疏特征差异度的动态抽象聚类方法[J].系统工程与电子技术,2004,26(4):426-429.

[8] 何晓群.现代统计分析方法与应用[M].北京:中国人民大学出版社,1998.

[9] 武 森,高学东.高维稀疏聚类知识发现[M].北京:冶金工业出版社,2003.