

# 基于相似匹配的液体火箭发动机 故障模式挖掘

袁晓峰<sup>1</sup>, 许化龙<sup>1</sup>, 陈淑红<sup>2</sup>, 杨卫军<sup>1</sup>

(1 第二炮兵工程学院 西安 710025; 2 第二炮兵装备研究院第三研究所 北京 100085)

**摘 要:** 将时间序列相似性匹配方法引入到液体火箭发动机故障模式挖掘中。针对发动机试车数据的特点, 提出了一种基于序变换的时间序列相似匹配算法。该算法具有对时间序列幅值和持续时间不敏感、抗噪声能力强等优点。对某型液体火箭发动机故障数据的相似匹配实验表明: 该算法能够为液体火箭发动机的故障检测和诊断提供较好的技术支持。

**关键词:** 时间序列; 序模式; 相似性搜索; 发动机

中图分类号: V434

文献标识码: A

文章编号: (2008) 04-0048-07

## Similarity-search based mining for fault mode of liquid rocket engine

Yuan Xiaofeng<sup>1</sup>, Xu Hualong<sup>1</sup>, Chen Shuhong<sup>2</sup>, Yang WeiJun<sup>1</sup>

(1 Second Artillery Engineering Institute, Xi'an 710025, China;

2 NO.3 Institute of the Second Artillery Armament Academy, Beijing 100085, China)

**Abstract:** The method of time series similarity matching (TSSM) is introduced in the fault mode mining of the liquid rocket engine (LRE). According to the feature of fire-test data, an ordinal transformation based algorithm is proposed, which is insensitive to amplitude, time span and noises of the time series. The TSSM experiments on the data of some LRE indicated that it is quite supportive to fault detection and diagnosis.

**Key words:** time series; ordinal pattern; similarity search; engine

收稿日期: 2008-03-10; 修回日期: 2008-06-13。

作者简介: 袁晓峰 (1975—), 男, 博士生, 工程师, 研究领域为数据挖掘与故障诊断。

## 1 引言

液体火箭发动机的工作状态参数可以有几十甚至上百个,诸如流量、压力、推力、温度、振动、转速、应力等等<sup>[1]</sup>。随着计算机技术的发展和数据采集设备性能的提高,液体发动机的每次试车获得的数据量激增,目前已达到数十兆甚至更多。然而伴随着试车数据不断累积,试车数据库不断增大,这也带来了一些问题。

液体火箭发动机结构和工作过程复杂,很难建立精确的数学模型对发生的故障进行有效描述,且模型中涉及大量非线性方程,求解十分困难。这导致基于模型的诊断方法在实际应用中出现实时性差、故障分离困难、稳定性差等问题。因此,当前基于数学模型的应用研究进展较缓慢,很难在短期内获得突破<sup>[2]</sup>。Tennessee大学的研究人员提出将发动机低频数据分析法用于研究和提取故障信号的时域特征,并将其与存放在数据库中的故障模式进行对比,以检测故障<sup>[3]</sup>。该方法在实际的故障分析和诊断中很具有代表性。然而,这种方法必须将给定的故障模式与数据库中大量的历史数据进行相似匹配和对比分析,这样才能全方位地挖掘故障知识。然而该过程很难通过手工方式实现。因此,必须寻找合适的挖掘算法来解决此问题。

如何挖掘这些宝贵发动机试车数据资源的潜力已成为一个亟待解决的现实问题。对此问题诸多业内专家已逐渐形成共识,比如,2002年北京航空航天大学张振鹏教授曾指出“把我们四十年来液体火箭研制中能够发生的故障情况、故障数据和典型正常数据管理起来,把宝贵的专家经验保存起来,把故障仿真所得的数据贮存起来”<sup>[4]</sup>。文献[4]阐述了液体火箭发动机故障检测和诊断中可能采用的数据挖掘策略。这表明业内专家已认识到这些宝贵数据得不到有效利用的不良后果,也表明该问题已引起业界专家的高度关注。近年来液体火箭发动机试验数据管理系统的日趋完善,为该问题的解决提供了一个良好平台。液体火箭发动机的试车数据从数据形式上讲为时间序

列。近年来出现的面向相似性匹配的时间序挖掘方法有Agrawal提出的DFT法,Chan等人提出的DWT法,Korn等人提出的SVD法,Keogh等人提出的PAA法与APCA法,Peng等人提出的LM模型等。这些方法各有所长,针对不同的应用背景,仍有许多问题有待进一步研究。

本文从时间序列相似匹配角度,研究了大规模数据库中适合发动机故障模式匹配搜索的算法,并进行了实验分析。

## 2 时间序列相似匹配及其应用概述

时间序列相似匹配(亦称相似匹配搜索)既是一类独立的挖掘技术,又是整个时间序列数据挖掘的基础,所有的序列挖掘工作,比如分类、聚类,都是在匹配搜索的基础上开展进一步的研究工作。

### 2.1 时间序列相似匹配问题描述

时间序列相似匹配的基础性工作最早是由Agrawal等人提出的<sup>[5]</sup>。该问题可概括为对给定的某个时间序列,要求从一个大型时间序列数据库中找出与之最相似的序列。

根据查询序列与数据库中序列的长度大小关系,时间序列相似匹配按照匹配类型可以分为完全匹配和子序列匹配;依据用户的查询要求不同,按照匹配类型可以分为范围查询、全部配对查询和 $k$ 最近邻查询三种。本文主要研究基于子序列匹配的范围查询情况下的发动机故障模式挖掘,即查询序列的长度小于时间序列数据库中序列的长度,从数据库中时间序列的所有子序列中查找与查询序列 $q$ 满足相似性要求的所有子序列。

时间序列匹配查询涉及三个关键点,即特征提取(时间序列的表示)、相似性度量(各种距离的定义)和索引结构(时间序列的索引)。特征提取是对原始时间序列的降维近似表示,它取决于用户对相似性的认识和理解,并直接关系匹配查询的精度、可靠性以及运算量。因此,它是时间序列匹配查询的重要基础,相似性度量和索引结构的选择均不能脱离此基础单独考虑。

## 2.2 应用现状

目前, 时间序列相似性匹配技术在地震预报<sup>[6]</sup>, 电力系统故障诊断<sup>[7]</sup>, 航天飞机内部传感器信号离线分析<sup>[8]</sup>等领域有了一定的应用。从目前的应用现状分析, 时间序列相似性匹配技术特别适用于在数据库中进行指定模式的快速匹配搜索, 以满足技术人员对其感兴趣的模式进行分析和比较的需要。因此, 将其用于液体火箭发动机的故障模式的挖掘, 必将对发动机的故障检测和诊断提供较好的技术支持。

## 3 基于序变换的时间序列相似匹配

针对液体火箭发动机试车数据及其应用需求的特点, 通过对相似性度量、如何降低对偏移和噪声的敏感性、算法效率等问题的进一步综合分析, 本节给出了一种基于序变换的时间序列相似匹配搜索算法。

### 3.1 序模式与时间序列的序相似性

序模式分析是对复杂时间序列进行研究的一种新方法<sup>[9]</sup>。其主要思想是通过序模式对时间序列中距离均匀分布的时间序列值排序关系进行描述。从另一角度讲, 序模式是对时间序列幅值(或其它特征值) 随时间变化过程中蕴含的量值相对关系的刻化, 因此它能够比较全面地反映时间序列中蕴含的信息。通常, 两个时间序列相似, 它们的序模式必然在某种程度上相同或相近; 反之, 序模式在某种程度上相同或相近的两个时间序列必然相似。由此可见, 将序模式引入到相似性描述中具有理论上的可行性。

序模式 (Ordinal Pattern) 描述了时间序列的幅值随时间的变化。对于时间序列  $x(t)$ , 其阶次  $d \in \mathbb{N}$ , 且延时  $\tau \in \mathbb{N}$ , 则在  $t$  时刻获得的唯一的排序为:

$$\pi_d^\tau \equiv \begin{pmatrix} 0 & 1 & 2 & \cdots & d \\ r_0 & r_1 & r_2 & \cdots & r_d \end{pmatrix} \equiv (r_0, r_1, r_2, \cdots, r_d) \quad (1)$$

其中,  $\{r_0, r_1, \cdots, r_d\}$  满足:

$$x_{t+r_0\tau} \leq x_{t+r_1\tau} \leq \cdots \leq x_{t+r_d\tau} \quad (2)$$

如果  $x_{t+r_{l-1}\tau} = x_{t+r_l\tau}$ , 则取  $r_{l-1} < r_l$ 。公式(1)中的排序, 当

$l=1, 2, \cdots, d$  时, 其对应的序模式  $P=(i_1, i_2, \cdots, i_d)$  的第  $l$  个元素表示如下:

$$i_l = i(\pi_d^\tau) = \|\{r \mid r \in \{0, 1, \cdots, d-1\}, \pi_d^\tau(r) < \pi_d^\tau(l)\}\| \quad (3)$$

式中,  $\pi_d^\tau(r)$  表示数值  $r$  对应的序值; 运算符 “ $\|\cdot\|$ ” 表示集合的基数, 这里表示集合中满足不等式约束条件的元素  $r$  的个数。

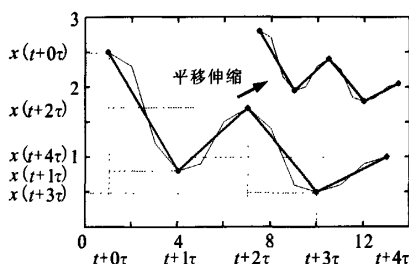


图1 时间序列序模式示意图

Fig.1 Ordinal pattern of a segment of time series

如图1所示, 参数设置为  $t=1$ ,  $d=4$ ,  $\tau=3$ , 得出  $x(t+3\tau) < x(t+\tau) < x(t+4\tau) < x(t+2\tau) < x(t+0\tau)$ ; 因此有唯一的排序:

$$\pi_d^\tau \equiv \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 & 0 \end{pmatrix} \equiv (3, 1, 4, 2, 0) \quad (4)$$

$\pi_d^\tau=(3, 1, 4, 2, 0)$  对应的序模式  $P$ , 可通过如下分析求得: 对于  $l=1$ ,  $\pi_d^\tau(0)$  对应位置为 3, 而  $\pi_d^\tau(1)$  对应的位置为 1,  $\pi_d^\tau(0)=3 > \pi_d^\tau(1)=1$  不满足式(3)中的条件, 有  $i_1=0$ ; 对于  $l=2$ , 有  $\pi_d^\tau(0)=3 < \pi_d^\tau(2)=4$ ,  $\pi_d^\tau(1)=1 < \pi_d^\tau(2)=3$ ,  $r=0$  和  $r=1$  两个元素满足条件, 所以  $i_2=2$ ; ……依此类推, 可得出  $\pi_d^\tau$  的序模式为  $P=(0, 2, 1, 0)$ 。

对图1分析, 可归纳出序模式具有如下性质:

性质1 (序模式的平移伸缩不变性) 对于给定的子序列  $x_1(t)$ , 其对应序模式为  $P_1$ ; 在坐标系内, 将  $x_1(t)$  沿时间轴方向进行任意伸缩, 沿幅值方向进行整体伸缩和平移后得到  $x_2(t)$ , 对应序模式为  $P_2$ 。则有  $P_1=P_2$ 。

在实际的序模式计算中, 通常采用由公式(1)和(3)导出的如下公式进行计算:

$$i_r = i(\pi_d^r) = \|\{r | r \in \{0, 1, \dots, l-1\}, x_{i+r} \geq x_{i+r}\}\| \quad (5)$$

序变换是在序模式定义的基础上,将给定的时间序列变换为处于 $[0, (d+1)!-1]$ 区间的整数,即把时间序列变换为序值。序模式 $(i_1, i_2, \dots, i_d)$ 与序值 $n_d$ 之间形成双射关系,即 $\{0, 1\} \times \{0, 1, 2\} \times \dots \times \{0, 1, \dots, d\}$ 取值空间的元素排列与集合 $\{0, 1, \dots, (d+1)!-1\}$ 中的整数一一对应,且该对应关系描述如下。

$$n_d(P) = \sum_{i=1}^d i_i \frac{(d+1)!}{(l+1)!} \quad (6)$$

对于前面的例子中的序模式 $P=(0, 2, 1, 0)$ ,按照公式(6)可得序值 $n_d(P)=45$ 。同时,通过分析可知当整个序列为升序时,序值 $n_d(P)$ 达到最大值 $(d+1)!-1$ ;当整个序列为降序时,序值 $n_d(P)$ 达到最小值0。

结合序模式的性质,同理可以得出序值的伸缩不变性。

性质2 (序值的平移伸缩不变性) 对于给定子序列 $x(t)$ ,将其沿时间轴向进行任意伸缩,沿幅值方向进行整体伸缩和平移后,序值保持不变。

### 3.2 相似性度量

目前,在时间序列的匹配中,主要采用单一的距离对相似性进行度量。然而,单一的以距离为基础进行相似性度量存在一些无法克服的缺陷。诸如,①对数据库中所有可能的子序列均进行试匹配,从而导致计算量繁重;②点距离为基础的时间序列误匹配(两个明显不同的子序列间的距离可能小于两个明显相似的子序列间的距离);③不符合人类对波形模式的识别中由粗到精、由全局到局部的认知特点;④缺乏明确的物理概念,等等。

序模式和序值的平移伸缩不变性,以及序变换值与序模式一一对应的优良特性,不仅为时间序列相似性提供了更加符合人类思维和视觉特点的刻化,也为时间序列的快速索引提供了一个有效途径。因此,采用序模式与距离相结合的二级度量方法进行相似性度量,即先按照序模式/序值对子序列进行分类,而后在一个分类内部通过距离进行度量,从而实现时间序列的相似匹配。

在基于序模式的时间序列相似匹配中,需要权衡以下两个具体问题:

(1) 若直接采用序模式对原始时间序列进行描述,显然这样不但引入繁重的计算量,而且对相似性匹配搜索没有实际意义。因此,这里采用时间序列分段线性表示,先将时间序列数据基于时间表示成多段相邻的直线,这样不但可以获得时间序列分段拐点作为特征点,而且实现了对原始时间序列的压缩、降噪和趋势特征的提取。

(2) 在基于序模式的分类中,如果直接采用公式(3, 4)的序值计算方法,存在算法灵敏度过高的问题。例如,两个子序列可能因为微小的差异而具有不同的序模式,从而属于不同的分类,这种情况在实际应用中需要避免。为此,定义了一种 $\varepsilon$ 序模式,即带有灵敏度 $\varepsilon$ 约束的序变换,以提高序值分类的鲁棒性。 $\varepsilon$ 序模式中,设定一个实数 $\varepsilon > 0$ ,在序模式生成过程中,同一子序列中的任意两点 $x_{i+r}$ 和 $x_{i+l+r}$ ,如果 $|x_{i+r}-x_{i+l+r}| \leq \varepsilon$ ,则认为 $x_{i+r}$ 和 $x_{i+l+r}$ 是相等的,即对小于 $\varepsilon$ 的变化不敏感。显然, $\varepsilon$ 序模式通过在序模式定义的基础上附加了灵敏度 $\varepsilon$ 约束;提高了其分类的鲁棒性。因此, $\varepsilon$ 序模式在性质上兼备了序模式的平移伸缩不变性和分类鲁棒性。由此, $\varepsilon$ 序模式进行序变换得到的序值称为 $\varepsilon$ 序值。

子序列相似性度量,通过 $\varepsilon$ 序模式分类和趋势分布距离匹配两个步骤实现。

#### (1) 子序列的 $\varepsilon$ 序模式分类

对数据库中的时间序列进行分段线性表示获得特征点序列,根据需要设定特征点子序列的长度,并对各个子序列进行 $\varepsilon$ 序变换,从而可以实现对特征点子序列的划分,将具有相同 $\varepsilon$ 序值的子序列归为同一类。

#### (2) 趋势分布距离

设两个序列 $S_1$ 和 $S_2$ 对应的分段数为 $n$ ,其对应各分段线段的斜率序列分别为 $[k_{11}, k_{12}, k_{13}, \dots, k_{1n}]$ 和 $[k_{21}, k_{22}, k_{23}, \dots, k_{2n}]$ ,定义两个序列的趋势分布距离为:

$$D_k(S_1, S_2) = \sqrt{\sum_{i=1}^n \left( \frac{k_{1i}}{K_1} - \frac{k_{2i}}{K_2} \right)^2} \quad (7)$$

$$K_m = \sum_{i=1}^n |k_{mi}|, \quad m=1,2 \tag{8}$$

按照公式(7)和公式(8)的距离定义，当  $S_1$  或  $S_2$  序列在水平方向进行缩放时，二者间的趋势分布距离不受影响。进一步讲，趋势分布距离是对两个序列内部相对趋势（或称之趋势分布关系）相似性的度量，这就保证了趋势分布距离匹配的合理性。

对上述两点分析可知，如果某特征点子序列  $P$  与待查询特征点序列  $Q$  属于同一个序模式类，则两者趋势分布距离愈短，相似匹配程度愈高。

序模式能够比较全面的描述原有序列中的信息，克服以点距离为基础的时间序列误匹配以及概念不明确等缺陷。

3.3 算法流程

基于序模式的时间序列相似匹配搜索，包括两大部分，即索引生成部分和搜索应用部分。

(1) 索引生成部分

Step1 初始化。设定分段线性化的规则与参数  $\varepsilon$ ，以及子序列的长度与匹配精度；

Step2 分段线性化。对原始时间序列分段线性化获得特征点序列；

Step3 子序列序模式提取及序变换；

Step4 建立索引。根据序变换值建立索引。

(2) 匹配搜索部分

Step5 对待匹配子序列的进行特征点提取、序模式计算及序变换。

Step6 根据变换的结果进行相似性搜索，得到同一序模式分类的匹配子序列。

Step7 在同一序模式分类内部，根据需要按照趋势分布距离进行二次匹配。

4 在发动机故障模式挖掘中的应用

时间序列的相似匹配方法，可以为发动机试车数据的分析提供了一个辅助工具。在应用中可采用两种方法：①先通过历史数据分析和过程知识建立用户感兴趣的模式库，然后利用时间序列挖掘技术将检测样本与兴趣模式进行匹配，并找出兴趣模式库中与之最匹配的样本，在此基础上对检测样本与兴趣模型进行比较分析；②对检测样本的某一段感兴趣，但模式库中没有与之相似的模式，需要对所有历史数据进行匹配搜索，查看是否存在具有相似特征的序列。通常，兴趣模式长度的选取根据情况综合考虑，不宜过短或过长，要选取最能代表模式特征的那一段。

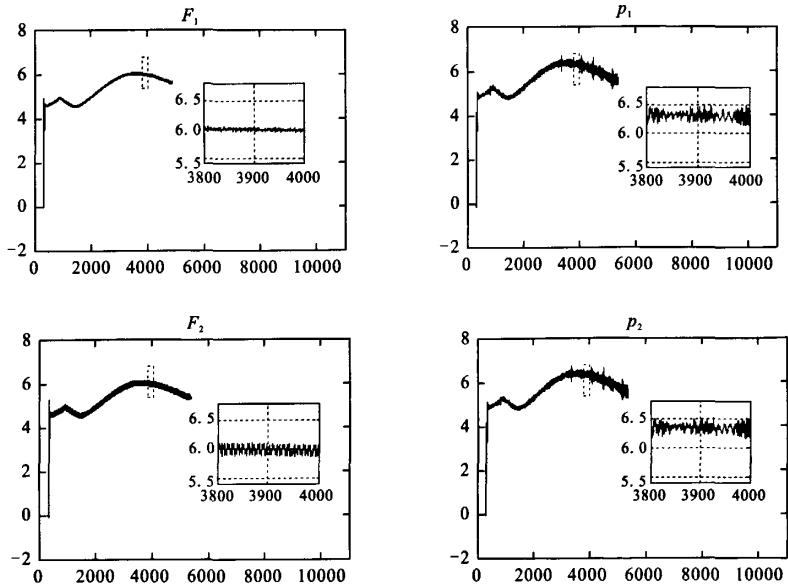


图 2 从 FDJ 截取的部分曲线及其局部放大图

Fig.2 A segment curve of FDJ and its local magnification

由于算法在应用中, 实际数据的噪声直接影响算法对时间序列的分割, 从而影响接下来的匹配质量; 在信噪比很低情况下, 算法甚至会不可用。针对此问题, 以某型液体发动机的某次试车获得的推力和压力数据 FDJ 为实验对象, 以评定算法的抗噪性能。图 2 给出了从 FDJ 数据中截取的异常数据段 (为电压值, 未通过传感器参数换算为实际单位, 并经过了适当处理, 在此不影响实验效果), 其中  $F_1$  和  $F_2$  为推力曲线,  $p_1$  和  $p_2$  为燃烧室压力。理论上讲, 推力  $F$  与推力系数  $C_f$  和喷管喉部截面积  $A_1$  以及燃烧室压力  $p_c$  有如下关系:

$$F=C_f A_1 p_c \tag{9}$$

通常认为推力系数  $C_f$  恒定, 理论上可认为推力  $F$  与燃烧室压力  $p_c$  呈线性关系。因此, 图 2 中的四条曲线具有相似性, 但信噪比差异较大 (图

中  $F_1$  和  $F_2$  为推力无量纲值;  $p_1$  和  $p_2$  为压力无量纲值)。其中推力的信噪比较大, 压力的信噪比相对较小<sup>[10]</sup>。

实验参数设置如下, 采用直线拟合法, 最小拟合长度 100 点, 以拟合的标准偏差  $\sigma=0.055$  为确定分割点的阈值, 灵敏度  $\varepsilon=0$ 。为了检验算法的抗噪能力, 在此不对曲线进行去噪预处理 (通常采用五点三次拟合)。经算法处理, 最后得到的曲线轮廓和关键点。图 3 中给出了与图 2 对应的轮廓和关键点, 虚线框内的曲线部分异常区域, 其拟合线段对应的序模式均为  $P=\{1,2,2,4\}$ , 对应的序变换值为 98。这说明对于四条不同噪声背景的曲线, 提取到了相同的序特征。异常区域内  $F_1$  与  $F_2$ 、 $p_1$ 、 $p_2$  之间趋势分布距离分别为 0.1210、0.0551 和 0.1220。由上述趋势分布距离体现的相似性差异与从原始曲线直观判断吻合。

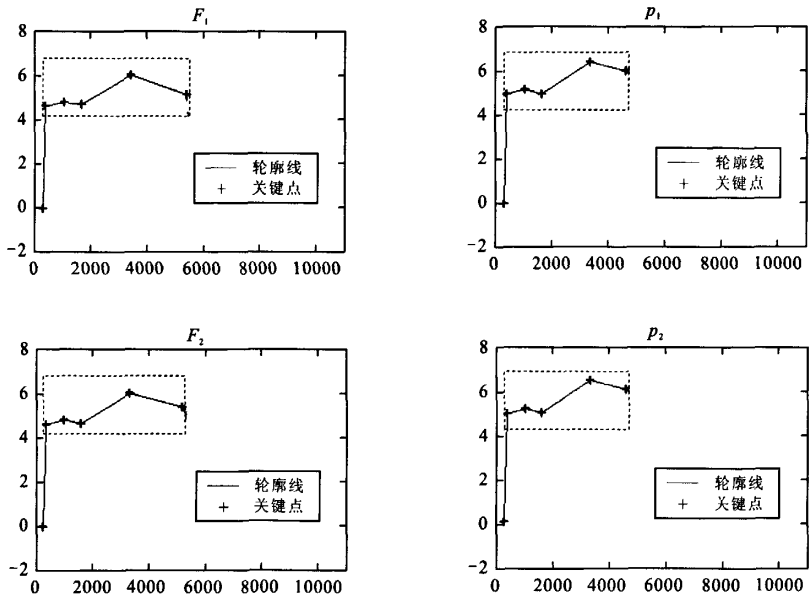


图 3 与图 2 中曲线对应的轮廓线和关键点  
Fig.3 Contour and key points of the curve in Fig.2

从本质上讲, 基于序变换的时间序列匹配方法是一种柔性模式匹配算法。它对噪声信号不敏感, 对检测样本的起始点和长度无严格要求, 只要检测样本中包含了故障信息即可。同时, 它无需对样本数据进行标准化处理, 因为序模式直接

体现了故障幅值的相对关系。因此, 该方法具有良好的鲁棒性。

该算法可辅助用户完成故障模式与检测样本的对比, 提高离线故障分析的效率。若用户感兴趣的曲线特征复杂, 可取其有代表性的局部特征

进行匹配搜索,从检索结果中找出满意的匹配曲线。之所以采用此方法,是因为满足给定全部特征的前提是必须满足局部特征,即满足局部特征是满足全部特征的必要条件。

显然,时间序列的“时间”具有广义坐标轴的含义,本文的时间序列相似匹配方法也可用于对诸如信号的频谱等曲线的匹配搜索。

## 5 结论

综上所述,本文借鉴人类对波形模式识别中由粗到精、由全局到局部的认知特点,将序模式的概念引入到相似性描述中,提出了一种适合发动机故障模式匹配搜索方法,该方法能够实现不同长度子序列的高效相似匹配搜索。对某型液体火箭发动机故障数据进行的相似匹配实验表明该方法对故障持续时间不敏感、不需要精确的故障起始时间、且具有很强的抗噪能力。

### 参考文献:

- [1] 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.
- [2] 张振鹏. 液体火箭发动机故障检测与诊断中的基础问题研究[J]. 推进技术, 2002, 23(5): 353-359.
- [3] 陈启智. 液体火箭推进系统健康监控技术的演变[J]. 推进技术, 1997, 18(1): 1-7.
- [4] 胡小平, 张丽娟, 王艳梅, 等. 液体火箭发动机故障检测和诊断中数据挖掘策略的分析 [J]. 国防科技大学学报, 2005, 27(3): 1-5.
- [5] Agrawal C, Faloutsos A. An efficient Similarity Search in Sequence Database [C]. Proceedings of the 4th International Conference of Foundations of Data Organization and Algorithms, Chicago, October 1993:69-84.
- [6] 吴绍春, 吴耿锋, 王伟, 等. 寻找地震相关地区的时间序列相似性匹配算法[J]. 软件学报, 2006, 17(2): 185-192.
- [7] Li Qiu-Dan, Chi Zhong-Xian, Wang Zhan-Chang. A Novel Method for Similarity Search Over Electric Time Series Data [C]. International Conference on Machine Learning and Cybernetics, Beijing, China, 2-5 Nov. 2003 (4): 2315-2318.
- [8] Fu Wai-chee, Eamonn Keogh, Leo Yung Hang. Scaling and Time Warping in Time Series Querying [C]. Proceedings of the 31st VLDB Conference, Trondheim, Norway, 2005: 649-660.
- [9] Keller K, Sinn M. Ordinal analysis of time series[J]. Physica A: Statistical Mechanics and its Applications (Germany) Vol.356, No.1, 2005:114-120.
- [10] 张育林, 吴建军, 朱恒伟, 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.

(编辑: 王建喜)



## 液体动力“国家队”组建

7月22日,伴随着航天科技集团公司新六院的挂牌成立,让中国航天飞得更高更远的航天液体动力专业,终于正式组建了“国家队”。

新六院整合了我国位于陕西、北京、上海三地的液体动力专业资源—原六院及其成员单位以及北京11所、101所、上海801所等,总部设在西安,共由10个研究所、工厂组成,并控股上市公司—陕西航天动力高科技股份有限公司。它集中了我国液体动力的专业、产品和生产能力,将在融合原有市场、技术、资源、人才、文化优势的基础上,逐步建立我国航天液体动力新体系。

摘自“中国航天报”