

基于决策树方法的液体火箭发动机 稳态段故障诊断

韩泉东, 胡小平, 王艳梅

(国防科技大学 航天与材料工程学院, 湖南, 长沙 410073)

摘 要: 以泵压式液体火箭发动机为研究对象, 针对故障特征难以提取的问题, 提出了在大量试车数据的基础上, 应用决策树方法提取故障特征并进行稳态段故障检测与诊断的思路, 对某型发动机大量热试车数据进行了实例分析, 将得到的故障检测和诊断结果与神经网络等方法所得结果进行了比较。并利用仿真数据对该方法的性能进行了进一步验证。研究表明, 决策树方法是一种有效、可靠的液体火箭发动机故障检测与诊断的新途径。

关键词: 液体火箭发动机; 故障检测与诊断; 数据挖掘; 决策树

中图分类号: V434

文献标识码: A

文章编号: (2007) 03-0026-06

Steady state fault detection and diagnosis of liquid propellant rocket engines based on decision tree method

Han Quandong, Hu Xiaoping, Wang Yanmei

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

Abstract: Because of noise and the coupling of many factors in the hot fire tests of rocket engines, the detection of fault features was always difficult. The decision tree method was applied to analyze steady test data for liquid rocket engines. Detection and diagnosis results were compared with those of ANN method. Then the simulated data were used to verify the performance of decision tree method. It showed that decision tree method was an efficient, reliable and novel approach for the FDD of LRE.

Key words: liquid propellant rocket engine; fault detection and diagnosis; data mining; decision tree

收稿日期: 2006-06-21; 修回日期: 2006-11-15。基金项目: 国家自然科学基金资助项目 50376073。

作者简介: 韩泉东 (1981—), 男, 硕士, 研究领域为火箭发动机故障诊断, 数据挖掘。

1 引言

有效地提取出故障特征是故障检测和诊断的关键。液体火箭发动机是复杂的流体-热动力系统,由于环境干扰的多样性、故障特征的多样性、故障的多样性以及内部因素的耦合表现出的强非线性,使得故障特征的提取难以进行,从而给发动机的故障检测和诊断带来了极大困难。目前常用的故障诊断方法主要有:基于信号分析的方法,基于动力学模型的方法以及基于人工智能的方法。由于液体火箭发动机的强非线性,难以建立精确的数学模型,传统的故障诊断方法由于不能有效地简化知识和压缩样本空间,从而影响了诊断的适应性和效率。基于信号分析的方法由于随机干扰以及各种瞬态过渡过程的存在,使得该方法在检测故障的及时性和准确性方面存在一定困难;基于动力学模型的诊断方法对模型过于依赖,而发动机系统难以建立精确的数学模型,从而诊断结果的可靠性不能严格保证;基于人工智能的诊断方法需要足够的典型故障样本和先验知识,而现实当中发动机故障样本很少。因而这些理论上的方法难以得到广泛的应用^[1-3]。

数据挖掘从大量数据中提取隐含在其中的事先未知的、但潜在有用的信息和知识。决策树方法是数据挖掘中重要的分类方法之一,它通过分析输入的数据,利用树结构将数据进行分类,从中找到有价值的、潜在的信息并通过训练得到判定规则,再用它来对更多的数据进行分类^[4,5]。

发动机试车数据多,但并未得到充分的利用。将决策树方法应用到液体火箭发动机的试车数据挖掘中,有望从大量的数据样本中挖掘出隐藏着的对故障检测和诊断具有重要参考价值的信息,对火箭发动机这种复杂系统的故障诊断具有重要意义。

2 决策树方法的基本原理

决策树是一个类似于流程图的树结构,其中每个内部节点表示在一个属性上的测试,每个分

枝代表一个测试输出,而每个树叶节点代表类或类分布。决策树学习是以实例为基础的归纳学习,它以分类和决策为目的。决策树知识表示就是把专家知识隐含地表示在决策树的内部节点和叶节点上,并根据内部结点上的属性及其取值表达知识的条件部分,而在叶节点上得到知识的结论部分^[4,5]。

目前常用的决策树算法是基于信息熵的算法,其核心算法目前主要有 ID3、C4.5、C5.0、CART、SLIQ、SPRINT 和 Rain Forest 等几种^[6]。根据火箭发动机试车数据是连续属性且现有训练集不是特别大的特点,选用了 C4.5 算法。

基于决策树方法的故障诊断的关键是决策树的构建。主要分为两步:树的建立和树的剪枝。

第一步,树的建立。决策树的建立一般采用自上而下的方式。找出最佳扩展属性是生长决策树的关键,它需要首先确定扩展属性的形式,找出符合这种形式的所有扩展属性,然后评价这些扩展属性,选出最具有判断能力的扩展属性作为最佳扩展属性。

定义 1 信息增益。设 $I(T)$ 为集合 T 根据类标号(即结论)属性(设有 m 个不同取值)被分为 m 个不同类 C_i ($i=1, \dots, m$) 所得的信息熵, t_i 表示 C_i 中的样本数,根据非类别属性 A (即条件)将集合 T 可划分为 n 个不同类 T_j ($j=1, \dots, n$), t_{ij} 表示子集 T_j 中类 C_i 的样本数,则根据 A 对 T 进行划分的信息增益为

$$\text{Gain}(A, T) = I(T) - \sum_{j=1}^n \frac{t_{j1} + \dots + t_{jm}}{t} \times I(t_{j1}, \dots, t_{jm}) \quad (1)$$

其中

$$I(T) = - \sum_{i=1}^m p_i \times \log_2(p_i)$$

$$I(t_{j1}, \dots, t_{jm}) = - \sum_{i=1}^m p_{ij} \times \log_2(p_{ij})$$

式中, $p_i = t_i/t$ 表示 T 中的样本属于类 C_i 的概率;

$p_{ij} = t_{ij}/t_j$ 表示 T_j 中的样本属于类 C_i 的概率。

定义 2 信息增益率。根据非类别属性 A 对训练集 T 进行划分的信息增益率为

$$\text{GainRatio}(A, T) = \text{Gain}(A, T) / \text{SplitInfo}(A, T) \quad (2)$$

其中

$$\text{SplitInfo}(A,T)=-\sum_{j=1}^n(t_j/t)\times\log_2(t_j/t)$$

C4.5 算法选择具有最高信息增益率的属性作为最佳扩展属性。C4.5 算法既可以处理离散值属性又可以处理连续值属性。对于发动机试车数据这种连续值属性，规定在属性 A 上的测试导致两个分支，分别对应于条件 $A\leq V$ 和 $A>V$ ，其中 V 称作局部阈值。若 A 是最佳扩展属性，则 V 称作阈值。要确定 A 的局部阈值，首先对 T 中 A 属性值已知的样本进行快速排序，依次考察排序后的每对相邻值的中间值 v ，以及对应的划分条件 $A\leq v$ 和 $A>v$ 。假定样本中 A 有 m 个不同取值，则存在 $m-1$ 个中间值 v ，分别对应 $m-1$ 个可能的信息增益率 GainRatio v' 。若某个 GainRatio v' 值最大，则此 v' 就是 A 的局部阈值。遍历完所有属性，便可得到最佳扩展属性^[9]。

第二步，树的剪枝。决策树构造时，许多分枝可能反映的是训练数据中的噪声或孤立点，树的剪枝就是试图检测和剪去这种分枝以解决过度适应数据问题，用一个叶子节点来代替一个子树，以提高在未知数据上分类的准确度。若一个决策规则检测到其子树中的误分比例大于单个树叶的比例时，就将该节点的子节点剪去。可采用交叉验证或人为选取部分数据进行验证的策略对树进行评估。

3 实例分析

对于泵压式液体火箭发动机而言，起动过程中参数变化非常剧烈，而稳态段的数据则相对稳定，因此对于启动段和稳态段的故障诊断，须区别对待。本文利用决策树方法对发动机稳态工作期间的试车数据进行了挖掘。

3.1 数据预处理

数据预处理是数据挖掘中一个重要环节。为数据挖掘所提供的原始数据往往缺乏一致性，存在大量冗余、空值及不完整性，数据预处理就是对所提供的原始数据进行加工，检查数据的完整性及数据的一致性，对其中噪声数据进行处理。

主要进行了以下的数据预处理：

(1) 数据集成与简化。发动机试车测量的参数种类很多，数据量非常大，数据采集的步长也可能不一致，从而给数据的选取带来了极大的困难。对此，研制了相应的数据提取算法，可以在同一次试车的多个数据文件中，快速提取出所需要的数据样本。根据多次试车数据中共有的且参数值相对稳定的原则，从 80 多个测量参数中初步提取出了 30 个参数的数据。然而，海量的实际数据中无意义的成分严重影响了数据挖掘算法的执行效率，故应该有效地缩减数据。相关统计表明，涡轮增压的故障率最高，对发动机的失效影响最大；频繁发生的推进剂输送管路故障也会对发动机工作产生致命后果^[1]，故主要选取与涡轮增压及流量调节器相关的参数，将上述 30 个参数进一步压缩到 19 个，用 $X_1\sim X_{19}$ 表示，如表 1 所示。试车实测数据可能变化较大，为了减小阈值的搜索范围以提高数据挖掘的效率，采用 $(X_{ij}-\bar{X}_j)/\bar{X}_j$ (简记为 $\text{dev_}X_{ij}$) 的方法对训练数据进行了变换，其中 X_{ij} 表示第 i 个样本中第 j 个参数的测量值， \bar{X}_j 表示该型号发动机多次正常试车稳态段第 j 个参数的统计均值。

表 1 训练集所选参数

Tab.1 Description of the selected parameters

参 数 名 称	代号	参 数 名 称	代号
氢泵转速	X1	氢泵出口压力	X11
氧泵转速	X2	氢泵出口温度	X12
氢泵流量	X3	冷却套出口压力	X13
氧泵流量	X4	发生器氢喷前压力	X14
氧泵入口压力	X5	发生器氧喷前压力	X15
氧泵出口压力	X6	燃气发生器压力	X16
氧泵出口温度	X7	氢涡轮入口压力	X17
燃烧室氧喷前压力	X8	氧涡轮入口压力	X18
燃烧室压力	X9	氢涡轮出口压力	X19
氢泵入口压力	X10		

(2) 一致性分析。为了提高数据挖掘结果的精度，必须保证数据是干净的、一致的。对于训练集而言，故障类型的定性要尽可能地准确，否

则会影响决策的正确性。测量传感器在极端的物理条件下有可能被损坏, 使测量的数据无效, 因此有时需要剔除严重错误的数据。对于个别变化剧烈的数据, 需分析是传感器故障还是发动机系统发生了故障, 若是传感器故障, 则需要对这一数据进行修正, 采用了滑动平均的方法, 即选取该点附近一段数据 (选取 3 个作为一组) 的平均值替代该点的测量值。

(3) 完整性分析。在训练集数据样本的选取上, 要保证数据选取的全面性。对于同型号发动机的正常或发生了同种类型故障的试车数据, 尽管其参数值可能不相同, 甚至差别较大, 但从整体上来看它们的确属于同一类型, 则在选取训练集时要尽可能地~~将~~不同范围的数据及反映不同故障类型的数据样本都包含进去。训练集中选取了 3 种稳态段故障, 加上正常状态共有 4 种类型, 样本总数为 6629。

3.2 热试车数据挖掘

基于前面的分析, 选取的训练集如表 2 所示。

表 2 训练集所采用的部分数据样本

Tab. 2 A sample subset of training examples for LRE test data

dev_X1	dev_X2	dev_X3	dev_X4	...	Engine state
-0.15120	-0.15481	-0.21310	-0.17774		A
-0.18411	-0.17866	-0.24668	-0.16456		A
-0.15168	0.03676	-0.24052	-0.03017		B
-0.25622	-0.06236	-0.33942	-0.12701	...	B
1.00637	0.09577	-0.16534	0.06927		C
0.98474	-0.09949	-0.61487	0.08681		C
-0.01030	-0.02568	-0.01256	-0.08173		D
0.01424	0.02257	0.00390	-0.04181		D

注: A—氧副文氏管出现多余物; B—氢泵次同步振动; C—燃料涡轮破坏; D—正常

对于上述训练集, 采用基于 C4.5 算法的决策树方法进行了挖掘, 并用 10-折交叉验证^[45]的方法进行了树的剪枝, 得到了图 1 所示的非常直

观的决策规则。在 1GB 内存、Pentium IV 3.0GHz CPU 的微机上共耗时 2.66 秒。

由图 1 可知, 只要利用其中 4 个参数, 即氢泵转速 (X1)、氧泵转速 (X2)、氢泵流量 (X3) 以及氢泵出口压力 (X11) 的相对偏差值, 就可以进行该型发动机的上述 3 类故障检测和诊断。可用 If-Then 语句对上述规则进行描述。如: 若 dev_X3 ≤0.13515 且 dev_X11 >-0.932494 且 dev_X2>-0.099489, 则故障类型为氢涡轮破坏。显然, 训练数据越丰富, 类型越多, 挖掘所得规则的适用范围就越广。稳态段发生故障的 T10-1 试车, T155 试车, T20 试车, 系统红线关机时间分别为 43.21 秒, 275.80 秒, 30.23 秒。上述 4 个参数的相对变化量随时间的变化关系如图 2 所示, 可见, 上述决策规则反映出了发动机的故障特征。

```
dev_X3 <= -0.13515
| dev_X11 <= -0.932494
| | dev_X1 <= 0.047424: 正常(13.0)
| | dev_X1 > 0.047424: 氢泵次同步振(3.0)
| dev_X11 > -0.932494
| | dev_X2 <= -0.099489: 氧副文氏管出现多余物(5.0)
| | dev_X2 > -0.099489: 氢涡轮破坏(4.0)
dev_X3 > -0.13515: 正常(6604.0)
```

图 1 试车数据挖掘得到的决策树
Fig.1 Decision tree generated from LRE test data

根据上述决策规则, 基于热试车稳态段数据对 23 次发动机试车 (包括上述 3 次故障试车数据和 20 次正常试车数据, 共 76005 个样本) 稳态段的状态进行了预测。稳态段故障试车的检测与诊断结果如表 3 所示。

表 3 不同方法对故障试车数据的诊断结果

Tab.3 Diagnosis results of fault test data using different methods

诊断方法	单位: s		
	T10-1	T115	T20
人工神经网络方法 ^[9]	42.64	275.20	4.48
综合决策与报警系统 ^[6]	42.70	275.60	30.20
决策树方法	42.70	275.50	30.00

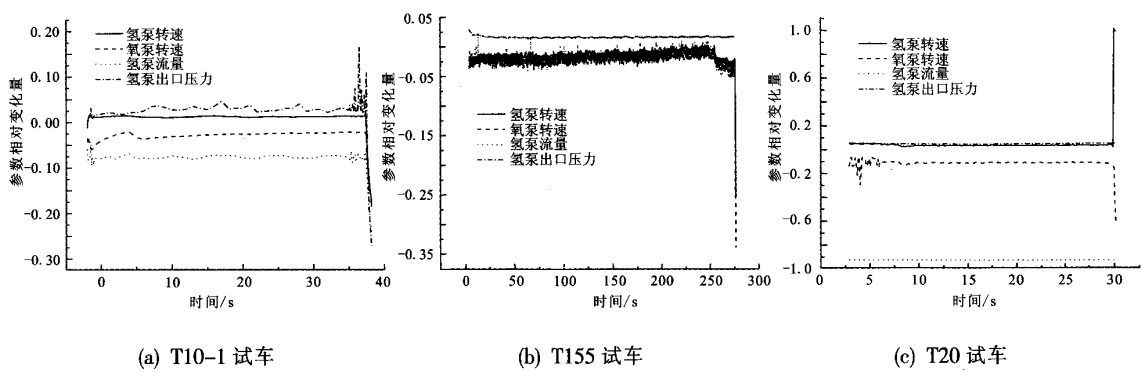


图 2 部分试车参数相对变化量随时间的变化关系图

Fig.2 Charts of parameters' deviation with time of some tests data

对正常试车数据检测结果如表 4 所示, 可见没有出现误报警。

表 4 对正常试车数据的检测和诊断结果

Tab.4 Detection and diagnosis results of normal data

试车代号	诊断结果	试车代号	诊断结果
T145	正常	T1852	正常
T146	正常	T1861	正常
T147	正常	T1911	正常
T1511	正常	T1912	正常
T1512	正常	T1921	正常
T152	正常	T1922	正常
T153	正常	T1931	正常
T183	正常	T1932	正常
T184	正常	T232	正常
T1851	正常	T233	正常

可见, 基于决策树的故障检测和诊断方法对于故障试车没有出现漏报警和误报警, 而对于正常试车没有出现误报警, 比文献 [6] 采用的液体火箭发动机实时故障检测与报警原型系统可更早地预报故障, 和文献 [3] 中采用神经网络方法所得结果也保持一致。因此, 决策树方法所得结果是可信的。

3.3 仿真数据挖掘验证

随着发动机设计与研制水平的不断提高, 故障数据越来越少, 因此, 人为设置和模拟故障将成为考核故障检测与诊断方法性能的重要途径。

利用氧稳压阀出口泄漏, 发生器氧副控阀泄

漏, 发生器氢副控阀泄漏, 氧涡轮入口燃气泄漏及氢涡轮入口燃气泄漏等火箭发动机可能发生的 28 种故障的仿真数据^[7]对决策树方法的性能进行了进一步的验证。选取仿真得到的 70 个参数的相对偏差值, 即 dev_X_j 作为训练集, 共有 1120 个样本, 即训练集的维数为: 1120×70 。采用决策树方法进行挖掘, 得到了如图 3 所示结果 (篇幅关系这里只截取了其中一部分)。图中 $X57$ 表示氧涡轮效率; $X45$ 表示发生器混合比; $X2$ 表示发动机总流量; $X44$ 表示推力室混合比; $X40$ 表示氢泵效率。

利用得到的决策规则对未经训练的大量故障仿真数据进行状态预测, 预测结果正确。因此, 决策树方法可用于液体火箭发动机的故障诊断。

```
dev_X57 <= -0.000714
| dev_X45 <= 0.001606
| | dev_X2 <= 0.002599
| | | dev_X44 <= -0.00405
| | | | dev_X40 <= -0.000017:Fault_5 (31.0)
| | | | dev_X40 > -0.000017
```

图 3 基于仿真数据挖掘得到的决策树

Fig.3 Decision tree generated from simulated data

4 结束语

基于决策树方法的液体火箭发动机故障检测和诊断技术不需要对信号进行特殊处理或建立发动机数学模型。

(下转第 42 页)

专业人员的通力合作。理论方面需掌握等离子体在各个环节中的行为规律,技术上需解决高性能的螺旋波等离子体源、大功率射频功率装置等问题,并提高各环节的效率特别是回旋射频天线耦合效率和射频能量转化为离子能量的效率。

参考文献:

- [1] Chang Díaz F R. The VASIMR Engine[J]. Scientific American, 2000, 283: 72-79.
- [2] Chang Díaz F R. An overview of the VASIMR engine: High power space propulsion with RF plasma generation and heating[C]. Proceedings of Radio Frequency Power in Plasmas: 14th Topical Conference, 2001: 3-15.
- [3] Squire J P, Chang Díaz F R, Jacobson V T, et al. Helicon plasma injector and ion cyclotron acceleration development in the VASIMR experiment[R]. AIAA 2000-3752.
- [4] Rey C M, Hoffman W C, Chang Díaz F R, et al. Design and fabrication of an HTS magnet for the VASIMR experiment[J]. IEEE Transactions on Applied Superconductivity, 2002, 12: 993-996.
- [5] Glover T W, Chang Díaz F R, Squire J P, et al. Principal VASIMR results and present objectives [R]. Space Technology and Applications International Forum, 2005.
- [6] Bering E A III, Squire J P. Progress toward the development of a 50 kW VASIMR engine[R]. AIAA 2005-0369.
- [7] Ilin A V, Chang Díaz F R, Squire J P, et al. Plasma heating simulation in the VASIMR system [R]. AIAA 2005-0949.
- [8] Arefiev A V, Breizman B N. Theoretical components of the VASIMR plasma propulsion concept[J]. Physics of Plasmas, 2004, 11: 2942-2949.
- [9] Squire J P. Acceleration of dense flowing plasmas using ICRF power in the VASIMR experiment [R]. 16th Topical Conference on Radio Frequency Power in Plasmas, 2005.
- [10] Chang Díaz F R, Squire J P, Bentson R D, et al. The physics and engineering of the VASIMR engine [R]. AIAA 2000-3756.
- [11] Sutton G P, Oscar Biblarz. Rocket Propulsion (Seventh Edition) [M]. John Wiley & Sons, Inc, 2001.
- [12] Rauwolf G A, Pelaccio D G, Patel S, et al. Mission performance of emerging in-space propulsion concepts for one-year crewed Mars missions[R]. AIAA 2001-3374.

(编辑:陈红霞)

(上接第30页)

基于大量试车数据从信息熵的角度对发动机的故障特征进行了提取,理论基础可靠。训练集中故障类型越丰富,系统对故障的识别能力越强,这就需要在实践当中不断的积累数据,完善训练集。由于训练集中选取的是参数的相对变化量,因此所得到的决策规则具有一定的普遍性,适用性强。

参考文献:

- [1] 张育林,吴建军. 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.
- [2] 张惠军. 液体火箭发动机故障检测与诊断技术综述[J]. 火箭推进. 2004, 30(5): 40-45.
- [3] 黄敏超. 液体火箭发动机故障的神经网络诊断研究[D]. 长沙:国防科技大学研究生院, 1998.
- [4] 韩家炜. 数据挖掘: 概念与技术[M]. 北京: 机械工业出版社, 2001.
- [5] 邵峰晶, 于忠清. 数据挖掘原理与算法[M]. 北京: 中国水利水电出版社, 2003.
- [6] 刘洪刚. 液体火箭发动机实时故障检测与报警原型系统的设计与实现[J]. 火箭推进. 2005, 31(4): 21-25.
- [7] 刘冰. 液氢液氧火箭发动机稳态数值仿真与故障分析[J]. 国防科技大学学报, 1997, 19(6): 15-19.

(编辑:马杰)