

DOI: 10.19416/j.cnki.1674-9804.2018.03.016

民用飞机健康诊断大数据应用研究

A Study on the Application of Big Data Technologies for Civil Aircraft Health Diagnosis

余 汇 唐大鹏 姚 斌 / YU Hui TANG Dapeng YAO Bin

(上海飞机设计研究院, 上海 201210)

(Shanghai Aircraft Design and Research Institute, Shanghai 201210, China)

摘 要:

该文对面向飞机健康诊断的几种大数据技术进行了综述。并行聚类分析方法可以抽取飞机维护系统并行计算机的基本特征,形成飞机各系统历史数据的抽象计算模型;而针对全机状态监控参数中如何从大量飞机参数中挖掘出有价值的信息问题,参数场域挖掘技术发挥了很大的作用;然而在实际应用过程中,经常会遇到多种复杂的故障,这些故障通常具有复杂的非线性关系,而要确定这些非线性相关规则,需要依托大数据技术确定相关的关系;而对于由某一部件失效形成飞机级联故障的链式效应,可利用可视化数据挖掘技术,描述故障数据,结合人类视觉感知能力,加强数据结果处理。

关键词: 飞机;健康诊断;大数据;聚类分析;数据挖掘

中图分类号: V247.5

文献标识码: A

OSID:



[Abstract] The article makes a review of big data technologies in civil aircraft health diagnosis. Parallel clustering analysis could get the basic characteristic of parallel computer to form a computing module of the historical data. The parameter field technology plays an important role in finding valuable parameters in a large number of aircraft parameters. In the actual application process, we will meet with all kinds of complex faults expressed by nonlinear functions. In order to make sure the nonlinear functions, the big data will be used to find out the quantization relationships. For the cascade failure caused by some part's failure, visualization data mining is used to figure the data, combined with human sense of vision to reinforce the result of data mining.

[Keywords] aircraft; health diagnosis; big data; clustering analysis; data mining

0 引言

飞机健康管理是一个动态过程,也可说是一个随机过程,重点是要通过各种传感器的集成来采集数据、获取相关特征量,结合算法建立计算模型对飞机的状态进行监控、诊断及健康管理,使维护和保障模式完成从状态监控向飞行管理的变更,是飞行测试和维修诊断的创新方案。健康管理与互联网技术的融合使得实时诊断飞机运行时产生的故障成为现实。有效的飞机健康管理,需要对飞机各个系统施以关注,由于各系统的数据形式及结构具有差异,导致了飞机故障诊断与健康管理的信号复杂多样且数量巨大,又由于以往数据分析手段不足以及数据利

用不当,使得大量数据的价值被埋没。

为了解决上述问题,面向飞机健康诊断的大数据技术应运而生,重点开展基于大数据健康诊断的机载维护系统的关键技术研究,特别是瞄准下一代民机机载维护系统的新技术和新应用。旨在通过空地综合的大数据技术分析和处理飞机的运行数据,并整合飞机全机系统、结构的运行信息^[1-2],以最大限度感知飞机健康状态,并对异常操作早预测。

基于大数据的飞机健康管理系统概念图,如图 1 所示。在飞机运行及维护过程中,从大量数据源和密集传感器中获得数据,同时需要不断地监控内部状态和外部环境及时进行故障诊断与预

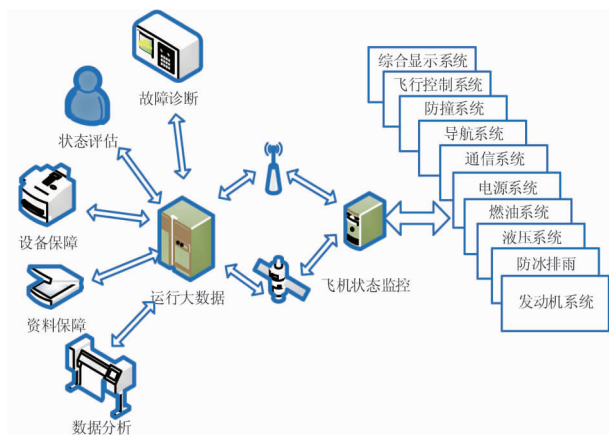


图1 基于大数据的飞机健康管理系统概念图

测,保障飞机的可靠性与安全性,确保经济效益最大化。

1 并行聚类分析方法

为了实时反映当前的健康状况,飞机会时刻产生大量动态数据。在一次航行中会产生几十甚至上百TB的数据,如何从飞机大量的历史健康数据中挖掘出有价值的信息,这就涉及到飞机大量历史数据并行聚类这一分析方法。

1.1 并行计算模型和体系结构

飞机并行计算模型通常是指并行算法的设计与分析,抽象出机载维护系统各并行计算机的基本特征,形成一个针对飞机燃油、电源、发动机等系统的历史大数据计算抽象模型。目前,所有机载并行数据计算机均属于多数据流多指令模式,即飞机上不同处理器可同时对不同类型数据执行不同的指令。指令流可同步或异步执行,如图2所示。

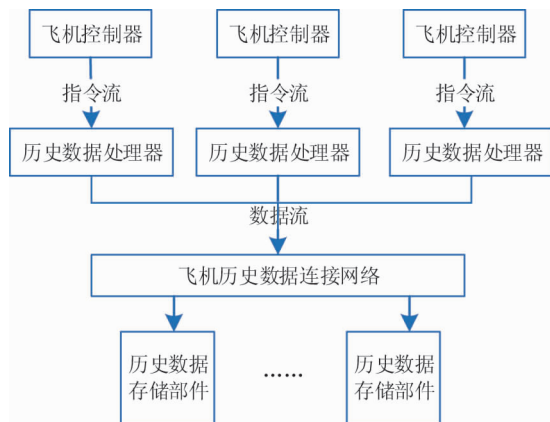


图2 多指令多数据流并行结构

1.2 并行聚类分析

以飞机燃油系统为例,并行计算技术同于飞机燃油系统中油温、油压等海量数据的数据挖掘^[3],这是飞机大数据挖掘技术逐步走向深入的一个重要层次。不仅可以加快飞机对燃油数据的处理速度,而且可以提高燃油系统故障数据挖掘的能力,方便飞机驾驶员合理操作。

首先,飞机燃油系统的历史数据集根据计算节点的数量进行划分,形成 p 个子数据集 D_1, D_2, \dots, D_p (p 是节点数)。然后,在 p 个节点运行聚类算法,形成局部数据集 $\{C_{11}, C_{12}, \dots, C_{1x}\}, \{C_{21}, C_{22}, \dots, C_{2y}\} \dots \{C_{p1}, C_{p2}, \dots, C_{pz}\}$ (x, y, z 是集群数, p 是子数据集的数量)和孤立点的集合 O_1, O_2, \dots, O_p ,并且将所有子数据集的孤立点进行聚类以获得新的数据集 $\{C_{o1}, C_{o2}, \dots, C_{om}\}$ (o 表示飞机孤立点数据, m 为集群数)和新的孤立点集合 $\{O_o\}$ 。最后局部聚类结果和孤立点聚类结果的合并便是最终的全局聚类结果。

根据上述思路,并行聚类算法一般分为以下三个步骤:

1) 根据飞机历史数据划分策略,待聚类的数据集被均等地或等差地划分为 p 个数据子集,其中 p 为当前可用节点的数量,并且,每个子集的数据都被发送到每个节点。

2) 在每个节点上对本地航班历史数据集进行局部聚类算法,各形成一组集群和一组飞机数据孤立点集合。每个集群完成后,将结果发送至主节点。

3) 主节点收到各个其他节点的聚类结果后,形成集群集和孤立点集,这个孤立点集就是飞机最终历史数据的孤立点。然后使用聚类合并技术将每个节点的聚类结果和集群的聚类结果进行合并,形成最终的集群集。整体流程如图3所示。

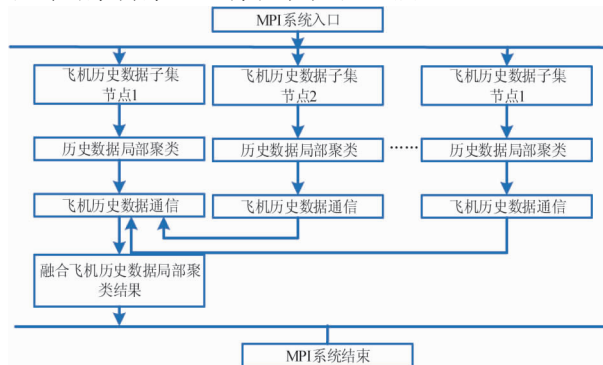


图3 大量历史数据并行聚类算法流程图

2 大量参数数据场挖掘技术

2.1 飞机参数场域定义

将原来飞机的全部相关参数都投射到数值域中,并规定每架飞机的所有相关参数都在数值域中向外扩散,并将所有飞机相关参数扩散范围称为飞机参数数据场。定义飞机相关参数之间的影响函数为场强函数,数值域中的值可以投射至场域中^[4]。

飞机参数的场域聚类的基本思想:以平滑连续的曲线将飞机参数数值域中所有相关参数的等势点连接起来,构成等势线(等势面),形成场域;然后从飞机各相关参数的场域中,可以找到相关参数的局部极值点,作为这些参数的聚类集群中心;根据各极值点向外扩散的梯度方向,找到这些参数的等势线,以此作为每个集群的分界线,从而划分成为不同的集群。

2.2 飞机大量参数数据场聚类算法

根据飞机参数数据场的特征,定义了气压高度、飞行速度、马赫数、飞行高速等 n 个飞行参数组成的飞机参数数据空间 $X = (x_1, x_2, \dots, x_n)$, m 个集群, $C = (c_1, c_2, \dots, c_m)$, Num Data 记录飞行参数样本数量, Num C_i 记录第 i 个集群中飞行参数样本的数量, outlier 记录为离散群点的数量。对于飞行参数数据空间中任意一个参数诸如马赫数 x , 通过接收飞行参数数据场中所有数据对象,例如气压高度、飞行速度、飞行高度的辐射而产生的势值 $f(x)$, 其中 p_i 表示第 i 个数据对象在点 x_i 处的场强。

数据场聚类算法的具体步骤如下:

输入 n 个飞行参数对象 $X = (x_1, x_2, \dots, x_n)$, 令 Num Data = n 。

1) 选择适当的场强函数并计算数值域中某一参数如气压高度的势值 $f(x)$;

2) 以平滑曲线连接数值域中势值相等的点即可构成等势线(等势面);令 $i = 1$;

3) 找出第 i 个集群中心势值最大的点;

4) 从第 i 个聚类中心展开,找到局部势能极小的等势线,然后将等势线内的所有数据对象移动到集群 C_i 中,以此形成第 i 个集群;

5) 更新数值域中数据对象的数量 Num Data = Num Data - Num C_i ;

6) 若算法结束, outlier = Num Data - Num C_i ; 否则执行步骤 7);

7) $i = i + 1$, 执行步骤 4)。

3 复杂故障的距离量化关联规则挖掘方法

在飞机健康诊断领域,需要进行复杂故障原因分析来确定机载健康数据的响应之间的关联关系,以飞机燃油系统故障为例,就是要确定该故障是发生在发动机供油系统还是在输油系统或是燃油箱的通气系统或增压系统、加油系统、放油系统还是燃油测量系统等。而这种关系不能表现为线性或非线性函数关系,只能在一定约束条件下用相关关系来表示。这种相关关系在大数据中客观存在,对于大数据分析更有意义,以解决飞机复杂故障的健康诊断问题。

3.1 复杂故障相关关系挖掘模型

飞机的维护系统包括两个主要的方面,分别是最小可信度和最小支持控制算法,如图 4 所示。

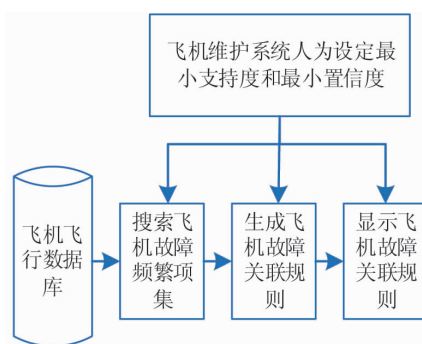


图 4 复杂的故障相关关系发掘流程

关联规则的挖掘^[5]主要有两步:(1)找出所有的飞机故障频繁项集;(2)找出第一步中产生的同时满足量化的最小可信度和最小支持度的相关关系。

3.2 复杂故障的基于距离算法设计

1) 飞机故障的集群效应将直接影响区间划分的合理性,并且对后面的产生的规则是有效的甚至是决定性的^[6]。飞机发动机振动数据聚类划分,不仅对一个或一对故障聚类,而且对所有不同种类的故障数据进行聚类,采用聚类算法,如基于密度的适应性密度可达性的集群算法(Clustering Algorithm Based on Density and Density Reachable,简称 CADD)和 K-means 集群算法,根据获得的飞机的健康数据,对所有故障数据进行聚类,根据聚类结果进行区间划分或形成簇。在所有飞机故障数据上聚类,得到故障簇或区间更能合理反映数据之间关系。

2)对所有故障数据采用聚类的方法,得到故障簇,若满足支持度,就可以减少限定基于距离的关联规则的条件,使方法更便于应用。

3)提出对故障参数 D_0 (故障最小置信度)取值的限定方法。将新获得的故障簇的半径作为 D_0 取值的参考标准。

由于采用故障簇半径的概念来控制聚类的方法十分便捷,基于故障簇中心的聚类算法,如K-means算法^[7-8],能够取得较好的挖掘效果。

在定义了基于距离的量化相关关系的规则之后,要使得规则成立,必须要满足以下三个条件:

1)前一个故障类别与当前故障类别的簇间距离不大于预设的值;

2)为了保证之前的每个故障簇中的飞机故障能够同时发生,必须要保证簇间距离不得超过预先设定的参数的距离;

3)为了保证各种类型的飞机的故障簇中的故障能同时出现,各簇间的距离不得超过预先设定的参数。

具体分析:对与飞机故障相关的的所有的相关数据进行集群聚类,这样做的好处是飞机的某一个系统的所有种类的故障数据通过聚类,都能得到非常完整的体现;另外,聚类得到的簇内,不同飞机故障类别数据必然形成高相似度的故障簇。因为所有的故障类别都会被聚类,所以最终的故障簇内将会包含故障数据的所有的属性。换句话说如果在最终的故障簇中的故障样本的数量满足了预先设定的最小支持度的话,那么由所有属性集构成的故障集就会成为最大频率故障集。

4 级联故障发掘方法可视化

飞机的级联故障是一种复杂的故障模式,以飞机发动机系统为例来直观解释,是指当其中的某个部分或某个组件发生故障或者失效(这里假设发动机轴承故障),进而导致系统内外与该部件相关联的部件(如定子)或者区域的失效,形成连锁反应并扩散,最终导致整个发动机系统失效。由以上描述可以看出飞机的级联故障具有以下几个特点:

1)飞机的级联故障的原因一般是由某个特定的源头引起的;

2)飞机级联故障的表现形式是较大范围的故障、系统级的故障或者多个故障同时发生;

3)飞机级联故障从源头到结果应该遵循一条因果链。

4.1 飞机级联故障的可视化数据挖掘

飞机的级联故障的可视化数据挖掘技术是随着数据挖掘技术和计算机可视化技术的发展而发展起来的^[9],它能有效地把人类的感知能力和相关知识应用到级联故障的信息挖掘中。它以刻画飞机级联故障数据的功能性,以及人类视觉的感知能力、倾向和关系的能力为基础,用可视化技术来加强数据处理。数据挖掘是将大量的飞机历史数据中潜在的、有价值的级联故障知识抽取出来,而可视化就是把故障数据、信息和挖掘出的级联故障知识转化成可视的表达形式的过程。

将数据进行可视化的具体的特点包括:

1)发现被监测数据的变化趋势,如飞机燃油系统中油压随时间的增长、下降等情况;

2)找出数据的奇异点,如时间维度下,某时刻压力传感器所测量的油压值相比于相邻采样时刻采集的值波动较大,通过可视化技术,就在很长的时间维度下可以看出异常点;

3)识别数据的边界,例如在整个飞行航段过程中油压的最大值、最小值等的边界值;

4)对故障数据的分类以及集群结果进行显示,识别出不同数据的不同特征;如以图像的模式向飞机驾驶员显示级联故障数据的特征,更易于飞机驾驶员找出正确规律,快速做出决策;

5)提供丰富的文本信息,通过飞机驾驶员的操作,显示器将显示驾驶员所需要的飞机状况信息,有助于驾驶员能够迅速找到需要的特定的参数数据,同时能够实现数据的选取、数据滤波、数据缩放以及其他功能。

4.2 飞机级联故障可视化数据挖掘系统设计

在飞机级联故障数据挖掘的整个过程中都伴随着可视化技术的使用,将数据挖掘的整个过程进行可视化具有很高的价值和重要性。系统结构如图5所示。

从图5可知,该级联故障可视化体系结构分为独立的四层。级联故障数据层从数据源载入数据后,可通过数据化查看飞机级联故障原始数据,经过数据清理、数据集成、数据变换、数据规约处理后,也可依托级联故障数据可视化查看整理后的数据,然后选择级联故障的数据挖掘算法进行数据挖掘,最后把挖掘结果以文件形式保存,并从文件读取级联

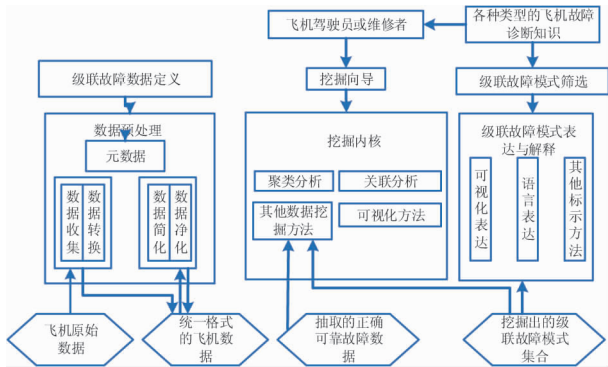


图5 对飞机的级联故障进行可视化的系统结构图

故障聚类结果,并将聚类结果显示出来,根据待诊断级联故障样本所处的故障类标,确定故障原因。便于飞机驾驶员进行快速适当地操作,了解级联故障机理,从而实现排故。

5 结论

飞机健康诊断的大数据技术的应用可提高飞机安全性、降低维修成本、提高维修性,实现对飞机故障的精确定位与诊断,减少故障检测及隔离时间,保持我国民机的可持续发展能力,是民机航空维护系统中不可缺少的部分。

参考文献:

- [1] 以涛. 飞行数据分析及译码[J]. 航空工程与维修, 2000(1): 25-26.
- [2] LIU Z, SUN G, BU S, et al. Particle learning framework for estimating the remaining useful life of lithium-ion batteries[J]. IEEE Transactions on Instrumentation & Measurement, 2017,

66(2): 280-293.

- [3] 胡善杰. 在云环境下的数据挖掘算法的并行化研究[D]. 成都: 电子科技大学, 2013.
- [4] 闫伟, 赵杨, 高原. 飞行时序数据相似性挖掘算法研究[J]. 计算机与网络, 2008, 34(21): 54-57.
- [5] 熊励. 基于关联规则的飞机机件故障因素分析与故障预防[C]//Intelligent Information Technology Application Association Proceedings of the 2011 International Conference on Future Computer Science and Application. [S. l.]: Intelligent Information Technology Application Association, 2011: 349-353.
- [6] LIU Z, JIA Z, VONG C M, et al. Capturing high-discriminative fault features for electronics-rich analog system via deep learning[J]. IEEE Transactions on Industrial Informatics, 2017, 13(3): 1213-1226.
- [7] 黄贤英, 张金鹏, 陈微微. 基于 K-means 聚类算法和信息熵的页面排序算法研究[J]. 计算机工程与设计, 2013, 34(5): 1695-1699.
- [8] LIU Z, LIU T, HAN J, et al. Signal model-based fault coding for diagnostics and prognostics of analog electronic circuits[J]. IEEE Transactions on Industrial Electronics, 2017, 64(1): 605-614.
- [9] 王开军. 多维数据的聚类结果可视化技术综述[J]. 福建师范大学学报(自然科学版), 2012, 28(4): 115-124.

作者简介

余 汇 男, 硕士, 高级工程师。主要研究方向: 民用飞机实时监控与故障诊断。E-mail: yuhui@comac.cc

唐大鹏 男, 本科, 工程师。主要研究方向: 民用飞机故障诊断。E-mail: tangdapeng@comac.cc

姚 斌 男, 硕士, 工程师。主要研究方向: 民用飞机故障诊断。E-mail: yaobin@comac.cc