

文章编号: 1001-2486(2007)04-0001-05

结合动态时间弯曲与决策树方法的液体火箭发动机故障诊断*

胡小平, 韩泉东, 李京浩

(国防科技大学 航天与材料工程学院, 湖南 长沙 410073)

摘要: 对某大型液体火箭发动机的热试车数据及通过发动机模型仿真得到的故障数据进行动态时间弯曲分析, 得到弯曲路径集, 然后结合决策树方法进行了故障检测和诊断。对于故障试车没有出现漏报警和误报警, 对于正常试车没有出现误报警。通过与神经网络、支持向量机等方法所得结果的对比, 证明该方法可以成功地应用于火箭发动机的故障检测和诊断。

关键词: 液体火箭发动机; 故障检测和诊断; 数据挖掘; 动态时间弯曲; 决策树

中图分类号: V448.1 **文献标识码:** A

Fault Diagnosis of Liquid Rocket Engine by Dynamic Time Warping Combined with Decision Tree Method

HU Xiao-ping, HAN Quan-dong, LI Jing-hao

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

Abstract: Through dynamic time warping analysis to the hot-fire test data and simulated fault data of a certain liquid rocket engine, the warped path sets were obtained. Combined with the decision tree method, fault detection and diagnosis were carried out. Results show that there were no failing report and no misstatement for fault tests, and no misstatement for normal tests. Compared with those results by ANN and SVM methods, the successful application of the dynamic time warping method combined with decision tree was proved.

Key words: liquid rocket engine; fault detection and diagnosis; data mining; dynamic time warping; decision tree

随着运载火箭的有效载荷不断增加, 发动机的工作参数值不断提高, 以致其工作环境越来越恶劣: 高温、高压、高速流动、大振幅及宽频谱的振动激励等。液体推进系统的可靠性和安全性的要求越来越高, 使得液体火箭发动机健康监控系统的研究在航天领域越来越受到重视^[1]。

数据挖掘是“从大量的数据中提取隐含在其中的人们事先未知的、但潜在有用的信息和知识的过程”, 提取的知识常表示为概念、规则、规律和模式等形式。动态时间弯曲(Dynamic Time Warping, DTW)和决策树(Decision Tree, DT)是数据挖掘中两种重要的方法。动态时间弯曲是一种相似性搜索算法, 通过测度两组时间序列之间的相似性参数, 得到一组动态时间弯曲路径集。在不同的工作状态下, 动态时间弯曲路径集具有不同的特征。然而这些特征之间的深层次联系通常难以把握, 故作者把动态时间弯曲与决策树方法结合在一起。决策树方法是一种数据特征提取判断的方法, 通过提取数据主要特征进行数次两-两分类, 从而形成了故障诊断规则, 达到故障检测和诊断的目的。

1 方法描述

1.1 动态时间弯曲方法

对时间序列数据动态时间弯曲的目的是根据给定的数据样本来求对具体样本的相似程度的估计, 使它能够对未知输出做出尽可能准确的预测。

设有两个时间序列 C 和 Q , 其数据长度分别为 m 和 n , 且:

* 收稿日期: 2007-01-28

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。

$$C = c_1, c_2, \dots, c_m, Q = q_1, q_2, \dots, q_n \quad (1)$$

为了利用动态时间弯曲将两个时间序列对准, 首先定义一个距离相异矩阵。

定义1 设有一个 m 行 n 列矩阵 D :

$$D = \begin{pmatrix} d(q_1, c_m) & \dots & d(q_n, c_m) \\ \vdots & \ddots & \vdots \\ d(q_1, c_1) & \dots & d(q_n, c_1) \end{pmatrix} \quad (2)$$

矩阵中的元素 $d(q_i, c_j)$ 是不同时间序列数据对象之间的点的欧几里德距离值 $d(q_i, c_j) = (q_i - c_j)^2$, 可以看作是对象 q 和对象 c 之间的相异性的量化表示。当对象 q 和 c 越相似或越接近, 其值越接近 0; 两个对象越不相同, 其值越大。将两个时间序列分别置于二维坐标的两轴, 就可以定义弯曲路径。

定义2 在两个不同时间序列间的距离矩阵中, 定义时间序列间相异性关系的一组连续的矩阵元素的集合 $W = w_1, w_2, \dots, w_k, \dots, w_k$ 为弯曲路径。弯曲路径应满足有界性、连续性、单调性和边界条件^[2]。

基于欧几里德距离测度与基于动态时间弯曲的距离测度进行时间序列分析时, 数据对准的表现方式是不相同的, 见图 1。

动态时间弯曲的目标在于使弯曲路径总长度最小(对应数据相似性程度最大), 即

$$DTW(Q, C) = \min \left[\frac{1}{K} \sum_{i=1}^K w_i \right] \quad (3)$$

在实际应用中还需要限制弯曲路径的行走区域。为了矩阵处理方便, 应该把弯曲路径的摆动范围限制在距离矩阵的斜对角线两侧附近, 因此, 在不考虑区域外的距离相似因子的情况下称这种弯曲路径的子集为弯曲窗口。弯曲窗口的设置主要有两个原因: (1) 可以加快动态时间弯曲算法的速度; (2) 防止逻辑上的病态弯曲路径的出现。在时间序列分析中, 时间因素是非常重要的, 时间偏移量过大是没有实际意义的, 从逻辑上来看应该予以忽略。

设点 (i, j) 在最佳路径上, 那么从点 $(1, 1)$ 到 (i, j) 的子路径也是局部最优解, 也就是说从点 $(1, 1)$ 到点 (n, m) 的最佳路径可以由时间起始点 $(1, 1)$ 到终点 (n, m) 之间的局部最优解通过递归搜索获得。即:

$$S(1, 1) = d(q_1, c_1), S = d(q_i, c_j) + \min\{S(i-1, j), S(i-1, j-1), S(i, j-1)\} \quad (4)$$

最终时间序列弯曲路径最小累加值为 $S(n, m)$ 。从 $S(n, m)$ 起沿弯曲路径按最小累加值倒退, 直到起始点 $S(1, 1)$ 即可找到整个弯曲路径。

1.2 决策树方法

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, 并根据内部节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^[3-6]。

根据火箭发动机试车数据是连续属性且训练集目前不是特别大的特点, 本文选用了 C4.5 算法。随着以后数据集的增大, 可考虑采用能处理较大数据集的 SLIQ 等算法。

决策树方法关键是决策树的构建, 主要分为两步: 树的建立和树的剪枝。

树的建立 通常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。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^[3]。它既

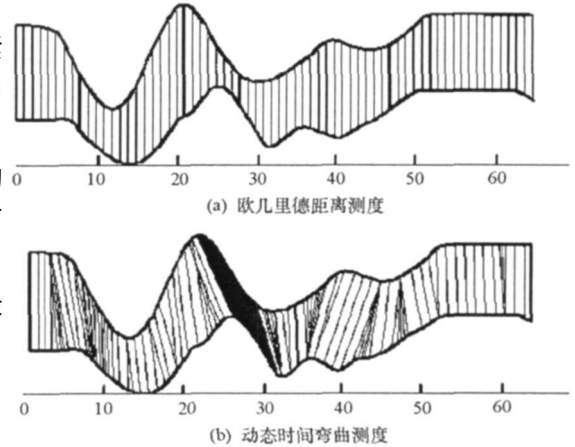


图1 两种测度的对比示意图

Fig. 1 Comparison of two kinds of measurement methods

可以处理离散属性的值, 又可以处理连续属性的值, 很适合发动机试车数据的挖掘。

树的剪枝 在决策树构造时, 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这种过分适应数据问题, 以提高在未知数据上分类的准确度。树的修剪就是使用一个叶子节点来代替一个子树。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验证两种策略。交叉验证是一种模型评估方法, 它是指在训练过程开始之前将一部分数据予以保留, 在训练之后利用这部分数据对学习的结果进行验证。其中 K -折交叉验证(K -Fold Cross Validation) 可对监督学习算法的结果进行评估。它将数据集分成 K 个子集, 将其中的一个用作测试集, 而其余的 $K-1$ 个数据集作为训练集, 最后对 K 个数据子集的错误数计算平均值。数据集的划分一般采取等均分或随机划分。

2 数据挖掘实例

2.1 数据预处理

数据预处理是数据挖掘中的一个重要环节。本文主要进行了以下预处理工作:

(1) 数据集成。发动机试车测得的参数种类很多, 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, 开发了相应的数据提取算法, 可以从多个数据文件中快速提取出所需的数据样本。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。

(2) 数据简化。实际测量得到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, 其中的噪声干扰还可能造成无效的归纳, 故有必要缩减数据维数。一般采用属性选择和数据抽样技术^[3]。本文采用属性选择的方法(主要选取与涡轮泵及流量调节器相关的易发生故障组件对应的参数), 将上述 30 个参数进一步压缩到 21 个, 用 $X_1 \sim X_{21}$ 表示, 如表 1 所示。

表 1 训练集所选参数

Tab. 1 Selected parameters for training sets

参数名称	符号	参数名称	符号	参数名称	符号
氢泵转速	X_1	燃烧室氧喷前压力	X_8	发生器氧喷前压力	X_{15}
氧泵转速	X_2	燃烧室压力	X_9	燃气发生器压力	X_{16}
氢泵流量	X_3	氢泵入口压力	X_{10}	氢涡轮入口压力	X_{17}
氧泵流量	X_4	氢泵出口压力	X_{11}	氧涡轮入口压力	X_{18}
氧泵入口压力	X_5	氢泵出口温度	X_{12}	氢涡轮出口压力	X_{19}
氧泵出口压力	X_6	冷却套出口压力	X_{13}	氧涡轮出口温度	X_{20}
氧泵出口温度	X_7	发生器氢喷前压力	X_{14}	氧贮箱压力	X_{21}

(3) 完整性分析。在数据样本的选取上要保证全面性。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到训练集中。

2.2 结合动态时间弯曲的决策树方法的数据挖掘结果

试车实测数据变化范围可能较大, 为了减小阈值的搜索范围以提高数据挖掘的效率, 本文采用了 $(X_{ij} - X_j)/X_j$ (简记为 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(对应某一个时间点)中第 j 个参数的测量值, 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。在保证训练集完整性的前提下, 选取 3 组稳态段发生了故障的数据以及正常试车的部分数据(共 6629 个样本)作为训练集。

对于所选训练集, 采用动态时间弯曲方法首先得到其弯曲路径, 在此基础上采用 C4.5 决策树算法进行了挖掘, 得到图 2 所示的决策规则。

```

13- TER <= 0.18024
| 11- POR <= 0.30607
| | 17- PFY <= 0.000165
| | | 1- NWR <= 0.000704: 正常 (226.0)
| | | 1- NWR > 0.000704
| | | | 2- NWY <= 0.006954: 正常 (16.0)
| | | | 2- NWY > 0.006954: 氢泵出现故障 (4.0)
| | 17- PFY > 0.000165: 正常 (1234.0)
| 11- POR > 0.03607
| | 4- GY <= 0.000934: 氧副文氏管出现多余物 (4.0)
| | 4- GY > 0.000934: 正常 (22.0)
13- TER > 0.18024
| 2- NWY <= 0.019692: 氢泵同步振动 (6.0)
| 2- NWY > 0.019692: 正常 (19.0)

```

图2 针对试车数据挖掘得到的决策树

Fig. 2 Decision tree mined for test data

可见,只要利用其中6个参数,即氢泵转速(X_1)、氧泵转速(X_2)、氧泵流量(X_4)、氢泵出口压力(X_{11})、氢泵出口温度(X_{13})、发生器氧喷前压力(X_{17}),就可以进行该型发动机的上述3类故障的检测和诊断。上述规则可以用 if-then 语句进行直观的描述,如:“if $TER > 0.18024$ and $NWY > 0.019692$ then 正常”,上述数值为动态时间弯曲路径上的值。显然,训练集所选的样本种类越多,挖掘所得规则的适用范围就越广。

根据上述决策规则,对该发动机23次热试车(即3次故障试车数据和20次正常试车数据,共76000个样本)稳态段的状态进行了预测。对故障试车数据预测的结果如表2所示。实际试车中,T10-1试车、T155试车和T20试车系统红线关机时间分别为43.21s、275.8s和30.23s,故障类型分别为氧副文氏管出现多余物、氢涡轮破坏以及氢泵次同步振动。对正常试车数据的预测结果如表3所示。

表2 对故障试车数据的检测和诊断结果

Tab. 2 FDD results for fault test data

诊断方法	试车代号		
	T10-1	T155	T20
结合动态时间弯曲的决策树方法	42.7s 预报故障	275.5s 预报故障	29.6s 预报故障
综合决策与报警系统 ^[4]	42.7s 预报故障	275.6s 预报故障	30.2s 预报故障
神经网络方法 ^[5]	42.64s 预报故障	275.2s 预报故障	4.48s 预报故障
支持向量机 ^[6]	42.6s 预报故障	275.3s 预报故障	30.0s 预报故障

表3 对正常试车数据的检测和诊断结果

Tab. 3 FDD results for normal test data

Test No.	Result	Test No.	Result	Test No.	Result
T145	Normal	T183	Normal	T1921	Normal
T146	Normal	T184	Normal	T1922	Normal
T147	Normal	T1851	Normal	T1931	Normal
T1511	Normal	T1852	Normal	T1932	Normal
T1512	Normal	T1861	Normal	T232	Normal
T152	Normal	T1911	Normal	T233	Normal
T153	Normal	T1912	Normal		

可见,结合动态时间弯曲及决策树方法的故障检测和诊断方法对于故障试车没有出现任何漏报警

和误报警, 对于正常试车也没有出现误报警, 比文献[4]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, 与文献[5]中采用神经网络方法以及文献[6]中采用的支持向量机方法所得结果也保持一致。因此, 故障诊断结果是可信的。

2.3 利用仿真数据的挖掘验证

随着火箭发动机设计和研制水平的不断提高, 有故障的试车越来越少, 故障数据也将越来越难以获得。因此, 应将人为设置和模拟的故障作为考核故障检测与诊断系统的检测准确性的一个重要手段。

通过建立液体火箭发动机的数学模型, 仿真了发动机可能发生的氧稳压阀出口泄露、发生器氧副控阀泄露、发生器氢副控阀泄露、氧涡轮入口燃气泄露、氢涡轮入口燃气泄露、氢涡轮出口燃气泄露、氧泵后泄露、氧主文氏管后泄露、氢泵后泄露、氢主文氏管后泄露等 28 种故障。选取仿真所得参数的相对偏差值作为训练集, 每种故障选取 40 个样本, 则共有 1120 个样本(每个样本包含 69 个参数)。采用结合动态时间弯曲的决策树方法进行挖掘, 得到了图 3 所示结果(限于篇幅, 只截取了前面一部分)。

```

X53 <= 0.000905: fault0 (14.0)
X53 > 0.000905
|   X40 <= 0.000011: fault1 (10.0)
|   X40 > 0.000011
|   |   X0 <= 3457059465: fault2 (20.0)
|   |   X0 > 3457059465
|   |   |   X69 <= 15211010266
|   |   |   |   X69 <= 6453162889: fault3 (8.0)
|   |   |   |   X69 > 6453162889
|   |   |   |   |   X0 <= 3930706726: fault2 (4.0)
|   |   |   |   |   X0 > 3930706726

```

图 3 针对仿真数据挖掘得到的决策树

Fig. 3 Decision tree mined for simulated fault data

利用得到的决策规则对未经训练的大量仿真数据样本进行了状态预测, 预测结果全部正确。这一验证结果表明, 结合动态时间弯曲的决策树方法可成功地应用于火箭发动机的故障诊断。

3 结束语

从数据挖掘的角度, 将结合动态时间弯曲的决策树方法应用于液体火箭发动机的故障检测和诊断中, 对试车数据和仿真故障数据进行了挖掘。结果发现, 该方法能够准确、及时地检测和诊断发动机工作过程中发生的故障, 对于理论分析和工程应用都具有重要的意义。本文方法对其它复杂流体机械系统的故障诊断也具有一定的参考价值。

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] 翁颖钧, 朱仲英. 基于动态时间弯曲的时序数据聚类算法的研究[J]. 计算机仿真, 2004, 21(3).
- [3] 韩家伟, 等. 数据挖掘: 概念与技术[M]. 北京: 机械工业出版社, 2001.
- [4] 刘洪刚, 等. 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).
- [5] 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, 1998.
- [6] 韩泉东, 胡小平, 等. 决策树和支持向量机方法在液体火箭发动机故障诊断中的应用[C]//2006年全国理论计算机科学学术年会(TCS2006)论文集, 2006.