

# 一种基于数据挖掘的无人飞行器故障分析方法

孟魁杰<sup>1</sup>, 董莹<sup>2</sup>, 赵宗涛<sup>2</sup>

(1. 陕西师范大学 计算机科学学院, 陕西 西安 710062;

2. 第二炮兵工程学院, 陕西 西安 710025)

**摘 要:** 无人飞行器的设计与应用包含了高科技技术的诸多方面, 是一个极为复杂的机电系统。为保障可靠飞行, 而对其进行故障检测是一件必不可少的工作。提出一种基于故障库(FRDB)的数据挖掘技术, 对无人飞行器进行了故障分析研究, 并设计了相应的算法。从FRDB数据库中采用选择、投影操作, 分离出相关部件的超差字段, 形成相应的数据文件, 以此建立联机统计分析的基础。而后, 又对发生故障的分布规律进行统计分析。测试中重点放在无人飞行器控制系统的各部件, 为实现统计分析的自动化和半自动化, 必须重视测试数据的规范化处理和数据库的建立。

**关键词:** 数据挖掘; 无人飞行器; 故障分析

**中图分类号:** TP311

**文献标识码:** A

**文章编号:** 1673-629X(2010)06-0225-03

## A Fault Analysis Method Based on Data Mining for Unmanned Aerial Vehicle

MENG Kui-jie<sup>1</sup>, DONG Ying<sup>2</sup>, ZHAO Zong-tao<sup>2</sup>

(1. Institute of Computer Science, Shaanxi Normal University, Xi'an 710062, China;

2. The Second Artillery Engineering Institute, Xi'an 710025, China)

**Abstract:** Design and application of unmanned aerial vehicle include many aspects of high technology, and it is a highly complex electro-mechanical system. To ensure the reliability of flight, it is an essential work to perform fault detection. A data mining technology based on FRDB is proposed to study fault analysis to unmanned aerial vehicle, and also the corresponding algorithm is designed. Selection and projection operation from FRDB are used to separate the super-bad fields of the relevant part and form the corresponding data files to establish the basis of statistical analysis. Besides, statistic analysis is performed to the distribution of the fault. In testing procedure, the components of the unmanned aerial vehicle control system are paid high attention. In order to achieve the automation and semi-automation of statistical analysis, normalized treatment to the testing data and the database building must be paid attention.

**Key words:** data mining; unmanned aerial vehicle; fault analysis

## 0 引言

无人飞行器是一种极为复杂的机电系统, 其类型很多, 文中主要以航天火箭为例进行讨论。据资料统计, 一枚中等规模的三级火箭, 它包括的部件多达两万多<sup>[1]</sup>个之多。如果作为武器应用, 连同战斗部, 它的造价可达上百万美元<sup>[1,2]</sup>。因而, 对其进行故障分析是一件非常重要的工作, 它关系到设计、试验和应用的全过程。文中提出一种基于故障库 FRDB 的数据挖掘技术, 对火箭进行故障分析, 不仅支持了火箭应用工程, 而且为设计、维修、部件采购提供了相应的理论支持。

在技术上, 这种基于 FRDB 的故障分析诊断方法, 将数据挖掘技术的模型设计与故障诊断相结合, 在工程应用中取得了良好的效果。

## 1 算法设计与分析

故障库(FRDB)是采用关系数据库的系统软件, 如 Oracle。库内所存放的数据是年来对火箭部件进行测试的历史数据记录, 也包括产品出厂时的一些原始记录数据。数据大体按不同部件分类统计, 主要有战斗部系统、控制系统、动力系统、计算机系统、地面发射系统、电源系统、稳定基准系统等<sup>[3]</sup>。

数据挖掘的模型抽取有多种方法<sup>[4]</sup>。以下研究在数据挖掘中最常用的过滤筛数学模型, 以此来挖掘可能发生故障的数据, 并找出故障在各部器件上的分布。

收稿日期: 2009-09-17; 修回日期: 2009-12-16

基金项目: 中国人民解放军第二炮兵预研项目(2007014016)

作者简介: 孟魁杰(1983-), 男, 吉林通化人, 硕士研究生, 研究方向为数据库技术; 赵宗涛, 教授, 研究方向为数据库技术。

为了抽取相关的数学模型,采用矩阵形式来表达测试数据文件、出厂数据文件和误差限文件,以便于归纳其检索的算法<sup>[5]</sup>。

### 1.1 定义 1

称矩阵  $A$ 、 $B$ 、 $\Delta$  分别为测试数据文件  $F_A$  出厂所标定的数据文件  $F_B$  和误差限文件  $F_\Delta$  所对应的测试数据矩阵:

$$A = (a_{ij}), i = 1, 2, \dots, n, j = 1, 2, \dots, m$$

$$B = (b_{ij}), i = 1, 2, \dots, n, j = 1, 2, \dots, m$$

$$\Delta = (\delta_{ij}), i = 1, 2, \dots, n, j = 1, 2, \dots, m$$

其中任一行  $a_{i1}, a_{i2}, \dots, a_{ij}, \dots, a_{im}$  代表某一部件一个属性值记录,任一行  $b_{i1}, b_{i2}, \dots, b_{ij}, \dots, b_{im}$  代表某一部件一个标准值记录,  $\delta_{ij} (i = 1, 2, \dots, m, j = 1, 2, \dots, n)$  是部件属性值与标准值所允许的超差限,一般是根据出厂产品测试时所设定的误差限和长期使用的经验所确定的。

### 1.2 过滤筛模型

$$|a_{ij} - b_{ij}| \leq \delta_{ij} \quad \delta_{ij} (i = 1, 2, \dots, m, j = 1, 2, \dots, n) \quad (1)$$

$$v(t) = v(0) + \int_0^t a(t) dt \approx \sum_{i=1}^m \Delta t_i * a(t_i) \quad (2)$$

(1) 式是检查系统特征值是否超差,并作记录。

(2) 式是诊断计算机系统是否能作正常的积分运算,对动态仿真测试的参数值  $a(t)$  进行累加,求得时间  $t$  时的  $v(t)$  值,用以控制动力系统的工作。弹道式导弹的主动段飞行主要是用速度  $v(t)$  进行控制,其在时间  $t$  的值是通过加速度  $a(t)$  进行积分求得的。通常,在给定误差限制厦,积分运算近似变成求和累加积分。

前述一些数学模型,如(1)式所示主要是检测给定的仪器仪表是否能在给定的误差条件下工作,也就是属于单元仪器的静态测试。(2)式是属于仿真条件下所进行的系统综合测试。文中用零矩阵来记录超差字段,而计算机系统另行记录出故障的次数,用以评估计算机工作的性态。

### 1.3 数据仓库建立

#### 1.3.1 数据仓库的建立

所谓数据仓库是将矩阵  $A$ 、 $B$ 、 $\Delta$  中所涉及的数据从 RDB 中分离出来,以应用为原则来加以集中<sup>[6]</sup>,通常要以表格形式编排成数据文件,写入内存中。可以用以下 RDB 的选择运算来组建数据仓库。

$S \text{ RDB} = F_A \quad (1, 2, \dots, l_1 \text{ 为测试数据文件中的所选项段值})$

$S \text{ RDB} = F_B \quad (1, 2, \dots, l_2 \text{ 为出厂数据文件中的所选项段值})$

$S \text{ RDB} = F_\Delta \quad (1, 2, \dots, l_3 \text{ 为误差限数据文件中的所选项段值})$

$F_A$ 、 $F_B$ 、 $F_\Delta$  的文件格式基本相同,如表 1 所示。

表 1 基本存储格式

编号	名称	出厂厂家	出厂时间	...	额定电流电压
1	000001	计算电源	199 厂	2001.3	... +5V,60mA
2	000002	频率计	299 厂	2002.7	... 220V,20mA
3	000003	计时器	258 厂	2001.11	... 220V,15mA
...	...	...	...	...	...

为了统计分析方便,对选择所得的表格记录还需按特征值进行投影运算,截取相应的字段,例如对文件  $F_A$ 、 $F_B$ 、 $F_\Delta$  按超差的特征值字段进行投影。以下是其运算公式:

$$P_{a=1,2,\dots,l_j} F_i = F_{i_d}, i = A, B, \Delta; j = 1, 2, 3 \quad (1, 2, \dots, l_j \text{ 为所需的属性名})$$

其中对仿真测试的动态  $a(t)$  值所对应的文件如表 2 所示。

表 2 动态  $a(t)$  值

	$a(t_1)$	$a(t_2)$	$a(t_3)$	...
1	1.003	1.001	1.000	...
2	1.010	1.000	1.001	...
3	1.008	1.005	1.006	...

表中取时间序列  $\Delta t = 0.0625s$ ,  $a(t)$  值为相应时间内的速度改变量,其单位为  $m/s^2$ ,表中所填的小数精确到小数点后三位。其中  $S$  - 选择运算,  $P$  - 投影运算, RDB 是给定的关系数据库。一般情况下,数据仓库相应文件数据应写入内存或相当的快速存储中,以便于进行统计分析。

#### 1.3.2 算法设计

利用公式(1)逐个比较测试数据  $a_{ij}$ ,若有误差,逐个比较求差值,并与相应误差相比较,将其差值记入矩阵  $D$  中。

$$D = (d_{ij}) \quad i = 1, 2, \dots, m, j = 1, 2, \dots, n$$

对于计算机系统,将出错次数和所设置的测试数据集另行记录,以便进行故障分析。现将测试数据误差检测与记录算法(DTRA)概括如下:

DTRA:

Step1:测试各系统部件表达特征的数值,逐个记录到矩阵  $A$  中,这些特征值一般是:  $a_{i1}$  是某仪器仪表的编号,  $a_{i2}$  是其名称,  $a_{i3}$  是出厂的厂家,  $a_{i4}$  是出厂时间.....。

Step2:整理测试矩阵  $A$ ,以最长的字段为准,记录短的字段,后补零与长字段看齐。按从左自右,从上到下的顺序检查各字段的值。如:  $a_{11}, a_{12}, \dots, a_{1n}, a_{21}, a_{22}, \dots, a_{2n}, \dots, a_{m1}, a_{m2}, \dots, a_{mn}$ 。

Step3:将表征  $a_{ij}$  性能的数值与标准值的差求绝

对值。

Step4:将差值再与允许的标准规定误差进行比较。将超差值记录到矩阵  $D$  中。

Step5:设置数据列,检查计算机的积分功能。数据列的设置及测试路径选择另行讨论。

在实测计算机工作过程中登记以下数值:输入数据出错位置与个数,模拟加速度表实时工作的数据,也就是计算机接收脉冲的个数及换算数值等,此外,还有计算机加载解算相关方程的功能。

Step6:结束。

### 1.3.3 复杂度分析

按上述算法 DTRA,记录时的次数为  $mn$  ( $m$  是数据矩阵的行数,  $n$  为其列数,下同)。求其误差的次数为  $mn$ ,求绝对值的次数为  $mn$ ,求再记录的次数为  $mn$ ,累计为  $4mn$ 。因而其计算复杂度为:

$$T_{DTRA}(m, n) = O(4mn) \quad (3)$$

由此可以看出,求测试数据差值的计算复杂度取决于检测仪表个数与表征特征值字段的多少相关。一般来说,字段的长短一般是用 20 ~ 50 的两位数表达,充其量取最大值 99;但  $m$  所表征的是火箭部件中仪表的个数,往往多达数千乃至上万个。因而将复杂式可以改成:

$$T_{DTRA}(m, n) = O(c * n) \quad m - \text{const}, c \leq 99 \quad (4)$$

也就是说,其复杂度为线性式的阶。因而是要节省计算机检测记录的时间,一般来说只检测经常出故障的一些重点部件,可以有效地节省时间。若考虑动态仿真测试数据的处理,则相应的表格要反复访问  $p$  遍,此时相应的计算复杂度(3)式为:

$$T_{DTRA}(m, n, p) = O(4mnp) \quad (5)$$

(5) 式中  $p$  往往变化范围可达三位数,因而在仿真动态的情况下,其计算量往往比较大。

## 2 统计分析

在数据挖掘的基础上,一是已经找出单元静态测试超差仪器仪表的数据文件,二是将其动态综合测试的一些数据也已挖掘出来并建立了相应的文件。这二者是统计分析的基础<sup>[7]</sup>。有了静态故障检测结果文件,动态仿真测试数据文件以及相关的统计计算公式,就可编程进行各种数据分析<sup>[8]</sup>,找出故障分布规律、各部件故障所占百分比,以及计算量统计等。根据统计分析情况主要结果如下:

在静态单元仪器仪表测试中,超差部器件数量主要分布是:导航控制系统占 41%,动力系统占 9.2%,弹头占 7.3%,计算机系统占 4.2%,发射设备占 19.

7%,其它占 18.6%。在动态合成模拟测试中,采用的手段是混合仿真组合,部分单元仪器仪表以实物接入,而部分是以模拟器件或嵌入式软件模块接入,所得的结果与静态测试不完全相同,主要是发生故障的部器件数量百分比有所变化,制导软件由 2% 上升到 7%,电源系统由 3% 上升到 4.2%,制导控制系统降低到 18%,详情见图 1。

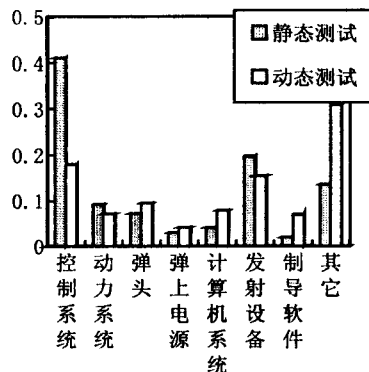


图1 静态与动态测试时各部件的超差百分比

在速度  $v(t)$  仿真试验中,模拟飞行总共 360 秒,取  $\Delta t = 0.0625s$  的间隔传送其改变量  $a(t)$ ,经积分  $v(t)$  增长的数量曲线如图 2 所示。由图可见,其变化规律与预先设计的曲线基本相符。

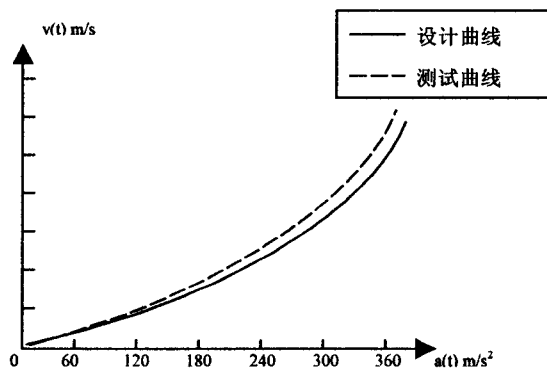


图2  $v(t)$  增长数量曲线

静态单元测试,若将误差值  $\Delta$  均缩小一半,则各部件发生故障的数量均有较大的增长;若将  $\Delta$  放宽一倍,则故障的数量均有较大的减少,相应的统计图见图 3。

## 3 结束语

(1)制导控制系统所占故障数约为 40% 左右。其主要原因是电子系统嵌入机械设备中,其接口工艺难以保障高成功率的结合。

(2)从动态测试所列数据和相关图像可以看出,制导软件的设计缺陷均容易发生故障。设计缺陷是经过大量的运行才能暴露的。

(下转第 232 页)

### 3 结束语

由于采用了回答集程序设计,文中的自动排课系统能够灵活地定义排课问题的基本需求,同时也能根据用户的需求定义各种约束。通过一个用户界面,用户定义的需求可以直接被解释为回答集程序的约束进行求解。高效的回答集求解器的使用也使得系统的效率和直接实现的求解算法相当。

实验结果显示该系统的运行效率和求解的问题规模与目前已有的系统相当<sup>[10]</sup>。实验中处理了最大规模为 240 课元的直接排课过程。如果考虑到实际系统中对年级和院系的分层,这个规模已足够适用于一般的中学和大学的排课系统。

由于排课系统能灵活地处理各类约束,获得了较高的教师满意度,因此在实践和理论研究中都是非常有意义的。

#### 参考文献:

- [1] Even S, Itai A, Shamir A. On the complexity of timetable and multicommodity flow problems[J]. SIAM J. of Computing, 1976(4):691-703.
- [2] Cooper T B, Kingston J H. The complexity of timetable construction problems[C]//Proc. ICPTAT'95. New Orleans, Louisiana, United States:[s. n.],1995:183-295.

- [3] Gueret C, Jussien N. Building university timetables using constraint logic programming[J]. Information and Computation, 1995(8):12-23.
- [4] Gelfond M. Answer sets[M]//In Handbook of Knowledge Representation. [s. l.]:Elsevier,2007:285-316.
- [5] Dantsin E, Eiter T, Gottlob G, et al. Complexity and expressive power of logic programming[J]. ACM Computing Surveys,2001,33(3):374-425.
- [6] Neumann A, Schaub T, Gebser B, et al. Conflict-driven answer set solving[C]//In Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI2007). Menlo Park: AAAI Press,2007:386-392.
- [7] Simons P, Niemela I, Soeninen T. Extending and implementing the stable model semantics[J]. Artif. Intell.,2002, 138 (1-2):181-234.
- [8] Syrjanen T. Lparse 1.0 user's manual[M]. [s. l.]:GPL, 2000.
- [9] Gottlieb C. The construction of class-teacher timetable[C]//Information processing, Proc. IFIP Congress. Vancouver, British Columbia, Canada:[s. n.],1962:73-77.
- [10] 陈章辉, 黄小晖, 任文艺. 基于双倍体遗传算法求解大学排课问题[J]. 计算机应用,2008,28(12):3074-3076.
- [11] 王 璐, 邱玉辉. 基于协商的智能排课系统的研究[J]. 计算机科学,2006,33(6):214-217.

(上接第 227 页)

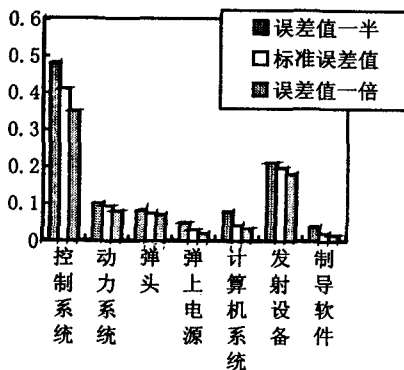


图 3 静态测试时各部件随  $\Delta$  变化的超差百分比

(3) 关于  $v(t)$  动态测试与积分运算,根据实践经验,主要是精度控制。影响精度的主要因素是精度步长  $\Delta t$  的选择。必须理论分析与实践相结合来优选步长大小,在积累误差和积分舍入误差中间求其优化点。

通过利用数据挖掘来对故障进行分析,其优点是可以利用长期积累的测试海量数据分析各部件在可能发生故障中所占的比例,也就是故障的分布规律。这可以有效地指导人们的测试任务的分配,也可为维修零件的采购提供建议。

#### 参考文献:

- [1] 吴文正, 周慧中, 于广汇, 等. 导弹引论[M]. 北京: 国防工业出版社, 1990.
- [2] 余超志. 导弹概论[M]. 北京: 国防工业出版社, 1982.
- [3] 崔吉俊. 火箭导弹测试技术[M]. 北京: 国防工业出版社, 2000.
- [4] Painter M K, Erraguntla M, Jr Hogg G L. Using Simulation, Data Mining, Knowledge Discovery Techniques for Optimized Aircraft Engine Fleet Management[C]//Proceedings of the 2006 Winter Simulation Conference. Monterey, USA: IEEE Press, 2006:1253-1258.
- [5] Han Jiawei, Kamber M. 数据挖掘: 概念与技术[M]. 范明, 孟小峰, 等译. 北京: 机械工业出版社, 2001.
- [6] Kumaz S, Cetin O, Kaynak O. Fuzzy Logic Based Approach to Design of Flight Control and Navigation Tasks for Autonomous Unmanned Aerial Vehicles[J]. Journal of Intelligent and Robotic Systems, 2009,54(3):230-236.
- [7] 宋远芳. 基于本体的数据挖掘技术在商务智能中的应用[J]. 计算机技术与发展,2009,19(1):184-186.
- [8] 邵晓巍, 邵长胜, 赵长安. 一类军用无人飞行器协作效能评估方法研究[J]. 导弹与航天运载技术,2005(1):60-62.