

液体火箭发动机故障检测和诊断中数据挖掘策略的分析*

胡小平¹, 张丽娟², 王艳梅², 李舟军²

(1. 国防科技大学 航天与材料工程学院, 湖南 长沙 410073; 2. 国防科技大学 计算机学院, 湖南 长沙 410073)

摘要 :分析了液体火箭发动机的工作特点,提出了应用数据挖掘方法从数据仓库的角度对液体火箭发动机进行故障检测和诊断的策略。对在液体火箭发动机故障检测和诊断的不同问题中可能应用的数据挖掘方法进行了分析比较。分析表明,聚类、分类、关联、时间序列分析和孤立点检测等数据挖掘方法适用于液体火箭发动机的故障检测和诊断。

关键词 :液体火箭发动机;故障诊断;数据挖掘;聚类;分类;关联;时间序列分析;孤立点检测

中图分类号 :V211;V433 文献标识码 :A

The Analysis of Data Mining Strategy in Fault Detection and Diagnosis of the Liquid Rocket Engine

HU Xiao-ping¹, ZHANG Li-juan², WANG Yan-mei², LI Zhou-jun²

(1. College of Aerospace and Material Engineering, National Univ. of Defense Technology, Changsha 410073, China;

2. College of Computer, National Univ. of Defense Technology, Changsha 410073, China)

Abstract :By analyzing the working characteristics of the liquid rocket engine, strategies of applying data mining from the point of view of the data warehouse for FDD of LRE are proposed. The data mining methods possibly applicable to different topics in the fault detection and diagnosis of LRE are compared. Primary research results show that such data mining methods as clustering, classification, association, time-series analysis and outlier analysis are feasible in the FDD of LRE.

Key words :liquid rocket engine; fault detect and diagnosis; data mining; clustering; classification; association; time-series analysis; outlier analysis

数据挖掘就是从大量的、不完全的、有噪声的、模糊的、随机的数据中,提取隐含在其中的人们事先未知、但潜在有用的信息和知识的过程^[1]。液体火箭发动机系统结构复杂,具有互相耦合的高度非线性热力学特性,难以得到精确的动力学模型或层次知识模型。对于这种复杂系统的故障进行检测与诊断,可以基于功能主义的“黑箱”理论,从数据仓库的角度,应用数据挖掘中的分类、聚类、关联、孤立点分析以及时间序列分析等方法,从表征系统行为的输入输出测量信号中直接提取系统的正常与故障特征信息。

1 火箭发动机试车数据的预处理

在进行数据挖掘以前,必须根据火箭发动机试车数据的特点及挖掘任务的需要对数据进行预处理^[2,4]。在此阶段,需要做的处理工作包括数据的补缺、除噪、压缩、维规约和规范化。

1.1 数据的除噪处理

实际试车中由于测量或执行错误,导致所测得的发动机工作参数往往带有随机错误和偏差,这就是噪声。噪声数据的存在影响挖掘结果的正确性,因此在对发动机的试车数据进行挖掘之前,必须去掉噪声。可以考虑使用的除噪方法有:

(1) 聚类中的孤立点分析 噪声数据与正常的的数据不一样,采用孤立点分析可以将正常数据聚类成

* 收稿日期:2004-11-15

基金项目:国家自然科学基金资助项目(50376073)

作者简介:胡小平(1963—),男,副教授,博士。

群,而落在群之外的孤立的噪声数据可以被去除掉。需要注意的是,落在群之外的孤立点也可能反映了重要的故障信息,若视为噪声而除掉有可能会丢失有用的信息,因此需要正确区分噪声信息和故障信息。

(2)回归方法:通过使数据适合于一个函数,进行数据平滑,找出适合数据的数学方程式,来帮助消除噪声。

(3)小波除噪:小波阈值除噪是一种很好的除噪方法,不仅对缓变信号效果好,对脉冲和冲击瞬变信号的恢复效果也很理想^[6]。其步骤是:将信号做小波变换,得到变换系数,确定变换系数阈值,进行阈值处理,将处理后的系数重构,得到除噪信号。

(4)独立分量分析(ICA):其实质是一个最优化问题,以混合信号系数矩阵的负熵最大为判据,通过估计系数矩阵来获得源信号矩阵,使得各源信号之间的独立性最强。ICA是盲源分离问题最主要和最有效的方法之一^[8]。

1.2 数据压缩

在海量的热试车数据中进行复杂的挖掘将需要很长时间,这使得挖掘不现实或不可行。数据压缩可以在大致保持原数据完整性的条件下,使用编码技术压缩数据集。可采用的压缩方法有:

(1)小波变换:离散小波变换(DWT)是一种线性信号处理技术,它适合高维数据。当用于数据向量 D 时,将它转换成数值不同的小波系数的向量 D' ,两个向量具有相同的长度。小波变换后,数据得到裁减,仅存放一小部分最强的小波系数,就能保存近似的压缩数据。

(2)主成分分析(PCA):若原数据是 N 维的,该方法搜索 M ($M \leq N$)个最能代表原数据的正交向量,将数据投影到较小的空间,导致数据压缩。PCA也可作为一种维归约方法使用。

1.3 维归约

反映发动机工作状态的参数有上百个,在热试车数据中也就可能包含了数以百计的属性,其中有些属性根本或可能与挖掘任务不相关,是冗余的。冗余属性的存在增大了数据量,可能减慢挖掘进程,导致发现的模式质量差,所以有必要使用维归约删除冗余属性。可能采用的维归约方法有:

(1)用粗糙集理论进行属性约简:粗糙集理论(RS)作为一种全新的数学概念,为处理具有不完整、不一致及不确定性特征的信息提供了新的有效工具。由于粗糙集理论提供了不同于其它常规方法的新思想,以及它不需要先验知识,只依赖数据内部关系,适宜处理不完备数据等特性,被广泛用于数据挖掘领域以解决其它方法难以解决或解决效果欠佳的问题。

(2)用判定树算法进行属性子集选择:构造判定树,出现在树中的属性是相关的。数据立方体和存放在概念分层中的知识可以用来在不同的抽象层归纳判定树。此外,一旦导出判定树,概念分层可以用来概化或特化树的节点,可以在属性上进行上卷或下钻,并对新的特定抽象层的数据重新分类。

(3)主成分分析:PCA通过创建一个替换的、较小的变量集来组合属性的精华,原数据可以投影到该较小的集合中。

1.4 数据的规范化

将属性数据按比例缩放,使之落入一个小的特定区间。对于一些分类算法(如BP神经网络算法、SVM或最临近分类和聚类的距离度量分类等算法),规范化特别有用。

2 可能用于故障检测与诊断的数据挖掘方法

经过对发动机不同时段和不同类型的故障进行分析表明,在液体火箭发动机的故障检测和诊断中可能采用的数据挖掘方法有:分类、聚类、关联、孤立点分析和时间序列分析等。

2.1 分类方法

故障检测是通过分析热试车数据,来确定发动机工作是否正常,即将数据分为正常和异常两类,所以故障检测可以看作是分类问题。可考虑用数据挖掘中的分类技术来进行故障检测,分类方法如下:

(1)支持向量机(SVM)

SVM是基于统计学习理论的模式分类器^[5]。它通过结构风险最小化准则和核函数方法,较好地解

决了分类器复杂性和推广性之间的矛盾。它主要针对两类分类问题,在高维空间中寻找一个超平面作为两类的分割,以保证最小的错误分类率。它既可用于线性可分问题,又可用于非线性问题——这很适合液体火箭发动机的故障检测与诊断。对非线性问题,SVM首先通过用核函数定义的非线性变换将输入空间变换到一个高维空间,然后再在这个空间中求广义最优分类面。

(2) 判定树分类(decision tree)

判定树是比较经典和成熟的分类方法^[3],目前已经开发出的判定树算法有:ID3和C4.5(用于较小数据集)、SLIQ、SPRINT和Rain Forest(可用于大规模数据集,可伸缩性好),还有与其它方法相结合的判定树算法(与粗集、模糊数学结合等)。判定树最大的优点是易生成可理解的规则,但对于连续值数据和有时间顺序的数据,它的处理效果并不理想。而热试车数据大多是连续的,且为了测控和故障检测的需要,往往在多个时间点上采样数据。因此,对热试车数据进行分类,采用判定树的效果可能不会好。

(3) 贝叶斯分类

贝叶斯分类是统计学分类方法,可以预测给定样本属于一个特定类的概率。贝叶斯分类基于贝叶斯定理,它有朴素贝叶斯分类和贝叶斯信念网络两种模型。朴素贝叶斯分类假定各属性的值相互条件独立,对于液体火箭发动机的热试车数据而言,各工作参数之间有一定的关系,满足参数平衡模型,而不满足朴素贝叶斯分类的假定,所以朴素贝叶斯分类用于热试车数据分类可能不太合适。贝叶斯信念网络则可以表示属性子集间的依赖,所以可以考虑采用贝叶斯信念网络,但训练贝叶斯网络比较困难。

(4) 神经网络分类

神经网络是由大量简单的神经元相互连接而成的自适应非线性动态系统,它具有很高的容错和抗干扰能力,在某些信息丢失的情况下,仍能作出正确决策,这很适合液体火箭发动机的故障检测与诊断。目前,已经有很多神经网络模型用于液体火箭发动机的故障检测与诊断。用于分类的最著名的神经网络学习算法是基于多层前馈神经网络的后向传播算法(BP算法)。但需要说明的是训练神经网络需要很长时间,建立网络需要做大量的数据准备工作。

(5) 粗糙集分类^[7]

粗糙集(rough set)是研究不完整数据和不确定知识的表达、学习、归纳等的理论。它基于集合中对象间的不可分辨思想。这一理论为处理具有模糊的、不精确的或不完全信息的分类问题提供了一种新的工具。其主要思想是在保持信息系统分类能力不变的前提下,通过知识约简,导出问题的决策或分类规则。因此可以直接将粗糙集理论用于分类。初步研究表明,粗糙集方法在涡轮泵故障的检测中是可行的。

(6) 模糊数学分类方法^[7]

模糊集理论也是研究信息系统中知识不完善、不准确问题的。但是它的着眼点在于集合的模糊性,它的基石是隶属函数。可以考虑用模糊数学中的模糊综合评判来进行故障检测与诊断。模糊综合评判的输出结果不再是一个对象明确地属于某一类或不属于某一类,而是以一定的隶属度属于各个类别的形式,这样的结果往往更真实,具有更多信息。

(7) k-最临近分类

最临近分类基于类比学习。训练样本用 n 维数值属性描述。每个样本代表 n 维空间的一个点,所有训练样本都存放在 n 维模式空间中。对于给定的未知类标号的样本, k -最临近分类法搜索模式空间,找出最接近未知样本的 k 个训练样本,然后将未知类标号的样本分配到 k 个最临近者中最公共的类。

(8) 基于案例的推理

基于案例的推理(CBR)是人工智能发展较为成熟的一个分支。它是一种基于过去实际经验或经历的推理。基于案例的推理用复杂的符号来描述样本或“案例”,案例记录了过去发生的种种相关情节。当给定一个待分类的新案例时,基于案例的推理首先检查案例库中是否存在一个与当前案例一样的案例,如果存在,则返回附在该案例上的解。如果找不到同样的解,则基于案例的推理就搜索与当前案例最相关的案例,并对该案例作必要的改动以适合当前案例。显然,案例库的建立对于推理是否成功至关

重要。

(9) 遗传算法分类

遗传算法是指人类仿照生物学进化所产生的仿真进化系统。它是一种优化算法,能模仿生物体(染色体)进行遗传、变异、交叉,体现了适者生存的原理,最终产生优化解。遗传算法易于并行,可以有效地防止搜索过程陷入局部最优解,具有较好的全局搜索性能。它处理的是计算对象编码,对搜索空间没有任何特殊要求,仅需用目标函数的值来进行遗传操作,不需导数或其它信息,采用变迁规则而非确定性规则来指导其搜索空间,有明确的搜索方向,比随机搜索方法有更高的搜索效率。

故障诊断根据已有的异常状态信息对故障的类型、程度、部位作出判断,以分离故障模式并估计故障发生的程度,其实质仍然可看作是分类问题,所以可以采用与故障检测类似的分类方法来解决。需要说明的是当故障模式未知时,可以采用数据挖掘中的聚类方法,利用样本热试车数据来确定故障模式,然后再用分类方法对未知故障模式的数据进行分类,以确定未知数据的故障模式。

2.2 孤立点分析

当发动机工作异常时,相应的热试车数据与正常工作时的热试车数据不一致,即故障数据相对正常数据是孤立点,可采用数据挖掘中的孤立点分析方法来进行故障检测。可以考虑采用的孤立点检测方法如下:

(1) 基于统计的孤立点检测

要求已知数据分布,且绝大多数检验针对单个属性,计算量也大,可能不太适合用于故障检测与诊断。

(2) 基于距离的孤立点检测

假定一个数据分布或概率模型,并使用距离度量,到其它聚类的距离(欧几里得距离或曼哈坦距离)很大的对象被视为孤立点。

(3) 基于偏离的孤立点检测

通过检查一组对象的主要特征的差别而不是使用统计或距离度量来识别孤立点,这种方法用于故障检测与诊断可能较合适。

2.3 关联挖掘

关联规则技术是数据挖掘领域的一个重要方法,侧重于确定数据中不同域之间的联系,找出满足给定支持度和置信度阈值的多个域之间的依赖关系。关联分析发现关联规则,这些规则展示属性—值频繁地在给定数据集中一起出现的条件。通过关联分析,可能发现液体火箭发动机故障模式之间的关联关系以及发动机不同工作参数之间潜在的关联关系。

2.4 时间序列分析

液体火箭发动机的试车数据是随时间变化的时序数据,可以考虑采用时间序列分析的数据挖掘方法^[5],包括趋势分析、相似性搜索、与时间有关数据的序列模式挖掘和周期模式挖掘等方法,动态时间弯曲和时域特征提取等数据挖掘方法也可能有用武之地^[6]。

3 讨论

由于发动机启动、关机过程同主级稳态工作过程的控制机理不尽相同,因此进行故障检测与诊断时要将启动、关机过程同主级稳态工作过程分开考虑,采用不同的故障检测与诊断方法。

(1) 启动/关机过程的故障检测与诊断

液体火箭发动机启动过程和关机过程的本质是非线性时变随机过程,其复杂性主要是由于启动器、活门开启的非线性动态特性以及启动器和燃气发生器交叠工作段的非线性特性所造成。此阶段可能出现的故障类型有:火药启动器工作异常、启动活门未全打开、主活门未全关闭、副断流活门不动作和电爆管失效等。

启动/关机过程的故障检测与诊断可考虑采用的数据挖掘方法有:支持向量机、隐马尔科夫模型、关联挖掘、非线性自回归移动平均模型(NARMAX)等。

(2) 主级稳态工作过程的故障检测与诊断

在主级的稳态工作过程中,主要受供应系统的流体及机械运动和燃烧室、燃气发生器的能量转换过程的控制,此阶段易发生阻塞、泄漏和涡轮泵系统故障等。发动机在稳态正常工作过程中,重要监测参数表现出较好的平稳特性,将 ARMA 模型用于故障监测目的,一般只需要发动机正常工作时的数据。时序分析方法在其它领域(如机械系统)的故障诊断中已有成功的应用^[6],基于时序分析的故障检测方法具有良好的实时在线工作能力。

每一种方法都有局限性,而实际系统又很复杂,想利用单一的任何一种方法达到高的故障检测率、低的误报警率及漏报率,准确地判断故障发生的时间、部位及大小都是不现实的,所以必须综合应用各种挖掘方法来进行故障检测与诊断。

4 结 论

在液体火箭发动机的故障诊断中应用数据挖掘的方法是一种新的思路。可以综合采用分类、聚类、关联、时间序列分析和孤立点检测等多种数据挖掘的方法,从数据仓库的角度,挖掘隐藏在发动机试车数据中的知识。

参 考 文 献:

- [1] 韩家炜,等. 数据挖掘:概念与技术[M]. 北京:机械工业出版社,2001.
- [2] 张育林,等. 液体火箭发动机健康监控技术[M]. 长沙:国防科技大学出版社,1998.
- [3] Bajwa A, Kulkarni D. Engine Data Analysis Using Decision Trees[R]. AIAA 2000-3627,2000.
- [4] 刘洪刚. 液体火箭发动机智能故障诊断理论与策略研究[D]. 长沙:国防科技大学,2002.
- [5] David V, Sanchez A. Advanced Support Vector Machines and Kernel Method[J]. Neurocomputing, 2003, 55: 5-20.
- [6] 刘东. 工程诊断中数据挖掘应用的若干问题[D]. 西安:西安交通大学,2001.
- [7] 张丽娟. 基于模糊数学和粗集理论的分类方法的研究、实现及应用[D]. 长沙:国防科技大学,2003.
- [8] 任海峰,胡小平,等. 独立分量分析在液体火箭发动机故障检测中的应用[J]. 国防科技大学学报,2004,26(1):13-16.

