

农业智能系统开发平台的知识表示与推理策略

魏圆圆^{1,2}, 王儒敬^{1,2}, 张英^{1,2}

(1. 中国科学院合肥智能机械研究所, 安徽 合肥 230031; 2. 中国科学技术大学 信息科学技术学院, 安徽 合肥 230026)

摘要: 针对农业智能系统的知识特点, 提出基于框架知识单元和求解知识单元的知识组织结构, 同时采用 XML 作为知识描述语言的知识表示方法, 在此基础上采用混合推理的推理策略, 并实现了不确定推理. 该知识表示方法和推理策略已经在工作中得到实际应用, 较以往知识表示方法, 在知识获取、知识的直观充分表达、知识扩充、知识模块化、知识共享等方面都取得了很好的效果.

关键词: 知识表示; 知识单元; XML; 推理

中图分类号: TP182 **文献标识码:** A **文章编号:** 1673-4785(2008)06-0523-06

A knowledge representation and inference strategy for a development platform of agricultural intelligence system

WEI Yuan-yuan^{1,2}, WANG Ru-jing^{1,2}, ZHANG Ying^{1,2}

(1. Institute of Intelligent Machines, Chinese Academy of Science, Hefei 230031, China; 2. School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China)

Abstract: In view of the knowledge characteristics of an agricultural intelligence system, a knowledge representation structure was formulated based on frame knowledge units and knowledge solving units. XML was chosen as its knowledge description language and a bidirectional reasoning strategy was employed to support uncertain reasoning. This knowledge representation mode and inference strategy was applied in our research work. Compared with other methods, it produced better results in areas such as knowledge acquisition, intuitionistic expression of knowledge, knowledge expansion, knowledge modularization and knowledge sharing.

Keywords: knowledge representation; knowledge unit; XML; inference

由于农业知识的复杂性、地域差异性和丰富性等特点, 如何直观充分地表示知识并进行准确高效的推理决策始终是农业智能系统开发平台所要解决的主要问题. 在建立某个特定任务的智能系统时, 首先要考虑的就是要选用合适的知识表示策略, 知识表示方法的合适与否不仅关系到知识库中知识的有效存储, 而且也直接影响着系统的知识推理效率和对新知识的获取能力.

农业智能系统中, 通常的数据、方法、模型、经验、案例等信息都是可能的知识源. 谓词逻辑、产生式系

统、语义网络、框架、过程表示等人工智能中常用的知识表示方法不能满足农业领域知识表示的需要^[1].

在 XML 广泛使用以前, 农业知识的管理多采用定制格式的文本管理方式^[2-3]或直接将知识存入数据库系统的管理办法^[4-7]. 定制格式的文本形式符合人们一般的阅读思维和语言习惯; 但因为结构化程度弱, 文本解析比较麻烦, 推理和知识检索效率较低, 同时不能实现跨平台, 限制了知识的共享. 数据库保存知识的优势在于利用成熟的数据库管理工具, 能够便捷稳定地管理、保存、提取知识; 但关系数据库的二维表结构, 使得在描述非结构化问题时存在弊端^[8]. 随着 XML 技术的成熟和普及, 也出现了一些新的相关的知识表示方法^[9-10], 在软硬件无关性、知识结构性、可扩展性等方面都体现出了优势,

收稿日期: 2008-04-29.

基金项目: 国家自然科学基金资助项目(60774096); 国家十一五科技支撑计划资助项目(2006BAD10A1410); 国家 863 计划资助项目(2006AA10Z237).

通信作者: 魏圆圆. Email: jsjwy@126.com.

但对复杂问题的知识描述能力还非常有限。

本文从知识组织结构和知识描述语言两个层次研究农业智能系统的知识表示,结合农业知识库的特点,提出基于框架知识单元和求解知识单元的知识表示模型,并以 XML 作为知识描述语言,通过语法规则的定义,将半结构化和非结构化的知识进行尽可能的结构化存储,形成 XML 知识库文档,推理机依据该知识库文档进行推理。平台采用 VC.NET 2003 作为开发工具,目前已投入正式使用,较以往开发平台越来越多地体现出在知识表达、知识扩充、知识模块化、知识共享等方面的优势。

1 农业智能系统知识库的知识组织结构

知识表示方法可以从以下几个方面来衡量^[11]:

1) 表示能力,具备将问题求解中的各类形式化知识完全表示出来的能力;2) 可理解性,便于理解和实现;3) 可操作性,有利于知识的利用,能使基于知识的推理有效地、符合逻辑地进行;4) 结构性,便于知识系统的维护、管理及扩充,并有利于推理的进行。

知识表示方式有两大基本类型:陈述性表示和过程性表示^[1,12]。这是知识理论中的经典问题。陈述性表示方式强调知识的静态,即描述事物的属性及其相互关系;过程性表示方式则强调知识的动态,即表示推理和搜索相关事实等运用知识的过程。与知识表示方式的这种分类吻合,本文提出的知识表示方法,引入框架知识单元(frame knowledge unit, FKU)和求解知识单元(solving knowledge unit, SKU)的概念,将陈述表示对应框架知识单元,过程表示对应于求解知识单元。

1.1 框架知识单元

描述问题逐层分解的框架体系树的知识单元,体现了求解问题所包含的知识内容和知识间的层次关系。将要求解的问题按照领域特征和专家知识结构进行层层分解与细化,形成一个树状框架结构,称为框架树。这种树状结构很直观地描述了整个智能系统知识结点之间的层次关系。以棉花栽培管理专家系统为例,开发平台构建的框架体系树片段如图1所示。

框架树中每个节点都是一个框架知识单元,与其关联的上层节点是父节点,所以框架树对应的框架知识单元是一个单元层层嵌套的结构。框架树过于复杂的时候,树的叶子节点还可以引出一个子框

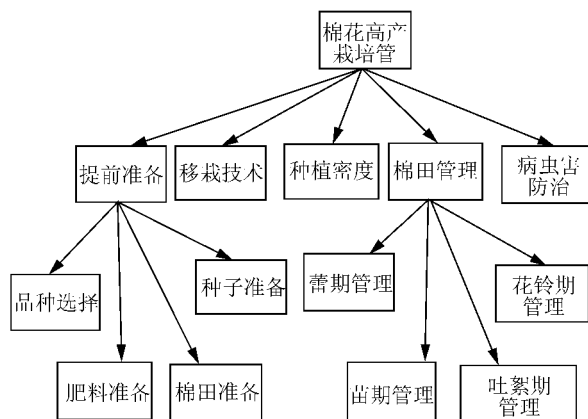


图1 框架树片段

Fig. 1 Segment of frame tree

架,该叶子节点就是子框架的根节点。否则叶子节点就为具体求解的求解目标,图1中,品种选择、肥料准备、蕾期管理等是要具体求解的目标。每个求解目标都有一个“求解类型”属性,指明求解目标的求解方式。简单的形式可以是直接显示描述问题解决内容的网页、播放视频等,复杂问题的求解形式可以通过求解知识单元求解,也可以直接调用外部构件进行求解,比如神经网络构件、数据挖掘构件等。

一个求解目标可以包含一个或多个求解资源,或者叫做求解内容。比如目标“种子准备”的求解类型是 webpage,推理机根据该属性,提取求解资源,即网页“zzzb.htm”,在界面中显示。目标“肥料准备”的求解类型是 sku,推理机根据这个属性,分别找到包含的3个求解资源对应的求解知识单元进行求解。

1.2 求解知识单元

求解知识单元指的是描述一个问题求解的过程性知识。通过框架树,描述了问题的层次结构以及求解目标,接下来求解目标如何实现,就要通过求解知识单元来完成。

类似于程序设计中完成一个基本功能的函数。求解知识单元包括变量声明部分(VARLIST)和执行部分(BODY)。变量声明部分定义问题求解过程中所需的变量,按照变量的用途、变量值的获取方式等属性,将变量分为一般变量(general)、sku变量和临时变量3种类型(type)。一般变量又包含3种数据类型(datatype):枚举型、实数型、字符串型。推理过程中用到且其值未知的情况下,通过人机交互界面提问用户。对sku变量,其变量值通过调用其他sku得到。临时变量是在计算过程中保存临时结果的变量。所以这里,sku变量和临时变量是求解知识单元的局部变量,一般变量的值的作用范围是其所在的求解目标。执行部分描述了问题求解过程所需

的知识,表示为有序的语句(SENTENCE)序列.语句类型包括:赋值语句、if...then...else...语句、输出语句、数据库访问语句等.输出语句又可以输出字符串、变量值、网页、网页超链接等形式.开发平台中包含的这些变量类型和语句类型基本满足了农业智能系统对知识表示的需求,通过傻瓜化人机交互界面从领域专家头脑中获取规则、计算模型等各种类型知识.

求解知识单元之间既相互独立,又相互联系.求解过程通过知识单元的形式来描述,很好地体现了知识的模块化、独立性,可维护性很高.求解知识单元之间通过相互调用联系起来,很好地实现了知识的共享.

从上面的描述,很容易看出,一个完整的知识库由一个递归嵌套的框架知识单元和多个并列的求解规则单元组成.

2 基于XML描述的知识表示方法

2.1 XML简介

扩展标记语言(extensible markup language)XML是标准通用标记语言(standard generalized markup language, SGML)的一个子集,继承了SGML的优点,但比SGML简单,用以创建可相互转换的结构化文本文档和数据文档.XML是W3C确定的Web上的标准数据格式,和HTML不同,它是一种元标记语言(meta-markup language),也就是说它没有一套能够适用于各个领域所有用户的固守的标签和元素;相反,它允许开发者根据自己的需要定义自己的元素,极大地强化了保存信息和处理信息的能力.XML的标记可以用来描述你要表达的概念,而属性则用来控制他们的结构^[13].它的主要特点有:

1) XML文档是一种结构化的文档,可用树的形式表示出来.

2) 可利用DTD(document type declaration)或Schema(模式)管理一致性问题.DTD或Schema为XML文档提供了一个语法规范,如果XML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.DTD主要用于文档型文档,Schema主要用数据型文档.

3) 具有很好的扩展性,可定义自己的元素和属性.

4) 通过XML可从关系数据库管理系统中提取数据到结构化文档.它还被设计成可对各种数据对象进行操作.

5) 在一个设计良好的XML应用中,XML标记不涉及文档如何显示,只表示文档的结构.

从数据描述语言的角度看,XML是灵活的、可扩展的,有良好的结构和约束;从数据处理的角度看,它足够简单且易于阅读,同时易于被应用程序处理^[14-15].XML的这些优越性,也使得近年来基于XML的知识表示方法成为知识表示领域研究的热点.

2.2 知识库的XML描述

一个XML知识库文件包含<FKU>元素和<SKULIST>元素两大部分.<FKU>元素对应一个递归嵌套的框架知识单元,<SKULIST>元素对应多个求解知识单元组成的知识单元序列.每个求解知识单元对应一个<SKU>元素,作为SKULIST的子元素.用浏览器打开“棉花栽培管理专家系统”的XML知识库文档,并将内容收起后的形式如下:

```
<? xml version = "1.0" encoding = "GB2312" ? >
- <EXPERTSYSTEM esname = "棉花高产栽培管理专家系统">
+ <FKU name = "棉花高产栽培管理专家系统"
  hastarget = "false">
```

```
+ <SKULIST>
```

```
</EXPERTSYSTEM>
```

取FKU片断如下:

```
<EXPERTSYSTEM esname = "棉花高产栽培管理专家系统">
  <FKU name = "棉花高产栽培管理专家系统"
    hastarget = "false">
    <FKU name = "播前准备" hastarget = "false">
      <FKU name = "种子准备" hastarget = "true">
        <TARGET targettype = "webpage">
          <TARGETRES> zzzb. htm
        </TARGETRES>
      </TARGET>
    </FKU>
    <FKU name = "肥料准备" hastarget = "true">
      <TARGET targettype = "sku">
        <TARGETRES> 氮肥用量
      </TARGETRES>
        <TARGETRES> 磷肥用量
      </TARGETRES>
        <TARGETRES> 钾肥用量
      </TARGETRES>
      </TARGET>
    </FKU>
    <FKU name = "棉田准备" hastarget = "true">
      <TARGET targettype = "webpage">
        <TARGETRES> mtzb. htm
```

```

</TARGETRES >
</TARGET >
</FKU >
</FKU >
<FKU name = “移栽技术” hastarget = “true” >
.....

```

取 SKU 片断如下:

```

<SKULIST >
<SKU name = “氮肥用量” resulttype = “real”
    unit = “千克” cf = “1.00” >
<VARLIST >
<VAR name = “土壤含氮量” type = “SKU”
    res = “土壤含 N(K) 量” / >
<VAR name = “生长需氮量” type = “SKU”
    res = “生长需氮量” / >
<VAR name = “氮肥种类” type = “general”
    datatype = “enum” >
<VALUE > 尿素 </VALUE >
<VALUE > 硝酸铵 </VALUE >
<VALUE > 硫酸铵 </VALUE >
<VALUE > 碳酸氢铵 </VALUE >
</VAR >
<VAR name = “x” type = “midvar”
    datatype = “real” unit = “” / >
</VARLIST >
<BODY >
<SENTENCE type = “exp” >
<LVALUE > x </LVALUE >
<RVALUE > 生长需氮量 - 土壤含氮量 *
    0.45 </RVALUE >
</SENTENCE >
<SENTENCE type = “rule” >
<IF >
<PREM varname = “氮肥种类”
    rel = “=”
    opval = “尿素” / >
</IF >
<THEN >
<SENTENCE type = “exp” >
<LVALUE > 氮肥用量 </LVALUE >
<RVALUE > x/0.4 </RVALUE >
</SENTENCE >
</THEN >
</SENTENCE >

```

为说明上述知识表示的优点,取作者在以往开

发平台中常采用的基于文本的知识库片断如下:

```

FKU 棉花高产栽培管理专家系统
{
棉花高产栽培管理专家系统 = (播前准备,
移栽技术,种植密度,……);
播前准备 = (品种选择,肥料准备,种子准备,
棉田准备);
病害防治 = (苗期病害,铃期病害,枯黄萎病);
.....
}
FKU 肥料准备(氮肥用量,磷肥用量,钾肥用量)
{
肥料准备 = (氮肥用量,磷肥用量,钾肥用量,
生长需氮量,土壤含氮量,……);
REAL 氮肥用量(MIN,0,MAX,10000,UNIT,单位);
REAL 磷肥用量(MIN,0,MAX,10000,UNIT,单位);
REAL 钾肥用量(MIN,0,MAX,10000,UNIT,单位);
REAL 生长需氮量(MIN,0,MAX,10000,UNIT,单位);
REAL 土壤含氮量(MIN,0,MAX,10000,UNIT,单位);
ENUM 氮肥种类;
氮肥种类 = {“尿素”,“硝酸铵”,“硫酸铵”,
“碳酸氢铵”};
.....
}
RKU 氮肥用量(生长需氮量,土壤含氮量,
    氮肥种类;氮肥用量)(CF=1)
{
IF 氮肥种类 == “尿素” THEN 氮肥用量 =
(生长需氮量 - 土壤含氮量 * 0.45)/0.4;
IF 氮肥种类 == “硝酸铵” THEN 氮肥用量 =
(生长需氮量 - 土壤含氮量 * 0.45)/0.34;
.....
}

```

可以看出,纯文本的知识库形式篇幅较小,更接近自然语言;但也正因为缺少了标签和属性的标识和规范,使得文本知识库必须要经过严格的语法检查,推理机在推理过程中对文本的解析工作相当繁琐.而 XML 表示的知识库尽可能使知识进行结构化描述,加上各种开发语言都有较成熟的 XML 文档解析器,所以推理机的实现要相对容易,而且本文研究的知识表示较以往基于文本的知识库在描述能力上要强很多.

3 推理策略

知识库和推理机是专家系统的主要部分^[16].推

理机是整个专家系统的控制中心,整个系统靠它进行运转。一般推理有两种方式:一种是由下而上,即从用户提出因素向目标节点一步一步推导;另一种是从上向下,即从目标节点到叶节点,反过来一步一步推导。常将前者称作正向推理,后者称作反向推理。正向推理是从事实向目标的推理,又叫事实驱动;反向推理是从目标向事实的推理,也叫目标驱动。在复杂的推理网络中,还有一种正反向推理或双向混合推理方式。知识表示策略与推理策略两者是相互配合的,不同的知识表示策略需有相应的推理机制。本文采用混合推理模式,同时支持不确定推理。

3.1 混合推理模式

基于知识库采用的知识组织结构,首先,根据框架树引导用户选择求解目标,这个过程是正向推理;然后,是反向推理过程,根据求解目标包含的资源名称找到对应的求解知识单元,顺序执行单元中的语句序列,搜索事实知识或者通过人机交互界面向用户提问。

3.2 推理的不确定性

推理的不确定性体现在求解知识单元的置信度(CF值)上,在求解资源对应多个SKU的时候,就是多路径推理,即问题有多种解决方式。这时就根据CF值优先求解CF值高的SKU,若该路径中有必要的事实知识搜索不到且用户不能给出,则此路径无法求解;取CF值次之的SKU进行求解,直到求解完成为止。

3.3 推理机运行过程

推理依据XML知识库文档进行,XML语言的良好结构化特性,大大方便了推理机对知识内容的解析。整个推理步骤如下:

1) 根据框架树引导选择求解目标。

2) 找到求解目标对应的SKU,建立求解队列,并按照CF值由大到小进行排序。

3) 优先求解CF值高的SKU,在某一个SKU的求解过程中,先依次求解该知识单元变量列表中调用的其他SKU,赋值给对应的sku变量,sku变量求解完毕后,推理机顺序执行求解语句。如果求解过程中遇到未知变量,就通过人机交互界面向用户提问,用户回答后推理继续进行,直到执行部分运行结束,得到求解结果,推理完毕。

4) 如果步骤中用户针对提问不能给出答案,则上述SKU无法求解,从求解队列中选择CF值次之的SKU进行求解,其过程与步骤3)相同。直到有SKU求解完成,推理完毕。若求解队列中所有SKU都无法求解,则求解目标求解失败,推理完毕。

3.4 推理结果

仍以“棉花栽培管理专家系统”为例,推理机基于XML知识库文档,对“肥料准备”的求解结果如图2所示。



图2 推理机运行结果

Fig. 2 Inferring results

4 结束语

本文针对农业智能系统知识体系的特点,提出了基于框架知识单元和求解知识单元的知识组织结构,同时采用XML作为知识描述语言的知识表示方法。在此基础上采用混合推理的推理策略,通过对求解知识单元赋以CF值,实现了不确定推理。该知识表示方法和与之对应的推理机制已经在作者的新的智能系统开发平台中得到实际应用,较以往知识表示方法,具有以下特点:1) 知识库由框架知识单元和求解知识单元组成的思路,与知识表示分为陈述表示和过程表示这一分类法吻合;2) 框架知识单元的树状结构和XML文档的树状结构对应,采用XML描述知识,两者优势互补;3) XML描述的知识库在形式上以元素为基本单元,一个<SKU>元素即对应一个求解知识单元,这就为基于知识库的推理提高了检索的效率;4) 推理采用混合推理方法,目标选定后只考虑对假设目标有用的知识,缩小了搜索空间,推理效率高;5) 支持多路径推理,同时也解决了多路径求解问题;6) 针对农业智能系统开发平台,知识库独立于推理机,有利于知识的更新与扩充,同时生成的XML知识库文档,适用于不同软硬件环境下的推理和知识交换。

参考文献:

- [1] 王珏,袁小红,石纯一,等. 关于知识表示的讨论[J]. 计算机学报,1995,18(3): 212-224.
WANG Jue, YUAN Xiaohong, SHI Chunyi, et al. Discussions on knowledge representation[J]. Chinese Journal of

- Computers, 1995, 18(3): 212-224.
- [2] 熊范纶, 乔克智, 胡海赢. 雄风专家系统开发工具[M]. 北京:清华大学出版社, 1999.
- [3] 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.
XUE Jun, WANG Rujing. A method of reasoning based on text knowledge-base[J]. Computer Engineering and Applications, 2003, 39(21): 189-191.
- [4] 席 磊, 张 丽, 张 慧, 等. 农业专家系统中知识表示技术的研究[J]. 河南师范大学学报: 自然科学版, 2006, 34(3): 43-47.
XI Lei, ZHANG Li, ZHANG Hui, et al. Knowledge representation research for agriculture expert system[J]. Journal of Henan Normal University: Natural Science, 2006, 34(3): 43-47.
- [5] 涂运华, 王东辉, 赵春江. 基于 Windows CE 的 HPC/PDA 农业专家系统开发平台的研究与开发[J]. 高技术通讯, 2000, 10: 28-31.
TU Yunhua, WANG Donghui, ZHAO Chunjiang. The research and development of a development platform on HPC/PDA for agricultural expert system based on Windows CE [J]. High Technology Letters, 2000, 10: 28-31.
- [6] 张文学, 朱乃立. 知识表示在牡丹栽培技术专家系统中的应用[J]. 计算机工程, 2005, 31(17): 210-212.
ZHANG Wenxue, ZHU Naili. Application of knowledge representation in peony growing technology expert system [J]. Computer Engineering, 2005, 31(17): 210-212.
- [7] YAN Hongsen. A new complicated-knowledge representation approach based on knowledge meshes[J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(1): 47-62.
- [8] 周桂红, 郑 磊, 黄丽华, 等. 农业专家系统生成工具的设计与实现[J]. 农业工程学报, 1999, 15(3): 53-59.
ZHOU Guihong, ZHENG Lei, HUANG Lihua, et al. Design and implementation of agricultural expert system tool [J]. Transactions of the Chinese Society of Agricultural Engineering, 1999, 15(3): 53-59.
- [9] 黄 海, 王儒敬, 黄 河. 一种基于 HornML 的 Web 知识表示方法[J]. 计算机工程与应用, 2006, 42(1): 53-55.
HUANG Hai, WANG Rujing, HUANG He. A HornML-based knowledge representation for web knowledge [J]. Computer Engineering and Applications, 2006, 42(1): 53-55.
- [10] 徐 伟, 王儒敬, 杨化峰. 基于 RuleML 的多级知识单元知识表示方法[J]. 计算机工程与应用, 2005, 41(1): 174-177.
XU Wei, WANG Rujing, YANG Huafeng. RuleML-based knowledge representation of multi-level knowledge units [J]. Computer Engineering and Applications, 2005, 41(1): 174-177.
- [11] DAVIS R, SHROBE H, SZOLVITS P. What is a knowledge representation[J]. AI Magazine, 1993, 14(1): 17-33.
- [12] 石纯一. 人工智能原理[M]. 1 版. 北京:清华大学出版社, 1993.
- [13] 丘光华, 张文敏. XML 编程实例教程[M]. 北京:科学出版社, 2004.
- [14] FARSI R. XML[J]. Informatik - Spectrum, 1999, 22(6): 436-438.
- [15] 孙 伟, 刘大昕. 一种 XML 代数及其查询优化方法[J]. 哈尔滨工程大学学报, 2007, 28(8): 899-904.
SUN Wei, LIU Daxin. An XML algebra and query Optimization based on algebra [J]. Journal of Harbin Engineering University, 2007, 28(8): 899-904.
- [16] NILSSON N J. Artificial intelligence: a new synthesis[M]. San Francisco: Morgan Kaufmann Publishers, 1998.

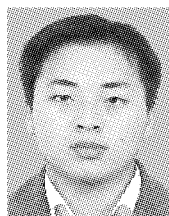
作者简介:



魏圆圆, 女, 1980 年生, 助理研究员, 博士研究生, 主要研究方向为知识表示、智能决策. 发表学术论文 4 篇.



王儒敬, 男, 1964 年生, 研究员, 博士生导师, 主要研究方向为知识工程. 参与或主持国家自然科学基金、863 项目、国家支撑等课题 10 余项. 发表学术论文 20 余篇.



张 英, 男, 1983 年生, 硕士研究生, 主要研究方向为知识获取、知识表示. 发表学术论文 1 篇.