

基于本体的几何学知识获取及知识表示

钟秀琴¹⁾ 符红光^{1),2)} 余 莉¹⁾ 黄 斌³⁾

¹⁾(电子科技大学计算机科学与工程学院 成都 610054)

²⁾(中国科学院成都计算机应用研究所 成都 610041)

³⁾(成都信息工程学院 成都 610225)

摘 要 文中研究了基于本体的几何学知识获取,并运用此方法获取几何学领域中的类、属性、关系及公理等;同时,研究了基于本体的知识表示,并将几何学知识转化为该领域本体,用 RDF/OWL 的形式进行描述.在此基础上构建了一个可共享、可重用、可扩展的几何学本体,形成了一套较完整的几何学知识获取和知识表示体系,将其应用于建立几何学知识库,可大大提高知识搜索和知识推理的效率,并为几何学专家系统、几何学信息检索、几何学教育系统、自然语言理解等领域提供智能基础.

关键词 几何学;本体;知识获取;知识表示;RDF(S)

中图法分类号 TP18

DOI号: 10.3724/SP.J.1016.2010.00167

Geometry Knowledge Acquisition and Representation on Ontology

ZHONG Xiu-Qin²⁾ FU Hong-Guang^{1),2)} SHE Li¹⁾ HUANG Bin³⁾

¹⁾(School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054)

²⁾(Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu 610041)

³⁾(Chengdu University of Information Technology, Chengdu 610225)

Abstract Knowledge acquisition is the bottleneck of knowledge engineering, and knowledge representation is the key technology in knowledge engineering. Ontology is a powerful implement for applications in many fields, which is so convenient that it is the promising method to resolve knowledge acquisition and representation. It'll mainly depends on semantic Web unless it has great breakthrough in Chinese natural language processing. In order to construct ontology for formalizing the concepts precisely in the very domain, and define the relationships between concepts in machine processing clearly, the authors research knowledge acquisition on domain ontology to acquire the category, the instance, the attribute, the relation and the axiom such as axioms between categories, axioms between attributes, axioms of range, axioms of semantic types and so on. Meanwhile, we research knowledge representation based on ontology to convert geometry knowledge into domain ontology represented in RDF/OWL. Furthermore, the authors establish a sharable, reusable and extensible geometry ontology and induce several geometry knowledge acquisition and representation modes, which can improve the efficiency of knowledge searching and reasoning in constructing geometry knowledge base, and also it can be the foundation of geometry expert systems, geometry information search systems, geometry educational systems and so on.

Keywords geometry; ontology; knowledge acquisition; knowledge representation; RDF(S)

收稿日期:2007-08-24;最终修改稿收到日期:2008-12-12. 本课题得到国家“九七三”重点基础研究发展规划项目基金(2004CB318003)资助. 钟秀琴,女,1976年生,博士,讲师,主要研究方向为自动推理及本体论. E-mail: zhongxiuqin@hotmail.com. 符红光,男,1965年生,教授,博士生导师,主要研究领域为计算机代数及知识科学. 余 莉,女,1978年生,博士,讲师,主要研究方向为数据挖掘与知识工程. 黄 斌,男,1965年生,硕士,副教授,主要研究方向为组合数学与图论.

1 引 言

知识获取是指从知识源获得知识来构建知识库,是知识工程一个公认的瓶颈问题.知识获取方法主要有3种:(1)知识工程师从领域专家那里获取知识,然后以正确的形式储存到知识库里;(2)领域专家通过知识编辑器将自己的知识和经验存入知识库,如IBM的TextMiner^[1]、OntoPrise的OntoEdit^[2]、StanFord的Protégé^[3]等;(3)通过知识学习器从数据库中自动获取知识,在这种方法中,知识编辑器提供一个具有一定格式的对话界面,领域专家按照对话要求输入知识.

知识表示是知识工程的一个核心问题.知识的表示方式主要有逻辑的知识表示、基于关系的知识表示、面向对象的知识表示、基于规则的知识表示、语义网络、基于模型的知识表示和基于本体的知识表示等.知识表示形式将直接影响到知识库系统的性能.基于语义的知识表示是当前研究的热点,RDF/OWL(Resource Description Frame/Web Ontology Language)是语义万维网技术知识表示的标准,是目前公认的开放网络环境下知识表示的规范.

在中文自然语言理解没有取得重大突破之前,主要采用语义Web的思想,通过构建本体来精确规范某个领域的概念,明确定义机器可以处理的概念以及概念之间的关系.本体可在不同的建模方法、范式、语言和软件工具之间进行翻译和映射;比数据库表达的知识丰富得多;是领域内重要实体、属性、过程及其相互关系形式化描述的基础;可以为知识库的构建提供一个基本的结构,从而去表达现实世界中浩如烟海的知识 and 常识;可以将对象知识的概念和相互间的关系进行较为精确的定义,大大提高知识搜索、知识重用、知识共享的效率.其最大贡献在于它可以将某个或多个特定领域的概念和术语规范化,为其在该领域或领域之间的实际应用提供便利.

领域本体(Domain-Specific Ontologies)指的是通过定义类(categories)、实例(instances)、属性(attributes)、关系(relations)、公理(axioms)等元素^[4],刻画领域中的类和实例及其之间的层次关系,对领域知识进行归纳和抽象.目前为止,许多领域都出现了大量的本体,如医学本体UMLS^[5]、中医药本体TCMLS^[6]等;已经实现的本体主要有:CYC^[7]、Enterprise^[8]、SENSUS^[9]、NKI^[10]等;国内陆汝钊院士等研制的常识知识系统,中国科学院计算技术研究

所曹存根研究员主持的大规模知识系统、医学知识库^[11],以及北京大学金芝教授研究的基于本体的软件需求获取方法等.

总之,本体论已成为知识工程中的一种常用工具,适用于基于本体的学科知识的知识表示,可用以分析一个论域中的类(categories)、关系(relations)以及它们所满足的限制(constraints),从而解决知识获取这个知识工程的瓶颈问题.

目前还没有一个大规模的、可共享的、可重用的、可扩展的几何学本体,因此建立一个良好的几何本体具有重要的意义.本文基于形式本体的理论建立了几何学形式本体及其知识表示,用TBC(Top-Braid Composer)本体构建工具来辅助构建本体,以此为基础进行基于文本的几何学知识获取和知识分析,功能模块已在软件GO(Geometry Ontology)中得以实现.它将为几何学专家系统、几何学信息检索、几何学教育系统、自然语言理解等领域提供智能基础.

2 几何学本体体系概述

2.1 几何学本体论

几何学本体论从知识工程的角度来说,指的是一个几何学本体结构以及对几何学概念的本体知识描述.在这里本体知识既可以在概念层次上进行描述,也可以在语言层次上进行描述.它包含了概念的所有相关知识,且知识必须与几何学领域相符合,又要完整地 and 全面地认识概念域,对隐式的或不确定的内容进行形式刻画.

在本体设计中,基本元素是类(categories),表示专业领域中的归属集体,每一个类中都有相应的个体(individuals)及实例(instances).类间的基本关系是继承(inheritances)和实现(implements),为了描述类及类的个体及实例,我们使用属性(attribute)及关系(relationship)的描述符,并对属性及关系的值进行限制.例1是关于类“平行四边形”的描述(要说明的是:本文所有关于本体的类描述都只是部分引用而非全部引用).

例 1.

类:平行四边形 继承 四边形

英文名:parallagram

属性:对边相等;对边平行;对角相等;同旁内角互补;对角线相互平分

数据类型:字符数组

注释: 两组对边分别平行的四边形 或 一组对边平行且相等的四边形

2.2 几何学本体结构

我们从知识工程的角度出发, 在参考众多分类法的基础上形成了几何学本体体系. 在设计结构时, 既考虑到几何学专业的通用性, 同时也考虑到知识库处理的方便性、可共享性、扩展性以及知识库知识的动态增量和再提取, 因此设计了如图 1 所示的几何学本体结构(注: 立体几何和解析几何中均包含有平面几何中的所有对象和属性).

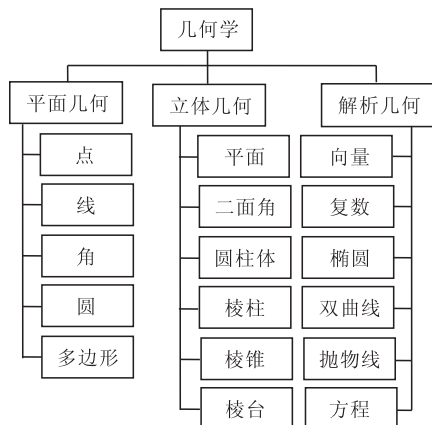


图 1 几何学本体结构

3 几何学本体设计及几何学知识获取

3.1 几何学知识表示形式

仔细分析几何知识源(几何学辞典、教科书、试卷、数学辞海等), 其基本的知识对象主要概括如下:

- (1) 几何符号: //、 \perp 、 $\triangle ABC$ 等;
- (2) 概念的定义知识, 几何知识最主要的描述对象, 如点、平面、垂直等;
- (3) 断言知识, 包括公理、引理、定理、猜想等. 如勾股定理是指在直角三角形中, 两条直角边的平方和等于斜边的平方, 即 $a^2 + b^2 = c^2$;
- (4) 几何命题;
- (5) 断言的证明知识: 几何定理、引理证明方法、使用技巧、涉及知识点、证明过程等.

这些基本的几何知识对象之间存在着层次依赖关系. 首先, 概念的定义是几何知识描述的最基本单位, 每个概念有相应的断言. 几何知识的一般表达方式为: 一个新概念的引入通常是以先前定义过的概念为基础; 一个定理的阐述依赖于定义过的概念; 一个几何实例是与一个概念或断言相关的; 定理的证明知识是给出相应的定理的证明过程、方法等.

通过对几何知识描述对象进行分类, 构建的几何学科知识本体主要包括:

(1) 概念定义. 几何知识描述的主要对象, 包括概念的基本知识(名称、简称、符号表示等)、参数及参数类型、各种定义方式(形式定义、非形式定义、等价定义等)、概念之间的关系表示等.

(2) 断言知识. 以几何概念定义为基础, 给出概念的性质描述、概念与概念之间的深层关系表达、断言与断言之间的关系表达, 通过“关于”建立与相应概念的联系;

(3) 实例. 概念或断言的实例, 以适当的方式有效地继承概念或断言的知识;

(4) 证明过程. 给出断言的推导方式和具体步骤;

(5) 几何方法. 给出各种证明方法、解题方法等;

(6) 其它. 关于几何的历史、人文等背景知识.

RDF 是一种声明语言, 它使用明确定义的尤其是 RDFS 定义的概念和特性来对资源进行描述. 从这种意义上来说, RDF 是一种资源内容语义的描述框架, 通过这个描述框架, 可以将概念体系中的概念和现实世界知识联系起来, 实现机器可理解的数据语义. RDF 数据模型主要包含下面的 3 个对象类型:

(1) Resources. 资源可以是整个网页、网页的一部分、页面的全部集合等不同粒度的可以用 URI (Unified Resource Identification) 来指称的对象;

(2) Properties. 特性可以用来描述资源特定的方面、特征、属性或某种关系;

(3) Statements. 一个特定的资源和特性名称加上该特性的值一起构成了一个 RDF 声明.

一个声明中的这 3 个部分分别称为: 主体(subject)、谓词(predicate) 和对象(object). 以基于 GO 的页面资源描述为例: 为了表达页面: <http://127.0.0.1/Triangle.html> 是有关三角形的信息, 可以采用图 2 所示的图形化描述方法:

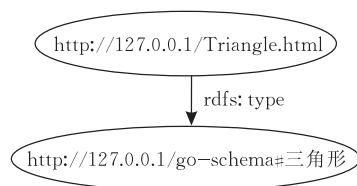


图 2 资源描述的图形表示

知识本体的构造经历了有向图模型、RDFS 描述、XML 序列化等 3 个阶段. 知识本体驱动的应用

程序使用知识本体的过程正好与此相反,首先需要经历 RDFS 解释,然后向应用程序提供应用编程接口.其中 RDF 解释包括了 XML 解析和 RDFS 解析两个过程. RDFS 是 W3C 所提供的一种简单的知识本体表示语言,它可以定义一些简单的概念(类)和关系(特性). RDFS 所使用的主要构造子(constructor)如下:

(1) 核心类. 包括 `rdfs:Resource`、`rdf:Property`

以及 `rdfs:Class`;

(2) 核心特性. 包括 `rdfs:type`、`rdfs:subClassOf` 和 `rdfs:subPropertyOf`;

(3) 核心约束. 包括 `rdfs:ConstraintResource`、`rdfs:ConstraintPropety`、`rdfs:range` 和 `rdfs:domain`.

采用 RDFS 规范所提供的构造子,可以对领域概念体系中的概念之间的关系进行描述,生成特定领域的知识本体. 一个平行四边形本体描述如图 3 所示.

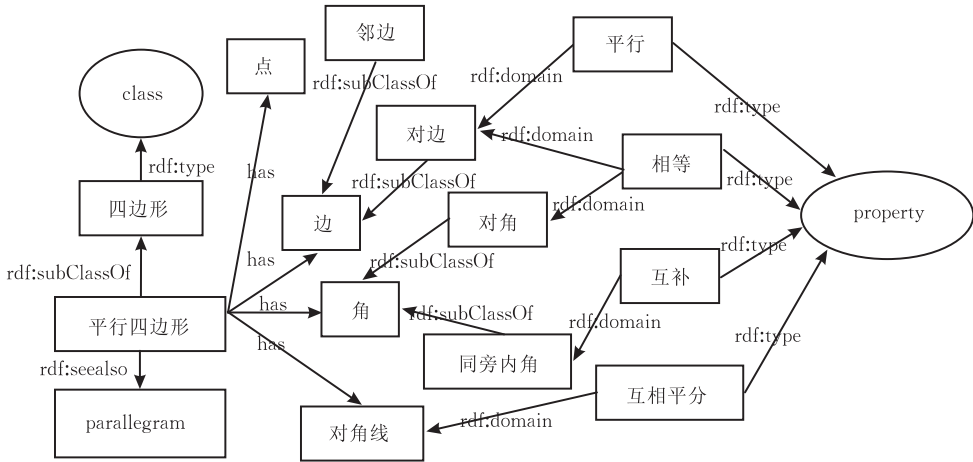


图 3 平行四边形本体描述

3.2 几何学本体的属性及关系

不失一般性,下面以平面几何为例来进行说明.

3.2.1 几何学本体中的属性

几何学知识中,核心的知识就是关于点的知识.点是一个最基本的几何学概念,根据几何学学科的范畴,可分为定点属性类、自由点属性类、半自由点属性类和约束点属性类.如图 4 所示.

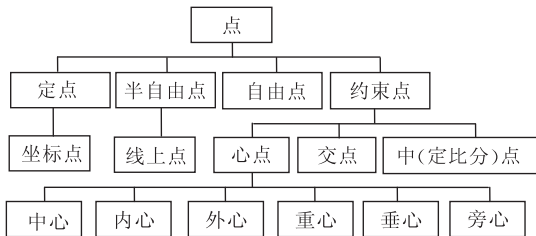


图 4 点属性类分类树

图 4 为平面上点属性类的分类结构,提取出了描述点类时所使用到的属性及关系集.由于属性之间固有的层次关系、点的属性及关系本身也可形成一个本体体系,称为点属性类.这个本体不含有任何实例,只能由点类来实现它,在这个属性本体中反映的类别知识也是类公理的一部分.

3.2.2 几何学本体中的关系

几何学本体中定义了一些关系,这些关系将几

何学中的概念与概念连接在一起,具有很重要的作用.

表 1 几何本体中的若干典型关系

关系	注释
平行	表示一条直线与另一条直线平行
垂直	表示一条直线与另一条直线垂直
垂直平分	表示一条直线过另一线段的中点且与该线段垂直
外接	表示一个多边形与另一个圆外接
内切	表示一个多边形与另一个圆内切

3.3 几何学本体设计方法及知识获取

我们在设计几何学本体时提出了以下原则:

- (1) 具体类的下面不应该有抽象类;
- (2) 类的整体分布应该比较均匀;
- (3) 类的继承关系不能有循环,可以有重复继承,但不能有冗余继承;
- (4) 类的命名唯一;
- (5) 专业知识的描述词汇专业化.

具体设计本体中的类时,我们采用的是从文本中直接获取知识的方法.对几何学知识源要求如下:知识可信度高(如标准化的知识源)、知识覆盖面要广、知识不能陈旧过时等.除了分类结构,描述设计几何学领域的属性和关系时,原则上尽量采用几何学术语,这样在从形式化还原到自然语言时,我们可

以更自然地以几何学语言的方式进行还原。

由于在类中所涉及到的属性及关系较多, 会形成一个层次关系. 这种层次关系反映了属性的类别知识, 并随之成为几何学公理的一部分. 我们采用属性本体结构以及聚类属性的方式来表示这种属性关系间的分类关系.

4 几何学公理的表达和获取

几何学公理具有很高的抽象性, 我们可以对其进行具体分析.

4.1 公理表示方法

几何学知识的分析建立在几何学公理上面, 几何学公理使用一阶逻辑来表示公理. 使用“ $X: \langle \text{类} \rangle$ ”表示 X 是一个类变量, X 的取值是类中的任意一个实例; 用“ $X \cdot S$ ”表示 X 的某个属性或关系的取值. $S(X, Y)$ 表示 X 的在属性或关系 S 的取值为 Y . 同时, 我们定义了一些函数和谓词来表示公理, 如表 2 所示.

表 2 部分函数和谓词表

谓词或函数	含义
平行(A, B)	谓词: 直线 A 平行于直线 B
垂直(A, B)	谓词: 直线 A 垂直于直线 B
外接(A, B)	谓词: 圆 B 外接于多边形 A
内切(A, B)	谓词: 圆 B 内切于多边形 A
Concatenate(x, y)	函数: 将 y 加到 x 的后面, 返回增加后的 x 串

几何学公理分为类间公理和类内公理、类间公理约束一个类和另一个类以及一个类的属性或关系和另一个类的属性或关系. 类内公理约束类中的属性或关系以及类中属性或关系与属性或关系之间的关系. 下面我们对类间公理和类内公理的获取分别描述.

4.2 公理获取方法

4.2.1 类间公理获取

类间公理分为类与类公理和类属性或关系与类属性或关系间公理. 我们在获取这一类公理的方法是对每个类都从几何学领域的涵义上, 找出该类与其它类的关系, 再把这种关系用一阶逻辑的方式表示.

(1) 类与类公理

其用来约束一个类与另一个类的关系. 为一个类与另一个类建立联系提供描述. 表 3 是这一类公理的实例.

表 3 类与类公理

公理	实例
公理 1	所有 X : 三角形, 存在 Y : 圆[是外接圆(Y, X)]
公理 2	所有 X : 角, 存在 Y : 线[是角平分线(Y, X)]

(2) 类属性(或关系)与类属性(或关系)间公理

其用于约束一个类的属性或关系及另一个类的属性或关系. 为多个类之间的属性或关系建立联系. 我们在获取这一类公理时, 将一个类与另一个类的同名属性或关系或分类层次上的上下位进行比较, 从几何学意义上得出它们之间的关系. 表 4 是这一类公理的一个实例.

表 4 类属性或关系与类属性或关系公理

公理	实例
公理 1	所有 X : 等分线, 所有 Y : 线段[是等分线(Y, X)]

4.2.2 类内公理获取

几何学公理中的类内公理包括类中关于属性几何学涵义的公理以及类中属性之间关系的公理, 是建立在几何学本体之上的对本体中的属性(或关系)及属性(或关系)间关系的一种约束. 我们在获取这一类公理时, 采用对每一个属性或关系进行分析, 找到表示属性或关系含义和约束的公理. 我们把类内公理分为值域公理、语义类型公理、属性关系公理、属性关系内涵公理.

(1) 值域公理

其用以对属性或关系值进行限定, 将属性或关系的取值限定在一个合理的范围内, 可以有助于进行知识的错误检查, 以发现一些常识上的矛盾. 如表 5 所示的公理 2 就是对三角形的面积的取值做出限定.

表 5 值域公理

公理	实例
公理 1	所有 X : 第一象限中的点[大于(X . 横坐标, 0) \wedge 大于(X . 纵坐标, 0)]
公理 2	所有 X : 三角形[大于(X . 面积, 0)]

(2) 语义类型公理

其用于对属性或关系值的语义类型进行说明, 实质上也是对属性或关系的意义的一种解释. 这有助于知识获取时对一个概念进行语义分析. 如表 6 所示的公理 1 表明: 圆的相切圆一定是圆. 如前面的值域公理一样, 如果属性或关系值是可枚举的, 我们也可以值域来固定语义类型, 而不用再定义语义类型. 需要说明的是, 在语义类型公理的 $isa(X, Y)$, 其中 X 可能是一个概念集合.

表 6 语义类型公理

公理	实例
公理 1	所有 X : 圆[相切圆(X, Y) $\rightarrow isa(Y, 圆)$]
公理 2	所有 X : 三角形[外接圆(X, Y) $\rightarrow isa(Y, 圆)$]

(3) 属性关系公理

其用于对属性与属性之间、关系与关系之间的元关系进行限定, 以便于在知识分析时进行诸如知识查错纠错、知识推理. 如表 7 所示的公理 1 表明了三角形任意两边之和与第三边之间的关系.

表 7 属性关系公理

公理	实例
公理 1	所有 X : 三角形[大于(X . 任意两边之和, X . 第三边)]
公理 2	所有 X : 三角形[小于(X . 任意两边之差, X . 第三边)]

属性关系公理、属性本体结构、聚类属性构成了整个属性本体体系, 它在整个几何学本体体系中占有很重要的位置.

(4) 属性关系内涵公理

它是对属性及关系的内在意义的一种解释性公理, 这是一种比较难以定义的公理, 我们的目的在于把属性关系的内涵尽可能地用一阶语言公理的形式固定下来, 如表 8 所示的公理 1 表明了三角形 X 的外接圆如果是 Y 的话, 那么 Y 对应的面积应大于 X 对应的面积.

表 8 内涵公理

公理	实例
公理 1	所有 X : 三角形, 所有 Y : 圆[是外接圆(Y, X) \rightarrow 大于(Y . 面积, X . 面积)]
公理 2	所有 X : 三角形, 所有 Y : 圆[是内切圆(Y, X) \rightarrow 小于(Y . 面积, X . 面积)]
公理 3	所有 X : 三角形[等于(周长 $p, \frac{a+b+c}{2}$), 等于(面积 $S, \sqrt{p \times (p-a) \times (p-b) \times (p-c)}$)]

属性关系内涵公理是一种综合性公理, 在内涵公理的形式化中, 可以包含值域公理、语义类型公理、属性关系公理.

综上, 本文采用的是半自动化几何知识获取方式, 基本的概念和关系是在几何领域专家的帮助下整理获取的; 定义、公理和定理等是根据几何学的内容进行抽取的; 其余的知识则是通过相关的性质(如传递性、对称性)等来进行推导, 自动获取的. 算法 1 为新概念知识的自动获取过程^[12-13].

算法 1. 概念知识获取.

输入: 概念集 CS

输出: 包含概念 CS 的知识库 Δ

1. 如果 $CS = \emptyset$, 概念集 CS 的知识获取结束; 否则, 任

取概念 $C, CS \leftarrow CS - \{C\}$;

2. 获取概念 C 的知识 $K(C)$, 并将其加入到已有的知识库 Δ 中: $\Delta \leftarrow \Delta \cup \{K(C)\}$; 若 $DefSet(C) = \emptyset$ (概念 C 是一个原概念), 则转步 1; 否则转步 3;

3. 对所有概念 $C' \in DefSet(C), CS \leftarrow CS \cup DefSet(C')$, 转步 1.

4.3 基于公理的几何学知识分析

在几何学本体建立之后, 一个非常重要的任务就是对几何学知识进行分析. 几何学知识分析内容包括知识一致性分析和几何学知识推理两个部分. 一致性分析是对几何学领域知识矛盾的分析, 这种矛盾可能来自于文本知识源的错误、人为的疏忽、对几何学领域的理解不够深而导致的本体设计上的失误, 还可能来自对某个问题域描述的缺失. 推理分析是分析知识的内在涵义, 从而得出知识库中没有的且和几何学领域相一致的新知识. 几何学公理中的每一类公理都既可用于一致性分析, 也可用于知识推理.

4.3.1 基于公理的几何学知识一致性分析

公理库的每一类公理均可用于一致性分析. 此处的一致性分析包括知识完备性检查、值错误检测、知识矛盾检测等. 下面举例说明.

例 2. 表 3 所示的公理 1 告诉我们, 在一个三角形的外接圆的知识中, 必然包含圆的知识和三角形的知识, 如果该三角形的外接圆的知识中没有关于圆的知识, 则该知识不完备.

此外, 在本体构建的过程中, 以 domain 和 range 等约束来实现其一致性分析.

4.3.2 基于公理的几何学知识推理

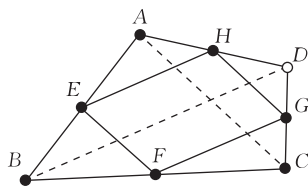
公理库的另一个重要应用在几何学知识推理方面. 根据公理, 我们可以从已知的概念出发, 推理出新概念的知识; 也可以从概念的已知属性出发, 推理出未知属性知识.

例如表 6 所示的公理 2 告诉我们一个三角形 A 的外接圆必然是一个圆 B , 这样就算库中没有关于 B 的知识, 我们也可知道 B 是一个圆; 表 8 的公理 1、公理 2 则告诉我们三角形的外接圆面积大于三角形面积, 内切圆面积小于三角形面积. 表 8 的公理 3 则告诉我们, 我们可以处理一些新出现的几何学概念, 如三角形半周长 $p = \frac{a+b+c}{2}$, 则此三角形的面积与半周长满足海伦公式

$$s = \sqrt{p \times (p-a) \times (p-b) \times (p-c)}.$$

几何学知识推理是由基于本体的推理与基于规

则的推理相结合来实现的. 基于本体的推理包括用本体描述语言中的约束等进行推理. 基于规则的推理如图 5 所示.



已知:

(1) 点E是AB的中点;

(2) 点F是BC的中点;

(3) 点G是CD的中点;

(4) 点H是DA的中点.

证明:

$EFGH$ 是平行四边形. 其中, 三角形的中位线、平行线传递、平行四边形的定义等是基于本体的推理, 而 $EH//DB$ 等知识是基于规则的推理.

证明过程:

(5) $EH//DB$ 由三角形的中位线根据(4)、(1)

(6) $FG//DB$ 由三角形的中位线根据(3)、(2)

(7) $EF//CA$ 由三角形的中位线根据(2)、(1)

(8) $GH//CA$ 由三角形的中位线根据(3)、(4)

(9) $EF//GH$ 由平行线传递根据(7)、(8)

(10) $EH//FG$ 由平行线传递根据(5)、(6)

(11) $\square EFGH$ 由平行四边形的定义根据(9)、(10)

图 5 基于规则的推理

5 结论和进一步的工作

本文从本体论的角度出发, 探讨了几何学知识的获取方法及知识表示, 构建了一个较完善的、可共享、可重用、可扩展的、实用的几何学本体, 总结出了若干几何学知识表示模式和获取模式, 形成了一套较完整的几何学知识表示方法和几何学知识获取体系, 可用于几何学专家系统、智能教学系统、语音识别、自然语言理解等方面.

目前, 本体还存在一些问题, 如本体要想获取有关几何学学科完整的知识是非常困难的; 又如由于人的理解以及知识源本身的不同, 不可避免地造成获取到的知识在知识库产生不一致. 这都给我们提出了关于本体研究更深入的课题, 诸如本体架构设计的优化、本体之间的相似性和等价性等.

为了加快几何学知识获取的效率, 完善其知识表示机制, 文本知识的自动获取^[14]也是目前正在进行的重要工作. 例如结合规则推理对框架模板进行扩展和改进, 发展和定义语义模板规范等. 又如对于毕达哥拉斯定理可进行如下扩展: 毕达哥拉斯定理的概念、毕达哥拉斯定理的试题、毕达哥拉斯定理的课件、毕达哥拉斯树等. 再则领域本体包含丰富的语义信息, 拟借助领域本体来动态生成抽取陈述所需的语义模板, 设计并开发语义模板的自动生成算法等.

致谢 感谢美国康奈尔大学 Wu Sheng-Chuan 博士的大力帮助和建议.

参 考 文 献

- [1] Tkach Daniel S. Information Mining with the IBM Intelligent Miner Family. An IBM Software Solutions White Paper, 1998
- [2] Sure Y et al. OntoEdit: Guiding ontology development by methodology and inferencing//Proceedings of the International Conference on Ontologies, Databases and Applications of SEmantics ODBASE 2002. University of California, Irvine, USA, 2002; 1205-1222
- [3] Knublauch H et al. The Protégé OWL plugin: An open development environment for semantic Web applications//Proceedings of the 3rd International Semantic Web Conference-ISWC 2004. Hiroshima, Japan, 2004; 229-243
- [4] Berners-Lee T, Hendler J, Lassila O. The semantic Web. Scientific American, 2001, 284(5): 34-43
- [5] Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Research, 2004, 32, Database issue D: 267-270
- [6] Zhou X, Wu Z, Yin A et al. Ontology development for unified traditional Chinese medical language system. Artificial Intelligence in Medicine, 2004, 32(1): 15-27
- [7] Lenat D B, Guha R V. Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project. Reading MA: Addison-Wesley Publishing Company, Inc., 1989
- [8] Uschold M, Grüninger M. Ontologies: principles, methods and application. The Knowledge Engineering Review, 1996, 11(2): 93-155
- [9] Knight K et al. Filling knowledge gaps in a broad-coverage Machine Translation system//Proceedings of the 14th International Joint Conference on Artificial Intelligence, IJCAI 95. Montréal, Québec, Canada, Morgan Kaufmann, 1995, 2: 1390-1396
- [10] Bernaras A et al. Building and reusing ontologies for electrical network applications//Proceedings of the 12th European Conf on Artificial Intelligence. ECAI 96. Budapest Hungary, 1996; 298-302
- [11] Zhou Xiao-Bin, Cao Cun-Gen. Medical knowledge acquisition: An ontology-based approach. Computer Science, 2003, 30(10): 35-39(in Chinese)
(周肖彬, 曹存根. 基于本体的医学知识获取. 计算机科学, 2003, 30(10): 35-39)
- [12] Zeng Qing-Tian. Research on ontology-based mathematical knowledge acquisition and knowledge heritage mechanism. Microelectronics & Computer, 2003, (9): 19-27(in Chinese)
(曾庆田. 基于本体的数学知识获取与知识继承机制研究. 微电子学与计算机, 2003, (9): 19-27)

- [13] She Li. Semantic-based geometric disciplinary knowledge platform. Chengdu Institute of Computer Applications, Graduate University of Chinese Academy of Sciences, Chengdu, 2006(in Chinese)
(余莉. 基于语义的几何学科知识平台[博士学位论文]. 中国科学院研究生院(成都计算机应用研究所), 成都, 2006)



ZHONG Xiu-Qin, born in 1976, Ph. D., lecturer. Her major research interests include automated reasoning and ontology.

FU Hong-Guang, born in 1965, Ph. D., professor. His

- [14] Wang Hai-Tao, Cao Cun-Gen, Gao Ying. Design and Implementation of a system for ontology-mediated knowledge acquisition from semi-structured text. Chinese Journal of Computers, 2005, 28(12): 2010-2018(in Chinese)
(王海涛, 曹存根, 高颖. 基于领域本体的半结构化文本知识自动获取方法的设计和实现. 计算机学报, 2005, 28(12): 2010-2018)

major research interests include computer algebra and knowledge science.

SHE Li, born in 1978, Ph. D., lecturer. Her major research interests include data mining and knowledge engineering.

HUANG Bin, born in 1965, master, associate professor. His major research interests include combinatorics and graph theory.

Background

This work is supported by the National Basic Research Program (973 Program) of China under grant No. 2004CB318003. The work aims at establishing a sharable, reusable and extensible geometry ontology, which reforms a system with geometry knowledge acquisition and representation. It can greatly improve the efficiency of knowl-

edge searching and reasoning in constructing a geometry knowledge base. Therefore, it is helpful for geometry expert systems, information retrieval systems, educational systems and Chinese natural language processing systems in geometry field.