



陈 果

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

陈 果¹, 宋兰琪², 陈立波², 张占纲²

(¹南京航空航天大学 民航学院, 南京 210016;

²空军装备研究院 航空装备研究所, 北京 100076)

摘 要:针对航空发动机滑油光谱诊断专家系统的知识获取问题, 本文建立了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模型。首先建立反映光谱元素浓度及元浓度梯度与发动机磨损故障之间关系的典型故障样本集; 然后运用粗糙集理论的离散、约简及规则提取等算法, 从大量的故障样本中自动获取知识规则, 并将知识规则存储于专家系统知识库中; 最后, 建立推理机, 运用一定的推理策略实现发动机磨损故障诊断。本文利用航空发动机实测的油样光谱数据对所建立的粗糙集知识获取方法, 进行了实例验证。

关 键 词:航空发动机; 滑油光谱诊断; 专家系统; 粗糙集理论; 知识获取

中图分类号: TH165.3

文献标识码: A

文章编号: 1003-8728(2007)07-0897-05

Knowledge Acquisition of Aero-engine Spectrometric Oil Diagnosis Expert System Using Rough Set Theory

Chen Guo¹, Song Lanqi², Chen Libo², Zhang Zhangan²

(¹ College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016;

² Aerospace Equipment Institute of Air Force Equipment Academy, Beijing 100076)

Abstract: The paper studies the knowledge acquisition of aero-engine spectrometric oil diagnosis expert system and establishes the knowledge acquisition model with the rough set theory. Firstly, it establishes the representative sample sets which reflect the relationship between the spectrum element content and spectrum element content gradient and the wear faults of an aero-engine. Secondly, it uses the discretization, reduction and rule extraction algorithms of the rough set theory to acquire automatically knowledge rules from abundant fault samples, and saves the rules into the knowledge base of the expert system. Finally, it establishes a reasoning machine and uses some reasoning strategies to carry out the wear fault diagnosis of the aero-engine. With aero-engine spectrometric oil data, the paper verifies the rough set's knowledge acquisition method.

Key words: aero-engine; spectrometric oil diagnosis; expert system; rough set theory; knowledge acquisition

航空发动机的磨损故障诊断主要是通过监测发动机滑油中的金属和非金属磨粒的含量、浓度、尺寸、形状和颜色等信息来诊断发动机的传动系统和具有相互运动的摩擦副的磨损性质、磨损类型和磨损部位。由于光谱分析不仅能够建立起元素含量与磨损部位和故障性质之间的关系, 而且具有精度高, 数据重复性好, 分析时间短以及对非铁磁性金属颗粒的灵敏性高等优点。目前, 在航空、铁路、军事、航海等领域得到了广泛应用^[1~3]。

光谱诊断专家系统是利用光谱分析进行航空发动机磨损故障诊断的高级阶段。目前, 美国和加拿大已有商品化的滑油检测专家系统^[4~7], 如加拿大针对柴油机开发的发动机诊断管理系统 EDMS, 美国和加拿大研制的预知性管理系统 BMP, 加拿大 GasTops 公司开发美国 JOAP 参与的状态监测和解释系统 CMIS, 美国 Mobil 润滑油公司开发的先进快速分析系统 PFALink, 美国和加拿大共同研究的润滑油分析专家系统 Lube Analyst 和 Atlas 等。这些

收稿日期: 2006-09-21

作者简介: 陈 果 (1972-), 男 (汉), 四川, 博士, 副教授, cgyzx@263.net

软件所提供的仅仅是一个框架和管理系统,其核心知识库要用户自己开发,另外还需用户提供所监控对象的磨损元素界限值。国内研究单位针对特定的对象也开发了许多滑油光谱分析专家系统^[8,9],并取得了许多成绩。

但是,应该看到,对于包括航空发动机在内的机器设备,在智能诊断专家系统的研究方面还存在很多不足之处,主要表现在目前的专家系统普遍存在知识获取能力弱,知识更新困难,知识适应性差等方面的缺点。目前的专家系统知识获取基本上是基于经验的机械式的学习方法,知识更新困难,知识规则经常会出现严重的不一致、冗余、甚至组合爆炸等问题。

为了解决故障诊断知识不完整和不确定问题及传统专家系统知识获取和更新等问题,本文将粗糙集理论引入到航空发动机油样光谱诊断专家系统,用粗糙集理论^[10,11]的一系列算法实现专家系统知识的自动获取,有效地提升了传统专家系统的智能化水平。

1 基于粗糙集理论知识获取的诊断专家系统流程

基于知识的专家系统^[12]主要由知识库、推理机、人机接口、知识获取子系统、解释子系统、全局数据库组成。传统的基于知识规则的专家系统由于知识获取和更新困难,本文提出基于粗糙集理论知识自动获取的专家系统。其工作流程如图 1 所示。

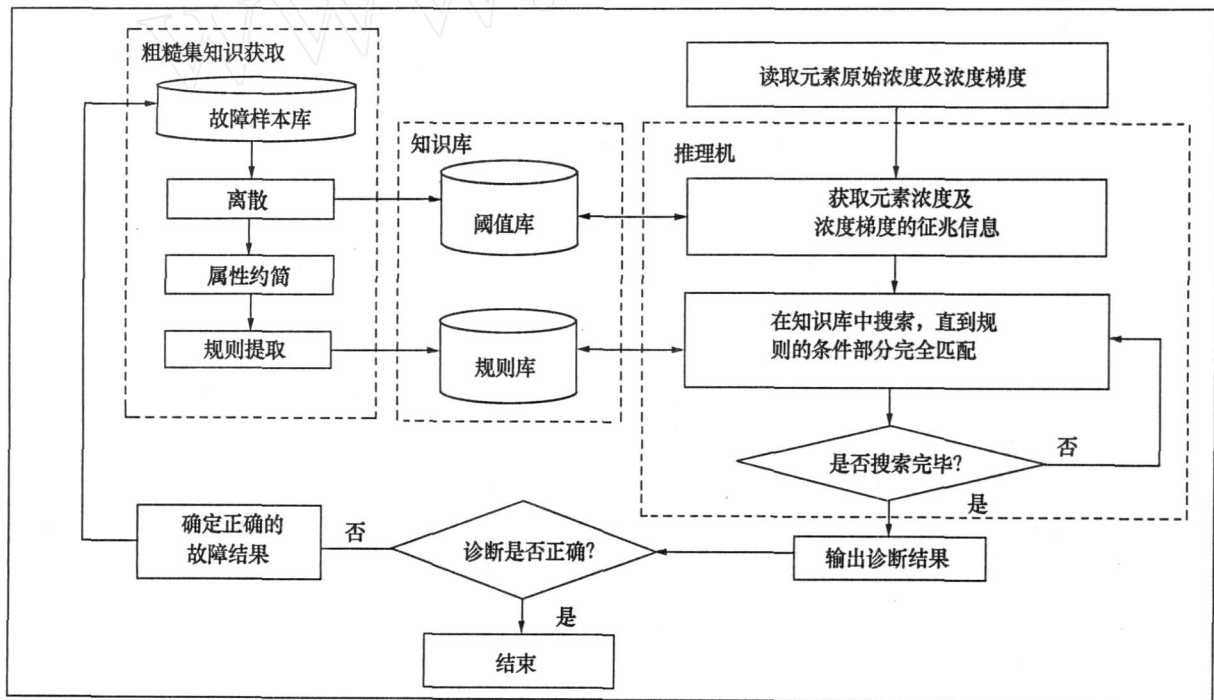


图 1 基于粗糙集理论知识自动获取的航空发动机光谱诊断专家系统

首先确定样本征兆集和故障模式集,建立由样本库构成的决策表,然后用粗糙集理论中数据补齐、属性约简、值约简及规则提取等方法进行规则提取,将获取的知识规则存入知识库。对于待诊断样本,将征兆特征值输入到推理机,按一定推理机制经过推理即可得到诊断结果。诊断样本得到验证后,又可将其作为案例,按一定原则输入到原始样本数据库中进行存储,进行增量式学习。如此往复,随着航空发动机油样光谱分析样本的不断积累,知识库的知识不断积累和更新,专家系统的诊断精度也将逐渐提高。

2 航空发动机光谱诊断专家系统中知识获取的关键技术

2.1 连续属性离散

粗糙集理论只能对离散数据进行处理,而光谱

分析数据为连续量,因此,需要离散处理。

定义 1 决策表。决策表是一个通过信息表来进行知识表达的系统,表的列表示属性,表的行表示实例对象。一般来讲,决策表由 4 部分构成, $S = U, R, V, f$, 其中 U 是论域, $R = C \cup D$ 是属性集合,子集 C 和 D 分别称为条件属性和决策属性。 V 为属性值域, $f: U \times R \rightarrow V$ 为一个信息函数,指定了 U 中每一个对象的属性值。

设决策种类的个数为 $r(d)$ 。属性 a 的值域 V_a 上的一个断点可以记为 (a, c) , 其中 $a \in R, c$ 为实数集。在值域 $V_a = [l_a, r_a]$ 上的任意一个断点集合 $\{(a, c_1^a), (a, c_2^a), \dots, (a, c_{k_a}^a)\}$ 定义了 V_a 上的一个分类 $P_a = \{[c_0^a, c_1^a), [c_1^a, c_2^a), \dots, [c_{k_a}^a, c_{k_a+1}^a]\}$ 。

因此,任意的 $P = \{P_a\}_{a \in R}$ 定义了一个新的决策表

$S^p = (U, R, V^p, f^p), f^p(x_a) = i \Leftrightarrow f(x_a) \in [c_i^a, c_{i+1}^a)$, 对于 $x \in U, i \in \{0, \dots, k_a\}$, 即经过离散后, 原来的信息系统被一个新的信息系统所代替。

在光谱分析数据中, 通常是根据元素的警告和异常界限值将浓度及浓度梯度离散为“正常”、“警告”和“异常”3类。这也是一种离散方法, 但其需要实现确定界限值, 但通常界限值的确定非常困难, 需要建立在大量的实验和统计基础上, 而且需要事先假设其统计分布规律。

本文从数据本身出发, 利用粗糙集理论的离散方法进行连续属性离散处理。目前, Nguyen S. H. 和 Skowron 提出的布尔逻辑与粗糙集理论相结合的离散方法^[13]是粗糙集理论中离散化思想的重大突破, 其基本思想是首先在保持信息系统的不可分辨关系不变的前提下, 尽量以最小数目的断点把所有实例的分辨关系区分开, Nguyen S. H. 和 Nguyen H. S. 在此基础上提出了贪心算法^[14], 大大降低了计算的空间和时间复杂度, 因此, 本文利用贪心算法来实现属性离散。

2.2 属性约简

定义 2 不可分辨关系。对于每个属性子集 $B \subseteq R$, 定义不可分辨二元关系 $ND(B)$, 即: $ND(B) = \{(x, y) \mid (x, y) \in U^2, \forall b \in B (b(x) = b(y))\}$ 。在粗糙集理论中, 不可分辨关系是定义其他概念的基础。

定义 3 基本集。由论域中相互间不可分辨的对象组成的集合, 是组成论域知识的颗粒。

定义 4 属性约简。如果 $B \subseteq A$, 属性集 A 和属性集 B 相对于决策属性的分类一致, 也就是具有相同的分类能力, 这就称 B 为 A 的相对约简。对于论域 U, P 和 Q 为定义在 U 上的两个等价关系簇且 $Q \subseteq P$ 。如果 $ND(Q) = ND(P)$; Q 是独立的, 则称 Q 是 P 的一个绝对约简。对于论域 U, P 和 Q 为定义在 U 上的两个等价关系簇, P 的所有 Q 不可省略的原始关系簇称为 P 的 Q 核, 记为 $CORE_Q(P)$ 。如果记 P 的所有 Q 约简关系簇为 $RED_Q(P)$, 则有 $CORE_Q(P) = RED_Q(P)$ 。

属性约简的目的就是保留重要的属性, 消除这些冗余属性, 因此属性重要程度是对信息决策表进行约简的重要依据。在粗糙集理论中, 对于一个属性重要程度的评判完全依赖于论域中的样本, 不依赖于先验知识。由于获取所有的属性约简被证明是一个 N-P 完全问题, 因此, 属性约简需要用启发性算法来实现。本文采用基于辨识矩阵的约简方法^[11]。

2.3 值约简与规则提取

定义 5 值约简。对于属性约简后的决策

表, 仍然含有部分冗余信息, 对于规则集中的每条规则的任意条件属性, 如果去掉该条件属性, 该规则不和规则集中的其他规则冲突, 则可以从该规则中去掉该条件属性。经过这样处理的规则集中所有规则均不含有冗余条件属性, 这一过程即为值约简。

定义 6 决策规则。对于决策表 $S = (U, R, V, f, \{a_1, a_2, \dots, a_n\}, P)$, 则公式 $(a_1, v_1) \wedge (a_2, v_2) \wedge \dots \wedge (a_n, v_n)$ 称为 P 基本公式, 如果 A 是 P 基本公式且 $B = (d, d_i)$, 则 $A \rightarrow B$ 为基本决策规则。

同样由于值约简也被证明是一个 N-P 完全问题, 因此也需要用启发性算法实现。目前已经提出了很多值约简方法, 如启发式值约简算法, 归纳属性值约简法, 基于辨识矩阵的值约简方法及缺省规则提取方法等。本文采用启发式值约简算法^[18]。规则提取后可以直接存入知识规则库, 航空发动机知识规则中应该包含发动机编号、发动机型号、飞机编号、飞机型号、规则编号、规则条件、规则结论、规则可信度及规则覆盖度等字段。

2.4 推理机

定义 7 规则绝对覆盖度及可信度。对于决策表 $S = (U, R, V, f)$, 决策规则 $A \rightarrow B$ 的不确定性可以用参数对 (α, β) 来表示, 其中 $\alpha = |X \cap Y|, \beta = |X|$, 规则可以表示为如下形式: $A \rightarrow B \mid (\alpha, \beta)$, 其中, $X = \{x \mid x \in U, A_x\}, Y = \{x \mid x \in U, B_x\}$ 。 A_x 表示实例 x 的条件属性值满足公式 A, B_x 表示实例 x 的决策属性值满足公式 B , 这里, 参数 α 表示了该规则在决策表中的(绝对覆盖度), β 就是该规则的可信度。

定义 8 多数优先的规则推理方法。多数优先的规则选择策略就是认为覆盖多数样本的规则(即根据多个样本得到的规则)具有更大的适应性, 具有得到合适结论的更高概率。其基本思想是: 假设有两条不一致的规则 R_1 和 R_2 同时与一个待识别样本匹配, 则:

若 $\alpha_1 / \beta_1 = \alpha_2 / \beta_2$, 则 $\beta = \beta_1, \beta_1 = \max\{\beta_i \mid i = 1, 2\}$; 若 $\alpha_1 / \beta_1 > \alpha_2 / \beta_2$, 则

(1) 若 $\beta_1 = \beta_2$, 则 $\beta = \beta_1, \beta_1 = \max\{\beta_i \mid i = 1, 2\}$; 即在频度一样的情况下选择可信度较大的那条规则的结论;

(2) 若 $\beta_1 < \beta_2$, 则若 $\alpha_1 > \alpha_2, \alpha_1 / \beta_1 > \alpha_2 / \beta_2$ (出现频度大的规则可信度高), 则 $\beta = \beta_1$; 若 $\alpha_1 > \alpha_2, \alpha_1 / \beta_1 < \alpha_2 / \beta_2$ (出现频度大的规则可信度低), 则选择 $\beta = \beta_2, \beta_2 = \max\left\{\frac{\alpha_i}{\beta_i} \mid i = 1, 2\right\}$ 。

在推理机设计中需要考虑规则的冲突解决问题。规则推理通常采用多数优先的推理机制, 其



匹配原则是首先选择可信度最高的规则,当最高可信度的规则有多条时,选取覆盖度最大(支持规则的样本数最多)的规则。而规则的可信度和覆盖度在规则提取中能够计算出来。因此定义正确识别为:根据多数优先的推理机制得到的规则结论正确;定义错误识别为:根据多数优先的推理机制得到的规则结论错误;拒识定义为:测试样本在规则集中找不到与之相匹配的规则;规则集中有多条规则与测试样本相匹配,具有最高可信度的规则有多条,且他们的支持样本数相同,同时规则结论不都一样。

在对具体的新样本进行诊断时,首先将原始待诊断样本的数值与阈值库中的阈值进行比较以实现连续值的离散。该阈值库中的阈值为粗糙集离散算法获得的断点;再由离散值形成征兆条件输入推理机;最后,运用推理机的推理机制,在知识库中进行搜索,获得所匹配的规则。

3 诊断实例

本文以某军用航空发动机油样分析数据为例,该数据包含了 10 台航空发动机在正常状态下和磨损状态下的 237 个样本。Fe、Al、Cu、Cr、Ag、Ti、Mg 这 7 种元素的含量作为样本实例的条件属性分别对

应于 ($A_1 \sim A_7$)。磨损状态“F”分为:“1”——正常状态、“2”——轴间轴承磨损、以及“3”——轴间轴承磨损且保持架断裂 3 种形式。磨损状态“F”作为实例的决策属性 D。

表 1 为其中的部分数据。用随机选取其中的 107 个样本作为训练样本进行规则提取,用其余的 130 个数据作为测试样本对规则进行验证。表 2 为利用贪心算法对训练样本进行属性离散的结果。表 3 为用表 2 中的断点对训练样本集进行离散后得到的结果。用基于辨识矩阵的约简方法进行属性约简得到约简后的属性集为 A_1 和 A_3 ,再利用启发式值约简算法进行规则提取,得到如表 4 的规则提取结果。表 5 为用测试样集对规则进行验证的结果。结果充分表明了所提取规则的正确性。本文利用粗糙集理论对光谱数据进行规则提取,用三条规则,两个属性就表达了数据中所蕴涵的规律。实现了对数据冗余特征的压缩和信息浓缩。

显然,将基于粗糙集理论的知识获取方法应用于航空机润滑油监控专家系统,实现知识的自动获取,利用粗糙集理论获取的知识规则直接写入知识库,避免了传统的机械学习方法,将有效地突破了知识获取的瓶颈问题。大大提升专家系统的智能化水平和知识获取能力。

表 1 光谱油样分析部分原始数据

Fe (A_1)	Al (A_2)	Cu (A_3)	Cr (A_4)	Ag (A_5)	Ti (A_6)	Mg (A_7)	F (D)
4.8	0.0	1.5	0.2	0.1	1.0	6.1	1
15.6	0.5	2.4	1.4	0.5	1.1	7.2	2
4.0	0.0	1.6	0.2	0.4	1.1	4.2	1
8.8	0.0	1.9	0.4	0.5	1.4	4.7	2
3.5	0.2	0.8	0.3	0.2	1.3	5.4	1
8.1	0.7	1.1	0.4	0.8	1.2	8.4	2
4.3	0.2	1.2	0.3	0.3	0.8	1.4	1
32.3	4.7	5.8	6.2	2.1	10.5	1.4	2
5.8	0.9	1.7	0.3	0.7	1.7	8.9	1
23.9	1.8	9.8	1.1	1.8	1.9	9.3	3
5.1	0.0	0.9	0.4	0.3	1.1	5.2	1
19.2	0.6	1.8	1.5	0.4	1.3	5.6	3
5.8	0.0	3.2	0.2	0.6	1.5	8.3	1
11.7	0.4	4.7	0.7	0.9	1.7	8.4	2

表 2 光谱元素离散断点结果

离散方法	Fe(A ₁)	Al(A ₂)	Cu(A ₃)	Cr(A ₄)	Ag(A ₅)	Ti(A ₆)	Mg(A ₇)
断点数	1	0	1	0	0	0	0
断点值	6.8500	—	7.8000	—	—	—	—

表 3 光谱元素离散结果

F	实例数	Fe(A ₁)	Al(A ₂)	Cu(A ₃)	Cr(A ₄)	Ag(A ₅)	Ti(A ₆)	Mg(A ₇)
1	102	1	1	1	1	1	1	1
2	4	2	1	1	1	1	1	1
3	1	2	1	2	1	1	1	1

表 4 规则提取结果

规则 1	$(A_1 = 1) \geq (D = 1) [1.00] [102, 102]$ 即: If Fe元素含量 < 6.85 Then 磨损状态为正常 (可信度 = 100%)
规则 2	$(A_1 = 2) \& (A_3 = 1) \geq (D = 2) [1.00] [4, 4]$ 即: If Fe元素含量 > 6.85 AND Cu元素含量 < 7.8 Then 轴间轴承磨损 (可信度 = 100%)
规则 3	$(A_3 = 2) \geq (D = 3) [1.00] [1, 1]$ 即: If Cu元素含量 > 7.8 Then 轴间轴承磨损且保持架断裂 (可信度 = 100%)

表 5 规则对测试样本的验证结果

实例数	样本数	识别数	误识数	拒识数	识别率	误识率	拒识率
测试样本集	130	129	1	0	99.2%	0.8%	0

4 结论

(1) 针对基于航空发动机光谱油样分析的磨损故障诊断,建立了基于粗糙集理论的航空发动机滑油监控专家系统知识自动获取模型。

(2) 介绍了包括连续属性离散、属性约简、值约简与规则提取等粗糙集理论知识获取的关键算法,同时,建立了专家系统的知识规则库和推理机。并运用实际的 107组航空发动机滑油光谱数据进行了知识获取,130组数据对规则进行验证,验证结果充分表明了本文粗糙集理论知识获取方法的有效性。将基于粗糙集理论的知识获取方法运用于航空发动机滑油监控专家系统的知识获取,将有效地提升专家系统的智能水平和知识获取能力。

[参考文献]

[1] 干敏梁,左洪福,杨忠等. 时序建模方法在滑油光谱分析中的应用 [J]. 光谱学与光谱分析, 2000, 20(1): 64 ~ 67
 [2] 杨虞微,陈果. 光谱油样分析监测技术中的神经网络预测法 [J]. 光谱学与光谱分析, 2005, 25(8)
 [3] 陈果,左洪福. 基于神经网络的机械磨损故障光谱定位诊断法 [J]. 摩擦学学报, 2004, 24(3)
 [4] Jayachandran T. Statistical Methods for the Joint Oil Analysis

Program [R]. ADA111736, 1982
 [5] FSILtd. Extended Diagnostic & Maintenance System [Z]. FSILtd, Canada, 1992
 [6] Bird Ltd. Predict Maintenance Program: Manual I & II [Z]. Bird Ltd, 1992
 [7] Kincaid R L. Advanced maintenance management: an expert system of applied tribology [A]. International Society of Tribology[C], 1993
 [8] 黄碧华,裘崇伟,谢友柏. 柴油机磨损监测及故障诊断专家系统知识库建立的研究 [J]. 摩擦学学报, 1994, 14(1)
 [9] 宋兰琪,汤道宇,陈立波,毛美娟. 航空发动机滑油光谱专家系统知识库建立 [J]. 航空学报, 2000, 21(5)
 [10] Pawlak Z. Rough set [J]. International Journal of Information and Computer Science, 1982, 11(5): 341 ~ 356
 [11] 王国胤. Rough集理论与知识获取 [M]. 西安交通大学出版社, 2001
 [12] 吴今培,肖建华. 智能故障诊断与专家系统 [M]. 北京:科学出版社, 1997
 [13] Nguyen H S, Skowron A. Quantization of real values attributes, rough set and Boolean reasoning approaches [A]. Proceeding of the 2nd Joint Annual Conference on Information Science [C], Wrightsville Beach, Nc, 1995: 34 ~ 37
 [14] Nguyen SH, Nguyen H S. Some efficient algorithms for rough set methods [A]. In: Proc of the Conference of Information Processing and Management of Uncertainty in Knowledge-Based Systems [C]. Granada, Spain, 1996: 1451 ~ 1456