

Doi: 10.11840/j.issn.1001-6392.2021.04.004

基于双注意力机制的台风轨迹预测模型

贺琪¹, 刘东旭¹, 宋巍¹, 黄冬梅², 杜艳玲¹

(1. 上海海洋大学 信息学院, 上海 201306; 2. 上海电力大学, 上海 200090)

摘要: 台风轨迹的准确预测对于减少台风灾害及风险评估意义重大。本文提出了一种基于双注意力机制的台风轨迹预测模型 (Dual-Attention-Encoder-Decoder), 首先根据台风轨迹数据计算台风轨迹的变化曲率, 将台风曲率序列与台风轨迹序列一同作为预测模型的特征输入, 充分考虑了台风轨迹中隐藏的转向、偏折信息; 然后构建双注意力机制增强的编码器-解码器网络 (Encoder-Decoder) 作为预测模型, 利用特征注意力机制和时间注意力机制分别对模型输入和隐藏状态进行权重分配, 能够有效解决编码器-解码器结构对过长序列预测的性能下降问题, 编码器和解码器均采用 LSTM 网络, 能够存储长时间依赖并且收敛性好, 不易发生梯度消失或爆炸; 最后, 本文使用 1949—2017 年中国气象局提供的西北太平洋台风最佳路径数据集, 将 DA-Encoder-Decoder 模型与 BP、SVR、LSTM、ELM 等模型进行对比, 分别对 24 h、48 h、72 h 台风轨迹进行预测。结果表明: DA-Encoder-Decoder 模型的均方根误差和实际误差距离指标均优于其他四种预测方法, 验证了本文方法的有效性。

关键词: 台风轨迹预测; 注意力机制; 曲率; 时间序列

中图分类号: P457.8; TP183

文献标识码: A

文章编号: 1001-6392(2021)04-0387-09

Typhoon trajectory prediction model based on dual attention mechanism

HE Qi¹, LIU Dongxu¹, SONG Wei¹, HUANG Dongmei², DU Yanling¹

(1. College of Information, Shanghai Ocean University, Shanghai 201306, China;

2. Shanghai University of Electric Power, Shanghai 200090, China)

Abstract: Accurate prediction of typhoon trajectory is of great significance for typhoon disaster deduction and risk assessment. This paper proposes a typhoon trajectory prediction model based on the dual attention mechanism, Dual-Attention-Encoder-Decoder. First, the typhoon trajectory change curvature is calculated based on the typhoon trajectory data. The typhoon curvature sequence together with the typhoon trajectory sequence are serving as the feature input to prediction model, taking full account of the hidden steering and deflection information in the typhoon trajectory. Then an Encoder-Decoder network enhanced by the dual attention mechanism is built as a predictive model. Using the feature attention mechanism and time attention mechanism for weight distribution of model input and hidden state respectively, we can learn the relationship between input features and prediction targets, and effectively solve the problem of performance degradation of the encoder-decoder structure for the excessively long sequences prediction. The encoder and decoder use LSTM respectively. The network can store long-term dependence and has good convergence, and it is not easy to cause gradient disappearance or explosion. Finally, the best path data set of the Northwest Pacific typhoon provided by the China Meteorological Administration (CMA) from 1949 to 2017 is used to compare the DA-Encoder-Decoder with BP, SVR, LSTM, and ELM to predict the typhoon trajectory at 24h, 48h, and 72h respectively. The results show that the root mean square error and actual error distance indicators of DA-Encoder-Decoder model are better than the other four prediction methods, demonstrating the effectiveness of this method.

Keywords: typhoon track prediction; attention mechanism; curvature; time series

收稿日期: 2020-12-17; 修订日期: 2021-03-06

基金项目: 海洋大数据分析预报技术研发基金 (2016YFC1401902); 上海市科委地方能力建设项目 (20050501900)

第一作者: 贺琪 (1979—), 教授, 博士, 主要从事于海洋大数据存储, 云计算。电子邮箱: qihe@shou.edu.cn

通讯作者: 宋巍, 教授, 博士, 主要从事于图像/视频处理、计算机视觉和海洋大数据分析。电子邮箱: wsong@shou.edu.cn

台风是一种典型的热带天气系统，是海洋-大气相互作用的重要形式之一（张志伟，2019）。它严重威胁着沿海地区人民的生命财产安全和经济的发展。因此，及时地对台风轨迹进行预测可以为防灾部门提供有效的信息支持，从而减少人员伤亡和经济损失。然而影响台风轨迹的因素众多，如台风热力学、台风动力学和台风期间环境场等（黄小燕等，2013）。并且在台风登陆之后，台风轨迹还会受到陆地地形地貌，以及海岸线水深的影响（余锦华等，2012）。所以，台风轨迹预测是一个十分重要而又充满挑战的研究课题。

早期的台风轨迹预测主要依靠热力学和空气动力学对台风环境场进行分析，结合台风登陆点对沿海地区复杂的海岸线以及陆地地形地貌影响因素的分析，建立台风轨迹预测领域特有的经验法则（王瀚，2020）。然而这种主观经验法，效率低下，并且需要大量人力物力，预测的精度和时效性都难以满足需求。随着台风监测技术和计算机技术的发展，为了解决台风轨迹预测精度不足和时效性滞后的问题，台风领域的专家提出了数值预报方法。袁炳等（2010）提出一种非对称台风 Bogus 方法，提升了台风强度预报精度。王康玲等（1996）提出了一套解决台风数值预报初始场的方案，并且已将其应用于中国南海台风预报。钱传海等（2012）讨论了不同初始场和一些边界条件对台风数值预报精度的影响。李泽椿等（2002）回顾了国家气象中心集合数值预报系统的开发过程，概述了国家气象中心现有的中期集合数值预报系统的组成、应用和发展趋势。虽然数值预报在预报性能上要远远优于主观经验法，但是数值预报的精度相较于主观经验法并没有明显提升（陈国民等，2019）。进一步提高数值预报模型的预报精度仍然是一个挑战。

目前，世界各国的气象局基本都建立了气象卫星、海洋观测站，以及地面观测站等台风三维观测系统。从 1949 年至今已经积累了大量台风数据（Gao et al, 2018）。随着台风数据的收集整理，部分研究者对台风轨迹数据的时空特征进行了分析，认为可以将台风轨迹数据视为时间序列数据，因此提出了将台风轨迹预测和时间序列预测技术相结合的方法，希望获得更好的预测效果。

循环神经网络（Recurrent Neural Networks, RNN）（Tokgöz et al, 2018）是一种适用于处理时

间序列数据的神经网络，已经在许多领域被广泛应用，但在迭代后期会出现“梯度消失”的问题。于是，Hochreiter 等（1997）最早提出了长短时记忆网络（Long Short-Term Memory, LSTM），通过增加输入门、遗忘门和输出门，使得网络自循环的权重可以变化，从而避免了“梯度消失”问题，适合处理和预报时间序列中延迟较长的事件。Ranzato 等（2014）提出了一种基于 LSTM 的视频解释算法。Sutskever 等（2014）采用了 LSTM 算法对时间序列进行预测，得到了很好的效果。研究表明，深度学习算法特别是 LSTM 算法，可以很好地应用在天气预报和海洋预报领域中。徐高扬等（2019）将 LSTM 网络应用在台风轨迹预测中，与动态时间规整算法（Dynamic Time Warping, DTW）相结合，在预测 6h 台风轨迹中效果很好。

随着神经网络的广泛应用（如金融、医学等），其包含的数据信息成指数增长，数据的长度、影响因子各有不同。传统的 LSTM 已经无法完美地解决数据不定长或者数据影响因子较多的分类或预测问题。在此基础上，20 世纪 90 年代，Forcada 等（1997）提出了一种编码器-解码器结构用来实现机器翻译。编码器-解码器网络结构无论输入和输出的长度是什么，中间的向量长度都是固定的，根据不同的任务可以选择不同的编码器和解码器。这种灵活性使得编码器-解码器网络迅速发展并且应用在各个领域。Chen 等（2018）提出了一种带可分离卷积的编码器-解码器网络并成功应用在图像分割上。Malhotra 等（2016）将编码器-解码器网络应用在时间序列的异常检测上，并取得很好的效果。Ekambaram 等（2020）利用编码器-解码器网络进行新产品的销售时间序列预测，得到了比较准确的预测结果。然而，随着时间序列的长度和特征数量增加，编码器-解码器网络的中间变量可能存储不了那么多信息，会造成精度下降，于是，一种基于注意力机制的编码器-解码器网络被提出。该注意力机制可以使网络更加集中于重要的特征，同时能够存储更长久的信息，亦有许多学者将其应用于时间序列预测领域，如：查铖等（2020）提出了一种结合注意力机制的区域海表面温度预报算法，明显提升了预报精度。Qin 等（2017）提出了一种双阶段的注意力机制模型，并将其应用在时间序列预测上，取得了很好的预测效果。Du 等（2020）

将注意力机制和编码器-解码器网络结合,对空气质量时间序列数据进行预测,取得了很好的效果。Sutskever (2014) 等针对注意力机制中的时间关系学习存在的问题,将注意力机制和编码器-解码器网络结合,发现使用注意力机制可以明显提高基于 LSTM 的编码器-解码器网络的预测能力。

虽然上述研究方法皆在一定程度上提高了预测精度问题,但目前针对台风轨迹预测还未考虑到台风轨迹数据的特征相关性和深层的时间相关性。为了解决该问题,本文基于长短时记忆网络和注意

力机制,提出一种结合双注意力机制的编码器-解码器网络模型 (Dual-Attention-Encoder-Decoder, DA-Encoder-Decoder), 首先使用台风轨迹数据计算得到台风曲率序列,充分考虑了台风轨迹数据中隐藏的转向和偏折信息。将曲率作为新的特征输入,然后通过特征注意力机制对输入特征分配权重,随后编码器将输入特征编码作为中间变量,而后再通过时间注意力机制对时间步长分配权重,最后通过解码器进行映射输出,得到预测结果。

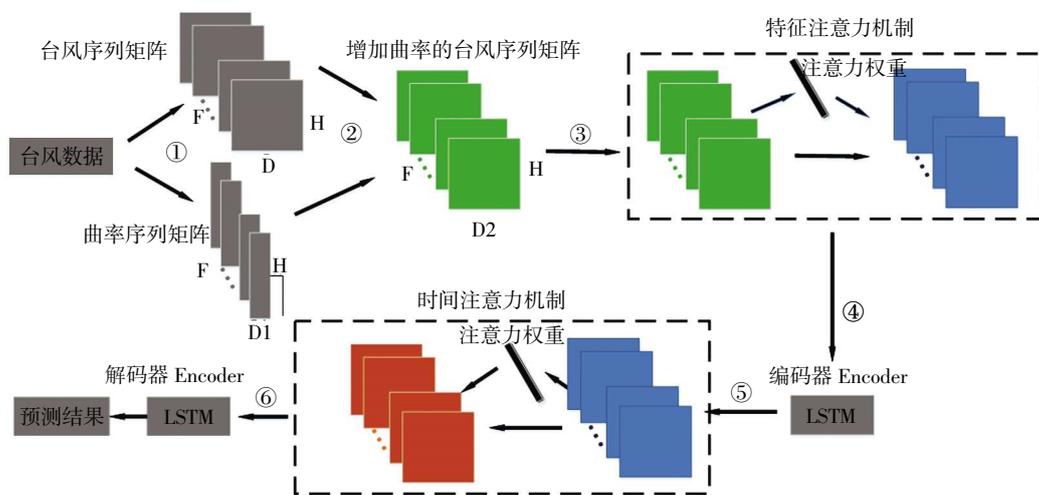


图 1 DA-Encoder-Decoder 模型流程图

1 DA-Encoder-Decoder 模型

使用台风数据集生成的时间序列矩阵, $M = M_1, M_2, \dots, M_F$, F 表示序列矩阵的时间长度 $M_i = D \cdot H$, 是台风时间序列的第 i ($1 \leq i \leq F, i \in Z$) 个输入样本。使用台风数据集计算出的曲率序列矩阵 $M^c = M_1^c, M_2^c, \dots, M_F^c$ 和台风时间序列矩阵一一对应。融合台风序列矩阵和台风曲率矩阵,得到模型的输入矩阵 $M^{input} = M_1^{input}, M_2^{input}, \dots, M_F^{input}$, $M_i^{input} = D_2 \cdot H$, i ($1 \leq i \leq F, i \in Z$) 为模型输入的一个样本。为充分考虑台风轨迹预测在时间和特征上相关性,以及历史台风轨迹数据对台风轨迹预测的影响程度。首次提出将双注意力机制和编码器-解码器模型 (DA-Encoder-Decoder) 应用到台风轨迹预测,获得了很好的效果。

图 1 给出了 DA-Encoder-Decoder 模型的流程图,两个灰色矩阵从上到下分别表示台风轨迹时间

序列矩阵和台风轨迹曲率矩阵,绿色矩阵表示融合台风轨迹矩阵和曲率矩阵的输入矩阵。蓝色矩阵表示经过特征注意力加权的特征矩阵,橙色矩阵表示经过时间注意力加权的特征矩阵。 D, D_1, D_2 表示矩阵宽度,即台风序列数据的特征个数, H 表示矩阵高度。DA-Encoder-Decoder 模型的具体步骤如下:

步骤①: 将西北太平洋台风数据库处理成台风轨迹矩阵和曲率矩阵。

步骤②: 将台风轨迹矩阵和台风曲率矩阵合并,依次按照台风轨迹时间先后进行排列,构成矩阵序列,作为 DA-Encoder-Decoder 模型的输入。

步骤③: 利用注意力机制为获得的输入矩阵特征分配注意力权重,然后将注意力权重乘上对应的特征,得到加权的特征矩阵。

步骤④: 将特征矩阵输入编码器,得到隐藏状态特征,保存特征信息。

步骤⑤: 利用时间注意力对隐藏状态特征按照

时间步长进行权重分配,然后将注意力权重乘上对应的特征,得到加权的隐藏状态特征。

步骤⑥:对获得的隐藏状态进行解码,通过映射,最终获得预测结果。

1.1 曲率特征的构建

本文中每一条台风序列可表示为 $S^i = [x_1, x_2, x_3, \dots, x_n]$, 每一个 $x_t \in S^i$ 表示台风序列的一个时间步长的输入 $[a, b]$, 其中 a 表示纬度坐标, b 表示经度坐标。然后将台风轨迹抽象到二维坐标系中,可以看作是一系列坐标点,通过二维坐标系中三点定圆原理,可以求得台风轨迹对应的曲率变化序列,求解步骤如下:

步骤 1: 首先设所求圆的半径为 R , 圆心坐标为 (x, y) , 则可由圆公式得:

$$(x_1 - x)^2 + (y_1 - y)^2 = R^2 \quad (1)$$

$$(x_2 - x)^2 + (y_2 - y)^2 = R^2 \quad (2)$$

$$(x_3 - x)^2 + (y_3 - y)^2 = R^2 \quad (3)$$

步骤 2: 由 (1) 式减去 (2) 式, 由 (2) 式减去 (3) 式, 整理可得出:

$$\begin{cases} (2x_2 - 2x_1)x + (2y_2 - 2y_1)y = x_2^2 - x_1^2 + y_2^2 - y_1^2 \\ (2x_3 - 2x_2)x + (2y_3 - 2y_2)y = x_3^2 - x_2^2 + y_3^2 - y_2^2 \end{cases} \quad (4)$$

步骤 3: 简化公式 (4) 可得:

$$\begin{cases} a = 2x_3 - 2x_2, & e = 2x_2 - 2x_1 \\ b = 2y_3 - 2y_2, & f = 2y_2 - 2y_1 \\ c = x_3^2 - x_2^2 + y_3^2 - y_2^2, & g = x_2^2 - x_1^2 + y_2^2 - y_1^2 \end{cases} \quad (5)$$

代入 (4) 式可得:

$$\begin{cases} ex + fy = g \\ ax + by = c \end{cases} \quad (6)$$

根据 (6) 式可求解出圆心:

$$\begin{cases} x = \frac{eb - af}{gb - cf} \\ y = \frac{af - be}{ag - ce} \end{cases} \quad (7)$$

步骤 4: 根据求解出的圆心坐标, 任选一点带入圆的公式可求解圆的半径, 曲率 (Curvature) 为半径的倒数:

$$R = \sqrt{(x - x_1)^2 + (y - y_1)^2} \quad (8)$$

$$\text{Curvature} = \frac{1}{R} \quad (9)$$

根据上述步骤求得每条台风轨迹的曲率序列, 加入输入轨迹序列中作为第三个特征。得到新的输入序列 $S^i = [x_1, x_2, x_3, \dots, x_n]$, 每一个 $x_t \in S^i$ 表

示台风序列的一个时间步长的输入 $[a, b, c]$, 其中 a 表示纬度坐标, b 表示经度坐标, c 表示当前坐标点对应的曲率。

1.2 编码器-解码器模型

1.2.1 编码器

编码器在本质上是一种循环神经网络, 它将我们的输入序列编码转化为一种特征。对于时间序列来说, 给定输入序列 $X = (x_1, x_2, x_3, \dots, x_T)$, 编码器可以利用下面公式学习从 x 到隐藏状态的映射。

$$h_t = f_i(h_{t-1}, x_t) \quad (10)$$

其中 $h_t \in R_m$, R_m 是 t 时刻编码器的隐藏状态, m 是隐藏状态的大小, f_i 是一个非线性激活函数, 在本文中我们使用 LSTM (长短时记忆网络)。每一个 LSTM 单元都有一个记忆细胞来记录 t 时刻的状态 S_t 。对于这个记忆细胞将会通过三个门: 遗忘门 f_t , 输入门 i_t 以及输出门 o_t 。LSTM 三个门的更新机制如下:

$$f_t = \sigma(W_f[h_{t-1}; x_t] + b_f) \quad (11)$$

$$i_t = \sigma(W_i[h_{t-1}; x_t] + b_i) \quad (12)$$

$$o_t = \sigma(W_o[h_{t-1}; x_t] + b_o) \quad (13)$$

$$h_t = o_t \odot \tanh(S_t) \quad (14)$$

$$s_t = f_t \odot s_{t-1} + i_t \odot \tanh(W_s[h_{t-1}; x_t] + b_s) \quad (15)$$

上式中 $[h_{t-1}; x_t]$ 表示当前输入和之前的隐藏层的连接。 W_f 、 W_i 、 W_o 、 b_f 、 b_i 、 b_o 、 b_s 是需要学习的参数。 σ 和 \odot 分别表示逻辑 sigmoid 函数和对应元素的相乘。

相较于传统的循环神经网络, LSTM 的细胞状态可以对一段时间内的活动进行求和, 可以克服梯度消失、梯度爆炸等问题, 并且能够更好地捕获时间序列的时序依赖问题。

1.2.2 解码器

为了预测输出经纬度, 本文使用了另外一个基于 LSTM 的循环神经网络去解码已编码的输入信息。然而 Cho 等 (2014) 提出, 随着输入时间序列长度的增加, 编码器-解码器结构网络的性能会下降。因此本文在解码之前结合了一个时间注意力机制, 针对时间步长去训练相应的注意力权重。简单地说, 就是根据前一个时间步的解码器隐藏状态计算当前编码器隐藏状态的权重。最终计算出我们期望的目标预测值。

1.3 注意力机制

注意力机制 (Attention mechanism) 是一种模仿

人们视觉神经的方法，当人们观看某个东西时，会将视线主要集中在重点关注的部分周围，获取这个部分更多的相关信息，减少对无用信息的获取。所以，注意力机制最早应用在图像分析领域，且表现出色。随后慢慢地应用到了自然语言处理领域，与深度学习相结合，显著地提升了模型的效果。目前，注意力机制经常会和 Encoder-Decoder 一起应用。注意力机制可以应用在多个领域，它们所作用的对象不同，对模型的性能提升也不尽相同，本文利用特征注意力机制和时间注意力机制来为输入特征和时间步长分配权重。

1.3.1 特征注意力机制

在使用编码器之前，使用特征注意力主要是为了获得输入特征对预测结果的影响大小的关联关系，以获得更好的预测结果。本文特征注意力的学习流程如图 2，对 t 时刻的第 n ($1 \leq n \leq k, n \in Z$) 个特征向量的隐藏状态的特征注意力权重表示方法如下：

$$e_t^n = v_e^k \tanh(W_e [h_{t-1}; s_{t-1}] + U_e x^n + b_e) \quad (16)$$

$$\alpha_t^n = \text{softmax}(e_t^n) = \frac{\exp(e_t^n)}{\sum_{l=1}^n \exp(e_t^l)} \quad (17)$$

其中， h_{t-1} 和 s_{t-1} 分别表示解码器 LSTM 前一时刻的隐含状态和单元状态。 v_e^k 、 U_e 、 b_e 是需要学习的线性层的参数， e_t^n 是度量在输入矩阵上的第 n 个特征步长的重要性的注意力权重。对 e_t^n 使用一个 softmax 函数来确保所有的注意力权重 α_t^n 的和为 1，然后可以得到针对输入特征的注意力权重。

1.3.2 时间注意力机制

输入数据经过编码器编码后，将会得到一系列隐藏状态，为了得到预测结果，需要对隐藏状态进行解码。本文使用 LSTM 网络作为编码器和解码器。然而，编码器中输入序列长度过长的时候，解码器的性能会下降。所以，本文在解码器之前利用时间注意力机制来尽量提升解码器的性能。对编码器生成的隐藏状态来说，其对时间依赖性的关注不足，需要利用时间注意力机制来对编码器的隐藏状态进行权重分配，以提升网络整体性能。简单地说，时间序列的长期依赖性可以通过加权编码器中目标值最相关的隐含状态来学习 (Sagheer et al, 2019)。本文时间注意力的学习流程如图 2，对 t 时刻的第 i ($1 \leq i \leq T, i \in Z$) 个隐藏状态的时间注

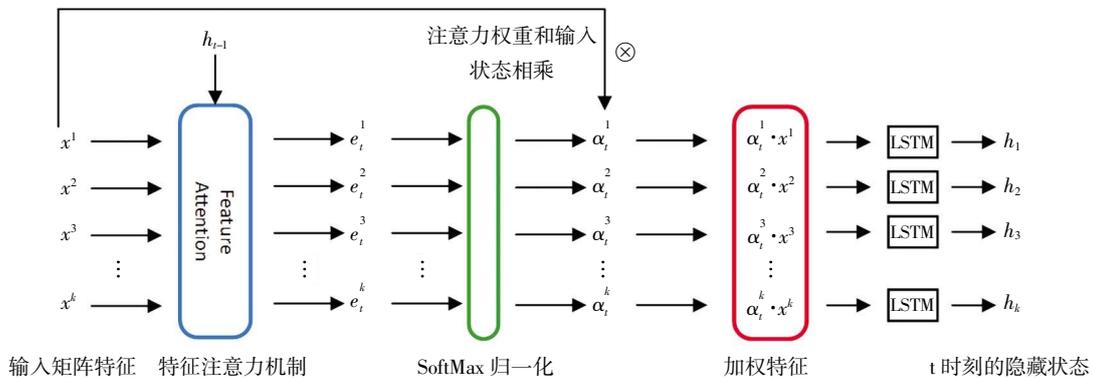


图 2 特征注意力机制流程图

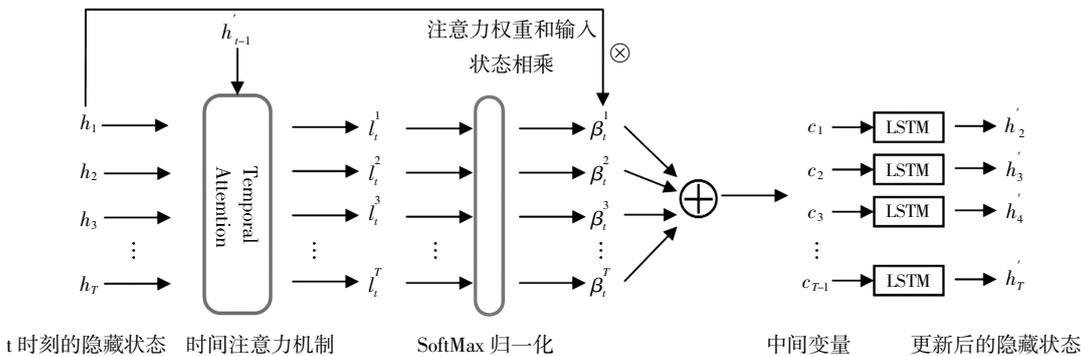


图 3 时间注意力机制流程图

注意力权重表示方法如下:

$$l_i^i = v_{h'}^T \tanh(W_h [h'_{t-1}; s_{t-1}] + U_h h_t + b_h) \quad (18)$$

$$\beta_i^i = \text{softmax} (l_i^i) = \frac{\exp(l_i^i)}{\sum_{j=1}^T \exp(l_j^i)} \quad (19)$$

$$\tilde{h}_t = (\beta_1^i h_1, \beta_2^i h_2, \dots, \beta_i^i)^T \quad (20)$$

$$c_t = \sum_{j=1}^T \beta_j^i \tilde{h}_j \quad (21)$$

其中, h'_{t-1} 和 s_{t-1} 分别表示解码器 LSTM 前一时刻的隐含状态和单元状态。 \tilde{h}_t 、 c_t 分别表示加权隐藏状态和中间向量, $v_{h'}^T$ 、 U_h 、 b_h 是需要学习的线性层的参数, l_i^i 是度量在输入矩阵上的第 i 个特征步长的重要性的注意力权重。对 l_i^i 使用一个 softmax 函数来确保所有的注意力权重 β_i^i 的和为 1, 然后可以得到针对时间步长的注意力权重。

2 实验

2.1 实验数据

本文中的实验数据来源于中国气象局热带气旋资料中心的西北太平洋最佳路径数据集 CAM-STI, 选取 1949—2017 年的路径数据, 包含经度、纬度、气压、风速, 记录间隔为 6 h, 精度为 0.1°。实验中剔除了台风持续时间小于 72 h 的台风序列数据, 保留了 2 087 条台风数据。将剩余台风序列数据按照 3:1 的比例划分为训练集和验证集, 原始输入特征包含经度和纬度。考虑到台风轨迹是由二维坐标表示, 所以对于时间步长为 t 的台风序列数据, 本文使用 $([x^1, y^1]^T, [x^2, y^2]^T, \dots, [x^t, y^t]^T)$ 表示, 实际中预测 m 个时间步长后的位置, 则输入可表示为 $\lceil [x^1, y^1], [x^2, y^2], \dots, [x^n, y^n] \rceil^T$, $\lceil [x^2, y^2], [x^3, y^3], \dots, [x^{n+1}, y^{n+1}] \rceil^T, \dots, \lceil [x^{t-n+1}, y^{t-n+1}], [x^{t-n+2}, y^{t-n+2}], \dots, [x^t, y^t] \rceil^T$, 与之相对应的输出可表示为 $[x^{n+m}, y^{n+m}], [x^{n+m+1}, y^{n+m+1}], [x^t, y^t]$ 。

本文实验使用 Keras 框架搭建, Keras 是一个高度模块化的神经网络库, 可以基于 Theano 或者 tensorflow 搭建, 并且支持扩展开发。实验硬件环境平台是 Windows10, Intel Core i7, 3.0 Hz, 8GBRAM。

2.2 精度评价指标

在使用模型进行预报时, 由于误差的计算是根据预测坐标和实际坐标的差值, 均方根误差 (Root Mean Square Error, RMSE) 能够很好地反应坐标点的偏移, 也是一种常用的精度评价指标。实际误差距离 (ErrDis) (km) 则是根据预测坐标点和实际坐标点来计算两点之间地理上的实际误差距离。能够更加明确清晰地体现预测误差, 对实际业务应用也有一定的作用。

因此, 为验证 DA-Encoder-Decoder 模型的有效性, 本实验使用 RMSE 和 ErrDis 来评估不同预报方法的性能, RMSE 和 ErrDis 的公式如下表示:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N ((x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2)} \quad (22)$$

$$\begin{cases} a = y_i - \hat{y}_i \\ b = x_i - \hat{x}_i \\ \text{ErrDis} = 2 \arcsin \sqrt{\sin^2 \frac{a}{2} + \cos x \cdot \cos \hat{x} \cdot \sin^2 \frac{b}{2}} \times R \end{cases} \quad (23)$$

其中, y_i 、 x_i 为经纬度真实值, \hat{y}_i 、 \hat{x}_i 为对应的预测值, t 表示预测的第 t 个值。 a 表示真实数据的纬度和预测数据的经度之差, b 表示真实数据的经度和预测数据的纬度之差。 R 表示地球半径, 取 6 378.137 km。

对于 RMSE 和 ErrDis 来说, 它们的值越小越好。当模型训练使得 RMSE 和 ErrDis 最小的时候, 就是最优模型。训练时, 通过观察 RMSE 和 ErrDis 的变化来确定合适的模型结构和模型参数。

2.3 实验结果分析

由于台风预测需要时效性, 实验选取训练时间步长为一天 (24 h) 的数据。然后将全部的台风轨迹数据进行划分, 将 75% 的数据作为训练集, 用于训练 DA-Encoder-Decoder 预测模型的参数, 剩下的 25% 的数据作为验证集, 用于验证模型的学习效果。

确定输入特征: 前文中提到利用台风轨迹计算出曲率序列。增加曲率的特征主要是为了模拟出台风的轨迹的曲线运动趋势, 包含了台风的转折角度和未来方向等信息。为了确定曲率特征对预测性能的提升能力, 使用台风轨迹原始数据和增加曲率特征的数据集分别预报 24 h、48 h、72 h 的轨迹坐标。对比了曲率在编码器-解码器 (Encoder-Decoder)

和基于双注意力机制的编码器解码器模型 (DA-Encoder-Decoder) 预测模型的作用。根据表 1 预测结果, 可以看到在 Encoder-Decoder 模型和 DA-Encoder-Decoder 模型上, 增加了曲率特征之后对

RMSE 和 ErrDis 都有一定的降低作用, 有利于提高台风轨迹预测的精度。因此, 本文在输入特征中增加曲率特征。

确定注意力机制: 针对 Encoder-Decoder 模

表 1 增加曲率特征性能对比

	Encoder-Decoder		Encoder-Decoder+Curvature		DA-Encoder-Decoder		DA-Encoder-Decoder+Curvature	
	RMSE	ErrDis/km	RMSE	ErrDis/km	RMSE	ErrDis/km	RMSE	ErrDis/km
预测 24 h	2.43	238.46	2.37	224	2.13	201.3	1.82	172.03
预测 48 h	4.01	381.02	3.48	328.91	3.37	319.8	2.63	245.14
预测 72 h	8.22	765.6	6.12	578.46	5.17	492.3	4.25	403.6

型, 增加不同的注意力机制, 来记录性能变化。从表 2 可以看到, 增加了注意力机制的模型预测精度要优于未增加的注意力机制的模型。增加了时间注意力的 TA-Encoder-Decoder (Temporal-Attention-Encoder-Decoder) 模型要优于增加了特征注意力的 FA-Encoder-Decoder (Feature-Attention-Encoder-Decoder) 模型。而本文提出的基于双注意力机制的 DA-Encoder-Decoder 模型效果明显优于其他注意力机制。因为利用特征注意力机制去计算输入特征的注意力权重会在更大程度上修正输入特征对预测结果的影响。第二部分的时间注意力机制会针对每次输入的样本计算时间步长的注意力权重, 尽管 LSTM 也具备这种能力, 但是在编码器解码器结构下, 长序列会大大降低模型预测精度, 而时间注意

力机制能够解决这个问题, 可以对输入数据进行长时期的权重计算, 将之前的信息存储在隐藏层中。

表 2 是针对不同的注意力机制与预测模型结合的预测结果。预测模型分别对 24 h、48 h、72 h 台风轨迹进行预测。可以看出, 增加了时间注意力的模型 (TA-Encoder-Decoder) 精度比单纯的 Encoder-Decoder 模型高, 证明了在时间维度上的注意力权重训练能够提高模型的预测精度。同理, 在特征维度上的注意力权重训练也能够提高模型预测精度。因此, 特征相关性和时间相关性对台风轨迹预测的精度均有影响。本文的 DA-Encoder-Decoder 模型既考虑了特征相关性, 又考虑了时间相关性, 从表 2 可以看出, DA-Encoder-Decoder 模型的精度要明显优于其他同类模型的预测精度。

表 2 增加不同注意力机制性能对比 (输入包含曲率特征)

	Encoder-Decoder		TA-Encoder-Decoder		FA-Encoder-Decoder		DA-Encoder-Decoder	
	RMSE	ErrDis/km	RMSE	ErrDis/km	RMSE	ErrDis/km	RMSE	ErrDis/km
预测 24 h	2.37	224	2.26	208.4	2.31	219.6	1.82	172.03
预测 48 h	3.48	328.91	2.95	289.1	3.04	299.82	2.63	245.14
预测 72 h	6.12	578.46	5.43	510.72	5.23	489.6	4.25	403.6

预报方法对比: 确定了预测模型的输入特征、注意力机制, 预报模型基本上已经确定, 为了验证本文提出的 DA-Encoder-Decoder 模型的有效性, 分别与传统预测方法以及神经网络方法进行了对比。对比方法包括: BP、SVR、LSTM、ELM。对于 SVR 网络使用的是径向基核函数 (Radial Basis Function, RBF), 该核函数能够实现非线性映射并且需要学习的参数较少。以上对比试验均采用 Keras 库函数进行搭建。训练数据使用数据集的 75%, 验证数据使用 25%。通过预测 24 h、48 h、72 h 的台风轨迹, 对比模型的 RMSE 和 ErrDis 指

标来比较模型之间的预测性能的差异。

表 3 是所有对比模型和本文模型的预测结果。可以看出, 在预测 24 h 台风轨迹时, BP、SVR、LSTM、ELM、DA-Encoder-Decoder 模型的 RMSE 指标分别为 3.48、2.96、2.81、5.62、1.82, ErrDis 指标分别为 330.6 km、279.24 km、263.14 km、539.55 km 和 172.03 km。在 24 h、48 h、72 h 上的预测中, 与其他机器学习、深度学习预测模型相比 DA-Encoder-Decoder 模型降低了预测误差, 具备最好的预测性能。

表 3 不同模型性能对比 (输入包含曲率特征)

	BP		SVR		LSTM		ELM		DA-Encoder-Decoder	
	RMSE	ErrDis/km	RMSE	ErrDis/km	RMSE	ErrDis/km	RMSE	ErrDis/km	RMSE	ErrDis/km
预测 24 h	3.48	330.6	2.96	208.4	2.31	279.24	2.81	263.14	1.82	172.03
预测 48 h	5.21	504.37	2.95	289.1	5.03	482.81	4.3	408.6	2.63	245.14
预测 72 h	8.73	822.62	5.43	510.72	7.9	764.3	6.42	612.74	4.25	403.6

3 结束语

本文提出的 DA-Encoder-Decoder 模型在台风轨迹预测上的应用具有一定的研究意义,它充分挖掘了台风轨迹数据的特征和时间信息,在输入上结合了台风轨迹的曲率序列,包含台风轨迹的转向、偏折等隐藏信息,与同类台风轨迹预测模型相比,提高了预测准确性。

本文主要的结论有三点:

(1) 之前的台风轨迹预测都是只单纯考虑台风时间序列经纬度之间的关系,而本文将包含转向、偏折信息的曲率特征作为预测特征,考虑了台风转向等隐藏因素对台风轨迹的影响。

(2) 利用了特征注意力机制为特征分配权重,得到不同的特征对要预测的轨迹点在特征维度上的影响。特征与权重相乘得到加权特征,突出了关键特征,提升了预测精度。

(3) 利用了时间注意力机制为特征分配权重,反映出历史台风轨迹数据对要预测的轨迹点在时间维度上的影响,历史台风轨迹数据距离预测的轨迹点的时间远近不同,对其的影响也不相同;特征与权重相乘可得到加权特征,使得关键信息被突出,加权特征包含的信息越明显,预报精度便越高。

对比不同方法在同一数据集下的预测结果,DA-Encoder-Decoder 分别预测 24 h、48 h、72 h 台风轨迹,误差分别为 172.03 km、245.14 km、403.6 km,实验结果表明 DA-Encoder-Decoder 模型在台风轨迹预测方面获得了一定的精度提升,从而验证了本文方法的有效性,该模型值得更深入地研究和应用。由于台风轨迹所受到的影响因子众多,如风速、压强、降雨量等,这些因子对台风轨迹的变化具有一定的影响,因此在后续的研究中也应该充分考虑其他相关要素对台风轨迹的影响。在下一步工作中,利用要素相关性或关联规则去筛选对台风轨迹影响较大的要素,利用这些要素建

立多要素预测模型,从而进一步提高台风轨迹预测精度。

参 考 文 献

- 陈国民,张喜平,白莉娜,等,2019.2017年西北太平洋和南海热带气旋预报精度评定[J].气象,45(4):577-586.
- 查铖,贺琪,宋巍,等,2020.结合注意力机制的区域型海表面温度预报算法[J].海洋通报,39(2):191-199.
- 黄小燕,金龙,2013.基于主成分分析的人工智能台风路径预报模型[J].大气科学,37(5):1154-1164.
- 李泽椿,陈德辉,2002.国家气象中心集合数值预报业务系统的发展及应用[J].应用气象学报,(1):1-15.
- 钱传海,端义宏,麻素红,等,2012.我国台风业务现状及其关键技术[J].气象科技进展,2(5):36-43.
- 王瀚,2020.基于深度学习的台风路径预测多模型算法研究[D].成都:电子科技大学.
- 王康玲,何安国,薛纪善,1996.南海区域台风路径数值预报业务模式的研究[J].热带气象学报,(2):113-121.
- 徐高扬,刘姚,2019.LSTM网络在台风路径预测中的应用[J].计算机与现代化,285(5):68-72.
- 袁炳,费建芳,王云峰,等,2010.一种非对称台风 Bogus 方法的数值模拟应用[J].海洋通报,29(2):187-193.
- 余锦华,唐家翔,戴雨涵,等,2012.我国台风路径业务预报误差及成因分析[J].气象,38(6):695-700.
- 张志伟,2019.西北太平洋上层海洋对台风响应的个例研究[J].海洋通报,38(5):562-568.
- CHEN LC, ZHU Y, PAPANDREOU G, et al, 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation[J]. Springer, Cham: 801-818.
- DU S, LI T, YANG Y, et al, 2020. Multivariate time series forecasting via attention-based encoder-decoder framework[J]. Neurocomputing, 388: 269-279.
- EKAMBARAM V, MANGLIK K, MUKHERJEE S, et al, 2020. Attention based Multi-Modal New Product Sales Time-series Forecasting[C]. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 3110-3118.
- FORCADA ML, ECO RP, 1997. Recursive hetero-associative memories for translation[C]. In International Work-Conference on Artificial Neural Networks. Berlin, Heidelberg: Springer.
- GAO S, ZHAO P, PAN B, et al, 2018. A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network[J]. Acta Oceanologica Sinica, 37(5): 8-12.

- HOCHREITER S, SCHMIDHUBER J, 1997. Long Short-Term Memory [J]. *Neural Computation*, 9(8): 1735–1780.
- MALHOTRA P, RAMAKRISHNAN A, ANAND G, et al, 2016. LSTM-based encoder-decoder for multi-sensor anomaly detection [J].? arXiv preprint arXiv: 1607.00148.
- QIN Y, SONG D, CHEN H, et al, 2017. A dual-stage attention-based recurrent neural network for time series prediction[J]. arXiv preprint arXiv:1704.02971.
- RANZATO M, SZLAM A, BRUNA J, et al, 2014. Video (language) modeling: a baseline for generative models of natural videos[J]. arXiv preprint arXiv:1412.6604.
- SUTSKEVER I, VINYALS O, LE QV, 2014. Sequence to sequence learning with neural networks[J]. arXiv preprint arXiv:1409.3215.
- SAGHEER A, KOTB M, 2019. Time series forecasting of petroleum production using deep LSTM recurrent networks[J]. *Neurocomputing*, 323, 203–213.
- TOKGÖZ A, ÜNAL G, 2018. A RNN based time series approach for forecasting turkish electricity load[C]. 2018 26th Signal Processing and Communications Applications Conference (SIU): 1–4.

(本文编辑：王少朋)