

基于复杂性测度的仿真模型验证

吴晓燕¹, 吴静², 陈永兴¹, 滕江川¹

(1. 空军工程大学防空反导学院, 陕西西安, 710051; 2. 94221 部队, 山东新泰, 271206)

摘要 复杂仿真系统模型验证中, 系统输出具有复杂、非线性和非平稳的特点, 传统模型验证方法解决上述问题存在较大风险。复杂性测度是描绘非线性时间序列特征的重要参数, 将复杂性测度理论引入到仿真模型验证, 在计算试验数据和仿真数据复杂性测度的基础上, 将仿真数据与试验数据之间的复杂性差异作为评价指标, 衡量模型的有效性。实例分析表明: 该方法计算简便, 对时间序列的长度、平稳性等方面没有特殊要求, 具有一定的实用性。

关键词 复杂性测度; 模型验证; Kolmogorov 复杂度; C_0 复杂度; 近似熵

DOI 10.3969/j.issn.1009-3516.2013.02.007

中图分类号 TP391.9 **文献标志码** A **文章编号** 1009-3516(2013)02-0028-04

Validation of Simulation Models Based on Complexity Measure

WU Xiao-yan¹, WU Jing², CHEN Yong-xing¹, TENG Jiang-chuan¹

(1. Air and Missile Defense College, Air Force Engineering University, Xi'an 710051, China;

2. Unit of 94221, Xintai 271206, Shandong, China)

Abstract: The outputs of complex simulation system are characterized by complex, nonlinearity and smoothness. Great risks come into existence when the traditional model validation method is used to solve the above problems. Complexity measure is an important parameter in describing the characteristic of non-linear time sequence, so the theory of complexity measure is introduced into simulation model validation. Based on calculating the complexity of experiment and simulation data, the validity of model is measured by comparing the complexity measures between the simulation results and experiment data. Finally, an example is given to analyze the conclusion. The result shows that the method is simple in computation, practical and without any special requirement to the length and stationarity of time sequence.

Key words: complexity measure; model validation; Kolmogorov complexity; C_0 complexity; approximate entropy

仿真模型验证是校核、验证和确认(Verification Validation and Accreditation, VV&A)工作的重要内容, 也是确保仿真模型可信性的重要途径。进行模型验证最基本、最直接的途径是考察在相同初始条件下仿真模型输出与实际系统输出(或理论

计算结果)的一致性程度^[1]。

然而在复杂系统建模与仿真中, 实际系统或仿真系统的各个元素或子系统中存在着非线性的相互作用, 而它们之间的函数关系大多是非线性的, 使得系统行为(模型或系统的输出时间序列)具有相当程

收稿日期: 2012-10-29

基金项目: 陕西省自然科学基金资助项目(2012JM8020)

作者简介: 吴晓燕(1957-), 女, 陕西西安人, 教授, 主要从事系统建模与仿真 VV&A 研究。

E-mail: x_ywu@126.com

度的复杂性。同时,在对实际系统量测过程中,由于测量噪声及各种干扰的存在,使得系统输出时间序列具有非线性、非平稳性、随机性和不确定性的特点^[2]。而传统的模型验证方法(如谱分析、数理统计、TIC不等式系数等)解决具有上述特点的模型验证问题存在较大的应用风险。其原因如下:对时间序列需要进行预处理,或者对样本数量、概率分布、平稳性等有多项要求,因而不可避免地造成部分信息的丢失或人为地加进一些信息,从而产生了误差。因此,如何有效处理复杂非线性的系统输出时间序列并确定序列间的一致性程度,成为当前模型验证研究的新内容之一。

复杂非线性动态系统的模型验证问题需要通过与复杂性有关的非线性、非平衡、动态的思想和系统方法来解决^[3]。复杂性测度是近年来发展起来的一种定量评价序列复杂性的方法,是描述非线性时间序列信息量的一个重要参数,其算法简单,具有较强的抗干扰能力,对时间序列的长度、平稳性等方面没有特殊要求,已经在生物信号分析、交通流时间序列分析等领域得到了广泛应用^[4-6]。

因此,本文将非线性动力学的复杂性测度理论引入到模型验证中,将 Kolmogorov 复杂度、 C_0 复杂度和近似熵作为系统输出时间序列的特征参数,分别提取试验数据和仿真数据的上述3种特征,计算综合复杂性测度,定义模型有效性指标来衡量模型的有效性,为复杂非线性系统的模型验证工作开辟了新的思路。

1 复杂性测度

复杂性测度简称复杂度,其最初定义是由 Kolmogorov 于 1965 年提出的,直到 Lempel 和 Ziv、Kaspar 和 Schuster 给出复杂度的算法,复杂度的应用才有了较为实质性的进展。目前常用的复杂度算法主要有:Kolmogorov 复杂度、 C_0 复杂度和近似熵。

1.1 Kolmogorov 复杂度

Kolmogorov 复杂度(Kolmogorov Complexity, KC)表征为能够产生某给定“0,1”序列最少的计算机程序的比特数^[7]。为了具体地用符号序列表示时间序列,以便进行 KC 的计算,Lempel 和 Ziv、Kaspar 和 Schuster 给出计算 KC 的具体算法,其主要原理如下^[8-9]:

1)多值粗粒化预处理。由于二值粗粒化方法可能会丢失动力学系统的一些有用信息,因此,本文采

用多值粗粒化方法重构原时间序列。设 $L(L>2, L \in \mathbb{N})$ 表示原序列中数据粗粒化的段数,建立一个具有 L 个不同字符的符号序列 $s(i)$ 。

2)Lempel-Ziv 复杂度的具体算法可参考文献^[7~9]。设长度为 n 的某时间序列按照 Lempel-Ziv 算法求得复杂度为 $c(n)$ 。

3)当 $n \rightarrow \infty$, $c(n)$ 会趋向于定值 $n/\log_L n$, 则归一化复杂度的计算公式为:

$$C_{KC} = \frac{c(n)\log_L n}{n} \quad (1)$$

复杂度的计算前提是各符号在字符串中出现的概率 $p_j (j=1, 2, \dots, L)$ 应该相等,当 p_j 相差较大时,应该考虑用归一化的信源熵 h 进行修正,则:

$$h = -\sum_{j=1}^L p_j \log_L p_j \quad (2)$$

修正后的多分段归一化复杂度计算公式为^[8]:

$$C_{KC} = \frac{c(n)\log_L n}{nh} \quad (3)$$

KC 反映了时间序列的随机程度,如果时间序列是周期性的,那么 C_{KC} 就会随时间序列的增加而趋向于 0;如果时间序列是随机的,则 C_{KC} 趋向于 1。 C_{KC} 越大,动力学系统就越复杂;相反越小,动力学系统的规律性就越明显^[7-9]。

1.2 C_0 复杂度

KC 复杂度算法中过粗粒化的预处理可能会导致原始信号中信息的大量丢失,不恰当的粗粒化甚至会改变原始时间序列的动力学特性,例如,有可能将随机时间序列改变成周期时间序列。为了消除这种潜在的危险,文献^[10]定义了 C_0 复杂度算法,它认为复杂运动是由规则运动和随机运动混合而成,而随机运动所占的份额就是 C_0 复杂度。文献^[11]指出对于零均值的时间序列,其计算结果与理论值不相符合,因此,引入参数 r ,提出了改进的 C_0 复杂度算法。

记 $\{f(k) \mid k=0, 1, \dots, N-1\}$ 是一个长度为 N 的时间序列,则:

$$F_N(j) = \frac{1}{N} \sum_{k=0}^{N-1} f(k) W_N^{kj}, j = 0, 1, \dots, N-1 \quad (4)$$

构成其傅里叶变换序列^[10],式中 $W_N = \exp\left\{\frac{2\pi i}{N}\right\}$, $i = \sqrt{-1}$ 表示虚数单位。

设 $\{F_N(j) \mid j = 0, 1, \dots, N-1\}$ 的均方值为:

$$G_N = \frac{1}{N} \sum_{j=0}^{N-1} |F_N(j)|^2 \quad (5)$$

引入参数 $r (r>1, r \in \mathbb{N})$,保留超过均方值 r 倍的频谱,而将其余部分置为零,即:

$$\tilde{F}_N(j) = \begin{cases} F_N(j), & |F_N(j)|^2 > rG_N \\ 0, & |F_N(j)|^2 \leq rG_N \end{cases} \quad (6)$$

根据文献[9]性质3,定义 C_0 复杂度简化为:

$$C_0(r) = \frac{\sum_{k=0}^{N-1} |F_N(k) - \tilde{F}_N(k)|^2}{\sum_{k=0}^{N-1} |F_N(k)|^2} \quad (7)$$

显然,当 $r=1$ 时, $C_0(r)$ 就是原来定义的 C_0 复杂度。在实际应用中 r 取 $5 \sim 10$ 为宜。

由式(7)可知: $C_0 \rightarrow 0$ 说明时间序列几乎是规则的,不含随机成分; $C_0 \rightarrow 1$ 说明时间序列几乎是完全随机的。所以,随着 C_0 的增加,意味着时间序列中的随机成分增加^[6]。

1.3 近似熵

近似熵(Approximate Entropy, ApEn)是一种不需要进行粗粒化的时间序列复杂度分析方法^[12]。该方法实际上是在衡量当维数变化时该时间序列产生新模式的概率大小,产生的新模式的概率越大,序列的复杂度越大,相应的近似熵也越大。近似熵具有一定的抗噪、抗野点能力,使用较短的数据就可以产生稳健的估计值,适合于工程应用,其定义和算法如下^[4-5,12]:

1) 对于一给定的时间序列 $\{x_i | 1 \leq i \leq N\}$,按顺序将其组成一个 m 维的向量集 $\mathbf{X}(i)$,即 $\mathbf{X}(i) = [x(i), x(i+1), \dots, x(i+m-1)]$, ($1 \leq i \leq N-m+1$);

2) 2个 m 维的向量 $\mathbf{X}(i)$ 与 $\mathbf{X}(j)$ 之间的距离定义为:

$$d[\mathbf{X}(i), \mathbf{X}(j)] = \max(|x(i+k) - x(j+k)|), 0 \leq k \leq m-1 \quad (8)$$

3) 按照给定阈值 r ($r > 0$),对每一个 i 值统计 $d[\mathbf{X}(i), \mathbf{X}(j)] < r$ 的数目,定义此数目与总的向量个数 $N-m+1$ 的比值为:

$$C_i^m(r) = \frac{\text{num}\{d[\mathbf{X}(i), \mathbf{X}(j)] < r\}}{N-m+1} \quad (9)$$

它表示以 $\mathbf{X}(i)$ 为中心,在窗口长度为 m 、容许偏差为 r 的情况下,其余向量 $\mathbf{X}(j)$ 与 $\mathbf{X}(i)$ 的距离小于 r 的概率,从而表示所有 $\mathbf{X}(j)$ ($j \neq i$) 与 $\mathbf{X}(i)$ 的关联程度,也就是表示矢量序列 $\{x(i)\}$ 的规律性程度。

4) 根据 $C_i^m(r)$,定义矢量序列的平均关联程度 $\Phi^m(r)$ 为:

$$\Phi^m(r) = (N-m+1)^{-1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \quad (10)$$

则,近似熵被定义为:

$$\text{ApEn}(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r) \quad (11)$$

ApEn 值与 N 、 m 和 r 的取值有关。Princus 指出 m 取 2, r 取 0.1~0.25 倍原始数据的标准差时, ApEn 值对序列长度 N 的依赖程度最小^[4-5]。

2 基于复杂性测度的模型验证

基于复杂性测度的仿真模型验证的基本原理是:在分别计算试验数据和仿真数据复杂性测度的基础上,将仿真数据与试验数据之间的综合复杂性差异作为评价指标,衡量模型的有效性,步骤如下:

1) 计算仿真数据的综合复杂性测度:

$$C(\text{sim}) = \frac{C_{\text{kc}}(\text{sim}) + C_0(\text{sim}) + \text{ApEn}(\text{sim})}{3} \quad (12)$$

2) 计算试验数据的综合复杂性测度:

$$C(\text{real}) = \frac{C_{\text{kc}}(\text{real}) + C_0(\text{real}) + \text{ApEn}(\text{real})}{3} \quad (13)$$

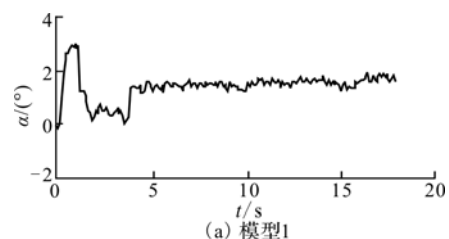
3) 比较仿真模型与试验结果的一致性,计算模型有效性指标(Model Validity Index, MVI):

$$\text{MVI} = |C(\text{sim}) - C(\text{real})| / C(\text{real}) 100\% \quad (14)$$

MVI 越小,则表明仿真模型与试验结果的一致性越好。

3 仿真分析

以某型飞行器制导控制系统的仿真模型验证为例,根据建模与仿真方案,分别建立了 2 个仿真模型。在气动数据处理方面,模型 1 采用了线性插值方法建立气动参数的解析表达式,而模型 2 采用多项式拟合方法。另外两者在进行控制系统设计时,所选特征点有所不同。攻角(α)是反映某型飞行器动态性能的重要参数之一,这里以攻角为研究对象,基于复杂性测度理论分析在相同初始条件下各可选仿真模型输出与实际系统输出的一致性。图 1 为 2 种仿真模型在某次仿真试验中的攻角曲线,图 2 为实际曲线。



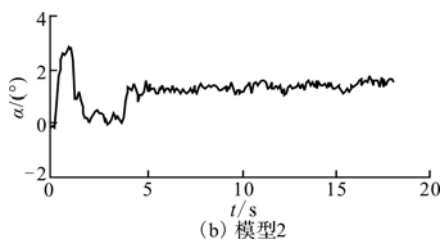


图 1 仿真模型攻角输出

Fig. 1 Attack angle output of simulation

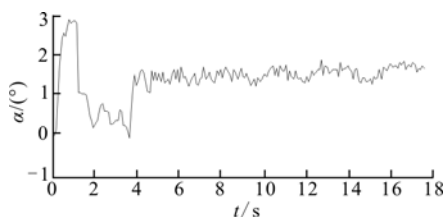


图 2 实际系统攻角输出

Fig. 2 Attack angle output of actual system

由图 1 和图 2 可看出攻角时间序列具有非线性、非平稳和随机性等特点,且序列中夹杂着一定程度的噪声。因此,将复杂性测度引入到仿真模型验证中,分别计算试验数据和仿真数据复杂性测度。3 种复杂度计算方法中,都需要设置相关参数,具体如下:KC 复杂度,选取段数为 $L=4$; C_0 复杂度,令 $r=10$; ApEn,令 $m=2$, $r=0.1$ 。则仿真模型和实际系统攻角时间序列的复杂度计算结果见表 1。

根据基于复杂性测度的仿真模型验证的基本原理,由式(12)、式(13)分别计算试验数据和仿真数据的综合复杂度,由式(14)计算 MVI,见表 1。

表 1 复杂度计算结果

Tab. 1 Result of complexity calculation

系统	KC	C_0 复杂度	ApEn	综合复杂度	MVI/%
仿真模型 1	0.663 8	0.888 3	0.724 5	0.758 9	5.468 0
仿真模型 2	0.563 2	0.890 9	0.741 8	0.732 0	8.815 5
实际系统	0.703 6	0.893 2	0.811 5	0.802 8	

MVI 越小,则仿真模型与实验结果一致性越好,显然,相对于实际系统,可选仿真模型 1 比模型 2 更为有效,其原因在于气动数据较为复杂,很难用多项式来精确描述。同时,多项式拟合的基函数的确定对拟合精度的影响较大,而在实际操作中,基函数的选取往往依赖于经验和人为试凑,相比之下,线性插值所得结果更接近真实值。另外,特征点选取是影响控制效果的重要因素,设计控制系统时特征点的选取也可能是模型 1 更为有效的因素之一。

4 结语

模型验证方法对提高模型的可信性至关重要。为有效解决复杂非线性系统的模型验证问题,针对复杂性测度能反映非线性时间序列特征的特点,研究了基于复杂性测度理论的仿真模型验证方法。通过计算试验数据和仿真数据的复杂性测度,并将两者之间的复杂性差异作为评价指标,检验仿真模型与实际系统的一致性。该方法从系统全局来分析仿真和试验数据的差异,不需要对数据进行预处理,对样本的分布没有要求,不引入人为主观判断因素,能够有效处理有噪声、复杂非线性的样本的数据,不失为一种较为有效的模型验证方法。但论文的研究还不够深入,比如将试验数据和仿真数据复杂性测度

的差异作为衡量模型有效性的评价指标,其理论依据不充分,有待于继续深入研究。

参考文献(References):

- [1] Sargent R G. Verification and validation of simulation models[C]//Proceedings of the 2011 winter simulation conference. Orlando, FL: IEEE press, 2011: 183-184.
- [2] Cor Jacques Kat, Pieter Schalk Els. Validation metric based on relative error[J]. Mathematical and computer modelling of dynamical systems, 2012, 18(5): 487-520.
- [3] Jacek Marczyk, Balachandra Deshpande. Measuring and tracking complexity in science[J]. Inter journal complex systems, 2006, 12(3): 101-107.
- [4] Hong H, Liang M. Fault severity assessment for rolling element bearings using the lempel-ziv complexity and continuous wavelet transform [J]. Journal of sound and vibration, 2009, 320(1/2): 452-468.
- [5] 庄建军, 宁新宝, 邹鸣, 等. 两种熵测度在量化射击运动员短时心率变异性信号复杂度上的一致[J]. 物理学报, 2008, 57(5): 2805-2812.

(下转第 36 页)

- [4] 姜毅. 独立自排导垂直发射技术燃气流场数值计算及试验研究[D]. 北京:北京理工大学,2005.
JIANG Yi. The numerical calculation and experiment research of the self-discharge vertical launching technology jet flow field[D]. Beijing: Beijing institute of technology, 2005. (in Chinese)
- [5] 刘琦,傅德彬,姜毅. 贮运发射箱内燃气射流的非定常冲击波流场数值模拟[J]. 弹箭与制导学报,2004,24(3):45-49.
LIU Qi, FU Debin, JIANG Yi. Unsteady simulation of shock wave in launcher[J]. Journal of projectiles rockets missile and guidance, 2004,24(3):45-49. (in Chinese)
- [6] 何景轩,余贞勇,孙利清. 固体火箭发动机内弹道性能仿真[J]. 固体火箭技术,2004,27(1):20-23.
HE Jingxuan, YU Zhenyong, SUN Liqing. Internal ballistic calculation of solid rocket motors [J]. Journal of solid rocket technology, 2004,27(1):20-23. (in Chinese)
- [7] Uwe R, Rosemarie M. CFD-simulation of the flow through a fluidic element[J]. Aerospace science technology,2004(4):111-123.
- [8] 赵承庆,姜毅. 气体射流动力学[M]. 北京:北京理工大学出版社,1998.
ZHAO Chenqing, JIANG Yi. Gas dynamics of jet flow [M]. Beijing: Beijing institute of technology press, 1998. (in Chinese)
- [9] 高峰,王建新,马岑睿. 高过载下固体火箭发动机长尾喷管[J]. 空军工程大学学报:自然科学版,2012,13(1):1-4.
GAO Feng, WANG Jianhui, MA Cenrui. Numerical simulation of the tail nozzle pipe of heat transfer in hgh acceleration[J]. Journal of air force engineering university:natural science edition, 2012,13(1):1-4. (in Chinese)

(编辑:田新华)

(上接第31页)

- ZHUANG Jianjun, NING Xinbao, ZOU Ming, et al. Agreement of two-entropy-based measures on quantifying the complexity of short-term heart rate variability signals from professional shooters [J]. Acta physica sinica, 2008,57(5):2805-2812. (in Chinese)
- [6] 佟春生,黄强,刘涵. 基于复杂性理论的径流时间序列动力学特征分析[J]. 系统工程理论与实践,2004,9(9):102-108.
TONG Chunsheng, HUANG Qiang, LIU Han. Analysis on runoff time series dynamics character based on complexity theory [J]. System engineering & practice, 2004,9(9):102-108. (in Chinese)
- [7] 高清维,李川奇,庄镇泉. 齿轮箱振动信号的复杂度分析[J]. 电子测量与仪器学报,2002,16(2):1-4.
GAO Qingwei, LI Chuanqi, ZHUANG Zhenquan. Complexity analysis of gearbox vibration signal[J]. Journal of electronic measurement and instrument, 2002,16(2):1-4. (in Chinese)
- [8] 张佃中. Lempel-Ziv 复杂度算法中粗粒化方法分析及改进[J]. 计算物理,2008,25(4):499-505.
ZHANG Dianzhong. Coarse graining method in Lempel-Ziv complexity arithmetic[J]. Chinese journal of computational physics, 2008,25(4):499-505. (in Chinese)
- [9] 张佃中. 非线性时间序列互信息与 Lempel-Ziv 复杂度的相关性研究[J]. 物理学报,2007,56(6):3152-3158.
ZHANG Dianzhong. Research on the correlation between the mutual information and Lempel-Ziv complexity of nonlinear time series[J]. Acta physica sinica, 2007,56(6):3152-3158. (in Chinese)
- [10] 沈恩华,蔡志杰,顾凡及. C_0 复杂度的数学基础[J]. 应用数学和力学,2005,26(9):1083-1090.
SHEN Enhua, CAI Zhijie, GU Fanji. Mathematical foundation of C_0 complexity[J]. Applied mathematics and mechanics,2005,26(9):1083-1090. (in Chinese)
- [11] 蔡志杰,孙洁. 改进的 C_0 复杂度及其应用[J]. 复旦学报:自然科学版,2008,47(6):791-797.
CAI Zhijie, SUN Jie. Modified C_0 complexity and applications[J]. Journal of fudan university:natural science edition,2008,47(6):791-797. (in Chinese)
- [12] 何航程,沈模卫,陈硕,等. 认知神经科学领域脑电复杂度测度方法的新进展[J]. 应用心理学,2002,8(3):51-55.
HE Hangcheng, SHEN Mowei, CHEN Shuo, et al. Recent development of the complexity analysis of EEG in cognitive neuroscience[J]. Chinese journal of applied psychology,2002,8(3):51-55. (in Chinese)

(编辑:田新华)