

扩展产生式规则知识表示方法

刘培奇, 李增智, 赵银亮

(西安交通大学计算机系统结构及网络研究所, 710049, 西安)

摘要: 通过对现有知识表示方法进行分析, 指出了产生式规则和概念图表示方法在自然语言理解中存在的问题, 提出了扩展产生式规则知识表示方法, 并对该规则的推理机制和具体实现进行了讨论. 扩展产生式规则表示方法是将产生式规则的前提、结论和处理分别用概念图表示, 是一种混合知识表示方法. 它既保持了产生式规则的模块性, 又揭示了自然语言中的深层次关系, 与自然语言形成自然映射. 通过分析表明, 扩展产生式规则的总体性能要优于传统的产生式规则, 并且特别适合于自然语言理解中的知识表示. 该规则已应用于网络故障诊断专家系统的自然语言接口设计中.

关键词: 扩展产生式规则; 自然语言理解; 知识表示; 概念图; 专家系统

中图分类号: TP182 **文献标识码:** A **文章编号:** 0253-987X(2004)06-0587-04

Knowledge Representation of Extended Production Rule

Liu Peiqi, Li Zengzhi, Zhao Yinliang

(Institute of Computer Architecture and Network, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A knowledge representation method that can be applied to the natural language understanding and knowledge reasoning conveniently is proposed. By analyzing current methods of the knowledge representation, the shortcomings of the production rule and conceptual graphs are discovered. In order to represent the knowledge in the natural language, the method of extended production rule and its reasoning mechanisms are presented. The extended production rule is a new production rule that the premise, the conclusion and the process of the rule are conceptual graphs. This method is a mixed representation that maintains not only the modularity of the production rule, but also the natural mapping between the natural language and the conceptual graphs. It can also reveal the deep-level relation exactly in the natural language. By analyzing the performance of the extended production rule, it is found that this rule is much better than the production rule, especially fits for the knowledge representation in the natural language understanding. It has been applied to designing the natural language interface of the expert system in the fault diagnosis of the network.

Keywords: extended production rule; natural language understanding; knowledge representation; conceptual graph; expert system

知识表示是研究知识在计算机中存储和处理的形式, 是人工智能的一个重要研究方向. 常用的知识表示方法有产生式规则、语义网络、框架等, 这些方法难于揭示自然语言中的深层次关系, 不利于自然语言理解. 1976 年计算机科学家 Sowa 在文献 [1],

2] 中提出的概念图解决了自然语言的知识表示问题, 但纯粹的概念图模块性较差, 不利于程序设计.

本文根据自然语言理解和知识推理的特点, 提出扩展产生式规则表示法, 该方法将产生式规则的前提、结论和处理分别用概念图表示. 它既便于自然

语言的知识表示,又保持了知识的模块性.

1 知识表示方法

1.1 概念图

概念图是在语义网络基础上提出的一种集语言学、心理学、哲学为一体的图形知识表示法,它是由概念结点和关系结点组成的有限、连通、有向的二分图.

概念结点和关系结点是概念图中的基本结点,概念结点表示研究领域中的一个概念、实体、属性等;关系结点揭示了概念之间的关系.概念结点用矩形框中的概念标识符表示,形式为“概念类:概念所指”(泛指概念的所指域可省略);关系结点用圆中的关系标识符表示.概念图中的有向弧表示结点间的作用关系,如语句“A cat eats meat with paw”的概念图如图1所示.

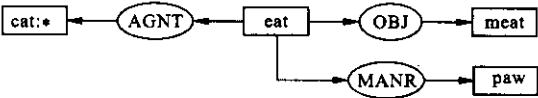


图1 自然语言的概念图表示

在图1中,关系标记AGNT表示eat动作发出者,OBJ为eat动作承受者,MANR表示eat发生的方式,概念标记CAT:*表示任意一只猫,eat、meat和paw分别表示吃、肉和爪子.从图1可以看出,若将关系结点看作概念结点间的转换条件,则概念图CG是一个4元组 (CS, RS, δ, P) ,其中 $CS = \{c_1, c_2, \dots, c_n\}$ 为概念结点集合, $RS = \{r_1, r_2, \dots, r_m\}$ 为关系结点集合, $P = \{c | c \in CS\} \wedge |c| = 1\}$, $\delta: CS \times RS \rightarrow P$ 为概念结点间的映射关系.

概念图与自然语言语句之间存在一对多映射.为了便于计算机处理,在概念图中用方括号代替矩形,圆括号代替圆,称这种表示为概念图的线性表示.图1可线性表示为

$$\begin{aligned} [eat] - (AGNT) \rightarrow [cat : *] \\ (OBJ) \rightarrow [meat] \\ (MANR) \rightarrow [paw] \end{aligned}$$

概念图表示方法除了能表示自然语言中的基本语法关系AGNT和OBJ之外,还能表示像MANR这样的深层次格关系,并且概念图与自然语言之间是一种自然映射,其显著优点是既可以表示语言中的时态、模态和格(CASE)关系,还可完成一阶谓词运算.另外,概念图中概念结点的所指域很丰富,它有

利于语义描述,可实现双向搜索和计算,但不便于模块化,不适于知识组织和程序设计.

1.2 扩展产生式规则知识表示方法

扩展产生式规则(EPR)是一种复杂的产生式规则.在EPR知识表示方法中,规则的前提、结论和处理分别用概念图表示,并引入规则强度RC、前提重要度IM和规则可信度CF.图2是EPR知识表示的形式化定义.图中“处理”表示出现故障后的处理办法、建议等,CG为概念图.图3是计算机故障诊断中第78条规则的表示形式.其中,规则强度为0.86,规则可信度为0.82.

规则::=NO IF RC<前提> THEN<结论>PROC<处理>
前提::=<简单条件>|<复合条件>
简单条件::=<条件,IM>
复合条件::={AND<条件,IM>}ⁿ
结论::=<CG,CF>{,<CG,CF>}ⁿ
处理::=<CG>{,<CG>}ⁿ
条件::=<CG>{OR<CG>}ⁿ
结论::=<CG>
RC∈[0,1], IM∈[0,1], CF∈[0,1], NO∈I⁺

图2 EPR知识的形式化定义

IF 0.86
([硬盘DOS版本]→(LOW)→[2.0版],0.8) AND
([计算机]←(AGNT)←[启动]→(LOC)→[硬盘],0.6)
THEN
[计算机故障]→(TYPE)→[DOS系统],0.82
PROC
[SYS命令]←(TOOL)←[更换]→(OBJ)→[DOS系统],
[冷启动]←(MANR)←[启动]→(OBJ)→[计算机]

LOW、AGNT、LOC、TYPE、OBJ、TOOL和MANR分别表示关系低于、动作发出者、动作作用位置、故障类型、动作作用对象、使用的工具和方式

图3 计算机故障诊断的第78条规则

2 推理机制

根据事实概念图与EPR前提的匹配情况,可将推理分为不确定推理、完全匹配推理、投影匹配推理和最大连接匹配推理.

2.1 不确定推理

为了处理不确定性知识,系统中引入了MYCIN系统的C-F模型^[3].

设e为一给定证据,则假设h成立的可信度可定义为

$$CF[h|e] = MB[h|e] - MD[h|e]$$

式中: $MB[h|e]$ 为因证据 e 而对结论 h 的信任增长度, $MD[h|e]$ 为因证据 e 而对结论 h 的不信任增长度. $MB[h|e]$ 和 $MD[h|e]$ 可分别定义为

$$MB[h|e] = \begin{cases} 1 & \text{若 } \mu(h) = 1 \\ \frac{\max\{\mu(h|e), \mu(h)\} - \mu(h)}{1 - \mu(h)} & \text{否则} \end{cases}$$

$$MD[h|e] = \begin{cases} 1 & \text{若 } \mu(h) = 0 \\ \frac{\min\{\mu(h|e), \mu(h)\} - \mu(h)}{-\mu(h)} & \text{否则} \end{cases}$$

根据规则的可信度和静态强度, 规则的可信度更新方法可总结如下.

(1) “与”结点. 若有规则 IF e_1 and e_2 THEN e_3 CF, 则可信度 $CF'(e_3 | e_1 \wedge e_2) = CF \times \min\{CF(e_1), CF(e_2)\}$.

(2) “或”结点. 若有规则 IF e_1 THEN e_3 CF¹ 和 IF e_2 THEN e_3 CF², 则可信度 $CF'(e_3 | e_1 \vee e_2) = CF(e_1)CF^1 + CF(e_2)CF^2 - CF(e_1)CF^1CF(e_2)CF^2$.

特别对于规则 IF e_1 or e_2 THEN e_3 CF, 可信度 $CF'(e_3 | e_1 \vee e_2) = (CF(e_1) + CF(e_2))CF - CF^2 CF(e_1)CF(e_2)$.

根据 CF 的基本定义及规则的可信度, 在推理树中自底向上逐层计算, 可得到结论的可信度.

2.2 完全匹配推理

完全匹配推理同一般产生式规则的推理方式类似, 是一种最理想的推理方式, 只要 EPR 规则的前提概念图与事实概念图完全相同, 则匹配成功, 得到需要的结论概念图.

2.3 投影匹配推理

定义 1 设 $c_1, c_2 \in CS$, 若 $c_1 \leq c_2$, 则称 c_2 是 c_1 的概化, c_1 是 c_2 的特化.

在定义 1 中定义了概念之间的 \leq 关系, 它反映了概念之间的种、属关系, 形成概念层次. 概念有内涵和外延之分, 概念的内涵是概念对象本质属性的反映, 而概念的外延是指具有概念所反映本质的对象集合. 在概念图中, 概念的类标识符是概念内涵的体现, 概念的所指域为概念外延的所指, 它可以是值、变量、个体或集合, 如 $[cat : \{a_1, a_2, a_3\}]$ 特指名为 a_1, a_2, a_3 的 3 只猫. 概化就是增大概念的外延, 减小概念内涵, 相反, 特化就是减小概念的外延, 增加概念的内涵.

定理 1 对于概念集合 CS 和关系 \leq , CS, \leq 为偏序集.

证明 设 $c_1, c_2, c_3 \in CS$, 按照 \leq 关系的定义, 有: ① $c_1 \leq c_1$ 显然成立, 这是一种平凡关系; ② 若 $c_1 \leq c_2$ 且 $c_2 \leq c_1$, 则 $c_1 = c_2$, 所以 \leq 是反自反的关系; ③ 若 $c_1 \leq c_2$ 且 $c_2 \leq c_3$, 则 $c_1 \leq c_3$, 所以 \leq 是传递关系. 因此, CS, \leq 为偏序集.

根据概念的 \leq 关系, 对于两个概念图也可定义概念图的 \leq 关系.

定义 2 设 u, v 是概念图. 若 $(\exists c_1 \in CS_u) \wedge (\exists c_2 \in CS_v) c_1 \leq c_2$, 则称 v 是 u 的概化, u 是 v 的特化, 记为 $u \leq v$.

定义 3 设 u, v 是概念图. 若 $u \leq v$, 则存在映射 $\pi: v \rightarrow u$, 称 πv 为 v 在 u 上的投影, 记 πv 为 u' . 显然, u' 是 u 的子图.

定理 2 设 u, v 是概念图, $u \leq v$. 若 u' 为 v 在 u 上的投影, 则对于概念 $c' \in CS_{u'}$ 和 $c \in CS_u$, 有 $c' \leq c$.

从上述定理可以看出, 投影中的概念或为 u 中的概念, 或为 v 中的概念在 u 上的特化. 例如, 对于图 4 中的规则 (v 为规则的前件) 和事实 u , 可得投影 πv .

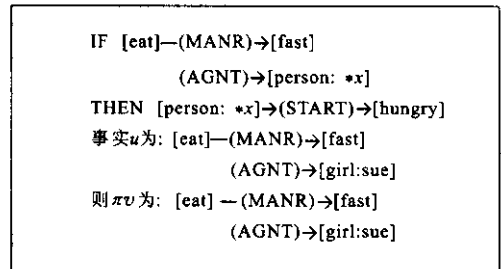


图 4 投影匹配推理

由规则和事实 u 可以得到结论 $[girl:sue] \rightarrow (START) \rightarrow [hungry]$ 概念 $[girl:sue]$ 为 u 中概念.

通过该例子的推理过程可以看出, 投影匹配推理实现了不完全推理.

2.4 最大连接匹配推理

最大连接匹配推理又称相容匹配推理, 定义如下.

定义 4 若概念 c_1 和 c_2 存在最大公共子类 c_3 , 即对任意的概念节点 c , 如果 $c \leq c_1, c \leq c_2$, 则 $c \leq c_3$, 称概念 c_1 与概念 c_2 相容.

定义 5 对于概念图 u_1, u_2 和 u_3 , 若 u_3 中的关系是 u_1 和 u_2 的并集, u_3 中的概念是 u_1 和 u_2 的相容概念, 则称 u_3 是 u_1 和 u_2 的最大连接概念图.

定理 1 保证了两个概念图相容. 最大连接就是建立在相容概念基础上的连接,它累加了原始图的相容概念,保留了概念图的不同部分. 例如

u [eat]-(MANR)-[fast]
(AGNT)-[person :sue]
 v :[eat]-(OBJ)-[pie]
(AGNT)-[girl : * x]

则 u 和 v 的最大连接为

[eat]-(MANR)-[fast]
(AGNT)-[girl :sue]
(OBJ)-[pie]

定理 3 对于任意两个概念图,则存在惟一的最大连接图.

证明 因为概念之间的关系 \leq 为偏序,则在任何两个概念之间必存在最大公共子类,所以任何两个概念图的最大连接图必然存在. 设 u_1 和 u_2 为两个概念图,则最大连接图 u_3 可通过以下几步得到:对 u_1 、 u_2 中的相同概念和相同关系,将 u_1 中的对应概念和关系加入到 u_3 中;对 u_1 、 u_2 中的不同概念和相同关系,将两概念的相容概念和关系加入到 u_3 中;对 u_1 、 u_2 中的不同概念和不同关系,将概念和关系分别加入到 u_3 中. 重复以上 3 步,可构造出惟一的概念图 u_3 . 按照定义 4 μ_3 是 u_1 和 u_2 的最大连接图.

3 扩展产生式规则的实现

在文献 [4]的基础上,EPR 在 Prolog 中可用以下谓词表示:3 元谓词 Concept(Ctype ,Cref ,CNo)记录概念的类标识符、所指域和概念号;4 元谓词 Relation(Rtype ,NodeInList ,NodeOutList ,RNo)记录关系标识符、入结点表、出结点表和关系号;3 元谓词 Graphs(GNo ,CidList ,RidList)记录概念图号、概念结

点表和关系结点表;7 元谓词 Rule(RNo ,PreList ,CFList ,ConList ,ProcList ,CF ,RC)记录规则号、前提列表、前提 CF 列表、结论列表、建议列表、CF 值和 RC 值. 根据以上谓词,图 3 可表示为:Concept(“ 硬盘 DOS 系统 ”,NULL ,1) ,... ,Concept(“ 冷启动 ”, NULL ,10) ;Relation(“ LOW ” [1] [2] ,1) ,... ,Relation(“ OBJ ” [4] [3] ,8) ;Graphs(1 [1 ,2] , [1]) ,... ,Graphs(5 [3 ,4 ,10] [7 ,8]) ;Rule(78 , [1 ,2] [0.8 ,0.6] [3] [4 ,5] 0.82 ,0.86). 其中, NULL 表示概念的所指域为空.

4 扩展产生式规则的分析、评价与结论

本文利用 Prolog 语言在 P4 计算机上分别实现了基于 EPR 规则和产生式规则的动物识别专家系统. 通过运行测试,两种表示方法的性能比较详见表 1.

在表 1 中,EPR 规则的信息量略高于产生式规则,但推理效率很接近. 由于 EPR 提供了一个符合人类自然语言形式的知识表示方法,所以在建立的专家系统中,EPR 规则在知识组织、获取和人机接口方式等方面具有明显的优势,因此 EPR 的总体性能优于产生式规则表示法.

本文针对自然语言处理中的特殊要求,在产生式规则和概念图的基础上,提出了扩展产生式规则知识表示方法,并介绍了扩展产生式规则的推理机制和实现过程. 性能分析表明,扩展产生式规则的总体性能优于传统的产生式规则. 该表示方法已应用于网络故障诊断专家系统的自然语言接口设计中. 关于扩展产生式规则的推理效率,有待进一步研究、提高.

表 1 EPR 规则同产生式规则的性能比较

	规则数 /条	谓词数 /个	信息量	知识组织方式	知识获取	知识获取效率	平均推理 时间/s	人机接口
EPR 规则	15	4	26 个概念图及语言信息	由算法将描述性知识转换成概念图	自然语言方式	高	1.32	自然语言
产生式规则	15	4	53 条语句	用人工方式从描述性知识中提取规则	按系统要求	一般	1.25	菜单或命令

参考文献:

[1] Sowa J F. Conceptual graphs for database interface [J]. IBM J Res & Dev ,1976 20(4) 336 ~ 357.

[2] Sowa J F. Conceptual structure : information processing in mind and machine [M]. Boston :Addison-Wesley Publishing Co Inc ,1984.

[3] 施鸿宝,王秋荷. 专家系统 [M]. 西安 :西安交通大学出版社,1990.

[4] 刘晓霞. 概念图知识处理器的设计 [J]. 小型微型计算机系统 2001 22(3) 351 ~ 354.

(编辑 苗 凌)