

基于案例推理和粗糙集的商用飞机故障诊断研究

晏震乾¹, 曹磊^{2,3,4}, 陈金¹

- (1. 上海民用飞机健康监控工程技术研究中心, 上海 200241; 2. 北京航天测控技术有限公司, 北京 100041;
3. 北京市高速交通工具智能诊断与健康重点实验室, 北京 100041;
4. 轨道交通装备全寿命周期状态监测与智能管理技术与应用北京市工程实验室, 北京 100041)

摘要: 为了实现国产商用飞机维修和自动化故障诊断的需要, 将商用飞机在役维修积累的维修故障描述记录文档进行信息化和智能化故障诊断; 采用基于案例推理和粗糙集结合的方法进行国产商用飞机故障诊断的研究; 其中, 用 MongoDB 存储飞机故障案例库, 通过构建中文航空词本进行同义词替换减少故障案例描述的中文表述差异, 使用粗糙集表示故障案例, 粗糙集动态决策计算案例的属性权重并结合海明距离实现案例检索, 最后是案例库学习扩充案例库的案例数量和质量, 提高基于案例推理国产商用飞机故障诊断系统的可靠性; 并通过实验重点验证粗糙集动态决策计算案例的特征属性权重, 和结合海明距离相似数计算两个案例间的相似性, 证明基于案例推理和粗糙集进行国产商用飞机故障诊断的可行性和有效性。

关键词: 基于案例推理; 粗糙集; 海明距离; MongoDB; 同义词替换

Research on Fault Diagnosis of Commercial Aircraft Based on Case Base Reasoning and Rough Set

Yan Zhenqian¹, Cao Lei^{2,3,4}, Chen Jin¹

- (1. Shanghai Engineering Research Center of Civil Aircraft Health Monitoring, Shanghai 200241, China;
2. Beijing Aerospace Measure & Control Corp. Ltd, Beijing 100041, China;
3. Beijing Key Laboratory of High-speed Transport Intelligent Diagnostic and Health Management, Beijing 100041, China;
4. Beijing Engineering Laboratory of Rail Transportation Equipment Life Cycle Condition Monitoring and Intelligent Management Technology and Application, Beijing 100041, China)

Abstract: In order to realize the needs of domestic commercial aircraft maintenance and automatic fault diagnosis, the maintenance fault description record files accumulated by the commercial aircraft in service maintenance are used for information and intelligent fault diagnosis. A case-based reasoning and rough set combination method is used to study the fault diagnosis of domestic commercial aircraft. Among them, MongoDB is used to store the aircraft failure case database, and the Chinese aviation wordbook is used to replace synonyms to reduce the difference in the Chinese expression of the fault case description. The rough set is used to represent the fault case. The rough set determines the attribute weight of the case and combines the Hamming distance for case retrieval. Finally, the case library learns to expand the number and quality of cases in the case library, and improve the reliability of the fault diagnosis system for domestic commercial aircraft based on case-based reasoning. And through experiments, the feature attribute weights of the rough set dynamic decision calculation case and the similarity between the two cases combined with the Hamming distance similarity number are calculated to prove the feasibility and effectiveness of fault diagnosis of domestic commercial aircraft based on case reasoning and rough set.

Keywords: Case-based reasoning; Rough set; Hamming distance; MongoDB; Synonym replacement

0 引言

商用飞机的维修和自动化故障诊断, 具有非常的技术意义和商业价值。本文着重研究用于商用飞机的故障诊断技术, 助力中国商用飞机的维修^[1]与自动化故障诊断技术的积累和维修成本的节约。本文将采用基于案例推理和粗

糙集结合的方法进行商用飞机故障诊断的全流程一体化解决方案。其中, 使用 MongoDB 存储故障案例库, 通过构建中文航空词本进行同义词替换减少故障中文描述的差异, 粗糙集表示飞机故障案例且粗糙集动态决策计算故障案例的特征属性权重并结合海明距离来度量案例间的相似度实现案例的检索, 最后是案例库学习实现案例的扩充, 提高系统的可靠性以及技术流程的闭环。

1 基于案例推理和粗糙集实现飞机故障诊断

1.1 基于案例推理

基于案例推理^[2-4] (case-based reasoning, CBR) 是指

收稿日期: 2020-06-05; 修回日期: 2020-06-15。

作者简介: 晏震乾(1982-), 男, 湖南益阳人, 研究生, 高级工程师, 主要从事民用飞机健康监控、飞机状态监控机载软件客户化、飞行大数据分析方向的研究。

在解决新案例时，通过在案例库中检索过去已解决的类似案例，并比较新旧案例的异同，推理出新案例的解决方法的一种决策推理方法，常用于决策系统、专家系统中。

推理新案例时，先将新案例通过案例表示输入基于案例推理系统中，通过相似度计算检索出与目标故障案例相似的案例集，若有与目标故障案例描述完全一致的案例，则将其故障排除方法着重突出提交给用户；若没有则把目标案例的相似案例按相似程度高低进行排序输出给用户，若实践验证成功则对此次案例诊断进行评价打分并加入案例库，完成以此次故障诊断。基于案例推理的故障诊断流程如图 1 所示。

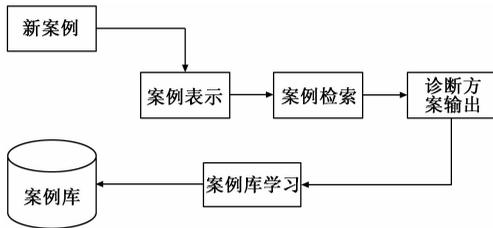


图 1 基于案例推理的故障诊断过程

基于案例推理的关键在于案例库的数量和质量。随着案例库中案例的增量式学习，案例库的覆盖程度随着系统的使用会逐渐增大，新案例推理的效果就会越来越好。

1.2 粗糙集

粗糙集^[5-7]是用来研究不完整数据，不精确知识的表达、学习，归纳等的一套理论方法。它是一种新的处理模糊和不确定性问题的数学工具，被广泛应用于知识发现、机器学习、决策支持、模式识别、专家系统及归纳推理等领域。

1.3 结合案例推理与粗糙集实现飞机故障诊断

基于案例推理进行飞机故障诊断的关键是案例表示和案例检索即案例间相似度的计算。之前的基于案例推理故障诊断系统案例间相似度的计算大多采用专家指定案例的特征属性权重去进行相似性度量，这具有很大的主观性。本文采用粗糙集表示案例进而粗糙集动态决策计算出案例的特征属性权重，避免案例间相似度计算时特征属性权重确定的主观性。

2 案例的知识表示

案例的知识表示是案例推理的基础。首先，应确定案例应该描述哪些信息。

2.1 案例内容

从多种多样的飞机故障案例报告中抽取其最重要最有效的共同特征，将案例抽象成这些共同特征来进行案例描述。飞机故障案例 1 如图 1 所示。

案例描述内容包括：机型、故障来源、发生位置、故障现象或代码、故障排除方法等。

然后，是确定案例的存储方式，将过去已解决的飞机故障实例存储到数据库即案例库中。

案例编号	***
案例名称	***
发生时间	***
机型	***
故障来源	外站短停期间
发生位置	电子舱风扇
故障现象或代码	"VENT AVNCS SYS FAULT"、"VENT BLOWER FAN FAULT"、超控操作后，电子舱抽风扇也故障"VENT EXTRACT FAULT"，飞机不能按 MEL 放行
故障排除方法编号	***
故障排除方法	对倒抽风扇和吹风扇，清洁气滤后测试 AEVS 工作正常，可正常放行飞机。 飞机返回基地后继续完成电子舱通风故障排除工作： (1) 参照 AMM TASK 21-26-43-920-001 更换电子舱通风气滤； (2) 参照 AMM TASK 21-26-51-000/400-001-A 更换电子舱抽风扇，自此故障彻底排除。
解释	气滤故障和太脏
参考资料	AMM TASK 21-26-43-920-001、AMM TASK 21-26-51-000/400-001-A

图 2 飞机故障案例 1

2.2 MongoDB 存储案例

传统使用结构化查询语言 SQL 把故障案例的各特征编码成传统的数据结构保存在例如 MySQL 或 Oracle 等关系型数据库中，其具有牵涉信息的编解码、增加复杂度和信息的不直观等缺点。

由于故障案例来自航空机修工程师一手的 pdf 格式的飞机故障案例报告文件，故而本文采用一种最新的 NoSql 非关系型数据库 MongoDB^[8] 数据库直接保存文本信息，克服上述缺点。

我们感兴趣的关键内容通过专门定制的 pdf 读取软件，将 pdf 感兴趣内容读取后以如下 BSON 的形式存储在 MongoDB 数据库中：

```

    {"案例编号": "xx",
    "案例名称": "xx",
    "发生时间": "xx",
    "机型": "xx",
    "故障来源": "xx",
    "发生位置": "xx",
    "故障现象或代码": "xx",
    "故障排除方法编号": "xx",
    "故障排除方法": "xx",
    "解释": "xx",
    "参考资料": "xx"}注:内容略
  
```

飞机故障诊断案例一般主要由故障现象及故障排除方法两部分组成。对飞机故障案例进行知识表示时，有必要联系案例的上下文环境，即机型、故障来源和发生位置。因为是通过搜索故障现象得到目标故障排除方法实现故障诊断功能，所以检索的故障案例描述有（机型；故障来源；发生位置；故障现象），其中故障现象中又有不确定数量的特征描述。

最后，是从案例数据库中读取的案例，该以什么方式表示，并有利于后续实现案例间的检索。

2.3 粗糙集表示案例

粗糙集用 $S = (U, A, V, f)$ 表示，其中 U 是非空有限对象集，称为论域， A 是非空有限属性集， $V = \cup_{a \in A} V_a$ ，

V_a 是属性 a 的值域, $f: U \times A \rightarrow V$ 是一个信息函数。对于任意 a 属于 A , 并且 x 属于 U , $f(x, a) \in V_a$ 。如果 $A = C \cup D$, 则 C 是条件属性集, D 是决策属性集, $S = (U, A, V, f)$ 也被称为决策表。

以上述飞机故障案例 1 为例, 此案例为 U (只有一个案例的案例集), 机型、位置、故障现象等故障描述为条件属性集 C , 故障排除方法为决策属性集 D , V 为属性的具体取值, f 为粗糙集系统归纳总结的决策规则。

3 案例的检索

基于案例推理关键环节在于相似案例的检索, 案例的检索是一个查找与匹配的过程, 即从案例库中搜索出与搜索案例相似的有帮助的案例集供用户查阅。

3.1 同义词替换

同义词替换作为自然语言处理数据预处理的一种手段, 在比较两个中文句子的相似性时, 利用同义词替换技术把几个同义词转换为同一个词, 可以提高中文相似度计算的可靠性, 也就提高了故障描述相似度计算的可靠性。

由于中文语言的博大精深与表述多元化, 同一事物的描述词汇较之英文更多种多样, 不同航空工程师对故障现象的描述也会使用不同的近义词。比如: 供气与通风、排气与抽风、二号发动机与右发。

所以, 首先需要对故障案例的中文描述语言进行分词, 然后再利用同义词词本进行同义词替换^[9]。

同义词词本一方面可以来自通用中文同义词词林, 比如哈尔滨工业大学的大词林^[10], 并一方面, 航空领域的专业词汇通常高频出现在故障案例描述中, 其呈现高频和有限的特点, 应该由航空领域专业人士构建并维护一个航空领域专用同义词词本“同义词航空领域专用词本.txt”并随时收集与增改。

其词本内容格式及示例如下:

- 右发|二号发动机|右发动机
- 故障|正常位 OFF 灯亮
- 钢缆|钢索
- 通风|供气
- 抽风|排气
- 地面|外站短停

每行为一个同义词序列, 同义词之间用指定分隔符“|”隔开, 每行后面的同义词都会替换为该行第一个指定的最优表述词。

由于飞机故障大多是由指示灯指示的, 而指示灯出于惯性大多是英文, 而英语的分词很简单, 就是空格和标点, 但需要中英描述之间的匹配, 此时案例检索时遇到英文要翻译成中文再进行中文故障描述的匹配。例如: VENT AVNCS SYS FAULT (电子舱航电系统故障)、VENT BLOWER FAN FAULT (电子舱通风扇故障)、VENT EXTRACT FAULT (电子舱抽风扇故障)。

3.2 海明距离计算相似度

向量或样本之间的相似性度量, 一般通过计算两个向

量之间的距离, 即转化为数学上的向量之间距离的计算, 它可以反映出事物间在距离上接近或远离的程度。案例的故障描述视为一个粗糙集, 粗糙集间的相似度可以用案例描述属性间的海明距离来测量。

信息编码学中, 海明距离是两个码流对应位上编码不同的位数, 比如比特流 1011 与 1110, 第二位与第四位不同, 其海明距离为 2。本文将海明距离取反作为相似度衡量标准, 记为海明相似数, 比如特征描述 1 {特征 A、特征 B、特征 C} 和特征描述 2 {特征 A、特征 D、特征 E}, 特征描述 1 与特征描述 2 中, 有一个相同的特征 A, 其海明相似数为 1。海明相似数越大, 案例特征描述越匹配, 案例间越相似。

案例编号	***
案例名称	***
发生时间	***
机型	***
故障来源	地面
发生位置	电子舱风扇
故障现象或代码	供气正常位 OFF 灯亮, 排气正常位 OFF 灯亮
故障排除方法编号	***
故障排除方法	航前过站可按 MEL 放行的情况: 一个风扇可以不工作, 如果核实其余的风扇工作正常且核实两个低流量探测器工作正常。 飞机返回基地后: 更换供气和排气风扇, 或更换或清洁低流量传感器
解释	供气气滤故障和太脏 风扇故障 跳开关跳出
参考资料	***

图 3 飞机故障案例 2

例如上述飞机故障案例 2 (见图 3) 的故障现象为“供气正常位 OFF 灯亮, 排气正常位 OFF 灯亮”。标点分隔每一句话为一个特征描述 (特征集), 中文分词为一个特征。

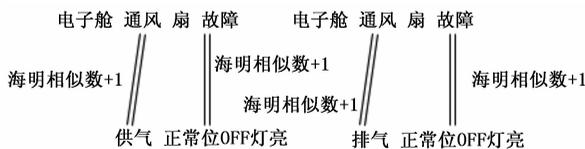


图 4 海明相似数计算

如图 4 所示, 结合航空词本同义词替换, 两个案例的故障现象描述语句相似数为 4。因为中文表达的开放又统一, 可以认为两个故障现象描述语句中有一个同义词以上 (相似数大于 1), 两个故障现象描述语句的相关性就强, 需要被检索出来, 从而把故障现象的值设为“存在”或“不存在”两种取值更有意义, 分别以 1 或 0 表示。

在案例的检索中, 案例的特征属性对搜索案例与其他案例区分时所起到的作用是不同的, 即特征权值反映了特征的不同重要程度, 因此赋予特征合适的权值对案例间相似度的计算是非常有意义的。

3.3 粗糙集确定属性权重

现有基于案例推理系统中, 属性的权重是由专家指定的, 比如故障现象、发生时间、发生位置权重分别设为

0.75、0.05、0.2，存在非常大的主观性。事实上，发生时间、发生位置、故障现象，同为故障描述的三者，本质上并没有高低轻重之分。所以，采用自动地属性权重计算方法，通过粗糙集的决策规则自动提取功能，动态决策计算出属性的权重。

粗糙集理论中，属性重要度反映去掉某个属性后决策的变化程度。设属性 $\{a\}$ 为条件属性集 C 的一个属性，属性 $\{a\}$ 关于条件属性集 C 相对于决策属性集 D 的属性重要度为：

$$Sig(a, C, D) = \gamma_C(D) - \gamma_{C-\{a\}}(D)$$

其中： $\gamma_C(D)$ 表示决策属性 D 对条件属性集 C 的依赖程度。

$$\gamma_C(D) = \frac{|POS_C(D)|}{|U|}$$

其中： $POS_C(D)$ 表示 D 关于 C 的正域， $|U|$ 表示 U 的数量。

而 $\gamma_{C-\{a\}}(D)$ 表示决策属性 D 对去掉属性 $\{a\}$ 的剩余条件属性集的依赖程度，两者相减即为决策 D 对属性 $\{a\}$ 的依赖程度，即属性 $\{a\}$ 对决策的权重。

4 实验结果与分析

下面先用有 6 个案例的案例库实验论证粗糙集动态决策计算出案例的特征属性权重的可行性和有效性；然后利用计算出的属性权重，结合海明距离相似数，对文中列出的两个示例案例间的相似性进行度量，表明基于案例推理飞机故障诊断系统案例检索的有效性；最后对基于案例推理和粗糙集结合的飞机故障诊断系统存在的问题和需要改进的地方进行论述。

表 1 中，设某机型 1=1，某机型 2=2，故障来源 1=飞行中，故障来源 2=地面，故障现象 1=电子舱航电系统故障，发生位置 1=电子舱航电系统，发生位置 2=电子舱风扇，故障现象 1=电子舱航电系统故障，故障现象 2=电子舱通风扇故障，故障现象 3=电子舱抽风扇故障，故障现象是否存在用 1 或 0 表示，排故方案以编号表示。

表 1 飞机故障诊断决策表

机型	属性					决策
	故障来源	发生位置	故障现象 1	故障现象 2	故障现象 3	排故方案编号
1	1	2	0	1	1	A001
2	2	1	1	0	0	B002
1	1	1	1	0	0	B002
2	2	2	1	1	1	A001
1	2	2	0	1	1	A001

按照上述粗糙集属性重要性计算公式，得出属性 $\{$ 机型 a_1 ，故障来源 a_2 ，发生位置 a_3 ，故障现象 $1a_4$ ，故障现象 $2a_5$ ，故障现象 $3a_6$ $\}$ 的权重分别为：

$$Sig(a_1, C, D) = \gamma_C(D) - \gamma_{C-\{a_1\}}(D) = \frac{1}{5} - 0 = \frac{1}{5}$$

$$Sig(a_2, C, D) = \gamma_C(D) - \gamma_{C-\{a_2\}}(D) = \frac{1}{5} - \frac{1}{5} = 0$$

$$Sig(a_3, C, D) = \gamma_C(D) - \gamma_{C-\{a_3\}}(D) = \frac{1}{5} - 0 = \frac{1}{5}$$

$$Sig(a_4, C, D) = \gamma_C(D) - \gamma_{C-\{a_4\}}(D) = \frac{1}{5} - 0 = \frac{1}{5}$$

$$Sig(a_5, C, D) = \gamma_C(D) - \gamma_{C-\{a_5\}}(D) = \frac{1}{5} - 0 = \frac{1}{5}$$

$$Sig(a_6, C, D) = \gamma_C(D) - \gamma_{C-\{a_6\}}(D) = \frac{1}{5} - 0 = \frac{1}{5}$$

即属性 $\{$ 机型 a_1 ，故障来源 a_2 ，发生位置 a_3 ，故障现象 $1a_4$ ，故障现象 $2a_5$ ，故障现象 $3a_6$ $\}$ 的权重分别为 $\{0.2, 0, 0.2, 0.2, 0.2, 0.2\}$ 。

结合上节海明距离相似数乘以属性权重，上述两个飞机故障示例案例 1 和 2 案例描述之间的相似性度量为：

$$1 \times 0.2 + 1 \times 0 + 1 \times 0.2 + 1 \times 0.2 + 1 \times 0.2 + 1 \times 0.2 = 1$$

则故障案例 2 可以采用故障案例 1 的排故方案，从而实现飞机故障诊断功能。从故障案例 1、2 的故障排除方法也可以看出，故障案例 1、2 的排故方案也确实是大体相同的。

实验结果证实了结合海明距离和粗糙集动态计算属性权重实现案例间相似性度量（即案例检索）的可行性和有效性，但由于案例库中的案例来自飞机在役飞行航空维修工程师的一手数据，获取困难。目前实验的案例库中的案例数量还不够，不仅影响粗糙集对案例的特征属性权重的动态决策的准确性，而且众所周知检索的关键、效果好坏在于内容的多寡，基于案例推理系统必须足够数量案例的案例库才能发挥其基于案例推理专家系统的效果，下一步项目实施的重点在于案例库的扩充与学习。

随着本系统的使用和故障案例的发现，故障案例库会记录越来越多的故障案例，这样可以积累国产商用飞机自己的宝贵维修故障数据，并且对商用飞机自己的故障更有针对性，基于案例推理和粗糙集的国产商用飞机故障诊断系统故障诊断的效果也会越来越好。

5 结束语

本文基于案例推理和粗糙集算法对中国商用飞机故障诊断提供了全链路的技术解决方案和实验可行性验证。一是用 MongoDB 存储商用飞机故障案例库；二是构建航空同义词词本替换案例描述中的中文同义词；三是用粗糙集表示故障案例，并用粗糙集动态决策计算故障案例描述属性的权重，并结合海明距离计算故障案例间的相似度，来实现案例检索；四是案例库扩充学习以完成解决方案的闭环和提高基于案例推理和粗糙集的故障诊断系统的可靠性与准确性。

值得指出的是，只有在实际应用中不断扩展案例库中案例的数量和质量，基于案例推理和粗糙集的商用飞机故障诊断系统才能越来越可靠和有效。