

一种基于 HornML 的 Web 知识表示方法

黄海 王儒敬 黄河

(中国科学院合肥智能机械研究所,合肥 230031)

E-mail:huanghai@ustc.edu

摘要 论文提出了基于 HornML 的多级知识单元表示方法,它可以满足 Web 上对知识表示的要求,它适合于 Web 上分布式知识的交换和统一存储,具有比较强的表达能力且易于通过扩展 Schema 增加新的知识表示能力,方便表示 Horn 子句规则的同时介绍了在其上的推理方法,我们将其应用在网络农业专家系统中,取得了良好的应用。

关键词 HornML 知识表示 XML

文章编号 1002-8331-(2006)01-0053-03 文献标识码 A 中图分类号 TP18

A HornML - Based Knowledge Representation for Web Knowledge

Huang Hai Wang Rujing Huang He

(Institute of Intelligent Machines of Chinese Academy of Sciences, Hefei 230031)

Abstract: In this paper we have described the method of HornML - based knowledge representation of multi-level knowledge units which caters to demands of web environment. This knowledge representation with strong expression power is able to exchange and store assorted and distributed knowledge on the web. It is apt to express horn clause rules and increase expression power by extending Schema. We have applied it to web expert system of agriculture and achieved a better result.

Keywords: HornML, knowledge representation, XML

1 引言

知识表示问题是人工智能要研究的根本问题之一^[1],为了便于把知识和智能传授给计算机,必须有合适的知识表示语言。传统的知识表示语言主要有三个来源。第一个来源是计算机科学家们对可计算性理论的研究。例如,PROLOG 语言的理论基础是一阶谓词演算的子集 Horn 子句演算,它可实现于任何可计算函数。第二个来源是认知科学的研究成果,产生式表示、框架表示、语义网表示等实际上都有其认知模型作为背景。第三个来源是知识工程的实际需要。当前 Web 是一个巨大的知识源,有大量的知识需要在 Web 上表示,但 Web 有其自身的特点。Web 的一些重要特征主要有:Web 是分布式的;Web 发展的推动力量是自由而非集中控制;Web 是动态的;Web 是开放的等等。Web 的这些特性对传统的知识表示方法提出了新的要求。XML 是 Web 环境下的一种可扩展置标语言。XML 文档以树型结构包含和描述数据、数据类型以及文档结构,用它描述具有良好结构的知识。DTD 或 schema 为 XML 文档提供了一个语法规范,如果 XML 文档交换的双方事先在 DTD 或 Schema 的使用和意义理解上都达成一致,则可实现对文档的共享。基于 XML 的知识表示系统具有以下优点^[2]:(1)语法独立,能够为分布于网上的知识提供一种统一的存储和交换格式。(2)可扩充性,可以通过对 Schema 的扩展增加新的知识表示能力。(3)可综合多种知识表示方法,可用相同的 XML 语言重写多种传统知识表示方法。(4)可以对不同信息源的信息进行集成,并形成统一的文档。(5)可以实现数据的结构化,允许

不同 AI 应用之间知识库的交换。(6)标准化,XML 是 W3C 确定的 Web 上的标准数据格式。采用 XML 的知识表示可以在世界范围内定义标准化的公用且具备自我描述功能的数据。针对不同的知识形式,出现了多种知识标记语言,其中 HornML 就是其中的一种,它主要用来描述 AI 领域的 Horn 类型子句。由于基于 Horn 子句演算的 PROLOG 语言是传统 AI 的一种经典的知识表示语言,现实中有大量的规则知识采用 PROLOG 语言表示,但其不适合在 Web 环境下描述这些规则知识,而 HornML 就是一种能适应 Web 特点的知识标记语言。我们在研究网络农业专家系统过程中,结合 HornML 以及我们早先提出的多级知识单元,提出了一种新的基于 HornML 的知识表示方法。

2 Horn 子句及 HornML 简介

2.1 Horn 子句

原子公式或其否定称为文字。文字的析取叫子句。而 Horn 子句是一种特殊的子句,它只包含一个正文字。例如: $P(x,y) \vee \neg Q(x) \vee \neg R(y)$ 是一个 Horn 子句,它只有一个正文字 $P(x,y)$,写成 PROLOG 的语法就是: $P(x,y):-Q(x),R(y)$ 。

2.2 HornML 简介

HornML 是基于 XML 的规则标记语言,主要针对的是 Horn 逻辑语句的描述。一个 HornML 的 DTD 定义如下^[3]:

```
<! ELEMENT kb (hn*)>
```

```
<! ELEMENT hn (relationship,relationship*)>
```

基金项目:国家 863 高技术研究发展计划资助项目(编号:2001AA118070)

作者简介:黄海(1980-),男,硕士研究生,主要研究方向:语义网,本体,人工智能的逻辑基础。王儒敬(1964-),男,研究员,主要研究方向:智能决策支持系统。黄河(1980-),男,硕士研究生,主要研究方向:J2EE 企业级系统架构,语义 Web 服务。

```
<! ELEMENT relationship (relator,(ind | var | struc)*)>
<! ELEMENT struc (constructor,(ind | var | struc)*)>
<! ELEMENT relator (#PCDATA)>
<! ELEMENT constructor (#PCDATA)>
<! ELEMENT ind (#PCDATA)>
<! ELEMENT var (#PCDATA)>
```

其中单个常量采用<ind>元素描述,变量用<var>元素描述,谓词或关系符号用<relator>元素描述,Horn子句的规例如:grandfather(x,z):-father(x,y),mother(y,z);用HornML表示为:

```
<hn>
<relationship>
  <relator> grandfather </relator>
  <var> x</var>
  <var> z</var>
</relationship>
<relationship>
  <relator> father </relator>
  <var> x</var>
  <var> y</var>
</relationship>
<relationship>
  <relator> mother </relator>
  <var> y</var>
  <var> z</var>
</relationship>
</hn>
```

其中第一个<relationship>元素表示的是正文字,其余的项在后面。可用RDF图^[4]来表示,例如对于事实:Teacher(Gruber, Tom)。

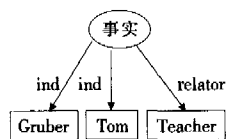


图1 Teacher{Gruber, Tom}的RDF图

3 基于HornML的多级知识单元知识表示方法

3.1 多级知识单元的知识表示方法^[5]

按照领域知识的特点和知识表达的功能,知识单元可划分为框架单元、规则单元等多种知识单元模式。知识单元是知识单元模式的具体实现,模式与单元的关系等同于面向对象知识表示方法中类与实例的关系。知识单元的逻辑组成及行为结构表示如下:

```
FKU 框架知识单元(输入参数;输出参数)
{ 单元状态;
  缺省行为;
  根据特征属性的不同状态值,执行相应的行为;
}
RKU 规则单元名称(输入参数序列;输出参数)
{
  输入参数序列;该单元所需要的事实项和数据
  输出参数;该单元推理的结论
  RULE
  {产生式规则,如:A:- B,C,D }
}
```

例如:

```
FKU 棉花品种选择(适宜品种)
{ STRING 适宜品种;
  ENUM 茬口;//枚举类型
  茬口=(“早茬地”,“中茬地”);
  ENUM 土壤肥力;
  土壤肥力=(“高”,“中”,“低”)
  ENUM 品质要求;
  品质要求=(“能纺细支纱”,“能纺粗、中支纱”);
}
RKU 适宜品种(茬口,土壤肥力,品质要求;适宜品种)
{ //产生式组
  品种熟性(早熟):- 茬口(早茬地);
  适宜品种(酒棉一号):- 品种熟性(早熟),土壤肥力(高),品质要求(能纺细支纱);
  适宜品种(抗虫棉二号):- 品种熟性(早熟),土壤肥力(中),品质要求(能纺粗中支纱);
}
```

多级知识单元知识表示方法的特点:能综合集成多种知识表示方法,能够有效地表示规则型、运算型、描述型、多媒体型、不确定知识和模糊知识等多种知识。是一种逐渐分层的知识表示机制,符合人们思维习惯,易于表达多学科交叉的、较为复杂的领域知识。面向对象的思想技术,应用到领域知识的表达中,使得知识表示对自然实体的映射更为自然,知识表达的可理解性、透明度增大。

3.2 基于HornML的多级知识单元知识表示方法

标签定义如下:

```
<FKUNIT> 描述框架单元 FKU
<FKUnitID> 描述框架单元 FKU 的 ID 编号
<FKUnitName> 描述框架单元 FKU 的名称
<RKUNIT> 描述规则单元 RKU
<RKUnitID> 描述规则单元 RKU 的 ID 编号
<RKUnitName> 描述框架单元 RKU 的名称
<StringVar> 描述框架单元中的字符串变量
<EnumVar> 描述框架单元中的枚举型变量
<RealVar> 描述框架单元中的实型变量
```

另外,规则单元RKU中的产生式规则采用HornML中的描述方法。上面的例子可描述成如下形式:

```
<FKUNIT>
<FKUnitID>1</FKUnitID>
<FKUnitName>棉花品种选择</FKUnitName>
<StringVar>适宜品种</StringVar>
<EnumVar>茬口
  <value>早茬地 中茬地</value>
</EnumVar>
<EnumVar>土壤肥力
  <value>高 中 低</value>
</EnumVar>
<EnumVar>品质要求
  <value>能纺细支纱 能纺粗中支纱</value>
</EnumVar>
<RKUNIT> <! —规则单元描述-->
  <RKUnitID>2</RKUnitID>
  <RKUnitName>适宜品种</RKUnitName>
  <kb> <! —描述规则产生式组-->
```

```

<hn>
<relationship>
  <relator>品种熟性 </relator>
  <ind>早熟</ind>
</relationship>
<relationship>
  <relator>茬口 </relator>
  <ind>早茬地</ind>
</relationship>
</hn>
<hn>
<relationship>
  <relator>适宜品种 </relator>
  <ind>酒棉一号</ind>
</relationship>
<relationship>
  <relator>品种熟性 </relator>
  <ind>早熟</ind>
</relationship>
<relationship>
  <relator>土壤肥力 </relator>
  <ind>高</ind>
</relationship>
<relationship>
  <relator>品质要求</relator>
  <ind>能纺细支纱</ind>
</relationship>
</hn>
.....
</kb>
</RKUNIT>
</FKUNIT>

```

4 推理方法

由于 HornML 描述的是 Horn 子句, 因此利用 SLD 反驳原理, 即可方便地实现推理工作。以上面为例来说明推理过程。当选择: 茬口(早茬地), 土壤肥力(高), 品质要求(能纺细支纱)时, 将这些作为事实 fact 加入到知识库中。即加入以下事实:

```

<hn>
<relationship>
  <relator> 茬口 </relator>
  <ind>早茬地</ind>
</relationship>
</hn>
<hn>

```

```

<relationship>
  <relator>土壤肥力 </relator>
  <ind>高</ind>
</relationship>
</hn>
<hn>
<relationship>
  <relator>品质要求 </relator>
  <ind>能纺细支纱</ind>
</relationship>
</hn>
若此时询问: 适宜品种(X), 即推理条件为:
<relationship>
  <relator>适宜品种</relator>
  <var>X</var>
</relationship>
根据 SLD 反驳原理, 可得到推理结论为:
<ind X="酒棉一号">
  true
</ind>
即推理出的结果是: 适宜品种为酒棉一号。

```

5 总结

我们讨论了 Web 特性对传统人工智能知识表示的影响, 提出了基于 HornML 的多级知识单元表示方法。它利于对知识提供一种统一的存储和交换格式, 方便 Internet 上分布式知识的交换, 适合农业领域知识的特点, 有较强的表达能力且易于扩充。我们下一步的工作主要是在原来工作的基础上, 集成更多的知识表示方法。同时我们也注意到, 这种表示方法缺少对语义的描述, 今后我们的工作要在这方面加强。

(收稿日期: 2005 年 8 月)

参考文献

1. 陆汝钤. 人工智能[M]. 上册, 北京: 科学出版社, 1995: 1~4
2. 张维明. 语义信息模型及应用[M]. 北京: 电子工业出版社, 2002: 107~108
3. Harold Boley, Stefan Decker. Michael Sintek Tutorial on Knowledge Markup Techniques. <http://www.dfki.uni-kl.de/km/knowmark/know-mark-25-apr>
4. Harold Boley, DFKI GmbH. The Rule Markup Language: RDF-XML Data Model, XML Schema Hierarchy, and XSL Transformations. <http://www.ruleml.org>
5. 薛军, 王儒敬. 一种基于文本知识库的推理方法研究[J]. 计算机工程与应用, 2003; 39(21): 189~191

(上接 43 页)

- one processor[J]. Real-Time Systems, 1990; (2): 173~179
10. Baruah S K, Mok A K, Rosier L E. Preemptively scheduling hard-real-time sporadic tasks on one processor[C]. In: Proceedings of the 11th IEEE Real-Time Systems Symposium, 1990: 182~190
 11. 刘云生, 何冰, 冉龙波. 混合实时事务的延期单调速率调度算法及其可调度性分析[J]. 计算机学报, 2004; 27(3)
 12. K Jeffay, D L Stone. Accounting for Interrupt Handling Costs in Dynamic Priority Task Systems[C]. In: Proceedings of IEEE Real-Time

- Systems Symposium, 1993: 212~221
13. G C Buttazzo, G Lipari. Scheduling Analysis of Hybrid Real-Time Task Sets[C]. In: Proceedings of the 9th IEEE Euromicro Workshop on Real-Time Systems, Toledo, Spain, 1997: 200~206
 14. T P Baker. Stack-based Scheduling of Real-Time Processes[J]. The Journal of Real-Time Systems, 1991; 3(1): 76~100
 15. Spuri M, G Butazzo. Scheduling aperiodic tasks in dynamic priority systems[J]. Real-Time Systems Journal, 1996; (10): 179~210