



# 基于改进双阶段注意力机制的降水智能预报

## 摘要

为提高现有时间序列算法降水预报的准确率,本文提出一种基于改进双阶段注意力机制的时间序列降水预报模型(DeepAMogLSTM)。该算法分为两部分,在输入注意力机制中,使用三层注意力机制对输入序列进行多重关注,选择稳定的输入特征;在时间注意力中,通过选择与目标值最相关的隐层状态,捕获时间序列的长期相关性。算法同时引入形变长短时记忆网络(Mogrifier LSTM),增强模型特征表示能力。模型使用2016—2019年预处理的自动站点特征数据和欧洲中期天气预报中心(ECMWF)气象场模式资料进行集成预报,并利用同期实况观测资料进行模式预报订正。实验结果表明:该模型在时效为2 h的降水预报中,各项数值评价指标均有改善,其中均方根误差为1.877 mm,平均绝对误差为0.727 mm,拟合优度( $R^2$ )为0.783;同时与其他模型预报订正效果相比,该模型较好地拟合了实际降水空间分布。

## 关键词

时间序列预测;降水量预测;形变LSTM;双阶段注意力机制

中图分类号 TP399

文献标志码 A

收稿日期 2021-03-27

资助项目 国家自然科学基金(61773220);国家重点研发计划(2016YFC0203301);江苏省自然科学基金(BK20150523);国家自然科学基金联合重点项目(U20B2061)

## 作者简介

戈苗苗,男,硕士生,主要从事数据挖掘、深度学习.1135920987@qq.com

陆振宇(通信作者),男,博士,教授,主要研究方向为数值预报、人工智能.nuist10113016@163.com

1 南京信息工程大学 电子与信息工程学院,南京,210044

2 南京信息工程大学 人工智能学院,南京,210044

## 0 引言

降水实时预报是指在很短的时间内对区域降水进行分析和预报,可为管理决策提供及时有效的环境信息,防止灾害的发生。现阶段天气预报主要依靠数值模式预报<sup>[1]</sup>,预报员首先通过模式产品进行数据分析,然后根据自己多年的预报经验进行模式订正,但是这需要花费大量时间和成本对预报员进行培养。由于降水影响因素比较特殊,依靠模式预报很难抓取降水与环境要素之间的联系<sup>[2]</sup>。李启芬等<sup>[3]</sup>通过选取预报因子,建立了夏季降水和气温预报动力统计混合模型,然而,模型订正效果在不同地区有明显的差异。降水量预测需要一种启发式的方法,这是典型的数值天气预报和统计学习方法所不能解决的,而机器学习、深度学习的迅速发展有效地弥补了数值预报的缺陷。任海青<sup>[4]</sup>利用灰度预测模型构建了预报模型,程敏等<sup>[5]</sup>基于时间序列模型ARIMA对区域降水进行建模,构建了当地降水预报模型,张帅等<sup>[6]</sup>构建了递归神经网络的降水预报模型,上述模型在降雨量预测任务上取得了不错的结果,但因降雨形成因素具有动态性和非线性特性,仅依靠线性外推技术,是无法根据历史数据捕捉其繁衍趋势的。临近预报的研究领域涉及大量数据,基于序列的深层神经网络缺少非线性关系,在降水预报订正中效果欠佳。双阶段注意力机制网络是由Qin等<sup>[7]</sup>提出的一种多变量时间序列神经网络,它能长时间捕获时间序列的长期依赖性,改善模型的预测效果。本文所使用的降雨数据由与降雨量关系密切的环境因子组成,和双阶段注意机制网络的输入变量有着相似的结构。

本文基于双阶段注意力机制网络进行了一定修改:首先在第一阶段输入注意力机制中引入目标序列,通过连续三层注意力机制提取稳定的权重,让网络稳定抓取降雨环境因子的空间关系;其次,在网络输入阶段映射到编码阶段中,通过加入形变LSTM<sup>[8]</sup>缓解循环神经网络长时间记忆流失问题,进一步提高模型预报性能。

## 1 研究区数据及处理

广东省以亚热带季风气候为主,全境地势北高南低,北部、东北部和西部都有较高山脉,中部和南部沿海地区多为低丘、台地或平原。广东省海拔分布如图1所示。南部城市平均海拔较低,其中江门、阳江等临海城市受海洋气团的影响,普遍吹偏南风,从而带来了丰沛的雨

水.6月是广东前汛期的降雨高峰期,而每年的6—10月又是热带气旋影响广东的主要时段,各地出现暴雨的次数较多.如图2所示,临海城市普遍年平均降雨量相对北部城市而言较高,海拔高度、温度分布、风向等气象环境因素对不同地区降雨有着显著影响.

由于自动站设备老化或极端雷击等因素造成观

测数据出现缺失、异常值的情况,在实验前,必须对脏数据进行去噪,过滤掉异常值.盒须图定义了识别异常值的标准:异常值一般为小于 $Q_L - 1.5R_{I_Q}$ 或大于 $Q_U + 1.5R_{I_Q}$ 的值. $Q_L$ 为下四分位数, $Q_U$ 为上四分位数, $R_{I_Q}$ 为四分位数间距,是上四分位数与下四分位数的差值.

由盒须图识别出的异常值,直接删除容易丢失

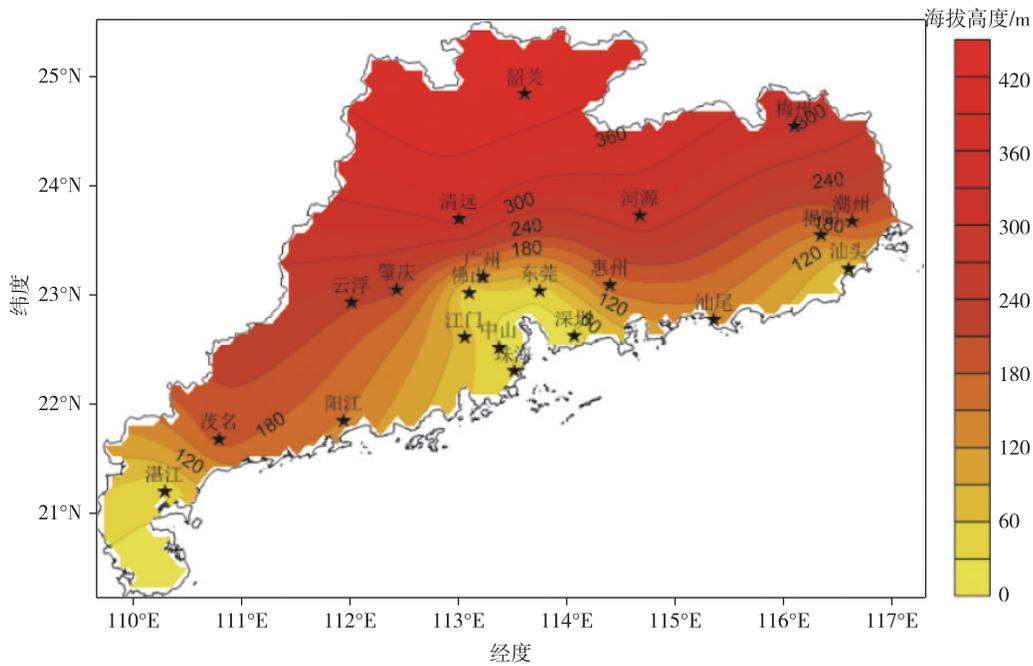


图1 广东省海拔分布

Fig. 1 Altitude distribution map of Guangdong province

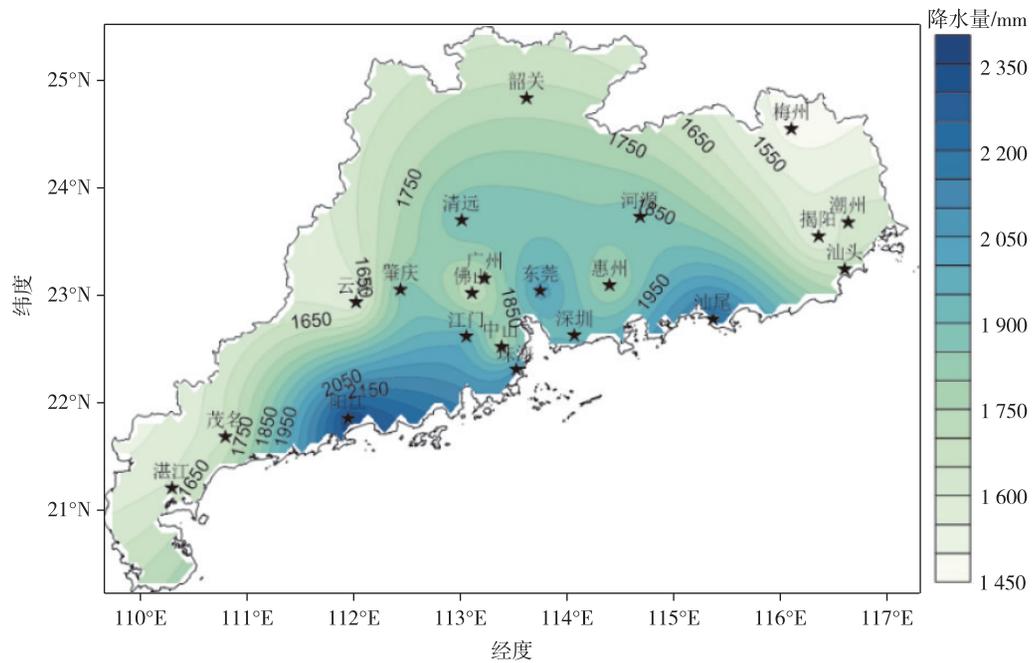


图2 广东省平均年降水量分布

Fig. 2 Annual average precipitation distribution of Guangdong province

众多隐藏在上下文数据中的信息,削弱数据之间的时间相关性.本文将异常值作为缺失值,利用缺失值的解决方案进行插值.常用的数据插补方法大致有:众数插补、最近邻插补、回归方法、插值法.由于降雨量数据前后关系紧密,使用众数、中位数等插补法插值的数据只能单一解决缺失值问题,并不能平滑数据,因此本文对缺失值前后的数据进行拉格朗日插值<sup>[9]</sup>,其公式定义如下:

$$y = a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1}, \quad (1)$$

$$L(x) = y_1 \frac{(x - x_2)(x - x_3) \dots (x - x_n)}{(x_1 - x_2)(x_1 - x_3) \dots (x_1 - x_n)} + y_2 \frac{(x - x_2)(x - x_3) \dots (x - x_n)}{(x_2 - x_1)(x_2 - x_3) \dots (x_2 - x_n)} + \dots + y_n \frac{(x - x_2)(x - x_3) \dots (x - x_n)}{(x_n - x_1)(x_n - x_3) \dots (x_n - x_{n-1})}, \quad (2)$$

式(1)为已知平面上  $n$  个点形成的  $n - 1$  次多项式,式(2)为将  $n$  个点的坐标代入多项式函数,让缺失值对应的点  $x$  插值多项式,得到缺失值的近似值  $L(x)$ .

对数据进行质量分析能有效剔除噪声数据,得到更加合理的观测资料,但这些数据并不能直接输入模型.气象台的观测特征较多,数据未经处理容易造成特征冗余的现象.因此,本文引入统计学方法<sup>[10]</sup>来研究变量之间的关系密切程度.

Lasso 变量选择具有稳定性,变量选取较为准确,其公式定义如下:

$$\hat{\beta} = \arg \min \frac{1}{2} \sum_{i=1}^n (y_i - \sum_j \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j|, \quad (3)$$

其中  $x_{ij}$  为矩阵中对应位置的值,  $y_i$  为目标变量对应值,  $\beta$  为回归系数向量,  $\lambda$  为非负正则参数.通过选择合适的  $\lambda$  对模型进行调整,可以减少与因变量关系小的系数,同时对变量进行显著性检验,证明变量之间的相关性不是偶然因素导致的.实验中算法寻得最优  $\lambda$  为 0.01.表 1 展示了部分 Lasso 变量选择和显著性检验结果.

表 1 Lasso 变量选择系数及显著性检验

Table 1 Lasso variable selection coefficient and significance

特征变量	Lasso 系数	$p$ 值
小时最高温	0.62	0.00
小时最高气压	0.39	0.00
小时最低气压	-0.28	0.00
平均风速	-0.18	0.00
天气现象	0.37	0.41
最高气温出现时间	0.00	0.00

由表 1 看出:最高气温出现时间相关系数为 0,同时  $p$  值小于 0.05,说明该因素对降水量是不相关的,可视为无关因素;小时最高温、小时最高压等影响因子与小时降水量之间有着很强的相关性且  $p$  值小于 0.05,说明这些因素是相关的;天气现象相关系数表现出弱相关性,但  $p$  值大于 0.05,说明两个因素的高相关性可能是偶然因素导致的.

皮尔逊相关系数可以分析目标与变量之间的线性相关程度,其计算公式如下:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (4)$$

其中:  $x_i, y_i$  为两个特征变量因素,  $\bar{x}, \bar{y}$  为特征变量均值.结合 Lasso 系数与皮尔逊相关系数选取相关性高的气象环境因子有助于搭建能耗更低、运算更快的模型结构.最终根据降雨形成因素与统计学方法计算结果<sup>[11-13]</sup>筛选与降雨量关系密切的气象环境因子,对数据进行预处理,总体流程如图 3 所示.

## 2 模型原理

### 2.1 形变长短时记忆神经网络

循环神经网络(RNN)被设计成可以利用历史的信息来辅助当下的判断,但是 RNN 主要依靠最后的输入信号来决策最后的结果,更早之前的输入信号

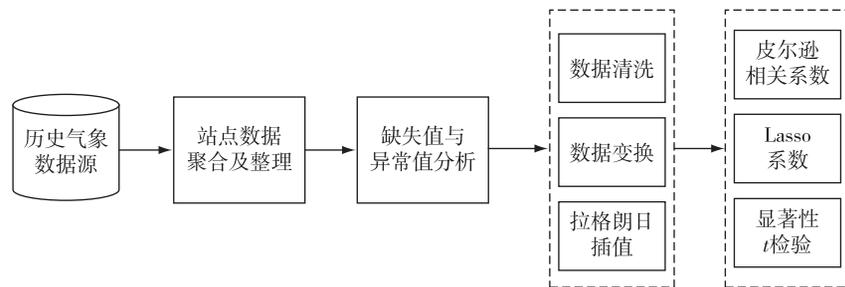


图 3 数据预处理流程

Fig. 3 Data preprocessing flowchart

会随着时间的推移被忽略,所以 RNN 在长期依赖问题上显得差强人意.长短时记忆神经网络(LSTM)的出现解决了这一难题,它主要在网络结构中新增了一个判别器,这个判别器的结构被称为“细胞”(cell).每个 cell 中包含三个门结构,依次为输入门、遗忘门和输出门.网络计算公式如下:

$$f_t = \sigma(\mathbf{W}_f \cdot [h_{t-1}, x_t] + \mathbf{b}_f), \quad (5)$$

$$i_t = \sigma(\mathbf{W}_i \cdot [h_{t-1}, x_t] + \mathbf{b}_i), \quad (6)$$

$$\tilde{C}_t = \tanh(\mathbf{W}_C \cdot [h_{t-1}, x_t] + \mathbf{b}_C), \quad (7)$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t, \quad (8)$$

$$O_t = \sigma(\mathbf{W}_O \cdot [h_{t-1}, x_t] + \mathbf{b}_O), \quad (9)$$

$$h_t = O_t \tanh(C_t), \quad (10)$$

式中:  $f_t, i_t, C_t, O_t$  分别为遗忘、输入、记忆细胞和输出门,  $\mathbf{W}_f, \mathbf{W}_i, \mathbf{W}_C, \mathbf{W}_O$  为相应权重矩阵,  $\mathbf{b}_f, \mathbf{b}_i, \mathbf{b}_C, \mathbf{b}_O$  为相应偏置向量.

LSTM 通过加入输入、遗忘、输出门的方式,可以缓解梯度消失和信息遗忘问题.但是,在 LSTM 计算公式中,当前时刻输入  $x_t$  和隐藏状态  $h_{t-1}$  是没有关联的,而形变 LSTM 则在现有 LSTM 计算之前,通过门机制让输入  $x_t$  和隐藏状态  $h_{t-1}$  进行充分交互,得到新的输入和新的隐藏状态,这样两者能互相关联,达到增强上下信息联系的目的<sup>[8]</sup>.具体公式如下:

$$x^i = 2\sigma(Q^i h_{t-1}^{i-1}) \odot x^{i-2}, \quad \text{for odd } i \in [1, \dots, r], \quad (11)$$

$$h_{t-1}^i = 2\sigma(R^i x^{i-1}) \odot h_{t-1}^{i-2}, \quad \text{for even } i \in [1, \dots, r], \quad (12)$$

式中:  $i=1$  时,  $x^{-1}=x, h_{t-1}^0=h_{t-1}$ ;  $\mathbf{Q}$  和  $\mathbf{R}$  为额外设置的两个随机矩阵;交互次数  $r$  作为超参数也影响着模型的效果,选取合适的  $r$  能让模型达到最优,如果  $r=0$ ,则形变 LSTM 变为普通的循环神经网络;常数 2 是为了避免激活函数(0,1)分布而造成数值变小的问题,乘以常数 2 能稳定数值.具体结构如图 4 所示.

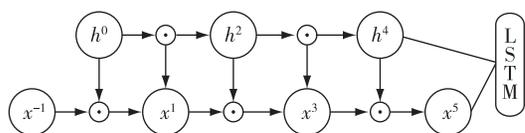


图 4 Mogrifier LSTM 交互示意  
Fig. 4 Schematic of Mogrifier LSTM interaction

对于给定的序列信息  $\{x_1, x_2, \dots, x_{t-1}, x_t\}, x_i \in \mathbf{R}^n$ , 定义激活函数(形变 LSTM)为  $F$ , 则能计算每个时间步  $h_t = F(h_{t-1}, x_t)$ .

## 2.2 输入注意力机制

目前,注意力机制广泛应用于深度学习,在自然语言处理中,传统 RNN 注意力机制<sup>[14]</sup>通常在隐藏层信息中选择对结果有重要影响的时间步,而忽略了输入序列之间的影响.因此本文先对输入序列进行自适应选取,与双阶段注意力机制不同的是,本文在其第一阶段注意力中引入目标序列  $y$  和加深注意力模块层数得到一个稳定的输出权重,用来提取输入序列和目标序列的空间相关性,称该模块为深度输入注意力机制,结构如图 5 所示.

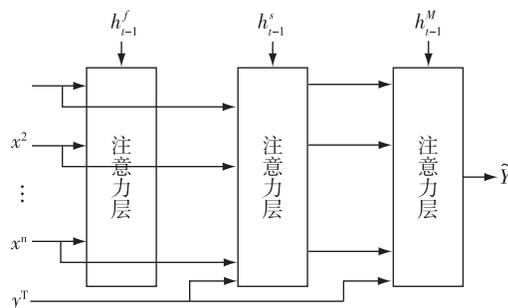


图 5 深度输入注意力机制示意  
Fig. 5 Schematic of deep input attention mechanism

对于给定的第  $k$  个输入序列信息  $\{x_1^k, x_2^k, \dots, x_t^k\}$ , 第一层注意力采用如下方式:

$$f_t^k = V_f^T \tanh(W_f [h_{t-1}^f; s_{t-1}^f] + U_f x^k + b_f), \quad (13)$$

$$\alpha_t^k = \text{soft max}(f_t^k), \quad (14)$$

其中  $[h_{t-1}^f; s_{t-1}^f] \in \mathbf{R}^{2m}$  为形变 LSTM 的先前隐藏状态和细胞状态的拼接向量.  $V_f, b_f \in \mathbf{R}^T, W_f \in \mathbf{R}^{T \times 2m}, U_f \in \mathbf{R}^{T \times T}$  则为学习参数.  $m$  为注意力模块中的隐藏层大小. 根据得到的注意力权重, 自适应提取输入序列:

$$\tilde{x}_t = (\alpha_t^1 x_1^1, \alpha_t^2 x_2^2, \dots, \alpha_t^n x_t^n)^T. \quad (15)$$

深度第二层注意力的输入是将输入序列  $\{x_1^k, x_2^k, \dots, x_t^k\}$  和目标变量  $y^k$  结合得到  $Z^k = [x^k; y^k] \in \mathbf{R}^{(n+1) \times T}$ , 通过第二次选择输入序列与目标序列的空间关系, 得到联系更加紧密的空间相关性. 具体方式如下:

$$S_t^k = V_s^T \tanh(W_s [h_{t-1}^s; s_{t-1}^s] + U_s [x^k; y^k] + b_s), \quad (16)$$

$$\beta_t^k = \text{soft max}(S_t^k). \quad (17)$$

类似地  $V_s, b_s \in \mathbf{R}^T, W_s \in \mathbf{R}^{T \times 2q}, U_s \in \mathbf{R}^{T \times T}$  也分别为学习参数,  $[h_{t-1}^s; s_{t-1}^s] \in \mathbf{R}^{2q}$  分别为形变 LSTM 隐藏状态和细胞状态. 根据得到的第二阶段注意力权重, 得到输出如下:

$$\tilde{\theta}_t = (\beta_t^1 \tilde{x}_t^1, \beta_t^2 \tilde{x}_t^2, \dots, \beta_t^n \tilde{x}_t^n)^T. \quad (18)$$

深度第三层注意力机制的输入是将第二层注意力模块的输出  $\tilde{\theta}^k$  和目标变量  $y^k$  结合起来得到  $Y^k = [\tilde{\theta}^k; y^k] \in \mathbf{R}^{(n+1) \times T}$ , 重复类似第二层注意力模块操作, 得到如下注意力结果:

$$M_t^k = V_M^T \tanh(W_M[h_{t-1}^M; s_{t-1}^M] + U_M[\tilde{\theta}^k; y^k] + b_M), \quad (19)$$

$$\gamma_t^k = \text{soft max}(M_t^k), \quad (20)$$

$$\tilde{Y}_t = (\gamma_t^1 \tilde{\theta}_t^1, \gamma_t^2 \tilde{\theta}_t^2, \dots, \gamma_t^n \tilde{\theta}_t^n)^T. \quad (21)$$

### 2.3 时间注意力机制

多变量序列提取到的特征是影响预测结果的重要因素, 依靠输入序列和目标序列之间的空间相关性还不足以预测准确的结果. 由于输入序列长度的增加, 网络的运行效果会被严重影响, 所以还需要加入时间注意力机制选择与目标序列最相关的隐藏状态, 具体结构如图 6 所示. 相同地, 这里再次使用形变 LSTM 解码上一层注意力机制的编码信号. 根据激活函数  $F$ , 得到新隐藏状态  $h_t = F(h_{t-1}, \tilde{Y}_t)$ ,  $h_t \in \mathbf{R}^p$ , 记忆细胞状态记为  $C_t \in \mathbf{R}^p$ , 则时间注意力层关注度计算如下:

$$l_t^i = V_d^T \tanh(W_d[h_{t-1}; C_{t-1}] + U_d h_t), \quad (22)$$

$$\beta_t^i = \text{soft max}(l_t^i), \quad (23)$$

式中:  $[h_{t-1}; C_{t-1}] \in \mathbf{R}^{2p}$  为上一个时间步隐藏状态和细胞状态的拼接,  $V_d \in \mathbf{R}^m$ ,  $W_d \in \mathbf{R}^{m \times 2p}$ ,  $U_d \in \mathbf{R}^{m \times m}$  分别为学习参数,  $\beta_t^i$  代表了第  $i$  个隐状态对于预测结果的重要性.

由求得的  $\beta_t^i$  和新隐藏层状态  $h_t$  便能自适应得到不同时间步的语义向量  $c_t = \sum_{i=1}^T \beta_t^i h_i$ . 通过组合目标

序列  $\{y_1, y_2, \dots, y_{t-1}\}$  和向量  $c_t$  求得  $t-1$  时刻的输出:

$$\tilde{y}_{t-1} = \tilde{w}^T [y_{t-1}; c_{t-1}] + \tilde{b}, \quad (24)$$

式中:  $[y_{t-1}; c_{t-1}] \in \mathbf{R}^{m+1}$  为  $t-1$  时刻的真实值与向量的拼接,  $\tilde{w} \in \mathbf{R}^{m+1}$  为学习参数.

根据  $t-1$  时刻的隐藏状态和求得目标  $\tilde{y}_{t-1}$ , 得到  $t$  时刻的隐层状态  $d_t \in \mathbf{R}^p$ , 最后进行回归预测, 如下所示:

$$d_t = F(d_{t-1}, \tilde{y}_{t-1}), \quad (25)$$

$$\hat{y}_T, \hat{y}_{T+1}, \dots, \hat{y}_{T+\lambda} = V_y^T (W_y [d_t; c_t] + b_y). \quad (26)$$

$W_y \in \mathbf{R}^{p(p+q)}$ ,  $b_y \in \mathbf{R}^p$ ,  $[d_t; c_t] \in \mathbf{R}^{p+q}$ .

### 2.4 基于双阶段注意力机制的降水量智能预报

首先获取预处理的气象环境因子逐时实况资料, 其次将 ECMWF 模式资料处理成对应站点数据, 并对数据进行集成和归一化处理, 再次将集成的预报因子输入到本文构建的改进双阶段注意力机制模型 (DeepAMogLSTM) 中进行训练、调优, 优化模型效果, 选取最优超参进行降水预报, 最后进行数值检验和时空分布检验.

## 3 实验结果及分析

### 3.1 实验评价指标

本文选择回归任务中的均方根误差 (RMSE, 量值记为  $\eta_{\text{RMSE}}$ )、平均绝对误差 (MAE, 量值记为  $\eta_{\text{MAE}}$ ) 和决定系数 ( $R^2$ ) 评价指标对模型进行数值检验. RMSE 是均方误差的平方根, 通常代表预测值的离散程度, 即标准误差,  $\eta_{\text{RMSE}}$  越小, 结果拟合越好; MAE 是平均绝对误差, 由于预测误差有正有负, MAE 能防止预测误差正负抵消, 可以更好地反映预

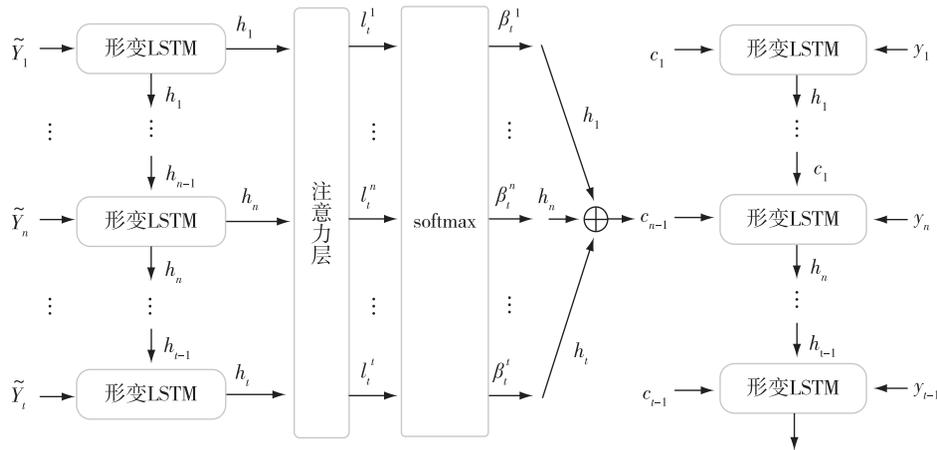


图 6 时间注意力机制

Fig. 6 Time attention mechanism

测值误差的实际情况; $R^2$  反映观测值与预测值的拟合优度,其数值越靠近 1,拟合程度越好.它们定义分别如下:

$$\eta_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}, \quad (27)$$

$$\eta_{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - y_i|, \quad (28)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (29)$$

式中, $n$  代表样本总数, $x_i$  代表观测值, $y_i$  代表预测值, $\bar{x}$  为观测值均值.

### 3.2 实验结果分析

#### 3.2.1 降水预报数值检验

为验证模型在多因素降水预测上的有效性,本文对广东省沿海城市降水量数据进行网络评估,该数据经过上文相关预处理步骤,提升了数据质量.并以湛江市 3 h 累计降水量测试数据为例,图 7、图 8 给出了本文模型降水量预报数值与其余方法 (AR、SVR<sup>[15]</sup>、LSTM<sup>[16]</sup>、LSTnet<sup>[17]</sup>、DARNN<sup>[7]</sup>) 降水量预报数值的对比结果,预报时效分别为 1 h 和 2 h,实验中对累计降水量不大于 0.1 mm 的数据进行了清零处理.

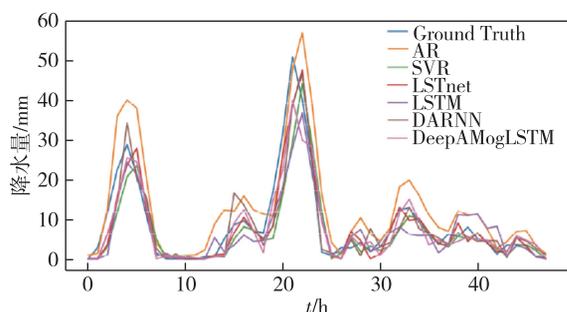


图 7 各模型下 1 h 降水量预报值

Fig. 7 The 1 h precipitation forecasted by different models

不难发现,AR 线性模型作为时间序列算法之一,在时间序列问题的处理上有一定优势,但缺少非线性因素的影响,容易出现过大的峰值.SVR 支持向量回归模型采用线性核函数做回归预测,但是在传统的多变量方法中,冗余的信息容易导致模型过拟合问题.作为近些年比较热门的深度学习网络 LSTM 和 LSTnet 能有效利用其长时间的记忆性能,取得更好的结果,但这两种方法更多的是关注时间相关性,没有考虑空间相关性.DARNN 模型与单一的 LSTM

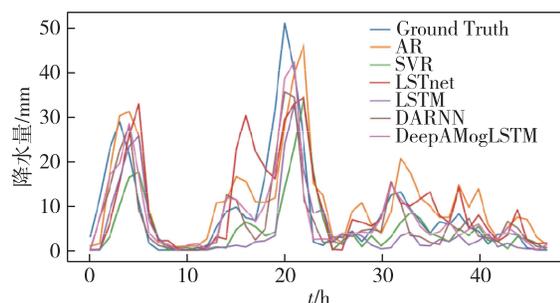


图 8 各模型下 2 h 降水量预报值

Fig. 8 The 2 h precipitation forecasted by different models

和 LSTnet 相比拟合更好,但本文提出的算法 (DeepAMogLSTM) 在曲线趋势上更加拟合,它能有效缓解单注意力机制权重变化分散的问题,通过第一阶段连续三层注意力机制提取平稳的权重,以更好地表示输入与目标序列的空间相关性.

为直观检验本模型在降水量预报上的性能,表 2 给出了各算法在测试集上的误差对比结果.

表 2 六种预测方法的评价指标

Table 2 Precipitation forecast comparison between six methods

预报时长	方法	RMSE/mm	MAE/mm	$R^2$
1 h	AR	2.091	0.832	0.694
	SVR <sup>[15]</sup>	2.069	0.441	0.735
	LSTM <sup>[16]</sup>	1.947	0.543	0.745
	LSTnet <sup>[17]</sup>	2.054	0.490	0.759
	DARNN <sup>[7]</sup>	1.917	0.623	0.783
	DeepAMogLSTM	1.796	0.412	0.801
2 h	AR	3.147	0.987	0.370
	SVR <sup>[15]</sup>	3.351	0.766	0.310
	LSTM <sup>[16]</sup>	3.303	0.856	0.329
	LSTnet <sup>[17]</sup>	3.067	0.644	0.422
	DARNN <sup>[7]</sup>	2.086	0.873	0.693
	DeepAMogLSTM	1.877	0.727	0.783

表 2 结果表明,本文提出的模型在各评价指标上都有较好的成绩.当预报时长为 1 h,模型的 RMSE 为 1.796 mm,MAE 为 0.412 mm, $R^2$  为 0.801;当预报时长为 2 h,模型的 RMSE 为 1.877 mm,MAE 为 0.727 mm, $R^2$  为 0.783,其误差指数低于其他模型.

为验证改进模块在降水数据集上的影响结果,对 DARNN 的两个改动模块进行消融实验,主要结果如表 3 (模块 1 是改进第一阶段输入注意力机制,简称 DeepA,模块 2 是加入形变 LSTM,简称 MogLSTM) 所示.

表3 各模块实验评价指标对比  
Table 3 Comparison of experimental evaluation indexes for each module

预报时长	DeepA	MogLSTM	RMSE/mm	MAE/mm	R <sup>2</sup>
1 h			1.917	0.623	0.783
	✓		1.828	0.412	0.793
		✓	1.875	0.509	0.791
	✓	✓	1.796	0.412	0.801
2 h			2.086	0.873	0.693
	✓		1.946	0.719	0.767
		✓	2.042	0.802	0.743
	✓	✓	1.877	0.727	0.783

相较于 DARNN,加入其余模块后预测效果都有小幅提升,其中加入输入注意力机制模块的改动对实验结果提升较大,加入形变 LSTM 后,模型能取得更优的结果。

### 3.2.2 降水预报空间分布分析

每年的夏季都是全国降雨高峰期,出现暴雨天气次数较多.本节以广东省 2019 年 7 月 31 日 18 时降水为例,对累计降水量进行空间预报分析.图 9a—f 和 g—l 分别是各模型预测广东省未来 1 h 和 2 h 累计降水量的空间分布,图 10 表示的是对应时间广东省观测资料的降水量空间分布.由于克里金插值法考虑了空间属性的变异分布,它可以有效剔除由于空间采样不均匀造成的误差,使结果更加贴合真实情况,所以作图过程中使用克里金插值法对广东省的降水预报空间分布进行插值。

由图 10 观测资料降水空间分布来看,广东省各城市降雨状态有明显分层,沿海城市(汕头、湛江等临海城市)累计降雨量多于西北山地地区,西北地区可能因为山地海拔高等原因,降水趋势往东北方向扩散.由图 9 可知,各模型整体预测累计降雨量空间分布状态是趋向于真实降雨空间分布的,其中 DeepAMogLSTM 和 DARNN 的降雨空间预报结果相对其余的降雨空间预报而言更符合实际情况,部分线性模型出现较高峰值,预报订正效果较差.由图 9e—f 和 k—l 可知,本文模型的降雨空间预报结果相对而言更拟合实际,从而间接反映加强输入注意力机制的空间特征提取能力对模型降水预报的订正十分有效。

## 4 结论

本文以降雨量预测为例,根据 Lasso 系数、显著性  $t$  检验和皮尔逊相关系数,筛选气象环境因子,并

对部分缺失的特征数据进行插值填补.针对降水因素非线性关系复杂,预测较难的问题,提出一种改进双阶段注意力机制网络模型,并对该模型的性能进行数值检验和时空分布检验.通过实验分析得到如下结论:

1) 本文对双阶段注意力机制模型进行了改进,相比 DARNN 与其余模型而言,DeepAMogLSTM 模型在降水预报上的 RMSE 与 MAE 都取得了更好的结果.其中,模型预报时长 1 h 的均方根误差为 1.796 mm,平均绝对误差为 0.412 mm,拟合优度为 0.801,模型预报时长 2 h 的均方根误差为 1.877 mm,平均绝对误差位 0.727 mm,拟合优度为 0.783.

2) 以广东省 2019 年夏季各市降水空间预报为例,本文模型降水空间预报分布更拟合实际降水空间分布,局部分布相对其余模型而言更具优势。

## 参考文献

### References

- [1] Buisán S T, Smith C D, Ross A. The potential for uncertainty in numerical weather prediction model verification when using solid precipitation observations [J]. *Atmospheric Science Letters*, 2020, 21 (7). DOI: 10.1002/asl.1976
- [2] 智协飞,赵忱.基于集合成员订正的强降水多模式集成预报[J].*应用气象学报*,2020,31(3):303-314  
ZHI Xiefei, ZHAO Chen. Heavy precipitation forecasts based on multi-model ensemble members [J]. *Journal of Applied Meteorological Science*, 2020, 31 (3): 303-314
- [3] 李启芬,刘婷婷,陈海山,等.基于土壤湿度和年际增量方法的中国夏季气温预测试验[J].*气象科学*, 2016, 36(5):629-638  
LI Qifen, LIU Tingting, CHEN Haishan, et al. Prediction of summer temperature in China based on soil moisture and interannual increment approach [J]. *Journal of the Meteorological Sciences*, 2016, 36 (5): 629-638
- [4] 任海青.基于灰色系统的年降雨量预测模型探析[J].*水资源开发与管理*,2017(3):61-63  
REN Haiqing. Discussion of annual rainfall forecast model based on grey system [J]. *Water Resources Development and Management*, 2017 (3): 61-63
- [5] 程敏,张耀文,姜纪沂,等.基于时间序列模型的降雨量预测分析[J].*水科学与工程学报*,2019(1):1-5  
CHENG Min, ZHANG Yaowen, JIANG Jiyi, et al. Rainfall prediction based on time series model [J]. *Water Sciences and Engineering Technology*, 2019 (1): 1-5
- [6] 张帅,魏正英,张育斌.递归神经网络在降雨量预测中的应用研究[J].*节水灌溉*,2017(5):63-66,71  
ZHANG Shuai, WEI Zhengying, ZHANG Yubin. Forecasting rainfall with recurrent neural network [J]. *Water Saving Irrigation*, 2017 (5): 63-66, 71
- [7] Qin Y, Song D J, Chen H F, et al. A dual-stage attention-based recurrent neural network for time series prediction

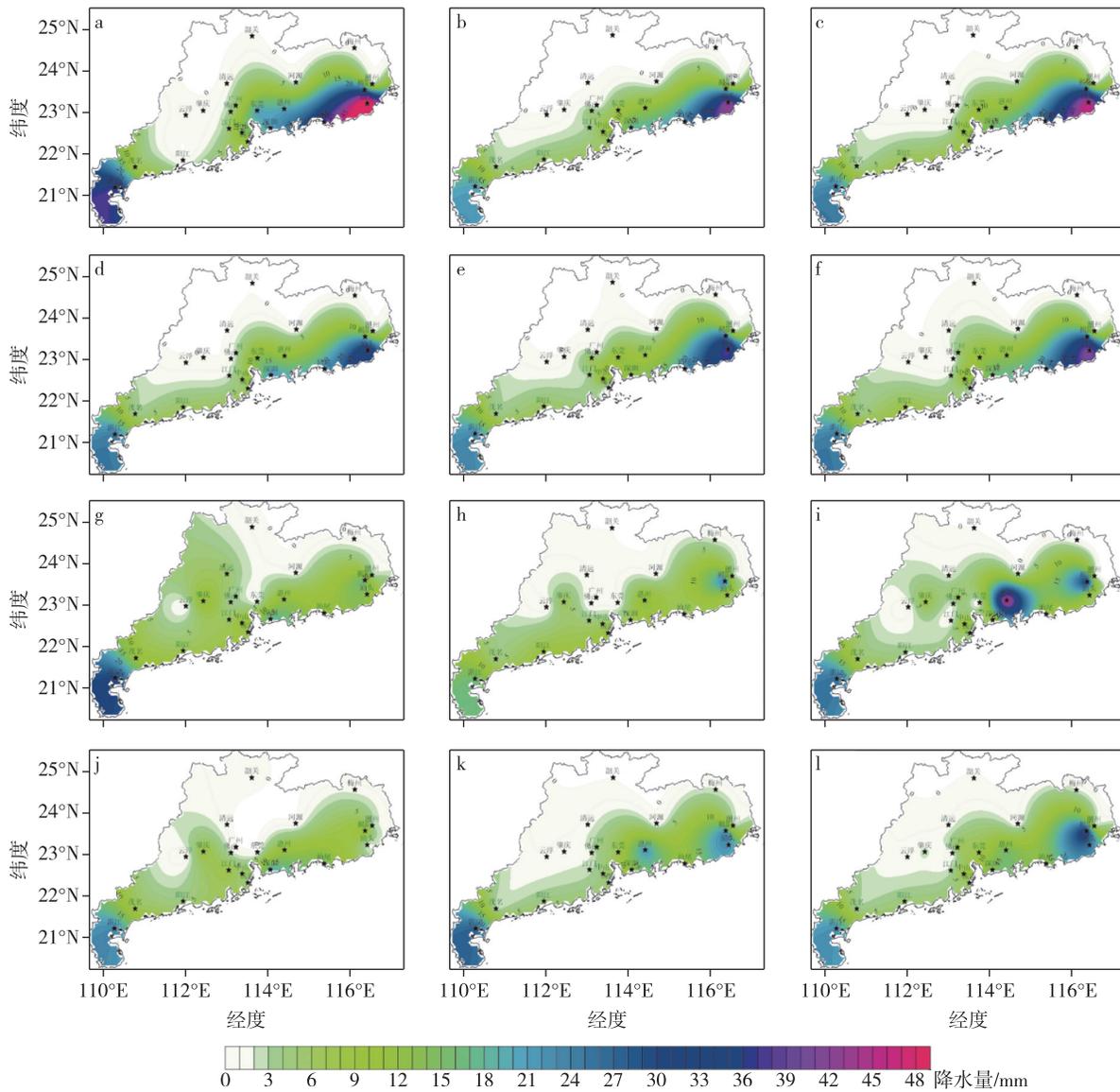


图9 2019-07-31 18:00 各模型降水量预报空间分布

a—f.预测 1 h 累计降水量;g—l.预测 2 h 累计降水量;a,g.AR 模型预报;b,h.SVR 模型预报;c,i.LSTM 预报;  
d,j.LSTnet 模型预报;e,k.DARNN 模型预报;f,l.DeepAMogLSTM 模型预报

Fig. 9 Spatial distribution of precipitation forecast at 18:00 on July 31, 2019,

a—f.the 1 h cumulative precipitation;g—l.the 2 h cumulative precipitation;a,g.AR model;b,h.SVR model;  
c,i.LSTM forecast;d,j.LSTnet forecast;e,k.DARNN model;f,l.the proposed DeepAMogLSTM model

[ C ]//Proceedings of the 26th International Joint Conference on Artificial Intelligence,2017:2627-2633

[ 8 ] Melis G, Ko ě iský T, Blunsom P. Mogrifier LSTM [ J ]. arXiv e-print, 2019, arXiv:1909.01792

[ 9 ] Le K T,Chaux C, Richard F J P, et al. An adapted linear discriminant analysis with variable selection for the classification in high-dimension, and an application to medical data [ J ]. Computational Statistics and Data Analysis, 2020, 152 ( 1 ) : 107031

[ 10 ] Cai L Q, Maiti T. Variable selection and estimation for high-dimensional spatial autoregressive models [ J ]. Scandinavian Journal of Statistics, 2020, 47 ( 2 ) : 587-607

[ 11 ] Dai Y L, Lu Z Y, Zhang H D, et al. A correction method of environmental meteorological model based on long-short-term memory neural network [ J ]. Earth and Space Science, 2019, 6 ( 4 ) : 2214-2226

[ 12 ] 尹珩,张珂,孙辰,等.武汉城区臭氧时空分布及其与气象因子相关性研究 [ J ].环境与可持续发展, 2017, 42 ( 6 ) : 150-152

YIN Heng, ZHANG Ke, SUN Chen, et al. Study on the spatial-temporal distribution of ozone and its correlation with meteorological factors in urban district of Wuhan [ J ]. Environment and Sustainable Development, 2017, 42 ( 6 ) : 150-152

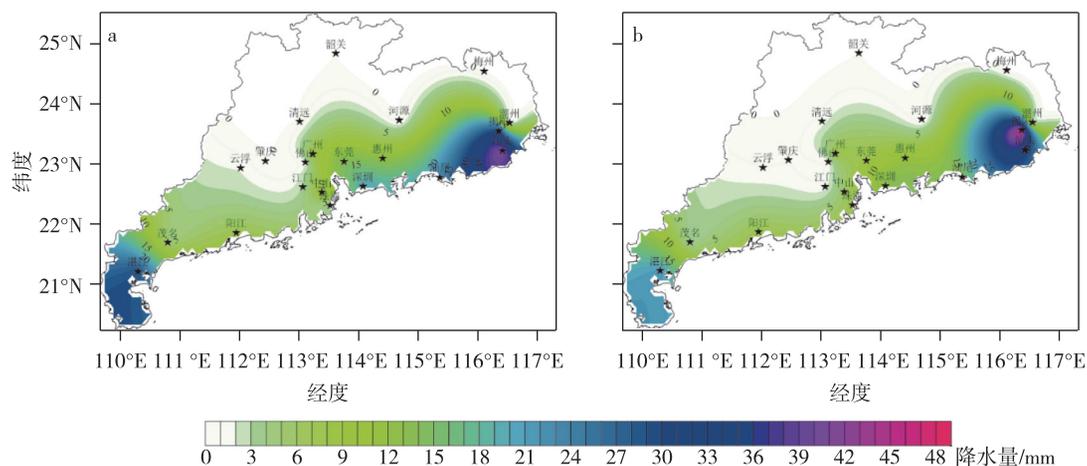


图 10 2019-07-31 18:00(a), 2019-07-31 19:00(b) 观测资料日降水量空间分布

Fig. 10 Spatial distribution of daily precipitation observed at 18:00 (a) and 19:00 (b) on July 31, 2019

- [13] Tian B Q, Fan K. Climate prediction of summer extreme precipitation frequency in the Yangtze River valley based on sea surface temperature in the southern Indian Ocean and ice concentration in the Beaufort Sea [J]. International Journal of Climatology, 2020, 40 (9): 4117-4130
- [14] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [J]. Computer Science, 2014: 1-15
- [15] Vapnik V, Golowich S E, Smola A. Support vector method for function approximation, regression estimation, and signal processing [J]. Advances in Neural Information Processing Systems, 2008, 9: 281-287
- [16] Greff K, Srivastava R K, Koutnik J, et al. LSTM: a search space odyssey [J]. IEEE Transactions on Neural Networks & Learning Systems, 2016, 28 (10): 2222-2232
- [17] Lai G K, Chang W C, Yang Y M, et al. Modeling long- and short-term temporal patterns with deep neural networks [C] // The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018: 95-104

## Intelligent precipitation forecast based on improved dual-stage attention mechanism

GE Miaomiao<sup>1</sup> LU Zhenyu<sup>2</sup> LIANG Shaoyang<sup>2</sup> XIA Yingru<sup>2</sup>

1 School of Electronic & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044

2 School of Artificial Intelligence, Nanjing University of Information Science & Technology, Nanjing 210044

**Abstract** In order to improve the existing time series algorithms for precipitation forecasting, this paper proposes a time series precipitation forecasting model (DeepAMogLSTM) based on improved dual-stage attention mechanism. The algorithm can be divided into two parts. In the input attention stage, a three-layer attention mechanism is designed to pay multiple attention to the input sequence; while in the time attention stage, the hidden state most relevant to the target value is selected to calculate the long-term correlation of the time sequence. In this manner, input features can be stably selected and input into the prediction structure. The algorithm also introduces Mogrifier LSTM (Long Short-Term Memory) to enhance the feature representation ability. The model uses preprocessed automatic station data from 2016 to 2019 and ECMWF weather field model data for integrated forecast, and corrects the model forecasts using observation data of the same period. The experimental results show that the evaluation indexes of the model are improved in the 2-hour precipitation nowcasting, in which the maximum square root error is 1.877 mm, the maximum average absolute error is 0.727 mm, and the goodness of fit ( $R^2$ ) is 0.783. At the same time, the modeled precipitation fits actual precipitation in spatial distribution, which is better than the correction effect of other models.

**Key words** time series prediction; precipitation forecast; Mogrifier LSTM; dual-stage attention mechanism