

基于粗糙集的属性约简在数据挖掘中的应用研究

张颖淳 苏伯洪 曹娟

(重庆交通大学信息科学与工程学院 重庆 400074)

摘要 属性约简是粗糙集理论知识获取中的关键问题之一。先利用差别矩阵求得核属性,再利用属性的重要度作为启发式去求约简,可取得合理的属性组合,避免了基于代数方法与基于信息熵方法的复杂运算。最后通过实例分析验证了算法的有效性与可行性。

关键词 数据挖掘,属性约简,算法,粗糙集

中图分类号 P274 文献标识码 A

Study on Application of Attributive Reduction Based on Rough Sets in Data Mining

ZHANG Ying-chun SU Bo-hong CAO Juan

(College of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract The attribute reduction is one of the key issues on knowledge acquisition in rough set theory. In this paper, the discernibility matrix was utilized to obtain the core attribute first, then a new heuristic algorithm for attribute reduction was proposed based on the importance ratings of the attribute. Our algorithm can give proper combination of attributes for effective attribute reduction and therefore the complex computing is avoided, which is different from some methods relying on algebra and information entropy. Finally, the example analysis demonstrates the validity and feasibility of our algorithm.

Keywords Data mining, Attribute reduction, Algorithm, Rough set

近几年,数据库研究的热门领域是从大型数据库中自动提取未知的、隐含的、有潜在应用价值的信息或模式。粗糙集理论作为一种处理不完整、不一致、不精确等各种不完善信息的有效工具,与基于证据理论、基于模糊理论以及基于概率论的数据挖掘方法等其他处理不确定性问题理论的方法相比较,优势第一在于不需要提供问题所需处理的数据集合之外的先验知识,且它的数学基础成熟;第二在于它的易用性,与处理其他不确定性问题的理论有很强的互补性(特别是模糊理论)。因为直接对数据进行推理和分析,进而找出潜在的知识,发现隐含的规律,是粗糙集理论创建的目的和研究的出发点,所以粗糙集理论是一种天然的知识发现或者数据挖掘方法^[1]。

1 粗糙集基本知识

粗糙集理论作为一种数据分析处理理论,其主要思想是:将数据库中的属性分为条件属性和结论属性,根据不同的属性将数据库中的元组分成相应的子集,然后根据由条件属性划分的子集与由结论属性划分的子集之间的上下近似关系生成判定规则^[1-3]。

1.1 信息集

在粗糙集理论中,信息集用一个4元组 $S=(R,A,V,f)$ 来表示。其中: R,A 是一个非空集合, R 表示数据库中的所有

记录; A 表示数据库中的所有属性, C 表示条件属性, D 表示结论属性, $A=C\cup D,C\cap D=\phi;V$ 是属性值组成的集合; f 是属性 a 和记录 e 的函数, $f(a,e)$ 的值确定记录 e 关于属性 a 的取值。

1.2 等价类及相关概念

1.2.1 等价

对于属性集 A 中的任意一个属性 a ,如果记录 e_i 和 e_j 对于属性 a 的取值相等,则称 e_i 和 e_j 基于属性集 A 等价。

1.2.2 等价类

在数据库记录集合 R 中,基于某个属性集 A 的所有等价记录的集合称为基于属性集 A 的等价类。

1.2.3 划分

在数据库记录中,属于同一等价类的记录归为一类,此分类称为 R 基于属性集 A 的划分,表示为 $R=(E_i | E_i \text{ 为 } R \text{ 中基于属性集 } A \text{ 的等价类}, i=1,2,\dots)$ 。

1.2.4 最大包含(上近似)

设 R 基于 A 的划分是 E ,基于 A' 的划分是 Y,Y_j 最大包含 E_i 是指等价类 E_i 中的所有元素可能包含在 Y_j 中,即:

$$\overline{AY}_j = \cup \{E_i | E_i \in E \text{ 且 } E_i \cap Y_j \neq \phi\}$$

1.2.5 最小包含(下近似)

设 R 基于 A 的划分是 E ,基于 A' 的划分是 Y,Y_j 最小包含 E_i 是指等价类 E_i 中的所有元素全部包含在 Y_j 中,即:

到稿日期:2012-10-23 返修日期:2013-02-26 本文受重庆市科学技术研究基金项目(KJ120413)资助。

张颖淳(1967-),女,副教授,主要研究方向为数据挖掘、计算机基础教学,E-mail:786435880@qq.com;苏伯洪(1965-),男,主要研究方向为信息处理;曹娟(1980-),女,硕士生,讲师,主要研究方向为机器视觉、计算机仿真。

$$\underline{AY}_j = \cup \{E_i | E_i \in E \text{ 且 } E_i \subseteq Y_j\}$$

最大包含和最小包含是粗糙集中不精确概念的表示。上近似是指所有元素可能被包含,下近似是指所有元素都一定被包含。

1.2.6 确定度

确定度的定义为:

$$K = \alpha_A(Y_j) = \frac{|R| - |\overline{AY}_j - \underline{AY}_j|}{|R|}$$

式中, $|R|$ 和 $|\overline{AY}_j - \underline{AY}_j|$ 分别表示集合 R 以及 $\overline{AY}_j - \underline{AY}_j$ 中的记录数。

$\alpha_A(Y_j)$ 值反映了 R 中根据 A 中各属性的值就能确定其属于或不属于 Y_j 比例,即对 R 中任意一条记录,根据 A 中各属性的值确定其属于或不属于 Y_j 的可信度。

1.2.7 正域

假设 P 和 Q 是集合 U 上的等价关系, Q 是 P 的正域,记为 $POS_P(Q)$ 。则

$$POS_P(Q) = \cup_{x \in U/Q} PX$$

1.2.8 依赖

令 $K = (U, R)$ 是一个知识库, $P, Q \subset R$, 当 $K = r_p(Q) = POS_P(Q) / |U|$ 时,称知识 Q 是 K ($0 \leq K \leq 1$) 度依赖于知识 P 的,记作 $P \Rightarrow Q$ 。

- 1) 当 $K=1$ 时,称知识 Q 完全依赖于知识 P ;
- 2) 当 $0 < K < 1$ 时,称知识 Q 部分(粗糙)依赖知识 P ;
- 3) 当 $K=0$ 时,称知识 Q 完全独立于知识 P 。

2 基于粗糙集的属性约简算法

属性约简是根据不同的条件在数据库中对于决策属性的重要性不一致,去掉不必要的条件^[4-6]。由于属性约简方法一般采用 $r_p(Q) = 1$ 作为判断条件,当数据较多时这个判断条件几乎不能满足,因此在本算法中我们增加了对 $r_p(Q)$ 的排序,然后选取前 m 个值。通过研究,我们构造如下一种属性约简算法:

输入:相容的原始决策表

输出:约简属性的决策表

1) 先求核

Core = C; // Core 表示核, C 表示条件属性集;

{ P[j] = C - {C_j}; // C_j 表示第 j 个属性值;

r_p(Q) = POS_P(Q) / |U|; // 集合 Q 对 P 的依赖度;

A[j] = r_p(Q)

}

For(i=0; i < n; i++);

{ For(j=i+1; j < n; j++);

{ if A[i] < A[j]

A[i], A[j] 互换;

P[i], P[j] 互换;

}

}

For(j=0; j < m; j++) // m 表示核所需的属性数;

{ Core = Core ∩ P[j];

}

2) 约简

R ← {Core}

do

T ← R

∀ y ∈ (C - R)

If r_{RU(y)}(Q) > r_T(Q)

{ T ← R ∪ {x}, 若满足条件的有多个属性,则选择属性取值范围最小的属性;若属性取值范围相同,则选择属性值最多的属性;若都相同,则任取一属性 c

}

R ← T

until r_R(Q) ≥ r_c(Q)

return R

3 实例分析

为了测试算法的实用性和有效性,下面通过一个实例来进行分析。为了处理方便,用 c_1, c_2, \dots, c_{13} 来代替评价指标表示条件属性(见表 1), d 表示结论属性^[7,8]。表 2 为教学质量评价原始决策表,由于篇幅有限,这里只列举 24 条,其中的 1、2、3、4 分别表示优良、中、及格、不及格。用上述属性约简算法求其相对核和相对约简。

表 1 属性列表

属性	属性含义	属性	属性含义
c1	遵守纪律,无脱岗现象;上课不做与教学无关的活动	c8	职称
c2	概念清楚,论证严谨,突出重点、难点	c9	教态自然、仪表端庄
c3	方法灵活、师生互动、讲求实效;学生思维活跃、学习兴趣浓	c10	教学手段运用科学、有效,恰当使用多媒体
c4	教学思路清晰,条理性强、知识面宽、精通教学内容	c11	工龄
c5	目标明确,知识容量恰当,按照学生实际来要求	c12	语言表达简洁、清晰、准确
c6	教学设计合理、整体把握好、安排紧凑	c13	通过学习能掌握一定的知识、方法与技能
c7	学历	d	总分

表 2 教学质量评价原始决策表

工号	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	d
1	1	2	1	4	1	1	1	2	1	1	1	1	2	2
2	3	3	3	2	2	3	1	2	1	3	2	1	3	2
3	2	3	2	1	1	1	1	1	1	1	2	1	2	1
4	1	2	1	3	1	1	1	2	1	1	1	1	2	1
5	1	3	1	2	1	2	2	2	2	2	1	3	2	2
6	1	3	1	2	1	2	2	2	2	1	1	3	2	1
7	1	1	2	1	1	1	1	2	1	1	1	1	2	1
8	2	3	2	1	1	1	1	1	1	3	2	1	2	2
9	2	2	2	1	3	1	1	1	3	1	1	1	2	2
10	2	2	2	1	2	1	1	1	3	1	1	1	2	1
11	2	2	3	2	1	1	3	2	1	1	1	2	3	2
12	3	1	4	1	1	3	1	1	1	3	2	1	1	1
13	3	2	3	2	1	1	2	2	2	1	1	3	2	2
14	3	3	3	2	3	3	1	2	1	3	2	1	3	3
15	3	1	4	1	3	3	1	1	1	3	2	1	1	2
16	4	3	2	1	3	3	1	2	1	3	1	1	1	3
17	4	2	3	2	1	2	1	2	2	1	1	3	2	2
18	4	2	3	2	1	2	1	2	2	1	1	2	2	1
19	4	4	1	1	1	2	1	2	1	1	1	1	2	2
20	4	3	2	1	3	3	1	2	1	2	1	1	1	2
21	1	2	2	3	1	1	1	2	1	1	1	1	2	2
22	4	3	1	1	1	2	1	2	1	1	1	1	2	1
23	1	3	3	2	3	3	1	2	1	3	2	1	3	2
24	2	1	2	1	3	1	1	1	3	1	1	1	2	1

利用算法的第 1 步可得

$$U/\{c1\} = \{\{1,4,5,6,7,21,23\}, \{3,8,9,10,11,24\}, \{2,$$

$12,13,14,15\},\{16,17,18,19,20,22\}$
 $U/\{c2\}=\{\{4,7,12,15,21,24\},\{1,9,10,11,13,17,18\},\{2,3,5,6,8,14,16,20,22,23\},\{19,22\}\}$
 $U/\{c3\}=\{\{1,4,5,6,19\},\{3,7,8,9,10,16,20,21,24\},\{2,11,13,14,17,18,23\},\{12,15\}\}$
 $U/\{c4\}=\{\{1,3,7,8,9,10,12,15,16,19,20,22,24\},\{2,5,6,11,13,14,17,18,23\},\{4,21\}\}$
 $U/\{c5\}=\{\{1,3,4,5,6,7,8,11,12,13,17,18,19,21,22\},\{2,10\},\{9,14,15,16,20,23,24\}\}$
 $U/\{c6\}=\{\{1,3,4,7,8,9,10,11,13,20,21,24\},\{5,6,17,18,19,22\},\{2,12,14,15,16,23\}\}$
 $U/\{c7\}=\{\{1,2,3,4,7,8,9,10,12,14,15,16,17,18,19,20,21,22,23,24\},\{5,6,13\},\{11\}\}$
 $U/\{c8\}=\{\{3,8,9,10,12,15,24\},\{1,2,4,5,6,7,11,13,14,16,17,18,19,20,21,22,23\}\}$
 $U/\{c9\}=\{\{1,2,3,4,7,8,11,12,14,15,16,19,20,21,22,23\},\{5,6,13,17,18\},\{9,10,24\}\}$
 $U/\{c10\}=\{\{1,3,4,6,7,9,10,11,13,17,18,19,21,22,24\},\{5,20\},\{2,8,12,14,15,16,23\}\}$
 $U/\{c11\}=\{\{1,4,5,6,7,9,10,11,13,16,17,18,19,20,21,22,24\},\{2,3,8,12,14,15,23\}\}$
 $U/\{c12\}=\{\{1,2,3,4,7,8,9,10,12,14,15,16,19,20,21,22,23,24\},\{11,18\},\{5,6,13,17\}\}$
 $U/\{c13\}=\{\{12,15,16,20\},\{1,3,4,5,6,7,8,9,10,13,17,18,19,21,22\},\{2,11,14\}\}$
 $U/c-\{c1\}=\{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\},\{11\},\{12\},\{13\},\{14,23\},\{15\},\{16\},\{17\},\{18\},\{19\},\{20\},\{21\},\{22\},\{24\}\}$
 $U/c-\{c2\}=\{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9,24\},\{10\},\{11\},\{12\},\{13\},\{14,23\},\{15\},\{16\},\{17\},\{18\},\{19,22\},\{20\},\{21\},\{23\}\}$
 $U/c-\{c3\}=\{\{1\},\{2\},\{3\},\{4,21\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\},\{11\},\{12\},\{13\},\{14\},\{15\},\{16\},\{17\},\{18\},\{19\},\{20\},\{22\},\{23\},\{24\}\}$
 $U/c-\{c4\}=\{\{1,4\},\{2\},\{3\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\},\{11\},\{12\},\{13\},\{14\},\{15\},\{16\},\{17\},\{18\},\{19\},\{20\},\{21\},\{22\},\{23\},\{24\}\}$
 $U/c-\{c5\}=\{\{1\},\{2,14\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9,10\},\{11\},\{12,15\},\{13\},\{16\},\{17\},\{18\},\{19\},\{20\},\{21\},\{22\},\{23\},\{24\}\}$
 $U/c-\{c6\}=U,U/c-\{c7\}=U,U/c-\{c8\}=U,U/c-\{c9\}=U,U/c-\{c11\}=U,U/c-\{c13\}=U$
 $U/c-\{c10\}=\{\{1\},\{2\},\{3,8\},\{4\},\{5,6\},\{7\},\{9\},\{10\},\{11\},\{12\},\{13\},\{14\},\{15\},\{16,20\},\{17\},\{18\},\{19\},\{21\},\{22\},\{23\},\{24\}\}$
 $U/c-\{c12\}=\{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\},\{11\},\{12\},\{13\},\{14\},\{15\},\{16\},\{17,18\},\{19\},\{20\},\{21\},\{22\},\{23\},\{24\}\}$
 $POS_{c-1}(D)=\{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\},\{11\},\{12\},\{13\},\{15\},\{16\},\{17\},\{18\},\{19\},\{20\},\{21\},\{22\},\{24\}\}$

$POS_{c-2}(D)=U/c-\{c2\}=\{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{10\},\{11\},\{12\},\{13\},\{15\},\{16\},\{17\},\{18\},\{20\},\{21\},\{23\}\}$
 $POS_{c-3}(D)=\{\{1\},\{2\},\{3\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\},\{11\},\{12\},\{13\},\{14\},\{15\},\{16\},\{17\},\{18\},\{19\},\{20\},\{22\},\{23\},\{24\}\}$
 $POS_{c-4}(D)=\{\{2\},\{3\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\},\{11\},\{12\},\{13\},\{14\},\{15\},\{16\},\{17\},\{18\},\{19\},\{20\},\{21\},\{22\},\{23\},\{24\}\}$
 $POS_{c-5}(D)=\{\{1\},\{2,14\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{11\},\{13\},\{16\},\{17\},\{18\},\{19\},\{20\},\{21\},\{22\},\{23\},\{24\}\}$
 $POS_{c-6}(D)=POS_{c-7}(D)=POS_{c-8}(D)=POS_{c-9}(D)=POS_{c-11}(D)=U$
 $POS_{c-10}(D)=\{\{1\},\{2\},\{3,8\},\{4\},\{5,6\},\{7\},\{9\},\{10\},\{11\},\{12\},\{13\},\{14\},\{15\},\{17\},\{18\},\{19\},\{21\},\{22\},\{23\},\{24\}\}$
 $POS_{c-12}(D)=\{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9\},\{10\},\{11\},\{12\},\{13\},\{14\},\{15\},\{16\},\{19\},\{20\},\{21\},\{22\},\{23\},\{24\}\}$
 $r_c(D)=1,r_{c-c1}(D)=11/12,r_{c-c2}(D)=9/12,r_{c-c3}(D)=11/12,r_{c-c4}(D)=11/12,r_{c-c5}(D)=10/12$
 $r_{c-c6}(D)=r_{c-c7}(D)=r_{c-c8}(D)=r_{c-c9}(D)=r_{c-c11}(D)=1,r_{c-c10}(D)=9/12,r_{c-c12}(D)=11/12$
 决策表的核可由上排序所得:
 $Core=C-\{c6\}\cap C-\{c7\}\cap C-\{c8\}\cap C-\{c9\}\cap C-\{c11\}=\{c1,c2,c3,c4,c5,c10,c12,c13\}$
 再利用算法的第 2 步可得,剩余属性为 $\{c6,c7,c8,c9,c11\}$,然后我们分析属性集 $Core\cup\{c6\}=\{c1,c2,c3,c4,c5,c6,c10,c12,c13\}$
 $U/\{c1,c2,c3,c4,c5,c6,c10,c12,c13\}=\{\{1,4,7,19,21\},\{16,20\},\{22\},\{2,14,23\},\{3,8\},\{12,15\},\{5,6\},\{13\},\{9,10,24\},\{11\},\{17,18\}\}$
 $U/\{c1,c2,c3,c4,c5,c10,c12,c13\}=\{\{1,4,7,16,19,20,21,22\},\{2,14,23\},\{3,8,12,15\},\{5,6,13\},\{9,10,24\},\{11\},\{17,18\}\}$
 $POS\{c7,c8,c9,c11\}(D)=\{11,13,22\}$
 $POS\{c6,c7,c8,c9,c11\}(D)=\{11\}$
 $r\{c1,c2,c3,c4,c5,c6,c10,c12,c13\}(D)=3/24$
 $r\{c1,c2,c3,c4,c5,c10,c12,c13\}(D)=1/24$
 $r\{c1,c2,c3,c4,c5,c6,c10,c12,c13\}(D)>r\{c1,c2,c3,c4,c5,c10,c12,c13\}$
 同理可得
 $r\{c1,c2,c3,c4,c5,c7,c10,c12,c13\}(D)>r\{c1,c2,c3,c4,c5,c10,c12,c13\}$
 $r\{c1,c2,c3,c4,c5,c8,c10,c12,c13\}(D)>r\{c1,c2,c3,c4,c5,c10,c12,c13\}$
 $r\{c1,c2,c3,c4,c5,c9,c10,c12,c13\}(D)>r\{c1,c2,c3,c4,c5,c10,c12,c13\}$
 $r\{c1,c2,c3,c4,c5,c11,c10,c12,c13\}(D)>r\{c1,c2,c3,c4,c5,c10,c12,c13\}$

由计算结果可知,属性 $c_6, c_7, c_8, c_9, c_{11}$ 同时满足条件,虽然它们的属性取值范围都为 $\{1, 2, 3, 4\}$,但是 c_6 和 c_9 在表中的取值有 3 个, c_7, c_8, c_{11} 在表中的取值只有 2 个,按照算法,我们选择 c_6 ,算法终止。所得最后相对属性约简集为: $\{c_1, c_2, c_3, c_4, c_5, c_6, c_{10}, c_{12}, c_{13}\}$ 。说明教师在平时教学过程中应该注意以下方面:①遵守教学纪律,上课时不进行与教学无关的任何活动;②概念清楚,论证严谨,重点突出,难点明确;③方法灵活,讲求实效,学生学习兴趣浓,思维活跃,注意力集中;④教学思路清晰,条理性强,精通教学内容,知识面宽;⑤目标明确,知识容量恰当,切合学生实际,要求适度;⑥教学设计合理,教学安排紧凑,整体把握好;⑦教学手段运用科学、有效,恰当使用多媒体;⑧语言表达准确、清晰、简洁;⑨通过学习能掌握一定的知识、方法与技能。结果表明,学校在招聘新教师时,不应忽略教师的基本素质而一味只注重学历。

结束语 在数据挖掘中,基于信息熵方法与基于代数方法都需在遍历决策表内的每一个属性后,再用数学运算得到核属性集,进而约减决策表。这两种方式都不能直接求解核属性,需要引入复杂的计算公式才能实现算法^[9]。本文采用基于差别矩阵的方法,用属性约简方法得到决策表的核属性。实验分析验证了算法的可行性与有效性,为决策表的属性约简提供了一条高效的途径。

(上接第 199 页)

图形人工检验统计。统计结果如表 3 所列。

表 3 完整命题理解的准确率

章节	题目数量	正确数	准确率
相交线与平行线	15	15	100.00%
三角形	19	18	94.74%
勾股定理	14	12	85.71%
全等三角形	17	16	94.12%
四边形	17	15	88.24%
相似	18	17	94.44%
圆	20	17	85.00%

表 3 的统计结果显示,自然语言几何作图接口正确完整地作出了 110 套几何习题对应的几何图形,其余 10 套几何习题对应的几何图形均不完整,需要人工补充或修订,系统整体准确率达到 91.67%,每一章的平均准确率达到 91.75%。自然语言几何作图接口相比加入融合算法前,多作出了 9 道题的正确图形,准确率从原来的 84.17% 提高到了 91.67%。由此可以看出,本文提出的几何命题分句间形式化作图命令融合算法是有效的,可以在一定程度上提高几何作图和推理证明的准确率。

结束语 现有的几何命题理解方法没有考虑几何命题中分句间的逻辑关系,将几何命题中的每个分句进行单独处理,无法解决分句间形式化命令的冗余和逻辑冲突,严重影响了后期作图和推理的准确率。本文针对这一问题,在前期的工作基础上提出了一种几何命题分句间作图命令融合算法,并将其集成应用于自然语言几何作图接口中进行实验论证。实验结果表明,本文提出的几何命题分句间形式化作图命令融合算法是有效的,可以在一定程度上提高几何作图和推理证明的准确率。

- [1] Berry G, Boudol G. The Chemical Machine[J]. Theoretical Computer Science, 1992(96): 217-248
- [2] 张文东,李明壮,石小艳. 基于粗糙集理论的属性约简算法[J]. 计算机工程与设计, 2008, 11(29): 95-97
- [3] He X. coefficient of variation and its application to strength prediction of adhesively bonded joints[C]// Proceedings of International Conference on Measuring Technology and Mechatronics Automation. 2009: 602-605
- [4] 张长胜. 基于决策表的区分矩阵增量属性约简算法[J]. 计算机工程与应用, 2012, 48(35): 110-113
- [5] 程京,朱婧,张帆. 一个基于差别矩阵的属性约简改进算法[J]. 湖南大学学报, 2009, 36(4): 85-88
- [6] 杜晓昕,徐慧,任长伟,等. 基于粗糙集的属性约简在数据挖掘中的研究[OL]. hyyp://www.paper.edu.cn
- [7] 刘香怡. 粗糙集在教学质量评价中的应用[J]. 中国管理信息化, 2011, 11(12): 86-87
- [8] 付海艳,符谋松,张诚一. 粗糙集理论在高校教学质量评价分析中的应用[J]. 计算机工程与应用, 2007, 43(46): 214-216
- [9] 张任伟,白晓颖,郁莲,等. 决策表的属性约简算法综述[J]. 计算机科学, 2011, 11(38): 1-6
- [10] 陈媛,杨栋. 基于信息熵的属性约简算法及应用[J]. 重庆理工大学学报:自然科学版, 2013, 27(1): 42-46

但是,本文提出的融合算法建立在几何命题分句的形式化作图命令之上,无法解决几何命题分句理解不准确带来的问题。事实上,几何命题是对几何图形性质的描述,要生成正确的形式化作图命令,还需要运用几何知识进行必要的逻辑推理。因此,如何运用几何知识本体进行推理,将几何命题作为一个整体进行理解,最后生成正确的形式化作图命令序列,将是我们下一步研究的目标。

参考文献

- [1] Fabricz K, Alexin Z, Gyimothy T, et al. THALES; a Software Package for Plane Geometry Constructions with a Natural Language Interface[C]// Proceedings of the 13th conference on Computational linguistics. Vol. 1, 1990: 44-46
- [2] 余莉,符红光,方海光. 几何命题处理中的中文分词技术[J]. 计算机工程, 2005, 31(18): 180-182
- [3] 余莉,符红光. 基于自然语言处理的计算机几何作图[J]. 计算机应用, 2005, 25(1): 7-10
- [4] 石磊,亿珍珍,赵克. 一种自然语言理解系统的研究与实现[J]. 微电子学与计算机, 2006, 23(7): 83-86
- [5] Liu Qing-tang, Huang Huan, Wu Lin-jing. Using Restricted Natural Language for Geometric Construction[J]. Applied Mechanics and Materials, 2012, 145: 465-469
- [6] 郭海燕,刘清堂,陈矛,等. 面向平面几何的自然语言作图研究[J]. 计算机科学, 2012, 39(6): 503-506
- [7] 义务教育课程标准实验教科书——数学(七年级上—九年级下)[M]. 北京:人民教育出版社, 2006
- [8] 吴振奎. 世界数学奥林匹克解题库[M]. 石家庄:河北少年儿童出版社, 2000
- [9] 高小山. 数学期械化自动推理平台 MMP 用户手册[M/OL]. ht-tp://www.mmrc.iss.ac.cn/mmp/, 2012-10-08