

飞机实时监控数据挖掘方法研究

朱 睿, 郭隐彪*

(厦门大学机电工程系, 福建 厦门 361005)

摘要: 通过数据挖掘技术分析与飞机运行和维护相关的数据资源来研究一套能及时察觉、分类并预报故障的飞机实时监控(Aircraft Real-Time Monitoring & Troubleshooting System, AMTS), 使其能够在优化飞机关键部件寿命的同时减少飞机运行和维护的费用。分析着手于历史维护经验库, 结合飞机维护技术文档和相关数据挖掘算法, 介绍了构建该系统的原理和方法, 并初步实现利用飞机实时故障报文对飞行状态进行实时监控及故障诊断。

关键词: 数据挖掘; 实时监控; 技术文档; 历史经验

中图分类号: V 267; V 247.5 **文献标识码:** A

文章编号: 0438-0479(2007)05-0657-04

现代的飞机已经装备了先进的机载数据采集、管理系统并具备空地数据通讯系统。飞机的运行和维护产生大量的信息和数据, 随着时间的推移, 信息和数据量越来越大, 以及相关的维护经验不断地累计, 达到了很高的数量级。然而在实际维护工作中, 这些信息和数据往往得不到充分有效地利用。如果能够有效地利用现有的相关维护资源(如中央维护计算机系统(Central Maintenance Computing System, CMCS)产生的实时报文、飞机维护技术文档, 以及历史维护经验)^[1], 利用数据挖掘技术对飞机进行实时状态监控, 及时地发现和预测飞机故障, 在飞机落地前提供相关的维护方案就能够极大地减少非例行工作量所占的比例, 从而提高维修的效率, 减少飞机运行和维护的费用以及提高飞机的利用率和飞行安全性, 实现对飞机的实时监控。

本文对数据挖掘技术在飞机实时监控方面的应用进行了有益的探索, 设计了一个基于数据挖掘技术的飞机实时监控(Aircraft Real-Time Monitoring & Troubleshooting System, AMTS)。下文从体系结构、功能模块、关键技术等方面对该系统进行了详细的介绍^[2]。

1 系统构架

1.1 AMTS 体系结构

AMTS 以飞机维护技术文档和历史维护经验为数据源, 通过数据清洗、转换、汇总、抽取等技术手段, 使用数据挖掘技术中的基于事例的推理方法(Case

Based Reasoning, CBR) 构建 CBR 数据库。数据库建成之后, 飞行实时数据在经过必要的格式化之后, 就可以实时地调用 CBR 数据库, 通过特定的数据挖掘算法, 实现发现飞机故障与维护方案同时到位的过程。实时维护方案的不断积累, 不仅可以补充历史维护经验, 同时可以生成维护方案库, 以进一步完善飞机维护技术文档。

如图 1 所示, AMTS 在实现对飞机进行实时状态监控的同时不断地使用数据挖掘技术丰富和完善自身的原始数据源, 从而实现系统的自我升级。

1.2 功能模块

AMTS 由 4 个功能模块组成, 包括: CBR 模块、维护方案模块、KDD (Knowledge Discovery in Database) 模块、数据源反馈模块。CBR 模块的任务是清洗、转换、汇总、抽取飞机维护技术文档(FIM、AMM 等)以及历史维护经验(重复故障库)中的数据和信息, 并使用 CBR 方法构建 CBR 数据库^[3]。维护方案模块可以格式化飞行实时数据(实时故障报文), 使其能够实时地调用 CBR 数据库, 通过特定的数据挖掘算法程序, 实现发现飞机故障与维护方案同时到位的过程, 并

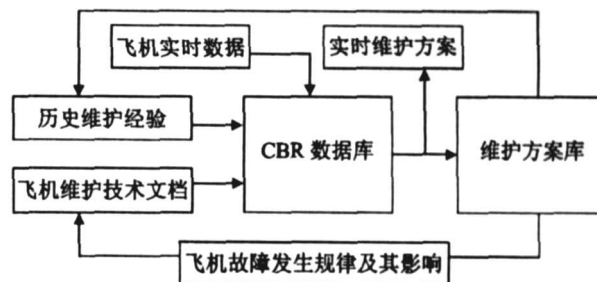


图 1 飞机实时监控系统体系结构图

Fig. 1 Aircraft real-time monitoring & troubleshooting system structure

收稿日期: 2006-11-17

* 通讯作者: guoyb@jingxian.xmu.edu.cn

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

生成维护方案库. KDD 模块基于维护方案库, 使用关联规则、异常检测、聚集等数据挖掘技术挖掘潜在的知识 and 规律. 数据源反馈模块有 2 个主要任务, 一是利用不断积累的维护方案补充和丰富历史维护经验, 二是使用 KDD 模块挖掘出的潜在知识和规律(如飞机的故障规律、影响、特征等)来完善飞机维护技术文档.

2 关键技术

2.1 数据挖掘技术

数据挖掘分为两类: 描述性数据挖掘和预言性数据挖掘. 描述性数据挖掘是以概要的方式对数据信息进行描述, 提供数据的有趣的一般性质; 预言性数据挖掘是进行数据分析, 建立一个或一组模型, 并根据模型产生关于数据的预测^[4].

在 AMTS 中, 主要采用了基于事例的推理方法、关联规则、决策树、异常检测、聚集等数据挖掘技术来实现相应的系统功能.

基于数据挖掘的分类, AMTS 的模块信息见表 1.

表 1 AMTS 模块信息表
Tab.1 AMTS modules' information

模块名称	模块类型	采用方法
CBR 模块	描述性	基于事例的推理、信息提取
维护方案模块	预言性	决策树
KDD 模块	描述性	关联规则、异常检测、聚集

2.2 CBR 模块设计

CBR 模块的任务是清洗、转换、汇总、抽取飞机维护技术文档以及历史维护经验中的数据和信息, 并使用 CBR 方法构建 CBR 数据库.

范例库是历史经验的反射集合, 但它不能像其他的知识资源如厂商技术手册一样包罗万象. 维护工程师依靠这些资源在解决新的问题方面协助他们. 在 AMTS 中, 构建范例库方法以历史排故经验为起点, 飞机维护技术文档为构建提供修正支持. 因此, 我们分两个阶段来构建 CBR 数据库. 首先, 基于飞机历史排故经验, 构建一个总体的、全面的范例库; 其次, 借助飞机维护技术文档来不断地修正范例库(图 2).

飞机制造商所提供的飞机维护技术文档(Fault Isolation Manual(FIM)、Aircraft Maintenance Manual(AMM)等)是飞机维护最主要和重要的参考信息. 重复故障库以及 AMTAC(Aircraft Maintenance Tracking And Control)维护记录则是历史维护经验的主要来源^[5]. 上述维护资源是构建 CBR 数据库的前提条

件.

CBR 数据库构建可分为以下 3 步:

(i) 范例库构建: 构造基于事例的推理系统的主要任务是确定适当的事例特征, 包括领域的专有名词的定义和专家解决的有代表意义的问题的收集. 要构建一个总体、全面的范例库, 需要邀请多位飞机维修行业的资深工程师, 依据他们多年的飞机维护经验来选取航线维护中典型飞机故障及其最有效的排故措施.

(ii) 索引机制的可操作化: 这个阶段的主要任务是为事例库中的事例的重要特征提供一个好的索引机制. 要使得范例库中的各个字段名便于检索且具有飞机维护特点以便于理解, 如 CBR_FIM_TASK 代表故障隔离任务号.

(iii) 范例库修正: 根据已有的事例库以及索引机制, 构造出完整的基于事例推理的专家系统. 当 CBR 依据一个输入检索范例库后无法找到完全匹配飞机故障时, 系统根据索引机制结合飞机维护技术文档对其最相似故障进行修正, 同时可视问题的典型程度来决定该问题是否作为一新事例存入事例库中.

在范例库构建的过程中, 我们还使用了信息提取技术(Information Extraction, IE), 它能够从一段文本中抽取指定的一类信息(事件、事实), 并将其(形成结构化的数据)填入一个数据库中供用户查询.

CBR 数据库应包含以下字段, 见表 2.

表 2 CBR 数据库结构表
Tab.2 CBR database structure

含义	字段名称	类型
故障代码	CBR_CODE	INT
故障信息	CBR_MESSAGE	INT
故障描述	CBR_DISCRIPTION	STRING
FIM 任务号	CBR_FIM_TASK	INT
AMM 任务号	CBR_AMM_TASK	INT
排故措施	CBR_ACTION	STRING
排故经验	CBR_EXPERIENCE	STRING

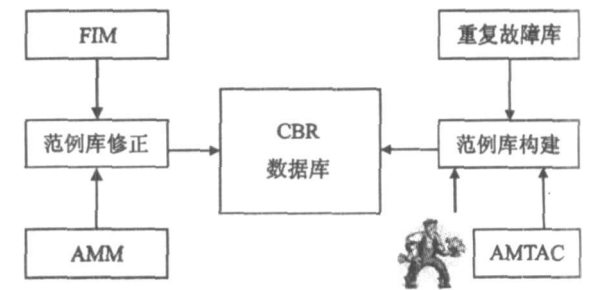


图 2 CBR 数据库构建流程图
Fig.2 CBR database construction process

2.3 维护方案模块设计

维护方案模块可以格式化飞行实时数据(实时故障报文),使其能够实时地调用 CBR 数据库,通过特定的数据挖掘算法程序,实现发现飞机故障与维护方案同时到位的过程,并生成维护方案库。

CMCS 产生的实时故障报文即(Computational Fluid Dynamics, CFD)包含了报文头、FDE (Flight Deck Effect) 部分、NCMM (Non- Correlated Maintenance Message) 部分以及 EOR(End Of Report) 标志(这是报文有效的标志)。在 CFD 报文中, FDE 部分中的故障代码和故障信息是进行飞机实时状态监控的关键,也是调用 CBR 数据库的关键。格式化飞行实时数据,就是要提取 CFD 报文中的故障代码(CFD_CODE)或故障信息(CFD_MESSAGE),使其关联 CBR 数据库中的 CBR_CODE 或 CBR_MESSAGE,并能够实时调用 CBR 数据库中的排故措施以及排故经验,通过特定的数据挖掘算法程序(建议使用 Java 编程语言),生成实时的维护方案,实现发现飞机故障与维护方案同时到位的过程,并生成维护方案库(图 3)^[6]。在维护方案模块设计中,我们使用了决策树算法。决策树算法有很好的可解释性,易于理解,所以在 AMTS 中用来建立预言模块,生成维护方案。

决策树中最上面的节点称为根节点,是整个决策树的开始。本文中根节点是“CFD 报文是否有效”,对此问题的不同回答产生了“是”和“否”两个分支。每个分支要么是一个新的决策节点,或者是树的结尾,称为叶子。在沿着决策树从上到下遍历的过程中,每个节点都会遇到一个问题,对每个节点上问题的不同回答导致不同的分支,最后会到达一个叶子节点,这个过程就是利用决策树进行分类的过程。

2.4 KDD 模块设计

KDD 模块基于维护方案库,使用关联规则、异常检测、聚集等数据挖掘技术挖掘潜在的知识 and 规律。

维护方案的不断积累为进一步使用数据挖掘技术提供了大量的数据资源。维护方案中包含了各种故障现象、发生原因、具体排故措施等元素,相同的故障发生原因可能有不同的故障现象,本文主要利用关联规则算法,结合异常检测、聚集等数据挖掘技术来发现其故障本质。

关联规则算法思路如下:假设某飞机故障 A 出现的概率为 45%,故障 B 出现的概率为 42.5%,故障 C 出现的该概率为 40%,B 和 C 同时出现概率 15%,A、B、C 同时出现概率为 5%。

支持度:就是一个故障事件在整个数据库中出现的概率。如表 1 的例子中 $S(A)=0.45$ 。

可信度:它是针对规则而言的。对于一般的规则,它的可信度= $p(\text{condition and result})/p(\text{condition})$ 。例如有如下规则: If B and C then A, 则它的可信度是: $p(B \text{ and } C \text{ and } A)/p(B \text{ and } C)=5\%/15\%=0.33$ 。

兴趣度:对于上面的一个规则,可以发现,当从数据库中直接取 A 的时候,概率是 45%;但在我们的规则中,取到 A 的概率却只有 33.3%,这种情况是我们不愿意见到的,因此应该略去这样的一些规则。所以我们引入了兴趣度的概念,具体的公式如下:兴趣度= $p(\text{condition and result})/p(\text{condition}) * p(\text{result})$ 。当兴趣度大于 1 的时候,这条规则就是比较好的;当兴趣度小于 1 的时候,这条规则就是没有很大意义的。兴趣度越大,规则的实际意义就越好^[7]。在该模块中,运用计算机语言程序来实现上述关联规则算法,计算飞机故障之间的频繁模式、关联、相关性、或因果结构,通过分析数据或记录间的关系,确定数据库中飞机故障的发生条件或原因。

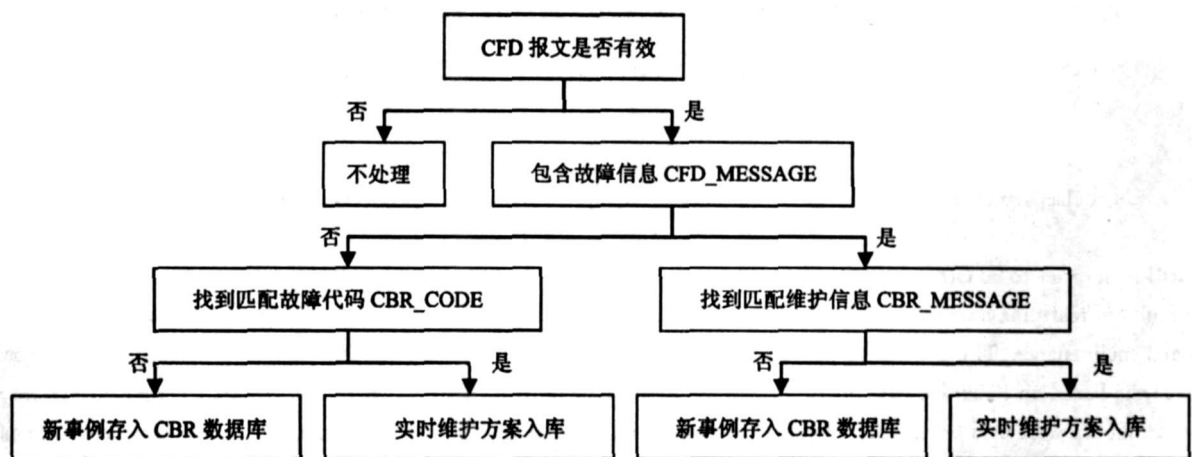


图 3 维护方案决策树算法

Fig. 3 Decision tree arithmetic for maintenance project

采用异常检测算法为每一个维护方案计算一个异常因子,该异常因子越大表示这条记录的异常越大,以突出该故障原因.利用聚集算法,将维护方案分为不同的聚集,在同一聚集内部的维护方案具有相同的故障原因,不同聚集之间的故障原因差异较大.

2.5 数据源反馈模块设计

利用不断积累的维护方案补充和丰富历史维护经验以及使用 KDD 模块挖掘出的潜在知识和规律来完善飞机维护技术文档是数据源反馈模块的主要功能.主要依靠专家意见并通过一定的数据库(如 Oracle)语言程序实现其模块功能.

2.6 基于数据挖掘飞机实时监控系统的初步实现

根据上述步骤构建基于数据挖掘飞机实时监控系统的软件雏形已初步实现,该系统可以实时读取飞机机载系统所产生的故障代码和维护信息,明确飞机运行状态,及时发现飞机故障并明确故障等级,并以故障代码和维护信息作为输入源,查询 CBR 数据库寻找匹配历史排故经验措施,在飞机落地前提供相应维修方案以提高维修效率;在查询不匹配的情况下借助飞机维护手册对该故障事例进行修正由用户决定是否将其作为新事例存入事例库.

3 结束语

数据挖掘是一个多学科交叉的新型研究领域,是近几年来研究的热点.在这个新兴领域中,汇集了来自机器学习、模式识别、数据库、统计学、人工智能等各学科的成果,多元化的投入,使得这一学科蓬勃发展.本文对其基本知识做了介绍,并对其在飞机实时监控中的应用进行了研究.数据挖掘技术应用于民用航空,目

前才起步不久,还存在着许多问题有待于进一步深入和广泛的研究.我们将基于飞机维护的实际工作,不断地积累经验,进一步改进算法并完善现有的模块.

参考文献:

- [1] 张学军,罗喜伶,张其善,等.中国国际航空公司 ACMS 系统[EB/OL]. 2002-08-22. http://www-900.ibm.com/cn/services/news/news_2002082601.shtml.
- [2] Michael Halasz, Francois Dub, Bob Orchard. IDS: remote monitoring and decision support for commercial aircraft-putting theory into practice[C]//Integrated Reasoning Group National Research Council of Canada Institute for Information Technology. Canada: Commercial Aviation Publishing company, 1999: 37.
- [3] Amit Mathur. Data mining of aviation data for advancing health management[C]//SPIE's 16th International Symposium on AeroSense, Aerospace/Defence Sensing, Simulation and Controls. American: Aerospace& Defence Publishing Company, 2002: 48.
- [4] Amit Mathur. Data mining of aviation data for advancing health management[C]//SPIE's 16th International Symposium on Aero Sense, Aerospace/Defence Sensing, Simulation and Controls. American: Aerospace& Defence Publishing Company, 2002: 153.
- [5] 许春生 选编. 发动机机队管理[M]. 天津: 中国民航出版社, 1999: 10.
- [6] Air China Group. B777 fault isolation manual[M/CD]. D633W103-BEJ. Washington: Boeing Commercial Airplanes Group, 1999.
- [7] Marvin Zaluski. Case authoring from text and historical experiences[EB/OL]. 2003-07[2004-04]. <http://www.site.uottawa.ca/~nat/Papers/692aluski.doc>.

Study for the Application of Data Mining in Aircraft Real-time Monitoring

ZHU Rui, GUO Yin-biao*

(Department of Mechanical and Electrical Engineering, Xiamen University, Xiamen 361005, China)

Abstract: In order to construct an aircraft real-time monitoring and troubleshooting system that could detect, classify, and predict aircraft faults in real-time, this paper presented the applicability of data mining technology for analyzing data generated by aircraft operation and maintenance. The system could reduce operation and maintenance costs while optimizing critical aircraft components' lives. Analysis based on historical experiences and technical documents, principles and processes of constructing the system with data mining arithmetics were also presented. The primary system could use aircraft real-time faults reports to monitor and diagnose aircraft condition.

Key words: data mining; real-time monitoring; technical documents; historical experiences