

引用格式:杨其,陈水忠,沈淑梅,等. LSTM网络和 ARMA 模型对惯性器件随机误差预测适应性分析[J]. 电光与控制, 2018, 25(3):68-72, 91.  
YANG Q, CHEN S Z, SHEN S M, et al. Adaptability of LSTM network and ARMA modeling to random error prediction of inertial devices[J]. Electronics Optics & Control, 2018, 25(3):68-72, 91.

## LSTM 网络和 ARMA 模型对惯性器件随机误差 预测适应性分析

杨 其<sup>1a,1b</sup>, 陈水忠<sup>2</sup>, 沈淑梅<sup>2</sup>, 朱振华<sup>2</sup>

(1. 火箭军工程大学, a. 士官学院, 山东 青州 262500; b. 初级指挥学院, 西安 710025;

2. 中国航空工业集团公司洛阳电光设备研究所, 河南 洛阳 471000)

**摘 要:** 结合惯性器件随机误差研究的实际, 针对传统的基于时间序列的 ARMA 建模方法和深度学习 LSTM 网络进行了适用性和实时性对比分析, 通过获取具体型号的惯性器件输出数据设计了算例。研究认为在未做实时性要求的情况下, ARMA 建模和 LSTM 网络均可以达到较好的拟合效果, 而建立 LSTM 网络方法可以减少提取趋势项和周期项的环节; 实时在线预测情况下 LSTM 网络优势明显, 但预测精度会随时间序列缩短而明显下降, 可以在一定程度上反映噪声变化的趋势并据此对整体控制系统的滤波算法进行优化。

**关键词:** 惯性器件; 随机误差; LSTM 网络; ARMA 建模

**中图分类号:** V448 **文献标志码:** A **doi:**10.3969/j.issn.1671-637X.2018.03.015

## Adaptability of LSTM Network and ARMA Modeling to Random Error Prediction of Inertial Devices

YANG Qi<sup>1a,1b</sup>, CHEN Shui-zhong<sup>2</sup>, SHEN Shu-mei<sup>2</sup>, ZHU Zhen-hua<sup>2</sup>

(1. Rocket Force University of Engineering, a. Sergeant College, Qingzhou 262500, China; b. Primary Command College, Xi'an 710025, China; 2. Luoyang Institute of Electro-Optical Equipment, AVIC, Luoyang 471000, China)

**Abstract:** For the random error of inertial devices, a contrastive analysis was made to the applicability and real-time performance of the traditional ARMA modeling method and the popular deep learning LSTM network. A simulation example was designed by obtaining the output data of a specific inertial device. The study showed that: 1) Without the real-time performance requirement, both ARMA modeling and LSTM network can achieve accurate prediction results, and the LSTM network is advantageous since it doesn't need to extract the signal trend and cycle terms; and 2) In real-time prediction, the LSTM network has obvious advantages, but the prediction accuracy decreases with the shortening of time series, whereas the output can still reflect the trend of noise change and can be used for optimizing the filtering algorithm of the whole control system.

**Key words:** inertial device; random error; LSTM network; ARMA modeling

### 0 引言

惯性器件可以在不借助外部信息的孤立系统中直接测量比力和角速率值, 因此在飞行器、舰船、智能设备等众多领域得到了广泛应用, 同时作为一种自推算

制导方式, 惯性制导本身也不可避免地存在缺陷, 其误差会随时间不断累积并放大。依靠地面静态环境建立误差模型并标定系数, 可以在一定程度上对误差进行补偿。但地面标定环境总是与实际工作的动态环境存在一定差别, 误差模型也难以对真实误差完全补偿, 一般将两者的差距归入随机误差。

造成惯性器件产生随机误差有众多因素<sup>[1-5]</sup>, 按照目前已知的物理原因一般将其分为量化噪声、零偏不稳定性、角度随机游走、速率斜坡等。解决随机误差

收稿日期: 2018-01-03

修回日期: 2018-01-25

作者简介: 杨 其(1983—), 男, 山东菏泽人, 讲师, 博士生, 研究方向为飞行器动力学与制导。

的最根本方法是对误差来源的物理机制深入研究,提高制造工艺消除误差,但这样的做法不仅导致费效比较高,而且目前对某些惯性器件随机误差成因并未形成统一的结论。例如,惯性器件的零偏不稳定性多认为由电子元件的闪烁噪声导致,但对闪烁噪声形成机理和分析方法仍未形成确定性共识<sup>[6-7]</sup>。因此,目前对随机误差的补偿方法多基于将各种噪声作为滤波项加入系统方程进行在线实时滤波<sup>[8-9]</sup>,但滤波方法本质上是一种信号融合工具,需要建立在对实际系统描述基本准确的基础上,不加任何验证直接对滤波系统方程进行扩展极易导致对实际系统的欠拟合,使滤波难以收敛,这也是大量改进形式滤波算法难以实际应用的重要原因。

本文研究主要针对惯性器件随机误差预测问题,对比了成熟的基于时间序列的自回归-滑动平均(Auto-Regressive and Moving Average Model, ARMA)建模方法和流行的深度学习的长短时记忆(Long Short-Term Memory, LSTM)网络在该问题上的适用性和实时预测效率。

## 1 随机误差 ARMA 模型拟合局限性分析

按照时间先后顺序记录的一系列惯性器件输出的数据可以作为一个时间序列。由于惯性器件受到各种不确定因素的作用,输出的时间序列中包含各种随机误差,同时惯性器件作为实际存在的物理系统,相邻时间点上的输出会存在状态延续性,可以认为时间序列反映了系统的内在规律,并通过拟合序列建模的方法来预测系统进一步的输出。在无实时性要求情况下,常用方法为建立 ARMA 模型,即认为惯性器件时间序列  $y(n)$  包含了以下分解项

$$y(n) = f(n) + s(n) + x(n) \quad (1)$$

式中,  $f(n)$ ,  $s(n)$ ,  $x(n)$  分别表示趋势项、周期项和随机项。其中,趋势项提取可采用差分法或最小二乘法,对周期项亦可通过功率谱密度分析并差分去除。ARMA 模型主要针对随机项  $x(n)$ , 并认为当前时刻的测量值与之前的  $p$  个测量值、白噪声  $w(n)$  以及之前的  $q$  个时间平移白噪声  $w(n-1)$ ,  $w(n-2)$ ,  $\dots$ ,  $w(n-q)$  相关,即

$$x(n) = a_1 x(n-1) + a_2 x(n-2) + \dots + a_p x(n-p) + w(n) + b_1 w(n-1) + b_2 w(n-2) + \dots + b_q w(n-q) = \sum_{k=1}^p a_k x(n-k) + \sum_{k=0}^q b_k w(n-k) \quad (2)$$

但在工程应用中,惯性器件误差输出最为突出的特征表现为性能可重复性差,即存在一次通电误差和逐次通电误差,因此必须获取足够长的时间样本才能够真实反映惯性器件的特性。假设时间序列样本服从 AR(1) 模型且时间序列长度为  $N$ , 可以根据式(2)构建

$N-1$  个方程,即

$$\begin{cases} x(2) = a_1 x(1) + w(2) \\ x(3) = a_1 x(2) + w(3) \\ \vdots \\ x(N) = a_1 x(N-1) + w(N) \end{cases} \quad (3)$$

式中:  $a_1$  为 AR(1) 模型参数;  $w(n)$  为白噪声。

由一元线性回归对  $a_1$  进行求解,其估计值  $\bar{a}_1$  应服从正态分布,即

$$\bar{a}_1 \sim N\left(a_1, \frac{\sigma^2}{l_{xx}}\right) = N\left(a_1, \frac{\sigma^2(1-a_1^2)}{(N-1)\sigma^2}\right) = N\left(a_1, \frac{1-a_1^2}{N-1}\right) \quad (4)$$

式中,  $l_{xx} = (N-1)\sigma^2/(1-a_1^2)$ ,  $\sigma^2$  为白噪声分布的方差。

对式(4)中部分参数进行近似,令

$$\begin{cases} \bar{a}_1 \approx e^{-1/\bar{\tau}} \approx 1 - \frac{1}{\bar{\tau}} \\ N-1 = m\tau \quad m=1, 2, \dots \end{cases} \quad (5)$$

则  $\bar{a}_1$  服从的正态分布可表示为

$$\frac{1}{\bar{\tau}} \sim M\left(\frac{1}{\tau}, \frac{2/m}{\tau^2}\right) \quad (6)$$

按照正态分布的  $3\sigma$  准则,分布位于区间  $[u-3\sigma, u+3\sigma]$  的概率为 99.74%, 也就是说  $1/\bar{\tau}$  有极大概率在区间

$\left[\frac{1-3\sqrt{2/m}}{\tau}, \frac{1+3\sqrt{2/m}}{\tau}\right]$  范围内,近似于确定性事件,即

$$\bar{\tau} \in \left[\frac{\tau}{1+3\sqrt{2/m}}, \frac{\tau}{1-3\sqrt{2/m}}\right] \quad (7)$$

式(7)成立必须保证  $1-3\sqrt{2/m} > 0$ , 即  $m > 18$ , 同时结合式(5)可得时间序列长度  $N$  最小值应为  $18\tau + 1$ , 此时  $\bar{\tau} \in [0.5067\tau, 37.7358\tau]$ , 即利用当前长度的时间序列  $N$  进行的估计仍存在 50% 以上的误差。一般  $N$  的取值至少为  $N > 100\tau$ , 此时  $\bar{\tau} \in [0.7021\tau, 1.7370\tau]$ , 估计误差仍大于 30%。较为稳妥的时间序列长度应大于 1000 倍的相关时间, 此时  $\bar{\tau} \in [0.8817\tau, 1.1548\tau]$ , 相应的估计误差在 11.83% ~ 15.48% 范围内。

以某陀螺仪上电输出为例进行时间序列分析, 获取长度为 500 的时间序列(第 1 组), 如图 1 所示。

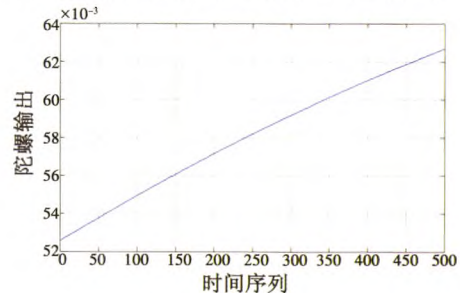


图1 陀螺输出序列样本

Fig. 1 Original series of gyro output

对获取时间序列逐步回归提取趋势项,其趋势项



表达式为

$$f(t)=59.3652+68.7115t-8.3752t^{-2}-766.3564t^{1/2}+667.4568t^{-1/2}+86.2657e^{-t}+675.354\ln t。$$
 (8)

提取趋势项后的样本序列如图 2 所示。

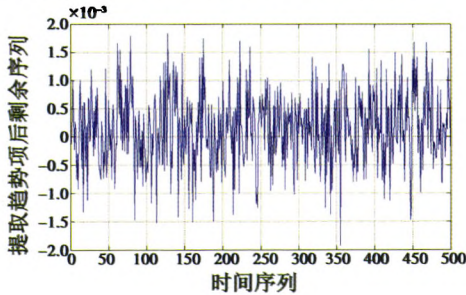


图 2 提取趋势项后序列样本

Fig. 2 Time series after extracting the trend term

经功率谱检验,序列中存在的周期项并不显著。对其进行 ARMA 建模,分别求解自相关函数和偏自相关函数,如图 3 所示。

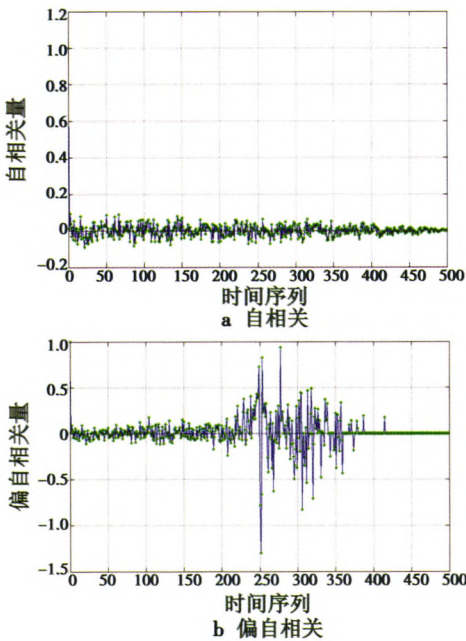


图 3 自相关和偏自相关函数(第 1 组)

Fig. 3 Autocorrelation function and the partial autocorrelation function (Group 1)

时间序列应识别为一阶 MA 模型,参数估计值为  $b_1=0.7318$ 。

在同样试验条件下,对同一个陀螺再次上电获取长度为 500 的时间序列(第 2 组),提取趋势项为

$$f(t)=65.3415+67.2534t-6.6743t^{-2}-638.2364t^{1/2}+681.2591t^{-1/2}+90.3002e^{-t}+585.5673\ln t$$
 (9)

其自相关函数和偏自相关函数如图 4 所示。

时间序列识别为一阶 MA 模型,参数估计值为  $b_1=0.5671$ 。

两组时间序列拟合后的残差均近似为白噪声,对

比如图 5 所示。

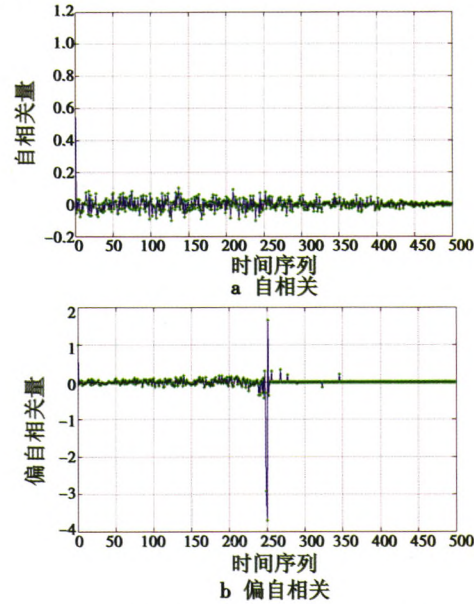


图 4 自相关和偏自相关函数(第 2 组)

Fig. 4 Autocorrelation function and the partial autocorrelation function (Group 2)

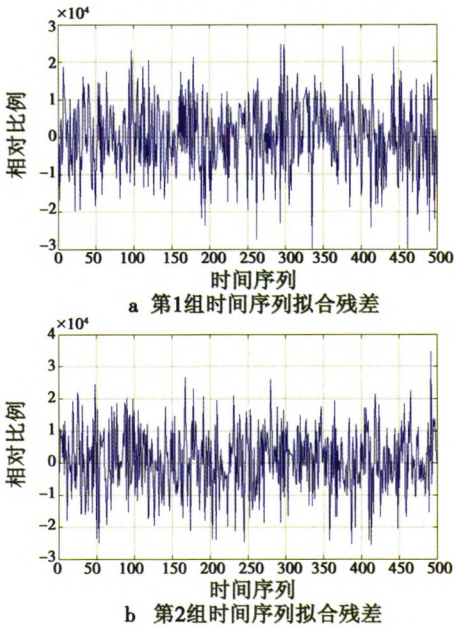


图 5 拟合残差(第 1 组和第 2 组)

Fig. 5 The fitting residuals (Group 1 and Group 2)

通过算例对比可以看到,依据不同样本序列对同一个惯性器件建立的 ARMA 模型在趋势项和模型参数上仍差别较大。由于算例直接获取的时间序列并不长,有可能导致周期性误差未能完全体现,但算例对比足以说明在惯性器件逐次通电误差影响下,即使 ARMA 模型构建的十分精确,使用该模型对另一组数据的预测效果也并不理想,同时,模型建立的时间序列数据在静态环境中获取,用其拟合动态环境激励出的随机误差结果将更不理想。



## 2 随机误差预测 LSTM 网络构建

通过分析可以看到,由于惯性器件本身存在逐次通电误差,时间序列样本相当于在随机过程中提取了一个片段,而且该随机过程的平稳性和遍历性无法确定,对“片段数据”进行高精度的 ARMA 建模极易导致过拟合,使模型泛化能力下降。而提高预测精度需要付出巨大的时间成本,并集中体现在时间序列样本获取长度上,因此难以应用于实时性要求高的场合。但借鉴其思路,如果能够建立具体的预测模型替代为建立深度学习网络,事先利用大量的标签数据对其进行训练,不仅可以解决 ARMA 模型泛化能力不够的问题,而且训练完善的网络在实际使用时具有较高的实时性。

惯性器件作为实际物理系统,当前时刻输出与相邻时刻的输出必然存在一定的相关性,变化规律并不像公共场合客流数据一样存在无穷方参与博弈,出现无规则跳变,因此对惯性器件输出序列的预测采用 LSTM 网络是一种合适的构建形式,其基本组成单元如图 6 所示。

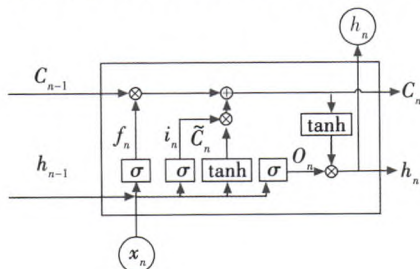


图 6 LSTM 网络基本单元

Fig. 6 Basic unit of LSTM network

LSTM 网络是循环神经网络 (RNN) 的一种<sup>[10-11]</sup>, 每个基本单元输入  $x_n$  与前一单元输出  $h_{n-1}$  连接后通过线性单元  $\sigma$  (Sigmoid 函数运算) 转化为区间  $[0, 1]$  之间的系数  $f_n$ , 同样  $\tanh$  单元表示将数值映射到区间  $[-1, 1]$  之间, 这一部分的特征映射表达式为

$$\begin{cases} f_n = \sigma(W_f \cdot [h_{n-1}, x_n] + b_f) \\ i_n = \sigma(W_i \cdot [h_{n-1}, x_n] + b_i) \\ \tilde{C}_n = \tanh(W_c \cdot [h_{n-1}, x_n] + b_c) \end{cases} \quad (10)$$

式中,  $W_f, b_f, W_i, b_i, W_c, b_c$  为待训权重。

同时, 前一个基本单元传递系数  $C_{n-1}$  与式 (10) 的计算结果  $\tilde{C}_n$  作线性叠加, 即

$$C_n = f_n \cdot C_{n-1} + i_n \cdot \tilde{C}_n \quad (11)$$

式 (11) 实际上包含了“遗忘门”的设计, 通过线性叠加确定本次输入信息的权重和对之前信息的遗忘比例<sup>[12-14]</sup>。遗忘门控单元的设计主要用于控制数据长期依赖性信息的流动, 近似于滑动平均的思想: 可令  $\alpha h_{n-1} + (1 - \alpha) h_n \rightarrow h_n$ , 如果  $\alpha$  接近于 1, 则可以记忆久

远地传递信息, 而  $\alpha$  接近于 0, 则完全遗忘, 防止训练中出现由长期依赖性导致的梯度消失。

基本单元的输出向外的传递系数  $C_n$  再次经过一个遗忘门并向同层和其他层传递, 即

$$\begin{cases} O_n = \sigma(W_o \cdot [h_{n-1}, x_n] + b_o) \\ h_n = O_n \cdot \tanh C_n \end{cases} \quad (12)$$

将基本单元按照问题的规模和预测输出的长度连接为拓扑结构即可完成网络主要部分的搭建。

## 3 LSTM 网络预测效果对比

LSTM 网络实现采用 Keras 序贯模型 (Sequential), 增加多个 LSTM 网络层接收输入信息并作为隐藏层训练。针对随机误差实时估计, 深度学习网络设计目的在于对同一型号的惯性器件能够根据在线敏感到的数据, 依靠训练完成的神经网络实时估算即将出现的随机误差, 因此在 LSTM 网络输出时, 应根据需要估算的随机误差序列长度的规模设计一个全连接层 (Dense) 来连接隐藏层和输出层<sup>[15-18]</sup>。

首先对某陀螺仪获取 1000 组序列长度为 420 个时序点的标签样本, 在这些样本中既包含多次通电获取到的序列, 也包含在一次通电过程中对陀螺仪输出截取获得的序列样本。保留样本的后 10 组作为预测数据, 前 990 组作为训练样本, 首先设计 5 层 LSTM 网络进行训练, 输入数据按照 Keras 三维格式组织并全部标准化为区间  $[0, 1]$  之间数值, 选择平均绝对误差百分比 (Mean Absolute Percentage Error, MAPE) 作为误差比较值。

训练完成后对预测数据的 MAPE 值为 0.2708, 拟合结果与实际数据对比如图 7 所示。

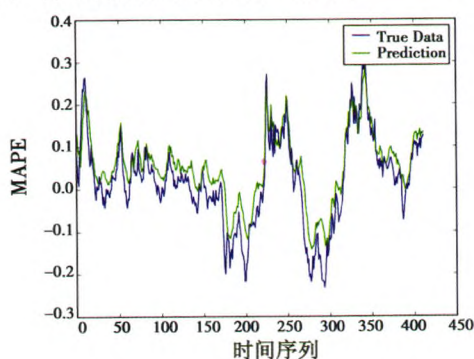


图 7 预测拟合效果对比 (5 层网络)

Fig. 7 Comparison of the predicted results of LSTM (5 layers) networks

LSTM 网络如需满足惯性器件在线预测的要求, 必须具备较高的实时性并能依据较短时间序列完成预测的能力。为了验证网络的实时性, 对网络输入输出规模修改后进行在线预测仿真, 训练完成后预测结果与实际数据对比如图 8 所示。



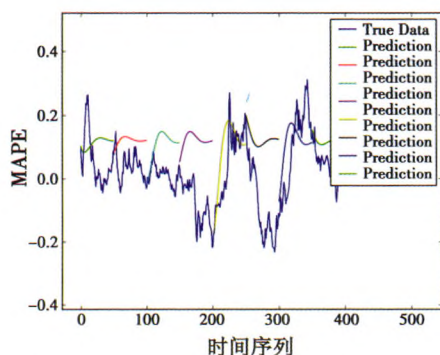


图 8 在线实时预测效果对比(5 层网络)

Fig. 8 Online predicted results of LSTM(5 layers) networks

可以看到,由于样本序列缩短却需要预测更远的输出,LSTM 预测能力出现退化而且预测结果存在延迟。为了改善结果,尝试构建更深的隐藏层,进一步增加 LSTM 层数至 15 层,引入 Dropout 机制概率为 0.1,多次对超参数进行调整后,网络对测试样本的拟合以及在线预测结果如图 9 所示。

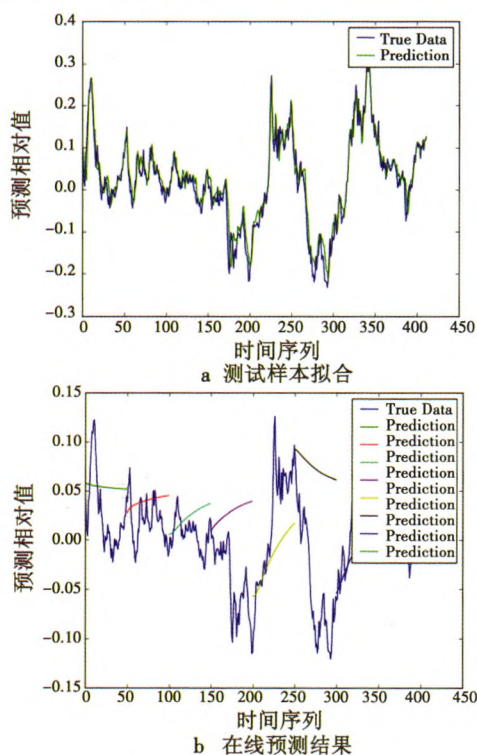


图 9 15 层 LSTM 网络预测效果

Fig. 9 Predicted results of LSTM(15 layers) networks

可以看到,在采用更多 LSTM 层数时,拟合和预测精度得到了一定的改善。但在线预测过程中由于时间序列规模缩小,预测准确率下降明显,当然准确率下降的原因仍有训练样本不够庞大和参数设置的原因。但通过观察图 8、图 9,可以看到预测对噪声变化趋势仍保证相当的契合度,仍可以将这个结果作为有色噪声范围对滤波算法进行优化设计。

## 4 结论

对惯性器件实际应用环境而言,ARMA 建模可以提取出惯性器件输出变化的趋势项、周期项并建立其中随机变化部分的自回归滑动平均模型,但需要较长的时间样本和一定规模的计算复杂度,因此方法多应用于不计时效比和实时性的设计、精度验证和工艺提高阶段;而深度学习的 LSTM 网络通过训练,在样本拟合预测中同样可以达到较好的效果,而且减少了 ARMA 方法中趋势项和周期项判断,在线实时预测中 LSTM 网络预测能力会随着时间序列的缩短出现较大的退化,但可以通过叠加更多层 LSTM 并减小预测输出的规模使预测更为精准,再根据其输出对控制系统的滤波算法进一步优化设计。

## 参考文献

- [1] RAÚL D C, LUIS C, JESÚS S. Guidance and control for high dynamic rotating artillery rockets[J]. Aerospace Science and Technology, 2017, 64(5):204-212.
- [2] MICHAEL A, BLENDER, HEIDI W. Flight dynamics of a hypersonic vehicle during inlet un-start[C]//The 16th AIAA/DLR/DGLR International Space Planes and Hypersonic Systems and Technologies Conference, 2009; AIAA-2009-7292.
- [3] BO W, ZHI H D, CHENG L, et al. Estimation of information sharing error by dynamic deformation between inertial navigation system[J]. IEEE Transactions on Industrial Electronics, 2014, 61(4):2015-2023.
- [4] SAVAGE P G. Analytical modeling of sensor quantization in strapdown inertial navigation error equations[J]. Journal of Guidance, Control and Dynamics, 2002, 25(5):833-842.
- [5] 张瑞民,杨其,魏诗卉,等.捷联惯测组合标定及误差补偿技术[M].北京:国防工业出版社,2017.
- [6] STHAL F, DEVEL M, IMBAUD J, et al. Study on the origin of  $1/f$  noise in quartz resonators[J]. Journal of Statistical Mechanics Theory & Experiment, 2016(5):054025.
- [7] JONES B K. Electrical noise as a measure of quality and reliability in electronic devices. [J] Advances in Electronics and Electron Physics, 1993, 87(1):201-257.
- [8] BROWN R G, HWANG P Y C. Introduction to random signals and applied Kalman filtering[M]. 3rd ed. New York: John Wiley & Sons, Inc., 1997.
- [9] 韩松来. GPS 和捷联惯导组合导航新方法及系统误差补偿方案研究[D].长沙:国防科学技术大学,2010.

(下转第 91 页)

各自得到的曲线相比较,全局快速终端滑模控制算法可在更短的收敛时间内运行更高的高度。对比文献[5]可以看出,本文的控制算法无切换项,比传统滑模算法更加有效地消除了抖振,受到风扰时曲线不会产生大的波动影响;与文献[10]采用的 PD 控制方法相比,本文算法得到的6条位姿曲线图的线条更加平滑。与其他文献比对后可得,本文采用的全局快速终端滑模控制器具有较强的鲁棒性与可靠性,适用于四旋翼无人机控制系统的设计。

## 4 结论

本文分析了四旋翼无人机的飞行控制原理与在风场作用下的受力情况,建立四旋翼无人机在紊流风场作用下的动力学模型,采用可自动消除抖振的全局快速终端滑模的控制算法计算各个控制律,并引入过渡输入控制律实现欠驱动耦合控制。通过 Matlab/Simulink 软件的仿真结果表明,该控制器算法具有可靠性高、收敛速度较快和鲁棒性较强的特点,能够有效地实现四旋翼无人机的飞行任务。

## 参考文献

- [1] 王贞琪,马洁. 四旋翼几种控制方法比较[J]. 北京信息科技大学学报,2016,31(2):35-39.
- [2] 孟佳东,赵志刚. 小型四旋翼无人机建模与控制仿真[J]. 兰州交通大学学报,2013,32(1):63-57.
- [3] 江杰,冯旭光,苏建彬. 四旋翼无人机仿真控制系统设计[J]. 电光与控制,2015,22(2):27-30.
- [4] 牛洪芳,吴怀宇,陈洋. 基于反步法的四旋翼飞行器滑动模态控制[J]. 高技术通讯,2016,25(12):1083-1090.
- [5] 许喆. 基于 SMC 的四旋翼无人机抗风扰研究[J]. 电光与控制,2017,24(1):67-71.
- [6] 王小莉. 面向桥梁检测的四旋翼飞行器控制系统研究[D]. 重庆:重庆交通大学,2013.
- [7] NIKOLA G S, ANDON V T, OKYAY K. Comparative results on stabilization of the quad-rotor rotorcraft using bounded feedback controllers[J]. Journal of Intelligent & Robotic Systems, 2012, 65(1):389-408.
- [8] 孙妙平,刘静静,年晓红. 基于区间矩阵的四旋翼无人机鲁棒跟踪控制[J]. 控制理论与应用,2017,34(2):1-11.
- [9] BENALLEGUE A, MOKHTARI A, FRIDMAN L. High-order sliding-mode observer for a quadrotor UAV[J]. International Journal of Robust and Nonlinear Control, 2008, 18(3):427-440.
- [10] 吴翰文. 四旋翼飞行器抗风控制研究[D]. 哈尔滨:哈尔滨工业大学,2016.
- [11] GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to forget: continual prediction with LSTM[J]. Neural Computation, 2000, 12(10):2451-2471.
- [12] HOCHREITER S, BENGIO Y, FRASCONI P. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies[J]. Field Guide to Dynamical Recurrent Networks, 2001, 28(2):237-243.
- [13] BENGIO S, BENGIO Y. Taking on the curse of dimensionality in joint distributions using neural networks[J]. IEEE Transactions on Neural Networks, Special Issue on Data Mining and Knowledge Discovery, 2000, 11(3):550-557.
- [14] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[J]. Computer Science, 2014(4):310-311.
- [15] LIPTON Z C, BERKOWITZ J. A critical review of recurrent neural networks for sequence learning[J]. Computer Science, 2015(1):104-112.
- [16] HASIM S, ANDREW S, FRANCOISE B. Long short-term memory current neural networks architecture for large scale acoustic modeling[C]//Proceeding of Interspeech, ISCA, 2014:338-342.
- [17] GRAVE A. Generating sequences with recurrent neural networks[J]. Computer Science, 2013(1):234-243.
- [18] PALANGI H, DENG L, SHEN Y, et al. Deep sentence embedding using long short-term memory networks: analysis and application to information retrieval[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2015, 24(4):694-707.

(上接第72页)

- [10] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [11] GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to forget: continual prediction with LSTM[J]. Neural Computation, 2000, 12(10):2451-2471.
- [12] HOCHREITER S, BENGIO Y, FRASCONI P. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies[J]. Field Guide to Dynamical Recurrent Networks, 2001, 28(2):237-243.
- [13] BENGIO S, BENGIO Y. Taking on the curse of dimensionality in joint distributions using neural networks[J]. IEEE Transactions on Neural Networks, Special Issue on Data Mining and Knowledge Discovery, 2000, 11(3):550-557.

- [14] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[J]. Computer Science, 2014(4):310-311.
- [15] LIPTON Z C, BERKOWITZ J. A critical review of recurrent neural networks for sequence learning[J]. Computer Science, 2015(1):104-112.
- [16] HASIM S, ANDREW S, FRANCOISE B. Long short-term memory current neural networks architecture for large scale acoustic modeling[C]//Proceeding of Interspeech, ISCA, 2014:338-342.
- [17] GRAVE A. Generating sequences with recurrent neural networks[J]. Computer Science, 2013(1):234-243.
- [18] PALANGI H, DENG L, SHEN Y, et al. Deep sentence embedding using long short-term memory networks: analysis and application to information retrieval[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2015, 24(4):694-707.