

程鹏宇,赵嘉,韩龙哲,等.双向多尺度 LSTM 的短时温度预测[J].江西师范大学学报(自然科学版),2022,46(2):134-139.
CHENG Pengyu,ZHAO Jia,HAN Longzhe,et al.The short-term temperature prediction based on bidirectional multi-scale LSTM[J].
Journal of Jiangxi Normal University(Natural Science),2022,46(2):134-139.

文章编号:1000-5862(2022)02-0134-06

双向多尺度 LSTM 的短时温度预测

程鹏宇¹,赵嘉^{1*},韩龙哲¹,张翼英²,武延年³

(1.南昌工程学院信息工程学院,江西 南昌 330099;2.天津科技大学人工智能学院,天津 300457;
3.深圳市国电科技通信有限公司,广东 深圳 518000)

摘要:针对长短期记忆网络(long short-term memory, LSTM)无法有效提取温度数据的多尺度特征和反向特征的问题,该文提出了一种双向多尺度跳跃 LSTM(bidirectional multi-scale skip long short-term memory, BMS-LSTM)的短时温度预测模型.该模型以 LSTM 为核心单元,采用双向深层网络结构提取反向特征;根据温度数据日的周期性设置跳跃连接数提取多尺度特征,解决了指数增长的跳跃连接数后期跳跃尺度过大的问题;最后使用全连接层进行特征融合预测.实验结果表明:BMS-LSTM 成功提取了温度数据的多尺度特征和反向特征,预测均值误差仅为 3.890,优于对比模型,是一种有效的短时温度预测模型.

关键词:长短期记忆网络;温度预测;多尺度;反向特征;跳跃连接;特征融合

中图分类号:TP 183 **文献标志码:**A **DOI:**10.16357/j.cnki.issn1000-5862.2022.02.04

0 引言

全球变暖引起气候变化,导致极端天气事件急剧增加^[1].极端天气对人类生产和生存环境的影响逐渐增大,极端高温与低温带来的损失非常严重.极端高温可能会引发中暑、心脑血管疾病、系统呼吸疾病等病症^[2].因此,深入研究温度变化规律,提高温度预测的准确度,具有重要的现实意义,有关温度预测的研究正受到人们的普遍关注^[3].

传统的温度预测是通过建立大气模型模拟天气,搭建温度预测系统^[4],建立的模型需要精确模拟真实的气象变化,但该方法实际操作比较复杂.目前,大多数研究方法是基于数学模型的温度预测,如支持向量回归^[5]、随机森林^[6]、概率图^[7]等,但这类模型的预测精度不高.随着数据分析^[8]、回归分析及深度学习^[9-10]在众多领域中的发展,温度预测也有了新方法.为了提升模型的预测精度,学者们提出了基于神经网络的温度预测模型^[11-13].

神经网络(recurrent neural network, RNN)^[14]是以序列数据为输入,进行递归且按链式传递的递归神经网络.神经网络具有记忆性、参数共享的特点,在处理时序数据时具有一定的优势,但存在梯度消失、爆炸和长期依赖等问题.1997年 S. Hochreiter 等^[15]提出长短期记忆网络(long short-term memory, LSTM).LSTM 在 RNN 的隐藏状态中添加了门控单元,解决了长期依赖和梯度消失、爆炸问题.近年来,学者们开始关注神经网络如何提取序列数据的多尺度特征,如时钟神经网络(clockwork recurrent neural network, CW-RNN)^[16]、分层多尺度循环神经网络(hierarchical multiscale recurrent neural networks, HM-RNN)^[17]、多尺度 RNN(dilated recurrent neural networks, DiRNN)^[18]、多尺度跳跃深度 LSTM(multi-scale skip deep long short-term, MSD-LSTM)^[19]等.DiRNN 通过添加多个指数增长多尺度循环跳跃连接提取在数据中的多尺度特征;MSD-LSTM 选择数据的年、月、日的周期数作为跳跃连接数,上述 2 种模型的预测精度都得到了明显提升.但这些模型

收稿日期:2021-09-25

基金项目:国家自然科学基金(52069014,61962036)和江西省杰出青年基金(2018ACB21029)资助项目.

通信作者:赵嘉(1981—),男,安徽桐城人,教授,博士,主要从事大数据分析、人工智能理论、深度学习研究. E-mail:zhaojia925@163.com

依然存在如下缺陷:(i)只考虑数据的多尺度特征,忽视了数据的反向特征;(ii)月周期或按照指数增长的跳跃连接数,在循环神经网络隐含层中跳跃周期太长,存在较大的局限性,不适合对短期数据进行建模。

针对上述问题,本文提出了双向多尺度跳跃 LSTM (bidirectional multi-scale skip long short-term memory, BMS-LSTM) 短时温度预测模型。BMS-LSTM 模型使用双向深层网络结构,以 LSTM 为核心,设置多个跳跃连接数提取多尺度时序特征和反向特征。区别于指数增长的跳跃连接数,该模型使用日周期对温度数据进行建模;使用双向网络结构,提取数据的反向特征。

1 双向 LSTM

1.1 LSTM

RNN 具有参数共享、记忆等特点,对序列数据建模具有巨大优势。RNN 隐含层通过迭代处理信息,且在更新迭代时依赖前一隐含状态,保存了数据之间的信息。然而,在使用 RNN 对序列数据进行建模时,依然存在如下缺陷:(i)无法或者很难记忆长期依赖;(ii)在反向传播训练 RNN 时会发生梯度消失和爆炸。为了克服这些困难,S. Hochreiter 等^[15]提出的 LSTM 弥补了 RNN 的缺陷。LSTM 与 RNN 区别在于细胞状态(见图 1)。LSTM 设计了 3 个门结构来控制状态信息的流通和损失,包括遗忘门 f_t 、输入门 i_t 及输出门 o_t 。

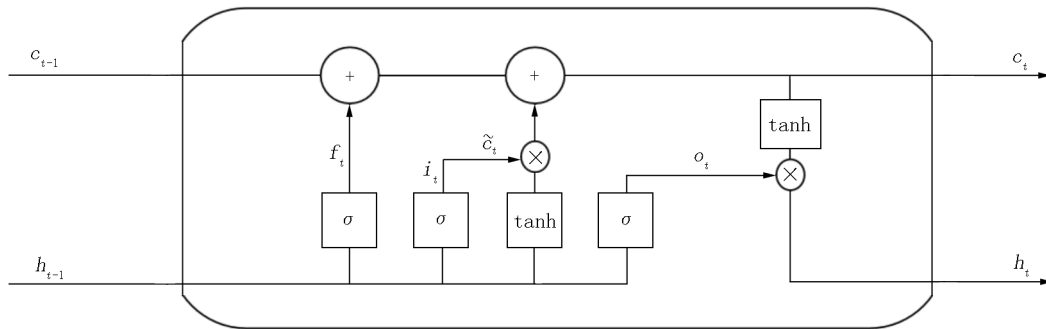


图 1 LSTM 细胞状态

遗忘门 f_t 表示在 c_{t-1} 中有多大比例计算 c_t , 输入门 i_t 控制当前时间步的输入状态, 输出门 o_t 控制当前时间步的输出信息, \tilde{c}_t 表示细胞单元更新状态, 它们的计算方法分别为

$$\begin{aligned} f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f), \\ i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i), \\ \tilde{c}_t &= \tan(W_c[h_{t-1}, x_t] + b_c), c_t = f_t c_{t-1} + i_t \tilde{c}_t, \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o), h_t = o_t \tanh(c_t), \end{aligned}$$

其中 σ 为 sigmoid() 函数, $[\]$ 为矩阵拼接, W_f, W_i, W_c, W_o 分别为遗忘门、输入门和输出门的权重, b_f, b_i, b_c, b_o 分别是各门的偏置值。sigmoid(x) 函数将实数 x 映射到 (0,1) 区间, 计算方法为

$$\text{sigmoid}(x) = 1/(1 + e^{-x}).$$

tanh(x) 为双曲正切函数, 它将实数 x 映射到 (-1,1) 之间, 计算方法为

$$\tanh(x) = (e^x - e^{-x})/(e^x + e^{-x}).$$

1.2 双向 LSTM

RNN 处理时序数据的顺序是由前向后单向进行的, 即当前时刻的输出依赖之前时刻的状态。但某些数据当前时刻的输出不仅与之前的状态相关, 而且和未来的状态也息息相关。对于此类数据, RNN

无法提取其反向特征, 这导致预测精度降低。为弥补此缺陷, M. Schuster 等^[20]提出了双向 RNN, 双向 RNN 的隐藏状态分为正序和逆序 2 种状态。正序状态是指当前状态依赖之前时刻的隐藏状态, 逆序状态是指当前状态依赖后面时间步的隐藏状态。因此, 双向 RNN 能提取数据的反向特征, 从而提高了预测精确度。图 2 为按时间步展开的双向 RNN 结构图, 各层中计算方法分别为

$$\begin{aligned} \vec{h}_t &= H(w_{x\vec{h}}x_t + w_{h\vec{h}}\vec{h}_{t-1} + b_{\vec{h}}), \\ \overleftarrow{h}_t &= H(w_{x\overleftarrow{h}}x_t + w_{h\overleftarrow{h}}\overleftarrow{h}_{t-1} + b_{\overleftarrow{h}}), \\ y_t &= G(w_{\vec{h}y}\vec{h}_t + w_{\overleftarrow{h}y}\overleftarrow{h}_t + b_y), \end{aligned}$$

其中 $H(\)$ 、 $G(\)$ 表示激活函数, 根据实际情况可任意选取。在本文中 $H(\)$ 选用 tanh() 函数, $G(\)$ 选用 sigmoid() 函数。

A. Graves 等^[21]提出了双向 LSTM (bidirectional long short-term memory neural network, BiLSTM) 模型, 并将其应用于语音信息建模。双向 LSTM 结合了双向 RNN 和 LSTM 的优势, 在提取双向特征的同时, 能有效解决长期依赖和梯度消失、爆炸问题, 提升了模型的预测精度。双向 LSTM 凭借其优势, 得到了广泛应用, 如股票预测^[22]、负荷预测^[23]和温度预测^[24]等。

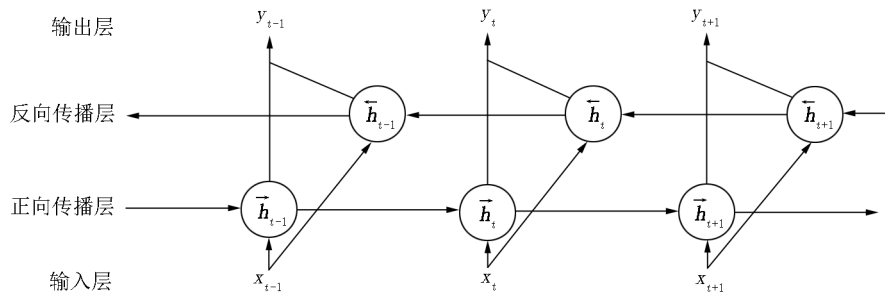


图2 双向 RNN 结构图

2 BMS-LSTM 短时温度预测模型

BMS-LSTM 短时温度预测模型可同时提取在序列数据中的双向特征和多尺度特征. 温度数据存在明显的周期特性, 主要体现在年、月、日时间期限上的循环特性. 本文实验使用 720 h (30 d) 的温度数据来预测未来 24 h 的温度, 即模型输入的时间步长为 720, 当模型的输入只有 30 d 的数据时, 无法体现年、月的循环特性, 因此设计出的模型仅考虑提取日循环周期的多尺度特征. 模型的第 1 层跳跃连接数设为 1, 即标准的 BiLSTM; 第 2 层的跳跃连接数设为 24, 即时间步更新状态依赖前 24 个时间步的隐藏状态; 第 3 层跳跃连接数设置为 48, 这表示当前时间步更新依赖前 48 个时间步的隐藏状态; 第 4 层跳跃连接数设置为 72, 这表示当前时间步更新依赖前 72 个时间步隐藏状态. 把模型的 4 层跳跃连接数依次设为 1、24、48、72, 就可提取数据的多尺度特征. 最后

添加 2 个全连接层, 拼接每层的输出作为全连接层的输入, 实现多尺度特征融合. 最后, 全连接层的输出作为模型的预测值.

BMS-LSTM 模型架构如图 3 所示, 每层设置不同的跳跃连接数^[19], 使每个时间步的状态信息不是依赖前一时间步的隐含状态, 而是依赖多个时间步之前的隐含状态. 时间步之间的更新方法为

$$h_t^i = f(h_{t-p}^i, x_t),$$

其中, 上标 i 表示所在网络层数, 下标 t 表示所在时间步数, p 表示所在层的跳跃连接数, $f()$ 表示 h_t^i 状态取决于 h_{t-p}^i 和 x_t 信息的依赖关系.

全连接层的传播的计算方法为

$$Y_1 = w_{d_1} H + b_1, Y_{out} = w_{d_2} Y_1 + b_2,$$

其中 w_{d_i} 为第 i 层全连接层的权重, b_i 为第 i 层的偏置. H 为第 1 层全连接层的输入, H 的计算公式为

$$H = [h_1^1, h_1^2, h_1^3, h_1^4],$$

其中 h_i^j 是第 i 层网络层的输出.

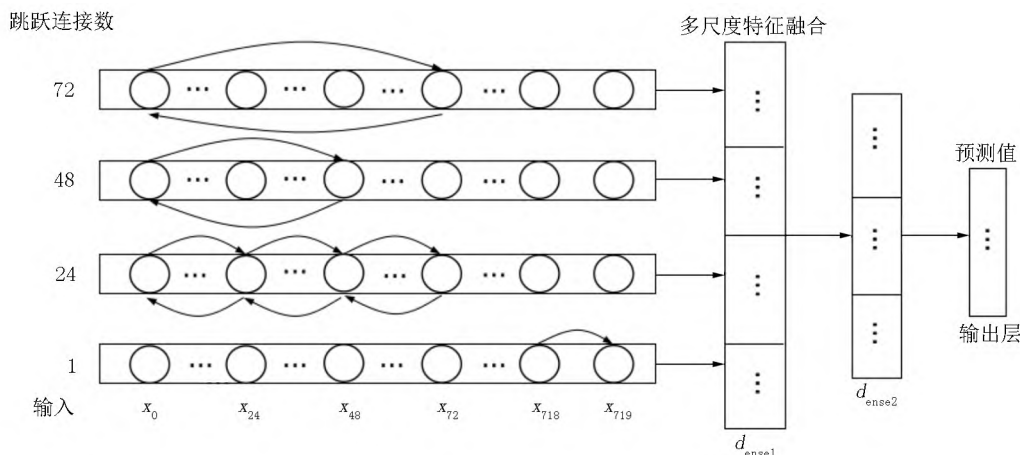


图3 BMS-LSTM 模型

3 实验分析与比较

3.1 数据预处理

实验使用的数据来源于欧洲某城市 2015 年 1 月 1 日—2018 年 12 月 31 日的天气数据, 包括温度、风速、压强、湿度、风向等. 将上述 5 个天气数据拼接为

一个 5 维向量, 通过融合其他天气数据, 提高短时温度预测精度.

模型使用 720 个时间点的数据预测未来 24 h 的温度, 样本数据维度为 720×5 . 天气数据的数值差异较大, 不利于模型的预测, 因此实验前需要对数据进行归一化处理. 将数据随机分为 14 组, 前 13 组样

本作为训练集,最后一组为测试集.每次实验均将数据重新随机分组,使每次实验的训练集与测试集都不相同.样本的原始数据为 $X = [x_1, x_2, x_3, x_4, x_5]^{720}$,数据归一化处理的方法为

$$X'_i = (x'_i - \min(X_i)) / (\max(X_i) - \min(X_i)),$$

其中 $t \in [0, 1, 2, 3, \dots, 718, 719]$,即对应的是网络层的时间步; i 表示列数,即某一天气数据; $\min()$ 、 $\max()$ 是最小值和最大值函数.

3.2 实验设置

本实验环境的主要参数:处理器是 Intel(R) Core(TM) i7-9750H CPU@2.60GHz;显卡为 NVIDIA GeForce RTX2060 8G;内存为 16 GB;操作系统是 Windows10 家庭中文版;使用 Tensorflow 构建模型并进行训练与测试.

实验使用多层 LSTM(MLSTM)模型、双向 MLSTM^[20]和多尺度跳跃深度 LSTM^[19](Multi-scale skip deep long short-term memory,MSD-LSTM)与 BMS-LSTM 模型进行对比.

MLSTM 模型层数设为 4,与 MBiLSTM 模型对比,检验双向深层网络结构相比于深层网络结构性能是否得到了提升.MBiLSTM 模型层数设为 4,检验 BMS-LSTM 模型是否有效提取数据的多尺度特征.MSD-LSTM 和 BMS-LSTM 模型层数均设为 4 层,将它们进行对比,检验 BMS-LSTM 是否可以提取数据中反向特征.所有模型的神经单元个数均设为 32,采用 RMSprop 算法对参数进行优化,学习率为 0.002,损失函数为平均绝对误差函数.

MLSTM、MBiLSTM、MSD-LSTM和BMS-LSTM

模型在训练时采用 mini-batch 技术(batch_size 为 100),可加快模型的学习速度,同时有利于参数更加鲁棒性的收敛.除此之外,为了防止模型过拟合,所有模型都使用 Dropout 技术,Dropout 率设为 0.5.

3.3 实验结果及分析

本文进行了 8 次实验,采用平均绝对误差(mean absolute error,MAE) e_M 作为评判标准. e_M 计算方法为

$$e_M = |y^i - y^i_{\text{pred}}| / M, \quad (1)$$

其中 y^i, y^i_{pred} 是测试集中相对应的真实值与预测值, M 为样本数.

对比实验结果如表 1 所示,表 1 中的数据是利用式(1)计算的平均绝对误差,数值越小说明模型预测精度越高.从表 1 可以看出,由于 BMS-LSTM 采用双向深层网络结构,提取了数据的反向特征,设置多个跳跃连接数提取了多尺度特征,所以对比其他 3 种模型,BMS-LSTM 均表现出较好的结果.BMS-LSTM 模型实验的平均预测误差约为 3.890,远小于 MLSTM 和 MSD-LSTM 模型的误差.虽然 MBiLSTM 模型的平均预测误差为 4.980,与 BMS-LSTM 的误差仅差 1.1 左右,但由表 1 数据显示:MBiLSTM 的预测精度忽高忽低,极不稳定.而 BMS-LSTM 模型 8 次实验预测精度较稳定,一直保持在 3.890 左右.综上所述,BMS-LSTM 预测精度最高,预测效果最稳定.

图 4 直观展示了 MLSTM、MBiLSTM、MSD-LSTM 和 BMS-LSTM 这 4 种模型的预测值与真实值之间的对比结果,从图 4 可以看出 BMS-LSTM 与真实值最为接近,预测精度最高,平均误差仅为 3.097.

表 1 各模型温度预测的平均绝对误差对比

实验次数	BMS-LSTM	MLSTM	MBiLSTM	MSD-LSTM
1	4.246	7.216	4.698	5.098
2	3.097	7.929	3.718	8.412
3	4.338	7.820	6.685	4.450
4	3.623	7.839	5.884	8.501
5	4.437	7.672	5.030	8.425
6	3.848	7.582	4.112	9.488
7	3.862	7.777	5.327	1.173
8	3.671	7.765	4.401	8.470
平均值	3.890	7.700	4.980	8.070

3.4 超参数探究

为研究层数、跳跃连接数以及神经单元数量对 BMS-LSTM 性能的影响,本文设置了 4 组使用不同超参数的对比实验,5 次实验结果如表 2 所示.对比实验 1 和实验 2,可以发现实验 2 误差较低,这说明

合理地增加神经单元的数量有利于提升短期温度预测的精度;对比实验 3 和实验 4 可以得到同样的结果.进一步,将实验 1、实验 2 和实验 3、实验 4 进行对比,在神经单元数量相同的情况下,实验 3 和实验 4 的误差更低,这表明跳跃连接数影响模型的性能,

根据数据实际的多尺度时间结构来选择合适的跳跃数尤为关键. 综上所述, 在实验 4 选取的超参数下短

期温度预测的精确度最高. 因此, 高效地使用 BMS-LSTM 需要综合考虑多个超参数对模型的影响.

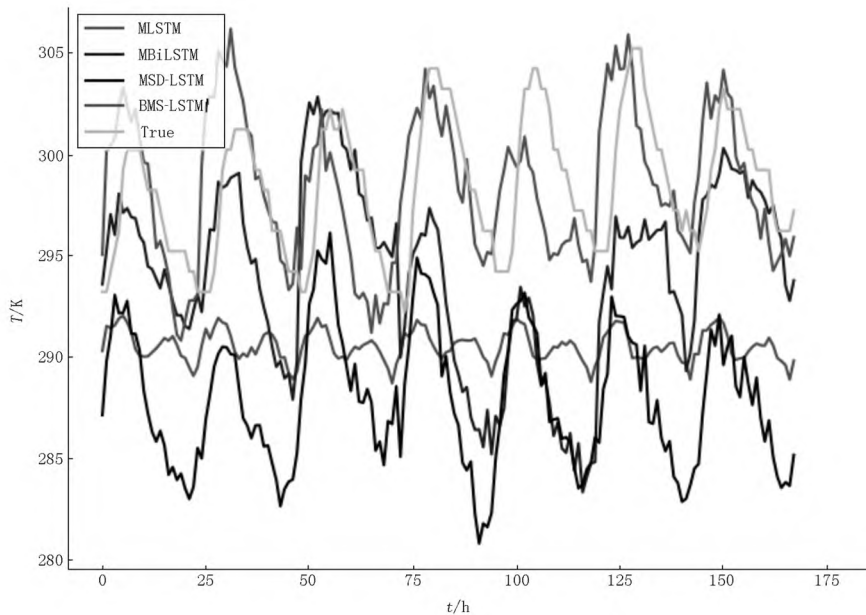


图 4 各模型预测结果对比

表 2 不同超参数的预测的平均绝对误差对比

实验	层数	跳跃数	神经单元数量	预测的平均绝对误差对比				
				1	2	3	4	5
1	3	1,24,48	16	294.021	293.776	293.719	294.015	293.070
2	3	1,24,48	32	292.119	289.710	290.934	292.195	292.114
3	4	1,24,48,72	16	291.301	293.540	291.307	291.293	293.725
4	4	1,24,48,72	32	289.646	289.547	290.261	290.245	291.842

4 结语

近年来,关于循环神经网络提取序列数据的多尺度特征的研究,成果显著,但是依然存在如下缺陷:(i)忽视了数据的反向特征;(ii)跳跃连接数不利于深层网络.这2个缺陷使普通多尺度循环神经网络不适用于短时温度预测.因此,本文提出了BMS-LSTM短期温度预测模型,它使用双向深层网络架构,提取反向特征,设置多个日周期跳跃连接数提取多尺度特征.实验结果表明:BMS-LSTM能有效提取温度数据的多尺度特征和反向特征,提升了预测精度.该模型也存在一定不足,在模型中的跳跃连接数依靠先验知识手动设置,无法自适应的学习时序数据的多尺度特征.如何自适应学习数据的多尺度特征是接下来的研究重点.

5 参考文献

[1] 陈广仁. 气候变化关乎人类未来 [J]. 科技导报, 2011,

29(34):8.

- [2] 王灏晨,路风,武继磊,等. 中国气候变化对人口健康影响研究评述 [J]. 科技导报, 2014, 32(28/29): 109-116.
- [3] 夏亮. 气温的统计分析及其短期预测 [D]. 武汉: 华中科技大学, 2010.
- [4] 杨成荫,赵苏璇,程立国,等. 短期区域气候预测系统及回报试验 [J]. 解放军理工大学学报(自然科学版), 2016, 17(3): 289-295.
- [5] 王定成,曹智丽,陈北京,等. 日气温多元时间序列局部支持向量回归预测 [J]. 系统仿真学报, 2016, 28(3): 654-660.
- [6] HUANG Chiyo, LIU Yuwei, TZENG Weichang, et al. Short term wind speed predictions by using the grey prediction model based forecast method [EB/OL]. [2021-03-16]. <https://ieeexplore.ieee.org/abstract/document/5754856>.
- [7] 刘丽丹. 基于概率图模型的天气预测研究 [J]. 计算机技术与发展, 2019, 29(7): 103-107.
- [8] 张曦,李璠,付雪峰,等. 随机学习萤火虫算法优化的模糊软子空间聚类算法 [J]. 江西师范大学学报(自然科学版), 2021, 45(2): 137-144.
- [9] 陈开阳,徐凡,王明文. 基于知识图谱和图像描述的虚

- 假新闻检测研究 [J]. 江西师范大学学报(自然科学版),2021,45(4):398-402.
- [10] 赵嘉,谢智峰,吕莉,等. 深度学习萤火虫算法 [J]. 电子学报,2018,46(11):2633-2641.
- [11] 陶晔,杜景林. 基于随机森林的长短期记忆网络气温预测 [J]. 计算机工程与设计,2019,40(3):737-743.
- [12] MIGUEZ-MACHO G, STENCHIKOV G L, ROBOCK A. Regional climate simulations over North America: interaction of local processes with improved large-scale flow [J]. Journal of Climate,2005,18(8):1227-1246.
- [13] 李芬,杨程,赵晋斌,等. 基于天气类型聚类 and LSTM 的 PM_{2.5} 短期预测模型 [J]. 水电能源科学,2021,39(3):199-202,151.
- [14] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization [EB/OL]. [2021-06-16]. <http://www.arxiv.org/pdf/1409.2329.pdf>.
- [15] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation,1997,9(8):1735-1780.
- [16] KOUTNIK J, GREFF K, GOMEZ F, et al. A clockwork RNN [EB/OL]. [2021-06-16]. <http://www.arxiv.org/pdf/1402.3511v1.pdf>.
- [17] CHUNG J Y, ANH S J, BENGIO Y S. Hierarchical multi-scale recurrent neural networks [EB/OL]. [2021-07-16]. <https://arxiv.org/pdf/1609.01704.pdf>.
- [18] CHANG Shiyu, ZHANG Yang, HAN Wei, et al. Dilated recurrent neural networks [EB/OL]. [2021-06-16]. <https://arxiv.org/abs/1710.02224>.
- [19] 肖勇,郑楷洪,郑镇境,等. 基于多尺度跳跃深度长短期记忆网络的短期多变量负荷预测 [J]. 计算机应用,2021,41(1):231-236.
- [20] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks [J]. IEEE Transactions on Signal Processing,1997,45(11):2673-2681.
- [21] GRAVES A, SCHMIDHUBER J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures [J]. Neural Networks,2005,18(5/6):602-610.
- [22] 曾安,聂文俊. 基于深度双向 LSTM 的股票推荐系统 [J]. 计算机科学,2019,46(10):84-89.
- [23] 杨龙,吴红斌,丁明,等. 新能源电网中考虑特征选择的 Bi-LSTM 网络短期负荷预测 [J]. 电力系统自动化,2021,45(3):166-173.
- [24] TABRIZI S E, XIAO Kai, GRIENSVEN J V, et al. Hourly road pavement surface temperature forecasting using deep learning models [J]. Journal of Hydrology,2021,603(1):126877.

The Short-Term Temperature Prediction Based on Bidirectional Multi-Scale LSTM

CHENG Pengyu¹, ZHAO Jia^{1*}, HAN Longzhe¹, ZHANG Yiying², WU Yannian³

(1. School of Information Engineering, Nanchang Institute of Technology, Nanchang Jiangxi 330099, China;

2. College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China;

3. Shenzhen Guodian Technology Communication Corporation Limited, Shenzhen Guangdong 518000, China)

Abstract: Because at the LSTM (Long Short-Term Memory) cannot effectively extract the multi-scale features and inverse features of temperature data, the BMS-LSTM (Bidirectional Multi-scale Skip Long Short-Term Memory) short-term temperature prediction model is proposed. The model uses LSTM as the core unit, and uses a two-way deep network structure to extract reverse features. The multi-scale feature is extracted according to the daily cycle setting of the number of jump connections in the temperature data, which solves the problem of the jump scale being too large in the later stage of the exponentially increasing number of jump connections. Finally, the fully connected layer is used for feature fusion prediction. The experiment's results show that BMS-LSTM has successfully extracted the multi-scale features and reverse features of temperature data, and the average prediction error is only 3.890, which is better than the comparison model and is an effective short-term temperature prediction model.

Key words: Long Short-Term Memory (LSTM); temperature prediction; multi-scale; reverse feature; skip connection; feature fusion

(责任编辑:冉小晓)