

文章编号: 1001 - 2486(2003)04- 0079 - 03

基于符号树信息熵的机械振动瞬态信号特征提取*

张 雨¹, 胡 芑庆²

(1. 长沙理工大学, 湖南 长沙 410073; 2. 国防科技大学, 湖南 长沙 410073)

摘 要: 用简单的二进制划分简要说明了信号符号化问题, 给出了二进制划分下符号树结构及其 Shannon 熵计算方法, 叙述了符号序列编码方法和符号序列直方图的作用, 提交了符号时间序列分析的计算机流程图, 对内燃机机身振动信号和转轴振动信号用符号树信息熵进行了特征提取。

关键词: 符号; 时间序列; 熵; 信号; 分析
中图分类号: O235; TH17 **文献标识码:** A

Extraction of the Characteristic of Mechanism Vibration
Transient Signal Based on Entropy of Symbolic Tree

ZHANG Yu¹, HU Niao-qing²

(1. Changsha Univ. of Science and Technology, Changsha 410076, China;
2. National Univ. of Defense Technology, Changsha 410073, China)

Abstract: The question how to translate a time series into a symbolic series is briefly answered by the simplest binary partition. The frame of symbolic tree and calculation method of Shannon entropy are presented depending on the binary partition. The coding method and histogram function of symbolic series are described. The calculation flow chart for symbolic time series analysis is referred to. The signal characteristics, which are from the vibration of both IC engine block and mechanical bearing, are extracted by the measurement of entropy of the symbolic tree.

Key words: symbolic; time series; entropy; signal; analysis

1 符号化方法简要说明^[1]

引入划分 $P = \{P_1, P_2, \dots, P_q\}$, 并把系统状态空间划分为 $m = (q + 1)^d$ 个单元, 其中 d 是状态空间的维数, q 是划分个数, $q = 1$ 为最简单的二进制划分。每个单元用一个符号 $S_r \in \{S_1, S_2, \dots, S_m\}$ 进行标记。从而, 在相空间为连续曲线的系统的轨迹将被转换为符号序列。这些符号标记了被轨迹所访问的单元。这一符号模型以最简单的方式完全地描述了系统的动力学特征, 它提供了一种研究复杂动态系统运行状态的手段。

对时间序列进行符号统计量方法分析时, 通过引入相空间离散划分并分配符号给每个划分单元, 将模拟时间序列转换为符号序列。一旦建立了符号序列, 就可以对表示不同符号序列出现概率的符号统计量进行计算。最简单的方法是通过一个阈值函数, 模拟信号序列 $x_n \equiv x(t_n)$ 转换为:

$$\{x_n\} = (x_1, x_2, \dots, x_N) \mapsto \{s_n\} = (s_1, s_2, \dots, s_N) \tag{1}$$

其中 $s_n \in (0, 1)$ 。图 1 所示为在 $q = 1$ 的二进制划分下, 时间序列转化为符号序列的过程。

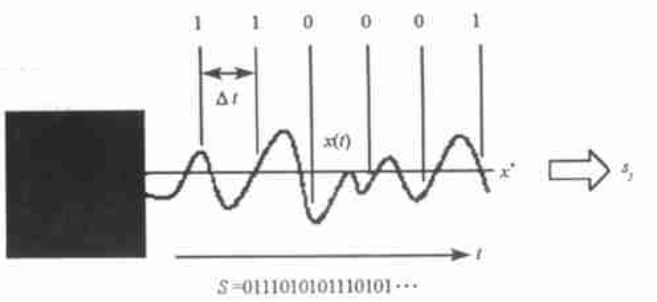


图 1 对时间序列 $x(t)$ 与其符号化序列 s_j
Fig. 1 Time series $x(t)$ and its symbolic series s_j

* 收稿日期: 2003 - 01 - 05
基金项目: 国家自然科学基金资助项目(50176006)
作者简介: 张雨(1958—), 男, 教授, 博士后。

2 符号序列编码、符号序列直方图

对于每个符号序列,更为方便的数值表达是把基 n 的序列变换为基 10 的十进制数序列,这一过程称作序列编码过程。例如,对于基 2 的序列 000,其编码为 0,001 为 1,010 为 2,011 为 3,等等。一般地,选择了树层 L 和时延 τ 后,把序列 $s(n)$ 分为短序列组:

$$s(k) = [s(k), s(k + \tau), \dots, s(k + (d - 1)\tau)] \tag{2}$$

其中 $k = 1, 2, \dots, N - (d - 1)\tau$ 。为方便起见,符号序列 $s(k)$ 可以转换为十进制数序列:

$$S(k) = \sum_{i=1}^d (q + 1)^{d-i} s_i(k) \tag{3}$$

其中 $s_i(k) = s(k + (i - 1)\tau)$ 。符号序列经编码后,可把它们的统计量表示为直方图,称为符号序列直方图。

3 符号树图解及其信息熵

符号树是符号统计量的图形表示,后者是符号序列长度 L 的函数。在获得符号树后可以方便地提取符号序列统计量。图 2 是一棵三层的符号树,其中 p_{001} 是观测符号序列 001 的概率。显然,树中每一层对应着符号序列的特定长度。符号序列长度 $L = 3$,表示 3 个连续符号出现的情况。 L 也称为树的层数,因为它表达了用以评估符号统计量的符号树的深度。在树的第 L 层有 $(q + 1)^L$ 种不同的符号序列。

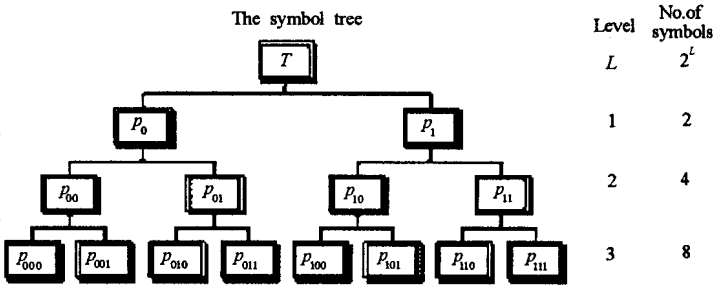


图 2 一棵三层符号树的符号序列长度,符号序列概率

Fig. 2 The length of symbolic series in a three-layer tree and the frequency of symbolic series

可用 Shannon 熵来描述符号统计量。对于二进制划分,在树的第 L 层可定义 Shannon 熵:

$$H(p(\tau, L)) = - \frac{1}{L \ln(q + 1)} \sum p_{s_1 s_2 \dots s_L} \ln(p_{s_1 s_2 \dots s_L}) \tag{4}$$

式(4)中变元 p 表示树的个数,是时延 τ 和树层数 L 的函数。所有的熵都隐式地依赖于划分的选择。

对于混沌和紊流数据,相关性随着时延 τ 的增加而衰减且其 Shannon 熵渐近达到上限,即:

$$\lim_{\tau \rightarrow \infty} H(p(\tau, L)) = H^\infty(p(L)) \tag{5}$$

由符号时间序列分析得到的 Shannon 熵可以对复杂系统状态进行评估。因为从几何观点看,感兴趣信号实质上是动力学系统相空间中的轨迹,不论它是混沌的还是非混沌的,它必定是确定性的。如果使用生成划分把相空间符号化,并将轨迹转换为符号序列,那么特定符号串将更为频繁地出现。而对于纯粹的随机噪声,各种符号出现的可能性在统计意义上相等。因此,这一特性可以用来检测时间序列是否包含了确定性的结构,或者在总体上是否是随机的。

4 基于符号树信息熵的瞬态信号特征提取

4.1 提取内燃机机身振动信号特征

高速高功率柴油机活塞环胶结的故障并不鲜见,它使得整机的动力性下降。活塞环胶结故障在机身振动信号中并无明显的故障特征和差别。经机理分析可以证明当活塞环发生胶结时^[2],由于第 I 道活塞环的迷宫密封作用丧失,作用在第 II、III 道活塞环上的激励力(气体力和油膜力)增强,使得活塞环的径向振动加剧,而且第 I 道活塞环胶结较之第 I 和 II 道活塞环同时胶结时机身振动更甚。

表 1 给出了某 4135G 型柴油机第四缸活塞环处于正常、第 I 道活塞环胶结、第 I 和 II 道活塞环同时胶结的三种状态时,机身振动加速度实测波形(工况为 1500r/min、75% 负荷,采样频率 6.4kHz)。

采用二进制静态符号化方法并进行十进制编码后,得到其符号时间序列直方图亦如表 1 所示,同时

给出了对应的 Shannon 熵值。三种状态的 Shannon 熵值呈现“大 一小 一大”的“下凹”情况, 在第 I 道活塞环胶结时有极值点, 这与文献[2]的机理分析结果是一致的。

4.2 提取转轴振动信号特征

航空涡轮发动机和工业透平机中转子叶轮一定子间触碰故障时有发生, 其诱发原因是多方面的, 与叶轮裂纹、受热不均衡、叶片加工质量等因素有关。发生触碰故障的严重性不言而喻, 问题在于对其监测相当困难性, 因为转子位于定子厚壳体内部, 转子叶片一定子间最初的触碰程度又极其轻微, 其症状无论是通过定子的振动强度或是声音信号都难以辨识, 而一旦触碰发生, 其故障的扩展过程非常迅速。实验模型的转子内置于支承, 工作转速 3240r/min, 空负荷, 采样频率 200Hz, 试验中使转子一定子间处于正常、早期尖锐型触碰、中期半尖锐型触碰、晚期平钝型触碰四种状态^[3]。表 2 给出了转子 $x-y$ 两向的时域振动位移轨迹、其垂直方向振动位移信号的符号序列直方图以及对应的 Shannon 熵值。可见早期尖锐型触碰较之其他状态, 其熵值有明显的差异, 其他各类工况的熵值几乎相同。

表 1 某 4 缸柴油机的机身振动加速度信号、符号时间序列直方图及其 Shannon 熵

Tab. 1 The vibration acceleration, symbolic time series histogram and its Shannon entropy of a 4 cylinders IC engine block




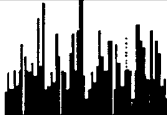
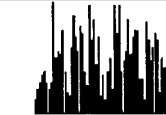
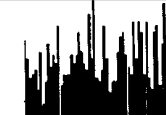




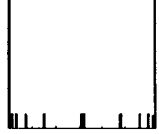
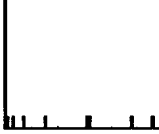
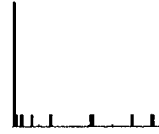
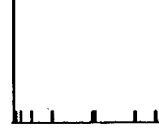
柴油机工况	正常	第 I 道活塞环胶结	第 I、II 道活塞环同时胶结
机身振动加速度信号			
符号时间序列直方图			
Shannon 熵	0.9754 ↑	0.9692 ↓	0.9795 ↑

表 2 某模型转子的振动位移信号、符号时间序列直方图及其 Shannon 熵

Tab. 2 The vibration displacement, symbolic time series histogram and its Shannon entropy of a model rotor shaft

转子一定子工况	正常	早期尖锐型触碰	中期半尖锐型触碰	晚期平钝型触碰
转子 $x-y$ 两向的时域振动位移轨迹				
y 方向(垂直)振动位移信号的符号序列直方图				
Shannon 熵	0.7429 ↓	0.7552 ↑	0.7424 ↓	0.7429 ↓

5 小结

- (1) 信号的符号标记以最简单的方式提供了一种研究复杂动态系统运行状态的手段。
- (2) 由符号序列直方图得到的 Shannon 熵是对瞬态信号整体形态综合特征的一种有效度量。
- (3) 对活塞环胶结时机身振动信号和转子一定子发生触碰时转子振动信号进行符号树信息熵分析, 结果证明了上述结论, 即可以用符号树信息熵来区分判断设备运行状态的变化。
- (4) 符号序列直方图中具体序列码的状态内容及其对应的概率值所表征的含义还有待进一步破解。

参考文献:

[1] Lin. Use of Symbolic Time Series Analysis for Stall Precursor Detection[C]. AIAA 1998-3310.
[2] 张雨. 内燃机主运动副结构响应与运行工况的研究[D]. 武汉交通科技大学, 1997.
[3] Zhang Yu. Analysis and Identification on Fault of Rub-impact between Rotor and Stator[J]. Journal of Southeast University, 2000(2).