

文章编号: 2095-2163(2022)11-0092-05

中图分类号: TN929.5

文献标志码: A

# 长短期记忆网络在气温预测中的应用

花凡, 李莉, 蔡鑫楠, 徐健

(天津职业技术师范大学, 天津 300222)

**摘要:** 气温变化对人们的正常出行、社会发展与生态环境都有着很大的影响,而天气和气候具有一定的随机性和无序性,是难以预测的复杂模型。本文采用长短期记忆网络深度学习模型对气温时间序列进行建模及预测,并将其与门循环单元模型、神经网络模型预测结果进行比较。结果显示, LSTM 预测的均方误差比其它 2 个模型预测精度更高,且在处理较大数据集时具有更好的性能。

**关键词:** LSTM; 气温预测; GRU; DNN; 深度学习

## Application of long short-term memory networks in temperature prediction

HUA Fan, LI Li, CAI Xinnan, XU Jian

(Tianjin University of Technology and Education, Tianjin 300222, China)

**[Abstract]** Temperature change has a great impact on people's normal travel, social development and ecological environment, while weather and climate have certain randomness and disorder, which are the complex model difficult to be predicted. In this paper, long short-term memory (LSTM) networks deep learning model is used to construct the model and predict temperature time series, and is compared with the prediction results of Gated Recurrent Unit (GRU) model and Deep Neural Networks (DNN) model. The simulation show that the Mean Squared Error (MSE) of LSTM prediction is more accurate than the other two models, and has better performance in processing large data sets.

**[Key words]** LSTM; temperature prediction; GRU; DNN; deep learning

## 0 引言

长期以来,空气温度都是影响着人们生活的重要条件。一方面,会影响到人们的日常活动,人们通过天气预报提前知道未来的天气,计划自己的出行、着装及其它活动。另一方面,空气温度也影响着农业、社会产业、制造业等行业的发展。特别是农业生产,面对突如其来的气候变化,如果不能采取及时、恰当的保护措施,就有可能导致农作物减产、甚至使当季的农作物全毁。

人类对空气温度的预测始于建立大气运动方程组,在一定的初始值和边界值的条件下,对这个方程组进行积分,来预测未来的天气温度。到了 20 世纪,人们开始思考如何让天气温度预测的准确性变高。Richardson<sup>[1]</sup>动用了大量的人力进行了天气预报的数值化研究,但是却未能取得理想结果,与实际天气温度有较大的差距。随着现代科技的不断发展,世界上第一台计算机 ENIAC 成功问世,1950 年,Charney 基于滤去高频波后的大气运动方程组,利用 ENIAC 首次实现了对天气温度 24 h 的数值预

测<sup>[2]</sup>。此后,数值天气预报也取得可观进展,已然成为预报天气的主要手段。此后不久,科研人员又把统计学应用到天气温度研究领域,采集历史天气温度的数据,通过数学统计等原理建立更准确的预测模型。正如 2021 年获得诺贝尔物理学奖的 3 位科学家的研究结果指出的一样:天气多变且混乱,是一个充满随机性和无序性的复杂模型<sup>[3-5]</sup>。

在智能化浪潮奔涌而至的背景下,深度学习模型在人脸识别、语音识别应用中的准确度已经赶超人类<sup>[6]</sup>。而在天气预测方面,通过深度学习模型,可以对观测到的天气数据进行更加全面的分析。并且对未来的气温进行更加精准的预测。

## 1 气温预测模型

### 1.1 LSTM 模型

长短期记忆网络(Long short-term memory networks, LSTM)<sup>[7-9]</sup>是一种特殊的递归神经网络(Recurrent Neural Network, RNN),由于 RNN 在训练长序列的过程中会产生梯度消失与梯度爆炸的问题,人们就在 RNN 的基础上引入了细胞状态,于是

作者简介:花凡(1998-),男,硕士研究生,主要研究方向:人工智能与深度学习。

通讯作者:花凡 Email:499227148@qq.com

收稿日期:2022-02-13

产生了 LSTM。所以 LSTM 在处理长序列时有较好的表现。LSTM 与 RNN 的结构很相似,最大的不同就是隐藏层。LSTM 的隐藏层具有特殊的记忆单元 (cell) 结构和三门机制,通过门控状态来控制传输状态,记住需要长时间记忆的,忘记不重要的信息,而不像普通的 RNN 那样仅有一种记忆叠加方式。故而 LSTM 在一定程度上克服了 RNN 的梯度问题。所有 RNN 都具有一种重复神经网络模块的链式的形式。在标准的 RNN 中,这个重复的模块只有一个非常简单的结构,比如一个 tanh 层。LSTM 也是同样的结构,但是重复的模块拥有一个不同的结构。不同于单一神经网络层,这里是有 4 个,以一种非常特殊的方式进行交互。RNN 与 LSTM 的结构对比如图 1 与图 2 所示。

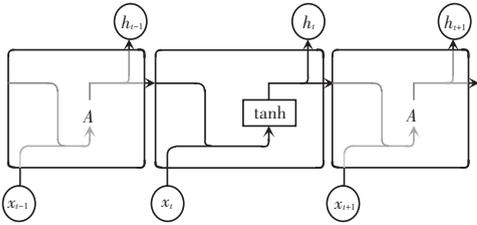


图 1 RNN 结构

Fig. 1 Structure of RNN

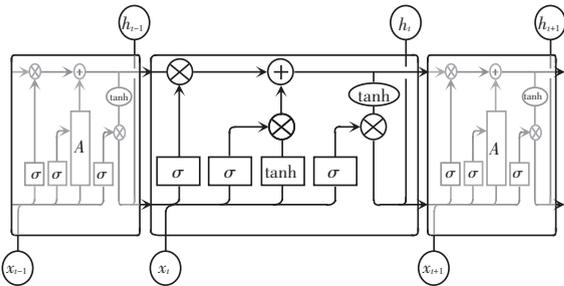


图 2 LSTM 结构

Fig. 2 Structure of LSTM

LSTM 的记忆模块主要包含了遗忘门 (forget gate)、输入门 (input gate)、输出门 (output gate) 和记忆单元 (cell)。LSTM 的工作流程可做分述如下:

**Step 1** 遗忘门会读取  $t-1$  时刻的隐藏层输出  $h_{t-1}$  和当前输入  $x_t$ , 再通过 *sigmoid* 函数计算出遗忘门的输出  $f_t$ , 来决定需要丢弃的信息。其中,  $f_t$  的计算公式见如下:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

**Step 2** 输入门根据  $t-1$  时刻的隐藏层输出  $h_{t-1}$  和当前输入  $x_t$  计算当前输入有多少保存到  $C_t$ , 并通过 tanh 生成新的  $\tilde{C}_t$ 。其中,  $\tilde{C}_t \in (-1, 1)$ 。研究后推得的计算公式分别如下:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \sigma(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

进一步,还可推得  $C_t$  的数学定义式具体如下:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

**Step 3** 输出门控制  $C_t$  输出到  $h_t$ , 使用 tanh 对  $C_t$  压缩到  $(-1, 1)$ , 再通过 *sigmoid* 函数得出最终输出  $h_t$ 。为此可推导得到的数学公式表示如下:

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = O_t * \tanh(C_t) \quad (6)$$

其中,  $W_f, W_i, W_c, W_o$  均为权重矩阵;  $b_f, b_i, b_c, b_o$  均为各个网络层的偏执向量;  $\sigma$  是 *sigmoid* 函数。

## 1.2 GRU 模型

门循环单元 (Gated Recurrent Unit, GRU) [10-12] 和 LSTM 一样,都是基于门控制的循环神经网络,但是结构比 LSTM 更简单,并且也可以解决 RNN 中的梯度问题,其结构如图 3 所示。

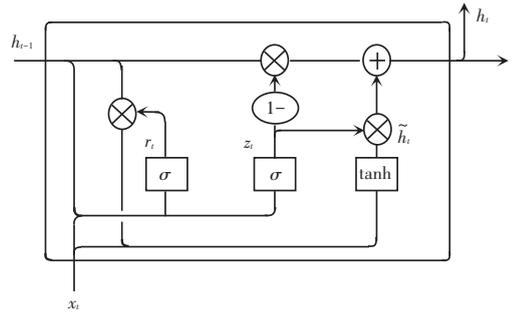


图 3 GRU 结构图

Fig. 3 Structure of GRU

GRU 将 LSTM 的遗忘门与输入门结合起来,所以 GRU 只有 2 个门:更新门 (update gate) 和重置门 (reset gate)。由图 3 可知,  $z_t$  与  $r_t$  分别表示更新门与重置门,  $r_t$  控制  $\tilde{h}_t$  保留  $h_{t-1}$  传递的记忆,  $z_t$  在功能上实现了 LSTM 中遗忘门和输入门的功能,决定丢弃与保存的信息,两者结合得到最后输出的隐藏层信息。这里,将依次给出计算当前输出状态的数学公式重点如下:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (7)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (8)$$

$$\tilde{h}_t = \tanh(W_h \cdot [h_{t-1}, x_t] + b_h) \quad (9)$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \quad (10)$$

$$y_t = \sigma(W_o \times h_t) + b_o \quad (11)$$

其中,  $\sigma$  表示 *sigmoid* 函数;  $W_z, W_r, W_h$  和  $W_o$  表示各权重矩阵;  $h$  表示的是隐藏层信息;  $b_z, b_r, b_h$  和  $b_o$  均为各个网络层的偏执向量; tanh 表示输出的激活函数。

### 1.3 DNN 模型

深度神经网络(Deep Neural Networks, DNN)<sup>[13-15]</sup>内部的神经网络层可以分为3类:输入层、隐藏层和输出层,如图4所示。总的来说,第一层是输入层,中间数层都是隐藏层,而最后一层是输出层。

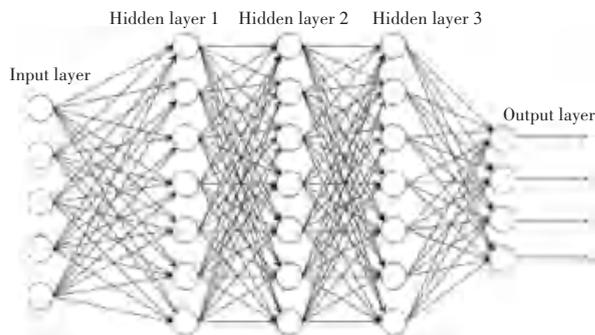


图4 DNN 的结构

Fig. 4 Structure of DNN

一般来说,神经网络的输入层并不计算在内,隐藏层与输出层共有几层,这个神经网络就是几层。故分析可知,图4中的神经网络就是4层。

DNN 中层与层之间是全连接的,第  $n$  层中的任意一个神经元在第  $n + 1$  层中都有与之连接的神经元。且彼此之间的关系都是由一个线性关系  $z = \sum w_i x_i + b_i$  和激活函数 *sigmoid* 函数连接而成。

假设  $l - 1$  层有  $m$  个神经元,则对于第  $l$  层的第  $j$  个神经元的输出  $a_j^l$  可由如下公式计算求得:

$$a_j^l = \sigma(z_j^l) = \sigma\left(\sum_{k=1}^m w_{jk}^l a_k^{l-1} + b_j^l\right) \quad (12)$$

## 2 气象数据来源和数据预处理

### 2.1 气象数据来源

气象时间序列来自于 Max Planck Institute for Biogeochemistry<sup>[16]</sup> 收集的气象时间序列数据集,该数据集含有 14 个不同的特征,例如气温、压强、温度与密度等。从 2003 年开始,每隔 10 min 就收集一次天气信息。本文使用 2009~2016 年之间收集的数据。

### 2.2 数据预处理

为了减少 LSTM 深度学习模型的训练时长,使 LSTM 深度学习模型的预测更加准确,需要缩放特征,对数据进行归一化处理,将值缩放到(0,1)范围内。本文采用的是  $z$ -score 标准化,相应的数学公式可写为:

$$Y = \frac{X - \bar{X}}{S} \quad (13)$$

其中,  $\bar{X}$  是所有数据的平均值;  $S$  是所有数据的标准差;  $Y$  是所有数据经过  $z$ -score 标准化后的值。

## 3 算法结构与流程

### 3.1 算法结构

本文提出的 LSTM 深度学习模型是由一个输入层、一个隐藏层、一个输出层构成。输入层由 32 个神经元组成,经过 72 个神经元组成的隐藏层处理后输出为 16 个神经元。为了比较 LSTM 模型对气温预测的准确性,本文设置了 DNN 与 GRU 两个模型为对照组,保证 DNN 与 GRU 模型输入层、隐藏层、输出层的层数与神经元个数和 LSTM 模型相同,比较 3 个模型的预测精度。

本文在 LSTM 模型中加入了取最大值(Rectified Linear Units, *relu*) 激活函数,如式(14)所示:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases} \quad (14)$$

可以在一定程度上减少运算量,避免层数增加的问题。当神经元取值为负数时, *relu* 会使其输出为 0,这样就会使网络变得稀疏,简化了参数之间的联系,可以有效地减少过拟合。

### 3.2 算法流程

至此,研究给出的算法流程具体如图5所示。这里,对算法的设计流程可进行阐释分述如下:

- Step 1** 原始数据输入,将气象数据集按照 7 : 3 划分成数据集与验证集。
- Step 2** 数据标准化,将数据缩放到(0,1)范围内。
- Step 3** 构建 LSTM 模型、DNN 模型、GRU 模型。
- Step 4** 将 Step2 中的数据集分别输入 3 个模型进行训练,寻找最优的训练次数。
- Step 5** 输出每个模型的 *MSE* 结果进行比较。

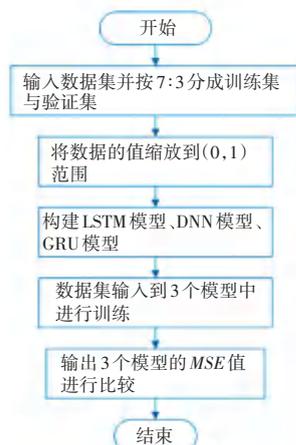


图5 算法具体流程

Fig. 5 Specific flowchart of the algorithm

## 4 实验结果与分析

### 4.1 评价指标

对于预测结果好坏的判定方法,选用 *MSE*。指

标运算可由下式求得:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (15)$$

其中,  $\hat{y}_i$  是时间序列预测值;  $y_i$  是待预测时间序列的真实值。

$MSE$  的值越小, 说明 LSTM 预测就越精确。

### 4.2 预测结果分析

本文搭建的 LSTM 预测模型由 1 个输入层、3 个隐藏层、1 个输出层构成。每 10 min 收集一次数据, 所以 1 h 有 6 个数据。本文用 5 天的数据, 也就是 720 个数据作为训练集来预测, 用 12 h、共 72 个数据作为测试集。实验使用服务器配置为 i5 8 GRAM, 1050TI 显卡, 系统为 Windows10。深度学习框架使用 Tensor Flow。在训练模型的过程中, 为了加快 LSTM 深度学习模型的学习速率与收敛速率, 本文加入 *RMSprop* 优化器进行优化。训练次数 (*epoch*) 如果太少, 将不足以提取全部特征, 太多又会发生过拟合的情况。经过多次实验, 最终确定 *epoch* 为 10。训练过程中损失函数的变化如图 6 所示。图 6 中, 横坐标为 *epoch* 的次数, 可以看到 LSTM 模型训练较好。

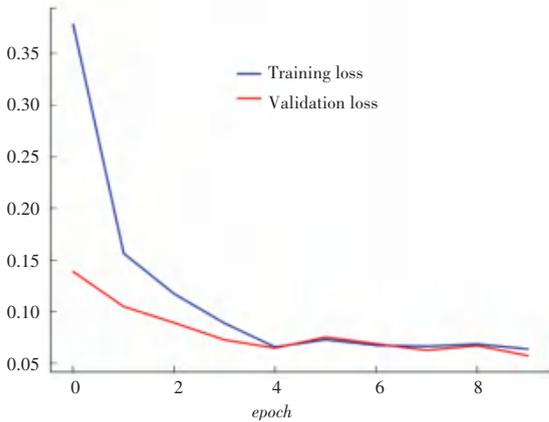


图 6 损失函数变化曲线

Fig. 6 Variation curve of loss function

为了验证模型的精确度, 分别用了 DNN 模型、GRU 模型对天气温度进行预测。训练集与测试集保持一致。DNN 模型、GRU 模型与 LSTM 模型预测效果如图 7 所示, 横坐标为预测的 12 h。整个训练过程中, 3 个模型的性能指标见表 1。

表 1 3 种模型的  $MSE$  结果

Tab. 1  $MSE$  results for the three models

预测方法	训练集	测试集
LSTM	0.064 1	0.057 7
DNN	0.073 8	0.065 1
GRU	0.070 1	0.061 9

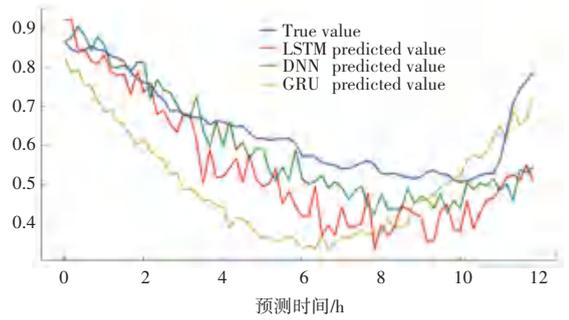


图 7 3 种模型预测结果

Fig. 7 Prediction results of three models

由图 7 可知, 3 种方法均在真实值附近波动, 都可以预测天气温度。通过表 1 可以看出, LSTM 的训练集与测试集的  $MSE$  都是最低的, 说明 LSTM 深度学习模型的预测效果更好, 预测精度更高。

### 5 结束语

本文采用 LSTM 建立模型, 通过前 5 天的数据来预测未来 12 个小时, 并通过加入 *RMSprop* 优化器, 有效提高各性能指标来预测精度。从结果上来说, 与 GRU 和 DNN 模型相比, 基于 LSTM 建立的模型预测的结果具有更高的预测精度。本文研究表明, LSTM 在处理长时间序列上具有更好的性能, 适合用于气温预测等应用。并且在处理较大数据集时具有更好的性能。

### 参考文献

- [1] RICHARDSON L F. Weather prediction by numerical process[M]. Cambridge: Cambridge University Press, 1922.
- [2] 孙健, 曹卓, 李恒, 等. 人工智能技术在数值天气预报中的应用[J]. 应用气象学报, 2021, 32(01): 1-11.
- [3] BROCCOLI A J, MANABE S. Syukuro manabe receives 2010 William Bowie Medal [J]. Eos, Transactions American Geophysical Union, 2011, 92(1): 1- 5.
- [4] MANABE S, WETHERALD R T. Thermal equilibrium of the atmosphere with a given distribution of relative humidity [J]. Journal of the Atmospheric Sciences, 1967, 24(3): 241-259.
- [5] HASSELMANN K. Optimal fingerprints for the detection of time-dependent climate change[J]. Journal of Climate, 1993, 6(10): 1957-1971.
- [6] 杨函. 基于深度学习的气象预测研究[D]. 哈尔滨: 哈尔滨工业大学, 2017.
- [7] PENG Zhichao, DANG Jianwu, MASASHI U, et al. Multi-resolution modulation-filtered cochleagram feature for LSTM-based dimensional emotion recognition from speech[J]. Neural Networks, 2021, 140: 261-273.
- [8] 王鑫, 吴际, 刘超, 等. 基于 LSTM 循环神经网络的故障时间序列预测[J]. 北京航空航天大学学报, 2018, 44(04): 772-784.