基于信息熵的复杂系统动态特征辨识及应用研究

张楼悦[1],王 曦[1],石多奇[1],王 信[2],白克强[3],钱秋朦[2]
(1. 北京航空航天大学 能源与动力工程学院,北京 100191;2. 中国航发四川燃气涡轮研究院 高空模拟技术重点实验室,四川 绵阳 621703;3. 西南科技大学 信息与控制工程学院,四川 绵阳 621010)

摘 要:大型复杂系统在运行中会采集并存储大量过程数据。复杂数据的特征提取、利用以及辨识水平对系统建模、仿真系统构建、控制算法设计等至关重要。以飞行环境模拟系统的试验数据为研究对象,首先,提出一种基于概率云空间的信息熵计算方法,该方法能够有效辨识出试验数据信号幅值变化率的动态特征。随后,以信息熵作为关键技术支撑,分别开展了积分步长参数自适应调整的跟踪微分器算法以及结合Apriori算法的信号关联性挖掘算法研究。结果表明,所开发的自适应跟踪微分器算法能够在运算周期受限的情况下实现稳态与动态兼优的信号提取品质,所提出的信号关联性挖掘算法能够有效反映不同试验参数间的内部规律。

关键词:飞行环境模拟系统;复杂系统;信息熵;特征辨识;Apriori算法;跟踪微分器
中图分类号:V231.1 **文献标识码**:A **文章编号**:1672-2620(2026)01-0041-13
DOI:10.3724/j.GTER.20250086

Dynamic feature identification and application of complex systems based on information entropy

ZHANG Louyue[1], WANG Xi[1], SHI Duoqi[1], WANG Xin[2], BAI Keqiang[3], QIAN Qiumeng[2]
(1. School of Energy and Power Engineering, Beihang University, Beijing 100191, China; 2. Science and Technology on Altitude Simulation Laboratory, AECC Sichuan Gas Turbine Establishment, Mianyang 621703, China; 3. School of Information and Control Engineering, Southwest University of Science and Technology, Mianyang 621010, China)

Abstract: In the operation of large complex systems, a substantial amount of process data is collected and stored. The extraction, utilization, and identification of features from complex data are crucial for system modeling, simulation system construction, and the design of control algorithm. The experimental data of the flight environment simulation system (FESS) was investigated. Firstly, an information entropy calculation method based on the probability cloud space was proposed, which could effectively identify the dynamic characteristics of the rate of change in the signal amplitude of the experimental data. Secondly, the information entropy was used to supportthe research on the adaptive tracking differentiator algorithm of integral step parameter and the signal correlation mining algorithm combined with the Apriori algorithm. The results show that the developed adaptive tracking differentiator algorithm can achieve superior signal extraction quality for both steady and dynamic states within the constraints of computational cycles. The proposed signal correlation mining algorithm can effectively reflect the internal patterns among different experimental parameters.

收稿日期:2025-03-19
基金项目:国家财政稳定支撑项目(GJCZ-0302-05)
作者简介:张楼悦(1997−),男,浙江金华人,博士研究生,研究方向为航空发动机飞行环境模拟系统建模与控制算法设计。
引用格式:张楼悦,王 曦,石多奇,等. 基于信息熵的复杂系统动态特征辨识及应用研究[J]. 燃气涡轮试验与研究,2026,39(1):41−53. ZHANG Louyue,WANG Xi,SHI Duoqi,et al. Dynamic feature identification and application of complex systems based on information entropy[J]. Gas Turbine Experiment and Research,2026,39(1):41−53.

Key words: flight environment simulation system; complex system; information entropy; characteristic identification; Apriori algorithm; tracking differentiator

0 引言

航空发动机高空模拟试车台(以下简称高空台)能在地面上模拟航空发动机空中飞行高度、速度和进气畸变条件,是研制先进航空发动机和推进系统必不可少、最有效的试验设备之一[1-3]。飞行环境模拟系统(FESS)是高空台的重要组成部分,是准确模拟被试发动机在宽广飞行范围内飞行高度、飞行速度的核心设备。可编程逻辑控制器(PLC)具备高可靠性、元器件高稳定性和抗强磁干扰性,是目前国内外普遍采用的FESS的控制器硬件。

FESS通过控制发动机进气压力、温度以及试验舱环境压力来建立发动机在不同高度、马赫数下的飞行条件[4]。FESS规模庞大且结构复杂,包含多个调节控制子系统,如主/副供气调节系统、发动机进口总压抗扰控制系统、发动机高空舱高度控制系统等。各调节控制子系统联合工作时会相互耦合、互为干扰,且呈现出强非线性特征,往往无法基于数学模型进行精确描述[5-8]。FESS在试验过程中会产生海量的试验数据,蕴含着整个系统的内部特性,包括耦合特性、频率特性等。这些数据样本是数据挖掘、模型辨识等机器学习技术研究的重要资源,也是我国未来高空台数字化、智能化的技术增量所在。

业界学者针对FESS试验数据开展了若干数据处理、分析、提取、挖掘等相关技术研究,但研究对象主要为稳态试验数据,且研究集中于对数学模型的修正[9-12]。原因在于,FESS体积庞大、结构复杂、内部参数变动剧烈、传感器点位繁杂,同时受限于硬件性能和算法复杂度,控制器的实际执行频率较低。目前,针对FESS过渡态试验数据的自动获取、信息处理、关键特征提取和关联关系挖掘的相关技术研究尚处于起步阶段[13-14],且面临多个技术难题。①针对大型复杂系统的试验数据,缺乏有效手段来高效辨识采样信号的动态特征信息,难以为各类基于数据驱动的技术应用提供有效支撑。②目前控制算法整体复杂度较高,导致控制器的执行频率较低,进而无法实现稳态与动态兼优的滤波去噪品质。另外,滤波器的参数设定仍依赖于技术人员经验进行手动调试。③高空台的试验采样数据种类繁杂,包含调节阀开度、容腔压力温度、截面流量等大量数据。针对模型辨识、故障诊断、耦合性分析等特定任务,需要充分挖掘数据间的关联性,但目前尚缺乏一种有效的评估手段和决策机制。

综上,高空台FESS目前亟需一种可以有效对系统内试验数据进行动态特征辨识的实用型工具,以及将其引入去噪算法、数据挖掘等依赖于系统动态特征的智能算法。为此,本文以高空台FESS过渡态试验数据为研究对象,提出了一种基于信息熵的复杂系统数据动态特征辨识方法。该方法针对FESS的试验数据特性,设计了一种将信号区间转化为概率云的数据预处理方案,进而提升了信息熵在低运算频率下的辨识能力。相较于已有的动态特征辨识方法,如基于线性假设的数据回归方法[15]、基于机器学习的信号模式识别方法[16]、基于复杂网络的数据分析方法[17-18]、基于频域变换的信号分析方法等[19],所提出的辨识算法具有不依赖于模型、算法简洁、迟滞性低、鲁棒性较强的优势,在FESS这类大型复杂系统中具有更强的适配性。

1 算法原理、设计及验证

1.1 试验数据信息熵特征提取

信息熵是一种用于量化信息中的不确定性或混乱性的概念[20],可表达为式(1)。可以看出,样本数据的不确定性或混乱性同$H(X)$成正比。基于这一特性,可以利用信息熵来表征或量化时序信号所包含的不确定性占比。

$$H(X)=-\sum_I p_I \log p_I \tag{1}$$

式中:$H(X)$为随机变量X的信息熵;p_I为事件I发生的概率;求和项遍历随机变量X的所有可能事件。

以高空台FESS试验数据为例,当某时间段内的试验数据处于稳态时,试验数据的数值波动主要来自噪声扰动,不确定性大,对应的$H(X)$值相对较高;当某时间段内的试验数据处于明显过渡态时,试验数据的数值波动主要来自信号本身的动态响应,不确定性小,对应的$H(X)$值相对较低。

1.2 信息熵概率云矩阵计算

假设:①采样信号的噪声幅值满足随机分布;②采样信号在时间尺度上存在不确定性且满足

随机分布。在以上2项假设下,运算器在时间t下接收到的采样信号$z(t)$可表示为

$$z(t)=v(\hat{t})+N \tag{2}$$

式中:v为真实信号;N为满足随机分布的噪声;$\hat{t}$为$(t-\mathrm{d}t,t+\mathrm{d}t)$范围内的随机值,$\mathrm{d}t$为运算器的运算周期。

基于式(2)可以将运算器接收到的采样信号在时间-幅值平面上转化为概率云。方法如下:

将单个信号点$z(t)$映射为$n\times n$的矩阵$\boldsymbol{Z}(t,z(t))$,即(以n=5为例)

$$z(t)\rightarrow \boldsymbol{Z}(t,z(t))=\begin{bmatrix} z_{22}^{t} & z_{21}^{t} & z_{20}^{t} & z_{21}^{t} & z_{22}^{t} \\ z_{12}^{t} & z_{12}^{t} & z_{10}^{t} & z_{12}^{t} & z_{22}^{t} \\ z_{02}^{t} & z_{01}^{t} & z_{00}^{t} & z_{01}^{t} & z_{02}^{t} \\ z_{12}^{t} & z_{11}^{t} & z_{10}^{t} & z_{11}^{t} & z_{12}^{t} \\ z_{22}^{t} & z_{21}^{t} & z_{20}^{t} & z_{21}^{t} & z_{22}^{t} \end{bmatrix} \tag{3}$$

基于二维高斯分布函数计算矩阵中的元素z_{ij}:

$$z_{ij}^{t}=\frac{1}{2\pi\sigma^{2}}\mathrm{e}^{-\left[(\mu i)^{2}+(\eta j)^{2}\right]/2\sigma^{2}},\ i,j=0,1,2 \tag{4}$$

式中:μ、η分别为$z_{(i+1)(j+1)}^{t}$与z_{ij}^{t}之间的幅值跨度和时间跨度;σ为二维高斯分布函数的标准差,其数值通过历史试验数据特性以及多次仿真调试确定。

采用式(3)、式(4)可以基于高斯分布函数将信号点映射为n阶矩阵,矩阵中的每个元素代表真实信号点落在该位置的概率。由此称$\boldsymbol{Z}(t,z(t))$为$z(t)$的概率云矩阵,$\boldsymbol{Z}(t,z(t))$在时间-幅值二维空间中的坐标为$(t,z(t))$。若将$\boldsymbol{Z}(t,z(t))$中各元素的数值视为灰度,可绘制出单个信号点的概率云图,如图1所示。

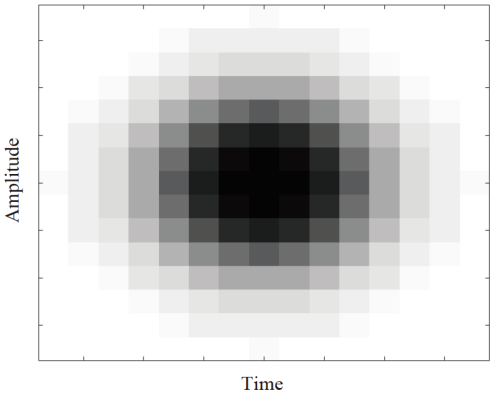


图1 单个信号点的概率云图(n=15)

Fig.1 Schematic of the probability cloud for a single signal point (n=15)

取$t-W\mathrm{d}t$至t时刻为动态辨识区间(W为区间宽度),生成用于描述概率云的二维空间(概率云空间),并令该空间内的横坐标范围为$(t-W\mathrm{d}t,t)$,纵坐标的上限$Z_{\max}$、下限$Z_{\min}$分别为

$$\begin{aligned} Z_{\max}&=\max\{z(t),z(t-\mathrm{d}t),\ldots,z(t-W\mathrm{d}t)\} \\ Z_{\min}&=\min\{z(t),z(t-\mathrm{d}t),\ldots,z(t-W\mathrm{d}t)\} \end{aligned} \tag{5}$$

W的取值会影响辨识的实时性、准确性以及运算速度。依托大量实际系统数据进行仿真测试,确认(300,400)为W在FESS试验数据环境下合理的取值区间,能够兼顾实时性、准确性以及运算速度。

由此可将$\boldsymbol{Z}(t,z(t))$按坐标放置于该空间中,若相邻矩阵存在重叠,则对重叠的部分进行元素求和,如图2所示。

两相邻矩阵间重叠部分的求和如式(6)所示(以n=5为例):

$$\begin{bmatrix} \boldsymbol{Z}(t-\mathrm{d}t,z(t-\mathrm{d}t)) & 0 \\ 0 & \boldsymbol{Z}(t,z(t)) \end{bmatrix}=\begin{bmatrix} z_{22}^{t-\mathrm{d}t} & z_{21}^{t-\mathrm{d}t} & z_{20}^{t-\mathrm{d}t} & z_{21}^{t-\mathrm{d}t} & z_{22}^{t-\mathrm{d}t} & 0 & 0 & 0 \\ z_{12}^{t-\mathrm{d}t} & z_{12}^{t-\mathrm{d}t} & z_{10}^{t-\mathrm{d}t} & z_{22}^{t}+z_{12}^{t-\mathrm{d}t} & z_{21}^{t}z_{22}^{t-\mathrm{d}t} & z_{20}^{t} & z_{21}^{t} & z_{22}^{t} \\ z_{02}^{t-\mathrm{d}t} & z_{01}^{t-\mathrm{d}t} & z_{00}^{t-\mathrm{d}t} & z_{12}^{t}+z_{01}^{t-\mathrm{d}t} & z_{12}^{t}+z_{02}^{t-\mathrm{d}t} & z_{10}^{t} & z_{12}^{t} & z_{12}^{t} \\ z_{12}^{t-\mathrm{d}t} & z_{11}^{t-\mathrm{d}t} & z_{10}^{t-\mathrm{d}t} & z_{02}^{t}+z_{11}^{t-\mathrm{d}t} & z_{01}^{t}+z_{12}^{t-\mathrm{d}t} & z_{00}^{t} & z_{01}^{t} & z_{02}^{t} \\ z_{22}^{t-\mathrm{d}t} & z_{21}^{t-\mathrm{d}t} & z_{20}^{t-\mathrm{d}t} & z_{12}^{t}+z_{21}^{t-\mathrm{d}t} & z_{12}^{t}+z_{22}^{t-\mathrm{d}t} & z_{10}^{t} & z_{12}^{t} & z_{12}^{t} \\ 0 & 0 & 0 & z_{22}^{t} & z_{21}^{t} & z_{20}^{t} & z_{21}^{t} & z_{22}^{t} \end{bmatrix} \tag{6}$$

经以上方法可以获得$(t-W\mathrm{d}t,t)$时间区间内采样信号的概率云空间,对该空间内所有元素进行信息熵计算,并将该区间的信息熵视为t时刻信号$z(t)$的数据动态辨识结果。

为验证前述信息熵与信号幅值变化率之间的关联性,取3种具有不同幅值变化率的信号(式(7))进行仿真验证。

$$\begin{aligned} v_1&=20t+N \\ v_2&=2t+N \\ v_3&=N \end{aligned} \tag{7}$$

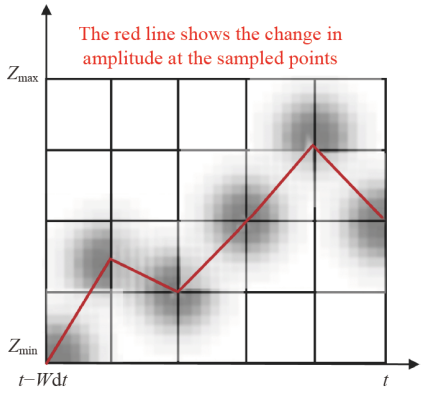


图2 $(t-W\mathrm{d}t,t)$时间区间内每个信号点的概率云图

Fig.2 Probability cloud diagram of each signal point within the time interval $(t-W\mathrm{d}t,t)$

N为(−0.2,0.2)范围内的高斯噪声。若基于信息熵的动态特征辨识算法能够对v_1、v_2、v_3进行有效区分,说明该算法具有对信号幅值变化率的动态特征辨识能力。v_1、v_2、v_3的信号曲线以及对应的概率云空间如图3所示。

设图3右侧每个概率云图中的像素点个数为$m\times n$,$P(j,k)$为每个像素点对应的概率密度,$En(i)$为第i个信号窗内信号段的信息熵,则有

$$En(i)=-\sum_{j=1}^{m}\sum_{k=1}^{n}P(j,k)\log P(j,k) \tag{8}$$

以图3右侧3个窗内信号点概率密度灰度图作为算例样本,利用式(8)计算各自的信息熵,结果如表1所示。可以看出,窗内信息熵与其相对应的信号曲线斜率呈正相关。

1.3 滚动动态辨识算法

为实现对输入到运算器内采样信号信息熵的实时计算,提出如下滚动动态辨识算法,具体计算流程如图4所示。滚动动态辨识算法步骤如下。

步骤1:在t时刻,以$(t-W\mathrm{d}t,t)$作为动态辨识区间进行采样数据提取;

步骤2:基于动态辨识区间内的数据生成概率云空间;

步骤3:基于式(8)计算该区间内的信息熵,并将其作为t时刻下的采样信号幅值变化率的辨识结果;

步骤4:$t=t+\mathrm{d}t$时重复上述步骤。

2 基于信号动态辨识的自适应跟踪微分器

基于第1节提出的信号动态特征辨识方法,设计

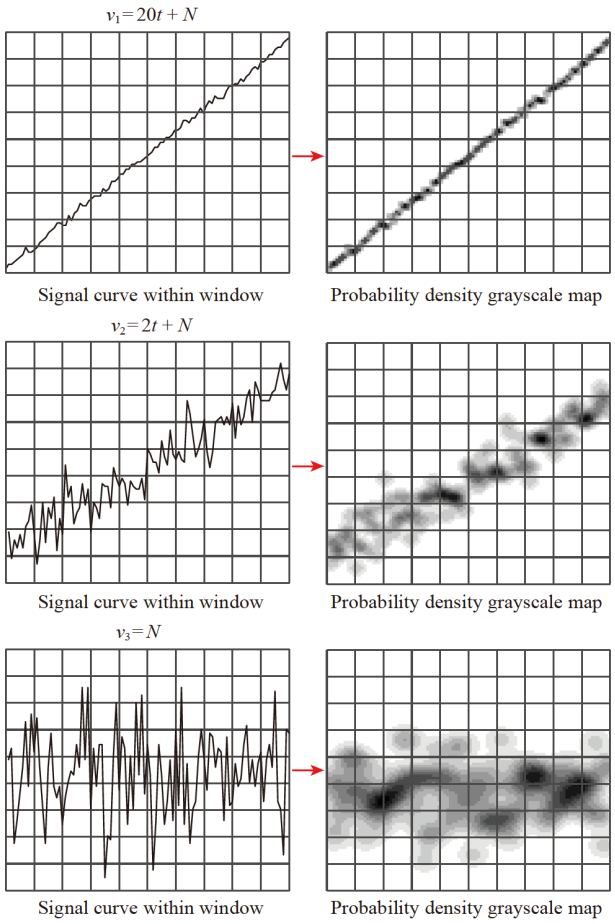


图3 窗内信号点到概率分布平面的变换示意

Fig.3 Illustration of transformation from signal points within the window to probability distribution plane

表1 信息熵计算结果

Table 1 Information entropy calculation results

信号表达式	信息熵
$v_1=20t+N$	154.75
$v_2=2t+N$	306.91
$v_3=N$	370.36

一种积分步长参数自适应调整的跟踪微分器(Tracking differentiator,TD)(以下简称自适应TD)算法。

TD是基于时间最优控制算法的非线性微分器。该微分器在结构上是由一个二阶积分串联型系统和一个时间最优控制函数(fhan)组成的负反馈非线性控制系统[21],如图5所示。图5中的算法可通过式(9)表示。

$$\begin{cases}u(k)=\mathrm{fhan}(x_1(k),x_2(k),v(k),r,h)\\x_1(k+1)=x_1(k)-v(k)+hx_2(k)\\x_2(k+1)=x_2(k)+hu(k)\end{cases} \tag{9}$$

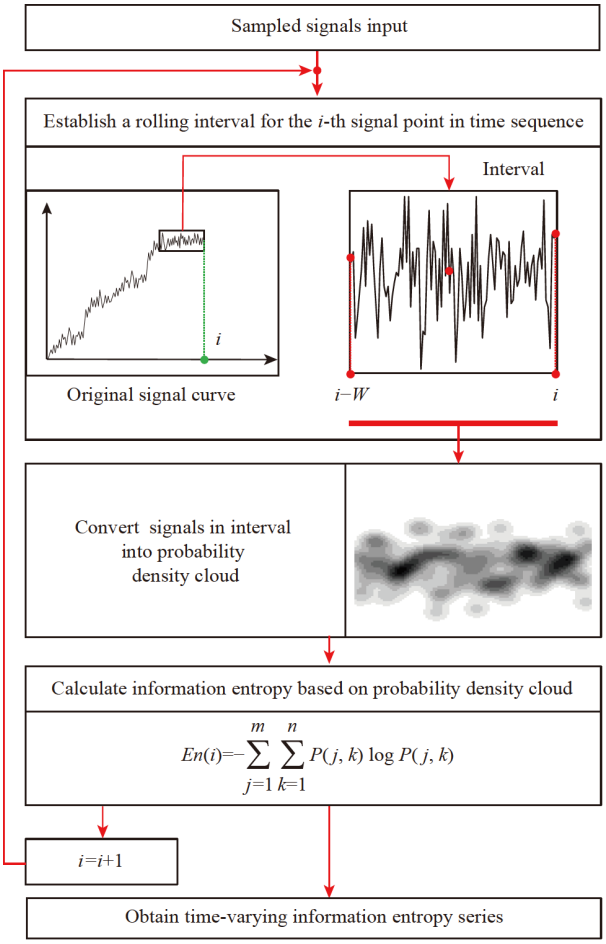


图4 动态信息计算过程

Fig.4 Calculation process of dynamic information

式中:u为控制输入,通过fhan函数计算获取;x_1为微分器的跟踪信号输出;x_2为微分器的微分信号输出;v为输入的试验数据;r为u的约束,$u\leqslant r$;h为积分步长;k为当前时间步的序号。

fhan函数如式(10)所示:

$$\begin{cases} u=\text{fhan}(x_1,x_2,r,h,v) \\ d=rh;d_0=hd \\ y=x_1-v+hx_2 \\ a_0=\sqrt{d^2+8r|y|} \\ a=\begin{cases} x_2+\dfrac{(a_0-d)}{2}\text{sign}(y), & |y|>h^2r \\ x_2+y/h, & |y|\leqslant h^2r \end{cases} \\ \text{fhan}=-\begin{cases} r\text{sign}(a), & |a|>d \\ r\dfrac{a}{d}, & |a|\leqslant d \end{cases} \end{cases} \tag{10}$$

TD的设计目标包括:①稳态下输出信号尽可能光滑;②过渡态下输出信号相位延迟尽可能小。

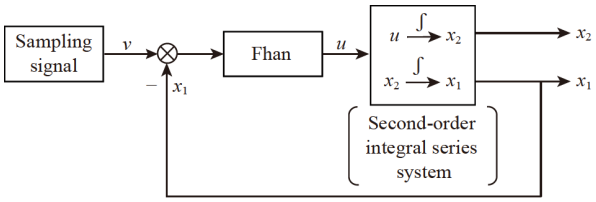


图5 基于时间最优算法的跟踪微分器结构

Fig.5 Tracking differentiator structure based on time-optimal algorithm

TD的运行原理决定了当信号采样频率一定时,若TD的运算周期及积分步长足够小,则TD的跟踪以及微分提取性能会随之增强。然而在实际系统中,控制算法的整体复杂度较高,导致TD的运算周期受限,进而限制了TD光滑、小相位滞后的信号提取能力。为此,引入基于信息熵计算的信号幅值变化率动态特征辨识方法,通过自适应函数在稳态时为TD设置较短的积分步长,在过渡态时为TD设置较长的积分步长,最终实现在控制器运算周期受限情况下的信号跟踪滤波,且兼顾稳态下的输出平滑性以及过渡态下的相位延迟抑制。

经数学推导及仿真验证发现,减小TD的积分步长可以有效提升TD输出的信噪比,但会产生一定的输出滞后。图6给出了采样信号$v(t)=10\sin 0.5t+N$(采样间隔为0.020 s,N为(−0.5,0.5)范围的随机噪声),积分步长分别取0.020 s和0.004 s时采样信号与跟踪信号的对比曲线。可以看出,积分步长0.020 s的跟踪曲线较积分步长0.004 s更为平滑,但存在一定的时间滞后。

为将可变积分步长引入实时TD系统,对积分步长与TD性能之间的关系展开研究。积分步长改变对TD的影响机制主要体现为以下方面。

对于二阶积分串联型系统(式(9)),当积分步长缩短至原步长的$1/K_t$(K_t为积分步长压缩系数,其定义式见式(11))时,TD接收到的采样信号自身的角频率等效提升为原来的K_t倍。

$$K_t=\frac{\mathrm{d}t}{h} \tag{11}$$

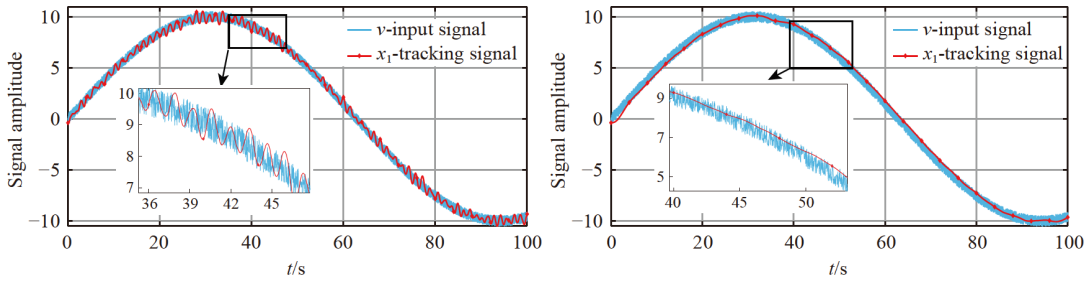
考虑式(12)所示正弦离散信号,

$$v(i)=A\sin(i\mathrm{d}t+\theta) \tag{12}$$

并将式(11)代入式(12),则有

$$v(i)=A\sin(K_tih+\theta) \tag{13}$$

式中:A为信号幅值;θ为信号初始相位。



(a) h=0.020 s (b) h=0.004 s

图6 采样信号与跟踪信号对比曲线

Fig.6 Comparative curves between sample and tracking signals

式(12)、式(13)的信号在相同序列上数值相同,但其各自形式在时域上的物理意义不同。将式(12)、式(13)分别转化为时域上的连续信号,则有

$$\begin{aligned} v(i) &= A\sin(i\mathrm{d}t+\theta) \xrightarrow{\text{convert}} \\ v(t) &= A\sin(\omega t+\theta) \end{aligned} \tag{14}$$

$$\begin{aligned} v(i) &= A\sin(K_t ih+\theta) \xrightarrow{\text{convert}} \\ v(t) &= A\sin(K_t\omega t+\theta) \end{aligned} \tag{15}$$

式中:ω为角频率。

用图像表示式(14)、式(15)所示的转化(以$K_t=2$为例),如图7所示。图中,矩形代表单独的离散信号点,矩形的横向长度表示采样间隔。可以看出,当K_t=2时,离散信号的采样频率和信号自身的角频率均上升为步长压缩前的2倍。基于以上结果可推导出:人为改变TD运算时的积分步长等效于改变TD所处理的采样信号的角频率和采样频率。

由于控制输入的约束在运算中是恒定的,因此当处理角频率更高的采样信号时会不可避免地带来更大的相位误差。且原采样信号的角频率越高,受相位滞后的影响越严重。该现象产生的原理如下:

考虑正弦信号

$$v(t)=A\sin(K_t\omega t) \tag{16}$$

假设该信号的跟踪信号y存在t_d的时间滞后,则有

$$\begin{aligned} y(t) &= A\sin(K_t\omega(t-t_\mathrm{d})) = \\ & A\sin(K_t\omega t-K_t\omega t_\mathrm{d}) \end{aligned} \tag{17}$$

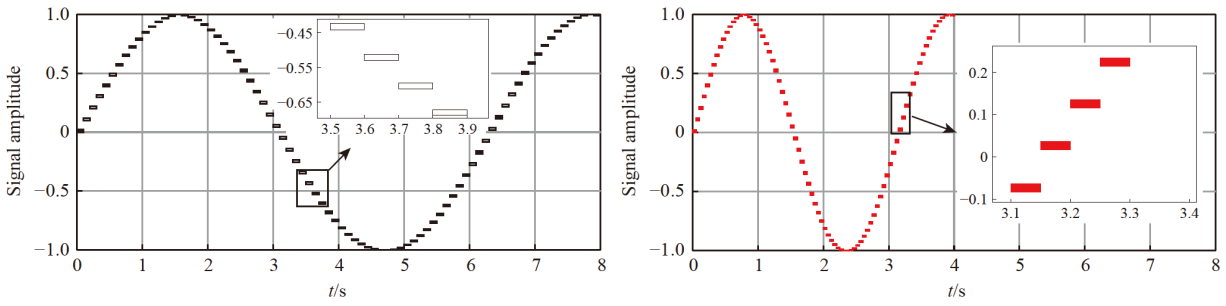
从式(17)可以看出,对于固定的时间滞后,原采样信号角频率越高则相位滞后越大。引入变积分步长后,该相位滞后将被放大K_t倍,进而使相位滞后超出工程可接受范围的风险显著提升。

在采样信号幅值变化率较低时适当提高K_t,在幅值变化率较高时适当降低K_t,可实现信号稳态时的噪声抑制以及过渡态时的TD信号输出滞后削减。该思路实现的关键在于,对采样信号幅值变化率进行实时辨识。基于信息熵的信号动态特征辨识方法与变步长TD相结合,提出一种积分步长自适应的新型TD算法,其运行框架如图8所示。图中,Adaptive fhan(自适应fhan)的形式如下:

$$\begin{cases} h=\dfrac{\mathrm{d}t}{K_t} \\ d=rh;d_0=hd \\ y=x_1-v+hx_2 \\ a_0=\sqrt{d^2+8r|y|} \\ a=\begin{cases} x_2+\dfrac{(a_0-d)}{2}\mathrm{sign}(y), & |y|>h^2r \\ x_2+y/h, & |y|\leqslant h^2r \end{cases} \\ \mathrm{fhan}=-\begin{cases} r\,\mathrm{sign}(a), & |a|>d \\ r\dfrac{a}{d}, & |a|\leqslant d \end{cases} \end{cases} \tag{18}$$

式(18)将h由固定值改为了随K_t变化的自适应参数,而K_t在本框架中以En为输入,通过自适应算法$K_t(En)$求得。

$$K_t(En)=\begin{cases} K_{t,\min}, En<En_{\min} \\ K_{t,\max}, En>En_{\max} \\ (En-En_{\min})\cdot\dfrac{(K_{t,\max}-K_{t,\min})}{(En_{\max}-En_{\min})}+K_{t,\min}, En_{\min}\leqslant En\leqslant En_{\max} \end{cases} \tag{19}$$



(a) 步长压缩前　　(b) 步长压缩后 (K_t=2)

图7 采样信号

Fig.7 Sample signal

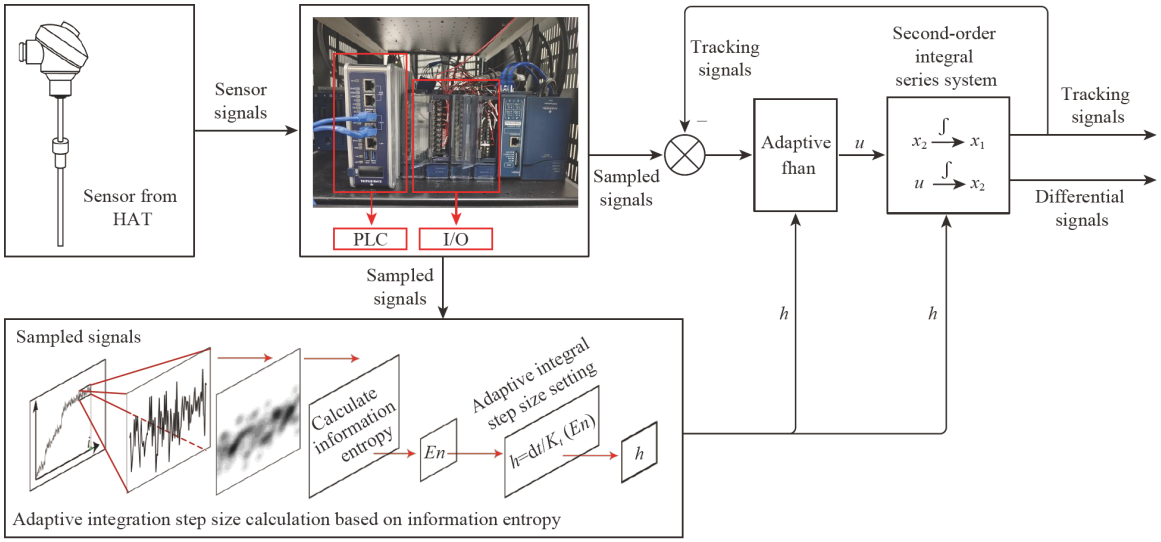


图8 自适应跟踪微分器运行框架

Fig.8 Adaptive TD operation framework

式中:$K_{t,\min}$、$K_{t,\max}$分别为K_t的最小值和最大值($K_{t,\min}\geqslant 1$);$En_{\min}$、$En_{\max}$分别表示En的人为设定上界与下界。

$En_{\min}$和$En_{\max}$决定了算法对试验数据稳态与最大过渡态的识别边界。即,若对试验数据辨识得到的信息熵大于$En_{\max}$,则将该数据视为稳态;若对试验数据辨识得到的信息熵小于$En_{\min}$,则将该数据视为最大过渡态。基于对FESS试验数据的广泛统计以及信息熵的计算,得出试验数据辨识得到的信息熵基本处于(200,400)区间。经仿真验证,$En_{\min}=250$,$En_{\max}=350$能够较好地兼顾TD对稳态数据的去噪性能以及对过渡态数据的抗滞后性能。

3 动态特征辨识Apriori关联性计算算法

Apriori是一种经典的数据挖掘算法,被用于挖掘数据集中的频繁项集(Frequent itemset)和关联规则[22]。在面对种类繁多的大量数据时,该算法可高效计算出不同种类数据间的关联性,并根据设定需求将具备共同特征的数据种类规划为统一子集。该算法在运算时,通过循环迭代不断地计算不同种类数据之间的关联性,并将相互关联性高于设定值的数据种类归为频繁项集,频繁项集的阶数表示该项集内的数据种类数,当无法寻找到阶数更高的频繁项集时计算结束,如图9所示。

对于FESS的试验信号集合,Apriori算法可以实现:①从大量不同种类的试验数据中将关联性强的数据种类各自分组;②对任意2种试验数据进行关联性量化。以上2种功能的意义在于,Apriori算法可以高效地对大规模的试验数据进行批量处理,按动态特征进行分类、分析、量化,进而显著提升样本数据准备工作效率。

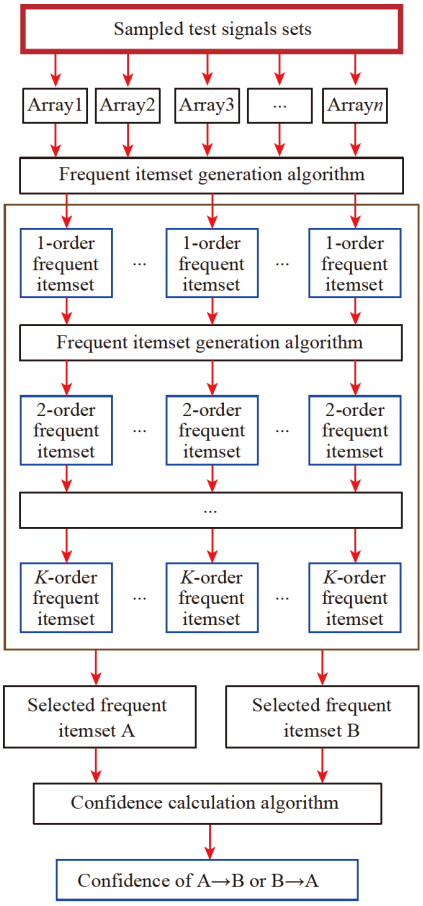


图9 Apriori算法运行流程

Fig.9 Operational flow of Apriori algorithms

无法直接使用Apriori算法对FESS的试验数据进行关联性分析。这是因为该算法要求样本数据满足2个条件:①样本数据能够表征目标特性,且可被量化和计算;②样本数据格式为一系列事件的“发生与否”,即通常为仅包含“1、0”的布尔值。为此,需要引入相应的方法将FESS的试验数据转化为Apriori算法可以直接识别的格式。

针对上述问题,提出动态特征辨识Apriori关联性算法。首先引入基于信息熵的辨识方法,将试验数据转化为能够反映数据幅值变化率的特征值;随后通过中间算法将该特征值转化为布尔表;最后将该布尔表作为输入,使用Apriori算法进行分析计算。

动态特征辨识Apriori关联性算法的运行框架如图10所示,具体步骤如下。

步骤1:载入试验数据集合(参数1~*N*的试验数据组);

步骤2:基于信息熵的数据动态特征辨识,生成各参数的*En*曲线;

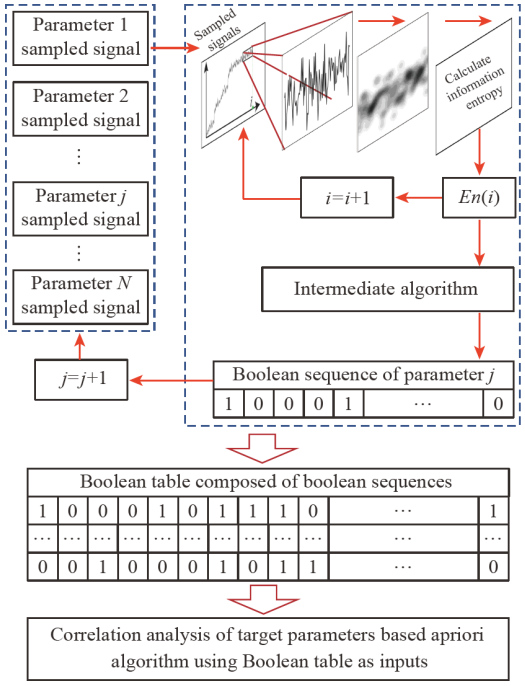


图10 信号相关计算框架

Fig.10 Framework of the signal correlation calculation

步骤3:将各*En*曲线输入中间算法,进而将其转化为布尔序列;

步骤4:整合布尔序列,形成试验数据集合对应的布尔表;

步骤5:以布尔表作为输入,通过Apriori算法对数据集合内的各参数进行关联性分析计算。

上述流程的关键在于对中间算法的设计。该算法要求在保留参数信号原动态特征的前提下,将其转化为布尔序列。为此,对所获得的*En*曲线的变化率进行分析转化,提出一种中间算法。该算法的运行流程如图11所示,具体步骤如下。

步骤1:载入*En*曲线。

步骤2:基于*En*曲线每段的幅值变化率,将该段标定为-1或1或0。

步骤3:对每条*En*曲线进行步骤1和步骤2的操作,整合得到*En*趋势表(*En* trend table)。

步骤4:将*En*趋势表中所有被标定为-1的元素设为1,其余元素设为0,令设置后的表为表A;将*En*趋势表中所有被标定为0的元素设为1,其余元素设为0,令设置后的表为表B;将*En*趋势表中所有被标定为1的元素设为1,其余元素设为0,令设置后的表为表C。将表A、B、C进行串联合并,形成布尔特征表。

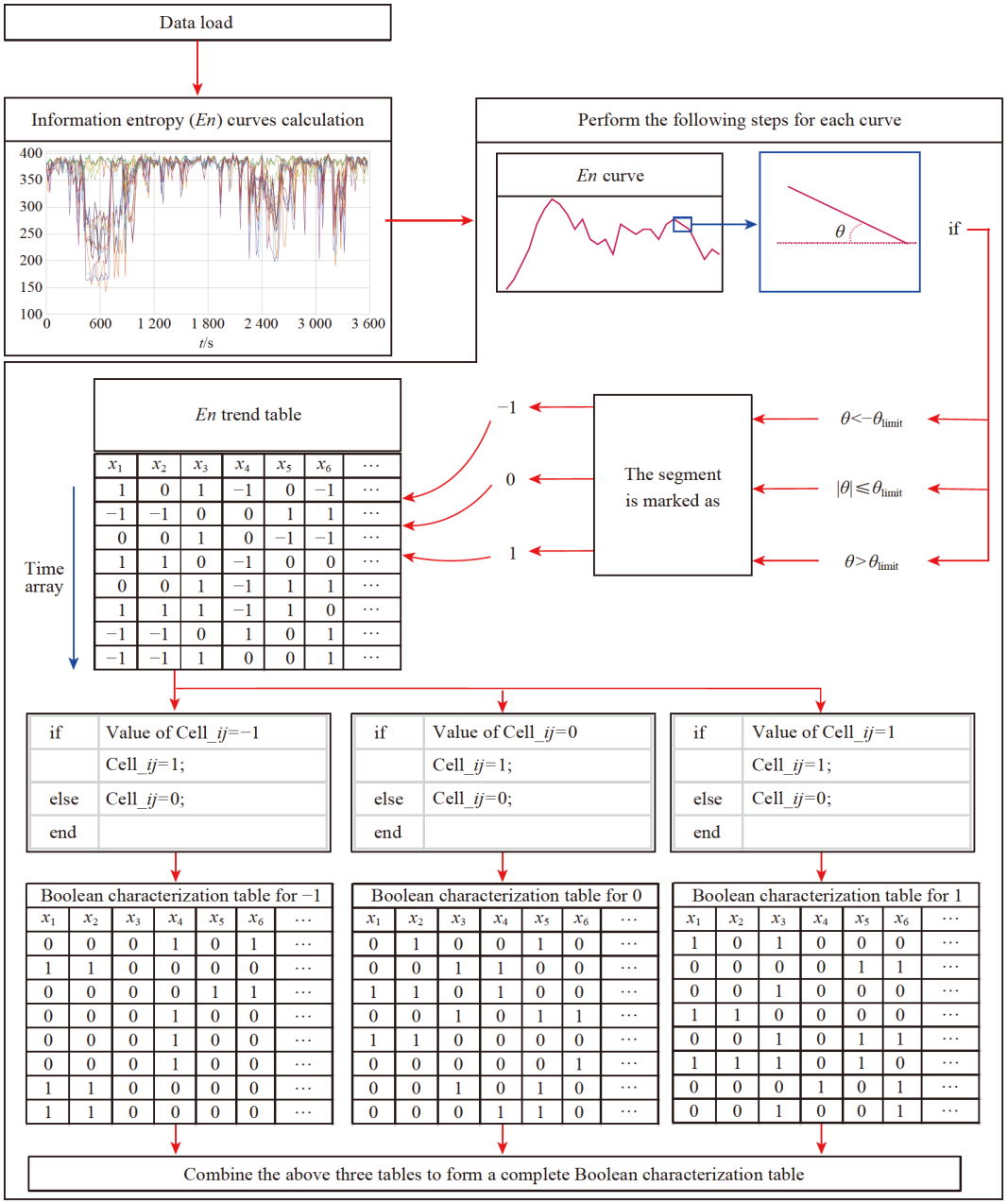


图11　中间算法运行流程

Fig.11　Operational flow of intermediate algorithm

步骤5:以布尔特征表作为输入,通过Apriori算法进行数据关联性分析计算。

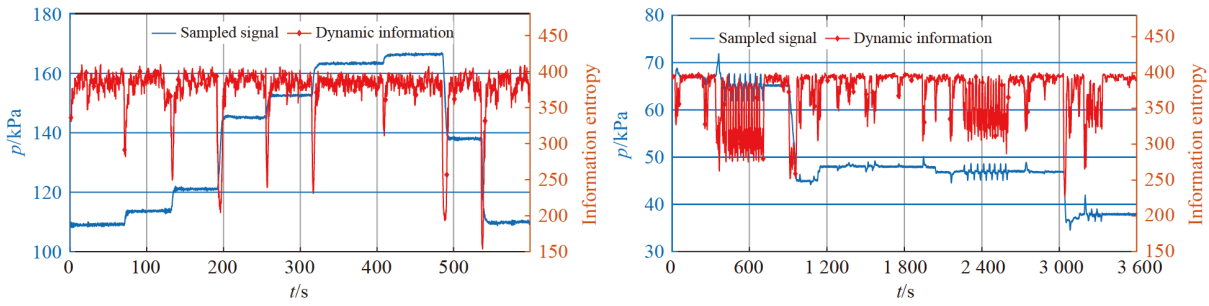
4　仿真验证

4.1　基于信息熵计算的信号动态特征辨识算法验证分析

选取高空台某2次试验中的某点位压力p数据作为算例样本,并基于第1节所述步骤进行相应的信号幅值变化率特征辨识,计算得到随样本信号同步变化的信息熵曲线,如图12所示。可以看出,当样本数据处于稳态时,信息熵幅值维持在400上下小幅波动;而当样本数据曲线发生明显幅值变化时,信息熵幅值能够做出快速减小的响应,且响应幅值与信号幅值变化率的快慢成正比。该结果表明,所提出的基于信息熵计算的数据动态特征辨识算法对过渡态信号幅值变化率具有较强的辨识能力。

4.2　基于信号动态辨识的自适应TD验证分析

分别取1组人工信号(图13,图中V_p为阀门开度)



(a) 第1次试验 (b) 第2次试验

图12 采样信号与动态信息的对比曲线

Fig.12 Comparative curve between sample signals and dynamic information

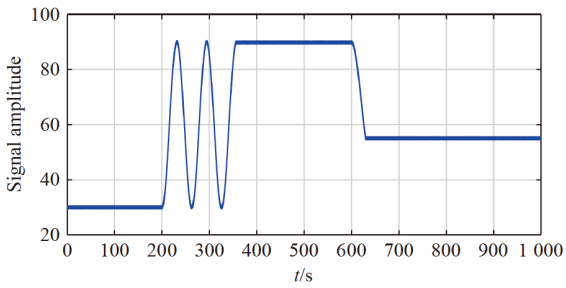


图13 人工信号

Fig.13 Artificial signals

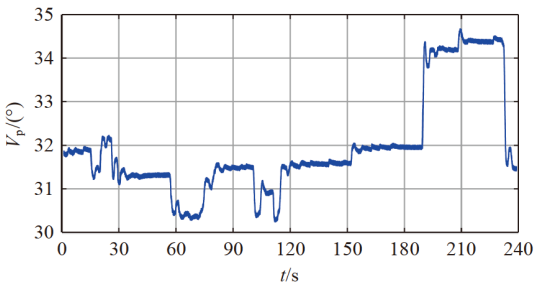


图14 试验信号(阀门开启信号)

Fig.14 Test signals (valve opening signals)

和1组实际试验信号(图14)进行算法验证。其中,人工信号在原信号的基础上添加了(−0.5,0.5)范围内的高斯噪声。2组信号的采样频率均为50 Hz。

分别基于所提出的自适应TD以及固定步长TD,对以上2组含噪声信号进行动态信息提取和信号跟踪滤波。自适应TD内自适应函数的参数设置如表2所示。2种TD的控制约束均设置为30。

表2 自适应TD内自适应函数的参数设置

Table 2 Parameters setting of the adaptive function in the adaptive tracking differentiator

信号类型	$K_{t,\max}$	$K_{t,\min}$	$En_{\max}$	$En_{\min}$
人工信号	10	1	350	150
试验信号	3	1	350	150

基于信息熵辨识与自适应算法得到的积分步长自适应选定结果如图15所示。可以看出,对于人工信号及试验信号,该算法均能根据信号动态特性准确选取合适的积分步长。

2种TD的跟踪效果如图16所示。可以看出,对于人工信号及试验信号,自适应TD在跟踪滤波方面均具有更高的信噪比,有效抑制了噪声,并且并未引入更大的滞后。

以均方误差(MSE,量符号为δ_{MSE})作为指标,进一步评价2种TD对采样信号的跟踪性能。

$$\delta_{\mathrm{MSE}}=\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2 \tag{20}$$

式中:N为信号点总数;y_i为不含噪声的原信号幅值;$\hat{y}_i$为滤波后的信号幅值。

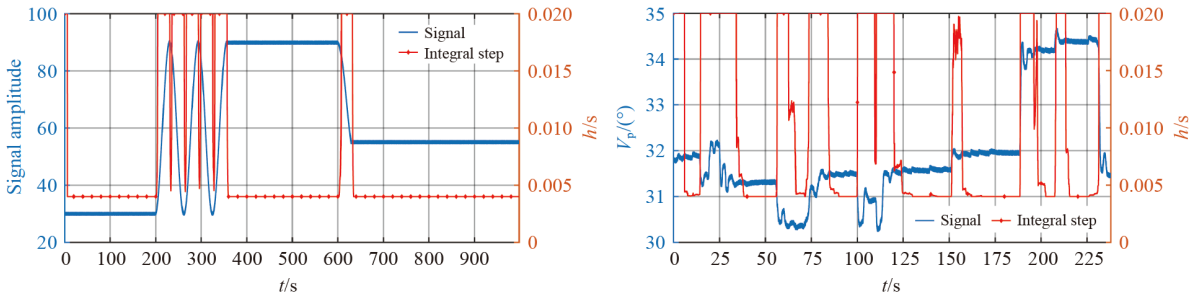
以人工信号及其仿真结果作为计算对象,结果如表3所示。可以看出,相较于固定步长TD,采用自适应TD的输出信号的MSE减小了约36.75%。

表3 MSE计算结果

Table 3 MSE calculation results

TD	MSE
自适应TD	0.071 95
固定步长TD	0.113 75

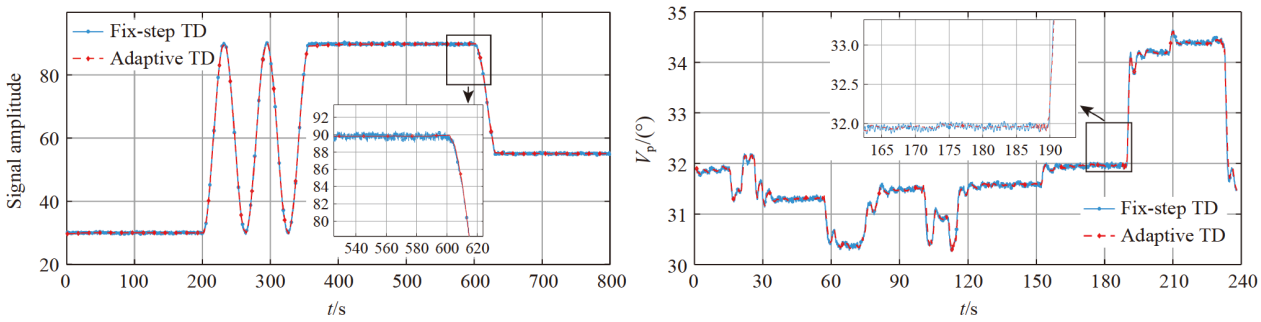
综上所述,针对较低频率的时序采样信号,所提出的自适应TD在跟踪去噪滤波方面实现了显著的



(a) 人工信号　　(b) 试验信号

图15 采样信号与积分步长对比曲线

Fig.15 Comparative curves between sample signals and integral step



(a) 人工信号　　(b) 试验信号

图16 2种TD输出对比

Fig.16 Comparison of the two TD outputs

性能优化。

4.3 基于动态特征辨识Apriori关联性计算算法验证分析

选取若干组试验数据并基于所提出的信号关联性计算算法进行分析,以验证算法对信号间关联性计算的准确程度。取试验台A某次试验下的阀门1开度$V_{p,1}$、阀门1所在腔的压力p_1,试验台B的阀门2开度$V_{p,2}$、阀门2所在腔的压力p_2以及试验台C的阀门3开度$V_{p,3}$作为试验样本,各组样本数据曲线如图17所示。

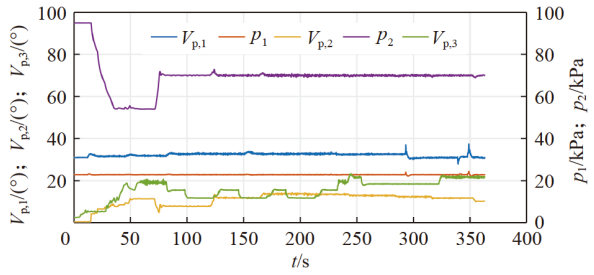


图17 样本数据曲线

Fig.17 Sample data curves

基于所提出的算法对以上5组数据进行关联性计算,结果如表4所示。表中,关联性通过一个(0,1.000)范围内的小数表示,该数越接近1.000,表示关联性越强。可以看出,$V_{p,1}$与p_1、$V_{p,2}$与p_2之间的关联性显著高于其他数据组合。在实际操作中,设定0.500为阈值,即可从总数据集合中提取存在明显关联性的数据。

表4 关联性计算结果

Table 4 Correlation calculation results

数据	$V_{p,1}$	p_1	$V_{p,2}$	p_2	$V_{p,3}$
$V_{p,1}$	1.000	0.825	0.256	0.298	0.228
p_1	0.825	1.000	0.246	0.281	0.211
$V_{p,2}$	0.256	0.246	1.000	0.666	0.328
p_2	0.298	0.281	0.666	1.000	0.351
$V_{p,3}$	0.228	0.211	0.328	0.351	1.000

5 结论

以航空发动机高空台的过渡态试验数据为研究对象,针对其规模庞大、结构复杂、耦合性强、采样频

率低、噪声来源多等特点,提出了一种基于概率云空间的信息熵计算方法,有效实现了对试验数据幅值变化率动态特征的量化辨识。该方法不依赖于系统模型,具备算法简洁、迟滞性低、鲁棒性强的特点,适用于FESS等大型复杂系统的动态特征提取。主要结论如下:

(1) 在控制技术层面,设计了一种基于动态特征辨识的自适应跟踪微分器。针对FESS多源异构噪声导致的低信噪比问题,利用动态特征辨识结果实时调节跟踪微分器积分步长,实现了滤波器参数的自适应优化。设计的自适应跟踪微分器使去噪跟踪信号的均方误差较固定步长跟踪微分器减小了约36.75%,显著提升了信号预处理品质,为后续自抗扰控制、鲁棒自适应控制等高精度算法在强噪声、非线性工况下的工程化应用奠定了坚实基础。

(2) 在建模技术层面,开发了一套融合动态熵权值与Apriori算法的数据关联性挖掘框架。该框架能够量化评估多变量数据间的关联性,自动识别容腔压力-调节阀开度等关键耦合关系,为建立高精度的数据驱动降阶模型或故障传播模型提供了可靠的特征筛选依据。该框架有效降低了传统建模对经验的依赖,为FESS的智能化建模分析提供了新范式。

(3) 形成了一套面向高空台FESS的"动态特征辨识-信号自适应处理-数据特征关联挖掘"的联合解决方案,提升了对复杂系统试验数据的利用深度与智能处理水平,也为高空台数字化、智能化升级提供了高效可行的技术路径。

参考文献:

[1] DAVIS M,MONTGOMERY P. A flight simulation vision for aeropropulsion altitude ground test facilities [J]. Journal of Engineering for Gas Turbines and Power,2005,127(1):8-17.

[2] MONTGOMERY P A,BURDETTE R,KRUPP B. A real-time turbine engine facility model and simulation for test operations modernization and integration[R]. ASME 2000-GT-0576,2000.

[3] MONTGOMERY P A,BURDETTE R,WILHITE L,et al. Modernization of a turbine engine test facilityt uilizing a real-time facilityt model and simulation[R]. ASME 2001-GT-0573,2001.

[4] 朱美印,王 曦,裴希同,等. 飞行环境模拟系统多容腔流-固传热建模[J]. 推进技术,2020,41(12):2848-2859.
ZHU Meiyin,WANG Xi,PEI Xitong,et al. Multi-volume fluid-solid heat transfer modeling for flight environment simulation system[J]. Journal of Propulsion Technology,2020,41(12):2848-2859. (in Chinese)

[5] 裴希同,张 松,但志宏,等. 高空台飞行环境模拟系统数字建模与仿真研究[J]. 推进技术,2019,40(5):1144-1152.
PEI Xitong,ZHANG Song,DAN Zhihong,et al. Study on digital modeling and simulation of altitude test facility flight environment simulation system[J]. Journal of Propulsion Technology,2019,40(5):1144-1152. (in Chinese)

[6] 但志宏,张 松,白克强,等. 基于扩张状态观测器的高空台进气环境模拟控制技术研究[J]. 推进技术,2021,42(9):2119-2128.
DAN Zhihong,ZHANG Song,BAI Keqiang,et al. Air intake environment simulation of altitude test facility control based on extended state observer[J]. Journal of Propulsion Technology,2021,42(9):2119-2128. (in Chinese)

[7] DAVIS M,MONTGOMERY P. A flight simulation vision for aeropropulsion altitude ground test failties [R]. ASME GT2002-30003,2002.

[8] PACHLHOFER P M,PANEK J W,DICKI D J,et al. Advances in engine test capabilitis at the NASA glenn research center's propulsion system laboratory[R]. ASME GT2006-90181,2006.

[9] 张楼悦,朱美印,王 曦,等. 基于有限试验数据与流场仿真的套筒阀流量特性获取方法[J]. 测控技术,2022,41(2):41-48.
ZHANG Louyue,ZHU Meiyin,WANG Xi,et al. Method for obtaining flow characteristics of sleeve valve based on limited test data and flow field simulation[J]. Measurement & Control Technology,2022,41(2):41-48. (in Chinese)

[10] 朱美印,张 松,但志宏,等. 一种大口径蝶阀流量特性的坐标定位回归算法[J]. 燃气涡轮试验与研究,2017,30(4):39-44.
ZHU Meiyin,ZHANG Song,DAN Zhihong,et al. A coordinate positioning and regression algorithm for the flow characteristics of a large butterfly valve[J]. Gas Turbine Experiment and Research,2017,30(4):39-44. (in Chinese)

[11] PEI Xitong,WANG Xi,LIU Jiashuai,et al. A review of modeling simulation and control technologies of altitude ground test facilities for control application[J]. Chinese

Journal of Aeronautics,2023,36(9):38−62.

[12] 缪柯强,王　曦,朱美印,等. 一种高空台特种调节阀通用特性修正方法[J]. 燃气涡轮试验与研究,2019,32(4):22−26.
MIAO Keqiang,WANG Xi,ZHU Meiyin,et al. A method on general characteristic map correction for special regulating valve of the simulated altitude test facilities[J]. Gas Turbine Experiment and Research,2019,32(4):22−26. (in Chinese)

[13] ZHU Meiyin,WANG Xi,PEI Xitong,et al. Modified robust optimal adaptive control for flight environment simulation system with heat transfer uncertainty[J]. Chinese Journal of Aeronautics,2021,34(2):420–431.

[14] 田金虎,但志宏,张　松,等. 高空台环境模拟控制技术[J]. 航空动力,2021,(3):64−68.
TIAN Jinhu,DAN Zhihong,ZHANG Song,et al. Environment simulation control technology for altitude test facility[J]. Aerospace Power,2021,(3):64−68. (in Chinese)

[15] LIU Junyan,KUMAR S,PALOMAR D P. Parameter estimation of heavy-tailed AR model with missing data via stochastic EM[J]. IEEE Transactions on Signal Processing,2019,67(8):2159−2172.

[16] TAVAKOLI N,SIAMI-NAMINI S,ADL KHANGHAH M,et al. An autoencoder-based deep learning approach for clustering time series data[J]. SN Applied Sciences,2020,2(5):937.

[17] ROSTAGHI M,AZAMI H. Dispersion entropy: a measure for time-series analysis[J]. IEEE Signal Processing Letters,2016,23(5):610−614.

[18] ZOU Yong,DONNER R V,MARWAN N,et al. Complex network approaches to nonlinear time series analysis[J]. Physics Reports,2019,787:1−97.

[19] KEIL A,BERNAT E M,COHEN M X,et al. Recommendations and publication guidelines for studies using frequency domain and time-frequency domain analyses of neural time series[J]. Psychophysiology,2022,59(5):e14052.

[20] SHANNON C E. A mathematical theory of communication[J]. ACM SIGMOBILE Mobile Computing and Communications Review,2001,5(1):3−55.

[21] 韩京清. 自抗扰控制技术[J]. 前沿科学,2007,(1):24−31.
HAN Jingqing. Auto disturbances rejection control technique[J]. Frontier Science,2007,(1):24−31. (in Chinese)

[22] AGRAWAL R,IMIELIŃSKI T,SWAMI A. Mining association rules between sets of items in large databases[J]. ACM SIGMOD Record,1993,22(2):207−216.