



# 基于混合学习算法 IHMCAP 的 故障诊断模型<sup>1)</sup>

周志华 陈兆乾 骆斌 陈世福

(南京大学计算机软件新技术国家重点实验室 南京 210093)

**摘要** 在故障集和差错属性集的基础上,通过结合了基于概率论的符号学习与神经网络学习的增量式混合型多概念获取算法 IHMCAP 寻找属性值与故障类型之间的对应关系,由此建立一个故障诊断模型.实验表明,该模型不仅精度高、速度快、学习能力强,而且在利用系统的先验知识与新增数据上也取得了均衡.

**关键词** 故障诊断,神经网络,机器学习.

## A FAULT DIAGNOSIS MODEL BASED ON HYBRID LEARNING ALGORITHM IHMCAP

ZHOU Zhihua CHEN Zhaoqian LUO Bin CHEN Shifu

(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093)

**Abstract** In this paper, a fault diagnosis model that uses an incremental hybrid multi-concept acquisition algorithm IHMCAP is proposed based upon fault set and defective attribute set. The model combines probabilistic based symbolic learning and neural learning to search for the relationships between attribute values and fault types. Experiment results show that this fault diagnosis model not only achieves high accuracy, fast speed, strong learning ability, but also well balances the utility of domain knowledge and fresh data.

**Key words** Fault diagnosis, neural networks, machine learning.

## 1 引言

IHMCAP<sup>[1]</sup>将基于概率论的符号学习与神经网络学习相结合,通过引入 FTART<sup>[2,3]</sup>神经网络,成功地解决了符号学习与神经网络学习精度之间的均衡性问题.该算法采用了

1)国家自然科学基金、江苏省自然科学基金资助项目.

一种独特的增量学习机制<sup>[4]</sup>,当增加新的示例时,只需对原结构进行快速调整即可提高学习精度.同时,该机制还可以降低算法对噪音数据的敏感度<sup>[5]</sup>,从而使得 IHMCAP 可以较好地适应实际应用领域的需要.本文首先在故障集和差错属性集的基础上建立故障模型,然后利用 IHMCAP 算法,运用分类的观点处理故障诊断问题,从而获得一个基于混合学习算法的故障诊断模型.该模型不仅可以合理地利用系统的先验知识,还可以凭借 IHMCAP 的增量学习能力自适应地对新增故障类型进行处理.

## 2 基于 IHMCAP 的故障诊断模型

### 2.1 故障集和差错属性集

令  $FS$  为系统  $S$  的故障集,  $FS = \{f_0, f_1, f_2, \dots, f_n\}$ ,  $f_0$  表示无故障.

$AS_i = \{A^i | A^i \text{ 属性作为 } S \text{ 的动态属性,能表示由故障 } f_i \text{ 导致的差错}\}$ ,

$IN\_AS = \{A^k | A^k \text{ 属性作为 } S \text{ 的输入属性,能决定 } S \text{ 动态特性}\}$ ,

$$AS' = \bigcup_{i=0}^n AS_i, AS = AS' \cup IN\_AS.$$

由此得到系统  $S$  的故障集  $FS$  对应的差错属性集  $AS$ ,不妨设  $AS = \{A_1, A_2, \dots, A_m\}$ ,  $t$  时刻诊断出系统  $S$  的故障  $f_i$  即可表示为  $\exists \Phi$  使得下式成立:

$$\Phi(A_1(t), A_2(t), \dots, A_m(t)) = f_i, \quad i = 0, 1, \dots, n,$$

式中  $A_j(t)$  表示属性  $A_j$  在时刻  $t$  的取值,  $j = 1, \dots, m$ .

$\Phi$  就代表具体的故障诊断方法.若把  $\Phi$  看作是系统专家进行的诊断,则其自变量的取值应是属性集  $AS$  在  $0 \sim t$  时间段上的所有取值,而诊断结果则可能不仅是某个  $f_i$ ,而是  $FS$  中若干个故障的逻辑运算.因此,从本质上讲,故障诊断就是寻找故障的表现形式到故障本身的映射.

### 2.2 IHMCAP 算法

IHMCAP<sup>[1]</sup>采用离散属性优先策略生成结合神经网络的混合型二叉判定树.算法首先根据示例集的离散属性进行处理,如果不能确定故障类型,就在需要划分的树叶结点上引入 FTART 神经网络模型<sup>[2,3]</sup>,根据示例的连续属性继续进行学习.由于该算法在进行判别时不涉及耗时的迭代运算,因此可以获得较快的处理速度.下面为算法的形式描述.

**定义 1.**  $E = \{x | x = \langle c, a_1, \dots, a_n \rangle, c \in C, a_i \in A_i\}$ .

$E$  称为示例集,  $E$  中元素  $x$  为示例,集合  $C$  为概念,  $A_i$  为属性;  $n$  为示例集  $E$  的属性数,  $|C|$  为示例集  $E$  中概念数,  $|E|$  为示例数;示例  $x$  的第一个分量为其概念值,其它分量  $a_i$  为其属性  $A_i$  的取值.

**定义 2.**  $RL = \{\langle X, v \rangle\}$ ,  $X \in \{C, A_1, \dots, A_n\}$ ,  $v \in X$ .  $RL$  称为条件集合,  $\langle X, v \rangle$  称为条件.

**定义 3.** 集合  $T = \{\langle i, r \rangle | i \in N^+, r \in RL\}$ , 如果满足以下条件,则称  $T$  为二叉判定树:  
 $\forall i, r \langle i, r \rangle \in T, \exists fr, ybr, ebr (\langle \lceil \frac{i-1}{2} \rceil, fr \rangle \in T) \wedge (\langle 2 \lceil \frac{i-1}{2} \rceil + 1, ybr \rangle \in T)$   
 $\wedge (\langle \lceil \frac{i-1}{2} \rceil + 2, ebr \rangle \in T)$ , 其中,  $\lceil \quad \rceil$  是取整函数.

算法描述:

输入:示例集合  $E$ .

输出:二叉判定树  $T$ .

0.  $N \leftarrow \{\langle E, 0 \rangle\}, T \leftarrow \emptyset$

1.  $\langle S, No \rangle \leftarrow \text{SelectE}(N), r \leftarrow \text{Formula}(S)$

2. WHILE NOT(Void( $r$ ))DO

(I)  $S^+ \leftarrow \{x | x \in S \text{ 且 } x \text{ 满足条件 } r\}, S^- \leftarrow \{x | x \in S \text{ 且 } x \text{ 不满足条件 } r\}$

(II)  $T \leftarrow T \cup \{\langle No, r \rangle\}; N \leftarrow N \cup \{\langle S^+, 2 * No + 1 \rangle, \langle S^-, 2 * No + 2 \rangle\} - \{\langle S, No \rangle\}$

(III)  $\langle S, No \rangle \leftarrow \text{SelectE}(N), r \leftarrow \text{Formula}(S)$

3. RESORT\_TO\_ANN

其中  $N = \{\langle Ex, n \rangle | Ex \subseteq E, n \in N^+\}$ ;

$\text{SelectE} \in f: \{N\} \rightarrow N$ ,  $\text{SelectE}(N)$  从当前判定树中选择一个需要继续划分的叶结点;

$\text{Formula} \in f: N \rightarrow RL$ ,  $\text{Formula}(S)$  根据示例集生成划分条件;

$\text{Void} \in f: RL \rightarrow \text{BOOLEAN}$ ,  $\text{Void}(r)$  判断条件是否为空;

RESORT-TO-ANN 引入神经网络模型进行进一步学习.

### 2.3 诊断过程

1) 选择学习样本. 随机抽取一定数量的属性-故障值对构成训练示例集, 尽可能覆盖故障及差错属性空间.

2) 初步学习训练. 对示例集中的数据用 FTART 网络进行学习, 然后, 从属性集中删除那些权值很小的作用不大的属性, 从而使属性个数得以压缩, 以减小判定树和神经网络的规模.

3) 部分属性离散化. 对示例集中取值范围较小的整型属性进行离散化处理, 减少连续属性的个数.

4) 学习训练和增量学习. 对示例集中的数据, 用 IHMCAP 算法进行学习, 形成混合型判定树. 对新增的属性-故障值数据以增量方式进行学习.

5) 验证诊断效果. 用剩余的属性-故障值对, 对构造出的混合型判定树进行验证, 如果精度不满足要求, 则将那些不能进行正确诊断的示例按一定的比例增加到训练示例集中, 返回 4) 重新进行学习, 直到诊断精度满足要求, 即得到一个最终的故障学习诊断模型.

## 3 应用实例

X 指控设备中情报处理软件分成两个部分, 分别驻留在指挥决策机和航迹处理机中, 如图 1 所示. 由于搜索雷达(SR)点迹数目过载或正北脉冲间隔严重超差、操作失误、错误的编码或不完善的设计而引起的情报软件中的错误状态, 称之为情报软件故障. 其错误状态包括因情报软件故障而导致的死机、异常中断或不能完成其应有的功能等.

根据中国船舶工业总公司 716 研究所提供的数据, 我们利用本文提出的故障诊断模型研制了一个情报软件故障诊断系统, 取得了较好的实际应用效果. 因篇幅所限, 在此仅作简要介绍.

原始示例集中共有 352 个示例, 每个示例有 22 个属性, 其中 8 个离散属性, 如“威胁

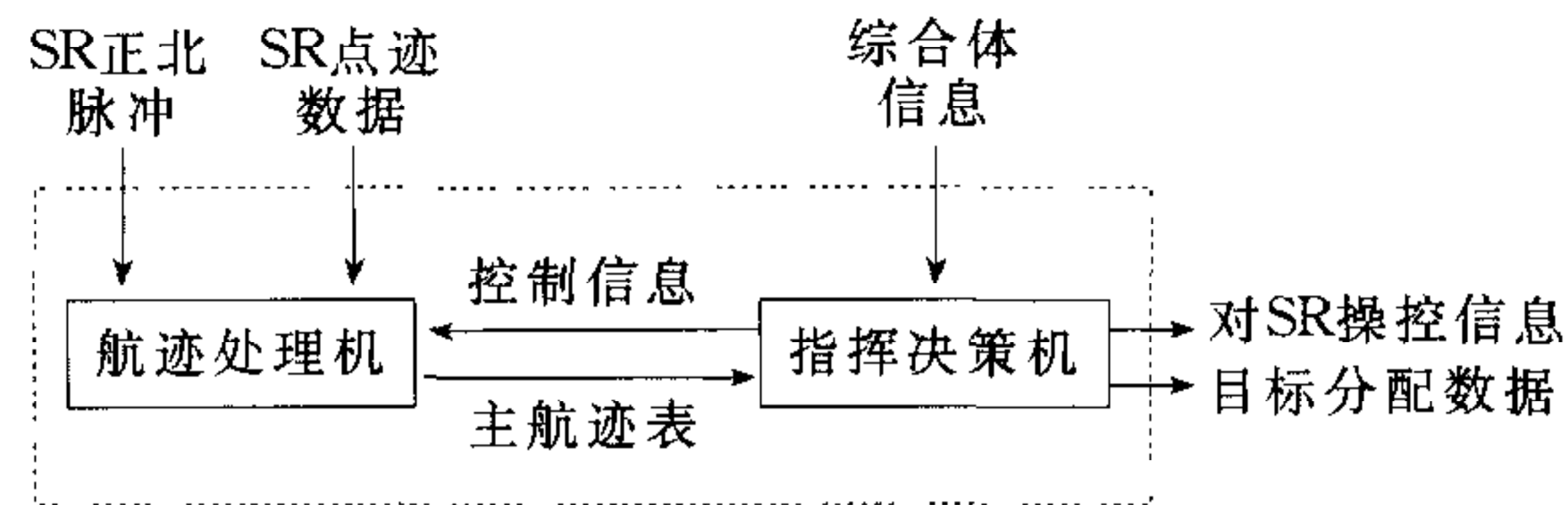


图1 指控设备中的情报处理软件系统

排序变化情况”等,14个连续属性,如“测试起始航迹数”等.要诊断10种故障情况,如“威胁排序模块故障”等.首先从原始示例集中随机抽取172个示例组成训练集对系统进行训练,再用剩余的180个示例组成测试集对其进行测试,诊断正确率为92.22%.为检验在线诊断效果,我们对训练好的系统增加18个示例进行增量学习后,用相同的测试集进行测试,正确率提高至96.11%.

## 4 结束语

本文在故障集和差错属性集的基础上,通过引入混合型学习算法 IHMCAP,以分类的观点寻找故障症状与故障类型之间的映射,从而建立了一种新型故障诊断模型,不仅具有较强的学习能力和较快的处理速度,还在利用新增数据和专家先验知识上取得了均衡.

## 参 考 文 献

- 1 陈兆乾,周志华等.增量式 IHMCAP 算法的研究及其应用.计算机学报,1998,21(8):759~764
- 2 陈兆乾,周戎等.一种新的自适应谐振算法.软件学报,1996,7(8):458~465
- 3 陈兆乾,李红兵等.对 FTART 算法的研究及改进.软件学报,1997,8(4):259~265
- 4 陈兆乾,周志华等.混合型学习模型 HLM 中的增量学习算法.软件学报,1997,8(11):875~880
- 5 周志华,陈兆乾等. IHMCAP 算法对噪音数据的处理.清华大学学报(自然科学版),1998,38(S2):118~122

**周志华** 男,1973年生.1996年在南京大学计算机系获学士学位,1998年获硕士学位并提前攻读博士学位.主要研究领域:机器学习、神经网络、遗传算法、数据挖掘.已参与编撰人工智能教材一部,发表论文10余篇.

**陈兆乾** 女,1940年生.教授.1963年毕业于南京大学数学系,现在南京大学计算机系任教.主要研究领域:人工智能、知识工程、机器学习、神经网络.近年来主持完成10余项国家及江苏省自然科学基金项目,多次获省部级科技进步奖,已出版著作11部,发表论文60余篇.