

利用关联规则检测液体火箭发动机 启动关机过程的故障

¹王艳梅, ²胡小平, ³李舟军

(¹国防科技大学计算机学院, 湖南 长沙 410073; ²国防科技大学航天与材料工程学院,
湖南 长沙 410073; ³北京航空航天大学计算机学院, 北京 100083)

摘 要: 提出了一种基于关联规则的故障检测方法, 根据发生故障时相应参数之间的关联关系是否有效来判断是否故障。对液体火箭发动机的启动过程试车数据进行挖掘的应用实例表明, 该方法能够有效地挖掘参数之间的关联关系, 有效地检测出液体火箭发动机启动过程中的故障。

关键词: 液体火箭发动机; 启动-关机过程; 故障检测; 关联规则; 数据挖掘

中图分类号: V434

文献标识码: A

文章编号: (2006)01-0019-05

Application of association rules to the fault detection of starting-up and shutdown process of liquid rocket engine

¹Wang Yanmei, ²Hu Xiaoping, ³Li Zhoujun

(¹College of Computer Engineering, National Univ. of Defense Technology, Changsha, 410073

²College of Aerospace and Material Engineering, National Univ. of Defense Technology, Changsha, 410073

³College of Computer Engineering, Beijing Univ. of Aeronautics and Astronautics, Beijing 100083, China)

Abstract: A fault detection method based on association rules algorithm is presented. By applying the method to the test data of liquid rocket engine in the starting-up and shutdown process, the association relationships among the parameters of liquid rocket engine were explored, the fault has been detected effectively.

Key words: liquid rocket engine; starting-up and shutdown process; fault detection; association rules; data mining

收稿日期: 2005-09-30; 修回日期: 2005-12-29。基金项目: 国家自然科学基金资助项目(50376073)。

作者简介: 王艳梅 (1977—), 女, 硕士研究生, 研究领域为数据挖掘。Email: tainjing1290@tom.com

1 引言

液体火箭发动机启动与关机过程的本质是非线性时变随机过程, 对其建立较为精确且能够满足在线实时故障检测的解析模型是相当困难的。模型的误差、复杂的计算加之随机噪声的影响常常使得基于模型的检测方法失效。目前, 在我国推进技术领域, 针对启动、关机过程等瞬变过程所进行的研究, 大多数是对其正常动态瞬变特性的仿真研究^[1]。这些研究除了受上述因素影响而导致其有效性的可信度不高之外, 还受到计算方法复杂性(如稳定性、收敛性及计算精度等)的强烈影响。因此在目前的情况下, 对启动关机过程采用基于解析模型的故障检测方法是很困难的^[1]。

国外, 联合技术研究中心(United Technologies Research Center, UTRC)的研究人员在研究航天飞机主发动机(Space Shuttle Main Engine, SSME)的启动和关机过程的故障检测问题时, 采用了双变量二次型非线性回归分析的方法, 提出并实现了非线性回归(Recursive Structural Identification, RESID)算法^[1]。但是, RESID算法目前在如何合理确定阈值上仍存在着困难。

黄敏超等把神经网络技术应用在液体火箭发动机启动阶段的故障诊断中^[2,3]。BP神经网络基于非线性辨识技术, 应用输入输出关系进行故障检测, 它的可靠性高一些, 但计算代价相对较高; 模糊超球神经网络基于模式识别技术, 直接应用匹配程度进行故障检测, 它的敏感性强一些, 但误报率相对较大^[3]。

本文将关联规则技术应用于液体火箭发动机的试车数据挖掘中。利用关联规则技术, 可有效地发现液体火箭发动机中各变量间的关联规则, 通过判断这些规则是否有效, 就能检测液体火箭发动机的故障^[4]。而且, 关联规则技术比神经网络技术计算代价低, 对数据的空缺值不敏感, 具有较好的鲁棒性^[5]。数据挖掘是从数据中获取尚未被人发现的知识, 是“唯数据的”、客观的。

2 关联规则技术的原理及其方法

关联规则挖掘的目的是发现大量数据中项集

之间有趣的关联关系^[6]。

2.1 基本概念

设 $I = \{i_1, i_2, \dots, i_m\}$ 为 m 个不同文字的集合, 其中的文字称为项(Item), 或者商品。称任何 $X \subseteq I$ 为一个项集(如果 $|X| = k$, 则称 X 为 k 项集)。记 $D = \{T_1, T_2, \dots, T_n\}$, 其中 $T_i \subseteq I$ 称为一个交易或事务(Transaction), 称 D 为 I 上的交易集或者数据集(Dataset), 简称交易集或者数据集。给定 I , 关联规则就是形如 $X \Rightarrow Y$ 的一个表达式, 其中, $X \subseteq I, Y \subseteq I, X \cap Y = \phi$ 。

衡量关联规则是否有意义的标准有两个: 一个是支持度(Support), 另一个是置信度(Confidence)。支持度用于衡量关联规则在整个数据集中的统计重要性, 置信度用于衡量关联规则的可信程度。一般来说, 只有支持度和置信度均有较高的关联规则对用户而言才是感兴趣和有用的关联规则。所以在挖掘时, 需要用户适当定义最小支持度和最小置信度的阈值。

2.2 Apriori 算法^[6]

Apriori 算法是一种最经典的挖掘关联规则频繁项集的算法。

Apriori 算法的伪代码如下:

```

 $L_1 = \{\text{频繁 1-项集}\};$ 
for( $k=2; L_{k-1} \neq \phi; k++$ ) do begin
 $C_k = \text{apriori-gen}(L_{k-1});$  //产生新的候选项集
for all transactions  $T \in D$  do begin
 $C_T = \text{subset}(C_k, T);$  //事务  $T$  中包含的候选项集
for all candidates  $c \in C_T$  do
 $c.\text{count}++;$ 
end
 $L_k = \{c \in C_k \mid c.\text{count} \geq \text{min\_sup}\}$ 
End
Answer =  $\bigcup_k L_k$ ;

```

L_k 为长度为 k 的频繁项集集合, C_k 为长度为 k 的候选项集集合。

3 基于关联规则的液体火箭发动机试车数据挖掘

从关联规则的描述可以看出它挖掘的直接作用对象是数据库中每一次交易中的项目(商品)。

所以,为了能从数据库中挖掘出有价值的关联规则,首先应该把液体火箭发动机的数据信息转换成这种形式。影响液体火箭发动机故障的因素很多,我们把所选参数作为项集,把每个时间点对应的各个参数随时间的变化作为一个事务,就可以利用关联规则技术,推导出在正常状态下液体火箭发动机各个参数之间的关联关系,这些关系可以看作描述液体火箭发动机正常工作状态的模型。如果在试车过程中发生了故障,那么,相应参数之间的关联关系必然会发生变化。如果原有的关联关系不再成立,我们就有理由推断它发生了故障;如果原有的关联关系全都成立,说明它是正常试车。由此可以检测到试车是否发生故障。

3.1 数据选择和属性编码

数据挖掘通常并不需要使用所拥有的所有数据,有些数据对象和数据属性对建立模型获得模式是没有影响的,这些数据的加入会大大影响挖掘效率,甚至还可能导致挖掘结果的偏差。因此,选择有效的数据是很有必要的。数据选择包括属性选择和数据抽样。

对于属性选择,结合发动机的工作特点,选择试车数据中的 27 个参数进行挖掘。首先对各个属性进行编码,如表 1 所示。

表 1 属性编码

Tab.1 Attribution code

编码	属性	编码	属性
A	氧贮箱压力	O	发生器氧喷前压力
B	氧泵前阀入口压力	P	燃气发生器压力
C	氧泵入口压力	Q	氢涡轮入口压力
D	氧泵出口压力	R	氧涡轮入口压力
E	燃烧室氧喷前压力	S	氢涡轮出口压力
F	氧泵出口温度	T	中减压器出口压力
G	氢贮箱压力	U	氧涡轮氢隔离腔压力
H	氢泵前阀入口压力	V	氢涡轮入口温度
I	冷却套出口压力	W	氧涡轮入口温度
J	氢气泄漏压力	X	氢涡轮泵转速
K	氢泵前阀入口温度	Y	氧涡轮转速
L	氢泵出口温度	Z	氢泵流量
M	氢冷却套入口温度	a	氧泵流量
N	发生器氧喷前压力		

对于数据选择,由于液体火箭发动机启动过程的影响因素极多,相应的正常试车的数据的离散度也很大,选取某种型号的液体火箭发动机启动阶段正常试车的数据 T146、T147、T183、T184、T1851 作为训练样本。由于试车数据各个参数的测量时间间隔不同,有些数据的测量时间间隔为 0.02 秒,而有些数据的测量时间间隔为 0.1 秒,所以选取最大的时间间隔 0.1 秒为选择数据的步长。当然,随着采样频率的提高,选择数据的步长可以越来越小,该方法检测故障的及时性也可以得到更好的改善。

3.2 空缺值处理

现实世界的数据也不总是完整的,有时可能在一些记录的属性上没有记录值。对于液体火箭发动机,也有空缺值现象出现。主要是因为测量数据的传感器故障,或者测量人员当时感觉那个数据不重要,没有进行测量或者没有保存,当然,也可能是因为存储数据的介质损坏造成的。

针对这些空缺值,进行如下处理:对于训练数据,我们尽量选择所有数据都有的属性来进行挖掘,因为有些属性是很重要的,对液体火箭发动机的工作状态起决定性的作用;对于要测试和检测的数据,对空缺值的处理方法,可以使用一个全局变量填充空缺值。

如果它是利用关联规则方法经过训练数据挖掘出来的规则当中包含的属性,那么该属性是对液体火箭发动机的工作状态起决定性作用的属性,此方法会显示不能判断。否则,如果它是挖掘出来的规则当中不包含的属性,那么它是不重要的属性,此方法就会进行正确的判断。

3.3 数据离散化

Apriori 算法可以用于挖掘布尔关联规则^[7],此时要求数据集中的属性都是布尔值。如果某些属性的值域为连续值(如浮点数表达),则在处理前需要进行离散化处理。液体火箭发动机的试车数据(经过归一化处理后的数据)如表 2 所示,为了表示数据的变化趋势,对数据进行了如下处理:当后一时刻的值大于等于前一时刻的值时,取值为 1;否则,取值为-1。其中,1 表示数据增加或不变;-1 表示数据减小。

表 2 发动机部分试车数据

Tab.2 Engine part test data

A	B	C	D	E
0.3668	0.4849	0.5159	0.5249	0.0998
0.368	0.4591	0.5094	0.6566	0.1001
0.3587	0.4763	0.4846	1.0727	0.0998
0.3551	0.516	0.4848	1.5026	0.0998
0.3621	0.516	0.5055	1.8389	0.1001
0.3695	0.4667	0.5225	2.068	0.0998
0.3708	0.4715	0.5224	2.2536	0.0998
0.3706	0.5002	0.5261	2.4702	0.0998
0.3681	0.5138	0.5394	2.5713	0.0998
0.3655	0.4897	0.5569	2.6448	0.1001
0.3624	0.492	0.56	2.6854	0.1001
0.358	0.4964	0.5671	2.7159	0.0994
0.3624	0.4992	0.5722	2.7125	0.0998
0.365	0.2801	0.5765	2.3891	0.0998
0.367	0.1966	0.5545	2.817	0.1046
0.3401	0.2634	0.5162	3.3862	0.1242

3.4 关联规则挖掘

由于关联规则是在给定支持度和置信度下挖掘出来的，不同的支持度和置信度对挖掘结果会有一定的影响。规则的置信度反映所发现规则的确定性，现在需要用发现的规则检测液体火箭发动机的工作情况，需要 100% 的确定性，所以最小置信度选择为 1。取各种大小不同的支持度阈值，对这些数据进行多次挖掘，通过分析比较，最后，把最小支持度选为 0.85，得到参数 D、G、X、Y 之间数据的变化关系，具体规则如下：

- (1) $X=1 \quad G=-1$
- (2) $Y=1 \quad G=-1$
- (3) $X=1,Y=1 \quad G=-1$
- (4) $D=1 \quad G=1$
- (5) $D=1,X=1 \quad G=-1$
- (6) $D=1,Y=1 \quad G=-1$
- (7) $D=1,Y=1 \quad X=1$
- (8) $D=1,X=1,Y=1 \quad G=-1$
- (9) $D=1,G=-1,Y=1 \quad X=1$
- (10) $D=1,Y=1 \quad G=-1,X=1$

3.5 故障测试

由于正常试车数据挖掘时，选择的置信度为

1，也就是说，如果条件成立，那么结论一定成立，所以可以用上面的规则进行故障检测，如果规则成立，说明试车是正常的，如果规则有 1 次以上不成立，说明发生了故障。代号为 T1852、T1861、T06、T1862 和 T21 的启动阶段试车情况测试的结果如表 3 所示。

表 3 启动段试车测试结果

Tab.3 Test results in starting-up process

试车代号	实际情况	检测结果	测试结果
T1852	正常	正常	正确
T1861	正常	正常	正确
T06	20 秒 红线关机	0.1 秒 预报故障	正确
T1862	1.32 秒 红线关机	0.2 秒 预报故障	正确
T21	1.32 秒 红线关机	0.2 秒 预报故障	正确

T06 试车的实际情况为启动段故障，二次启动工作参数低，发生了故障。T21 试车的实际情况为未达到启动条件（泵后出现两相流使涡轮泵转速过高）火药启动器爆炸，发生了故障。由于他们都是因启动的条件没有得到满足而发生的故障，因此，本文得到的检测结果 0.1 秒预报故障是合理的，它比红线系统分别提前了 19.9 秒和 1.12 秒关机。T1862 试车的实际情况为：在启动过程中，火药启动器爆炸，发生了故障。本文得到的结果为 0.2 秒预报故障，比红线系统提前 1.12 秒预报故障。预报故障的及时性比红线系统得到了较大的提高。

4 讨论

在参考文献[2]、[3]中，黄敏超用 BP 神经网络、模糊超球神经网络进行了液体火箭发动机的故障检测，关联规则检测结果和神经网络检测结果的比较如表 4 所示。

从表 4 可以看出，关联规则和神经网络的检测结果是一致的，但是，神经网络方法检测的时候需要测量 12 个参数，如果由于某种原因，其中的一个或几个参数无法测量，那么神经网络方法

就失效了。而利用关联规则方法在训练数据中，挖掘出了 4 个变量之间的关系，仅需要考虑这 4 个变量之间的关系即可，减少了测量参数的个数。而且只要其中的一个关系不满足，就可以说明这次试车是有故障的，因此，与神经网络相比，对数据的空缺值不敏感，具有较好的鲁棒性；另外，在故障检测时，由于关联规则只需要考察四个参数之间的关联关系，因此，计算代价比较低。

表 4 神经网络和关联规则检测结果比较

Tab.4 Comparison of detection results between neural network and association rules

试车代号	神经网络	关联规则
T18-52	正常	正常
T18-61	正常	正常
T06	未作检测	0.1 秒预报故障
T18-62	0.3 秒预报故障	0.2 秒预报故障
T21	未作检测	0.2 秒预报故障

当然，关联规则本身也有不足。由于关联规则是在一定的支持度和置信度的前提下进行的数据挖掘，因而不同的支持度和置信度对挖掘结果会有一定的影响。如果支持度过低，挖掘过程中将产生大量的规则，而且，产生的规则不全是感兴趣的；如果支持度过高，挖掘过程中将产生的规则太少，甚至不能产生任何规则，可能无法判断有些数据是否为故障数据。这就需要选取各种大小不同的支持度阈值，对数据进行反复的挖掘计算。

5 结束语

在将关联规则技术应用于液体火箭发动机的

试车数据挖掘中，发现了液体火箭发动机中各变量间的关联规则，通过判断这些规则是否成立，能检测到液体火箭发动机启动阶段的故障。而且检测结果比红线系统更及时，在相同的检测结果的情况下，比神经网络需要的参数更少。

本文所给出的实例是液体火箭发动机启动阶段的故障检测，但方法本身原则上也适用于关机阶段。

参考文献:

[1] 张育林等, 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] 黄敏超等, 神经网络在液体火箭发动机故障检测中的应用(I)非线性辨识技术[J]. 推进技术, 1999,19(2).

[3] 黄敏超等, 神经网络在液体火箭发动机故障检测中的应用()模式识别技术[J]. 推进技术, 1999,20(2).

[4] T Yairi, Y Kato, K Hori. Fault detection by mining association rules from house-keeping data [R]. In Proc. of International Symposium on Artificial Intelligence, Robotics and Automation in Space, 2001.

[5] Hu Hong, Li Jiuyong. Using association rules to make rule-based classifiers robust[R]. At the 16th Australasian Database Conference, Conferences in Research and Practice in Information Technology, University of Newcastle, Newcastle, Australia. Vol. 39, 2005.

[6] 韩家炜等. 数据挖掘: 概念与技术[M]. 北京: 机械工业出版社, 2001.

[7] 张云涛等. 数据挖掘原理与技术[M]. 北京: 电子工业出版社, 2004.

(编辑: 侯 早)