

基于简化型 LSTM 神经网络的时间序列预测方法

李文静^{1,2,3,4}, 王潇潇^{1,2,3,4}

(1. 北京工业大学信息学部, 北京 100124; 2. 计算智能与智能系统北京市重点实验室, 北京 100124;
3. 智慧环保北京实验室, 北京 100124; 4. 北京人工智能研究院, 北京 100124)

摘要: 针对标准长短期记忆(long short-term memory, LSTM)神经网络用于时间序列预测具有耗时长、复杂度高等问题, 提出简化型 LSTM 神经网络并应用于时间序列预测。首先, 通过耦合输入门与遗忘门实现对标准 LSTM 神经网络的结构简化; 其次, 从门结构控制方程中消除输入信号与偏差实现进一步精简; 然后, 采用梯度下降算法更新简化型 LSTM 神经网络的参数; 最后, 通过 2 个时间序列基准数据集及污水处理过程出水生化需氧量(biochemical oxygen demand, BOD)质量浓度预测进行实验验证。结果表明: 在不显著降低预测精度的情况下, 所设计的模型能够缩短训练时间, 减少 LSTM 神经网络的计算复杂度, 实现时间序列的预测。

关键词: 时间序列预测; 长短期记忆(long short-term memory, LSTM)神经网络; 门耦合; 参数精简; 梯度下降算法; 污水处理过程

中图分类号: U 461; TP 308

文献标志码: A

文章编号: 0254-0037(2021)05-0480-09

doi: 10.11936/bjtxb2020120032

Time Series Prediction Method Based on Simplified LSTM Neural Network

LI Wenjing^{1,2,3,4}, WANG Xiaoxiao^{1,2,3,4}

(1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;
2. Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124, China;
3. Beijing Laboratory for Intelligent Environmental Protection, Beijing 100124, China;
4. Beijing Artificial Intelligence Institute, Beijing 100124, China)

Abstract: To solve the problem that the standard long short-term memory (LSTM) neural network is time consuming and has high complexity for time series prediction, a simplified LSTM neural network was proposed and it was applied to time series prediction. First, the structure of the standard LSTM neural network was simplified by coupling input gate and forget gate. Second, the inputs and bias were removed from dynamic equation of the gates to further simplify the parameters. Third, the gradient descent algorithm was utilized to update the parameters of the simplified LSTM neural network. Finally, the validity of the proposed model was demonstrated by two time series benchmark problems and the prediction of biochemical oxygen demand (BOD) mass concentration in the wastewater treatment process. The experimental results show that the training time is shortened and the computational complexity is reduced without significantly reducing the prediction accuracy, which makes an efficient time series prediction.

收稿日期: 2020-12-30

基金项目: 国家自然科学基金资助项目(61603009, 62021003, 61890930-5); 国家重点研发计划资助项目(2018YFC1900800-5); 北京市教育委员会科技计划资助项目(KM201910005023)

作者简介: 李文静(1985—), 女, 副教授, 主要从事智能建模、模式识别、认知神经理论方面的研究, E-mail: wenjing.li@bjut.edu.cn

Key words: time series prediction; long short-term memory (LSTM) neural networks; gate coupling; parameter simplification; gradient descent algorithm; wastewater treatment process

时间序列预测可以判断事物发展趋势,高效的预测模型可为应用决策提供有力依据^[1]. 长短期记忆(long short-term memory, LSTM)神经网络对时间序列预测具有显著优势^[2-4],已广泛地应用于金融市场股票预测^[5-7]、石油产量预测^[8]、短时交通流预测^[9]等领域,但标准 LSTM 神经网络用于时间序列预测具有耗时长、复杂度高问题^[10-11]. 围绕 LSTM 神经网络结构设计,目前已有大量学者进行了研究.

LSTM 神经网络在训练过程中需要更新较多的参数,增加了训练时间^[12],故对其内部结构进行删减尤为重要. 一些研究者通过简化 LSTM 神经网络结构提出了多种基于标准 LSTM 神经网络的简化变体^[13],如去除遗忘门^[14]、耦合输入门与遗忘门^[15]、去除窥视孔连接^[16]等. Greff 等^[13]对多种 LSTM 神经网络简化变体的效果进行了评价,实验证明耦合输入门与遗忘门、去除窥视孔连接的简化变体可以在不显著降低性能的情况下减少 LSTM 模型的参数数量和计算成本. Cho 等^[15]提出一种包括重置门与更新门 2 个门结构的门控循环单元(gated recurrent unit, GRU),实验证明 GRU 可以达到与 LSTM 模型相当的效果,并且能够很大程度上提高训练效率. Zhou 等^[17]提出只有一个门结构的最小门控单元(minimal gated unit, MGU),实验证明 MGU 具有与 GRU 相当的精度,但结构更简单,参数更少,训练速度更快. Oliver 等^[18]通过耦合输入门与遗忘门以简化 LSTM 模型,使用一个门结构同时控制遗忘和选择记忆,该神经网络与其他 LSTM 模型简化变体相比能够减少对历史数据的依赖性,降低网络的复杂度,在网络性能不变的情况下缩短训练时间^[19]. 然而,以上介绍的 LSTM 简化模型,仍需要更新和存储较多的参数,导致网络计算冗余,训练时间较长.

针对以上问题,近几年一些学者提出精简门结构方程的方法,进一步减少训练过程中需要更新的参数,提高训练速度. Lu 等^[20]通过精简标准 LSTM 网络门结构方程减少参数更新,提出 3 个模型并将其与标准 LSTM 网络结构比较,实验证明该模型在较少参数的情况下可获得与标准 LSTM 模型相当的性能. Rahul 等^[21]通过减少重置门和更新门的参数,提出 GRU 的 3 种变体,并对其性能进行了评估.

结果表明,这些变体的性能与 GRU 模型相当,同时降低了计算开销. Joel 等^[22]介绍了 MGU 的 3 种模型变体,通过减少遗忘门动力方程中的参数数目,进一步简化了设计,这 3 种模型变体显示出与 MGU 模型相似的精度,同时使用较少的参数减少训练时间. 根据以上分析,在减少门结构数量的基础上精简门结构参数能够在保证网络性能的前提下减少网络的训练时间.

由于 Oliver 等^[18]提出的 LSTM 简化神经网络具有较短训练时间、较少参数数量等优点,本文基于该网络提出简化型 LSTM 神经网络,在耦合门结构的基础上继续对门结构方程中的参数进行简化,可以更大程度上减少 LSTM 神经网络在训练过程中参数更新的数量,提高网络的训练速度. 通过 2 个基准数据集及污水处理过程出水生化需氧量(biochemical oxygen demand, BOD)质量浓度预测的实验验证,将其在 3 个时间序列数据集上与标准 LSTM 网络及其他变体进行比较评价,结果说明本文提出的简化型 LSTM 神经网络在训练时间减少的同时能够达到较好的时间序列预测精度.

1 标准 LSTM 神经网络结构

标准 LSTM 神经网络结构包含一个状态单元及 3 个门结构(输入门、遗忘门、输出门),其中状态单元用于记录当前时刻的状态,各门结构用于控制信息的遗忘或记忆. 本文介绍的 LSTM 模型的结构是去除窥视孔连接的标准 LSTM 模型^[16],其内部结构图如图 1 所示. 标准 LSTM 模型的结构为

$$\begin{cases} z_t = g(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1} + \mathbf{b}_z) \\ i_t = \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i) \\ f_t = \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f) \\ c_t = f_t \odot c_{t-1} + i_t \odot z_t \\ o_t = \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o) \\ \mathbf{h}_t = o_t \odot g(c_t) \end{cases} \quad (1)$$

式中: \mathbf{x}_t 为当前时刻输入向量; \mathbf{h}_t 为当前时刻输出向量; $z_t, i_t, f_t, c_t, o_t, \mathbf{h}_t$ 分别为输入信号、输入门、遗忘门、状态单元、输出门、输出信号; $\mathbf{W}_z, \mathbf{W}_i, \mathbf{W}_f, \mathbf{W}_o$ 分别为 z_t, i_t, f_t, o_t 中的输入权重矩阵; $\mathbf{U}_z, \mathbf{U}_i, \mathbf{U}_f, \mathbf{U}_o$ 分别为 z_t, i_t, f_t, o_t 中的递归权重矩阵; $\mathbf{b}_z, \mathbf{b}_i, \mathbf{b}_f, \mathbf{b}_o$ 分别为 z_t, i_t, f_t, o_t 中的偏置矩阵; σ 为 sigmoid 激活函数;

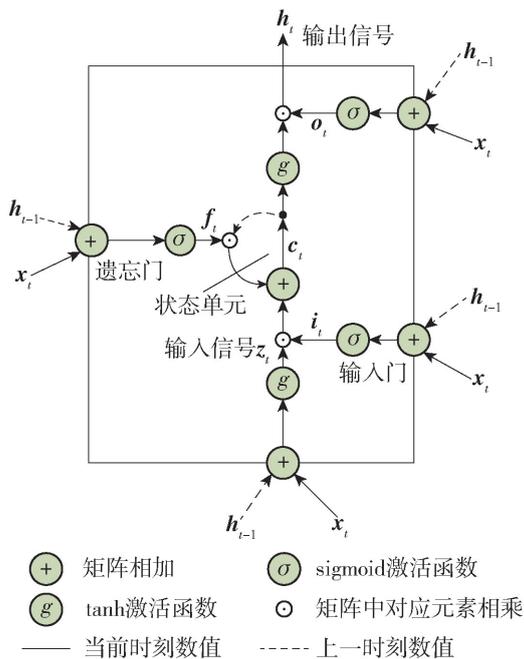


图1 LSTM神经网络内部结构详细示意图

Fig. 1 Detailed schematic diagram of the internal structure for LSTM neural network

g 为 tanh 激活函数; \odot 表示矩阵点乘操作.

对于只有一个重复隐含状态的递归神经网络 (recurrent neural network, RNN) 结构,若设定 m 为输入向量的维度, n 为隐含层单元的个数,则每次迭代过程需要更新的参数个数为 $(mn + n^2 + n)$. 由于标准 LSTM 神经网络存在 3 个门结构 (输入门 i_t 、遗忘门 f_t 、输出门 o_t) 与输入信号 z_t , 由式(1)可知,标准 LSTM 神经网络在每次迭代过程中需要更新的参数个数为 $4(mn + n^2 + n)$.

2 简化型 LSTM 神经网络设计

本文提出的简化型 LSTM 神经网络,首先通过耦合输入门与遗忘门简化标准 LSTM 神经网络结构,其次对门结构方程中的参数进行精简以进一步减少网络参数,从而提高网络训练速度.

2.1 LSTM 神经网络门结构简化设计

本文通过耦合输入门与遗忘门实现对标准 LSTM 神经网络的门结构简化,其结构由 1 个状态单元及 2 个门结构组成 (如图 2 所示),具体介绍如下.

1) 输入门:控制需要输入到网络中的信息,该结构与标准 LSTM 神经网络相同,通过

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (2)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (3)$$

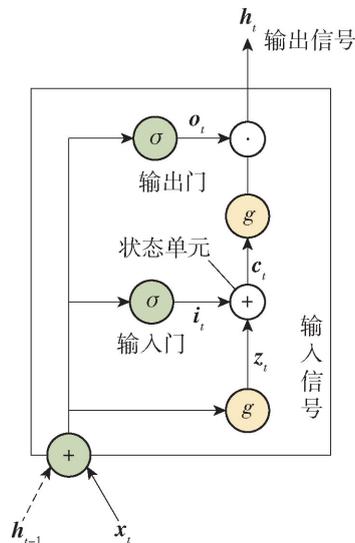


图2 简化 LSTM 神经网络内部结构

Fig. 2 Internal structure of the simplified LSTM neural network

实现.

2) 状态单元:状态单元 c_t 结合输入信号 z_t 与 $1 - i_t$ 控制的上一时刻的状态单元 c_{t-1} , 其更新公式为

$$c_t = (1 - i_t) \odot c_{t-1} + z_t \quad (4)$$

由此可见,与标准 LSTM 神经网络不同,式(4)由 $1 - i_t$ 代替遗忘门 f_t 对上一时刻的状态单元进行选择记忆,当 i_t 数值为 0 时,上一时刻的单元状态全部记忆,当 i_t 数值为 1 时,上一时刻的单元状态全部遗忘,从而实现了输入门与遗忘门的耦合.

3) 输出门:控制当前时刻状态单元信息 c_t 的输出程度,该结构与标准 LSTM 神经网络输出结构相同,通过

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t \odot g(c_t) \quad (6)$$

实现. 由此可见,输出门 o_t 控制神经网络的最终输出. 若 o_t 数值为 0,则当前时刻单元状态 c_t 全部不输出, h_t 输出值为 0; 若 o_t 数值为 1,则当前时刻单元状态 c_t 全部输出.

经过输入门与遗忘门的耦合, LSTM 网络在简化后由 2 个门结构组成,每次迭代过程需要更新的参数个数为 $3(mn + n^2 + n)$,与标准 LSTM 神经网络结构相比减少了 25%.

2.2 LSTM 神经网络门结构参数精简方法

虽然耦合输入门及遗忘门简化了标准 LSTM 神经网络的结构,然而在每次训练过程中均需对输入权重矩阵 W_z 、 W_i 、 W_o 进行更新,由此导致计算

量较大,训练时间较长.针对该问题,本文通过简化门结构方程的参数进一步对 LSTM 神经网络结构进行精简,在不损失精度的前提下缩短网络的训练时间.

本文主要通过 2 种方法精简门结构方程,包括:1) 去除输入权重矩阵 \mathbf{W}_i 、 \mathbf{W}_o ;2) 去除输入权重矩阵 \mathbf{W}_i 、 \mathbf{W}_o 与偏置矩阵 \mathbf{b}_i 、 \mathbf{b}_o . 本文将经过以上 2 种形式简化后的 LSTM 神经网络分别简称为 LSTM-简化型 I 神经网络和 LSTM-简化型 II 神经网络,以下分别对这 2 种简化型 LSTM 神经网络进行介绍.

1) LSTM-简化型 I 神经网络

该简化方法通过去除输入门与输出门中的输入权重矩阵 \mathbf{W}_i 、 \mathbf{W}_o 进一步简化 LSTM 神经网络,由

$$\begin{cases} \mathbf{z}_t = g(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1} + \mathbf{b}_z) \\ \mathbf{i}_t = \sigma(\mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i) \\ \mathbf{c}_t = (1 - \mathbf{i}_t) \odot \mathbf{c}_{t-1} + \mathbf{z}_t \\ \mathbf{o}_t = \sigma(\mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o) \\ \mathbf{h}_t = \mathbf{o}_t \odot g(\mathbf{c}_t) \end{cases} \quad (7)$$

构成.

由此可见,与标准 LSTM 神经网络的门结构控制信号不同之处为:该网络门结构控制信号由 $t-1$ 时刻输出信号 \mathbf{h}_{t-1} 、递归权重矩阵及偏置矩阵 2 项组成,在每次迭代过程中该网络需要更新的参数个数为 $3(mn + n^2 + n - 2mn)$,降低了计算复杂度.

2) LSTM-简化型 II 神经网络

该简化方法在去除输入门与输出门中输入权重矩阵 \mathbf{W}_i 、 \mathbf{W}_o 的同时,将偏置矩阵 \mathbf{b}_i 、 \mathbf{b}_o 去除,由

$$\begin{cases} \mathbf{z}_t = g(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1} + \mathbf{b}_z) \\ \mathbf{i}_t = \sigma(\mathbf{U}_i \mathbf{h}_{t-1}) \\ \mathbf{c}_t = (1 - \mathbf{i}_t) \odot \mathbf{c}_{t-1} + \mathbf{z}_t \\ \mathbf{o}_t = \sigma(\mathbf{U}_o \mathbf{h}_{t-1}) \\ \mathbf{h}_t = \mathbf{o}_t \odot g(\mathbf{c}_t) \end{cases} \quad (8)$$

构成.

由此可见,与标准 LSTM 神经网络的门结构控制信号不同之处为:该网络门结构控制信号仅由 $t-1$ 时刻输出信号 \mathbf{h}_{t-1} 、递归权重矩阵 1 项组成,在每次迭代过程中该模型需要更新的参数个数为 $3(mn + n^2 + n - 2mn - 2n)$,进一步降低了 LSTM 神经网络的计算复杂度.

2.3 简化型 LSTM 神经网络学习算法

本文采用梯度下降算法^[23-24]对提出的简化型 LSTM 神经网络的参数进行学习,定义损失函数计算公式为

$$E(t) = \frac{1}{2}(\mathbf{h}_{d,t} - \mathbf{h}_t)^2 \quad (9)$$

式中: $\mathbf{h}_{d,t}$ 为网络在 t 时刻的期望输出; \mathbf{h}_t 为网络在 t 时刻的实际输出.

下面以 LSTM-简化型 I 神经网络为例,介绍参数更新过程.

步骤 1 根据

$$\delta \mathbf{h}_t = \delta \mathbf{z}_{t+1} \mathbf{U}_z + \delta \mathbf{i}_{t+1} \mathbf{U}_i + \delta \mathbf{o}_{t+1} \mathbf{U}_o \quad (10)$$

$$\delta \mathbf{z}_t = \delta \mathbf{h}_t \odot \mathbf{o}_t \odot g'(\mathbf{c}_t) \odot \mathbf{i}_t \odot \mathbf{z}'_t \quad (11)$$

$$\delta \mathbf{i}_t = \delta \mathbf{h}_t \odot \mathbf{o}_t \odot g'(\mathbf{c}'_t) \odot \mathbf{z}_t \odot \mathbf{i}'_t \quad (12)$$

$$\delta \mathbf{o}_t = \delta \mathbf{h}_t \odot g(\mathbf{c}_t) \odot \mathbf{o}'_t \quad (13)$$

计算 t 时刻输出值 \mathbf{h}_t 及输入信号 \mathbf{z}_t 、 \mathbf{i}_t 、 \mathbf{o}_t 的误差项.其中,导数形式展开公式为

$$g'(\mathbf{c}_t) = 1 - g(\mathbf{c}_t)^2 \quad (14)$$

$$\mathbf{z}'_t = \mathbf{z}_t(1 - \mathbf{z}_t) \quad (15)$$

$$\mathbf{i}'_t = \mathbf{i}_t(1 - \mathbf{i}_t) \quad (16)$$

$$\mathbf{o}'_t = \mathbf{o}_t(1 - \mathbf{o}_t) \quad (17)$$

步骤 2 计算 t 时刻输入权重矩阵、递归权重矩阵、偏置矩阵的更新值公式为

$$\delta \mathbf{W}_{z,t} = \delta \mathbf{z}_t \otimes \mathbf{x}_t \quad (18)$$

$$\delta \mathbf{U}_{\Omega,t} = \delta \mathbf{\Omega}_t \otimes \mathbf{h}_{t-1} \quad (19)$$

$$\delta \mathbf{b}_{\Omega,t} = \delta \mathbf{\Omega}_t \quad (20)$$

式中: \otimes 为矩阵叉乘操作; $\mathbf{\Omega}$ 分别为 $\{\mathbf{z}, \mathbf{i}, \mathbf{o}\}$ 中的任意一个.

步骤 3 根据

$$\mathbf{W}_{z,t} = \mathbf{W}_{z,t+1} - \eta \times \delta \mathbf{W}_{z,t} \quad (21)$$

$$\mathbf{U}_{\Omega,t} = \mathbf{U}_{\Omega,t+1} - \eta \times \delta \mathbf{U}_{\Omega,t} \quad (22)$$

$$\mathbf{b}_{\Omega,t} = \mathbf{b}_{\Omega,t+1} - \eta \times \delta \mathbf{b}_{\Omega,t} \quad (23)$$

计算 t 时刻更新后的输入权重矩阵、递归权重矩阵、偏置矩阵.式中 η 为学习率.

步骤 4 计算训练样本的均方根误差 (root mean squared error, RMSE),如果训练样本的 RMSE 达到期望训练样本的 RMSE 或达到最大迭代次数,则参数更新结束,否则返回步骤 1.

对于 LSTM-简化型 II 神经网络,由于其门结构方程在 LSTM-简化型 I 神经网络的基础上进一步去除了偏置矩阵,其权重矩阵更新与 LSTM-简化型 I 神经网络相同,如式(21)(22)所示.

3 实验结果和分析

为了验证所提出的简化型 LSTM 神经网络在时间序列预测上的有效性,本文采用 RMSE 评价模型的预测准确性^[25-27],公式为

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^N (\mathbf{h}_{d,t} - \mathbf{h}_t)^2}{N}} \quad (24)$$

式中 N 为样本个数. 将其与标准 LSTM 神经网络、只进行门结构简化的 LSTM 神经网络(简称 LSTM-变体 I)、仅去除输入权重矩阵的 LSTM 神经网络(简称 LSTM-变体 II)、仅去除输入权重矩阵与偏置矩阵的 LSTM 神经网络(简称 LSTM-变体 III)等多种 LSTM 模型的性能进行比较,在参数设置(包括 LSTM 模型状态单元维度、学习率、期望训练样本的 RMSE、迭代次数)相同的情况下分别计算训练和测试 RMSE、训练时间及所需更新参数个数等,所有实验独立运行 20 次并求取均值.

3.1 时间序列基准数据集

在本节中采用 2 个时间序列基准数据集(Lorenz 时间序列、Mackey-Glass 时间序列)评估简化型 LSTM 神经网络的性能.

3.1.1 Lorenz 时间序列预测

Lorenz 系统是一种大气对流数学模型^[28],它被广泛地用作时间序列预测的基准实验以评价模型的有效性. 其系统方程为

$$\begin{cases} \frac{dx(t)}{dt} = a_1 y(t) - a_1 x(t) \\ \frac{dy(t)}{dt} = a_2 x(t) - x(t)z(t) - y(t) \\ \frac{dz(t)}{dt} = x(t)y(t) - a_3 z(t) \end{cases} \quad (25)$$

式中: $x(t)$ 、 $y(t)$ 、 $z(t)$ 为三维空间 Lorenz 系统的序列; a_1 、 a_2 、 a_3 为系统参数, $a_1 = 10$, $a_2 = 28$, $a_3 = 8/3$.

在本实验中,生成 5 000 组 Lorenz 样本,仅使用 y 维样本 $y(t)$ 进行时间序列预测. 前 2 000 组作为训练样本,后 3 000 组作为测试样本. 以 $[y(t)$

$y(t-1)$ $y(t-2)$] 为输入向量,预测 $y(t+1)$ 的值. 设定状态单元维度为 8,学习率 η 为 0.01,期望训练样本的 RMSE 为 0.060 0,最大迭代次数为 1 000 次. 当训练样本的 RMSE 达到期望训练样本的 RMSE 或最大迭代次数时,停止参数更新.

LSTM-简化型 I、II 的训练过程 RMSE 曲线分别如图 3、4 中的(a)所示. 从图中可以看出,本文提出的简化型 LSTM 神经网络的训练 RMSE 可以快速收敛. 其测试结果如图 3、4 中的(b)(c)所示,可以看出其均可以达到较好的拟合效果.

表 1 对比了不同模型的性能,可以看出,(LSTM-变体 I)或(LSTM-变体 II、III)均可以减少更新参数个数并缩短训练时间,但 LSTM-变体 I、II 的训练时间短于 LSTM-变体 III,同时 LSTM-简化型 I 在需要更新的参数个数比 LSTM-变体 III 较多的情况下训练时间显著缩短,均说明门结构精简相对于简化门结构方程对简化 LSTM 神经网络的效果更显著. 通过实验结果分析可以得出,本文提出的 LSTM-简化型 I、II 神经网络能够在不显著降低预测精度的情况下,进一步缩短训练时间,减少 LSTM 神经网络的计算复杂度,减少预测时间,更易对时间序列信息预测.

3.1.2 Mackey-Glass 时间序列预测

Mackey-Glass 时间序列预测问题已被公认为评估网络性能的基准问题之一^[29]. 时间序列预测由离散方程

$$x(t+1) = (1-a)x(t) + \frac{bx(t-\tau)}{1+x^{10}(t-\tau)} \quad (26)$$

产生. 式中: $a = 0.1$, $b = 0.2$, $\tau = 17$, $x(0) = 1.2$.

在本实验中,选取样本 1 000 组,其中前 500 组作为训练样本,后 500 组作为测试样本. 以 $[x(t)$ $x(t-6)$ $x(t-12)$ $x(t-18)]$ 为输入向量,预测

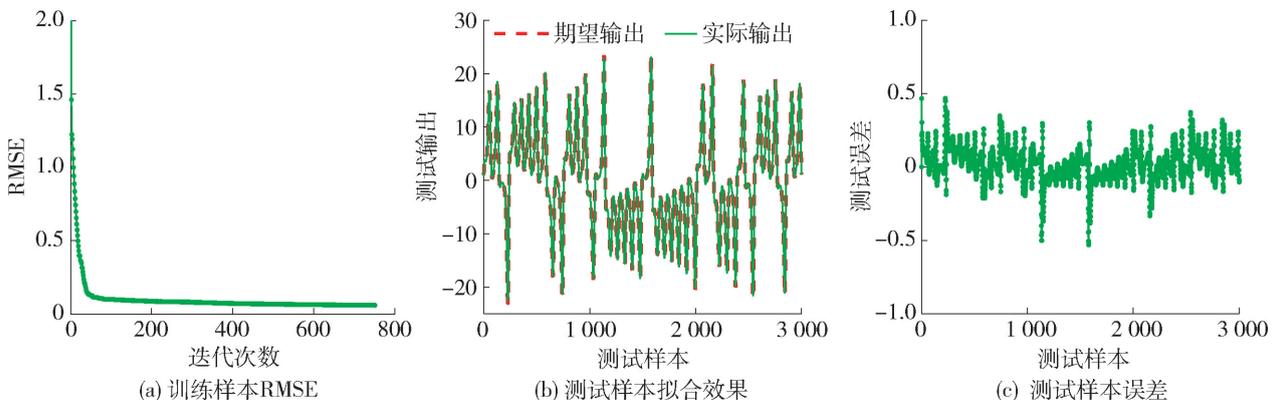


图3 LSTM-简化型 I 对 Lorenz 时间序列预测的训练过程及测试效果

Fig. 3 Training process and testing results for the simplified LSTM I in Lorenz time series

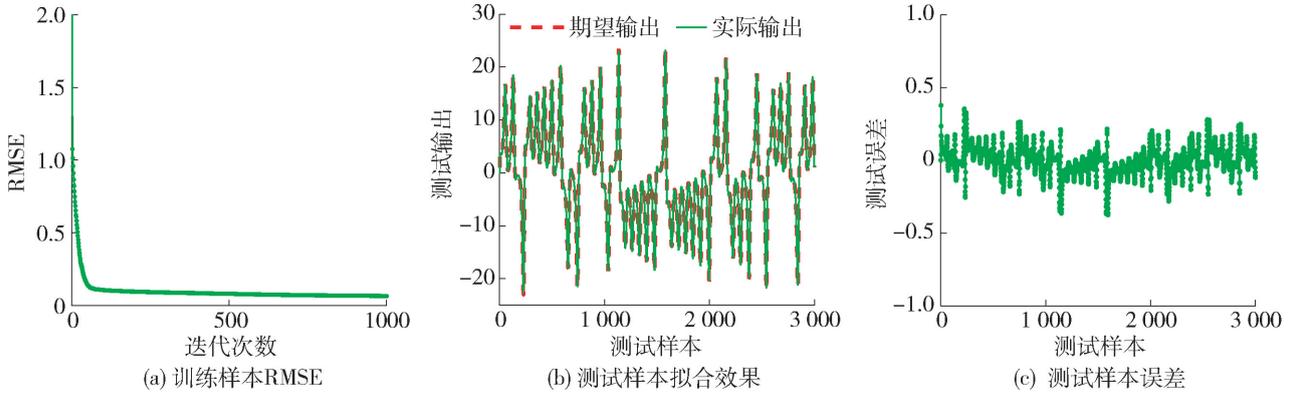


图4 LSTM-简化型 II 对 Lorenz 时间序列预测的训练过程及测试效果

Fig. 4 Training process and testing results for the simplified LSTM II in Lorenz time series

表1 Lorenz 时间序列预测模型性能对比

Table 1 Performance comparison of different models for Lorenz time series

模型	RMSE		训练时间/s	更新参数个数
	训练样本	测试样本		
标准 LSTM	0.060 0	0.079 3	98.67	384
LSTM-变体 I	0.060 0	0.074 1	73.91	288
LSTM-变体 II	0.060 0	0.073 3	72.68	192
LSTM-变体 III	0.060 0	0.070 8	92.62	128
LSTM-简化型 I	0.060 0	0.075 2	58.68	144
LSTM-简化型 II	0.066 0	0.078 1	69.05	96

注:黑体代表本文提出的简化型 LSTM 模型。

$x(t+6)$ 的值. 设定状态单元维度为 10, 学习率 η 为 0.01, 期望训练 RMSE 为 0.006 0, 最大迭代次数为 700 次. 当训练样本的 RMSE 达到期望训练样本的 RMSE 或最大迭代次数时, 停止参数更新.

LSTM-简化型 I、II 的训练过程分别如图 5、6 中的(a)所示. 从图中可以看出, 训练 RMSE 可以达到期望训练 RMSE. 测试结果、测试误差分别如图 5、6 中的(b)(c)所示, 从图中可以看出, 本文提出的简化型 LSTM 神经网络的测试结果可以达到较好的拟合效果.

从表 2 对不同模型进行比较的结果可以看出, 3 种 LSTM 变体(LSTM-变体 I、II、III)通过对门结构精简或简化门结构方程的方式, 均缩短了训练时间, 并且 LSTM-变体 I 在需要更新的参数个数比 LSTM-变体 II、III 较多的情况下训练时间缩短, 同时 LSTM-简化型 I、LSTM-变体 III 均可以减少更新参数个数并缩短训练时间, 但前者的训练时间短于后者, 说明门结构精简相对于简化门结构方程对简化 LSTM 神经网络的效果更显著. 通过实验结果分析, 可以得出, 本文提出的 LSTM-简化型 I、II 神经网络在不显著降低预测精度的情况下进一步缩短训练时间, 在时间序列预测过程中达到对时间序列信息简洁、快速预测的目的.

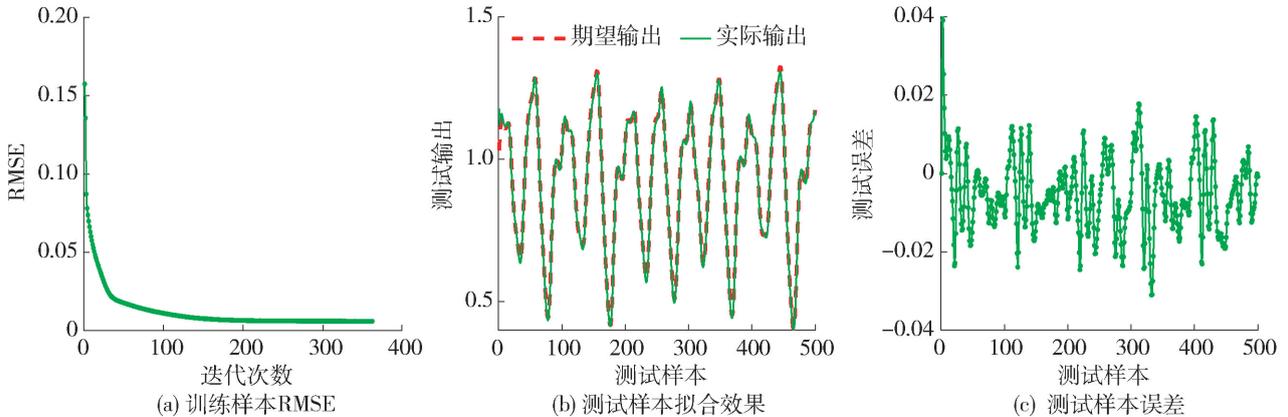


图5 LSTM-简化型 I 对 Mackey-Glass 时间序列预测的训练过程及测试效果

Fig. 5 Training process and testing results for the simplified LSTM I in Mackey-Glass time series

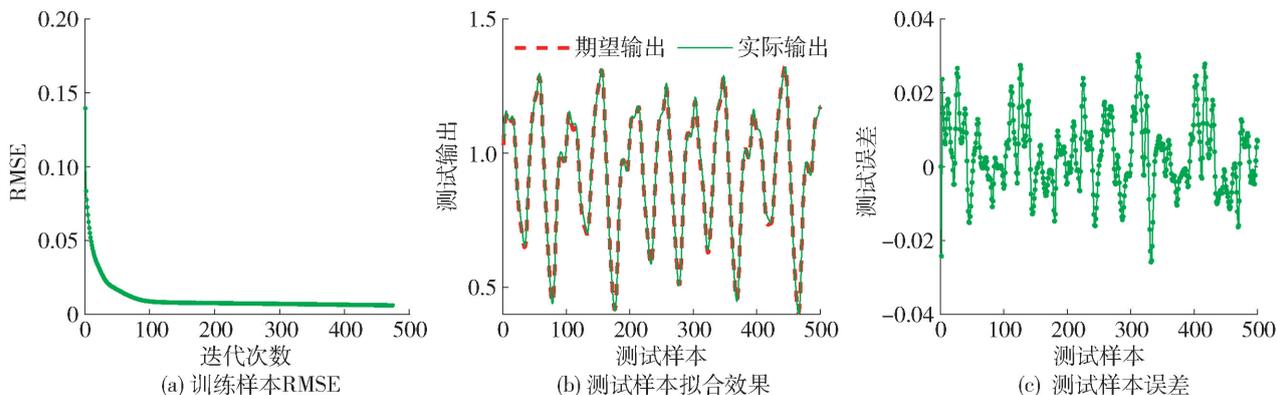


图6 LSTM-简化型 II 对 Mackey-Glass 时间序列预测的训练过程及测试效果

Fig. 6 Training process and testing results for the simplified LSTM II in Mackey-Glass time series

表2 Mackey-Glass 时间序列预测模型性能对比

Table 2 Performance comparison of different models for Mackey-Glass time series

模型	RMSE		训练 时间/s	更新参 数个数
	训练 样本	测试 样本		
标准 LSTM	0.006 0	0.007 1	55.27	600
LSTM-变体 I	0.006 0	0.007 2	38.11	450
LSTM-变体 II	0.006 0	0.006 6	38.63	280
LSTM-变体 III	0.006 0	0.006 5	43.64	200
LSTM-简化型 I	0.006 0	0.007 0	18.33	210
LSTM-简化型 II	0.006 0	0.007 1	21.58	150

注:黑体代表本文提出的简化型 LSTM 模型。

3.2 污水处理中 BOD 预测

BOD 是污水处理中评价水质的重要指标之一,具有高度的非线性、大时变的特征,很难及时准确地预测其质量浓度^[30]。本文利用 LSTM-简化型 I、II 神经网络对污水处理过程中的 BOD 进行建模,选取前 8 时刻的 BOD 质量浓度作为输入向量,下一时刻

的 BOD 质量浓度作为输出变量。

选取北京市某污水厂的数据进行仿真,获得 357 组按照时间顺序进行排列的样本,选取前 250 组作为训练样本,后 107 组作为测试样本,将所有样本归一化至 $[-1, 1]$ 输入模型,并将样本反归一化后输出。设定状态单元维度为 15,学习率 η 为 0.01,期望训练样本的 RMSE 为 0.060 0,最大迭代次数为 2 000 次。当训练样本的 RMSE 达到期望训练样本的 RMSE 或最大迭代次数时,停止参数更新。

LSTM-简化型 I、II 的训练过程分别如图 7、8 中的(a)所示。从图中可以看出,训练样本的 RMSE 能够达到期望训练样本的 RMSE。其测试结果反归一化后输出并计算测试误差,分别如图 7、8 中的(b)(c)所示。从图中可以看出,简化型 LSTM 神经网络的测试结果均可以达到较好的拟合效果。

从表 3 的对比结果可以看出,在达到期望训练样本的 RMSE、停止参数更新的情况下,LSTM-变体 I 比 LSTM-变体 II、III 需要更新较多的参数个数但需要较短的训练时间,同时 LSTM-简化型 I 在需要更新的参数个数与 LSTM-变体 III 相同的情况下训

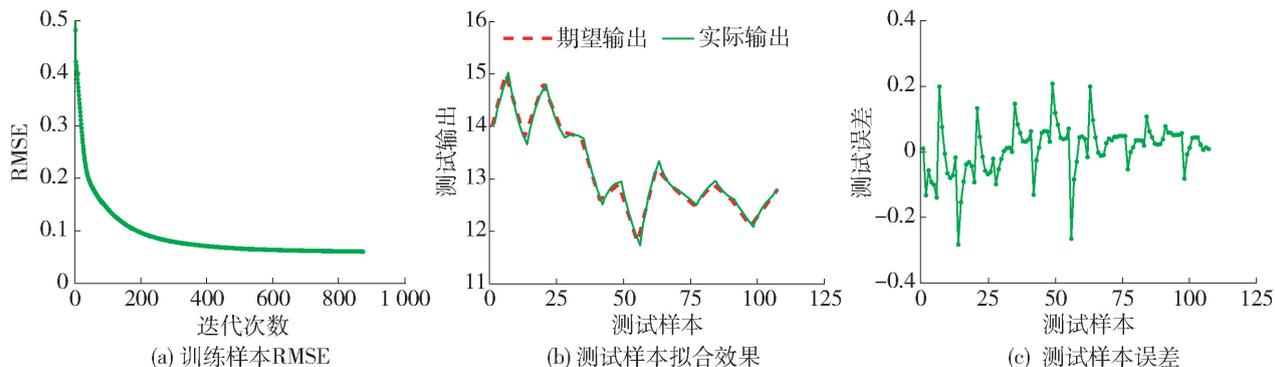


图7 LSTM-简化型 I 对 BOD 质量浓度预测的训练过程及测试效果

Fig. 7 Training process and testing results for the simplified LSTM I in BOD mass concentration prediction

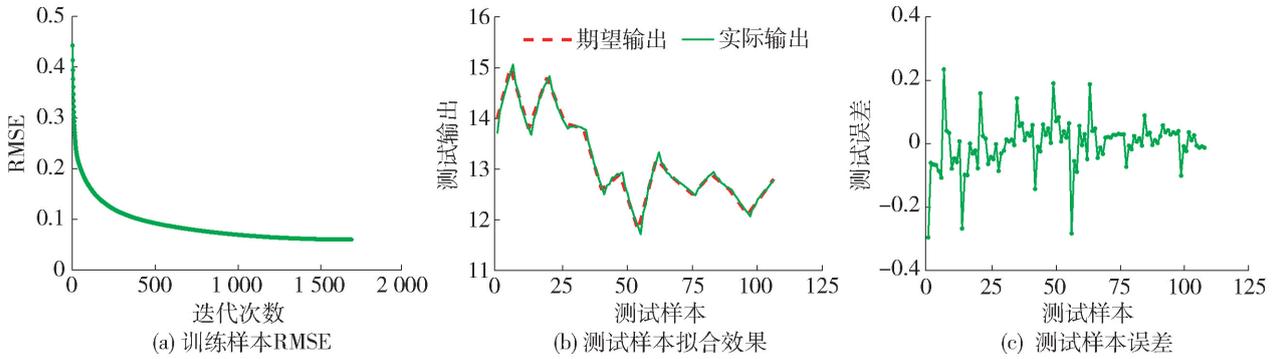


图8 LSTM-简化型 II 对 BOD 质量浓度预测的训练过程及测试效果

Fig. 8 Training process and testing results for the simplified LSTM II in BOD mass concentration prediction

练时间显著缩短,均说明门结构精简对简化 LSTM 神经网络的效果更显著. 通过实验结果分析可以得出,本文提出的 LSTM-简化型 I、II 神经网络能够在精度相当的情况下进一步缩短训练时间,对 BOD 质量浓度快速预测.

表3 BOD 质量浓度预测模型性能对比

Table 3 Performance comparison of different models for BOD mass concentration prediction

模型	RMSE		训练时间/s	更新参数个数
	训练样本	测试样本		
标准 LSTM	0.060	0.059	99.35	1440
LSTM-变体 I	0.060	0.064	30.11	1080
LSTM-变体 II	0.060	0.063	51.93	480
LSTM-变体 III	0.060	0.066	62.75	360
LSTM-简化型 I	0.060	0.062	14.44	360
LSTM-简化型 II	0.060	0.065	29.98	270

注:黑体代表本文提出的简化型 LSTM 模型.

4 结论

1) 简化型 LSTM 神经网络能够在不显著降低模型精度的情况下减少计算复杂度,缩短训练时间.

2) 基于简化型 LSTM 神经网络的时间序列预测方法能够实现时间序列的高效预测.

参考文献:

[1] TANG X, LI C, BAO Y. Seasonal forecasting of agricultural commodity price using a hybrid STL and ELM method: evidence from the vegetable market in China[J]. Neurocomputing, 2018, 275: 2831-2844.

[2] SPECHT D F. A general regression neural network[J].

IEEE Transactions on Neural Networks, 1991, 2(6): 568-576.

[3] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 1994, 5(2): 157-166.

[4] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[5] FISCHER T, KRAUSS C. Deep learning with long short-term memory networks for financial market predictions[J]. European Journal of Operational Research, 2017, 270(2): 654-669.

[6] YAN H J, OUYANG H B. Financial time series prediction based on deep learning [J]. Wireless Personal Communications, 2018, 102(2): 683-700.

[7] 史建楠, 邹俊忠, 张见, 等. 基于 DMD-LSTM 模型的股票价格时间序列预测研究[J]. 计算机应用研究, 2020, 37(3): 662-666.

SHI J N, ZOU J Z, ZHANG J, et al. Research of stock price prediction based on DMD - LSTM model [J]. Application Research of Computers, 2020, 37(3): 662-666. (in Chinese)

[8] SAGHEER A, KOTB M. Time series forecasting of petroleum production using deep LSTM recurrent networks [J]. Neurocomputing, 2018, 323: 203-213.

[9] 罗向龙, 李丹阳, 杨彧, 等. 基于 KNN-LSTM 的短时交通流预测[J]. 北京工业大学学报, 2018, 44(12): 1521-1527.

LUO X L, LI D Y, YANG Y, et al. Short-term traffic flow prediction based on KNN-LSTM [J]. Journal of Beijing University of Technology, 2018, 44(12): 1521-1527. (in Chinese)

[10] BANDARA K, BERGMEIR C, HEWAMALAGE H. LSTM-MSNet: leveraging forecasts on sets of related time series with multiple seasonal patterns [J]. IEEE Transactions on Neural Networks and Learning Systems,

- 2021, 32(4): 1586-1599.
- [11] 葛瑞, 王朝晖, 徐鑫, 等. 基于多层卷积神经网络特征和双向长短时记忆单元的行为识别[J]. 控制理论与应用, 2017, 34(6): 790-796.
- GE R, WANG Z H, XU X, et al. Action recognition with hierarchical convolutional neural networks features and bi-directional long short-term memory model [J]. Control Theory & Applications, 2017, 34(6): 790-796. (in Chinese)
- [12] HOUDT G V, MOSQUERA C, GONZALO N. A review on the long short-term memory model [J]. Artificial Intelligence Review, 2020(53): 5929-5955.
- [13] GREFF K, SRIVASTAVA R K, JAN K. LSTM: a search space odyssey[J]. IEEE Transactions on Neural Networks & Learning Systems, 2016, 28(10): 2222-2232.
- [14] 陈宙斯, 胡文心. 简化 LSTM 的语音合成[J]. 计算机工程与应用, 2018, 54(3): 131-135.
- CHEN Z S, HU W X. Speech synthesis using simplified LSTM [J]. Computer Engineering and Applications, 2018, 54(3): 131-135. (in Chinese)
- [15] CHO K, VAN MERRIENBOER B, BAHDANAU D, et al. On the properties of neural machine translation: encoder-decoder approaches [EB/OL]. [2020-06-27]. <https://arxiv.org/abs/1409.1259>.
- [16] GERS F A, SCHRAUDOLPH N N. Learning precise timing with LSTM recurrent networks [J]. Journal of Machine Learning Research, 2003, 3(1): 115-143.
- [17] ZHOU G B, WU J X, ZHANG C L, et al. Minimal gated unit for recurrent neural networks [J]. International Journal of Automation and Computing, 2016, 13(3): 226-234.
- [18] OLIVER O, RODRIGUEZ A. Simplified LSTM unit and search space probability exploration for image description [C] // IEEE International Conference on Information. Piscataway: IEEE, 2016: 1-5.
- [19] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [EB/OL]. [2020-07-30]. <https://arxiv.org/abs/1412.3555>.
- [20] LU Y Z, SALEM F M. Simplified gating in long short-term memory (LSTM) recurrent neural networks [EB/OL]. [2020-08-12]. <https://arxiv.org/abs/1701.03441>.
- [21] RAHUL D, FATHI M. Gate-variants of gated recurrent unit (GRU) neural networks [C] // IEEE International Midwest Symposium on Circuits and Systems. Piscataway: IEEE, 2017: 1597-1600.
- [22] JOEL C, FATHI M. Simplified minimal gated unit variations for recurrent neural networks [C] // IEEE International Midwest Symposium on Circuits and Systems. Piscataway: IEEE, 2017: 1593-1596.
- [23] GRAVES A, MOHAMED A R, HINTON G. Speech recognition with deep recurrent neural networks [C] // 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2013: 6645-6649.
- [24] WERBOS P J. Backpropagation through time: what it does and how to do it [J]. Proceedings of the IEEE, 1990, 78(10): 1550-1560.
- [25] RODRIGUES F, MARKOU I, PEREIRA F C. Combining time-series and textual data for taxi demand prediction in event areas: a deep learning approach [J]. Information Fusion, 2019, 49: 120-129.
- [26] YANG J, GUO Y, ZHAO W. Long short-term memory neural network based fault detection and isolation for electro-mechanical actuators [J]. Neurocomputing, 2019, 360: 85-96.
- [27] HONG J, WANG Z, YAO Y. Fault prognosis of battery system based on accurate voltage abnormality prognosis using long short-term memory neural networks [J]. Applied Energy, 2019, 251: 1-14.
- [28] CHEN H, GONG Y, HONG X. Online modeling with tunable RBF network [J]. IEEE Transactions on Cybernetics, 2013, 43(3): 935-947.
- [29] QIAO J F, HAN H G. Identification and modeling of nonlinear dynamical systems using a novel self-organizing RBF-based approach [J]. Automatica, 2012, 48(8): 1729-1734.
- [30] 韩改堂, 乔俊飞, 韩红桂. 基于自适应递归模糊神经网络的污水处理控制 [J]. 控制理论与应用, 2016, 33(9): 1252-1258.
- HAN G T, QIAO J F, HAN H G. Wastewater treatment control method based on adaptive recurrent fuzzy neural network [J]. Control Theory & Applications, 2016, 33(9): 1252-1258. (in Chinese)

(责任编辑 梁洁)