



基于混合神经网络和注意力机制的混沌时间序列预测

黄伟建 李永涛 黄远

Prediction of chaotic time series using hybrid neural network and attention mechanism

Huang Wei-Jian Li Yong-Tao Huang Yuan

引用信息 Citation: *Acta Physica Sinica*, 70, 010501 (2021) DOI: 10.7498/aps.70.20200899

在线阅读 View online: <https://doi.org/10.7498/aps.70.20200899>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

基于遗传算法优化卷积长短记忆混合神经网络模型的光伏发电功率预测

A hybrid model for photovoltaic power prediction of both convolutional and long short-term memory neural networks optimized by genetic algorithm

物理学报. 2020, 69(10): 100701 <https://doi.org/10.7498/aps.69.20191935>

基于大脑情感学习模型和自适应遗传算法的混沌时间序列预测

Chaotic time series prediction based on brain emotional learning model and self-adaptive genetic algorithm

物理学报. 2018, 67(8): 080502 <https://doi.org/10.7498/aps.67.20172104>

基于鲁棒极端学习机的混沌时间序列建模预测

Chaotic time series prediction based on robust extreme learning machine

物理学报. 2018, 67(3): 030501 <https://doi.org/10.7498/aps.67.20171887>

基于深度卷积神经网络的大气湍流相位提取

Extracting atmospheric turbulence phase using deep convolutional neural network

物理学报. 2020, 69(1): 014209 <https://doi.org/10.7498/aps.69.20190982>

变频正弦混沌神经网络及其应用

Frequency conversion sinusoidal chaotic neural network and its application

物理学报. 2017, 66(9): 090502 <https://doi.org/10.7498/aps.66.090502>

基于经验知识遗传算法优化的神经网络模型实现时间反演信道预测

Prediction of time reversal channel with neural network optimized by empirical knowledge based genetic algorithm

物理学报. 2019, 68(17): 170503 <https://doi.org/10.7498/aps.68.20190327>

基于混合神经网络和注意力机制 的混沌时间序列预测*

黄伟建 李永涛[†] 黄远

(河北工程大学信息与电气工程学院, 邯郸 056038)

(2020年6月12日收到; 2020年7月19日收到修改稿)

为提高混沌时间序列的预测精度, 提出一种基于混合神经网络和注意力机制的预测模型 (Att-CNN-LSTM), 首先对混沌时间序列进行相空间重构和数据归一化, 然后利用卷积神经网络 (CNN) 对时间序列的重构相空间进行空间特征提取, 再将 CNN 提取的特征和原时间序列组合, 用长短期记忆网络 (LSTM) 根据空间特征提取时间特征, 最后通过注意力机制捕获时间序列的关键时空特征, 给出最终预测结果. 将该模型对 Logistic, Lorenz 和太阳黑子混沌时间序列进行预测实验, 并与未引入注意力机制的 CNN-LSTM 模型、单一的 CNN 和 LSTM 网络模型、以及传统的机器学习算法最小二乘支持向量机 (LSSVM) 的预测性能进行比较. 实验结果显示本文提出的预测模型预测误差低于其他模型, 预测精度更高.

关键词: 混沌时间序列, 卷积神经网络, 长短期记忆网络, 注意力机制

PACS: 05.45.-a, 05.45.Tp, 42.79.Ta

DOI: 10.7498/aps.70.20200899

1 引言

非线性科学的研究涉及到社会科学和自然科学的各个领域, 如金融、电力、水文和天气等. 这些呈现出混沌现象的动力学系统中, 任一变量随时间演化而产生的混沌时间序列, 都蕴含着系统丰富的动力学信息, 而这些信息往往包含着具有重要价值的内容. 对混沌时间序列的研究, 就是要将其丰富的动力学信息提取出来, 并应用到实际中去. 在实际运用中, 对混沌时间序列进行预测一直以来都是研究的热点. 文献 [1] 提出一种分数阶最大相关熵的自适应滤波算法, 在混沌时间序列的预测中可以减少噪声影响, 预测结果准确率有所提升. 文献 [2] 提出了一种较为新颖的混沌时间序列预测方法, 使用自适应遗传算法优化大脑情感学习, 该方法能够

有效地反映混沌时间序列的发展趋势, 预测结果也优于传统神经网络. 文献 [3] 提出一种贝叶斯框架下的鲁棒极端学习机, 利用高斯混合分布和变分法使预测模型可以对异常点有较好的鲁棒性. 文献 [4] 将改进的内核递推最小二乘算法运用于多元混沌时间序列的预测, 并证明了该预测方法的有效性. 文献 [5] 提出使用向量投影来改进传统的灰色关联分析模型, 能够在多元混沌时间序列预测中减少输入维度, 并提高预测精度. 为了降低混沌时间序列的非平稳性和非线性带来的预测误差, 文献 [6] 提出使用集合经验模态分解和奇异谱分析对混沌时间序列进行分解, 该方法在风能预测中取得了较好的表现.

近年来, 随着深度学习技术和硬件技术发展, 已有研究人员将深度学习算法运用到混沌时间序列的预测中, 文献 [7] 用长短期记忆网络 (long

* 河北省自然科学基金 (批准号: F2015402077) 和河北省高等学校科学技术研究项目 (批准号: QN2018073) 资助的课题.

[†] 通信作者. E-mail: lyotard@163.com

short-term memory, LSTM) 搭配合适的组合策略, 验证了在低维情况下 LSTM 在混沌时间序列预测上能够取得较好的效果; 文献 [8] 则进一步验证了使用 LSTM 预测混沌时间序列不仅能够取得较好的预测效果, 且预测模型具有较好的鲁棒性. 结合混沌理论和经验模态分解, 文献 [9] 使用 LSTM 搭建预测模型, 取得了可靠性高的预测结果, 且预测模型具有一定的泛化能力. 文献 [10] 利用卷积神经网络 (convolutional neural networks, CNN) 训练连续和离散的低维混沌时间序列生成一个分类器, 使其能够识别并分类高维状态下的混沌时间序列.

相比最小支持向量机 (LSSVM)^[11,12]、极端学习机^[3,13] 等传统机器学习算法, LSTM 和 CNN 等深度学习算法具有更强大的学习能力, 在处理回归问题上, 对非线性数据具有很好的逼近能力. 因此, 为了进一步提高混沌时间序列的预测精度, 本文结合混沌系统的相空间重构理论, 提出一种基于深度学习算法的混沌时间序列混合预测模型 (Att-CNN-LSTM). 首先, 将获取的时间序列进行数据预处理, 即重构相空间和数据归一化. 然后提出利用注意力机制捕获 CNN 的相空间特征和 LSTM 的时间特征, 并对输出结果进行权衡, 给出预测结果. 最后将提出的预测模型分别应用于 Logistic、Lorenz 和太阳黑子混沌时间序列, 仿真实验结果表明本文提出的预测模型可以有效地对混沌时间序列进行预测, 且与其他预测方法相比, Att-CNN-LSTM 模型具有更高的预测精度.

2 数据预处理

在时间序列的分析中, 影响序列的因素很多, 而且在动力系统中, 这些影响因素的动力学方程往往都是非线性的, 更有可能是混沌的^[14]. 对于表现出混沌特性的时间序列, 对序列进行重构相空间是研究序列所在动力系统状态的重要方法. 因此, 在搭建和训练预测模型之前要对混沌时间序列进行相空间重构, 此外, 为消除量纲, 提升模型的收敛速度和预测精度, 再对原序列和重构的相空间数据进行归一化处理.

2.1 重构相空间

重构相空间理论的出现为混沌时间序列的预

测提供了理论基础, 在重构相空间的基本思想中, 系统中任一变量的变化是由相互作用着的其他变量所决定的, 因此任一变量的发展变化都蕴含着其他变量发展变化的信息^[14]. 根据 Packard 等^[15] 提出的理论, 使用动力系统中某一变量的延迟坐标即可重构相空间, Takens^[16] 则证明了在合适的嵌入维下, 即可恢复原动力系统的维数. 为实现相空间的重构, 使用互信息法 (mutual information)^[17] 确定延迟时间 τ , Cao 法^[18] 确定嵌入维数 m .

设有时间序列 $\{x_1, x_2, x_3, \dots, x_{N-1}, x_N\}$, 其长度为 N , 计算时间延迟 τ 和嵌入维数 m , 则相空间域数据集 $\bar{D} = \{X(t), Y(t)\}, t = 1, 2, \dots, M$, 其中 $M = N - (m-1)\tau$, $\mathbf{X}(t) = [X_t, X_{t+\tau}, \dots, X_{t+(m-1)\tau}]$, $\mathbf{Y}(t) = [X_{t+1}]$, 矩阵表示如下:

$$\mathbf{X} = \begin{bmatrix} x_1 & x_{1+\tau} & x_{1+2\tau} & \cdots & x_{1+(m-1)\tau} \\ x_2 & x_{2+\tau} & x_{2+2\tau} & \cdots & x_{2+(m-1)\tau} \\ x_3 & x_{3+\tau} & x_{3+2\tau} & \cdots & x_{3+(m-1)\tau} \\ \vdots & & & \cdots & \\ x_{M-1} & x_{M-1+\tau} & x_{M-1+2\tau} & \cdots & x_{M-1+(m-1)\tau} \end{bmatrix},$$

$$\mathbf{Y} = [x_2 \quad x_3 \quad x_4 \quad \cdots \quad x_M]^T.$$

2.2 数据归一化

在深度学习中, 对数据集尤其是对非线性的回归数据集进行归一化操作是非常有必要的预处理步骤, 因为归一化操作不仅可以消除量纲, 将数据统一到相同的尺度上, 还可以提升模型的收敛速度和预测精度. 本文利用归一化将相空间域数据集和原始混沌时间序列数据集都统一到 (0, 1) 范围之间, 归一化函数可表示为

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}. \quad (1)$$

相空间域数据集 \bar{D} 和混沌时间序列经过数据预处理后的数据集可表示为重构相空间域 D 和混沌时间序列 T . 数据预处理的过程表示如图 1 所示.

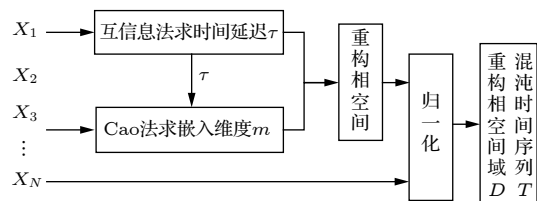


图 1 数据预处理过程

Fig. 1. The process of data preprocessing.

3 算法模型

在混沌时间序列的预测中,不同类型的神经网络具有的作用也不同. 序列的重构相空间表示了原动力系统空间特征,而序列中也蕴含着丰富的时间特征. 因此,本文提出 Att-CNN-LSTM 模型,利用 CNN 模型获取序列重构相空间的空间特征,通过 LSTM 得到空间特征下的时空特征,并采用注意力机制捕获混合神经网络提取到的关键时空特征.

本节主要介绍 Att-CNN-LSTM 模型的框架

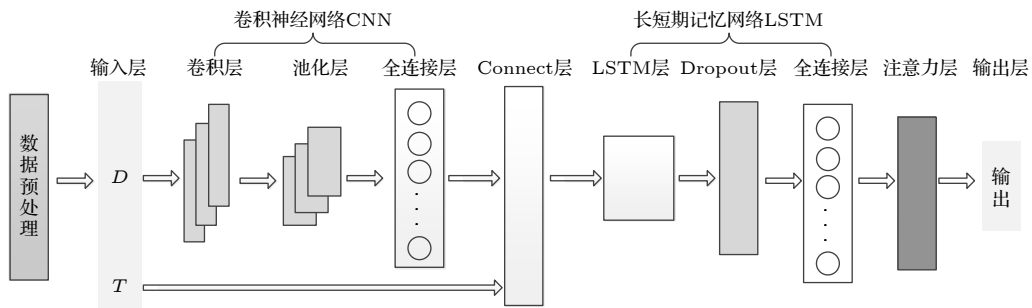


图 2 Att-CNN-LSTM 模型

Fig. 2. Att-CNN-LSTM model.

3.1 卷积神经网络 CNN

卷积网络^[19], 又称卷积神经网络, 是一种特殊的深度学习神经网络, 常用于处理具有已知网格拓扑的数据^[20], 在时间序列分析、计算机视觉和自然语言处理等领域有着广泛的运用. 根据处理数据流的不同, CNN 可以分为一维卷积、二维卷积和三维卷积, 其中一维卷积在时间序列分析和自然语言处理领域运用较多, 本文采取的 CNN 结构就属于一维卷积神经网络^[21], 如图 3 所示. 不管是何种 CNN, 其组成主要包括基本的输入层和输出层, 核心操作部分为卷积层、池化层 (亦称采样层) 和全连接层. 在一维卷积中, 卷积的作用可以理解为提取数据在某一方向的平移特征, 在时间序列的分析中则表现为提取序列的特征, 在这里卷积操作的本质是循环乘积与加和, 其数学表达如下:

$$y(k) = h(k) * u(k) = \sum_{i=0}^N h(k-i)u(i). \quad (2)$$

式中 y, h, u 均为序列, k 表示卷积次数, N 为 u 的长度.

结构, 如图 2 所示, 主要包括如下细节:

(1) 数据预处理: 先对序列进行重构相空间, 然后进行归一化处理;

(2) 卷积神经网络层: 处理网络输入的重构相空间域 D , 进行卷积操作, 获取空间特征并输入到 Concatenate 层;

(3) 长短期记忆网络层: 将 Concatenate 层融合得到的张量输入到 LSTM 中, 网络根据 CNN 提取的空间特征获取时空特征;

(4) 注意力层: 计算 CNN 和 LSTM 获取的时空特征的权重, 获得序列的完整特征表示, 根据特征计算预测数值.

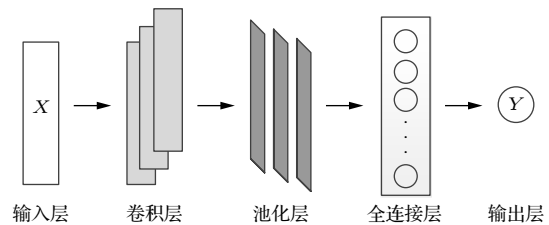


图 3 一维卷积神经网络

Fig. 3. One dimensional convolutional network.

3.2 长短期记忆网络

循环神经网络 (recurrent neural network, RNN) 是常规前馈神经网络的扩展, 能够处理可变长度序列输入. RNN 通过循环隐藏状态单元来处理可变长度序列, 但是 RNN 在训练时会出现梯度消失和梯度爆炸, 因此 RNN 很难捕获数据的长期依赖^[22]. 为解决 RNN 梯度消失的难题 Hochreiter 和 Schmidhuber^[23] 提出了长短期记忆网络, LSTM 是一种门控循环神经网络, 是 RNN 的一种特殊表现形式, 能够很好地捕获数据之间的长期依赖^[24].

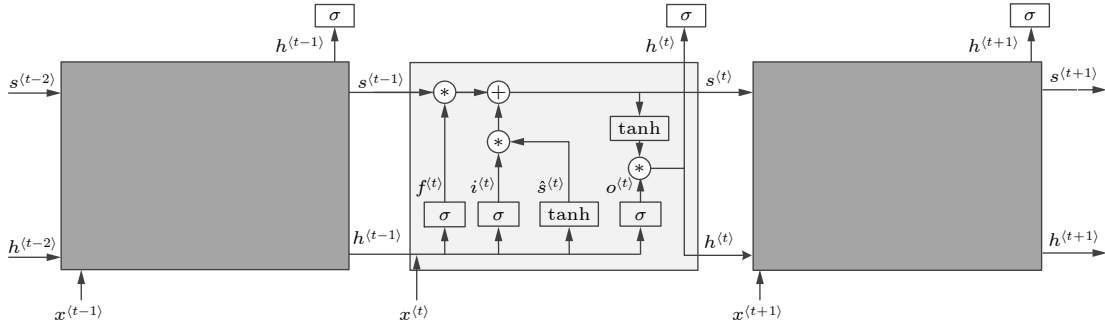


图 4 LSTM 结构

Fig. 4. The structure of LSTM.

在 LSTM 单元结构中一共有三个“门”：遗忘门、输入门和输出门，其结构如图 4 所示。在 LSTM 中，通过输入门 i 、遗忘门 f 、输出门 o 以及一个记忆单元 s 来实现长期记忆或遗忘信息，设当前时刻为 t ，则 $s^{(t-1)}$ 为前一时刻的记忆单元状态， $\hat{s}^{(t)}$ 为当前输入状态信息， $h^{(t)}$ 和 $s^{(t)}$ 为当前 LSTM 的输出值。LSTM 单元的计算公式如 (3) 式—(8) 式所示。

$$\Gamma_f = \sigma(W_f[h^{(t-1)}, x^{(t)}] + b_f), \quad (3)$$

$$\Gamma_i = \sigma(W_i[h^{(t-1)}, x^{(t)}] + b_i), \quad (4)$$

$$\Gamma_o = \sigma(W_o[h^{(t-1)}, x^{(t)}] + b_o), \quad (5)$$

$$\hat{s}^{(t)} = \tanh(W_s[h^{(t-1)}, x^{(t)}] + b_s), \quad (6)$$

$$s^{(t)} = \Gamma_f s^{(t-1)} + \Gamma_i \hat{s}^{(t)}, \quad (7)$$

$$h^{(t)} = \Gamma_o \tanh(s^{(t)}), \quad (8)$$

式中 W 为各个门的权重向量， b 为偏置向量， σ 为 sigmoid 函数， \tanh 为非线性激活函数。

3.3 注意力机制

人类大脑在观察某一事物时，往往会将注意力集中到事物的某些部分，这些注意力集中的部分往往也是从事物上获取信息的关键，这些信息对于认知同类事物有着很强的指引作用，而注意力机制 (attention mechanism)^[25] 就是一种模仿这一认知过程的特殊机制。注意力机制在计算机视觉和自然语言处理领域的运用已经取得了较好的结果^[26,27]，本文将注意力机制运用到时间序列的分析中。

在混沌时间序列的分析中，先用 CNN 提取序列的空间特征，在用 LSTM 根据空间特征提取时空特征后，过多或非关键的特征会影响最后的预测结果。因此，利用注意力机制来提取序列的关键特征。注意力机制可以简单地理解为一个加权求和器，也

可以理解为一个关键特征提取器，它主要进行的是加权求和操作。本文提出的注意力模型如图 5 所示，向量 c 即为提取的关键特征，其计算公式如下：

$$c = \sum_{i=1}^m \beta_i v_i, \quad (9)$$

式中 m 为输入 LSTM 网络的时间步总和； v 为 LSTM 网络输出的特征向量， β 为向量 v 的权值。

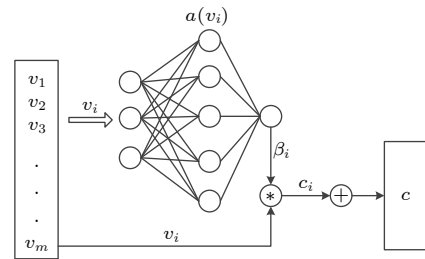


图 5 注意力模型

Fig. 5. Attention model.

为了获得权值 β ，在注意力模型中加入了一个小型的神经网络 a ，其输出层激活函数为 softmax，计算公式表示如下：

$$\beta_i = \frac{\exp(e_i)}{\sum_{k=1}^m \exp(e_k)}, \quad (10)$$

其中， e_i 可由 $e_i = a(v_i) = \sigma(Wv_i + b)$ 算得，且 σ 为 sigmoid 函数， W 是图 5 中输入层至隐藏层的权重矩阵， b 为偏置值矩阵。

3.4 其他

如图 2 所示，Att-CNN-LSTM 模型中添加了 Concatenate 层和 Dropout 层，其中 Concatenate 层是为了将卷积神经网络的输出，即序列的空间特征与原序列 T 进行特征融合。在深度学习模型中

加入 Dropout 层很常见,其作用主要是为了防止模型训练时出现过拟合,进而提升模型的鲁棒性和泛化能力.

4 实验分析与讨论

本文所有实验均在阿里云服务器 ECS 上进行, CPU 类型为 Intel Skylake Xeon Platinum 8163 2.5 GHz, 内存 8 GB, 所有程序代码均用 Python (3.7.7 版本) 语言编写, 使用的编辑器为 Jupyter notebook.

为验证 Att-CNN-LSTM 模型的预测能力, 利用 Logistic、Lorenz 和太阳黑子混沌时间序列进行实验, 并与未加入注意力机制的 CNN-LSTM 模型、单一的 LSTM 和 CNN 模型, 以及 LSSVM 模型进行误差对比. 为了评价预测误差, 除了采用常用的均方根误差 (RMSE)、平均绝对误差 (MAE) 和平均绝对百分比误差 (MAPE) 外, 再引入一个在评价时间序列预测性能中常用的均方根百分比误差 (RMSPE), 即:

$$RMSE = \sqrt{\frac{1}{I} \sum_{i=1}^I (y_i - \hat{y}_i)^2}, \quad (11)$$

$$MAE = \frac{1}{I} \sum_{i=1}^I |y_i - \hat{y}_i|, \quad (12)$$

$$MAPE = \frac{1}{I} \sum_{i=1}^I \frac{100 \times |\hat{y}_i - y_i|}{y_i}, \quad (13)$$

$$RMSPE = \sqrt{\frac{1}{I} \sum_{i=1}^I \left| \frac{y_i - \hat{y}_i}{y_i} \right|^2} \times 100, \quad (14)$$

式中 I 为预测样本数; y_i 和 \hat{y}_i 分别为预测值和真实值.

4.1 Logistic 时间序列预测

Logistic 混沌映射方程为

$$x_{n+1} = \mu x_n (1 - x_n). \quad (15)$$

当 $3 \leq \mu \leq 4$ 时 Logistic 映射的动力系统由周期转向混沌, 本文选取方程的初始值为 $x_0 = 0.32$, 参数 $\mu = 3.8$. 为保证混沌性, 舍弃前 10000 个数据, 选取后 3000 个数据作为模型数据样本. 利用互信息法和 Cao 法分别求得样本的延迟时间为 9, 嵌入维度为 15, 而后对样本进行相空间重构, 共得到 2873 个有效相点, 按照 8:2 划分训练集和测试集, 即前 80% 作为训练集后 20% 作为测试集. 数

据预处理后, 用 Att-CNN-LSTM 模型进行训练预测, 预测结果和预测相对误差如图 6 和图 7 所示, 表 1 列出了 Att-CNN-LSTM 模型和其他模型的预测结果对比.

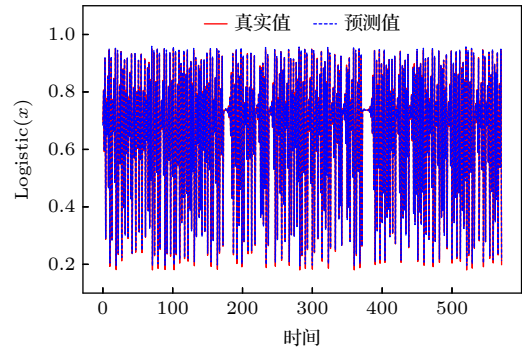


图 6 Logistic 序列预测

Fig. 6. Logistic series prediction.

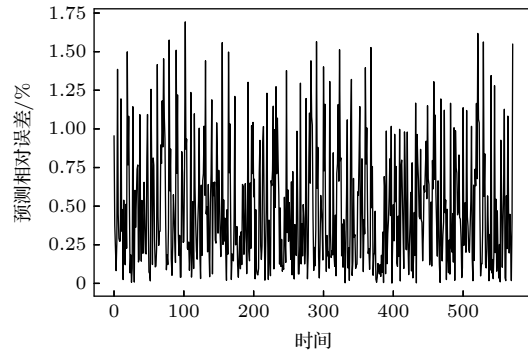


图 7 Logistic 序列预测相对误差

Fig. 7. Logistic series prediction relative error.

表 1 模型误差对比

Table 1. Model error comparison.

	RMSE	MAE	MAPE	RMSPE
Att-CNN-LSTM	0.003503	0.002935	0.5305	0.6767
CNN-LSTM	0.006856	0.005444	1.1064	1.7795
LSTM	0.006169	0.005316	1.1595	1.6887
CNN	0.004670	0.003849	0.8802	1.4019
LSSVM	0.009158	0.004307	1.3623	3.8604

从图 6 和图 7 中可以看出, Att-CNN-LSTM 对 Logistic 时间序列的预测取得了较好的结果, 不仅能够预测序列变化的趋势, 且真实值和预测值之间的误差较小, 说明本文提出的模型取得了较高的预测精度. 如表 1 所列, 与 CNN-LSTM, LSTM, CNN 和 LSSVM 等模型的预测误差对比, Att-CNN-LSTM 的各项预测误差均为最小, 说明本文提出的模型预测精度高, 预测性能较好. 与未引入注意力机制的

CNN-LSTM 预测模型相比, Att-CNN-LSTM对预测精度的提升较为明显, 说明注意力机制在模型中提取到了序列的关键时空特征, 使得预测结果更加准确. 值得注意的是, 单一的 CNN 和 LSTM模型的预测结果优于 CNN-LSTM 混合模型, 说明混合模型下的非关键特征在一定程度上降低了预测精度, 这也说明了注意力模型在 CNN-LSTM 混合模型中具有重要的作用. 此外单一的 CNN 预测模型在 Logistic 数据集上的预测表现优于单一的 LSTM 模型, 说明 CNN 在 Logistic 序列高纬度相空间下提取的空间特征好于 LSTM 提取的时间特征, 从 Att-CNN-LSTM 模型与这两个单一模型的对比中, 可以发现仅依赖单一的时间或空间特征, 难以获得精度较高的预测结果, 而注意力模型提取关键时空特征能够提升预测精度, 是一个较为理想的预测方法.

为了更全面地对比本文提出的 Att-CNN-LSTM 模型的预测性能, 表 2 中列出了各个模型的训练和预测时间. 不难看出, 在模型的训练阶段, 单一的 CNN, LSTM 和 LSSVM 预测模型所用的时间较小, 分别为 59.5, 48.8 和 215.4 s. 相比之下, 混合模型因为模型较为复杂, 所需进行的运算更多, 所以训练时间用时较多, 其中 Att-CNN-LSTM 用时 312.7 s, CNN-LSTM 用时 302 s. 但是在测试阶段, 各个模型所用时间相差不大, 虽然单一模型的预测用时仍比混合模型短, 但是都控制在 0.6 s 以内. 综合各个模型的预测误差表现, 本文提出的预测模型具有较高的性价比, 且预测优势明显.

表 2 模型运行时间对比

Table 2. Model running time comparison.

模型	Att-CNN-LSTM	CNN-LSTM	CNN	LSTM	LSSVM
训练时间 /s	312.7	302	59.5	48.8	215.4
预测时间 /s	0.53	0.49	0.25	0.21	0.47

4.2 Lorenz 时间序列预测

Lorenz 混沌映射方程为

$$\begin{aligned} \frac{dx}{dt} &= -a(x - y), \\ \frac{dy}{dt} &= -xz + cx - y, \\ \frac{dz}{dt} &= xy - bz. \end{aligned} \quad (16)$$

选取方程的初始值为 $x = y = z = 1$, 参数

$a = 10$, $c = 28$, $b = 8/3$. 为保证混沌性, 舍弃前 10000 个数据, 选取后 3000 个数据作为模型数据样本. 本文选取 Lorenz 系统的 x 分量进行预测分析, 先利用互信息法和 Cao 法分别求得样本的延迟时间为 17, 嵌入维度为 7, 而后对样本进行相空间重构, 共得 2898 个样本, 取前 80% 作为训练集, 后 20% 作为测试集. 同样在经过数据预处理后, 用 Att-CNN-LSTM 模型进行训练预测, 预测结果和预测相对误差如图 8 和图 9 所示, 表 3 列出了 Att-CNN-LSTM 模型和其他模型的预测误差对比, 表 4 则列出了各个模型的运行时间.

图 8 和图 9 所展示的 Lorenz(x) 序列预测结果很好地表现了 Att-CNN-LSTM 的预测性能, 在预

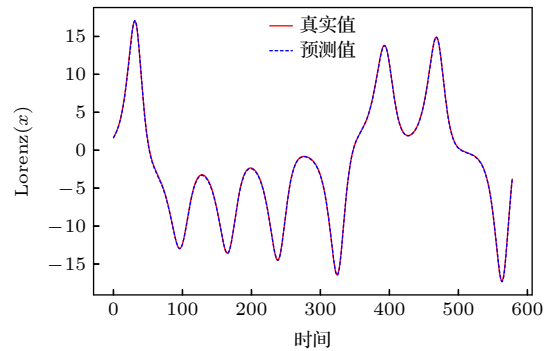


图 8 Lorenz(x) 序列预测

Fig. 8. Lorenz(x) series prediction.

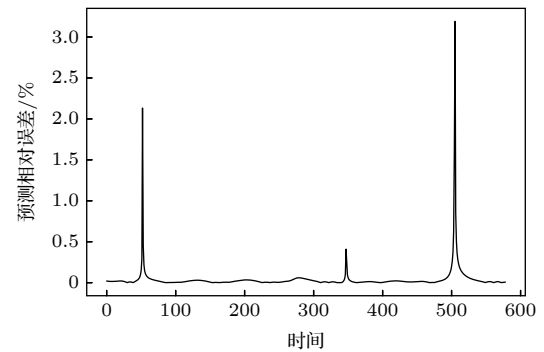


图 9 Lorenz(x) 序列预测相对误差

Fig. 9. Lorenz(x) series prediction relative error.

表 3 模型误差对比

Table 3. Model error comparison.

	RMSE	MAE	MAPE	RMSPE
Att-CNN-LSTM	0.0679	0.0521	1.2182	2.1102
CNN-LSTM	0.2445	0.1229	3.8849	14.7893
LSTM	0.5152	0.3901	13.6767	43.7676
CNN	0.5356	0.3811	11.0032	33.5251
LSSVM	0.5101	0.3652	9.3543	35.5644

测集上,模型的预测结果误差较小,精度较高,非常接近真实值.从表2中也能看出,在Lorenz样本集上本文提出的Att-CNN-LSTM模型表现依然出色,且Att-CNN-LSTM的预测性能明显优于CNN-LSTM, LSTM, CNN和LSSVM等模型, RMSE, MAE, MAPE和RMSPE四项预测指标分别降低至0.00679, 0.0521, 1.2182和2.1102, 相比其他模型精度提升较为明显.不同于在Logistic样本集上的表现, CNN-LSTM在Lorenz样本集上的表现虽然不如Att-CNN-LSTM模型,但是却优于单一的CNN和LSTM模型,各项预测指标也下降较多.在单一模型中,在RMSE指标下LSTM的预测性能略好于CNN,但在其他指标下其预测性能略低于CNN,此外, LSSVM的综合表现优于CNN和LSTM.在运行时间方面,单一的预测模型在训练和测试阶段所用时间依然低于混合模型,但时间差距有所缩小.此外,在Lorenz样本集上, Att-CNN-LSTM模型的运行时间低于CNN-LSTM模型,综合考虑下,本文提出的模型依然具有较高的预测性能.

表4 模型运行时间对比

Table 4. Model running time comparison.

模型	Att-CNN-LSTM	CNN-LSTM	CNN	LSTM	LSSVM
训练时间 /s	184.6	193.9	84.28	51.21	202.4
预测时间 /s	0.47	0.55	0.20	0.23	0.45

4.3 太阳黑子时间序列预测

在Logistic和Lorenz这两个理论混沌系统中, Att-CNN-LSTM表现出了较好的预测性能,为验证其在真实动力系统的表现,采用太阳黑子时间序列检验其预测性能.本文采集了1749-2019年的太阳黑子数据,共有3252条数据,经过相空间重构后,产生3094组样本数据,训练集和测试集仍以8:2比例划分,即前80%作为训练集后20%作为测试集.预测结果和误差如图10和图11所示,表5和表6则列出了Att-CNN-LSTM模型和其他模型的预测结果对比.

从图10和图11中可以看出, Att-CNN-LSTM模型在太阳黑子时间序列的预测上仍然具有较好的表现,能够很好地预测序列的变化趋势,尽管有部分时间点的预测误差达到了4.7%,但是总体误

差依然保持在较低的水平.如表3所列,虽然Att-CNN-LSTM模型在真实数据集上的预测提升不如对Logistic和Lorenz等理论数据集,但是其预测误差仍低于其他模型.相比其他预测模型, Att-CNN-LSTM至少降低了19%的RMSE值, 6%的MAE值, 23%的MAPE值和4%的RMSPE值,预测精度提升依然较为明显.此外, CNN, LSTM

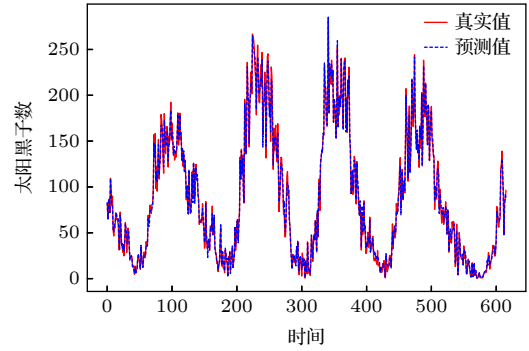


图10 太阳黑子序列预测

Fig. 10. Sunspot series prediction.

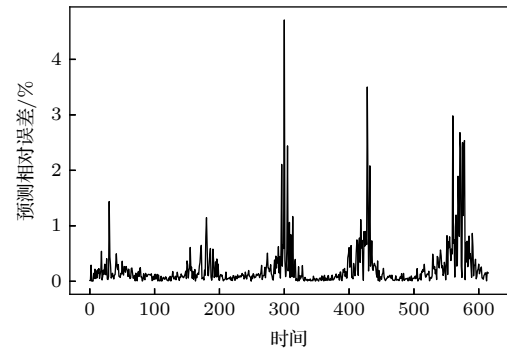


图11 太阳黑子序列预测相对误差

Fig. 11. Sunspot series prediction relative error.

表5 模型误差对比

Table 5. Model error comparison.

	RMSE	MAE	MAPE	RMSPE
Att-CNN-LSTM	20.1829	17.1827	32.8167	42.3529
CNN-LSTM	30.5652	21.4093	67.4343	56.7217
LSTM	24.9137	18.2815	49.5862	62.6939
CNN	24.8534	18.4677	65.3480	44.1892
LSSVM	27.3271	19.4373	43.0987	56.6781

表6 模型运行时间对比

Table 6. Model running time comparison.

模型	Att-CNN-LSTM	CNN-LSTM	CNN	LSTM	LSSVM
训练时间 /s	309.2	291.3	76.5	53.3	237.9
预测时间 /s	0.39	0.43	0.15	0.25	0.59

和 LSSVM 的综合预测性能相差不多, 但是都优于 CNN-LSTM 模型. 在运行时间方面, 虽然 Att-CNN-LSTM 模型的训练时间多于其他模型, 但是其预测时间为已低至 0.39 s. 综合各模型的预测误差和运行时间分析, 本文提出的模型依然具有较好的预测性能.

4.4 讨论

由 4.1—4.3 节的实验和分析可知, 本文提出的 Att-CNN-LSTM 模型在 Logistic, Lorenz 和太阳黑子三个混沌时间序列上取得了优于其他模型的预测性能. 在 Att-CNN-LSTM 模型中, CNN 模型可以提取混沌时间序列重构相空间下的空间特征, LSTM 则根据 CNN 提取的空间特征提取原时间序列的时空特征, 而注意力模型对 CNN-LSTM 提取的时空特征进行筛选, 对重要的特征赋予高权重, 并弱化非关键特征, 这对预测精度的提升起到了重要作用. 在混沌时间序列的高维重构相空间内, 混沌系统的内在信息被挖掘出来, 但是并非所有信息都有利于预测, 过多的非关键的信息会影响到预测精度. 原始的混沌时间序列虽然蕴含着混沌系统的时序特征, 但在复杂的混沌系统中, 仅仅依靠时序特征很难获取精度较高的预测结果. 此外, 若同时考虑混沌系统的空间和时间特征, 所涉及的信息和特征会很多, 那些不重要的信息和特征会对预测造成负面影响, 不利于预测精度的提升.

实验中发现, Logistic 和太阳黑子混沌时间序列拥有较高的相空间展开维度, 其所在的混沌系统较为复杂, 而 Lorenz 的展开维度相对较低, 其动力系统相对简单. 混合模型 CNN-LSTM 在高维相空间下的预测精度明显不如单一的 CNN 和 LSTM 模型, 但是在低维情况下, CNN-LSTM 的预测精度又明显优于单一预测模型, 这说明在复杂的混沌系统下, 过多的信息和非关键特征影响了预测精度, 而注意力模型能够通过提取关键特征解决这些问题, 使得预测精度有所提升, 三个预测实例也很好证明了本文提出的注意力模型能够提升混合模型 CNN-LSTM 预测性能.

5 结论

本文利用注意力机制捕获混合神经网络 CNN-LSTM 的时空特征, 提出了一种用于混沌时间序列

预测的 Att-CNN-LSTM 模型. 结合相空间重构理论, 用 CNN 提取动力系统重构相空间下的空间特征, 让 LSTM 根据空间特征提取序列的时间特征. 作为模型的关键部分, 注意力模型的设计关系到序列的特征获取, 进而影响到预测精度. 因此, 为更好地获取时间序列的特征, 可以通过添加非线性激活函数来获取加权参数. 仿真实验结果表明, 本文提出的模型在几个混沌时间序列的预测上都取得了较好的结果, 且预测精度高, 预测性能较好.

参考文献

- [1] Wang S Y, Shi C F, Qian G B, Wang W L 2018 *Acta Phys. Sin.* **67** 018401 (in Chinese) [王世元, 史春芬, 钱国兵, 王万里 2018 *物理学报* **67** 018401]
- [2] Mei Y, Tan G Z, Liu Z T, Wu H 2018 *Acta Phys. Sin.* **67** 080502 (in Chinese) [梅英, 谭冠政, 刘振焘, 武鹤 2018 *物理学报* **67** 080502]
- [3] Shen L H, Chen J H, Zeng Z G, Jin J 2018 *Acta Phys. Sin.* **67** 030501 (in Chinese) [沈力华, 陈吉红, 曾志刚, 金健 2018 *物理学报* **67** 030501]
- [4] Han M, Zhang S, Xu M, Qiu T, Wang N 2019 *IEEE Trans. Cybern.* **49** 1160
- [5] Han M, Zhang R, Qiu T, Xu M, Ren W 2019 *IEEE Trans. Syst. Man Cybern.* **49** 2144
- [6] Safari N, Chung C Y, Price G C D 2018 *IEEE Trans. Power Syst.* **33** 590
- [7] Xiong Y C, Zhao H 2019 *Sci. China, Ser. G* **49** 92 (in Chinese) [熊有成, 赵鸿 2019 *中国科学: 物理学 力学 天文学* **49** 92]
- [8] Sangiorgio M, Dercole F 2020 *Chaos, Solitons Fractals* **139** 110045
- [9] Li S X, Sun X K, Yin L, Zhang S S 2020 *Journal of Navigation and Positioning* **8** 65 (in Chinese) [李世玺, 孙宪坤, 尹玲, 张仕森 2020 *导航定位学报* **8** 65]
- [10] Boullé N, Dallas V, Nakatsukasa Y, Samaddar D 2019 *Physica D* **403** 132261
- [11] Tang Z J, Ren F, Peng T, Wang W B 2014 *Acta Phys. Sin.* **63** 050505 (in Chinese) [唐舟进, 任峰, 彭涛, 王文博 2014 *物理学报* **63** 050505]
- [12] Tian Z D, Gao X W, Shi T 2014 *Acta Phys. Sin.* **63** 160508 (in Chinese) [田中大, 高宪文, 石彤 2014 *物理学报* **63** 160508]
- [13] Wang Y X, Han M 2015 *Acta Phys. Sin.* **64** 070504 (in Chinese) [王新迎, 韩敏 2015 *物理学报* **64** 070504]
- [14] Lu J H 2002 *Chaotic Time Series Analysis and Application* (Wuhan: Wuhan University Press) pp57–60 (in Chinese) [吕金虎 2002 *混沌时间序列分析及其应用* (武汉: 武汉大学出版社) 第57—60页]
- [15] Packard N, Crutchfield J P, Shaw R 1980 *Phys. Rev. Lett.* **45** 712
- [16] Takens F 1981 *Dynamical Systems and Turbulence* (Berlin: Springer) pp366–381
- [17] Martinerie J M, Albano A M, Mees A I, Rapp P E 1992 *Phys. Rev. A* **45** 7058
- [18] Liangyue C 1997 *Physica D* **110** 43
- [19] Lecun Y, Boser B, Denker J, Henderson D, Howard R,

- Hubbard W, Jackel L 1989 *Neural Comput.* **1** 541
- [20] Goodfellow I, Bengio Y, Courville A 2016 *Deep Learning* (Cambridge: The MIT Press) p326
- [21] Kim Y 2014 arXiv: 1408.5882 [cs.CL]
- [22] Pascanu R, Mikolov T, Bengio Y 2013 *Proceedings of the 30th International Conference on International Conference on Machine Learning* Atlanta, USA, June 16–21 2013, p1310
- [23] Hochreiter S, Schmidhuber J 1997 *Neural Comput.* **9** 1735
- [24] Chung J, Gulcehre C, Cho K, Bengio Y 2014 arXiv: 1412.3555 [cs.NE]
- [25] Mnih V, Heess N, Graves A, Kavukcuoglu K 2014 *Proceedings of the 27th International Conference on Neural Information Processing Systems* Montreal, Canada, December 8–13 2014, p2204
- [26] Yin W, Schütze H, Xiang B, Zhou B 2015 arXiv: 1512.05193 [cs.CL]
- [27] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L, Polosukhin I 2017 *Proceedings of the 31st International Conference on Neural Information Processing Systems* Long Beach, USA, December 4–9 2017, p6000

Prediction of chaotic time series using hybrid neural network and attention mechanism*

Huang Wei-Jian Li Yong-Tao[†] Huang Yuan

(School of Information & Electrical, Hebei University of Engineering, Handan 056038, China)

(Received 12 June 2020; revised manuscript received 19 July 2020)

Abstract

Chaotic time series forecasting has been widely used in various domains, and the accurate predicting of the chaotic time series plays a critical role in many public events. Recently, various deep learning algorithms have been used to forecast chaotic time series and achieved good prediction performance. In order to improve the prediction accuracy of chaotic time series, a prediction model (Att-CNN-LSTM) is proposed based on hybrid neural network and attention mechanism. In this paper, the convolutional neural network (CNN) and long short-term memory (LSTM) are used to form a hybrid neural network. In addition, a attention model with *softmax* activation function is designed to extract the key features. Firstly, phase space reconstruction and data normalization are performed on a chaotic time series, then convolutional neural network (CNN) is used to extract the spatial features of the reconstructed phase space, then the features extracted by CNN are combined with the original chaotic time series, and in the long short-term memory network (LSTM) the combined vector is used to extract the temporal features. And then attention mechanism captures the key spatial-temporal features of chaotic time series. Finally, the prediction results are computed by using spatial-temporal features. To verify the prediction performance of the proposed hybrid model, it is used to predict the Logistic, Lorenz and sunspot chaotic time series. Four kinds of error criteria and model running times are used to evaluate the performance of predictive model. The proposed model is compared with hybrid CNN-LSTM model, the single CNN and LSTM network model and least squares support vector machine(LSSVM), and the experimental results show that the proposed hybrid model has a higher prediction accuracy.

Keywords: chaotic time series, convolutional neural network, long short-term memory network, attention mechanism

PACS: 05.45.-a, 05.45.Tp, 42.79.Ta

DOI: 10.7498/aps.70.20200899

* Project supported by the Natural Science Foundation of Hebei Province, China (Grant No. F2015402077) and the Scientific Research Foundation of the Higher Education Institutions of Hebei Province, China (Grant No. QN2018073).

[†] Corresponding author. E-mail: lyotard@163.com