

适应变化环境的元知识的研究

陈文伟^{1,2}, 黄金才¹, 毕季明²

(1. 国防科技大学 信息系统与管理学院, 湖南 长沙 410073; 2. 海军兵种指挥学院 作战指挥系, 广东 广州 510431)

摘要: 元知识是描述、使用一般知识的知识, 具有变化的特点. 提出了用变换产生式, 即可拓知识作为元知识的一种新的表示形式. 它比一般用规则式表示的元知识更适应变化环境. 在解决矛盾问题中用新元知识描述解决矛盾的本质是目标变换(T_c)或条件变换(T_L), 引起了关联函数值($T_{K(P)}$)的变换; 知识发现中用新元知识说明属性约简变换 T_{reduc} 是依赖于算子 $A_{\text{SGF}}(C_i)$ 的计算结果, 基于粗糙集的知识获取是依赖于算子 $A_{\text{updown}}(E_i, Y_j)$ 的计算结果; 专家系统中用新元知识作为控制知识推理的运行更体现了变化的特点. 这些应用例子表明了用变换产生式作为元知识的一种新的表示形式更适应变化环境.

关键词: 元知识; 变换产生式; 可拓知识; 矛盾问题; 知识发现; 专家系统

中图分类号: TP181 **文献标识码:** A **文章编号:** 1673-4785(2009)04-0331-04

A study on meta-knowledge suitable for a changing environment

CHEN Wen-wei^{1,2}, HUANG Jin-cai¹, BI Ji-ming²

(1. College of Information System and Management, National University of Defense Technology, Changsha 410073, China; 2. Department of Campaign and Command, Naval Arms Command Academy, Guangzhou 510431, China)

Abstract: Meta-knowledge is a form of knowledge for describing and applying general knowledge. It is in a continually changing state. The authors proposed a new method to represent meta-knowledge using transform-production, or extension knowledge. Compared with traditional meta-knowledge representational methods using rule implication, extension knowledge is more suitable for changing environments. In solving contradictory problems with new meta-knowledge, in essence, goal-transform (T_c) or condition-transform (T_L) triggers the transform of a related function's value ($TK_{(P)}$). The application of new meta-knowledge to knowledge discovery in a database shows that the transform of attribute reduction (T_{reduc}) depends on the computational result of operator $A_{\text{SGF}}(C_i)$. Knowledge acquisition based on a rough set is dependent on the computational value of operator $A_{\text{updown}}(E_i, Y_j)$. Extension knowledge is also suitable for representing meta-knowledge in order to control the operation of expert systems. It is obvious that the new method, which employs transform-production to represent meta-knowledge, plays an instructive role in summarizing and solving such problems.

Keywords: meta-knowledge; transform production; extension knowledge; contradictory question; knowledge discovery in database; expert system

元知识是知识的知识, 是对一般知识的描述、概括、处理、使用的知识. 提出一种新的元知识表示形式, 即可拓知识. 可拓知识是以可拓变换为基础, 是可拓变换的产生式, 它具有变化的特点, 能适应变化

的环境^[1-4].

1 适应变化环境的元知识概念

在可拓学中, 可拓变换定义为对对象(物元、事元、关系元)的变换, 表示成

$$T_u u = v.$$

即可拓变换 T_u 将对象 u 变为对象 v .

可拓变换可能由某个条件(原因)产生或者会引起某个结果,与可拓变换有关的具有因果关系的规则式,统称为可拓知识。

1) 可拓变换 T_u 由某一条件或原因所引起。

$$\text{Condition} \rightarrow (T_u u = v).$$

条件 Condition 可能是某一事实 $F=f$, 或是一个算子 A , 求出变量 X 的值 $A(x) = b$, 或是另一个可拓变换 $T_a a = b$, 具体表示为

$$(T_a a = b) \rightarrow (T_u u = v).$$

式中:可拓变换 T_u 称为可拓变换 T_a 的传导变换。

2) 可拓变换 T_a 产生一个结果。

$$(T_a a = b) \rightarrow \text{result}.$$

结果 result 同样可能是一个事实,或者是另一个可拓变换。

可见,可拓知识是包含可拓变换的规则知识。在人工智能中,一般知识是事实(变量的取值)的规则形式,即: $P \rightarrow Q$, 知识只体现了 P 与 Q 的静态关系。而可拓知识中的可拓变换将一对象变换成另一个对象,体现了变化的特点^[58]。

2 解决矛盾问题的元知识

2.1 解决矛盾问题的可拓模型和可拓知识

矛盾问题是问题 P 的条件 L 不能实现问题的目标 G 。量化矛盾问题 P 的不相容性用关联函数 $K(P) < 0$ 来表示。解决矛盾问题是利用条件 L 的可拓变换或目标 G 的可拓变换,引起关联函数 $K(P)$ 变换的可拓知识,使关联函数 $K(P) < 0$ 变为 $K(P) > 0$, 即关联函数发生质的变化,使矛盾问题得到解决。

解决矛盾问题的可拓模型包括3部分:

1) 量化矛盾问题 P 。对不同的问题需要利用不同的原理,建立判别矛盾问题是否解决的关联函数 $K(P)$ 。在解决矛盾问题的过程中,关联函数值是逐步变化的,设第 i 步关联函数值为 K_i 。

2) 找出化解矛盾问题的目标变换 T_G 或条件变换 T_L , 它将引起关联函数值的变换 $T_{K(P)}$ 的可拓知识。这同样需要利用原理来建立,它们也是逐步进行的。这样,解决矛盾问题的可拓知识为

$$(T_G(G_i) = G_{i+1}) \vee (T_L(L_i) = L_{i+1}) \rightarrow (T_{K(P)}(K_i) = K_{i+1}).$$

该可拓知识得到的关联函数值的变换 $T_{K(P)}$ 需要通过计算使 $K_{i+1} > K_i$, 这样才能保证矛盾问题的解决。

该可拓知识是解决矛盾问题的核心,它说明了解决矛盾问题的途径和可能性。

3) 设计算法,反复运行该可拓知识,且每运行

一次可拓知识,关联函数的值是逐步增加的,经过若干次计算后,会使 $K_i < 0$ 逐步变成 $K_{i+1} > 0$, 最终使矛盾问题得到解决。

2.2 解决“图形媒体与文本媒体”矛盾问题的元知识

图形与文本是2种不同媒体。矢量图形是在平面上展示各种图元素之间的分布关系,便于人的宏观思维和表达。但是,它不能反应出图元素之间的前后顺序,不便于有序的说明和表达。在作战中,一般利用图形来制定作战方案,它能体现宏观的平面布局。但是,在下达命令的时候,需要有序的表达,用图形来下达命令就不合适了。为解决这个矛盾问题,就必须将图形转换成为文本,文本是有前后顺序的,它适合以作战命令形式下达。

图形基元与文本基元分别表示为

$$M_A = (\text{图形 } A, \text{图元 } w_i, \text{信息 } u_i),$$

$$M_B = (\text{文本 } B, \text{句子 } s_j, \text{信息 } v_j).$$

这种图形 A 与文本 B 媒体的转换表示成宏观的可拓变换形式为

$$T_{gt}(M_A) = M_B.$$

图形 A 包括多个图元 w_i , 如圆点、直线、颜色等组成。从一个图元 w_i 找到另外一个图元 w_j , 需要通过图元之间的关系的知识推理来完成,知识表示为

$$\text{IF}(w_i) \text{ THEN}(w_j): \text{ACT}(g_k).$$

含义是若前提条件中的图元 w_i 已知,则通过执行某一识别函数 g_k , 找到结论中的图元 w_j 。该知识体现了从图元 w_i 到图元 w_j 的可拓变换:

$$T_{res}(w_i) = w_{i+1}.$$

文本是由句子组成,句子是由词组成。在作战文本中的句子 s_i 一般由固定词 sc_i 和变化的词 sd_i (关键文字)组合而成。固定的词在文本中形成了固定的格式和顺序。变化的词(关键文字)是随不同的作战方案而变化的。具体表示为

$$B = \bigcup_i s_i = \bigcup_i (sc_i + sd_i).$$

由图形转换成文本的关键在于将图元信息 u_i 变换成文本中变化词 sd_i 的信息,表示为

$$T_{gti}(u_i) = v_i.$$

图形与文本媒体转换矛盾问题的可拓模型为

1) 关联函数。

一幅图形中的图元总数设为 N 。图形转换成文本是逐个取出图元中的信息放入文本中形成词信息。这样转换的关联函数设计为

$$K_i = i - N.$$

式中:变量 i 为已完成图文转换的图元数目。每完成一个图元的寻找和信息获取,需要进行累加,即 $i = i + 1$ 。当 $K_i < 0$ 时,将继续进行图文转换,直到 $K_i = 0$ 时为

止.

2) 确定解决矛盾问题的可拓变换和可拓知识.

解决矛盾问题的可拓知识表示为

$$(T_{\text{res}}(w_i) = w_{i+1}) \wedge (T_{\text{gt}}(u_{i+1}) = v_{i+1}) \rightarrow (T(K_i) = K_{i+1}).$$

该可拓知识表示:经过图元 w_i 到图元 w_{i+1} 的可拓变换与图元信息 u_{i+1} 到文本词信息 v_{i+1} 的可拓变换,共同完成了一个新图元 w_{i+1} 的信息转换,它使关联函数值提高到 K_{i+1} .

3) 算法

当 $K_i < 0$ 时,反复计算可拓知识,通过 2 个可拓变换引起关联函数可拓变换,检查关联函数的值是否提高到 $K_i = 0$. 当 $K_i = 0$ 时,结束图文转换,输出文本信息,即 $\text{Out}(B)$.

3 知识发现的元知识

知识发现(KDD)是从数据中发现有用知识的整个过程.KDD 过程包括为 3 部分:数据准备(包括属性约简)、数据挖掘及结果的解释和评估.下面讨论基于粗糙集的属性约简和数据挖掘的元知识.

3.1 属性约简的元知识

属性约简问题是在数据库中保持分类效果不变的情况下,删除多余的属性.它的基础理论主要是粗糙集理论和信息论.按粗糙集理论,需要对数据库中的每个条件属性计算其重要度 SGF,为此引入计算重要度算子 A_{SGF} .对条件属性集 C 中的任一属性 C_i 相对决策属性 D ,计算其重要度 $A_{\text{SGF}}(C_i)$.

$A_{\text{SGF}}(C_i)$ 算子计算过程如下:

- 1) 计算条件属性集 $(C - \{C_i\})$ 的等价类;
- 2) 计算决策属性的 D 等价类;
- 3) 计算正域 $\text{Pos}(C - \{C_i\}, D)$;
- 4) 计算依赖度 $\gamma(C - \{C_i\}, D)$;
- 5) 计算 C_i 的重要度 $\text{SGF}(C - \{C_i\}, D)$.

在粗糙集属性约简中,若 $\text{SGF}(C_i, C, D) = 0$,表示属性 C_i 关于 D 是可省的,即可以对属性 C_i 进行约简.用可拓变换表示属性约简变换,如下式:

$$T_{\text{reduc}}(C) = (C - \{C_i\}).$$

该约简变换 T_{reduc} 是在算子 $A_{\text{SGF}}(C_i)$ 计算出 $A_{\text{SGF}}(C_i) = 0$ 时,才进行的变换.

算子 A_{SGF} 与可拓变换 T_{reduc} 之间的因果关系可以表示为可拓知识:

$$A_{\text{SGF}}(C_i) = 0 \rightarrow T_{\text{reduc}}(C) = (C - \{C_i\}).$$

该可拓知识表示为:若算子 A_{SGF} 对 C_i 属性计算出重要度 $\text{SGF}(C_i)$ 为零时,进行对属性 C_i 的约简变换.

该可拓知识概括了属性约简的原理和本质,是属性约简的元知识.

3.2 数据挖掘的元知识

数据挖掘是从大量数据中获取知识,这些知识实质上是这些数据的高度浓缩,但仍保留了数据的本质.这里讨论基于粗糙集理论的数据挖掘方法的元知识.该方法是通过条件属性集 E_i 与决策属性集 Y_j 之间的上下近似关系来获取知识.为此要建立一个求解集合 E_i 和 Y_j 之间上下近似关系的算子 A_{updown} .

1) $A_{\text{updown}}(E_i, Y_j)$ 算子的计算过程如下:

- ① 求条件属性集 C 中的等价类 E_i ;
- ② 求结论属性集 D 中的等价类 Y_j .
- 2) 求 E_i 和 Y_j 之间的交集,分别有 3 种情况:
 - ① $E_i \cap Y_j = E_i$;
 - ② $E_i \cap Y_j \neq E_i \neq \emptyset$;
 - ③ $E_i \cap Y_j = \emptyset$.

该算子对前 2 种情况能生成 2 类规则知识,它实际上是从数据库 D_{set} 中获取规则知识的数据挖掘变换,表示为

$$T_{\text{DM}}(D_{\text{set}}) = (E_i \rightarrow Y_j).$$

3) 基于粗糙集的数据挖掘方法的元知识.

数据挖掘的可拓变换 T_{DM} 是算子 $A_{\text{updown}}(E_i, Y_j)$ 的计算结果引起的,从中可得到 2 条可拓知识:

$$\begin{aligned} & (A_{\text{updown}}(E_i, Y_j) = E_i) \\ & \rightarrow (T_{\text{DM}}(D_{\text{set}}) = (E_i \rightarrow Y_j)), \\ & (A_{\text{updown}}(E_i, Y_j) \neq E_i) \wedge (A_{\text{updown}}(E_i, Y_j) \neq \emptyset) \\ & \rightarrow T_{\text{DM}}(D_{\text{set}}) = ((E_i \rightarrow Y_j), C_f). \end{aligned}$$

式中:可信度 C_f 为 $C_f = |E_i \cap Y_j| / |E_i|$.

这 2 条可拓知识是粗糙集获取知识的原理和本质的说明,是粗糙集数据挖掘的元知识.

4 专家系统的元知识

专家系统中的元知识主要是用来对专家系统运行的控制,用可拓知识来表示控制专家系统运行的元知识是很合适的.专家系统一般采用逆向推理,它运行控制的元知识主要包括:指定目标开始推理;检查当前变量是否处于推理树的叶结点,若是,则进行提问;提问回答符合要求时,推理进行回溯;提问回答不符合要求时,继续提问;目标求出值后,停止推理或转向另一推理树的目标等.下面对其中部分元知识改写为可拓知识表示形式,更能体现变化的特点:

1) 叶结点提问处理:当推理过程中发现当前结点 x 是叶结点 x_0 时,将叶结点变换成给定叶结点的提问句,元知识表示为

$(\text{Compare}(x, x_0) = \text{yes}) \rightarrow$

$(T_{\text{ques}}(x_0) = (\text{question}(x_0) = \text{"提问句"}))$.

2) 叶结点用户回答正确处理: 当用户回答的值 $v(\text{user})$ 属于叶结点取值 $v(x_0)$ 的范围时, 推理进行回溯, 即将上层结点 x' 置换叶结点 x_0 , 元知识表示为

$(\text{Compare}(v(\text{user}), v(x_0)) = \text{yes}) \rightarrow$

$(T_{\text{repl}}(x_0) = x')$.

3) 叶结点用户回答不正确处理: 当用户回答的值 $v(\text{user})$ 不属于叶结点取值 $v(x_0)$ 的范围时, 则继续提问, 元知识表示为

$(\text{Compare}(v(\text{user}), v(x_0)) = \text{no}) \rightarrow$

$(T_{\text{ques}}(x_0) = (\text{question}(x_0) = \text{"提问句"}))$.

4) 单推理树推理控制: 当目标结点 G 通过推理求出值 v_G 时, 停止推理. 元知识表示为

$\text{Check}(v(G) = v_G) \rightarrow (T_{\text{stop}}(x = G) = R_{\text{stop}})$.

5) 多推理树推理控制: 当一个推理树的目标结点 G 通过推理求出给定值 v_G^* 时, 控制推理机从该推理树转向另一推理树 i 的目标结点 G_i , 元知识表示为

$\text{Check}(v(G) = v_G^*) \rightarrow (T_{\text{repl}}(G) = G_i)$.

用可拓知识即变换产生式来表示专家系统中的元知识, 比原来采用的元知识表示更能体现变化的特点, 也便利专家系统程序容易控制专家系统有效运行.

5 结束语

通过以上的研究可知, 解决矛盾问题或者知识发现问题都是一个过程, 用可拓知识作为元知识来描述, 既适应了变化的需求, 又起到了把定量问题进行定性化描述, 浓缩了具体的定量计算过程. 在专家系统中的元知识用可拓知识表示, 更突出了运行专家系统的控制效果. 可拓知识作为元知识的一种新的表示形式, 是对元知识的扩充, 既能有效地把握问题的本质, 又能有效地起到指导和控制系统的运行的效果. 可拓知识作为元知识的表示形式具有广泛的应用前景.

参考文献:

- [1] 陈文伟. 挖掘变化知识的可拓数据挖掘研究[J]. 中国工程科学, 2006, 8(11): 70-73.
CHEN Wenwei. The research of mining the mutative knowledge with extension data mining[J]. Engineering Science in China, 2006, 8(11): 70-73.
- [2] 陈文伟, 杨春燕, 黄金才. 可拓知识与可拓知识推理[J]. 哈尔滨工业大学学报, 2006, 38(7): 1094-1096.
CHEN Wenwei, Yang Chunyan, HUANG Jincai. Research on extension knowledge and extension knowledge reasoning[J]. Journal of Harbin Institute of Technology, 2006, 38(7): 1094-1096.
- [3] 陈文伟. 基于本体的可拓知识链获取[J]. 智能系统学报, 2007, 2(6): 68-71.
CHEN Wenwei. Extension knowledge link acquisition based on ontology[J]. CAAI Transactions on Intelligent Systems, 2007, 2(6): 68-71.
- [4] 蔡文, 杨春燕, 陈文伟, 李兴森. 可拓集与可拓数据挖掘[M]. 北京: 科学出版社, 2008: 23-25.
- [5] 杨春燕, 蔡文. 可拓工程[M]. 北京: 科学出版社, 2007: 56-58.
- [6] 陈文伟. 数据挖掘的可拓知识与元知识[C]//中国人工智能进展. 哈尔滨, 2007: 942-946.
CHEN Wenwei. Extension knowledge and meta knowledge of data mining[C]//Progress of Artificial Intelligence in China. Harbin, 2007: 942-946.
- [7] 陈文伟. 数据仓库与数据挖掘教程[M]. 北京: 清华大学出版社, 2006: 37-51.
- [8] 陈文伟, 廖建文. 决策支持系统其开发[M]. 3版. 北京: 清华大学出版社, 2008: 16-21.

作者简介:



陈文伟, 男, 1940年生, 教授, 博士生导师, 中国人工智能学会机器学习专业委员会副主任, 中国人工智能学会可拓工程专业委员会副主任. 主要研究方向为决策支持系统、机器学习、可拓工程、数据仓库与数据挖掘. 获国家科技进步奖二等奖1项, 军队科技进步奖二、三等奖多项. 发表学术论文98篇.



黄金才, 男, 1972年生, 副教授, 博士, 中国机器学习学会常务委员和副秘书长, 湖南青年系统工程与管理研究会副会长, CCML2004程序委员会主席. 主要研究方向为决策支持、数据挖掘和作战模拟. 负责主持军队重点项目5项, 发表学术论文23篇, 被EI检索13篇.



毕季明, 男, 1972年生, 讲师, 主要研究方向为决策支持、指挥决策. 发表学术论文12篇.