

基于时间序列数据挖掘的故障检测方法

李海林¹, 郭崇慧², 杨丽彬¹

(1. 华侨大学工商管理学院, 泉州, 362021; 2 大连理工大学系统工程研究所, 大连, 116024)

摘要: 为了有效地检测发动机试车实验中性能参数发生的异常, 提出一种基于时间序列数据挖掘的发动机故障检测方法。通过基于形态特征的时间序列特征表示方法, 将发动机参数时间序列转化为符号序列, 再根据符号语义对发动机参数序列实现稳态特征和过渡态特征识别。同时, 根据稳态序列的数据特征, 利用基于统计特征的时间序列相似性度量结合最不相似模式发现方法实现发动机的故障检测。数值实验结果表明, 与传统方法相比, 本文方法能够有效地对发动机性能参数进行故障检测, 并且具有较强的鲁棒性。

关键词: 发动机参数; 故障检测; 异常模式; 时间序列数据挖掘

中图分类号: TP18 **文献标志码:** A

Fault Detection Algorithm Based on Time Series Data Mining

Li Hailin¹, Guo Chonghui², Yang Libin¹

(1. College of Business Administration, Huaqiao University, Quanzhou, 362021, China; 2. Institute of Systems Engineering, Dalian University of Technology, Dalian, 116024, China)

Abstract: To validly detect the anomalies of parameters in the engine test, a fault detection algorithm of engine based on time series data mining is proposed. The parameter time series are transformed into symbolic strings by a representation method based on shape features. The stable states and transition states are extracted from the parameter time series according to symbolic semantics. Meanwhile, the detection algorithm of abnormal pattern from the stable states is realized by similarity measurement between time series based on statistic features, combined with the most unusual pattern discovery method. The results of numerical experiments show that the new method validly detects the fault of engine and has the better robustness than the traditional method.

Key words: engine parameter; fault detection; abnormal pattern; time series data mining

引 言

故障检测是用来预防机械故障发生的一种有效手段^[1,2], 也常用于预防发动机运行故障的发生。近年来, 出现多种方法用于发动机故障的检测, 例如基于神经网络的方法^[3]、基于信号处理的方法^[4]和基于规则发现的故障诊断系统^[5]等。另外, 基于红线系统的检测方法早在 20 世纪 90 年代中期就已用于

故障检测,但其性能较低且具有较高的检测错误。后来提出的故障检测异常系统(System for anomaly and failure detection, SAFD)弥补了传统方法的部分不足之处,但其检测质量还依赖于具体参数的设定。对此,王珉等^[7]对参数采样值进行离散化,并结合模式知识发现故障规则,进而提出了一种自适应阈值的故障检测方法(Adaptive threshold algorithm, ATA)。在某种程度上讲,ATA弥补了SAFD的部分不足,但仍有其不够完善之处。例如,系统参数中具有明显特征的数据值可能引起较大的均值和标准差,进而影响阈值的确定;与此同时,ATA因设定参数门限将会导致它无法检测出测量值较小但数据波动形态异常情况。

由于发动机性能参数是与时间相关的数据,它可以被理解为时间序列数据,故可以利用时间序列数据挖掘算法和技术来实现发动机相关信息的分析,其主要包括数据处理^[8]、聚类^[9]、模式发现^[10]和异常检测^[11]等时间序列挖掘算法。为了提高发动机故障检测的效果,本文提出利用时间序列数据挖掘方法来实现故障检测。首先,根据前期研究的基于形态特征的时间序列特征表示方法对参数序列实现符号化特征表示,并且提出相应的算法来识别参数的两种状态,即稳态和过渡态。另外,结合基于统计特征的时间序列相似性度量方法,提出适用于故障检测的异常模式发现算法。数据仿真实验表明,本文提出的方法能有效地识别发动机参数稳态数据波动位于门限内但其观测数据值异常的情况,提高了对发动机性能参数的故障识别效果。

1 理论基础

1.1 时间序列符号化

符号化聚合近似(Symbolic aggregate approximation, SAX)^[8,12]是典型的时间序列符号转化方法,它利用分段聚合近似(Piecewise aggregate approximation, PAA)对时间序列进行特征表示。同时,将时间序列数值域按等概率划分成若干个子区域,每个区域由不同的符号来表示,聚合近似方法所得到的均值序列根据所在区域的符号被转化成字符串序列。

为了有效地对时间序列形态特征进行描述,使用一种基于形态特征的时间序列符号化表示方法(Shape-based SAX, S_SAX)^[13],它有别于传统符号化聚合近似方法。前者利用分段序列的斜率特征来描述形态变化,即基于形态特征的分段聚合近似(Slope based PAA, S_PAA)。斜率的计算通过直线拟合来确定,时间序列 Q 经分段表示被转化为序列 \tilde{Q} ,其长度从 m 降低到 w 。 \tilde{Q} 表示形态序列,被称为S_PAA序列。另外,其元素值 \tilde{q}_i 满足

$$\tilde{q}_i = \frac{k-1}{2} q'_i \quad i=1,2,\dots,w \quad (1)$$

式中 \tilde{q}_i 为时间序列 Q 被平均分成 w 个子序列中的第 i 个子序列段所对应拟合直线的斜率值,即

$$\tilde{q}_i = \frac{k \sum_{j=j_0}^{k \times i} j q_j - \left(\sum_{j=j_0}^{k \times i} j \right) \left(\sum_{j=j_0}^{k \times i} q_j \right)}{k \sum_{j=j_0}^{k \times i} j^2 - \left(\sum_{j=j_0}^{k \times i} j \right)^2} \quad (2)$$

式中 $j_0 = (i-1)k+1$, k 表示子序列段元素值的个数。

根据传统分段聚合符号化的原理,需要事先将数据进行z-score标准化,使其转化为服从标准正态分布的序列。为此可以将标准化数据空间等概率地划分成 α 个子空间或子区域,每个区域可用对应的字符进行表示,即可以完成S_PAA数值序列向S_SAX符号序列的转化。

如图1所示,可以将标准化后的时间序列数据分布空间按等概率的形式化分为3个区域,每个区域用相应的字符来表示,即“A, B, C”,获得相应的字符串“AAABBCCCC”。与传统SAX方法比较,S_SAX序列中的符号符合人的思维活动,具有一定的字符含义,例如,A为下降状态,B为平稳状态,C为

上升状态。因此, S_SAX 字符序列能够很好地反映时间序列的形态变化趋势, 图 1 中符号的描述说明了时间序列经过了连续下降、短暂平缓过度和慢慢上升 3 个阶段。

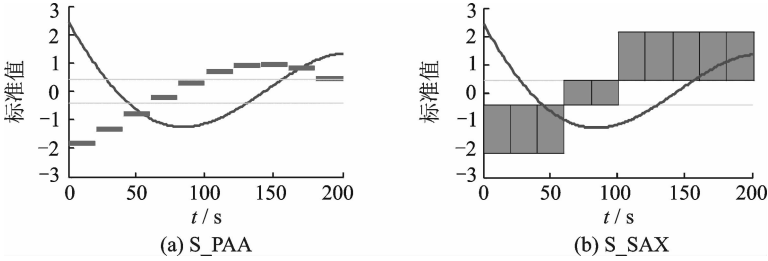


图 1 基于形态特征的时间序列符号化转化

Fig. 1 Time series symbolization based on shape feature

1.2 相似性度量

通过符号化特征表示后, 可以识别发动机参数的形态特征, 即稳态数据和非稳态数据。其中, 稳态数据是发动机试车实验的最重要来源, 也是特征识别的主要对象。稳态数据中的异常模式发现是试车实验中最主要的目标之一。在时间序列数据挖掘中, 特征表示相似性度量是数据分析过程中重要的方法^[14,15]。在异常模式发现任务, 通过选取合适的特征表示方法和相应的距离度量函数, 可以有效地发现数据异常。鉴于发动机参数具有时间序列的时间特性, 结合稳态数据序列的基本特征, 利用基于非线性统计特征的距离度量方法 (Non linear statistical feature based PAA, NLSF_PAA)^[16] 来描述特征序列之间的相似性。该方法不仅能有效地对特征序列进行相似性度量, 还具有良好的下界紧凑性和剪枝能力, 避免在相似性检索中发生漏报情况。

假设 $Q = \{q_1, q_2, \dots, q_m\}$ 被平分成 w 个子序列, 每个序列同时用相应的均值和方法来度量, 即得到均值特征序列 $\bar{Q} = \{\bar{q}_1, \bar{q}_2, \dots, \bar{q}_w\}$ 和标准差特征序列 $\hat{Q} = \{\hat{q}_1, \hat{q}_2, \dots, \hat{q}_w\}$ 。其中

$$\bar{q}_i = \frac{1}{k} \sum_{j=k \langle i-1 \rangle}^{k \times i} q_j \quad (3)$$

$$\hat{q}_i = \sqrt{\frac{1}{k} \sum_{j=k \langle i-1 \rangle}^{k \times i} (q_j - \bar{q}_i)^2} \quad (4)$$

传统 PAA 算法中, 给了基于均值特征序列的距离度量方法。对标准差特征序列的度量函数可以定义为

$$\hat{D}(\hat{Q}, \hat{C}) = \sqrt{k \sum_{i=1}^w (\hat{q}_i - \hat{c}_i)^2} \quad (5)$$

结合这两种距离度量方法, 使用非线性统计特征的时间序列相似性度量方法^[16], 即

$$D_{\text{NLSF_PAA}}(Q, C) = \sqrt{\bar{D}(\bar{Q}, \bar{C})^2 + \hat{D}(\hat{Q}, \hat{C})^2} \quad (6)$$

其中 $\bar{D}(\bar{Q}, \bar{C})$ 为度量 PAA 序列相似性的距离函数。

2 发动机故障检测算法

鉴于发动机稳态数据和过渡态数据的重要性, 结合基于形态特征的时间序列符号化表示方法提出特征识别算法对发动机参数进行这两种状态数据序列的识别。然而, 对于稳态数据序列特征, 使用基于非线性统计特征的发动机检测方法, 充分利用均值和方差两个统计量来描述稳态数据波动情况, 并且结合相应的度量函数来实现故障异常模式。

2.1 特征识别

发动机参数通常包括稳态数据和过渡态数据,其中过渡态数据可分成上升和下降两种状态。通过时间序列数据挖掘中的符号特征表示方法,可以有效地识别这3种状态。如图2所示,序列b和序列c分别为发动机参数的稳态和过渡态两种数据。同时,子图(b)与(c)分别详细显示了序列b和序列c的数据形态状况。从数据波动情况来看,稳态序列和过渡态序列中相邻数据之间没有一定的规律性,使得对它们的识别存在一定的困难。

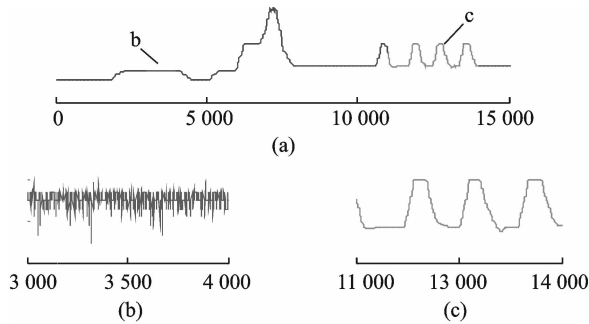


图2 发动机参数数据特征描述

Fig.2 Description of data feature for engine parameter

另外,由于发动机参数之间具有一定的相关性,这将引起参数序列中过渡态和稳态序列之间不存在明显的分界点,一般即有一事实上的缓冲过程,故在特征识别过程中,需要考虑此种情况,以便更有效地识别出这两大类特征。为了解决此问题,提出“去头尾”操作,即经符号特征表示后,删除字符串序列中相邻且相异的两个字符。如在图1中,先后删除相邻且相异子字符串AB再BC,最终获得的两种子字符串AA和CCCC分别表示参数序列的下降过渡态序列和上升过渡态序列。基于形态符号化表示的特征识别算法如下。

算法1 $[G, U, D] = \text{SSAX_FR}(T, k)$

输入:发动机参数 T ,子序列长度 k 。

输出:稳态数据序列集合 G ,上升过渡态数据序列集合 U ,下降过渡态数据序列集合 D 。

(1) 初始化相关数据。参数序列长度 $m = \text{Length}(T)$ 和子序列数目 $w = m/k$,字符(被划分子空间)个数 $\alpha = 3$,字符串 $R = \text{NULL}$ 。

(2) 将参数序列进行标准化处理并将其分成 w 个子序列,即 $S = \{S_1, S_2, \dots, S_w\}$,其中 $S_i = \{t_{\langle i-1 \rangle k+1}, t_{\langle i-1 \rangle k+2}, \dots, t_k\}$ 。

(3) 针对每个子序列 S_i ,利用式(1,2)计算它们形态特征,再结合SSAX方法,将参数序列转为字符串序列 R 。

(4) 对 R 进行“去头尾”操作,即从第2个字符开始至到最后一个字符,依次判断 r_{i-1} 与 r_i 是否相同,其中 $r_i \in R$ 。如两者字符不相同,则删除字符 r_{i-1} 和 r_i 。

(5) 根据符号意义以及相应子序列在参数中的时间位置,可将连续相同的字符合并,可以分别得到3种特征状态 G, U 和 D 。

2.2 异常检测算法

在时间序列数据挖掘中,异常模式是指与其他序列片段最不相似的子序列。目前一种较为流行的时间序列最不相似模式发现算法,即基于SAX的方法,通常也被称作异常模式发现算法。另外,为了提高最不相似模式算法的效率,文献[12]给合SAX的符号化过程和相应的启发式规则来实现最不相似模

式序列的识别,其算法过程描述如下。

算法 $[d, \text{loc}] = \text{HeuristicSearch}(\mathbf{Q}, n, \text{Outer}, \text{Inner})$

输入: 时间序列 \mathbf{Q} , 模式长度 n , 启发式规则 Outer 和 Inner。

输出: 最不相似模式与其他模式的最小距离 d 和该模式在时间序列 \mathbf{Q} 中的位置 loc。

(1) 初始化相关数据, 初始化 $d=0$ 和 $\text{loc}=-1$ 。分别利用启发式规则 outer 和 inner 控制本算法的内外两层循环, 用 p 和 q 来记录两个模式在 \mathbf{Q} 中的位置信息。

(2) 对于外层循环, 每个 p 将设定初始最近距离为 $d_0 = +\infty$, 并根据 q 值执行内层循环, 即:

(a) 若 $|p - q| > n$, 则两个模式为平凡匹配模式, 执行步骤(b); 否则, 返回步骤(2);

(b) 若 $\text{Dist}(\mathbf{Q}_p^n, \mathbf{Q}_q^n) < d$, 则跳出内层循环并返回步骤(2); 否则, 执行步骤(c);

(c) 若 $\text{Dist}(\mathbf{Q}_p^n, \mathbf{Q}_q^n) < d_0$, 则 $d_0 = \text{Dist}(\mathbf{Q}_p^n, \mathbf{Q}_q^n)$ 。

(3) 判断 d_0 是否大于 d 。若为真, 则 $d = d_0$ 且记录相应的位置信息 $\text{loc} = p$, 同时返回步骤(2), 直到遍历完外循环中的所有 p 值为止。

(4) 返回最小距离 d 和相应的位置信息 loc。

在上述最不相似模式发现算法中, \mathbf{Q}_p^n 和 \mathbf{Q}_q^n 分别表示在时间序列 \mathbf{Q} 中以起始位置为 p 和 q 且长度为 n 的模式序列。Dist 表示模式的距离度量函数, Outer 和 Inner 是改善模式发现算法效率的启发式规则^[12]。

由于基于统计特征的时间序列距离度量方法能较好地对序列片段进行相似性度量, 为了提高实际发动机参数稳态数据中的异常情况, 结合 HeuristicSearch 算法来进行发动机故障检测, 故在上述模式算法 HeuristicSearch 中, 利用式(6)来计算模式之间的距离, 即 $\text{Dist} = D_{\text{NLSF-PAA}}$ 。

通过特征识别方法 SSAX_FR, 可以获得发动机参数稳态特征序列片段集 $\mathbf{G} = \{g_1, g_2, \dots, g_k\}$, 其中任意 g_i 表示某一稳态特征序列片段。通常情况下, 这些特征存在同一水平状态下的序列片段, 故可以合并具有相同水平状态的稳态数据序列, 使得原稳态数据序列集合 \mathbf{G} 转化为水平状态各异的稳态特征序列集合 $\mathbf{G}' = \{g'_1, g'_2, \dots, g'_k\}$, 其中 $k < K$ 且 \mathbf{G}' 是按其特征序列的均值从小到大排列, 即 $\forall i, j, i < j$, 有 $\bar{g}'_i < \bar{g}'_j$ 。基于非线性统计特征表示方法的故障检测算法如下。

算法 3 $[P, \text{loc}] = \text{NLSF_AbnormSearch}(g'_i, n)$

输入: 某一水平状态下的稳态数据序列 g'_i , 故障检测模式的长度 n 。

输出: 故障模式 P 和在稳态数据序列中出现的位置 loc。

(1) 将基于启发式规则的最不相似模式发现算法 $[d, \text{loc}] = \text{HeuristicSearch}(\mathbf{Q}, n, \text{Outer}, \text{Inner})$ 中的距离度量函数 $\text{Dist}(*, *)$ 替换成式(6)所示距离度量函数 $D_{\text{NLSF-PAA}}(*, *)$, 即可获得最不相似模式序列 P 、最小距离 d 和位置信息 loc。

(2) 从 g'_i 中移除异常模式序列 P 。若 $\forall S_j^n \in g'_i - P$, 有 $D_{\text{NLSF-PAA}} > \epsilon$, 则视异常模式 P 为发动机稳态数据的故障序列; 否则, 返回 $P = \text{NULL}$ 和 $\text{loc} = -1$, 表示不存在故障情况。

通过 HeuristicSearch 算法, 获得模式 P 可能是时间序列中存在的异常模式, 也可能是与其他模式存在较大差异的正常序列片段。在发动机故障检测算法中, 通过给定的阈值 ϵ 来再次验证获取的最不相似模式 P 是否为存在异常。需要说明的是, 阈值 ϵ 可以根据具体情况人为设置, 也可以通过 HeuristicSearch 算法在 $g'_i - P$ 中再次搜索最不相似模式, 并得到相应的最小距离 d 值。将阈值 ϵ 取设定为最小距离 d 值, 即 $\epsilon = d$ 。

3 数值实验

为了验证发动机故障检测算法的可行性和有效性, 采用某型号发动机试车实验中参数的仿真数据进行故障检测实验。本次实验分为两个步骤: 特征识别实验和故障检测实验。前者通过基于形态特征

的时间序列符号化表示方法对仿真数据进行特征识别,进而说明基于形态特征表示方法用来进行发动机参数特征识别的可行性和有效性;后者利用本文提出的基于最不相似模式发现的发动机故障检测算法来对稳态数据进行异常检测,同时与传统方法相比,验证新方法对发动机稳态数据故障分析的有效性和检测效果。

3.1 特征识别实验

根据前面分析易知,发动机试车实验参数表现出时间序列数据的特征,故可以利用基于形态特征的符号化表示方法对其进行特征识别,即稳态特征和过渡态特征的识别。由于这两类特征可以细分成3种状态,即平稳状态、上升状态和下降状态,在识别过程中设置3种字符来分别表示3种状态的情况。根据算法要求,将长度为 m 的发动机参数数据平均分成 w 个序列片段,每个子列片段的数据量(即长度)为 k 。同时,根据具体需要将服从正态分布的数据区域划分成3份,即 $\alpha=3$,对应的符号分别为A,B和C。通过特征识别算法即可将发动机参数序列转化成相应的字符序列特征。另外,根据字符的具体语义,提取对应的特征序列。

在本次算法实验中,设定 $k=100$,其表示每个序列片段的长度为100,则通过特征识别算法即可获得发动机参数序列的特征识别。为了便于显示和描述,图3(a)中实线和其他3个子图中虚线表示同一发动机参数序列的数据信息。然而,在图3(b~d)中红色序列片段表示特征识别算法的运行结果。从图中显示效果容易判断,参数特征识别算法SSAX_FR能有效地对发动机参数序列进行3种主要状态特征的识别,即图3(b)显示稳态特征、图3(c,d)分别显示上升和下降等两种过渡态数据。

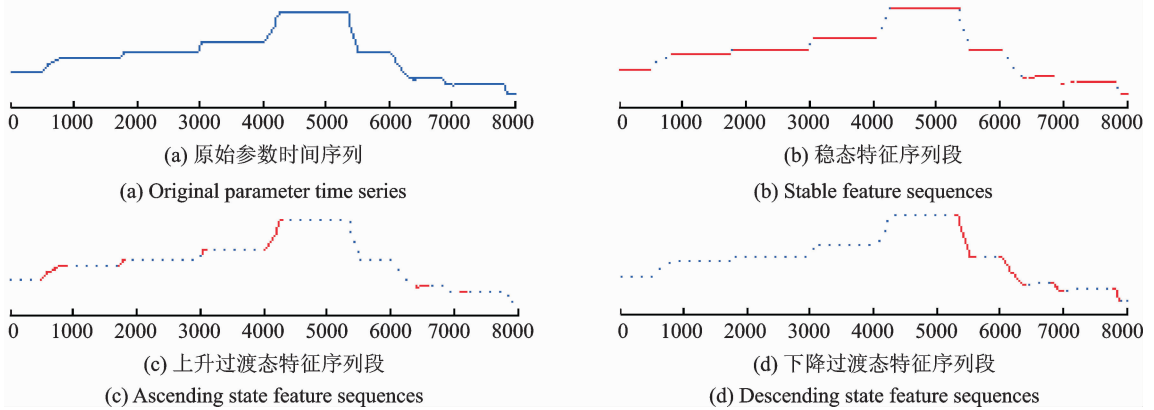


图3 发动机参数的特征识别

Fig. 3 Feature recognition of engine parameter

另外,通过对已标记稳态特征的参数序列进行参数特征识别算法SSAX_FR的执行,并考查不同 k 对稳态特征识别的质量,即稳态特征的识别率。通过对长度为8000的参数时间序列进行稳态特征识别,在 k 分别取值为[20,40,80,160,320]情况下,稳态特征识别的质量如图4所示。由图4性能分析结果易知,在同一采集频率下的发动机参数序列中, k 值在某一范围内能取得较好的识别效果。然而,若 k 值太大,则其稳态识别效果出现不好的现象,其原因在于较大的 k 值使得用某一字符来表示长度为 k 的序列段过于粗糙,

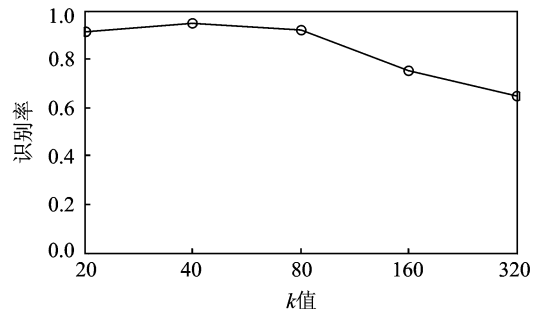


图4 SSAX_FR随 k 值变化的识别效果

Fig. 4 Recognition results of SSAX_FR with k

进而降低了稳态特征的识别质量。在发动机参数特征识别过程中,序列片段长度 k 还通常根据具体发动机参数试车实验中信息采集频率来决定。若试车信息采集频率越大,则取大值较为适宜;否则, k 取较小值。另外,由于特征识别算法中存在“去头尾”操作,故 k 值的选定也需考虑“去头尾”的长度。

3.2 异常检测实验

通过基于时间序列形态符号化表示的参数特征识

别算法,可以从发动机参数模拟数据中获得稳态数据序列集合 G 。针对稳态特征序列集合中的特征序列,提取同一水平状态下的稳态特征序列进行异常检测分析。为了更好地说明本文提出的基于统计特征的相似性度量^[15]进行发动机参数稳态特征数据异常检测方法的有效性和优越性,利用欧氏距离 Euclidean,基于分段聚合近似 PAA 的距离度量^[8]的异常检测方法来进行实验,即利用欧氏距离和基于 PAA 的距离度量函数来代替基于启发式的最不相似模式发现算法中的 Dist。

通过故障检测算法分析,其实验结果如图 5 所示。在图 5(a)中,给出了检测发动机参数中某个水平状态下的稳态数据序列片段逃逸出大部分子序列的数据波动范围,且序列片段是异常模式;图 5(b)则给出了发动机参数中另外某个水平状态下的稳态数据中,某段子序列的波动振幅突然小于大部分子序列的数据振幅,故该序列片段应该被看作是发动机的故障异常模式。

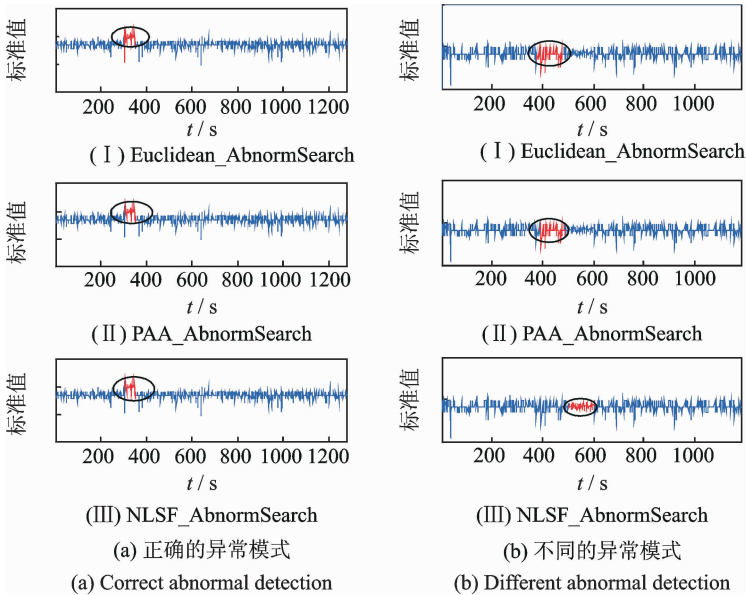


图 5 两种稳态特征数据的异常检测结果

Fig. 5 Abnormal detection for two kinds of stable feature data

实验结果显示,3种方法对于发动机参数中稳态数据片段波动性较大的异常模式都可以有效地识别,如图 5(a)所示。然而,对于波动性小于大部分子序列的情况,基于欧氏距离 Euclidean 和基于分段聚合近似 PAA 的异常检测方法无法进行有效识别,而本文提出的方法 NLSF_AbnormSearch 能够有效地发现该种情况的故障模式,如图 5(b)所示。因此,与传统特征表示和相似性度量相比,基于非线性统计特征表示的故障检测方法能较好地对发动机的异常模式进行识别。

针对上面两种情况,利用文献[7]提出的发动机故障检测方法 ATA 进行数据实验,该方法的发动机故障检测结果如图 6 所示。该实验结果表示,ATA 把超出门限范围的数据点视为了数据异常点,没能有效地对故障模式片段进行识别。与实验结果图 5(b)相比,本文提出的基于时间序列数据挖掘的发

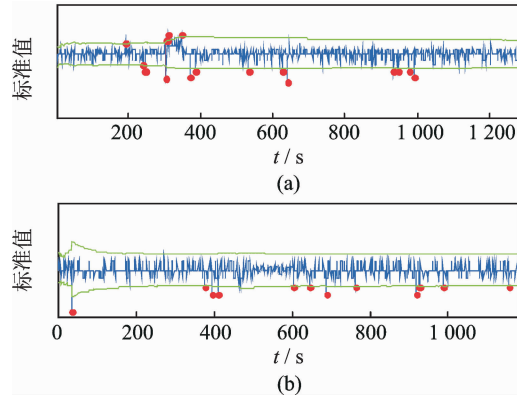


图6 ATA算法对两种数据情况的异常检测结果

Fig.6 Abnormal detection of ATA for the two kinds of data

动机故障检测方法 NLSF_AbnormSearch 能够很好地检测出发动机参数稳态特征序列中出现的异常模式,为预防发动机故障提供了可行的技术和方法。

4 结束语

通过对传统发动机故障检测方法的分析,本文提出一种基于时间序列数据挖掘的故障检测方法。根据发动机试车参数的时间序列特性以及参数中稳态数据和过渡态数据的重要性,利用基于时间序列数据挖掘中形态特征符号化表示方法对参数时间序列进行字符串转化,使其转化为具有实际语义的字符,并以此对发动机参数序列的两种状态特征进行有效的识别。另外,本文针对发动机试车实验中出现频率较高的稳态特征序列数据进一步分析,利用最不相似模式发现算法并结合基于非线性统计特征表示的时间序列距离度量方法实现该特征数据的异常模式检测。在数值仿真实验中,通过特征识别和异常检测结果的比较,验证了新方法对发动机参数故障检测的有效性。同时,与传统方法相比,新方法能够较好地发现和检测异常模式,具有较强的鲁棒性。然而,本文的重点主要在于发动机参数序列数据中状态特征的识别和稳态数据中和异常故障检测,对于过渡态数据序列中的模式发现尚未涉及到。因此,研究一种检测发动机参数时间序列过渡态特征数据的异常模式方法将是下一步工作的重要内容。

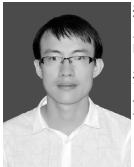
参考文献:

- [1] 许洁, 赵瑾, 刘如成, 等. 基于 KICA-KFDA 的集成故障识别算法[J]. 数据采集与处理, 2013, 28(6): 812-817.
Xu Jie, Zhao Jin, Liu Rucheng, et al. An integrated fault identification algorithm based on KICA and KFDA[J]. Journal of Data Acquisition and Processing, 2013, 28(6): 812-817.
- [2] 窦唯, 刘占生. 液体火箭发动机涡轮泵故障诊断的新方法[J]. 推进技术, 2011, 32(2): 266-270.
Dou Wei, Liu Zhansheng. A new fault diagnosis method for turbopump of liquid rocket engine[J]. Journal of Propulsion Technology, 2011, 32(2): 266-270.
- [3] 朱明悦, 李小申. 基于概率神经网络的发动机故障诊断方法研究[J]. 制造业自动化, 2012, 34(5): 90-92.
Zhu Mingyue, Li Xiaoshen. Research on fault diagnosis of engines based on probabilistic neural networks [J]. Manufacturing Automation, 2012, 34(5): 90-92.
- [4] 何皓, 覃道亮, 孔祥兴, 等. 基于 UIO 的航空发动机执行机构故障诊断[J]. 推进技术, 2012, 33(1): 98-104.
He Ai, Qin Daoliang, Kong Xiangxing, et al. UIO-based diagnosis of aeroengine actuator faults[J]. Journal of Propulsion Technology, 2012, 33(1): 98-104.
- [5] 陈果, 左洪福. 基于知识规则的发动机磨损故障诊断专家系统[J]. 航空动力学报, 2004, 19(1): 23-29.
Cheng Guo, Zuo Hongfu. Expert systems of engine wear fault diagnosis based on knowledge rule[J]. Journal of Aerospace

Power, 2004, 19(1):23-29.

- [6] Panossian H V, Kemp V R. Technology test bed engine real time failure control[R]. NASA-CR-192414, Canoga Park, CA: Rockwell International Corp, 1992.
- [7] 王珉, 胡鸾庆, 秦国军. 基于模式矩阵的液体火箭发动机试车台故障关联规则挖掘[J]. 宇航学报, 2011, 32(4): 947-951.
Wang Min, Hu Niaoqing, Qin Guojun. Association rules of liquid-propellant rocket engine test-bed based on pattern matrix [J]. *Journal of Astronautics*, 2011, 32(4): 947-951.
- [8] 霍铨宇, 倪黄晶, 宁新宝. 心率变异时间序列的预处理算法[J]. 数据采集与处理, 2013, 28(5):591-596.
Huo Chengyu, Ni Huangjing, Ning Xinbao. Preprocessing methods for heart rate variability time series[J]. *Journal of Data Acquisition and Processing*, 2013, 28(5):591-596.
- [9] 夏利, 王建东, 张霞, 等. 聚类再回归方法在机场噪声时间序列预测中的应用[J]. 数据采集与处理, 2014, 29(1):152-156.
Xia Li, Wang Jiandong, Zhang Xia, et al. Application of cluster regression in time series prediction of airport noise[J]. *Journal of Data Acquisition and Processing*, 2014, 29(1):152-156.
- [10] 万里, 廖建新, 朱晓民. 一种时间序列频繁模式挖掘算法及其在WSAN行为预测中的应用[J]. 电子与信息学报, 2010, 32(3): 682-686.
Wan Li, Liao Jianxin, Zhu Xiaomin. Time series frequent pattern mining algorithm and its application to WSAN behavior prediction[J]. *Journal of Electronics & Information Technology*, 2010, 32(3): 682-686.
- [11] Fujimaki R, Nakata T, Tsukahara H, et al. Mining abnormal patterns for heterogeneous time series with irrelevant features for fault event detection[J]. *Statistical Analysis and Data Mining*, 2009, 2(1): 1-17.
- [12] Keogh E, Lin J, Fu A. Hot Sax: Finding the most unusual time series subsequence: Algorithms and applications[J]. *Knowledge and Information Systems*, 2006; 11(1):1-27.
- [13] 李海林, 郭崇慧. 基于形态特征的时间序列符号聚合近似方法[J]. 模式识别与人工智能, 2011, 24(5): 665-672.
Li Hailin, Guo Chonghui. Symbolic aggregate approximation based on shape features[J]. *Pattern Recognition and Artificial Intelligence*, 2011, 24(5): 665-672.
- [14] 丁永伟, 杨小虎, 陈根才, 等. 基于弧度距离的时间序列相似度量[J]. 电子与信息学报, 2011, 33(1): 122-128.
Ding Yongwei, Yang Xiaohu, Chen Gencai, et al. Radian-distance based time series similarity measurement[J]. *Journal of Electronics & Information Technology*, 2011, 33(1): 122-128.
- [15] Li Hailin, Guo Chonghui. Piecewise cloud approximation for time series mining[J]. *Knowledge-Based Systems*, 2011, 24(4): 492-500.
- [16] Guo Chonghui, Li Hailin, Pan Donghua. An improved piecewise aggregate approximation based on statistical features for time series mining[C]//Proceedings of the 4th International Conference on Knowledge Science, Engineering and Management. Berlin, Heidelberg:Springer-Verlag, 2010: 234-244.

作者简介:



李海林 (1982-), 男, 博士, 副教授, 研究方向: 数据挖掘与智能决策, E-mail: hailin@mail.dlut.edu.cn.



郭崇慧 (1973-), 男, 博士, 教授, 研究方向: 数据挖掘与决策支持。



杨丽彬 (1982-), 女, 讲师, 研究方向: 数据挖掘与信息系统。