

基于混沌时间序列的水电机组状态短期预测

商志根, 姚志树

(盐城工学院 电气工程学院, 江苏 盐城 224051)

摘要:基于混沌时间序列短期可以预测的特点,构建水电机组状态短期预测。用采样周期确定相空间时延 τ , G-P 算法确定关联维数从而确定相空间的嵌入维数 m , 小数据量法证明水电机组振动状态的混沌特性。在重构相空间中, 运用加权一阶局域法构建水电机组状态短期预测模型。结果表明:混沌特性指数 $\lambda = 0.2605$ 的水电机组振动状态具有混沌特性, 可以在最佳嵌入维数 $m = 4$ 的情况下进行预测, 实例结果表明采用混沌理论进行水电机组状态短期预测是可行的。

关键词:水电机组; 混沌时间序列; 相空间重构; 状态预测

中图分类号:TV737; TP391 **文献标识码:**A **文章编号:**1671-5322(2010)02-0027-05

水电机组是水电生产系统的核心, 它可靠而又稳定的运行将直接影响到水电企业安全、可靠。为保障水电机组安全、稳定、可靠的运行, 依据收集到的水电机组的状态信息, 实现对水电机组状态的有效识别和状态趋势的准确分析, 将为水电企业实现状态检修与预知检修提供充实而可靠的支持。

机械设备故障的产生发展过程是一个从量变到质变的过程, 大部分水电机组故障会在振动信号特征量中得到反映。另一方面, 水电机组状态的可预测性要求相似样本呈现相似特性, 即设备缺陷及其扩展变化, 有时具有共同的特征。借鉴同类型设备或近似类型设备的相关故障发展变化特性, 将为设备状态预测提供必要的参考依据。

振动信号是时间序列特征量。对于水电机组振动时间序列而言, 可以认为是水电机组这一复杂的确定性系统中的一个分量, 它是由与之相关的水电机组其他分量所决定的, 这些相关分量的信息均隐含在这个振动时间序列之中。水电机组的运行受到水力、机械以及电气等多方面因素的影响, 直接导致振动信号特征量表现出明显的非线性与非平稳特性^[1,2], 并且在一定情况下表现出混沌的特性。因而, 水电机组振动信号序列预测问题是一个典型的非线性时间序列预测问题。因其混沌特性, 理论上讲水电机组振动时间序列

是可以短期预测的。

近年来, 对混沌时间序列预测模型的研究受到了广泛关注。常用的模型有回归模型、神经网络模型^[3]、支持向量机模型^[4,5]、自适应预测模型等。值得一提的是, 邹敏等人基于最小二乘支持向量机(LS-SVM)和小波分析建立了水电机组状态预测模型^[6], 对水电机组状态的预测模型进行了有益的探索, 但对水电机组振动信号特性的研究不够深入。

混沌时间序列分析为解决非线性问题提供了一个良好的思路^[7,8]。本文以单一混沌时间序列(水电机组振动信号)作为预测研究对象。首先证明水电机组振动信号的混沌特性, 继而找出该序列的分数维数和时间延时, 重构状态序列空间, 找出预测点的邻界状态点, 根据吸引子的吸引性和稳定性, 估计出邻界状态点与其后续序列之间的函数关系, 据此推导出预测点的预测值。

1 混沌序列识别

混沌运动对初始值非常敏感, 描述混沌运动特征的指标有 Lyapunov 指数、测度熵、拓扑熵。因 Lyapunov 指数计算相对简单, Lyapunov 指数法是混沌序列识别的常用方法。计算 Lyapunov 指数的方法有小数据量法、Wolf 方法、P-范数方法和奇异值分解方法等。本文采用小数据量法来提

收稿日期: 2010-03-31

作者简介: 商志根(1979-), 男, 江苏盐城人, 讲师, 硕士, 主要研究方向为智能信息计算。

取水电机组振动信号中的 Lyapunov 指数,小数据量法具有计算量小、计算可靠等优点^[9]。

在计算 Lyapunov 指数之前,要求对时间序列进行重构相空间。重构相空间的优良对于最大 Lyapunov 指数的计算精度具有重要影响。因此重构相空间的几个参数的确定就非常重要。对于确定的时间序列而言,重构相空间首先确定延时时间 τ 、最佳嵌入维数 m 和平均周期 p 。

1.1 延时时间与平均周期

对于延时时间的确定,常用的方法有自相关函数法、平均互信息法和 C-C 方法。自相关函数法不适合非线性问题,只适合于数据较短的时间序列,计算较为简便;而互信息法适合于非线性问题,且适合于大数据组,但计算较为复杂。C-C 方法是结合这两种方法的一种改进方法。本文直接采用周期确定延时时间,即 $\tau=1$ 。平均周期可通过能量光谱的平均频率的倒数估计出来,即可通过快速傅氏变换算法得到。

1.2 最佳嵌入维数

最佳嵌入维数 m 也是相空间重构的重要参数。如果 m 选得太小,则吸引子可能折叠以致在某些地方自相交,这样在相交区域的一个小领域内可能会包含来自吸引子不同部分的点。如果 m 选取的过大,理论是可行的,但实际应用中,随着 m 的增大会大大地增大吸引子的集合不变量(如 Lyapunov 指数)的计算工作量。

目前,嵌入维数的主要计算方法有 G-P 算法、虚假邻点法和奇异值分解法。本文采用 G-P 算法计算得到随着嵌入维数增加关联维数的变化情况,然后确定最佳嵌入维数。

现场得到的振动时间序列为 $x_1, x_2, \dots, x_N$, 其延迟向量可表示为:

$$X_i(M) = (x_i, x_{i+\tau}, \dots, x_{i+(M-1)\tau}), \\ i = 1, 2, \dots, N - (M-1)\tau \quad (1)$$

关联积分 $C(r, N, M, \tau)$ 采用如下定义:

$$C(r, N, M, \tau) = \frac{2}{(N - \omega + 1)(N - \omega)} \cdot \\ \sum_{N=\omega}^N \sum_{i=1}^{N-n} H\{|r - |X_{i+n} - X_i||\} \quad (2)$$

其中 H 为 Heaviside 函数,即为 $H(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$

同时要求条件:

$$\omega \geq \tau \left(\frac{2}{N} \right)^{\frac{1}{d}} \quad (3)$$

τ 为延时时间, M 为嵌入维数,时序的关联维数定义为:

$$d = \lim_{r \rightarrow 0} \lim_{N \rightarrow \infty} \frac{d[\log_2 c(r, \omega)]/dr}{d[\log_2 r]/dr} \quad (4)$$

G-P 算法基本思想是:增加 M , 上式中 d 变化逐渐减小,最终趋于平稳,此时的关联维数称为饱和关联维数,而最佳嵌入维则可取使趋于平稳的嵌入维数的下限。

1.3 最大 Lyapunov 指数

利用小数据量法提取振动时间序列的 Lyapunov 指数的计算步骤为:

第 1 步:确定延时时间 τ 、最佳嵌入维数 m 和平均周期 p ;

第 2 步:根据延时时间 τ 、最佳嵌入维数 m 重构相空间;

$$X_i(m) = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}), \\ i = 1, 2, \dots, N - (m-1)\tau \quad (5)$$

第 3 步:找相空间中每个点 X_j 的最近邻点 X_j , 并满足短暂分离条件,

$$d_j(0) = \min_{j_j} |X_j - X_{j_j}|, |j - j_j| > p \quad (6)$$

第 4 步:对相空间每个点 X_j 计算该邻点对的 i 个离散时间步后的距离 $d_j(i)$;

$$d_j(i) = |X_{j+i} - X_{j+i}|, \\ i = 1, 2, \dots, \min[N - (m-1)\tau - j, \\ N - (m-1)\tau - j] \quad (7)$$

第 5 步:对每个 i , 求出所有 j 的 $\ln d_j(i)$ 平均 $y(i)$, 即

$$y(i) = \frac{1}{q \Delta t} \sum_{j=1}^q \ln d_j(i) \quad (8)$$

其中 q 为非零 $d_j(i)$ 的数目, Δt 为样本采样周期,并用最小二乘法做出回归直线,该直线斜率就是最大 Lyapunov 指数。

2 加权一阶局域法预测模型

对混沌时间序列的预测模型有全局预测法、局域预测法、自适应预测模型、重构系统方程非线性自适应预测方法、神经网络以及支持向量机等模型。本文采用加权一阶局域法作为预测模型。加权一阶局域法能够反映最近邻域点对预测结果的不同影响,具有较好的预测精度,且其计算速度较快,建模较简单。

对于已知时间序列 $x_1, x_2, \dots, x_N$, 其延迟向量表示如式(5)。现假设为已知量, x_{n+1} 为需要预测量,则包含有这两个分量的最新相空间相点为:

$$X_{n-(m-1)\tau} = (x_{n-(m-1)\tau}, x_{n-(m-2)\tau}, \dots, x_n) \\ X_{n+1-(m-1)\tau} = (x_{n+1-(m-1)\tau}, x_{n+1-(m-2)\tau}, \dots, x_{n+1}) \quad (9)$$

按照相空间的演化轨迹,只要获得相点 $X_{n-(m-1)\tau}$ 的下一轨迹点 $X_{n+1-(m-1)\tau}$,并从中分离出唯一未知量 x_{n+1} ,就完成了该值的预测。 $X_{n+1-(m-1)\tau}$ 的获得常通过拟合预测中心点 $X_{n-(m-1)\tau}$ 的 k 个最近邻域点来得到,最近邻域点的确定采用欧氏距离法,即

$\|X_{n-(m-1)\tau} - X_j\| < \varepsilon, j = 1, 2, \dots, k \quad (10)$ 其中 $\|\cdot\|$ 为欧氏泛数, ε 为一个极小整数, $X_j (j = 1, 2, \dots, k)$ 为 $X_{n-(m-1)\tau}$ 的最近邻域点。确定各最近邻域点的下一个演化轨迹点 X_{j+1} , 即:

$$\begin{aligned} X_1 &\rightarrow X_2 \\ X_2 &\rightarrow X_3 \\ &\dots \\ X_k &\rightarrow X_{k+1} \end{aligned} \quad (11)$$

在基于最近邻域的局域预测中,不同的邻近点对预测的影响权重不同,距离预测点越近,其影响应该越大,为反映邻近点的不同影响程度在局域线性近似拟合中引入权重:

$$P_i = \frac{\exp[-a(d_i - d_m)]}{\sum_{i=1}^k \exp[-a(d_i - d_m)]}, \quad i = 1, 2, \dots, k \quad (12)$$

其中 d_i 为第 i 个邻近点 X_i 距预测中心点的距离, d_m 是 d_i 中的最小值。 a 为系数,一般取 $a = 1$ 。则一阶局域线性近似拟合函数为:

$$X_{i+1} = A + BX_i, i = 1, 2, \dots, k \quad (13)$$

A 和 B 是拟合参数,采用最小二乘法求解参数。

$$J = \sum_{i=1}^k P_i (X_{i+1} - A - BX_i)^2 \quad (14)$$

对 A 和 B 求偏导数得:

$$\begin{cases} \sum_{i=1}^k P_i (X_{i+1} - A - BX_i) = 0 \\ \sum_{i=1}^k P_i (X_{i+1} - A - BX_i) X_i = 0 \end{cases} \quad (15)$$

求解方程组得到参数 $\hat{A}$ 和 $\hat{B}$, 则预测点为:

$$\hat{X}_{n+1-(m-1)\tau} = \hat{A} + \hat{B}X_{n-(m-1)\tau} \quad (16)$$

从预测相点

$$\hat{X}_{n+1-(m-1)\tau} = (\hat{X}_{n+1-(m-1)\tau}, \hat{X}_{n+1-(m-2)\tau}, \dots, \hat{X}_{n+1})$$

分离出 x_{n+1} 的预测值 $\hat{x}_{n+1}$ 即可。

3 水电机组状态短期预测分析

本文振动序列数据共 128 个采样点,如图 1 所示。因水电机组的运行受到水力、机械以及电气等多方面因素的影响,导致振动信号特征量出现混沌特性是可能的。本文采用周期确定延时时间,即 $\tau = 1$ 。快速傅氏变换算法得到机组振动峰峰值功率谱曲线,并得到平均周期 $p = 32$, 如图 2 所示。利用第 1 部分提到的 G-P 算法得到关联维数 d 随着嵌入维数 m 变化而变化的情况,如图 3 所示。由图 3 可以看出,当 $m \geq 4$ 后,曲线就变得几乎重合,所以最佳嵌入维数取 $m = 4$ 。采用小数据量法进行振动序列混沌特性的鉴别,如图 4 所示,得到最大 Lyapunov 指数 $\lambda = 0.2605$ 。因为此序列最大 Lyapunov 指数 $\lambda > 0$,所以机组振动峰峰值序列具有混沌特性。

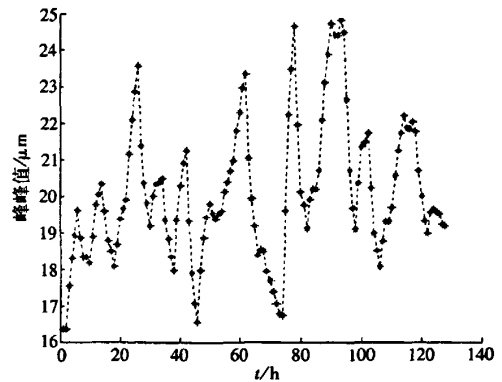


图1 机组振动峰峰值序列

Fig. 1 Vibration signal series of generating unit

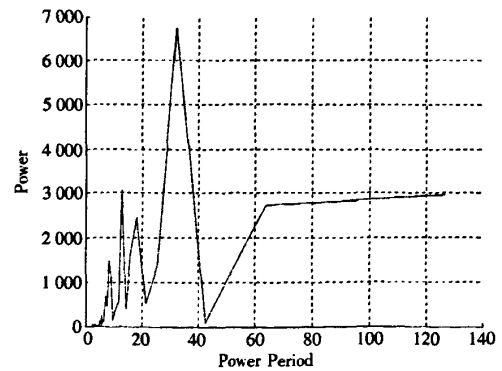


图2 机组振动峰峰值功率谱曲线

Fig. 2 The power spectrum curve of vibration signal

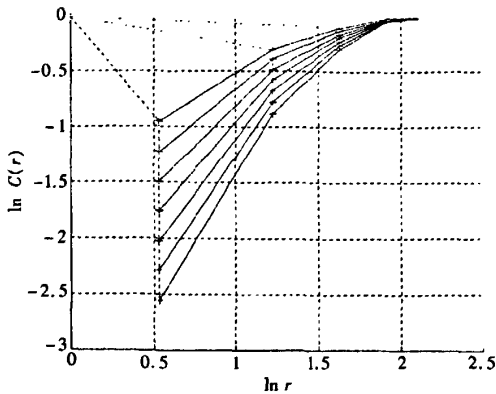


图 3 G-P 算法确定最佳嵌入维
Fig. 3 Determine the best embedding dimension based on G-P algorithm

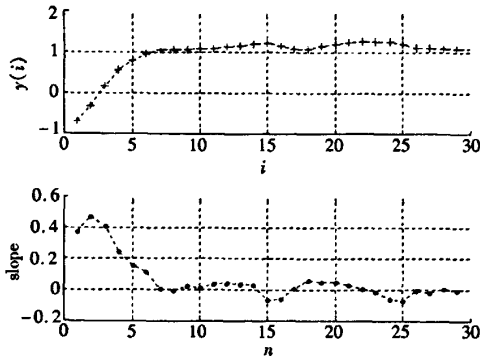


图 4 小数据量法确定最大 Lyapunov 指数
Fig. 4 Determine the largest Lyapunov exponent based on small data sets

在机组振动峰峰值的重构相空间之后,共产生 124 组数据。根据第 2 部分理论,利用前 100 组数据建立加权一阶局域近似模型,后 24 作为测试数据。从图 5 可以看出大部分的预测误差都在较小的范围,能够满足实际现场的预测需求。图 6 显示的是模型预测值减去实际值的情况,从图 6 中可以看出绝对误差值在 $-1.2 \sim +1.2$ 之间,所以这个误差值是在允许范围内。

4 预测误差分析

本文用测试数据集中预测值与实际值的均方差作为评判预测模型效果的一个指标,

$$ESS = \left[\frac{1}{L} \sum_{i=1}^L (x(n+i) - \hat{x}(n+i))^2 \right]^{\frac{1}{2}} \quad (17)$$

式中: $x(n+i)$ 表示实际值, $\hat{x}(n+i)$ 表示预测值。

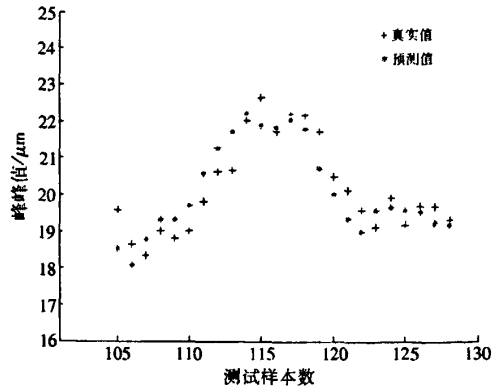


图 5 测试样本的真实值和模型预测值
Fig. 5 Actual value and model output of checking samples

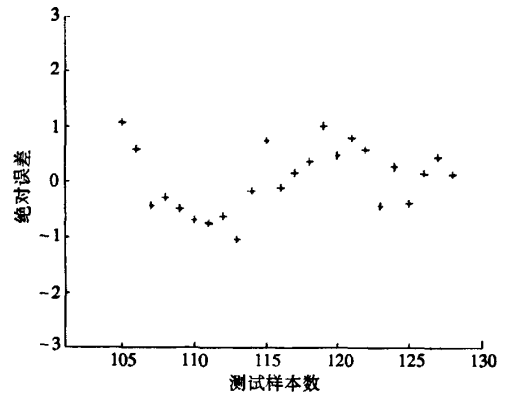


图 6 模型绝对预测误差值(预测值 - 实际值)
Fig. 6 Absolute value of prediction errors (forecasting value - actual value)

ESS 小,说明预测值偏离实际值的程度小,预测效果较好;ESS 大,说明预测值偏离实际值的程度大,预测效果就差。

从表 1 中可以看出 4 维为最佳嵌入维数。当增大嵌入维数时噪声可能占主导地位,使得均方差 ESS 变大;当减小嵌入维数时吸引子将折叠,无法恢复原系统的动力学特性,并且预测长度越长均方差 ESS 越大。

表 1 仿真结果误差分析

Table 1 Error analysis of simulation

嵌入维数 m	均方差 ESS
3	0.955 0
4	0.584 3
5	0.919 8

5 结束语

水电机组振动状态的预测结果直接关系到水电企业的安全。针对机组振动的非线性特性,本文证明了振动峰峰值序列为一混沌时间序列,以采样周期作为时间延迟,应用G-P算法确定关联维数的基础上确定最佳嵌入维数。依据小数据量法证明水电机组振动序列的混沌特性,然后采用加权一阶局域法实现了水电机组振动状态的短期预测。通过实例分析,结果表明:

(1) 利用重构相空间理论对水电机组振动状

态的时间序列进行了分析和预测,预测了未来短期内水电机组振动的一种整体变化趋势;

(2) 采用加权一阶局域法实现水电机组振动状态的短期预测,均方误差为0.5843,绝对误差在-1.2~+1.2之间,表明预测结果与实际情况符合较好,能够满足水电企业的实际需求。

研究表明应用混沌理论进行水电机组振动状态的分析与预测是可行的,它能够有效地反映水电机组振动状态的发展趋势,为水电企业提前采取安全措施提供科学的决策依据。

参考文献:

- [1] 董毓新. 水轮发电机组振动[M]. 大连:大连理工大学出版社,1989:224-254.
- [2] 梁武科,张彦宁,罗兴铨. 水电机组故障诊断系统信号特征的提取[J]. 大电机技术,2003,35(4):53-56.
- [3] 黄建国,罗航,王厚军,等. 运用GA-BP神经网络研究时间序列的预测[J]. 电子科技大学学报,2009,38(5):687-692.
- [4] Guoshao Su. A New Intelligent Model for Nonlinear Time Series Prediction[C]//2008 International Conference on Computer Science and Software Engineering,1:435-438.
- [5] 韩宁,闫德勤. 基于支持向量机的鲁棒盲水印算法[J]. 计算机工程与设计,2009,30(22):5273-5275.
- [6] 邹敏. 基于支持向量机的水电机组故障诊断研究[D]. 武汉:华中科技大学,2007:52-71.
- [7] 马军海. 复杂非线性系统的重构技术[M]. 天津:天津大学出版社,2005:46-79.
- [8] 程健,白静宜,钱建生,等. 基于混沌时间序列的煤矿瓦斯浓度短期预测[J]. 中国矿业大学学报,2008,37(2):231-235.
- [9] 张贵杰,高后秀,杨渝钦. 基于小数据法计算固体“类流态”的最大Lyapunov指数[J]. 天津大学学报,2006,39(增刊):185-189.

Short-term Prediction of Hydroturbine Generating Unit Condition Based on Chaotic Time Series

SHANG Zhi-gen, YAO Zhi-shu

(School of Electrical Engineering, Yancheng Institute of Technology, Jiangsu Yancheng 224051, China)

Abstract: Based on the characteristic of chaotic time series, a model was built to predict hydroturbine generating unit condition. The time delay τ was determined by sampling period, and the embedding dimension m was chosen according to correlation dimension, which was calculated by G-P algorithm. Chaotic characteristic of vibration signal series of hydroturbine generating unit was proved by small data sets arithmetic. The prediction model of hydroturbine generating unit condition was constructed by an adding-weight one-rank local-region method after the phase space was reconstructed. The results show that vibration signal series has a chaotic characteristic while the chaotic property exponent $\lambda = 0.2605$. Therefore, a prediction model can be carried out while the best embedding dimension m is 4. The results demonstrate that the prediction method is feasible.

Keywords: hydroturbine generating units; chaotic time series; phase space reconstruction; condition prediction

(责任编辑:沈建新;校对:张英健)