

旨在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,研究一套能及时察觉、分类并预报故障的飞机健康管理系统,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。

应用于飞机健康管理的数据挖掘方法研究

The Research for the Application of Data Mining in Aircraft Health Management

◎朱睿 / 厦门大学

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信系统。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,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的利用。如果能够有效地利用现有的相关维护资源如 CMCS (Central Maintenance Computing System) 产生的实时报文、飞机维护技术文档,以及历史维护经验,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,实现对飞机的健康管理。

本文对数据挖掘技术在飞机健康管理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健康管理系统 AHMS (Aircraft Health

Management System)。下面从体系结构、功能模块、关键技术等方面对该系统进行了详细的介绍。

系统构架

1. AHMS 体系结构

AHMS 以飞机维护技术文档以及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 (CBR-Case Based Reasoning) 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。接下来,进一步使用数据挖掘技术挖掘维护方案库中潜在的知识 and 规律,使其能够进一步完善飞机维护技术文档。

如图 1 所示,AHMS 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。

2. 功能模块

AHMS 由四个功能模块组成,包括 CBR 模块、维护方案模块、KDD (Knowledge Discovery in Database) 模

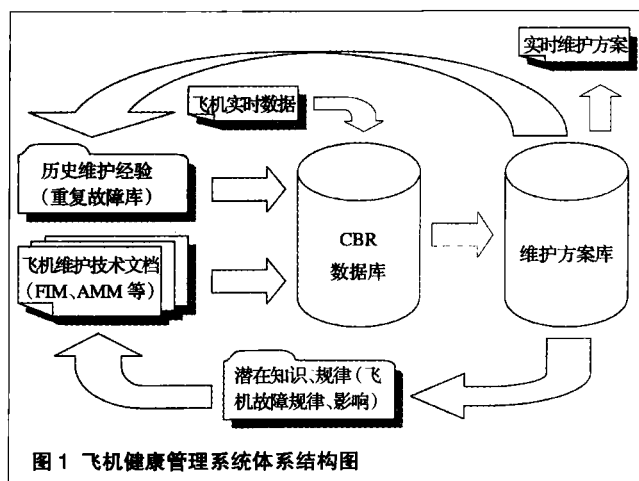


图 1 飞机健康管理系统体系结构图

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(Fault Isolation Manual, FIM; Aircraft Maintenance Manual, AMM 等)以及历史维护经验(重复故障库)中的数据和信息,并使用基于事例的推理方法构建 CBR 数据库。

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。

KDD 模块基于维护方案库,使用规则归纳、异常检测、聚集等数据挖掘技术挖掘潜在的知识和规律。

数据源反馈模块有两个主要任务,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。

关键技术

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。

描述性数据挖掘分析包括异常检测、聚集等多种数据挖掘方法。

预言性数据挖掘分析采用的主要方法是分类,分类是根据训练集数据找到可以描述并区分数据类别的分类模型,使之可以预测未知数据的类别。分类可以采用神经网络算法和决策树算法。

在 AHMS 中,主要采

用了基于事例的推理方法、规则归纳、决策树、神经网络、异常检测、聚集等数据挖掘技术。

基于事例的推理:这种方法思路非常简单,当预测未来情况或进行正确决策时,系统寻找与现有情况相类似的事例,并选择最佳的相同的解决方案,这种方法能用于很多问题求解,并获得好的结果。

规则归纳:即通过统计方法归纳、提取有价值的 if-then 规则。

决策树:即用树形结构表示决策集合,这些决策集合是通过对数据集的分类来产生规则。决策树方法是首先利用信息熵来寻找数据库中具有最大信息量的字段,从而建立决策树的一个结点,在根据字段的不同取值来建立树的分支;然后在每个分支子集中,重复建立树的下层结点和分支,即可建立决策树。

神经网络:这种方法主要是模拟人脑神经元结构,是一种通过训练来学习的非线性预测模型,它可以完成分类、聚类、特征规则等多种数据挖掘任务。

异常检测:异常检测是数据挖掘中一个重要方法,用来发现“小的模式”,即找到数据集中与大多数数据不同或不一致的数据对象。

聚集:聚集则是把数据集分为不同的簇,使得簇和簇之间的差别明显,而簇内个体之间的差异较小。

基于数据挖掘的分类,AHMS 的模块信息见表 1。

2. CBR 模块设计

CBR 模块的任务如前所述。

表 1 AHMS 模块信息表

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	规则归纳、决策树、神经网络
KDD 模块	描述性	规则归纳、异常检测、聚集

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 AHMS 中,构建范例库方法并不以历史经验为起点,而是基于飞机维护技术文档。因此,分两个阶段来构建 CBR 数据库。首先,基于飞机维护技术文档,构建一个总体的、全面的范例库。其次,借助历史维护经验来不断地修正范例库。

飞机制造商所提供的飞机维护技术文档 FIM、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。上述维护资源是构建 CBR 数据库的前提条件。

如图 2 所示,这种 CBR 数据库构建方式可归结为两个主要过程:范例库构建和范例库修正。在范例库构建的过程中,还使用了信息提取技术 IE(Information Extraction),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(可以是 FOXPRO、

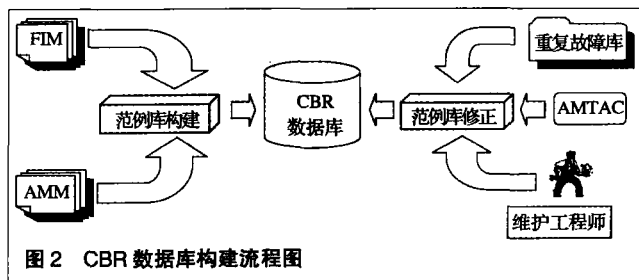


图 2 CBR 数据库构建流程图

ACCESS, 但建议使用 Oracle 数据库) 应包含表 2 中的字段。

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 CFD (Computational Fluid Dynamics) 报文包含了报文头、FDE(Flight Deck Effect)部分、NCMM(Non-Related 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。

程序的设计思路如下:

```
public class Maintenance{
    public static void main(String args[]){
        for(...;获取 CFD 报文;...){
            if(CFD 报文有效){
                提取 CFD 报文中的故障代码
                (CFD_CODE)或故障信息
                (CFD_MESSAGE);
                if(CFD_CODE==CBR_CODE||
                CFD_MESSAGE==CBR_MES
                SAGE){
                    调用 CBR_ACTION 以及
                    CBR_EXPERIENCE;
                    规则归纳、决策树、神经网络等
                    算法;
                    生成实时维护方案;
                    加入维护方案库;
                }
            }
        }
    }
}
```

在维护方案模块设计中,使用了规则归纳、决策树、神经网络等算法。规则归纳算法用于关联规则;神经网络算法的精确性高,所以在 AHMS 中用来建立预言模块,生成维护方案;而决策树算法的可解释性好,易于理解,适用于系统中维护方案的可视化。

4. KDD 模块设计

KDD 模块基于维护方案库,使用规则归纳、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、故障发生原因、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,正是利用数据挖掘技术来发现其故障本质。

我们使用规则归纳算法用于特征规则;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。

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。

我们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。

6. 基于数据挖掘飞机健康管理系统的初步实现

根据上述步骤构建基于数据挖掘飞机健康管理系统的软件雏形。

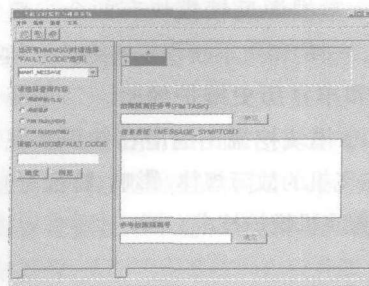


图 3 AHMS 主操作界面

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,并对其在飞机健康管理中的应用进行了研究。数据挖掘技术应用于民用航空,目前起步不久,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。 □

表 2 CBR 数据库结构表

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