DOI:10.16450/j.cnki.issn.1004-6801.2021.01.024

一种振测数据最佳分析长度的确定方法

张建伟^{1,2,3}. 李 洋^{1,2,3}. 马晓君^{1,2,3}. 程梦然^{1,2,3}

(1.华北水利水电大学水利学院 郑州,450046) (2.水资源高效利用与保障工程河南省协同创新中心 郑州,450046)(3.河南省水工结构安全工程技术研究中心 郑州,450046)

摘要 数据分析长度的选取是提取结构振动特征信息的关键,人为选取信号分析的数据长度会导致其计算结果存 在一定误差,为减少主观因素影响,提出一种基于改进多尺度排列熵(improved multi-scale permutation entropy,简称 IMPE)的振测数据分析长度的方法。对于获得的振动信号,将一维的时间序列数据多尺度化后进行粗粒化处 理,确定其相空间重构参数。选取不同长度的振测数据,分别计算多尺度排列熵(multi-scale permutation entropy, 简称 MPE)熵值,发现熵值对数据长度的变化敏感,随着数据长度的增加而变化,最后趋于稳定值。定义该稳定值 为标准熵值,满足标准熵值97%精度的熵值作为有效熵值,选出满足精度要求的熵值,将其所对应的最短数据长度 定义为振测数据的最佳分析长度。将该方法应用于仿真信号和具体泄流工程振动信号的最佳数据长度选取中,可 为结构监测选取准确的数据分析长度,有较好的普适性。

关键词 坝体振动;数据分析长度;改进多尺度排列熵;信号分析;白噪声 中图分类号 TH39

引 言

根据结构动力响应变化进行结构损伤诊断或安 全监测研究,是目前工程界研究的热点问题之一。 通常,结构在损伤诊断和在线监测等过程中的振测 数据长度大多为人为选取,而振测数据的分析^[1-3]作 为检测结构振动状态的关键依据,其长度的选取尤 为重要,在一定程度上决定着结构监测的有效性。 因此,需要一个有效的数据长度选取方法来确定数 据的最佳分析长度。

陈佳袁等^[4]对小样本水文数据序列进行优化处 理,将对时间序列的分析用于水文数据的预测,可提 高流量数据的预测精度。马佳妮等^[5]利用长时间序 列遥感数据,构建了一种反演净初级生产力(net primary productivity,简称 NPP)的耕地质量评价方法, 可同时获取耕地质量空间分布和时空演变过程,减 少数据断点导致的评价误差,提高了准确性。蒋增 林等^[6]利用趋势移动平均方法和指数平滑方法,对电 力系统负荷时间序列历史数据进行处理,取得较好 的预测效果。时间序列数据的分析及应用在许多领 域都有所涉及^[79],但对数据长度的合理性分析较少。 称 PE)的基础上提出的一种检测动力学突变和时间 序列排列的方法,具有灵敏度高、抗噪能力强和鲁棒 性强等特点,可以较敏感的反应非线性、非平稳信号 时间序列的微小变化^[11]。由于该方法在检测系统 动力学突变方面较为敏感,因此成为数据分析的热 点方法,被广泛应用于生物医学^[12-14]、机械损伤诊 断^[15-16]及股票分析^[17]等领域,具有较好的工程实用 性。MPE计算振测信号数据时,数据量的大小决定 了包含信号的丰富程度。数据长度过长,存在计算时 间长等缺点;数据长度过短,会导致信号的特征信息 不完全。因此,选取合适的数据长度是保证分析结果 正确的重要环节,笔者利用MPE算法求取振测数据 的最佳分析长度,并将其应用于具体泄流工程。

1 基本原理

1.1 多尺度排列熵

MPE算法相较于PE算法具有更好的鲁棒性, 能够很好地检测信号的复杂程度^[18]。将一维时间 序列进行粗粒化处理,选取合适的尺度对其进行多 尺度化,分别计算各尺度下时间序列的PE熵值,继

^{*} 国家自然科学基金资助项目(51679091);广东省水利科技创新基金资助项目(2020-18);广州市科技计划资助项目 (2020ky34);华北水利水电大学研究生教育创新计划资助项目(YK2019-20) 收稿日期;2019-06-02;修回日期;2019-12-19

而得到MPE。

首先,令 $\{X(i); i=1, 2, ..., n\}$ 为一维时间序列, 細粒化处理可得

$$y_{j}^{(s)} = \frac{1}{s} \sum_{i=(j-1)s+1}^{js} X(i) \ (j=1,2,\cdots,[n/s]) \ (1)$$

其中:y_j^(s)为粗粒化序列;s为尺度因子,s=1表示粗 粒化序列为原始序列;[n/s]表示对n/s取整。

由式(1)可知,时间序列长度与尺度因子s的大 小成反比。令m为嵌入维数, r为延迟时间,对粗粒 化序列 y_i^(s) 重构得到重构粗粒化序列 y_i^(s) 及重构 分量 Y_i^(s)

$$Y_{l}^{(s)} = \{ y_{l}^{(s)}, y_{l+t}^{(s)}, \cdots, y_{l+(m+1)t}^{(s)} \}$$
$$(l = 1, 2, \cdots, n(m-1)t)$$
(2)

重构分量 Y^(s) 中各元素所在列的索引分别为 l₁, l₂, …, l_m, 升序排列 Y^(s), 其中值相等的分量按先 后顺序排列

$$y_{l+(l_1-1)r}^{(s)} \leqslant y_{l+(l_2-1)r}^{(s)} \leqslant \cdots \leqslant y_{l+(l_m-1)r}^{(s)}$$
(3)

每个粗粒化序列 $y_l^{(s)}$ 都对应一组符号序列 $s(r) = (l_1, l_2, \dots, l_m), r = 1, 2, \dots, R(R \leq m!)$ 。计 算每一种符号序列出现的概率 P_r ,则PE定义为

$$H_{P(m)} = -\sum_{1}^{R} P_r \ln P_r \tag{4}$$

将PE归一化处理,可得

$$H_P = \frac{H_{P(m)}}{\ln(m!)} \qquad (0 \leqslant H_P \leqslant 1) \qquad (5)$$

*H_p*的值表示粗粒化后时间序列*y*^(s)的随机程度。时间序列的随机程度越低,*H_p*的值越小;反之,该时间序列的随机程度越高,复杂度越大。

由此可知,根据粗粒化处理后的原始时间序列 {X(i); i=1,2,...,n}可以得到s尺度下的粗粒化时 间序列,分别计算其PE熵值 H_p ,得到该原始时间序列 的MPE,即 $H_{mp}(X)$ ={ $H_p(1), H_p(2), ..., H_p(s)$ }。

1.2 改进的粗粒化方法

粗粒化过程的优化如图1所示。对时间序列进 行粗粒化处理的MPE计算,其目的是为了根据合适 的尺度因子,将长度为N的原始时间序列分成s个 不相重叠的窗口,分别计算每个窗口内数据点的均 值,构成一组新的时间序列。将原始时间序列直接 除以尺度因子时,若s值较大,则会导致新建窗口内 时间序列长度过短;若序列中数据个数太少,会造成 MPE熵值计算结果不精确。

针对此问题,将移动平均粗粒化过程应用于 MPE计算中,如图1(b)所示^[19]。已知尺度因子,通 过移动平均粗粒化过程得到对应序列

$$z_{j}^{(s)} = \frac{1}{s} \sum_{i=j}^{j+s-1} X(i) \ (j=1,2,\cdots,(n-s+1)) \ (6)$$

粗粒化处理后的序列长度为(n-s+1),若取 原始时间序列长度为600,s=10,相较于原始粗粒 化处理后得到的序列长度60,改进的粗粒化处理得 到的序列长度却为591,数据长度会直接导致包含 信息量的不统一。由此可知,改进后的粗粒化方法 可以提高计算结果的准确性。

1.3 相空间重构参数的选取

计算各粗粒化序列的 PE 熵值之前,需分别确 定嵌入维数*m*与延迟时间 τ 。*m*与 τ 值的计算有独 立确定与联合确定2种方法,虽然2种方法各有优 缺点,但在检测结构的异常情况时,独立确定方法有 更高的精确性^[20]。因此,分别以伪近临法(false nearest neighbor,简称FNN)与互信息法(mutual information,简称 MI)求取*m*与 τ 。参数选取的准则 是:恰当的维数*m*为相空间中伪近临点的百分比趋 于0时对应的维数,且该维数之后,伪近临点的值不 发生改变;最佳延迟时间 τ 为互信息第1次达到最小



Fig.1 Optimization of coarsening process

值所对应的延迟时间,实测数据中 $m \ge 2, \tau \ge 1$ 。

1.4 最佳数据分析长度的选取

基于 IMPE 的最佳振测数据分析长度的确定步骤如下:

1)通过设置传感器采集振测数据{X(i); i=
 1, 2, …, n};

2) 粗粒化处理信号数据,根据振动信号长度选择适合的尺度因子,利用移动平均粗粒化方法计算 粗粒化序列z_j^(s);

相空间重构,各粗粒化后数据的相空间重构
 参数*m*,τ根据FNN与MI选取;

4) 计算粗粒化后各时间序列的 PE 熵值 PE₁, PE₂, …, PE_s,得到 MPE_s = { PE₁, PE₂, …, PE_s},令 MPE = $\frac{PE_1 + PE_2 + \dots + PE_s}{S}$,以 MPE 作为衡量 振测数据复杂程度的依据;

5) 同一振动条件下,计算振测数据不同长度 的 MPE 均值 MPE(N_1), MPE(N_2), …, MPE(N_i), …, MPE(N_n),随着 N不断增大, MPE 熵值逐渐稳 定于一定值,当 MPE(N_n)一MPE(N_{n-1})≈0时,不 再增加数据长度,此时以 MPE(N_n)作为标准熵值, MPE(N_n)所对应的数据长度 N_n 作为标准数据 长度;

6)根据精度要求 MPE(N_i) ≥ 97% MPE(N_n), 将MPE(N_i)与MPE(N_1), MPE(N_2), …, MPE(N_i), …, MPE(N_n)进行比较,选出满足精度的 MPE(N_i),将MPE(N_i)所对应的最短数据长度定义 为振测数据最佳分析长度。

2 仿真试验

2.1 验证改进的粗粒化方法

改进粗粒化方法在MPE上有较大优势,选取数 据长度较短的白噪声,计算其在尺度因子s∈[1,15] 时,优化前后粗粒化方法的熵值,以检验其优异性, 白噪声熵值变化如图2所示。

由于白噪声是所有频率具有相同能量的随机噪声,因此其时间序列的理论熵值为1。在实测过程中,熵值的精度会受到数据长度的影响,在选取白噪声数据长度*N*=200,500时,其数据实测PE熵值在0.90~0.97范围内。由图2可知,经原始与改进两种



粗粒化方法处理后的白噪声熵值均随尺度因子的增 大而逐渐减小,可验证式(1)和式(6)的正确性,数据 长度的降低会增大理论值与真实值的差距。由于受 尺度因子的影响较小,随着数据长度的增加,熵值减 小的速度远低于原始粗粒化方法处理的白噪声的熵 值,故提高了计算精度。可见,当数据长度与尺度因 子相同时,改进粗粒化方法计算得到的实测熵值更 接近于理论值。此外,根据计算结果可得,不论粗粒 化方法改进与否,N=500的白噪声信号的熵值均比 N=200的白噪声信号熵值更精准。这表明不仅粗 粒化方法的选取会影响信号的熵值精度,数据信号 的序列长度也是影响熵值准确度的重要因素。

2.2 检验数据长度对 MPE 的影响

从仿真分析可知,序列长度会影响熵值的精准 性。序列长度过大,存在计算繁琐、耗时长和突变处 模糊化等弊端;序列长度过短,会导致包含特征信息 的数据缺少或丢失,导致其熵值的可信度偏低。因 此,需要寻找一个合适的序列长度以确保计算结果 的精确性。笔者选取具有标准熵值的白噪声信号进 行分析,不同数据长度下白噪声熵值变化曲线如 图3所示。可见,信号的数据长度越长,其熵值趋近 于真实值1;尺度因子越大,对数据长度的影响越 小,熵值越准确。这说明 MPE 熵值对信号的数据长 度以及尺度因子的变化较敏感,且当数据长度达到 4000~5000时, 熵值的变化极为轻微, 最后稳定于 0.998,说明实测数据的熵值只会无限趋近于真实 值,其精准度存在一定界限。因此,在信号分析中仅 在一定长度范围内存在数据越长、分析越准确的现 象,并不会无限精确。





Fig.3 Variation of white noise entropy value at different sequence lengths

当MPE熵值随数据长度的增加而趋于稳定后, 选取此时的数据长度与MPE熵值设定为标准序列 长度与标准熵值。为提高计算效率,在保证分析结 果准确的条件下,选取满足标准熵值97%精度的熵 值作为有效熵值,所对应的振测数据长度作为最佳 分析长度。当白噪声的数据长度*N*=5000时,熵值 趋于稳定值0.998,与理论值相差甚微。因此,选取 数据长度*N*=5000时所对应的熵值0.998作为标准 熵值,选择满足标准值97%精度的熵值即0.968所 对应的数据长度作为最佳分析长度,选取数据长度 *N*=3000(对应熵值为0.971)作为该白噪声数据的 最佳分析长度。

2.3 构造仿真信号

为验证选取数据最佳分析长度方法的可靠性, 构造了仿真纯净信号f₁(t)与白噪声f₂(t),其表达式 分别为

 $f_1(t) = 20e^{-t\pi/2}\sin(15t) + 15e^{-t/3}\sin(20t) \quad (7)$ $f_2(t) = \operatorname{randn}(m) \quad (8)$

其中:t为时间;m为样本数;randn(m)为白噪声,其服从标准正态分布;采样频率f = 100 Hz。

向纯净信号 $f_1(t)$ 中添加信噪比(signal noise ratio,简称 SNR)分别为5%,10%,20%,30%,40%和50%的白噪声 $f_2(t)$,根据上述方法,分别选取数据长度为N=500,800,1000,1500,2000,2500,3000,3500,4000,4500和5000的加噪信号,计算其熵值曲线。加噪信号熵值变化如图4所示。

由图4可知,加噪信号的熵值随数据长度的增 长而增加,当数据长度达到4000~5000时,熵值基 本稳定在0.998,与白噪声信号有相同的变化规律。 因此,选取满足标准熵值97%精度的熵值所对应的 数据长度 N=2000,作为加噪信号的振测数据最佳



Fig.4 Signal entropy values with different white noise

分析长度。在数据长度达到最佳数据长度后,不同 信噪比的加噪信号的熵值趋于同一稳定熵值,可证 明多尺度排列熵具有较强的抗噪能力和较好的鲁棒 性,可有效判断信号的状态。

3 工程实例分析

三峡大坝是三峡水利枢纽的主体建筑物,各种运行工况下大坝存在不同的激励源,使坝体振动较为复杂,对其监测及研究也较多^[21-22]。以三峡大坝5号溢流坝段为研究对象,测试现场及传感器布置如图5所示。信息采集由设置在结构表面的传感器完成^[23],为保证收集效果,避免遗漏系统特征信息,坝体共放置了6个传感器,如图5(b)所示。为全面反映坝体的振动特性,选取1#,2#,3#,4#测点水平向与垂向动位移传感器两类振动状态的1~6通道的振测信号作为研究对象,采样时长为40s,采样频率为100 Hz。

根据 FNN 与 MI 分别确定相空间重构参数 m 和 τ 。经计算,5号溢流坝段坝体振动在 1~6 通道的 6 种状态下的相空间重构参数分别稳定在 $m = 4, \tau =$ 4 状态。相空间重构参数的选取如图 6 所示。

图 7 为三峡 5 号溢流坝段 6 个通道的振动信号 熵值变化曲线,选取了 N=200,500,1 000,2 000, 3 000 和 4 000 等 6 种不同的序列长度。可见:各通 道振动信号的熵值各不相同,水平向动位移振动熵 值(1~4 通道)均比垂向动位移振动熵值(5,6 通道) 所测的熵值小;熵值随数据长度的增加而逐步递增 至平稳,当数据增加到一定长度时,熵值趋于一稳定 值,与白噪声信号具有相同的规律。这说明利用 MPE方法选取信号分析长度是可行的,各通道在数 据长度 N=2 000时的熵值均达到稳定状态,对应的



(a) 测试现场传感器(a) Field test sensors of dam



(b) Measuring point layout of dam (unit: m)

图5 三峡大坝5号溢流坝段测试现场及传感器布置

Fig.5 Field test and measuring point layout of No.5 overflow section of Three Gorges Dam









图7 坝体不同工况下的熵值变化图

Fig.7 Entropy change diagram of dam body under different conditions

熵值分别为0.773,0.774,0.789,0.766,0.847和0.928,依据97%的精度要求,计算各工况下最佳数据分析长度均为N=1000。

4 结 论

 1)该方法的提出为信号分析提供了可靠的数据长度,排除了人为选择数据长度导致的随机性与 主观性,提高了振动监测的精确程度,便于对结构振动状态进行准确判断。

 2)通过对信噪比不同的加噪信号计算分析,验 证了多尺度排列熵具有较强的抗噪能力和较好的鲁 棒性,可有效避免混杂噪声对计算结果精度的影响。

3) 仿真实验数据与具体工程的分析结果表明, 数据的长度与其熵值的精确性和稳定性密切相关, 最佳数据分析长度所对应的熵值满足标准熵值 97%,即可满足工程精度要求。

参考文献

- [1] DONG J, YI S, DEYU W, et al. Data analysis and processing for Maglev vibration test systems [C] // Electronic Measurement & Instruments (ICEMI). Qingdao, China; IEEE, 2015; 1162-1166.
- [2] 张国安,何平,余焕伟,等.基于广义延拓的电梯振动信号优化处理与分析[J].中国特种设备安全,2018,34
 (10):4-8.

ZHANG Guoan, HE Ping, YU Huanwei, et al. Optimal processing and analysis of elevator vibration signal based on generalized extension[J]. China Special Equipment Safety, 2018,34(10):4-8. (in Chinese)

- [3] WANG Z C, WU F, REN W X. Stationarity test of vibration signals with surrogate data and time – frequency analysis [J]. Advances in Structural Engineering, 2017, 20(8): 1143-1154.
- [4] 陈佳袁, 闫杰.基于ARMA模型的水文数据预测[J]. 浙江水利科技, 2017, 45(6):27-30.
 CHEN Jiayuan, YAN Jie. Hydrological data prediction based on ARMA model[J]. Zhejiang Hydrotechnics, 2017, 45(6): 27-30. (in Chinese)
- [5] 马佳妮,张超,吕雅慧,等.基于长时间序列遥感数据反 演NPP的耕地质量评价[J].农业机械学报,2019,50
 (1):202-208.

MA Jiani, ZHANG Chao, LÜ Yahui, et al. Evaluation of cultivated land quality based on long time series remote sensing data inversion of NPP[J]. Transactions of The Chinese Society of Agricultural Machinery, 2019, 50(1): 202-208. (in Chinese)

[6] 蒋增林,叶江明,陈昊.基于时间序列分析的负荷预测 方法的比较研究[J].南京工程学院学报(自然科学 版),2018,16(2):26-31.

JIANG Zenglin, YE Jiangming, CHEN Hao. A comparative study of load forecasting methods based on time series analysis[J]. Journal of Nanjing Institute of Technology (Natural Science Edition), 2018, 16(2):26-31. (in Chinese)

[7] 练继建,李成业,刘昉,等.基于HHT的泄流结构损伤 在线监测方法研究[J].中国工程科学,2011,13(12): 38-44.

LIAN Jijian, LI Chengye, LIU Fang, et al. Research on HHT-based on-line damage monitoring method for discharge structures [J]. Engineering Sciences, 2011, 13(12): 38-44. (in Chinese)

[8] 李火坤,杜磊,梁萱,等.基于原型振动测试的泄洪闸闸 墩动位移反演[J].振动、测试与诊断,2018,38(6): 1122-1129.

LI Huokun, DU Lei, LIANG Xuan, et al. Dynamic displacement inversion of spillway gate pier based on prototype vibration test[J]. Journal of Vibration, Measurement & Diagnosis, 2018, 38(6): 1122-1129. (in Chinese)

[9] 王洪波,罗贺,彭张林,等.股指时间序列的低维分形表示及相似性研究[J].系统工程学报,2019,34(1):46-56.
 WANG Hongbo, LUO He, PENG Zhanglin, et al.

Low-dimensional fractal representation and similarity of stock index time series[J]. Journal of Systems Engineering, 2019, 34(1):46-56. (in Chinese)

- [10] AZIZ W, ARIF M. Multi-scale permutation entropy of physiological time series [C] //2005 Pakistan Section Multitopic Conference. Karachi, Pakistan: IEEE, 2005: 1-6.
- [11] YIN Y, SHANG P, AHN A C, et al. Multiscale joint permutation entropy for complex time series [J]. Physica A: Statistical Mechanics and its Applications, 2019, 515: 388-402.
- [12] LIU T, YAO W, WU M, et al. Multiscale permutation entropy analysis of electrocardiogram [J]. Physica A: Statistical Mechanics and Its Applications, 2017, 471: 492-498.
- [13] CHOI Y S, HYUN K, CHOI J Y. Assessing multiscale permutation entropy for short electroencephalogram recordings[J]. Cluster Computing, 2016, 19(4): 2305-2314.
- [14] LI D, LI X, LIANG Z, et al. Multiscale permutation entropy analysis of EEG recordings during sevoflurane anesthesia[J]. Journal of Neural Engineering, 2010, 7 (4): 046010.
- [15] VAKHARIA V, GUPTA V K, KANKAR P K. A multiscale permutation entropy based approach to select wavelet for fault diagnosis of ball bearings [J]. Journal of Vibration and Control, 2015, 21(16): 3123-3131.
- [16] LI Y, ZHANG W, XIONG Q, et al. A rolling bearing fault diagnosis strategy based on improved multiscale

permutation entropy and least squares SVM[J]. Journal of Mechanical Science and Technology, 2017, 31(6): 2711-2722.

- [17] ZUNINO L, ZANIN M, TABAK B M, et al. Forbidden patterns, permutation entropy and stock market inefficiency[J]. Physica A: Statistical Mechanics and Its Applications, 2009, 388(14): 2854-2864.
- [18] ZHANG J, HOU G, CAO K, et al. Operation conditions monitoring of flood discharge structure based on variance dedication rate and permutation entropy[J]. Nonlinear Dynamics, 2018, 93(4): 2517-2531.
- [19] 马晓君. 泵站管道振动状态监测研究与应用[D]. 保定:华北水利水电大学, 2019.
- [20] 饶国强,冯辅周,司爱威.排列熵算法参数的优化确定 方法研究[J].振动与冲击,2014,33(1):73-79.
 RAO Guoqiang, FENG Fuzhou, SI Aiwei, et al. Method for optimal determination of parameter in permutation entropy algroithm [J]. Transactions of the Journal of Vibration and Shock, 2014, 33(1):73-79. (in Chinese)
- [21] 李火坤,刘世立,魏博文,等.基于方差贡献率的泄流结构多测点动态响应融合方法研究[J].振动与冲击,2015,34(19):181-191.
 LI Huokun, LIU Shili, WEI Bowen, et al. Research on fusion method of multi-point dynamic response of discharge structure based on variance contribution rate[J]. Journal of Vibration and Shock, 2015, 34(19):181-191. (in Chinese)
- [22] 李成业,练继建,刘昉,等.EMD与小波阈值联合滤波 方法的改进及其在泄流结构振动分析中的应用[J].振 动与冲击,2013,32(19):63-70.
 LI Chengye, LIAN Jijian, LIU Fang, et al. The improvement of combined EMD and wavelet threshold filtering method and its application in vibration analysis of discharge structure[J]. Journal of Vibration and Shock, 2013,32(19):63-70. (in Chinese)
- [23] 张建伟,暴振磊,刘晓亮,等.适用于梯级泵站压力管道 的传感器优化布置方法[J].农业工程学报,2016,32 (4):113-118.

ZHANG Jianwei, BAO Zhenlei, LIU Xiaoliang, et al. Sensor placement optimization method for pressure pipeline suitable forgrading pumping station [J]. Transactions of the Chinese Society of Agricultural Engineering, 2016, 32(4): 113-118. (in Chinese)



第一作者简介:张建伟,男,1979年3月 生,博士、教授。主要研究方向为水工结构 耦联振动与安全。曾发表《Operation conditions monitoring of flood discharge structure based on variance dedication rate and permutation entropy》(《Nonlinear Dynamics》2018, Vol. 93, No.4)等论文。 E-mail: zjwcivil@126.com