

# 面向航天信号数据的柔性分析平台设计与实现

姜南<sup>1</sup>, 史梅花<sup>1</sup>, 姚远<sup>2</sup>, 戴伟<sup>3</sup>

(1. 北京控制工程研究所, 北京 100080;

2. 北京途玛胜远科技有限公司, 北京 100195;

3. 北京航空航天大学, 北京 100191)

**摘要:** 航天信号数据是反映产品功能性能变化的重要载体, 基于数据的分析对于故障预测、运维使用乃至产品优化迭代具有重要作用。信号数据分析算法众多, 但存在组织应用层面中存在不能共享、不可灵活配置的瓶颈。该文在分析信号数据分析所面临的数据规范、数据存储和算法配置所面临问题的基础上, 提出并建立了一种.NET 框架技术设计的柔性的信号数据分析平台。在设计数据处理规范的基础上, 使用大对象技术实现了高效数据吞吐, 使用软驱动技术实现算法引擎从而实现对各种算法的调用和柔性算法配置, 从而使组织层面能够共享数据并形成算法生态。实现的信号数据分析平台得到了良好的应用, 内置的算法能够有效实现产品试验与在轨过程的异常监测。

**关键词:** 信号数据; 柔性; 分析平台; 算法配置; 算法引擎; 数据规范

中图分类号: TP315

文献标识码: A

文章编号: 1673-629X(2021)04-0182-05

doi:10.3969/j.issn.1673-629X.2021.04.031

## Design and Implementation of Flexible Analysis System for Signal Data of Aerospace Products

JIANG Nan<sup>1</sup>, SHI Mei-hua<sup>1</sup>, YAO Yuan<sup>2</sup>, DAI Wei<sup>3</sup>

(1. Beijing Institute of Control Engineering, Beijing 100080, China;

2. Beijing Tumashengyuan Technology Company, Ltd., Beijing 100195, China;

3. Beihang University, Beijing 100191, China)

**Abstract:** The signal data is an important carrier describing the performance of aerospace products, and the analysis based on the signal data plays an important role in products' prognostics, on-orbit management, even the optimization and upgrading of products. There are many kinds of available algorithms, but it is difficult to be shared and configured at the application level. By analyzing the bottlenecks of signal data analysis, including data normalization, data storage and algorithm configuration, a flexible platform framework for signal data analysis based on .NET framework is represented. The data normalization principles are designed, then the effective throughput capability is realized using BLOB technology. Based on these, the soft driven algorithms are used to realize the algorithm engines so the data analysis algorithms could be called on demand and the algorithms could be configured flexibly. So, signal data could be shared and the ecological system of analysis algorithms could be set up at the organization level. The signal analysis system is realized and the embedded algorithms could monitor the abnormal signals in test and on-orbit processes for products.

**Key words:** signal data; flexible; analysis system; algorithm configuration; algorithm engine; data normalization

### 0 引言

航天器是一类由机械、电子、材料、控制、能源、通讯、计算机技术等多学科协同工作的大型多功能复杂系统, 具有结构复杂、故障危害性大、意外干扰因素难以把握、星上资源配置约束条件苛刻、地面人工干预的机会和能力受限制等特点<sup>[1]</sup>。因此, 目前知识系统中欠缺的“黑盒子”机理使得设计师无法对产品进行充

分的认知, 也就难以进行有效的故障预测, 在航天器研制和使用中会造成不可估量的损失。在航天产品研制及飞行过程中, 随着自动化技术的持续发展, 采样频率的提高, 计算技术及存储容量的改善, 地面试验和在轨飞行已经积累了海量的数据(该文统称为信号数据), 这些数据反映了产品功能和性能的各种突发的或规律性的变化, 是了解产品性能、开展故障预测的重要依

收稿日期: 2020-05-23

修回日期: 2020-09-24

基金项目: 国防基础科研项目(JCKY2018203C005)

作者简介: 姜南(1974-), 女, 博士, 高级工程师, 从事基于人工智能的故障诊断技术研究。

据。若能获得这些参数的缓慢变化特点并对其趋势进行分析预测,便能够逐步实现对航天器健康状态的有效管理<sup>[2]</sup>。

目前国内外已开展了大量的关于趋势分析及参数预测等故障预测和健康管理研究<sup>[3-5]</sup>，“基于状态信息、基于异常现象信息、基于使用环境信息、基于损伤标尺、数据融合及综合诊断”、“统计预测、数字预测、智能预测、信息融合”<sup>[6]</sup>等分类方法可见诸文献。无论如何分类,基于数据的故障预测方法的基本思想都是在一定代价函数的约束下,对对象的历史数据进行学习,逼近数据中所隐含的映射机制,得到对象的数学模型关系,从而对故障进行预测。时间序列分析方法<sup>[7]</sup>、机器学习(概率趋势、神经网络、支持向量机)技术<sup>[8-10]</sup>都要基于一定的分析目的而设计和应用,不能直接复制已有的算法程序。另一方面,航天产品往往是一个产品族,同一产品族下的产品数据分析有其相通之处,并不能一事一议地临时编写算法获得分析结果,而必须要实现组织层面的分析算法共享。

该文研究了面向航天产品信号数据的分析平台的技术特点,构造了系统的架构,从而在组织层面进行算法的灵活管理与应用。

## 1 航天产品信号数据分析面临的技术问题

### 1.1 多阶段、长时空跨度的数据规范问题

航天产品从开展力学、温度等地面试验到在轨飞行,跨越了极大的时间尺度和空间尺度。虽然是小样本,但数据产生在不同的测试设备、不同传感器中,呈现典型的高维度、多源特征。主要表现在:

(1)不同测试设备对产品的定义不同。容易造成即使是同一个产品,由于经历不同测试设备开展的试验测试,其数据文件在数据库中也是一个一个的孤岛;

(2)测试设备形成的数据文件存储格式也各不相同。测试设备形成的文件有 Txt 文件、Excel 文件、Access 文件等,读取这些不同格式的数据文件,需要采用不同的文件读取技术,降低了数据使用的效率;

(3)不同传感器的采集频率各有不同,时间格式不同,造成不同应力环境下的信号数据无法进行有效的融合;

(4)数据解析方式不同,采集数据中有十进制数据,也有十六进制数据。系统无法自动识别十进制解码表示或是十六进制源码表示。

这些不规范、不标准的数据散落在各个测试系统中,会对后续的数据融合与分析造成极大的困扰。

### 1.2 高采集频率的数据存储和使用问题

航天产品经历众多试验,且设计寿命往往高达10年以上,在试验和在轨飞行阶段采集频率一般为毫秒

级,因此会在较短的时间内生产海量的数据。以1 k/s的采样率计算,连续采集一天,一个参数的数据量将达到7亿行。而这些数据必须能够及时地进行存储和管理。

在使用数据时,往往是以一个参数为单位获取数据,需要获取一个参数在一个大时间尺度的数据。传统的数据库存储方式是将数据按照表的结构进行存储,表的每一列是一个参数,每一行则是对应具体的参数值。但是这个表结构的存储方式在行数大于一亿行以后,存储和查询的效率将会急剧降低。导致的结果就是,将一个几GB的数据文件存进去速度就很慢,当要把其中某一个参数的数值全部获取出来时将更慢,时间将长达十分钟以上。

可见,要实现高速的数据吞吐能力,会对数据库的建设提出非常高的要求。

### 1.3 算法复杂且专业

事实上,无论是航天产品还是其他任何领域的工业产品,简单地应用某种发布的成熟算法进行聚类或异常检测等分析是不现实的。在专业领域的数据分析中,面向产品的物理或者机理知识不可或缺,因此分析算法往往并不能依赖于数据分析专家。航天企业设计师一般具有较高的综合素质与能力,往往可以独立编写分析算法;即使数据分析专家不可或缺,也是要与设计人员的工程知识结合,协同定制地编写算法并验证。因此,算法往往由专业领域人员编写并验证,其编写工具可以是C,可以是Python,也可以是R、MATLAB等等,在调用时需要能够调用不同的软件接口。而在一个分析项目中,也往往需要将多个算法用一个流程来管理和驱动。从另一方面看,不同规格的产品在数据分析中往往具有相似性,因此算法在一个组织内往往需要共享使用。

所以,对不同工具开发的算法进行封装、管理和调用也成为解决信号数据分析的关键技术之一。

## 2 信号数据分析平台的设计与实现

### 2.1 总体架构设计

为实现信号数据在不同目的和场景的数据分析功能,数据分析平台设计包括运行支撑环境层、数据层、应用支撑层、业务应用层和前台展现层,采用.NET框架技术设计,如图1所示。

在运行环境支撑层中,配置工业海量数据分析所需的网络、应用服务器、数据库服务器、数据备份设备等硬件及用于存储的Oracle 10g数据库。

数据层则是数据库中存储的所有数据,包括各类测试设备产生的多源数据和封装的数据分析算法库。

应用支撑层是数据层与应用层之间的功能组件,

包括数据预处理组件、数据存储组件和算法执行引擎。根据用户需求操作数据层中的数据,并将需要数据反馈给应用界面。

业务应用层是系统中实现的业务功能,包括算法库管理、算法分析接口、产品数据管理等,用于对数据分析的业务进行规范性管理。

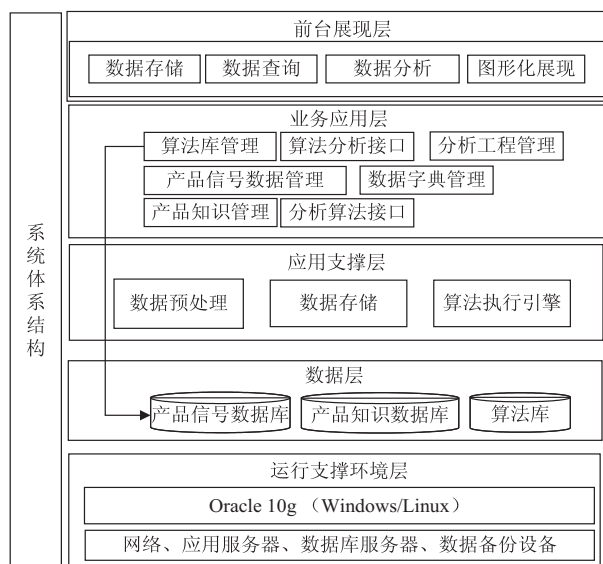


图 1 信号数据分析平台的总体架构

前台展现层则是用户使用和操作的系统软件界面,根据每个用户的特点、喜好和角色的不同,为特定用户提供量身定做的访问关键业务信息的安全通道和个性化应用界面。

## 2.2 关键技术的实现

### 2.2.1 数据预处理规范的设计

数据预处理规范的设计目的是使不同测试设备、不同阶段产生的信号数据相互握手,数据能够规范管理。通过对航天产品信号数据的分析,数据预处理规范设计包括编码、文件格式、时间格式、缺省值标识、数据解析方式等内容。

#### (1) 编码规范。

为使在不同业务系统、不同测试设备中产生的数据能够握手,需要在组织层面建立包括产品代号规范、版本编码规范、编号规范在内的各种编码规范。当信号数据文件存储到数据库时,便可自动解析该数据文件所属的型号及阶段、所属产品及版本,并将数据文件中测试结果与数据字典中的产品参数相关联,实现统一产品源下的数据存储。

#### (2) 数据文件格式标准化处理。

将来自于不同试验设备的不同格式的文件,如 Access 文件、Txt 文件、Csv 文件等统一为 Txt 格式数据文件,作为标准数据文件的文件格式。

#### (3) 时间格式标准化。

针对多源数据在采集时间、频率上的不一致,如时

间格式多种多样,“2018-09-08 12:06:07.245”、“2018/09/08 12:06:07”、“2016-10-09 11:17:37”各不相同。以典型航天执行机构为例,其数据文件至少包括十几种时间格式,因此也需要进行标准化,形成一个统一的时间格式。本平台设计将所有的时间格式标准化为“2018-09-08 12:06:07.245”形式。

#### (4) 缺失值标识标准化。

试验数据和在轨数据往往存在缺失值,而不同来源的数据文件缺失值有“ ”(空格)和“-”两种标识。由于在文件格式标准化处理中采用 txt 格式的数据文件,不能直观看到空格,因此将采用“-”作为数据缺省值的标准标识方式。

#### (5) 数据解析方式可配置。

由于测试设备输出文件有十进制解码表示方式和十六进制源码表示方式,面临着不宜解析的问题,而 XML 可以为解决各种类型的数据共享、交换和使用问题提供一种行之有效的技术解决方法<sup>[11]</sup>。并且,Oracle XML DB 数据库管理系统对 XML (extensible markup language) 文档提供了强大的支持<sup>[12]</sup>。因此,系统基于可扩展标记语言对文档和数据进行结构化处理,设计了一个 XML 配置文件 (Setting.xml) 作为用户对参数的数值表示类型进行配置的文件,如图 2 所示。

```
<?xml version="1.0" encoding="utf-8" ?>
<setting>
  <数据文件设置>
    <数据文件 Name = "CMG数据">
      <参数 Name = "时间" DataType = "datetime"></参数>
      <参数 Name = "编号" DataType = "float"></参数>
      <参数 Name = "转速" DataType = "float"></参数>
      <参数 Name = "位置" DataType = "float"></参数>
      <参数 Name = "电流" DataType = "float"></参数>
      <参数 Name = "电压" DataType = "float"></参数>
    </数据文件 Name = "CMG数据">
  </数据文件设置>
</setting>
```

图 2 setting.xml 文件示意图

### 2.2.2 高效率的数据吞吐技术

各类测试设备获取的信号数据均以文件形式传输。这些文件具有特定的格式和数据存储方式,在数据库系统内属于 BLOB 数据类型,这是一种用于存储无边界数据的数据类型,将大数据以 BLOB 方式存在数据库中便于管理且具有高可用性<sup>[13-14]</sup>。为实现高效率的数据吞吐,在数据层采用 Oracle 10g 数据库管理软件作为基础,利用大对象技术 (Blob),采用自定义数据存储结构存储产品数据。

系统在存储数据时,首先在内存中对数据进行缓存和聚集,将不同种类的参数的全部参数值进行分类聚集,如图 3 所示。然后再将聚集好的数据块,以数据块 Blob 技术存储到数据库中划分好的磁盘空间。

经测试,以 Blob 形式进行数据存储的速度大于 10 M/s;以 Blob 形式获取 100 M 数据,并绘制可视化时

序图的时间小于 1 s。

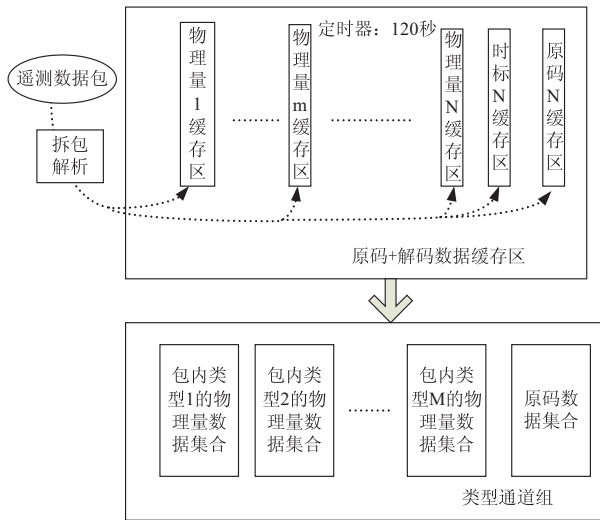


图 3 内存中参数值聚集示意图

### 2.2.3 柔性的算法配置

鉴于算法是由不同设计师采用不同的开发工具开发并验证,然后在平台上加载,因此需要在软件中实现算法接口调用的多样性,从而具备在一个分析项目中组合多种算法的柔性能力。

经调研,目前设计师进行算法编写的工具一般是 C#、Python、R、Matlab。因此解决方案是在 .NET 平台上由 C# 编程语言编写的软驱动。C# 是以事件为驱动

的可视化编程语言,可用于方便快捷地开发运行在 .NET 公共语言库上的应用<sup>[15]</sup>。其实现的软驱动便可天然地兼容调用 C# 编写的数据分析算法 dll 文件。当调用 Python 算法和 MATLAB 算法时,则使用混合编程技术来实现算法执行引擎。以 Python 算法为例,可以使用 IronPython 开发执行引擎。IronPython 是一个 .NET 平台上的 Python 实现,包括了完整的编译器、执行引擎与运行时支持,能够与 .NET 已有的库无缝整合到一起,所以 IronPython 和 C# 的交互也得以实现<sup>[16]</sup>。

通过数据分析引擎,系统就可以直接调用不同设计师采用不同工具开发的算法。在实际分析工作中,并不可能只用一个算法来实现分析,即便如简单的时间序列分析,也需要进行野值剔除、数据拼接、可视化等不同算法。数据分析引擎的实现使得用户得以对不同算法进行友好的组合,实现柔性的算法配置。

## 3 应用案例

信号数据柔性分析平台在笔者所在单位进行了良好的应用。以典型航天执行机构为例,其试验数据来自于多个测试设备,具有 Excel 格式、Access 格式、Csv 格式等,这些不同格式的数据文件经规范处理后,能够在系统中进行有效的管理,如图 4 所示。



图 4 不同测试设备的不同格式文件数据规范化存储

目前在系统中已经对各类常用数学函数、常用的机器学习算法(异常检测、聚类)实现了算法封装,设计师可以进行灵活的组合和调用。图 5 所示为系统

对算法的管理,图 6 为某产品使用基于动态窗口的异常监测算法进行异常判读的结果。



图 5 算法管理窗口

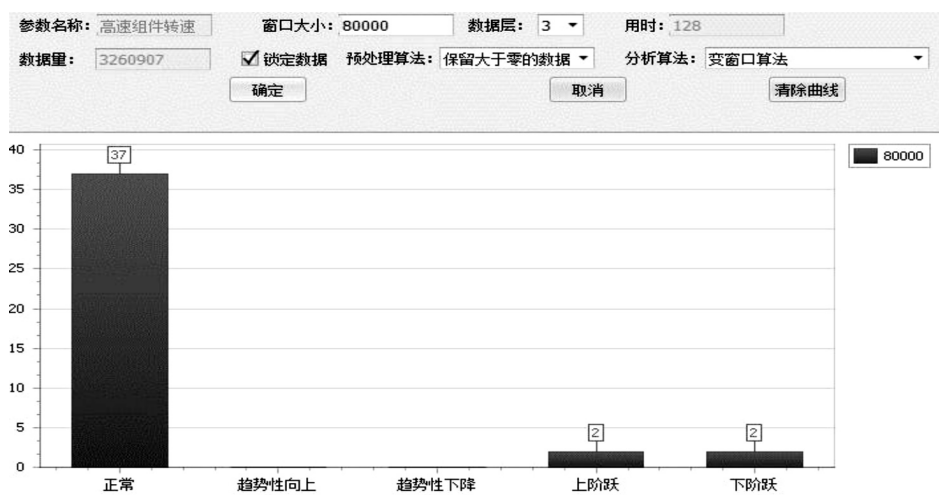


图 6 基于动态窗口的异常监测

#### 4 结束语

针对航天产品信号数据多源长时空跨度的特点,分析了将来使用机器学习等人工智能技术进行应用所必须解决的软件平台的关键技术,设计并实现的柔性分析平台解决了数据规范、数据吞吐和算法配置的关键技术。

此平台是应用人工智能技术进行各种目的的数据利用的基础。目前算法虽只包含较为基础的相关分析、聚类分析等,主要用于产品的健康管理。但平台提供的算法引擎有效支持了设计师系统提供智慧的工作,算法也将不断积累和完善,信号数据的分析将为航天产品的不断改进发挥更多的作用。

#### 参考文献:

- [1] 尹洪. 基于数据驱动的卫星故障诊断关键技术研究[D]. 长沙:国防科技大学,2015.
- [2] 周枫,皮德常. 一种时间序列分解的卫星周期性参数预测方法[J]. 计算机科学,2016,43(2):9-12.
- [3] LEE J. Measurement of machine performance degradation using a neural network model[J]. Computers in Industry, 1996,30(3):193-209.
- [4] HUANG R, XI L, LI X, et al. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods[J]. Mechanical Systems and Signal Processing, 2007, 21(1):193-207.
- [5] TALBOT D, LI S, KAHRAMAN A. Prediction of mechanical power loss of planet gear roller bearings under combined radial and moment loading[J]. Journal of Mechanical Design, 2013, 135(12):121007.
- [6] 左宪章,康健,李浩,等. 故障预测技术综述[J]. 火力与指挥控制, 2010, 35(1):1-5.
- [7] LIU Datong, PENG Yu, PENG Xiyuan, et al. Online adaptive status prediction strategy for data-driven fault prognostics of complex systems[C]//IEEE prognostics and system health management conference. Shenzhen: IEEE, 2011:1-6.
- [8] 王安娜,邱增,吴洁,等. 基于 SVM 多类分类算法的模拟电路软故障诊断[J]. 东北大学学报:自然科学版, 2008, 29(7):924-927.
- [9] CODETTA-RAITERI D, PORTINALE L. Dynamic Bayesian networks for fault detection, identification, and recovery in autonomous spacecraft[J]. IEEE Transactions on Systems, Man & Cybernetics Systems, 2015, 45(1):13-24.
- [10] 程月华,姜斌,杨明凯,等. 应用自组织模糊神经网络估计卫星姿态系统执行机构故障[J]. 应用科学学报, 2010, 28(1):72-76.
- [11] SU Houqin, SU Jinqian. A study and practice of report system techniques based on three-layer calculating architecture[C]//2nd international workshop on education technology and computer science. Wuhan, China: IEEE, 2010.
- [12] 任金铜. 基于 Oracle XML DB 的 GML 存储、查询及索引研究[D]. 赣州:江西理工大学, 2009.
- [13] 江凌,杨平利,齐芳丽,等. 基于 ADO.NET 的 BLOB 数据存取方法研究[J]. 软件导刊, 2014, 13(5):129-131.
- [14] 郝世艳. 在 Oracle 数据表中存取 BLOB 数据的方法研究[J]. 信息系统工程, 2014(9):143.
- [15] 宋艳朋,周兴华,付延光,等. Matlab.NET 混合编程在潮汐调和与分析预报中的应用[J]. 测绘与空间地理信息, 2017, 40(5):59-61.
- [16] FOORD M J, MUIRHEAD C J. IronPython: IronPython in action[M]. [s.l.]: Manning Publications Co., 2009.