

大数据分析技术在装备监测系统中应用研究

程继红¹, 阮传峰², 崔嘉³, 齐玉东⁴

(1. 海军航空工程学院 科研部, 山东 烟台 264001; 2. 海军航空工程学院 研究生管理大队, 山东 烟台 264001;
3. 海军航空工程学院 控制工程系, 山东 烟台 264001; 4. 海军航空工程学院 兵器科学与技术系, 山东 烟台 264001)

摘要: 针对装备状态监测系统中长期积累的大量数据缺乏有效分析手段、利用率低等问题, 对装备监测系统中存储的在线监测数据和积累的历史数据, 建立可分析挖掘的大数据集; 运用数据分析和机理分析的方法建立比较模型, 采用异步信息融合的算法对数据进行处理, 利用大数据的多尺度特性研究装备生命周期的分析预测、优化运行、预知维修等目标, 提高了装备运用知识的获取速度, 有利于构建完整的装备生命周期预测诊断体系。

关键词: 装备; 大数据; 数据分析; 数据挖掘

Research on the Application of Large Data Analysis Technology in Equipment Monitoring System

Cheng Jihong¹, Ruan Chuanfeng², Cui Jia³, Qi Yudong⁴

(1. Department of Scientific Research, Yantai 264001, China; 2. Graduate Students' Brigade, Naval Aeronautical and Astronautical University, Yantai 264001, China; 3. Department of Ordnance Science and Technology Engineering, Naval Aeronautical and Astronautical University, Yantai 264001, China; 4. Department of Control Engineering, Naval Aeronautical and Astronautical University, Yantai 264001, China)

Abstract: Aiming at the long-term accumulation of large amounts of data of the equipment condition monitoring system lack of effective means of analysis and low utilization of data, establish the large data set based on the equipment monitoring system in the storage of online monitoring data and long-term accumulation of historical data. Then, establish method analysis with data analysis and comparison of mechanism model, using asynchronous information fusion algorithm for data processing and adopting the multi-scale characteristics of big data to research the analysis and forecast, operation optimization, predictive maintenance of equipment life cycle etc. This method improves the acquisition speed of the equipment knowledge and is conducive to the equipment life to build a complete cycle prediction and diagnosis system.

Keywords: equipment; big data; data analysis; data mining

0 引言

随着装备信息化程度的不断提升, 信息系统对装备使用、维护各个环节的监测记录, 促进了大数据的形成。但对该大数据的知识挖掘比较欠缺, 未能有效地为装备的使用、维护等提供高质量的辅助决策。本文在建立大数据集的基础上, 运用模型算法, 通过不同尺度的分析和不同层面的知识挖掘, 所形成的结论对于装备良好工况的保持、维护费用的降低、安全性的提高具有重要的指导意义。

1 数据分析系统结构

在装备监测信息系统平台的统一数据元标准下, 装备生命周期中的在线状态监测数据、检维修数据、装备档案与装备变更、故障模式分析等各种形态的数字信息资源集合具有了统一的、标准的描述方法和检索手段。这使得监测系统中储存的大数据, 运用算法对它进行多尺度分析, 知识挖掘变得简便有效^[1]。有益于决策分析品质的提高, 使构建完整的装备生命周

期预测体系成为可能^[2]。装备监测系统中储存的大数据分析系统结构如图1所示。

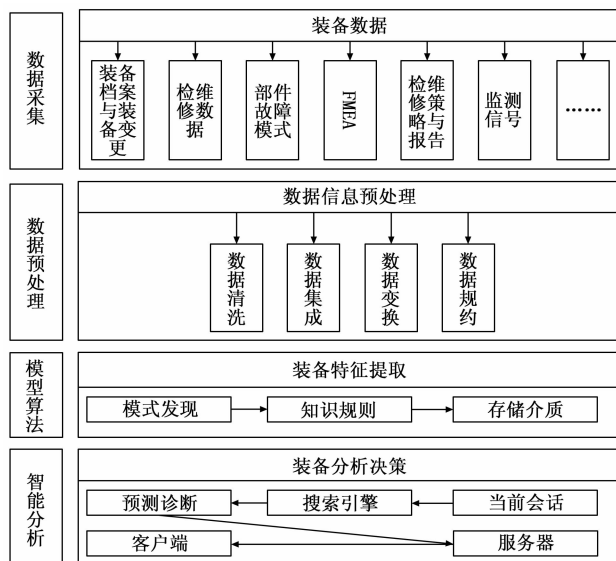


图1 数据分析系统结构图

收稿日期:2016-06-22; 修回日期:2016-08-02。

作者简介:程继红(1964-),男,安徽省桐城市人,教授,硕士研究生导师,主要从事海军航空、导弹装备综合保障方向的研究。

2 装备大数据的特点

当前科技进步日新月异,尤其是互联网、云计算、云存储等计算机网络技术的跨越发展,大数据呈现出海量性(Volume)、多样性(Variety)、高速性(Velocity)、易变性(Variability)四个典型特征^[3]。表现为数据量更多,不是片段的数据样本,而是全体数据;数据结构更混杂,非结构化数据量激增,所占比重越来越大;数据之间的关系更复杂,并非容易理解的因果关系,而是难懂晦涩的相关关系^[4]。而在高技术装备生命周期中,运行、维护、检修等全过程中采集累积的大数据,不仅具有上述特性,自身还有着以下独有的特点:

1) 数据变量多且维度高。各种复杂的物理、化学变化不可避免的发生在装备运行过程中,由于参数之间互相产生耦合,这使得系统将变得复杂且空间多维,致使描述系统需要的参数增多,对数据的理解难度也将相应的增加。

2) 数据呈复杂的非线性。描述装备监测系统状态的绝大部分变量的变化是非线性的。比如,在热力学中温度对应于焓值之间的变化并不是简单的线性关系。随着数据规模的不断增大,越来越复杂难懂的非线性关系,增加了对数据的理解分析、知识发掘的难度。

3) 样本数据采集分布不均匀。装备通常会长时间运行在状态良好且效率较高的工况下,考虑到装备磨损以及安全性,不会让系统运行在某些过渡过程。装备工作方式的这个特点使得监测数据点比较密集的落在较窄的时空区域内,不利于对装备状态的建模分析。

4) 高噪声。由于在装备的运行过程中难以屏蔽复杂电磁的干扰、恶劣环境的影响,状态监测的信息中必定会有噪声的产生。一般而言监测信号的信噪比越低,数据存在的误差越大,数据分析的可靠性也越差。

因此,监测系统中存储的大数据夹杂着错误决的办法是对大数据按照清洗、集成、变换、规约的步骤进行数据预处理^[1]。这个过程改善了数据的品质,为高效的分析数据创造了必要的前提条件。

3 大数据模型算法

3.1 常用算法

运用算法先把监测系统中存储的大数据分为两类,一类为提供建立基准模型所需的建模数据,另一类为作为基准模型输入的运行数据。根据装备运行在不同的工况下,使用数据建模或机理建模的方法,分别建立装备对象的动态和稳态的两种基准模型。把装备运行中的数据作为基准模型的输入,输出为装备状态的预测数据,与在线监测的实时状态数据相比较,分别获得两种工况下模型对应的残差值。按照信息粒化的概念,把残差分解为多层次的信息粒,由于信息粒存在的不同程度的延迟现象,采用异步信息融合进行信息处理的方法,大大提高融合残差的精确度。根据监测状态的差异性,对融合残差进行不同尺度的评测分析,构建出状态监测信号,为装备的分析决策提供参考依据^[5]。数据算法分析如图2所示。

3.2 基准模型的建立

建立基准模型的目的是获取装备运行状态参数的预测数据,为装备状态监测的数据提供比较的参考。在数据算法分析

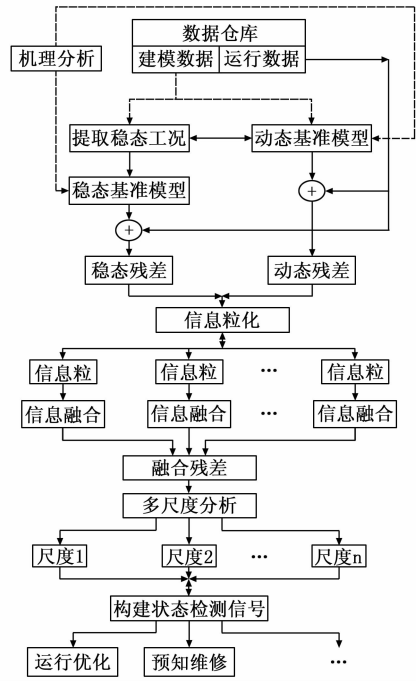


图2 数据算法分析

中,通过预测数据和监测数据两者相比较得到的差值,也称之为残差序列。该差值与装备状态存在某种对应的关系。严格的讲,当装备状态处于正常时,该差值为零,非正常时,该差值将出现较大的振动。

建立基准模型时应考虑以下几点:

1) 运行数据作为基准模型的输入直接影响到输出的预测数据。为保证模型预测的准确一致性,所以选取装备在良好状态下的运行数据是非常必要的。如使用新装备经过磨合期后处于性能较好且稳定状态的运行数据建模。

2) 基准模型并非对象的完备模型。由于参数相互耦合、干扰的存在,高技术装备运行时状态的变化极其复杂过程。因此,构建比较完备的模型是不易实现的。不过,构建模型的目的是获得仅仅与状态监测数据有对应关系的预测数据,要求模型输出的预测数据的变量参数只要能 and 选取的监测数据一一对应即可。因此,模型不必要是完备的,建模也将变得简单可行。

3) 建模方法的选择。可以通过机理分析、数据分析以及两者混合分析的方法建立模型^[6]。无论选择何种方法,都应首先对建立模型的难度、精确度、花费等因素加以权衡。

4) 建模变量的选取。一是如何确定模型的变量,应由建模对象的特性和模型的目的共同决定。通常选取那些具有容易监测且精度较高特性的状态变量。特别是温度、压强这类没有延迟且数据易获得且精准的变量。二是如何确定变量的个数,只要能够满足模型精度的要求即可,因为较多的变量不仅会造成模型的复杂程度,而且不利于模型对数据的处理效率,也会使模型的适应性变差。

5) 区分稳态工况与动态工况的基准模型。装备运行通常会控制在稳定的状态下,也称之为稳态工况。但根据具体情况相应的对装备的运行做出调整是必要且不可避免的。运行状态

的改变，与之对应的是多个参量的变化，这种过程称之为动态工况，它具有持续短暂、动态变化、临时过渡的特点。稳态工况和动态工况具有的不同特点，要求对它们分别建立模型，以便能更真实的准确的描述相对应的装备运行状态。其次，装备运行中，因为部分状态参数存在延迟、滞后，所以稳态工况模型往往适用于已较长时间在稳态运行下的状态描述。因此，区分稳态工况与动态工况分别建立基准模型，更能够较好描述装备运行的真实状态的变化。

3.3 残差及其信息粒化

针对比较模型得到的稳态和动态残差数据规模大的问题，一是采用并行方式对大数据进行处理，可以有效地减少运算时间。二是按各自的特征和性能对数据进行信息粒化，通过相互间的有机结合，根据需要按不同的层次划分出不同的信息粒，然后再对其进行并行的运算处理，这样就可以大大提高运算处理的效率^[7]。

3.4 单一信息粒的异步信息融合

经过信息粒化处理的稳态与动态残差并不是同步信息，而是在时域上的异步信息。运用卡尔曼滤波通过对输出和输入观测数据的信息处理，能够实现了对系统状态进行最优的估计^[7]。

3.5 融合残差的多尺度分析

对于由系统的扰动和观测而形成的噪声，可以通过分析其统计特性进行评估处理，从而使噪声对观测的状态参数的影响处于可控制的范围，即对状态监测精度造成的下降是微不足道的。根据相关的尺度对融合残差进行多尺度的分析、重构，可以得到与不同状态相对应的状态检测参数信号。多尺度分析信息处理过程如图 3 所示。

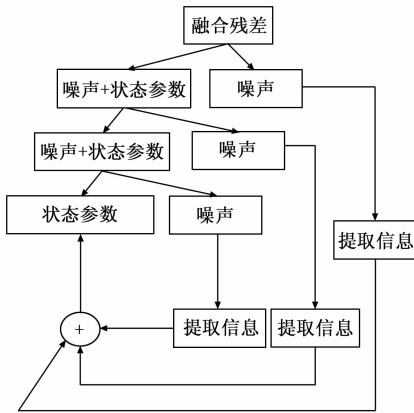


图 3 信息处理过程

4 数据分析挖掘

对数据进行分析挖掘首先应明确目标，特别要清楚需要获得那些信息知识。这个过程具有创造性，需要多种技术和大量知识的综合共同的参与^[8]。在装备运行状态监测系统中应用大数据分析挖掘技术，应时刻围绕者装备预测的目标对数据分析挖掘。在模型中应用大数据分析挖掘技术应明确以下三点，才能避免大数据分析挖掘应用的盲目性^[9]。

1) 理解和定义。数据规模大且变化微弱时，仅依靠人类专家分析处理大数据并发现其中所隐含的应用知识信息是不现实的。例如装备的使用者能够感知到装备总体系统特征的较大幅度的变化，但对部分系统状态特征的小幅度变化所引起的监测数据微乎其微的变化传递的信息知识却很难觉察到。

2) 建立可分析挖掘的大数据集。监测系统中存储的种类多样结构复杂大数据，要围绕数据分析挖掘的目标构建可分析挖掘的大数据集，之后，要对其中数据进行选择、清理、变换和归约四个方面的预处理。这样才能提高大数据集的品质，以便能够为模型的算法所高效处理。

3) 模型的应用。随着装备的长期不断的运行，积累的运行数据的规模越来越大，模型会在不断地数据累积中学习更新，从而智能的发现装备新的运行状态，机器的自我学习提高了大数据分析预测的准确度。

对装备状态监测和预测评估的过程是一个闭环系统，包括信号采集、特征提取、状态识别和预测分析四个步骤。通常一个往复循环并不能对装备状态进行可靠的预测评估，解决的办法就是增加预测评估的循环次数，才能够不断地深化认识和提高预测的准确度。不过随着装备监测系统中存储的数据规模越来越大，带来了新的问题使得对数据的运算难度也不断加大，进行循环的处理过程也将变得越来越困难。

5 结束语

随着大数据分析挖掘技术的不断进步，大数据所隐含的价值也将会慢慢凸显出来。本文把大数据分析挖掘技术应用于装备状态监测系统中，大大提高了装备预测诊断系统的知识获取速率，有益于构建装备生命周期的预测体系，具有很好的应用前景。

参考文献：

[1] 刘继伟. 基于大数据的多尺度状态监测方法及应用 [D]. 北京：华北电力大学，2013.
 [2] 尤明懿，孟光. 基于状态监测数据的产品寿命预测与预测维护规划方法研究 [D]. 上海：上海交通大学，2012.
 [3] 伊恩·艾瑞斯. 大数据思维与决策 [M]. 北京：人民邮电出版社，2014.
 [4] 维克托·迈尔，肯尼思·库克耶. 大数据时代 [M]. 浙江：浙江人民出版社，2013.
 [5] 王永卿，秦亮曦. 高维海量数据聚类算法研究 [D]. 南宁：广西大学，2007.
 [6] Mark Meerschaert, 刘来福. 数学建模方法与分析 [M]. 北京：机械工业出版社，2009.
 [7] 申彦，宋顺林. 大规模数据集高效数据挖掘算法研究 [D]. 镇江：江苏大学，2013.
 [8] 钱超. 高速公路 ETC 数据挖掘研究与应用 [D]. 西安：长安大学，2013.
 [9] 杨风召. 高维数据挖掘中若干关键问题的研究 [D]. 上海：复旦大学，2003.