

知识求精系统

HKR

知识求精

(15)

计算机科学 1995 Vol. 22 No. 2

73-75

HKR——一个启发式知识求精系统

钟 珞 李桂青

(武汉工业大学自动化系 武汉 430070)

TP18

摘 要 HKR is a system that generates classification rules from a (possibly incomplete) theory and refines the classification rules and the theory on the basis of counter examples. HKR integrates Explanation Based Learning, Similarity Based Learning and Error Propagation.

关键词 Heurist, Knowledge, Refinement.

一、引 言

知识库系统应提供有效的问题求解方法,并且能够解释求解过程。有效的求解方法应尽可能使求解过程简便,即步骤要少;而解释过程又要求尽可能的详细。影响知识工程师的另外一个复杂的因素是人类专家通常难以完全提供其领域知识的详细的细节信息。

我们解决这一非常重要的问题的方法是对领域知识使用二种表示:(a)领域理论的表示,其中一部分要用于(b)中的某些任务(如分类);(b)用外层规则形式表示分类规则集。外层规则用于有效的问题求解及领域理论的测试。

本文的目的是将归纳和演绎学习机制与知识求精以二种主要的知识表示形式集成在一起。

二、HKR 简介

非形式化地说,首先由用户提供一个理论,这一理论可以是不完备的,甚至是错误的。规则用 Prolog 描述。

然后,由用户给 HKR 提供一些训练例子。HKR 使用这些例子,并通过解释学习过程而产生分类规则,即领域中概念的再次形式化。再次形式化表示为项,这些项不通过计算就可以应用。

接下来,由用户提供更大一些的一组已分类训练的例子。这些例子按模拟学习过程的方式用于分类规则的求解。

若用户决定求精领域知识,调试程序将使用分

类规则的测试结果,并以理论描述句的形式给用户提供建议,这些描述句指出理论中的错误或者是错误的类型。

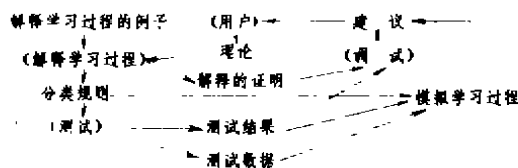


图 1 HKR 示意图

三、解释和模拟学习过程

解释学习过程的算法由文[1]提供。模拟过程首先构造一个分类规则“文本”,这些为真的分类规则由解释过程产生。由于这些规则提供了初始条件,用其可指导模拟过程产生最终的规则,“纯”模拟学习的过程不需任务例子。若因为理论是不完备的而无法得出规则,这时可用模拟过程简单地构造出规则。模拟学习过程的算法可通过文[2]的算法改进得到。

四、理论调试程序

理论的调试程序使用大量已分类例子对理论描述句进行检测。检测基于错误类型的元理论和通过引用链的错误传播。如,解释学习过程时的证明和构造。目前常用的两种不同的检测是,质量检测和共性检测。

1. 共性检测的计算

若给一个对象分配一个其不属于的类,我们称这个理论对这一对象和这个类是超共性的。至少,每

钟珞 副教授,博士生,硕士生导师,主要研究兴趣为软件工程、知识工程、神经网络。李桂青 教授,博士生导师,主要研究兴趣为结构工程、结构设计专家系统、专家系统开发环境。

一用于证明的描述句必须是具有共性的。HKR 这时假设全部描述句都是超共性的。

若对一个对象没有分配其属于某一类,而全部的理论描述句可用于这一类的一个子类证明之中,可假设其是超说明的。

对共性检测也存在以下例外:

(a)解释学习过程的证明中若出现否定句,并假定否定句存在某种类型错误,则可检测其相对错误。例如,若理论描述句“not male(x)”是超共性的,则 male(x)为超说明的。

(b)当理论用于证明某一对象属于一特别类时,理论中还可包含有算术函数。这时处理较困难。共性检测可利用表达式的多类型混合运算来进行。

(c)距离可通过共性检测和质量检测来确定。HKR 对质量检测赋予距离界值。若一个理论子句是共性很强的,则上界值一定很高或下界值一定很低。

2. 质量检测的计算

质量检测提供一个在证明中使用的数字,这个数字要么很小要么很大。质量检测从一个所观察到的质量错误,或一个通过函数的计算轨迹由距离的共性检测假定的错误开始回溯。要对函数的参变量检测函数调用的值,对数值常量不断地进行质量错误的传播。数值常量由理论定义或在对象描述中定值。

3. 理论描述句最终检测的计算

理论描述句的最终检测基于二个因素:

- (1)将若干证明中的理论规则的检测组合起来;
- (2)最终检测的类型。

每一理论规则都由类型来表示,其描述的规则表示专家的可信度。不同的类型包括定义、观察、假设和经验语句。定义表示规则,规则都假设为正确的。在经验语句之前设想一个假设。

启发式用于表示假设的理论子句为:

$\text{emax}(\text{Rule}) = \text{The maximum number of times the rule is suspected of one of the errors, defined for the type of the rule.}$

$\text{used}(\text{Rule}) = \text{The number of times the rule is used for the construction of a classification rule.}$

$\text{score}(\text{Rule}) = -(\text{emax}(\text{Rule}) - \text{used}(\text{Rule})) * \text{justification type factor.}$

因此给用户的信为:(a)理论规则是假设的,(b)错误的类型和其参数是假设的。

五、HKR 工作的例子

1. 学习的解释

领域理论说明了气象预报和解释,为什么一团

积云不是降雨云:一团积云含有小于0.7mm 的水滴不会产生雨滴。水滴的大小与云团中水滴存在的时间长短有关。

HKR 中学习的解释部分产生一个训练例子,其中包括云团、湿度、气压等的度量。训练例子描述了一种可由学习的解释过程证明的状态,天气保持良好:

Proof(produced by 学习的解释):

Numbers denote the numbers of the used theory clauses:

* $\leftarrow$ denotes how a goal is proven

* $=$ denotes the binding of an argument

niceweathercloud(c=cumulus) $\leftarrow$ 1 theory

convectioncloud(c=cumulus) $\leftarrow$ 2 definition

dropsize(c=cumulus,s=0.24975) $\leftarrow$ 3 theory

growspad(c=cumulus,G_s=0.000333) $\leftarrow$ 5 default

convectionspeed(c=cumulus,C_s=2) $\leftarrow$ 8 measure

S in H * G_s/C_s $\leftarrow$ system

S<G(G=0.3) $\leftarrow$ system

上述通过学习的解释证明过程构造了如下分类

规则:

niceweathercloud(c);convectioncloud(c),
convectionspeed(c,C_s),
H * 0.000333/C_s<0.3.

2. 增加学习的模拟文本求精解释规则

事实上使用通过学习的解释产生的规则是极具共性的。由反例,通过学习的模拟可归纳出一新的条件,规则形式如下:

not(atmosphericpressure(falling),
winddirection(southwest));
cloudedges(clear);
Humidity<70

3. 极具猜测的理论子句的计算

当分类规则‘niceweathercloud’成功时,开始训练,目标或通过学习解释过程的证明具有超共性错误类型。通过证明过程,该错误进行传播,得到一张表格,表明每一同类的规则或其参数具有猜测性的规则具有此类错误。

一个证明无法给出足够的信息表明极具猜测的理论子句,所以使用组合表格构造出证明,并测试其它分类规则。

六、相关工作

HKR 与文[3]中的 LAS 相似但也有不同之处,主要区别在于:

(a)HKR 是一个实现了的系统;

(b)HKR 从一组例子中获得知识,而不是讨论由用户提供的单一例子中隐含的知识;

(c)通过学习的模拟去构造一个学习的解释规

则文本,甚至在学习的模拟缺乏解释时,获取其分类规则。

(d)为辅助构造专家系统,LAS在非形式化的使用专家系统时要求精其理论。

另外,HKR中的调试程序类似于文[4],但我们是独立开发的,而且我们的调试程序是交互式的。

七、讨论

由上述关于HKR的描述可知,HKR还处于研制和完善期间,我们感觉到HKR的主要特点在于多种协同求精机制的可能性。

经验缺乏时,HKR过程在许多情况下产生“猜测链”,同时在知识库中“种植”一些错误。经验也表明,在不同的证明中组合错误传播结果的方法还很粗糙。例如,若一个证明A中的理论规则也出现于证明B中,则这一规则被猜测为超共性的,证明A规则的错误类型会不同于证明B中的错误类型,并且所增加的猜测也不总是正确的。

最好的方法可能是对学习的解释规则进行的学习的模拟所构造的文本事先关联一个错误类型,对理论规则进行的两种检测,计算出不同的解释学习过程的证明。

HKR调试程序对极具共性的子句计算非常合适,但对超说明的否定子句的处理就不是非常合适。

若一个证明失败,轨迹则极复杂,使得查询过程极困难。

我们准备扩充元级领域理论描述,目前,HKR仅含有限的元级知识,而某些领域可能会有丰富的元级知识,合理使用元级知识可使查询“猜测”更有效。

参考文献

- [1] R. S. Michalski, et al., The AQ15 Inductive learning System, An Overview and Experiment, ISG88-22, 1988
- [2] T. M. Mitchell, et al., Explanation Based Generalization: A Unifying View, Machine Learning 1, 1987, p. 47-80
- [3] R. G. Smith, et al., Representation and Use of Explicit Justifications for Knowledge Base Refinement, IJCAI-89
- [4] A. Ginsberg, Refinement of Expert System Knowledge Base: A Metalinguistic Framework for Heuristic Analysis, CBM-TR-147, 1988
- [5] Z. Luo, Xia Hongxia, Net-Based Knowledge Representation, Proc. of IFAC AES'92, p. 61-64
- [6] Z. Luo, Xia Hongxia, Development of Knowledge Abstraction, MICROCOMPUTER DEVELOPMENT, Vol. 3, No. 1-2, 1993, p. 50-55

全国第十三届数据库学术会议 征文通知

中国计算机学会软件专业委员会数据库学组与办公自动化专业委员会数据库学组拟于一九九六年一月在哈尔滨联合召开全国第十三届数据库学术会议。

一、征文范围

数据库理论、算法和模型;用户接口与数据库程序设计语言;数据库中的知识发现;存储器与缓冲器管理;查询优化;事务处理、并发控制和恢复;存储结构与存取方法;数据库技术的应用;数据库设计;面向对象数据库系统;多媒体数据库系统;异构和联邦数据库系统;知识库与演绎数据库系统;时态数据库系统;主动数据库系统;并行数据库系统;实时数据库系统;文本数据库系统;模糊数据库系统;科学与统计数据库系统;工程数据库系统;分布式数据库系统。

二、征文要求

所投论文应属上述范围并且是从未发表过的新研究成果,每篇论文字数不超过8000字(含论文摘要、图表、附录、参考文献),论文摘要不超过200字,论文字迹清楚、工整,一式两份,来稿无论录用与否均不退稿,故请自留底稿。

三、重要日期 1. 来稿截止日期:1995年7月15日(邮戳日期), 2. 录用通知日期:1995年8月15日, 3. 出版稿截止日期:1995年9月15日(邮戳日期)。

四、来稿请寄 150080 哈尔滨市黑龙江大学261信箱 常红 都微