

# 从专家分析实例中学习知识\*

曹存根

(中国科学院数学研究所, 北京 100080)

(国家智能计算机研究开发中心, 北京 100080)

**摘要** 专家分析实例中含有一般实例所没有的大量信息, 本文在进一步讨论这一概念的基础上, 给出一个算法, 它能从专家分析实例学得五种规则。

**关键词** 专家分析实例, 机器学习, 知识获取。

传统的以及目前尚流行的归纳学程序的输入实例一般是由“属性-值”对构成的, 或者是其变形. 这种实例不但结构性差, 而且一般没有实例记录者的分析过程以及重要的中介信息的说明. 我们似乎不能过高地期望学习程序能从这种实例中学得高品位的知识, 这正如我们无法期望一位具有一些气象学的人能从天气预报记录中学得记录者的经验以及所使用的分析天气测量数据的方法和策略。

文献[1]提出了一个称之为专家分析实例的概念, 它与普通实例(如病历, 气象预报, 机械故障诊断报表)的概念有很大的不同; 分析实例中包括普通实例中一般所没有的: 1. 关于实例所涉对象的重要的中间假设或状态说明; 2. 实例记录者的思维或认知状态的变迁说明; 3. 有关的认知状态变迁的理由, 等等。

另一方面, 我们知道, 基于知识的系统(如专家系统)中的推理过程与一般的自然演绎系统或公理系统中的推导不一样. 前者应是自然的, 即与领域专家(DE)的解题或认知过程相近; 还应该是可解释的. 要做到自然性和可解释性就意味着系统中要具有大量的、多方面的知识, 如系统交互策略(如何确定提问时间和次序, 怎么提问等等), 形成假设的方法, 何时以及如何转向分析另一个假设等等. 经过对专家分析实例的考察和分析, 我们认为, 从专家分析实例中可以归纳出许多这样的知识, 以保证系统行为的自然性(包含正确性)以及可解释性。

由于专家分析实例中所包含的知识和信息的种类很多, 因此, 本文只局限于讨论如何从专家分析实例中学习策略和控制性知识. 关于别的知识的学习问题, 我们将另文介绍。

第1节回顾 DECO 的知识表示, 第2节讨论学习算法所使用的领域背景知识, 第3、4节分别讨论实例分离和学习算法, 最后是与有关工作的比较和全文的总结。

\* 本文 1991-07-08 收到, 1991-09-24 定稿

作者曹存根, 29岁, 1992年博士毕业于中国科学院数学研究所, 主要研究领域为人工智能, 网论。

本文通讯联系人: 曹存根, 北京 100080, 国家智能计算机研究开发中心

### 1 专家分析实例的表示

专家分析实例的表示采用 DECO 中的 PNL 表示<sup>[1]</sup>. 下面, 我们首先回顾一下 PNL. PNL 是一个六元式  $(P, T; F, \alpha, \beta, P_0)$ , 它满足以下条件:

- (1)  $P$  是一个有限集, 其中的元素称为描述元符.
- (2)  $T$  是一个有限集, 其中的元素称为关系符.
- (3)  $F \subseteq (P \times T) \cup (T \times P)$ , 称为 PNL 的流关系符集
- (4)  $P_0 \subseteq P$ , 称  $P_0$  为初态.
- (5)  $\alpha$  为一一映射,  $\alpha: P \rightarrow D; D$  中的元素由下列语法给出 (称  $D$  中的元素为描述元):

〈描述子〉[[〈描述面〉: 〈面值〉{, 〈描述面〉: 〈面值〉}]]  
 〈描述子〉 ::= 〈字母数字串〉  
 〈描述面〉 ::= 〈字母数字串〉

(6)  $\beta$  为一一映射,  $\beta: T \rightarrow R$ ; 其中  $R = \{\text{导致}\} \cup (N \times S) \cup \{r | r \text{ 为一个领域关系}\}$ ,  $N$  为自然数集,  $S = \{\text{假设, 抽取信息, 转向, 验证, 排除, 确认, 鉴别, 类比, 限制, ...}\}$  ( $S$  中的元素为专家思维过程中的一些动作, 当然我们还不可能使  $S$  完备).

PNL 的语义已在 [1] 中有阐述, 有关上面定义中的  $N \times S$  的作用及引入的理由亦在 [1] 中给出. 我们需要补充说明的是面值的表达形式. 面值均采用集合表示, 集合的含义与描述面的类型 (参见第 2 节属性类型) 有关. 我们在下文中常将描述元与属性这两个术语交换使用, 称描述元的描述子后面的部分为附属说明或二级描述元.

### 2 领域背景知识

在学习算法中要使用一些领域背景知识, 这些知识将支持学习过程. 不同的学习系统需要不同的背景知识. 下面列出一些我们的学习算法中所需要的重要的背景知识, 略去其它常用的或常见的背景知识的说明 (例如, 概念的结构, 属性的值域, 等等).

在背景知识库中, 我们要说明属性的类型. 属性的类型由图 1 给出.

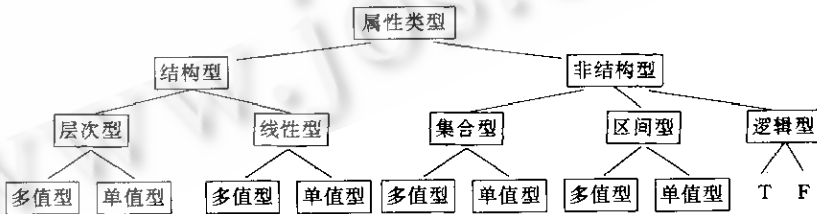


图1 属性的类型

层次型属性可在层次上作一般化或特殊化. 线性型属性的值集是全排序的 (如气温的值集). 区间型属性的取值是一个区间 (连续的或离散的).

令  $d$  是一属性,  $ds$  是它的一个可能的描述面,  $cost(d, ds)$  表示求出  $d$  的  $ds$  值的代价. 我们约定  $cost$  的值域是  $\{0, 1, 2, 3, 4, \infty\}$ ; 并且, 在缺省情况下, 它取 0 值. 当值为  $\infty$  时, 我们称  $d$  的  $ds$  是不可观察或测量的. 当  $d$  无描述面时,  $cost(d, )$  表示求  $d$  值的代价.

在学习过程中,学习系统可能会产生一些理论上有价值,但应用时代价特高的规则.因此,有必要指明下一条规则的所能容许的最高求值代价.我们约定用 MAX 表示这一值,它是一个综合指数.

学习算法在一般化时使用了一种领域启发式信息,我们以规则的形式表示这种信息(称为意义集规则,它是由领域专家给出).它的具体形式是:(令  $d$  是一个描述子, $ds$  是它的一个描述面).

$$\text{if } m_1 \wedge \cdots \wedge m_k, \text{ then } m(d, ds),$$

其中  $m_i$  是  $m$  中的点、子区间或子类.

### 3 实例分离

我们知道,如果在网上直接进行学习,那么在无人工帮助下,其效率是很低的.因此,我们在本节中介绍一种实例分离算法,将原专家分析实例进行分解,分离出一个一个小的例子.

#### 实例分离算法 CASA

输入:专家分析实例集  $E$ .

输出:假设、转向、鉴别、排除以及验证动作的正、反例集.

(1) 置  $CS_i = \emptyset$  ( $i=1, 2, \dots, 5$ ),  $CH_i = \emptyset$ .

(2) 如果  $E = \emptyset$ , 则结束本算法.

(3) 从  $E$  中任取一例  $e$ , 并置  $E = E \setminus \{e\}$ ,  $B_i = 0$ ,  $F_i = 1$ .

(4) 如果  $e$  中无标号大于  $F$  的变迁, 则转 2); 否则置  $A$  为  $e$  中标号等于  $F$  的变迁.

(5) 对  $A$  分情形讨论如下:

(5.1)  $A$  属假设类:以  $(F, A)$  在  $e$  中的前提描述元(不妨记为  $p_1, \dots, p_i$ )为证据, 以  $CH$  为历史, 分别以后提描述元(不妨记为  $H_1, \dots, H_j$ )为假设性结论(所谓假设性结论是指在  $H_1, \dots, H_j$  的描述面中加入“状态”, 其值为  $\{h\}$ , 而形成的描述元), 产生  $j$  个假设实例, 并将这  $j$  个实例推入  $CS_1$  的椎顶; 置  $B_i = F$ ,  $F_i = F + 1$ ,  $CH_i = CH \bar{\cup} \{j \text{ 个假设性结论}\}$ ; 转 4). \*

(5.2)  $A$  属转向类:同 5.1), 只是将所产生一个实例推入  $CS_2$  的椎顶(转向表示注意力的转移, 我们假设注意力只转移到一个位置, 而不是一批位置).

(5.3)  $A$  属验证类:以  $(F, A)$  的前提描述元为假设性结论  $H$ , 以  $CH \setminus \{H\}$  为证据, 产生一个验证实例, 记为  $ve$ ; 置  $B_i = F$ ; 在  $e$  中沿标号大于  $F$  的分支搜索(变迁发生); 若碰到抽取信息类动作, 则将其的后提放入临时队列  $Q$  之尾; 若碰到确认类动作, 则标志  $ve$  为“正例”, 将它推入  $CS_3$  椎顶, 并对  $CS_1$  和  $CS_2$  中与  $ve$  的假设性结论相同的实例均标以“正例”; 若碰到排除掉类动作, 则将  $ve$  标为“反例”, 推入  $CS_3$  顶, 并对  $CS_1$  和  $CS_2$  中与  $ve$  的假设性结论相同的实例标以“反例”.

置  $F$  为最后所碰到的动作(确认或排除掉)的标号,  $CH_i = CH \bar{\cup} \{H\} \bar{\cup} Q$ ; 转 4).

(5.4)  $A$  属排除类:同 5.3)类似, 但在对产生的排除类实例进行标志时, 方式相反(对  $CS_1$  和  $CS_2$  中的实例标志方式亦相反), 并将标志过的实例置于  $CS_4$  顶.

(5.5)  $A$  属鉴别类:类似 5.3), (包括对  $CS_1$  和  $CS_2$  的标志方式). 但对所产生的鉴别类实例作如下标志: 如果待鉴别的假设最后都被排除掉或被确认, 则标志此实例为反例, 否则标它为正例, 并将标志后的实例放入  $CS_3$  顶; 转 4).

在上述算法中, 我们只考虑了 5 类实例的分离, 对别的实例也可作相应的分离. 在下一节, 我们将讨论这几类实例的学习算法.

\*  $CH$  是一个队列,  $\bar{\cup}$  表示将第二个操者数加在第一个操作数之尾的联接操作.

### 4 学习算法

从第3节的 CASA 算法可以看出,我们将实例的前提部分分为两种不同作用的子部分,一是历史信息,另一是当前信息.下面将要介绍的学习算法的主要思想是:当仅根据当前信息不能学得规则时,就考虑历史信息的作用,以学得有用规则.

在介绍算法之前,先引进几个术语.令  $E$  是一实例集,  $H_i$  是  $E$  中实例的一个假设性结论,  $POS(H_i)$  为  $E$  中所有关于  $H_i$  的正实例.而  $NEG(H_i) = E/POS(H_i)$  为  $H_i$  的反例集.  $POS(H_i; d_1, \dots, d_k)$  为  $POS(H_i)$  中那些在证据部分含且仅含描述子  $d_1, \dots, d_k$  的实例所构成的集合,称为支持集.

下面,我们给出一个简化了的学习算法.

#### 学习算法 MILS

输入:  $CS_x(x; = 1, \dots, 5)$

输出:关于假设,转向,验证,排除和鉴别的规则

- (1)  $x_1 = 0$ .
- (2)  $x_i = x_i + 1$ ; 如果  $x_i \geq 6$ , 则结束本算法.
- (3) 建立  $CS_x$  的学习任务队列  $LQ$ :

$$\text{Learn } H_1, \text{Learn } H_2, \dots, \text{Learn } H_n.$$

其中,诸  $H_i$  为  $CS_x$  中实例的假设性结论,  $\text{Learn } H_1$  为  $LQ$  的顶.

(4) 如果  $LQ$  空,则转 2); 否则,从  $LQ$  顶取出一任务,记为  $\text{Learn } H_i$ .

(5) 从  $POS(H_i)$  中取出一实例  $e$ , 置  $POS(H_i)_i = POS(H_i) - \{e\}$ . 假设  $e$  的当前信息(即证据)有  $k$  个描述元, 它们的顺序如下:

$$d_1(ds_{11}, dsv_{11}, \dots, ds_{1j}, dsv_{1j}) \dots d_k(ds_{k1}, dsv_{k1}, \dots, ds_{km}, dsv_{km}) \quad (*)$$

(6) 以  $e$  为正例, 以  $POS(H_i; d_1, d_2, \dots, d_k)$  为支持集, 以  $NEG(H_i)$  为反例, 调用过程 LOR. 如果  $R$  为空, 则将  $e$  的历史信息加到 (\*) 式之前(如果有的话), 重新调用 LOR. 如果  $R$  又为空, 则报告("无关于" $H_i$ "的规则, 此时  $x$  值为" $x$ "), 并转 4). 置  $R(H_i)_i = R(H_i) \cup R$ .

(7) 如果  $R$  完全覆盖  $POS(H_i)$ , 则转 4).

(8) 置  $POS(H_i)_i = POS(H_i) \setminus \{POS(H_i) \text{ 中被 } R \text{ 覆盖的实例}\}$ , 转 5).

#### 过程 LOR:

输入: (a) 实例  $e$ ; (b) 支持集  $POS(H_i; d_1, \dots, d_k)$ ; (c)  $NEG(H_i)$

输出: 一组以  $POS(H_i; d_1, \dots, d_k)$  为支持集, 覆盖  $e$ , 但不覆盖  $NEG(H_i)$  中任何实例的规则.

(1) 置  $R_i = \emptyset$ .

(2) 从  $e$  的描述子  $(d_1, \dots, d_k)$  中找出一组满足下列条件的子描述的子元组  $(d_r, \dots, d_s)$ , 记为  $D'$ :

(i)  $D'$  的维大于 0;

(ii) 对  $D'$  的每一个分量  $d$ , 在  $e$  中有一组描述面及对应的面值(分别记为  $DS_d$  和  $DSV'_d$ ), 使得  $d_r(DS_{d_r}, DSV'_{d_r}) \wedge \dots \wedge d_s(DS_{d_s}, DSV'_{d_s}) \rightarrow H_i$  不覆盖  $NEG(H_i)$  中的任何例子. 其中,  $DS_{d_i}, DSV'_{d_i}$  等均指: 由前者与后者的对应两元组构成的序列,  $DSV'_{d_i}$  是  $DSV_{d_i}$  的一个子集;

$$(iii) \sum_{d \in D'} \sum_{d_s \in DS_d} cost(d, d_s) \leq MAX.$$

(3) 如果在上步中找不到  $D'$ , 则返回  $R$ .

(4) 将第 2 步产生的规则的结论当作假设性结论, 前提当作证据, 产生一个实例  $e'$ , 它与支持集  $POS(H_i; d_1, \dots, d_k)$  进行如下一般化: 令  $POS(H_i; d_1, \dots, d_k)$  中的实例在  $d_1, \dots, d_k$  的  $DS_{d_1}, \dots, DS_{d_k}$  上的面值分别之并为元组  $(DSV'_{d_1}, \dots, DSV'_{d_k})$ , 在保护  $(DSV'_{d_1}, \dots, DSV'_{d_k})$  的条件下, 对  $(DSV'_{d_1} \cup DSV''_{d_1}, \dots, DSV'_{d_k} \cup DSV''_{d_k})$  使用意义集规则. 闭区间规则和爬层次或线性结构规则进行一般化, 但一般化后所得规则  $r, d_r(DS_{d_r}, DSV'_{d_r}) \wedge \dots \wedge d_s(DS_{d_s}, DSV'_{d_s}) \rightarrow H_i$  不与  $NEG$

(H)相交;

(5)R:= {r},返回 R.

## 5 与有关工作的比较

我们的研究工作有其自身的特点,也与已有的工作有相似之处.本节将重点比较一下有关的研究成果.

我们知道,LEX<sup>[2]</sup>系列学习系统是学习积分公式(称为算子)的应用启发式条件.LEX对积分实例自动产生解释结构.在我们的工作中,实例的解释结构是由领域专家给出的,这种方法既可行又实用,而且可从专家那里学到许多专家的知识.专家分析过程中的思维动作(如假设,鉴别等)并不是算子,它们没有任何预先知道的前提或后提.

基于解释的学习(EBL)是目前比较热门的一个研究分支<sup>[3-5]</sup>.EBL需要的实例量很少(一个正例),但却需要大量的和正确的领域理论(domain theory)作为实例解释的依据和对解释结构进行一般化的背景知识.且不说如何获取这些领域理论,要保证它们的正确性<sup>[3]</sup>,以及可操作性(operability)的确定就很困难了.另外,实例的解释结构是实例的一个证明,它不含有领域专家的思维过程.而在我们的方法中,实例的解释结构由专家给出,而且并不需要很多的领域理论或背景知识.

Pearl<sup>[6]</sup>研究了从经验数据中学习隐藏着因果关系.但这一种学习中介因果知识的方法不易做到实用性与方法上的普通性.更难学到专家的分析策略.

与一般的归纳学习系统或工作相比,本文探讨了专家分析实例这一概念,以及分析实例的分离和归纳等问题.同时,所学得的知识均为策略性知识.

我们还可以与 Diederich 等人的 KRITON<sup>[7]</sup>,Breuker 等人的工作<sup>[8]</sup>以及 Wilkins 的工作<sup>[9]</sup>作比较,但限于篇幅,在此从略.

## 6 结束语

本文介绍了专家分析实例这一概念以及如何从专家分析实例中归纳出多种思维动作规则.

所学到的规则在知识系统推理过程中将有不同的用途.例如,假设规则可以当作框架系统中的触发条件(trigger),转向规则等可以引导推理机的控制走向.

我们认为,专家分析实例中蕴藏着大量的、多种的信息,从中可以学到许多不同用途的知识.我们将另文介绍如何从这种实例中学习对象描述以及冲突消解等方面的知识.

**致谢** 刘云光主任医生和李健主治医师为作者提供了一批分析实例,本文中的一些想法来源于对这些实例的分析以及与吴建敏和罗旭东的讨论,在此谨向他们表示感谢!

## 参考文献

- 1 曹存根.知识获取方法 DECO.计算机学报,1993,16(5).
- 2 Michalski R S, Carbonell J G, Mitchell T M. Machine Learning: an artificial intelligence approach. Palo Alto: Tioga Publishing Co., 1984.

- 3 马志方. 基于深层知识的学习模型. 博士论文, 吉林大学, 1989.
- 4 Dejong G. Explanation-based learning; an alternative view. *Machine Learning*, 1986(1):145-176.
- 5 Mitchell T M. Explanation-based generalization; a unifying view. *Machine Learning*, 1986(1):47-80.
- 6 Pearl J. Learning hidden causes from empirical data. *IJCAI*, 1985:567-572.
- 7 Diederich J, Ruhmann J, May M. KRITON; a knowledge-acquisition tool for expert system. *Int. J. Man-Machine Studies*, 1987,26:29-40.
- 8 Breuker K. KADS; structured knowledge acquisition for expert system. *Proc. Expert Systems and Their Applications*, 1985,2:887-900.
- 9 Wilkins D G. Knowledge base refinement using apprenticeship learning techniques. *Lecture Notes in Computer Science*, 1989,347:247-257.

## LEARNING FROM EXPERT ANALYTICAL CASES

Cao Cungen

*(Institute of Mathematics, The Chinese Academy of Sciences, Beijing 100080)*

*(National Research Center of Intelligent Computing Systems, Beijing 100080)*

**Abstract** A lot of useful and valuable information may be included in expert analytical cases, which is absent in ordinary cases. Based upon detailed observation and analysis analytical cases, the paper presents an algorithm which can learn five kinds of rules from them.

**Key words** Expert analytical cases, machine learning, knowledge acquisition.