

实时诊断中的无冗余动态知识存储技术研究

史永胜,蔡舒好,宋云雪

SHI Yong-sheng, CAI Shu-yu, SONG Yun-xue

中国民航大学 航空工程学院,天津 300300

College of Aeronautical Engineering, Civil Aviation University of China, Tianjin 300300, China

E-mail: yx_song@163.com

SHI Yong-sheng, CAI Shu-yu, SONG Yun-xue. Study on non-redundant storage technology of dynamic knowledge in real-time diagnosis. *Computer Engineering and Applications*, 2010, 46(9): 246-248.

Abstract: Current remote expert system does not make any control of redundancy when knowledge is stored. That makes the knowledge base redundant, and makes the management and utilization of the knowledge more difficult. In order to solve the knowledge redundancy problem during the process of remote real-time diagnosis, the non-redundant storage technology of dynamic knowledge is proposed. The concrete algorithm of this technology is realized on the platform of remote fault collaborative diagnosis of aircraft. And through data instance, the effectiveness and correctness of the algorithm are analyzed and demonstrated. The results show that the non-redundant storage technology of dynamic knowledge makes knowledge storage accurate and non-redundant, and helpful to improve the reasoning efficiency of remote real-time diagnosis.

Key words: dynamic knowledge storage; remote real-time diagnosis; expert system; redundancy control; knowledge base

摘要:当前的远程诊断专家系统在知识存储时并未做任何冗余控制,使得知识库中的信息出现较大冗余,给知识的管理和利用带来较大难度。为了解决上述问题,对无冗余的动态知识存储技术进行了研究,并基于飞机远程故障协同诊断平台,对该技术的具体算法进行了实现,通过数据实例对算法的有效性和正确性进行了分析论证。结果表明:无冗余动态知识存储技术可以精确地对知识信息进行无冗余存储,从而有效地提高远程实时诊断的推理效率。

关键词:动态知识存储;远程实时诊断;专家系统;冗余控制;知识库

DOI: 10.3778/j.issn.1002-8331.2010.09.070 文章编号: 1002-8331(2010)09-0246-03 文献标识码: A 中图分类号: TP391

1 引言

知识库是赋予专家系统智能性的重要组成部分。当前,故障实时诊断专家系统对海量、实时的动态知识信息进行推理诊断的需求增加了知识库管理的难度,这就需要较好的动态知识存储技术,以保证专家系统的性能和效率。

目前专门针对动态知识存储方面的研究还不多。如文献[1-3]中仅提出了知识管理的思想,并没有给出具体的结构和实施策略。如文献[4-5]中专家系统将提取出的知识直接存入知识库,并未对其进行一致性检测和冗余处理。这不仅造成了空间上极大的浪费,更会引起知识管理的混乱,最终影响专家系统的推理效率和准确度。因此,知识存储过程中进行有效的冗余控制是非常必要的。

2 动态知识无冗余存储技术

在知识的诸多表示方法中,产生式规则表示是目前专家系统中使用最为广泛的一种知识表示方法,很适合于故障诊断专

家系统的知识组织。下面以知识的规则表示形式具体介绍动态知识的无冗余存储技术。

待存储知识采用如下形式表示:

$$\text{if}(\text{condName}_1 \text{ op } \text{condClass}_1) \text{ and } \cdots \text{ and } (\text{condName}_m \text{ op } \text{condClass}_m)$$

$$\text{then}(\text{deciname}_1 \text{ op } \text{deciclass}_1) \text{ and } \cdots \text{ and } (\text{deciname}_n \text{ op } \text{deciclass}_n)$$

库中已存在知识表示为如下形式:

$$\text{if}(\text{scondName}_1 \text{ op } \text{scondClass}_1) \text{ and } \cdots \text{ and } (\text{scondName}_m \text{ op } \text{scondClass}_m)$$

$$\text{then}(\text{sdeciname}_1 \text{ op } \text{sdeciclass}_1) \text{ and } \cdots \text{ and } (\text{sdeciname}_n \text{ op } \text{sdeciclass}_n)$$

定义待存储知识规则中属性名为 condName_i 、分类为 condClass_i 的条件属性,其区间范围为 condValue_i ,属性名为 deciname_i 的决策属性所属分类为 deciclass_i ;库中已存在知识规则中属性名为 scondName_i 、分类为 scondClass_i 的条件属性,其区

基金项目:国家自然科学基金委员会与中国民用航空局联合资助项目(the National Natural Science Foundation of China and the Foundation of Civil Aviation Administration of China under Grant No.60879017);天津市自然科学基金项目(the Natural Science Foundation of Tianjin City of China under Grant No.08JCYBJC11600)。

作者简介:史永胜(1965-),男,博士后,教授,研究方向为结构智能化设计、故障诊断与结构修理、专家系统、知识表示;蔡舒好(1985-),女,硕士研究生,研究方向为远程故障诊断与状态监控;宋云雪(1968-),女,高级工程师,研究方向为结构智能化设计、故障诊断与结构修理的研究。

收稿日期:2008-11-25 修回日期:2009-02-23

间范围为 $scondValue_i$, 属性名为 $sdeciName_i$ 的决策属性所属分类为 $sdeciClass_i$ 。

当一条新的知识规则要进行存储时,要将其条件属性和库中已存在知识规则的条件属性进行比较。如图 1 所示,知识规则条件属性分类的区间关系有 3 种:相离、相交、包含。

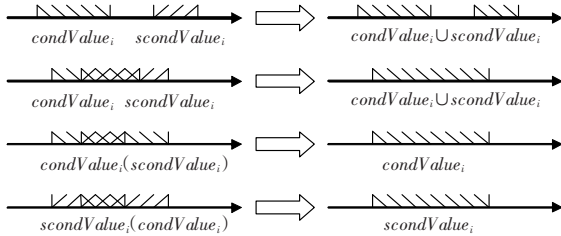


图 1 条件属性分类的区间关系

- (1) 若 $condValue_i \cup scondValue_i = \phi$, $condValue_i$ 和 $scondValue_i$ 的范围无交集,则需存储相应的条件属性;
- (2) 若 $condValue_i \cup scondValue_i \neq \phi$, $condValue_i$ 和 $scondValue_i$ 的范围相交, 则用 $condValue_i$ 和 $scondValue_i$ 区间范围的并集覆盖原 $scondValue_i$ 值;
- (3) 若 $condValue_i \supset scondValue_i$, $condValue_i$ 包含 $scondValue_i$ 且范围更大, 则用 $condValue_i$ 的区间范围覆盖原 $scondValue_i$ 值;
- (4) 若 $condValue_i \subset scondValue_i$, $condValue_i$ 包含于 $scondValue_i$, $scondValue_i$ 的范围大于 $condValue_i$, 则不需更改其区间范围。

3 基于飞机远程实时诊断中的算法实现

3.1 引入动态知识无冗余存储技术的必要性

飞机远程实时诊断中对数据的分析处理一般都在远程诊断中心完成^[6]。随着实时故障诊断的进行,大量的知识规则生成并直接存储于诊断中心的知识库中,对知识的存储和管理缺少相应的控制机制,这不仅浪费了存储空间,也容易引发知识的不一致、冗余、甚至组合爆炸等问题。这里将动态知识无冗余存储技术融入到飞机远程实时故障诊断中,作为知识存储和管理过程的控制机制,保证知识的一致性、完整性和无冗余性,解决当前飞机远程故障诊断在知识存储和管理方面的问题。

3.2 数据存储结构

在飞机远程实时诊断的知识库中建立 3 个表单来管理知识规则,分别为征兆字典、故障字典和知识库。

征兆字典用来存储知识规则的条件属性,即各征兆的具体信息。故障字典用来存储知识规则的决策属性,即各故障的具体信息。知识库用来存储知识规则信息,也就是征兆和故障的关联信息,这里不包含取值范围等具体属性信息,只关注征兆字典和故障字典之间的关联关系。3 个表单的具体结构和其间的依赖关联关系如图 2 所示。

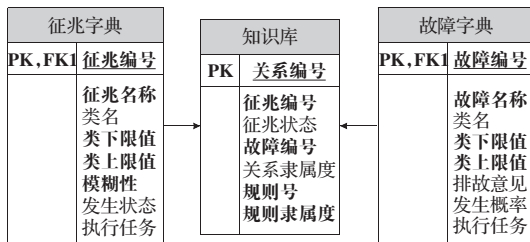


图 2 表单间的依赖关联关系图

3.3 无冗余知识存储技术的算法实现

对于待存储的每一条新的知识规则均按无冗余知识存储技术所对应的算法进行处理。根据数据存储结构对算法进行设计,伪代码如下所示:

```

for(int i=0;i<n;i++)//判断待存储知识规则的每个决策属性是否
已存在
{
    遍历故障字典的每一条记录;
    if(deciName!=sdeciName,&&deciClass!=sdeciClass)
        存储知识规则;
}
for(int i=0;i<m;i++)//判断待存储知识规则的每个条件属性是否
已存在
{
    遍历征兆字典的每一条记录;
    if(condName!=scondName,&&condClass!=scondClass)
        存储知识规则;
}
//判断与待存储知识规则完全相同的规则是否存在
遍历知识库的每一条记录;
if(count(condName)==count(scondName))
{
    if(condName==scondName,&&condClass==scondClass)
    {
        if(condValue⊃scondValue)condValue→scondValue;;
        else if(condValue∪scondValue≠ϕ)
            condValue∪scondValue→scondValue;;
        else if(condValue∪scondValue=ϕ)存储知识规则;
        else if(condValue⊂scondValue);
    }
}
    算法结束。
    
```

3.4 算法分析与验证

验证采用实验数据比较的方法以体现算法的功能性。

第一轮验证采用分类数均为 3 的等间距离散算法和参数为 0.8 的上分布变精度约简算法^[7]对数据进行处理,得到 4 个核,分别为: a, b, h, i ; 产生 1 组约简,对应 16 条规则。规则存储后,征兆字典新增记录 12 条;故障字典新增记录 2 条;知识库新增记录 64 条。

第二轮验证采用模糊聚类离散算法和参数为 0.7 的下分布变精度约简算法对数据进行处理;得到 3 个核,分别为: d, h, i ; 产生 2 组约简,每组约简对应 5 条规则。知识规则存储后,征兆字典新增记录 2 条;故障字典无新增记录;知识库新增记录 40 条。

第三轮验证采用模糊聚类离散法和概率粗糙集约简算法^[8]对数据进行处理;得到 3 个核,分别为: d, h, i ; 产生 1 组约简,对应 3 条规则。知识规则存储后,征兆字典和故障字典均无新增记录;知识库新增记录 9 条。

可以看出,随着验证次数的增加,征兆字典、故障字典和知识库中新增的记录数明显减少,这说明,该算法对存在信息冗余的知识规则不再重复存储;对存在信息重叠的知识规则进行了合并;仅对新产生的知识规则进行了存储,从而证明了该知识存储算法的有效性。

验证结果的各属性分类数据的物理存储情况如表 1 所示。

通过对验证结果数据在知识库中的存储情况进行集中比对,进一步分析算法的正确性。

表1 各属性分类数据的物理存储情况

	第一轮		第二轮		第三轮	
	分类区间	存储后结果	分类区间	存储后结果	分类区间	存储后结果
a1	-2,12	-2,12	-2,0	-2,12	-2,0	-2,12
a2	12,26	12,26	4,40	4,40	4,40	4,40
a3	26,40	26,40	-	26,40	-	26,40
b1	600,666.33	600,666.33	600,600	600,666.33	600,600	600,667.33
b2	666.33,732.66	666.33,732.66	690,770	666.33,770	690,770	666.33,770
b3	732.66,799	732.66,799	799,799	732.66,799	799,799	732.66,799
...
h1	1.4,2.2	1.4,2.2	1.4,1.4	1.4,2.2	1.4,1.4	1.4,2.2
h2	2.2,3	2.2,3	2.4,3.8	2.2,3.8	2.4,3.8	2.2,3.8
h3	3,3.8	3,3.8	-	3,3.8	-	3,3.8
i1	6.27,20.85	6.27,20.85	6.27,10	6.27,20.85	6.27,10	6.27,20.85
i2	20.85,35.42	20.85,35.42	15.3,50	15.3,50	15.3,50	15.3,50
i3	35.42,50	35.42,50	-	35.42,50	-	35.42,50
j1	0,0.33	0,0.33	0,0	0,0.33	0,0	0,0.33
j2	0.33,0.66	-	1,1	-	1,1	-
j3	0.66,1	0.66,1	-	0.66,1	-	0.66,1

选取属性分类 a_1 、 a_2 、 a_3 、 b_2 、 d_1 进行详细分析。对于属性分类 a_1 , $scondValue_i$ 为 $(-2, 12)$, $condValue_i$ 为 $(-2, 0)$, $condValue_i \subset scondValue_i$ 因此不需更改其区间范围。对于属性分类 a_2 , $scondValue_i$ 为 $(12, 26)$, $condValue_i$ 为 $(4, 40)$, $condValue_i \supset scondValue_i$, 因此将原分类区间 $(-2, 12)$ 扩展为 $(4, 40)$ 进行存储。对于属性分类 a_3 , $scondValue_i$ 为 $(-2, 12)$, $condValue_i$ 为空, $condValue_i \cup scondValue_i = \phi$, 因此为涉及对属性分类 a_3 的更改。对于属性分类 b_2 , $scondValue_i$ 为 $(666.33, 732.66)$, $condValue_i$ 为 $(690, 770)$, $condValue_i \cup scondValue_i \neq \phi$, 因此需用区间并集 $(666.33, 770)$ 代替原分类区间。对于属性分类 d_1 , $scondValue_i$ 为空, $condValue_i$ 为 $(73, 87.5)$, $condValue_i \cup scondValue_i = \phi$, 因此将第二轮产生的 d_1 的分类区间添加入库。以上的 5 个属性分类的实例全面地包含了无冗余知识存储算法中所涉及的各种情况, 上述实例分析充分证明了该知识存储算法的正确性。

4 结论

提出了远程实时诊断中的无冗余动态知识存储技术, 通过飞机远程实时故障诊断的实例验证可以看出:

(1) 算法通过冗余处理中比较并适时修改属性分类区间范围的方法, 大大减少了数据库征兆字典和故障字典所需添加的记录数目, 在保证平台稳定、精准、高效运行的基础上, 充分地提高了存储空间的利用率;

(2) 知识库所需添加记录的数目, 与新产生知识规则的数目和其中所含属性的个数有关, 由于通过冗余处理新产生的知识规则数有所减少, 知识库新增记录数也有明显的减少;

(3) 动态知识存储过程中, 通过无冗余存储技术的应用, 保证了数据库中知识的无冗余性, 从而为提高远程实时诊断系统后续推理过程的精度和效率提供了保证。

参考文献:

- [1] Elfeqy M G, Saad A A, Fouad S A. ODMQL: Object data mining query language[J]. Lecture Notes in Computer Science, 2000, 1944: 128-140.
- [2] Han J W, Fu Y J, Wang W, et al. DBMiner: A system for mining knowledge in large relational databases[C]//Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining. Menlo Park, USA: AAAI Press, 1996: 250-255.
- [3] Quinlan J R. C4.5 programs for machine learning[M]. San Francisco, USA: Morgan Kaufmann Publishers Inc, 1993.
- [4] 陈果, 宋兰琪, 陈立波, 等. 基于粗糙集理论的航空发动机滑油光谱诊断专家系统知识获取方法研究[J]. 机械科学与技术, 2007, 26(7): 897-901.
- [5] 陈展荣, 徐娟, 禹智涛. 柔性结构的自组织模糊主动振动控制[J]. 华南理工大学学报, 2004, 32(3): 61-65.
- [6] 刘建辉, 秦现生. 远程协同故障诊断综述及研究[J]. 制造技术与机床, 2006(12): 36-40.
- [7] 隋光磊, 史永胜, 张传超. 基于变精度粗集的神经网络故障诊断模型研究[J]. 计算机工程与设计, 2008, 29(15): 3998-4000.
- [8] 宋云雪, 张传超, 史永胜. 基于模糊粗集的航空发动机特征参数提取算法[J]. 航空动力学报, 2008, 23(6): 1127-1130.
- [9] 郑祖国. 投影寻踪回归(PPR)技术在水泥配方优化中的应用[J]. 八一农学院学报, 1995, 18(1): 20-24.
- [10] 王晓原, 杨新月. 基于决策树的不同驾驶行为决策机制研究[J]. 系统仿真学报, 2008, 20(2): 415-420.
- [11] 王晓原, 杨新月, 王凤群. 基于灰关联熵理论的驾驶决策影响因子研究[J]. 中国安全科学学报, 2007, 17(5): 126-132.
- [12] 李裕奇, 刘海燕, 赵联文. 非参数统计方法[M]. 成都: 西南交通大学出版社, 1998: 81-83.
- [13] 王晓原, 宿宝臣, 孟昭为, 等. 微观交通仿真跟驰模型影响因子选择研究[J]. 软科学, 2004, 18(2): 16-19.
- [14] 刘岩, 贾丽洁. 非参数统计中两样本比较方法研究[J]. 燕山大学学报, 2006, 30(5): 392-394.

(上接 207 页)

- [3] 王晓原, 隗志才, 贾洪飞. 开发和评价 ITS 的微观交通流仿真模型[J]. 交通运输工程学报, 2002, 2(1): 64-66.
- [4] 李庆丰, 王兆安, 杨建国. 混杂交通微观仿真原型系统的验证[J]. 系统仿真学报, 2005, 17(3): 545-547.
- [5] 王晓原. 微观交通仿真中的模型验证技术[J]. 计算机工程与应用, 2005, 41(5): 188-192.
- [6] Benekohal R F. Procedure for validation of microscopic traffic flow simulation model[J]. Transportation Research, 1990, 6F(3): 190-202.
- [7] Huber P J. Projection pursuit (with discussion)[J]. Ann Statist, 1985, 13: 435-525.
- [8] Friedman J H, Stuetzle W. Projection pursuit regression[J]. Amer Statist Assoc, 1981, 76: 817-823.