

硕士学位论文

面向海表温大数据的时间序列 预测技术研究



导师：乔百友

研究生：王永前

東北大學

二〇一九年十二月

分类号_____ 密级 _____

UDC _____

学 位 论 文

面向海表温大数据的时间序列 预测技术研究

作者姓名：王永前

指导教师：乔百友 副教授

东北大学计算机科学与工程学院

申请学位级别：硕士 学科类别：工学

学科专业名称：计算机应用技术

论文提交日期：2019 年 12 月 论文答辩日期：2019 年 12 月

学位授予日期：2020 年 01 月 答辩委员会主席：杨晓春 教授

评 阅 人 ： 王波涛 教授

东 北 大 学

2019 年 12 月

A Thesis in Computer Application Technology

Study on Time Series Prediction Technology for Sea Surface Temperature Data

By Wang Yongqian

Supervisor: Associate Professor Qiao Baiyou

Northeastern University

December 2019

独创性声明

本人声明，所呈交的学位论文是在导师的指导下完成的。论文中取得的研究成果除加以标注和致谢的地方外，不包含其他人已经发表或撰写过的研究成果，也不包括本人为获得其他学位而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：

日 期：

学位论文版权使用授权书

本学位论文作者和指导教师完全了解东北大学有关保留、使用学位论文的规定：即学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人同意东北大学可以将学位论文的全部或部分内容编入有关数据库进行检索、交流。

作者和导师同意网上交流的时间为作者获得学位后：

半年 ☐

一年 ☐

一年半 ☐

两年 ☐

学位论文作者签名：

导师签名：

签字日期：

签字日期：

摘 要

海水表面温度（海表温）是表示海水热力状况的一个物理量，也是海洋水文状况中最重要的因子之一，对海洋生态系统有着巨大的作用和影响。海表温被广泛应用于海洋动力学和海气相互作用的研究，在渔业经济研究和污染检测等方面也扮演着极其重要的角色，因此对海表温进行精确预测是极其重要的。目前海表温的预测方法多以经验预测方法、数值预测方法和统计预测方法为主，这些预测方法的预测精度有限。而现有的基于机器学习的方法大都针对单观测点（单点）进行海表温的短期预测，无法高效支持局域海表温的预测。为此，本文基于海表温大数据，结合深度学习技术对海表温预测技术进行了深入研究，本文将海表温预测问题划分为单点海表温预测和局域海表温预测两大类，每一类又根据预测步长分别提出了短期、中期、长期三类预测模型，具体工作如下：

(1) 提出了一种基于 XGBoost 的单点短期海表温预测模型。该模型在数据处理上考虑“邻域点”对预测点的影响，利用 XGBoost 模型的并行优化、损失精确、防止过拟合等优点，实现了对短期海表温的准确预测；在此基础上，提出了基于 XGBoost 结合 LSTM 的中期海表温预测模型，该模型利用 XGBoost 来提取观测点数据的时间特征，并结合周围邻域点的历史数据，作为 LSTM 模型的输入，充分利用 LSTM 模型具有提取长短期特征和多步输出的优势实现了海表温的中期预测。

(2) 针对单点长期海表温预测，提出了基于 GRU 和一维卷积（CNN-1D）结合注意力机制的预测模型。首先采用 CNN-1D 对历史数据进行卷积操作，提取数据特征并实现对历史数据的压缩，从而降低后续处理的复杂度；然后再使用 GRU 模型进行训练，提取序列的时序特征；最后采用注意力机制，为 GRU 每一步的输出设定一个合理的权重，修正 GRU 输出结果，从而进一步提高长期预测的精度。

(3) 针对局域海表温预测问题，提出基于二维卷积（CNN-2D）和 LSTM 融合注意力机制的长期预测模型。首先将局域同一时刻的数据看作是一张通道数为1的二维图像。针对短期、中期海表温预测，采用 CNN-2D 结合 LSTM 进行预测。CNN-2D 来提取观测点与观测点之间数据的空间特征，然后使用 LSTM 学习数据的时序特征，最后进行训练预测；针对长期预测，提出基于 CNN-2D 和 LSTM 融合注意力机制的模型，学习到数据的时空特征以后，通过注意力机制来对结果进行修正，从而减小预测的误差。

(4) 分别选取渤海海域和南海海域的观测点的海表温数据，从模型的预测精度和泛化能力等方面对提出的上述几种模型进行了实验分析。实验结果表明本文提出的几种预测模型具有较小的预测误差，特别是中长期预测模型的精度均优于现有的预测模型。

这一块可以借鉴，
预测一个面的海洋
平面温度。

综上所述，本文工作是对现有海表温预测研究工作的有益补充，所提出的几种预测模型具有较高的预测精度、广阔的应用价值和前景。

关键词：海表温预测；XGBoost；CNN；LSTM；注意力机制

Abstract

Sea surface temperature (SST) is a physical quantity that represents the thermal condition of seawater and is one of the most important factors in the ocean hydrological condition. SST is widely used in the study of ocean dynamics and the interaction between sea and air. It also plays an extremely important role in the research of fishery economy and pollution detection. Therefore, accurate prediction of SST is extremely important. At present, SST prediction is mainly based on empirical prediction methods, numerical prediction methods and statistical prediction methods, which have limited prediction accuracy. However, most of the existing methods based on machine learning aimed at short-term SST prediction of single observation point (single point), which could not effectively support local SST prediction. For this reason, this thesis is based on sea surface temperature data, combined with the deep learning of sea surface temperature forecasting technology were studied, this thesis could be divided into single point of sea surface temperature prediction problem of sea surface temperature and local sea surface temperature prediction two categories, each category and according to the prediction step length respectively short-term, medium-term and long-term forecasting model is put forward, and the specific work is as follows:

(1) A single point short-term SST prediction model based on XGBoost is proposed. This model considers the influence of "neighborhood points" on the prediction points in data processing, and makes use of the advantages of XGBoost model, such as parallel optimization, accurate loss and over-fitting prevention, to achieve accurate prediction of short-term SST. On this basis, put forward based on XGBoost LSTM mid-term sea surface temperature prediction model, the model using XGBoost to extract the time characteristics of observation point data, combined with the surrounding neighborhood points of historical data, as the LSTM model input, make full use of the LSTM model has the advantages of long-term feature extracting and multi-step output has realized the sea surface temperature in the middle of prediction.

(2) For single-point long-term SST prediction, a prediction model based on GRU and one-dimensional convolution (cnn-1d) combined with attention mechanism was proposed. First, cnn-1d was used to carry out convolution operation on historical data to extract data features and realize compression of historical data, so as to reduce the complexity of subsequent processing. Then the GRU model is used for training to extract the sequence characteristics.

Finally, the attention mechanism is adopted to set a reasonable weight for the output of GRU at each step, and to modify the GRU output result, so as to further improve the accuracy of long-term prediction.

(3) To solve the problem of local SST prediction, a long-term prediction model based on two-dimensional convolution (cnn-2d) and LSTM is proposed. Firstly, the local data at the same time is regarded as a two-dimensional image with a channel number of 1. For short-term and medium-term SST prediction, cnn-2d combined with LSTM was used for prediction. Cnn-2d was used to extract the spatial features of the data between the observation points and the observation points, then LSTM was used to learn the timing features of the data, and finally the training prediction was conducted. For long-term prediction, a model based on cnn-2d and LSTM combined attention mechanism was proposed. After learning the spatio-temporal characteristics of the data, the results were modified through the attention mechanism, so as to reduce the prediction error.

(4) The temperature data of observation points in bohai sea and south China sea were selected respectively. The experimental results show that several prediction models proposed in this thesis have small prediction errors, especially the accuracy of medium and long-term prediction models is better than the existing prediction models.

In conclusion, the work of this thesis is a useful supplement to the existing SST prediction research. The proposed prediction models have high prediction accuracy, broad application value and prospects.

Key words: SST prediction; XGBoost; CNN; LSTM; attention mechanism

目 录

| | |
|-----------------------------|----|
| 独创性声明 | I |
| 摘 要 | II |
| Abstract | IV |
| 第 1 章 绪 论 | 1 |
| 1.1 研究背景与意义 | 1 |
| 1.2 国内外研究现状 | 2 |
| 1.2.1 传统时间序列预测方法 | 2 |
| 1.2.2 基于机器学习的时间序列预测方法 | 3 |
| 1.2.3 基于深度学习的时间序列预测方法 | 4 |
| 1.2.4 现有方法的不足 | 4 |
| 1.3 主要研究内容 | 5 |
| 1.4 本文组织结构 | 6 |
| 第 2 章 相关理论与技术 | 9 |
| 2.1 时间序列预测方法 | 9 |
| 2.2 Boosting | 9 |
| 2.2.1 XGBoost | 9 |
| 2.2.2 AdaBoost | 12 |
| 2.2.3 CatBoost | 13 |
| 2.3 卷积神经网络 | 14 |
| 2.3.1 卷积层 | 14 |
| 2.3.2 池化层 | 16 |
| 2.3.3 全连接层 | 17 |
| 2.3.4 目标函数 | 17 |
| 2.4 循环神经网络 | 18 |
| 2.4.1 RNN 模型 | 18 |
| 2.4.2 LSTM 模型 | 19 |
| 2.4.3 GRU 模型 | 20 |
| 2.5 注意力机制 | 21 |
| 2.6 本章小结 | 22 |

这一块出个摘要

| | |
|--|-----------|
| 第 3 章 单点海表温预测 | 23 |
| 3.1 问题提出 | 23 |
| 3.2 数据选取及预处理 | 24 |
| 3.2.1 数据选取 | 24 |
| 3.2.2 数据预处理 | 25 |
| 3.2.3 时间特征的选择 | 27 |
| 3.3 基于 XGBoost 的短期海表温预测 | 28 |
| 3.4 基于 XGBoost 结合 LSTM 的中期海表温预测 | 29 |
| 3.5 基于 GRU 和 CNN-1D 结合注意力机制的长期海表温预测 | 31 |
| 3.6 本章小结 | 35 |
| 第 4 章 局域海表温预测 | 37 |
| 4.1 问题提出 | 37 |
| 4.2 数据选取及预处理 | 38 |
| 4.2.1 数据选取 | 38 |
| 4.2.2 数据预处理 | 39 |
| 4.3 基于 LSTM 结合 CNN-2D 的短中期海表温预测 | 41 |
| 4.4 基于 CNN-2D 和 LSTM 融合注意力机制的长期海表温预测 | 42 |
| 4.5 本章小结 | 44 |
| 第 5 章 实验与分析 | 47 |
| 5.1 实验环境 | 47 |
| 5.2 评估标准 | 47 |
| 5.3 单点海表温预测 | 48 |
| 5.3.1 实验相关设置 | 49 |
| 5.3.2 实验结果及分析 | 49 |
| 5.4 局域海表温预测 | 62 |
| 5.4.1 实验相关设置 | 62 |
| 5.4.2 实验结果及分析 | 63 |
| 5.5 本章小结 | 67 |
| 第 6 章 总结与展望 | 69 |
| 6.1 总结 | 69 |
| 6.2 进一步的工作 | 70 |

| | |
|---------------------|----|
| 参考文献..... | 71 |
| 致 谢..... | 75 |
| 攻硕期间参与项目及获奖情况 | 77 |

第1章 绪论

海水表面温度（海表温）是海洋水文状况中最重要的参数之一。研究海表温随时间地点的变化趋势，是海洋学中的重要内容之一。而近海对人们的生产生活有着直接的影响，且近海海表温变化幅度较远海海表温更大，所以本文致力于近海海表温的预测研究。本章将首先介绍课题的研究背景与意义，并对国内外研究现状进行系统综述，接着给出本文的研究内容和组织结构。

1.1 研究背景与意义

近年来地球环境持续恶化，有科学研究表明，如果全球继续变暖^[1]，在本世纪内，海水将吞没马尔代夫。如果不能拿出有效的策略控制全球变暖，很有可能造成重大灾难。从长远角度来看，海洋在调节全球气候方面扮演着重要角色，海表温是表示海洋热力状况的一个物理量，对海洋生态系统有着巨大的作用和影响。所以海表温变化的研究对人类保护自身生存环境有着重大的意义。从我们国家来看，研究近海海表温变化对于预报冬季海冰冰情^[2]，从而减少海冰造成的损失也非常重要。在每年冬季，渤海会有几个月结冰期，这些海冰的多少对海温状况影响很大，也会直接影响海洋养殖^[3]的发展。除此之外，海冰对我们的船只、海上运输都会产生巨大影响，严重情况下会毁坏海洋工程基础设置，使得航道无法通行。

现在，人类的经济发展对海洋的依赖极大。对我国来说，渤海、南海等都有着较为发达的沿海旅游，旅游业的收入是当地人收入的重要部分^[4]；发展海洋生物渔业也是沿海居民收入的主要来源之一^[5]。除此之外，沿海海底矿产^[6]资源也极其丰富。正是依靠这些天然的资源，沿海的发展极其迅速，而这些海洋资源的开发利用又受到海表温的制约。从各项数据来看，海表温对海洋生物的物种丰富度、海洋生物的生存区域具有决定性的意义，海表温的变化会直接影响沿海养殖业的产量^[7]。所以，研究海表温变化及其规律对做好沿海水产养殖、渔场天气预报以及促进我国沿海经济的可持续发展有着极其重要的作用。另外，人们对海洋旅游需求一直很旺盛，因此海表温的预测对于沿海海域环境水质检测，发布沿海环境预报，以及发展沿海旅游经济都是十分重要的。

海表温是海洋要素场中的基本量之一，是重要的海洋物理参数与基础信息，在全球和区域海洋的各个领域研究中起着重要作用。通过对特定区域的海表温的观测和分析，可获得研究区域的海洋物理过程、海洋生物变化以及海洋灾害等重要信息。一直以来，我国十分重视海洋的探索、开发、保护。现如今尝试利用当下先进科学技术来探索和保

护海洋。国家重点研发计划“海洋环境安全保障”重点专项“海洋大数据分析预报技术研发”是科技部“十三五”期间在海洋领域批复的第一个也是唯一一个大数据项目，而海表温时间序列预测是“海洋大数据分析预报技术研发”中的研究内容之一。

1.2 国内外研究现状

海表温是监测海洋现象的重要参量，具有重要研究价值，不仅在渔业发展研究和海气相互作用上有着广泛应用，还在海洋动力学和污染检测等方面有着不可替代的作用^[8]。因海表温变化有着明显的时间特征，所以本文利用时间序列方法来探索海表温变化情况。时间序列的任意时刻的值和它历史数据是有一定关系的，于是本文想利用这种关系，建立历史数据和未来数据的关系模型，从而能够通过历史数据预测未来数据。《Time Series Analysis Forecasting and Control》在 1970 年出版后，处理时间序列的方法逐渐形成。对于海表温时间序列预测问题，目前主要是三大类方法：第一类是数值时间序列预测方法，使用一系列复杂方程来描述海表温变化规律；第二类是机器学习时间序列预测方法；第三类是深度学习时间序列预测方法。下面将从这三方面来介绍关于时间序列预测的国内外研究现状。

1.2.1 传统时间序列预测方法

在传统时间序列预测中，将数据分为两种，一种是线性数据，另一种是非线性数据，通常用线性模型来处理线性数据，非线性模型处理非线性数据。就线性模型而言，目前比较成熟，使用简单且有很多研究应用。2008 年刘慧等人^[9]利用 ARMA 模型对海浪过程进行仿真研究，结果表明 ARMA 方法在平滑、分辨率和偏差等方面更加有效。2016 年段文洋等人^[10]提出一种复合的经验模态分解自回归预报模型（EMD-AR），解决了非线性非平稳波浪的极短期预报问题。实验中使用三个海洋波浪实测数据进行验证，这三个海浪数据是选自不同的地理位置，最终分析出非线性非平稳性对波浪预报的影响，结果表明 EMD-AR 模型的效果更好。自回归整合滑动平均模型（ARIMA）可以用来解决非平稳时间序列的处理问题，很多文献介绍了如何使用 ARIMA 模型解决问题。2016 年刘珊等人^[11]为最近 30 年的台风逐月生成频次时间序列建立合适的 ARIMA 模型，使用的数据是日本东京台风中心的逐月台风生成频次数据，并用该模型对 2010-2014 年的台风逐月生成频次进行了预测，结果表明 ARIMA 模型在预测台风频次方面和观测值基本一致。2019 年张莹等人^[12]使用高时空分辨率的海表温度数据，将集合经验模态分解和 ARIMA 结合起来建立海表温预测模型，实验结果表明该方法在丰富的数据支撑下，通过对各种预测方法预测结果的综合比较和分析，该模型的预测效果更好。

此外,也有一些人使用多变量场合、异方差场合和非线性场合的时间序列分析方法进行研究。在异方差场合,自回归条件异方差(ARCH)模型^[13]、广义自回归条件异方差(GARCH)模型^[14]、门限 ARCH(TARCH)模型^[15]、自回归条件异方差(EGARCH)模型^[16]、有偏 ARCH(APARCH)模型^[17]相继提出。这些异方差模型是对经典的 ARIMA 模型很好的补充。

1.2.2 基于机器学习的时间序列预测方法

传统的时间序列预测方法理论成熟、使用简单,基于线性方法进行预测,不能很好地处理非线性问题。为此,研究人员提出了很多机器学习预测算法,并通过逻辑推理和大量实验证明了这些机器学习算法可以很好的处理非线性数据。人工神经网络(Artificial Neural Network, ANN)^[18]是较早出现的机器学习方法,后期提出遗传算法(Genetic Algorithm, GA)^[19]等寻求最优解的算法,还有极限学习机(Extreme Learning Machine, ELM)^[20]等都属于基础的机器学习方法。

2010 年李雪等人^[21]从阈值角度出发建立 BP 人工神经网络模型,解决海水水质类别评价的问题。随后使用渤海湾近岸海域水环境评价,经过模型计算后得到水质类别,结果表明该模型结构的设计比较合理,实验测定具有较好的实用性、通用性。2012 年韩震等人^[22]建立了西北太平洋海水温度 LM-BP 模型并使用 2008 年和 2009 年的 Argo 数据对模型进行检验,结果表明该模型的预测效果较好。

遗传算法是模拟自然进化的过程,已经广泛应用于各个领域,它被认为是一个非常强大和成功的解决实际问题的方案,用来解决人类面临的各种实际性问题。2019 年李强等人^[23]建立了 GA-BP 神经网络预测模型用来测定海洋沉积物烧失量,使用遗传算法优化 BP 的初始权重和偏置,为烧失量的测定提供了一种新的有效途径。

极限学习机最初由 Huang 等人提出,是一种前馈神经网络,用于单层隐藏层节点的分类或回归,它可以克服基于梯度的方法面临的一些局限性。除此之外,极限学习机还有速度快、泛化能力强等优点,在解决实际问题时得到了人们的广泛关注。2015 年张颖等人^[24]利用 ELM 建立海水叶绿素浓度预测模型。通过与支持向量机回归等模型进行对比实验,结果表明 ELM 的回归预测效果误差更小、预测效率和泛化能力更好。2017 年崔翔鹏等人^[13]为预测船舶交通流量采用 ELM 建立预测模型,同时采用 GA 进行优化。利用上海洋山港船舶流量对该模型进行测试,结果表明该模型的预测结果比单纯使用 BP 模型、单纯使用 ELM 模型的效果更好、预测精度更高。2019 年吴忠强等人^[26]使用 ELM 建立水深反演非线性回归模型,以 Sentinel-2A 遥感影像为数据源测试模型,并与

BP 等其他模型进行比较, 结果表明 ELM 反演精度更高, 尤其是在水深 5~20 米处, 具有较高的反演精度和良好的效果。

1.2.3 基于深度学习的时间序列预测方法

深度学习近年来在语音识别、机器翻译等领域表现优异, 在图像识别、自然语言处理方面更是取得了很多成果, 处理了很多复杂的模式识别难题。随着深度学习的快速发展, 逐渐有人开始将深度学习用于海洋数据的相关研究。2017 年张勤等人^[27]提出使用 LSTM 来构建海表温预测模型, 利用 LSTM 学习数据的时序关系, 利用全连接层映射输出结果, 使用中国沿海数据进行实验, 结果表明该模型的预测效果更好。2017 年顾兴健等人^[28]提出一种基础 LSTM 神经网络的海洋环境短期预报模型, 在网络训练过程中使用粒子群算法优化网络权重参数, 以我国东海、南海、黄海数据进行测试, 结果表明该模型在短期范围内的预报取得了较好的效果。2018 年杨玉婷等人^[29]建立 CFCC-LSTM 模型来预测海表面温度, 采用渤海数据集进行验证, 结果表明 CFCC-LSTM 模型的预测误差更小。2019 年朱贵重等人^[30]通过 LSTM-RNN 构建西太平洋海表温预测模型, 结果表明 LSTM-RNN 的效果显著, 应用于海表温建模是可行的。

1.2.4 现有方法的不足

指数平滑法^[31]、移动平均法、加权移动平均法等方法属于传统时间序列建模的典型方法, 被称为平滑方法, 它们有一个共同点: 都要“消除”数据中不规则成分造成的随机波动。平滑方法的优势: 平滑方法基于严谨的数学公式推导而来, 理论成熟且使用非常简单; 平滑方法在近距离预测上的表现更好, 可提供较高的精准水平。平滑方法的劣势是: 对于没有明显的趋势、变动幅度较小的稳定序列来说, 平滑方法表现很好; **但是当数据的变化有明显的季节变差、趋势、循环等特点时, 平滑方法的预测效果很差。**

循环神经网络 (RNN) 是深度学习中的模型。它是一种反馈神经网络, 通过内部特殊构造来处理任意时序的输入序列。但 RNN 的内部特殊构造, 记忆的只是部分的序列, 所以在短序列上的表现优于长序列, 序列越长, 预测误差越大, 偏差越大。LSTM 模型在单元内部通过添加三个“门”来解决长距离依赖的问题, 是一种改进的 RNN 模型, **但仍然会出现随着预测步长增加, 预测误差越来越大的情况; 且 LSTM 计算更为复杂, 所以训练模型的时间更长。**

以上所有方法, 都是针对的单点的时间序列预测。除此之外, **因为海表温实际上是以“场”的形式存在, 点与点之间是相互影响的, 所以在模型处理的过程中, 只考虑历史数据是不充分的, 应该将邻域点的影响也考虑进去。**

1.3 主要研究内容

本文的主要研究内容是面向海表温数据的时间序列预测技术，主要包括数据预处理、单点海表温预测建模、局域海表温预测建模。图 1.1 是本文研究内容的总体框架。

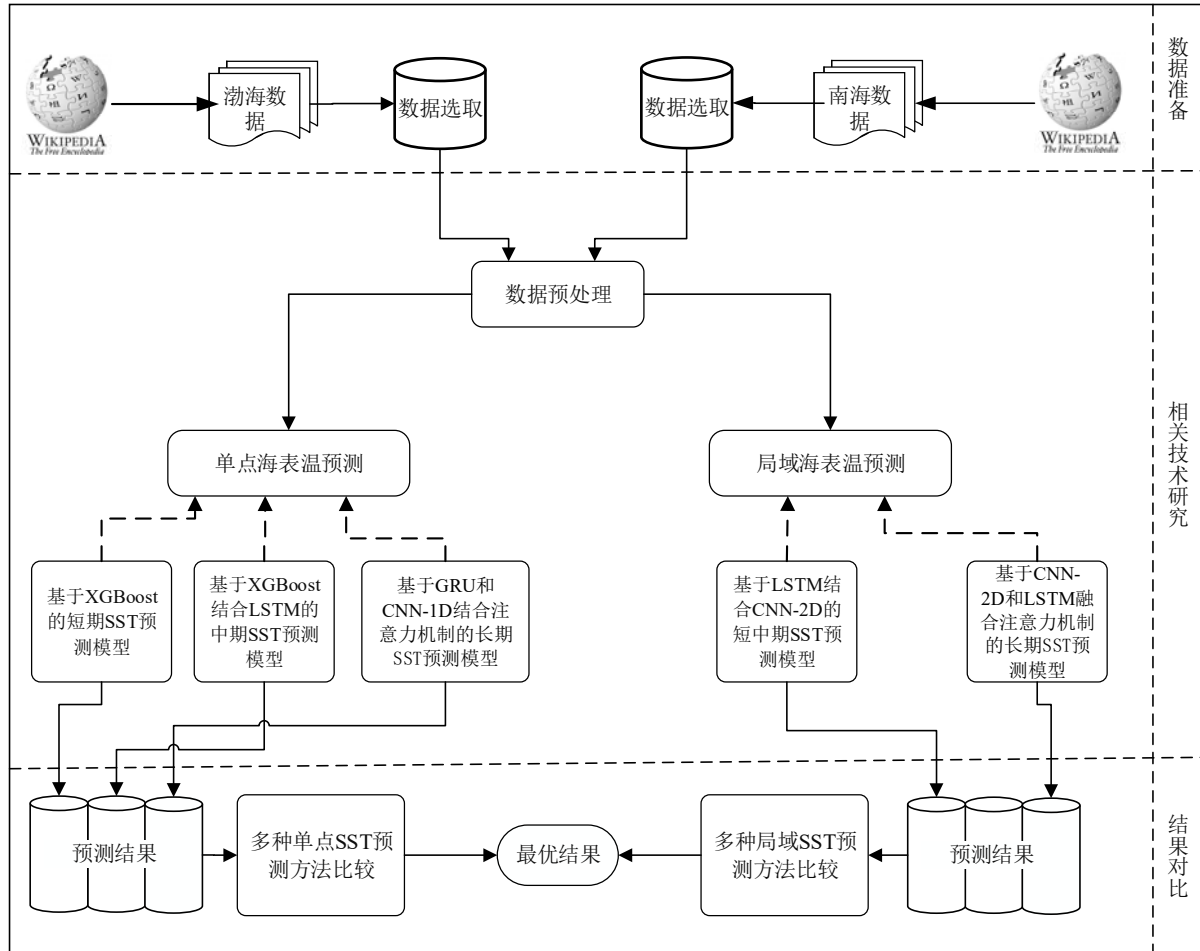


图 1.1 总体研究框架

Fig. 1.1 The overall research framework

由图 1.1 可知本文主要的研究内容如下：

(1) 数据集的预处理。本文使用的数据集是 NOAA/OAR/ESRL PSD 提供的高分辨率海温数据,该数据来自美国科罗拉多州博尔德市的网站 <http://www.esrl.noaa.gov/psd/>。该数据集包含 SST 日平均数据、SST 日异常数据、SST 周平均数据和 SST 月平均数据,范围从 1981 年 9 月 1 日至今,覆盖全球海洋 (89.875S-89.875N, 0.125E-359.875E),是 $0.25^{\circ} \times 0.25^{\circ}$ 网格表示。对数据的预处理主要包括根据经纬度选出我国渤海 1982 年到 2018 年共 37 年的渤海 SST 日平均数据。渤海范围 37.07N~41N, 117.35E~121.10E。这个范围包括渤海和一部分陆地,处理时提取部分海面数据进行训练和预测;选取南海海域数据的范围 105E~118E,4N~21N。采用 numpy 包中的 `log1p()` 函数对数据进行平滑处

理。通过 `pandas` 库给数据添加日期特征，从日期特征提取出年、季节、季度、月、周、日 8 种时间特征，作为输入数据的一部分。

(2) 单点海表温预测建模。首先提出了一种基于 XGBoost 的单点短期海表温预测模型。该模型在数据处理上考虑“邻域点”对预测点的影响，利用 XGBoost 模型的并行优化、损失精确、防止过拟合等优点，实现了对短期海表温的准确预测；在此基础上，提出了基于 XGBoost 结合 LSTM 的中期海表温预测模型，该模型利用 XGBoost 来提取观测点数据的时间特征，并结合周围邻域点的历史数据，作为 LSTM 模型的输入，充分利用 LSTM 模型具有提取长短期特征和多步输出的优势实现了海表温的中期预测。针对单点长期海表温预测，提出了基于 GRU 和一维卷积 (CNN-1D) 结合注意力机制的预测模型。首先采用 CNN-1D 对历史数据进行卷积操作，提取数据特征并实现对历史数据的压缩，从而降低后续处理的复杂度；然后再使用 GRU 模型进行训练，提取序列的时序特征；最后采用注意力机制，为 GRU 每一步的输出设定一个合理的权重，修正 GRU 输出结果，从而进一步提高长期预测的精度。

(3) 局域海表温预测建模。首先将局域同一时刻的数据看作是一张通道数为 1 的二维图像。针对短期、中期海表温预测，采用 CNN-2D 结合 LSTM 进行预测。CNN-2D 来提取观测点与观测点之间数据的空间特征，然后使用 LSTM 学习数据的时序特征，最后进行训练预测；针对长期预测，提出基于 CNN-2D 和 LSTM 融合注意力机制的模型，学习到数据的时空特征以后，通过注意力机制来对结果进行修正，从而减小预测的误差。本文在局域海表温预测中是通过自定义一个 Keras 层来实现注意力机制的。

1.4 本文组织结构

根据本文的研究内容，将本文分六章，其具体内容如下：

第 1 章是绪论。首先介绍本文的研究背景和意义。然后针对研究现状，介绍了传统时间序列预测方法、基于机器学习的时间序列预测方法、基于深度学习的时间序列预测方法等内容，然后给出目前仍然存在的一些问题和部分未充分考虑的因素。最后，根据存在的问题和不足，提出了本文的研究内容。

第 2 章是相关理论与技术。首先是介绍了时间序列预测相关概念和数学模型。然后介绍了机器学习中的提升算法 (Boosting)，主要介绍 XGBoost、AdaBoost、CatBoost 三种提升算法的相关理论和详细的推导公式。接着介绍了卷积神经网络，主要介绍卷积神经网络的组成部件。循环神经网络主要介绍 LSTM 和 GRU 的模型单元。本文使用注意力机制来尝试减少较长时间下模型的误差，所以在第 2 章的最后部分介绍了注意力机制相关的理论。

第 3 章是单点海表温预测。提出了一种基于 XGBoost 的单点短期海表温预测模型。该模型在数据处理上考虑“邻域点”对预测点的影响，利用 XGBoost 模型的并行优化、损失精确、防止过拟合等优点，实现了对短期海表温的准确预测；在此基础上，提出了基于 XGBoost 结合 LSTM 的中期海表温预测模型，该模型利用 XGBoost 来提取观测点数据的时间特征，并结合周围邻域点的历史数据，作为 LSTM 模型的输入，充分利用 LSTM 模型具有提取长短期特征和多步输出的优势实现了海表温的中期预测。针对单点长期海表温预测，提出了基于 GRU 和一维卷积（CNN-1D）结合注意力机制的预测模型。首先采用 CNN-1D 对历史数据进行卷积操作，提取数据特征并实现对历史数据的压缩，从而降低后续处理的复杂度；然后再使用 GRU 模型进行训练，提取序列的时序特征；最后采用注意力机制，为 GRU 每一步的输出设定一个合理的权重，修正 GRU 输出结果，从而进一步提高长期预测的精度。

第 4 章是局域海表温预测。首先将局域同一时刻的数据看作是一张通道数为 1 的二维图像。针对短期、中期海表温预测，采用 CNN-2D 结合 LSTM 进行预测。CNN-2D 来提取观测点与观测点之间数据的空间特征，然后使用 LSTM 学习数据的时序特征，最后进行训练预测；针对长期预测，提出基于 CNN-2D 和 LSTM 融合注意力机制的模型，学习到数据的时空特征以后，通过注意力机制来对结果进行修正，从而减小预测的误差。

第 5 章是实验与分析。实验分为两大类，一类是单点海表温预测实验，另一类是局域海表温预测实验。对于单点海表温预测，分为短期、中期、长期三种。每一种都有各自的模型和对比模型；对于局域海表温预测，分为短期、中期、长期三种，每种都有各自的模型和对比模型。所有的实验都会在两个数据集上进行，一个是渤海日平均海表温数据集，另一个是南海日平均海表温数据集。

第 6 章是总结与展望。首先对本文的工作进行整体性的总结和概括，然后指出本文工作中仍然需要改进的地方，并指出进一步的研究工作。

第2章 相关理论与技术

本文研究单点海表温时间序列预测和局域海表温时间序列预测问题，主要是按预测步长划分为短期预测、中期预测、长期预测，每种预测提出对应的方法。针对本文的研究内容，本章对涉及到的相关理论和技术进行介绍，主要包括时间序列预测、Boosting、卷积神经网络以及循环神经网络等相关理论知识。

2.1 时间序列预测方法

时间序列预测是一个规划工具，主要依靠过去和现在的数据，分析两者之间的关系，基于已经掌握的知识经验，运用合理的数学模型，对未来将要发生的事情做出相应的判断，人们通常使用一种或多种技术预计将在未来几个月或几年内发生的事情，传统的方法如 Box-Jenkins 模型^[32]、Delphi 方法^[33]、指数平滑、移动平均模型、回归分析和趋势预测^[34]。传统的时间序列预测技术主要分为两类，一类是根据历史数据特征预测未来数据的变化特征，是一种确定型的预测；另一类将预测变量看做随机过程，建立模型进而预测，是随机型时间序列预测方法。传统时间序列预测技术主要有自回归模型（AR）、移动平均模型（MA）、自回归移动平均模型（ARMA）以及这些模型的变异模型组成。

2.2 Boosting

Boosting^[35]是基于Kearns和Valiant提出的问题：一组弱学习器能创建一个强大的学习器吗？Boosting是一种能够将弱学习器转化为强学习器，从而提升各种学习算法的方法。理论上，Boosting可以显著减小弱学习器的偏差，这些弱学习器的效果只是稍微优于随机猜测，比如小决策树——数据加权模型。然后Boosting在运行时将更多的权重赋值给早期训练中错误最多的数据集，最终预测的产生主要是两种方式：一种是通过结合加权多数投票（分类）产生最终预测；另一种是通过加权求和（回归）以产生最终预测。Boosting的每个模型都是单独运行，每个模型决定下一个模型要关注的特征，最后在不偏向任何模型的前提下聚合输出结果。下面介绍Boosting中的三种算法：XGBoost、AdaBoost、CatBoost。

2.2.1 XGBoost

XGBoost^[36]算法思想：不断地添加树，树的生长依靠不断地进行特征分裂。每次添加一个树，其实是学习一个新的函数，去拟合上次预测的残差，训练完成后，我们得到

了 k 棵树。我们根据这个样本的特征预测样本值时，每个叶子节点对应一个分数，其实每个特征在每棵树中会落到对应的一个叶子节点，最后只需要将每棵树对应的分数加起来就是该样本的预测值。XGBoost是陈天奇于2016年提出的在梯度提升树（GBDT）^[37]基础上的改进算法，是以线性分类器或者分类回归树（CART）^[25]作为基学习器的梯度提升（Gradient Boosting）算法。相比与其他Boosting集成学习算法，XGBoost的理论优势在于：（a）对损失函数引入正则化项，控制了模型复杂度，防止过拟合；（b）对损失函数进行二阶泰勒展开，提高了收敛速度与收敛精度；（c）引入列抽样，进一步提高计算速度并防止过拟合。

假设样本集合为 $D = \{(x_i, y_i)\} (|D| = n, x_i \in R^m, y_i \in R)$ ，其中样本数目为 n ，样本维度为 m 。以 K 个加性函数构成的树融合模型的预测值 \hat{y}_i 的计算见公式（2.1）。

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), f_k \in \mathcal{F} \quad (2.1)$$

其中 \mathcal{F} 的计算见公式（2.2），

$$\mathcal{F} = \{f(x) = \omega_{q(x)}\} (q: R^m \rightarrow T, \omega \in R^T) \quad (2.2)$$

\mathcal{F} 是回归树所在空间，公式（2.2）中 q 为样本到叶节点索引的映射； T 为回归树的叶节点数目，每个 f_k 都对应于独立的树结构 q 与叶节点权重 ω 。

回归树的叶节点分数是连续的，用 ω_i 表示每个叶节点的分数。对于给定的一个样本 i ，如图2.1所示，其预测值 $\hat{y} = \sum_{j=1}^K \omega_{ij}$ ，即为其在每个树结构中相应叶节点的分数求和。

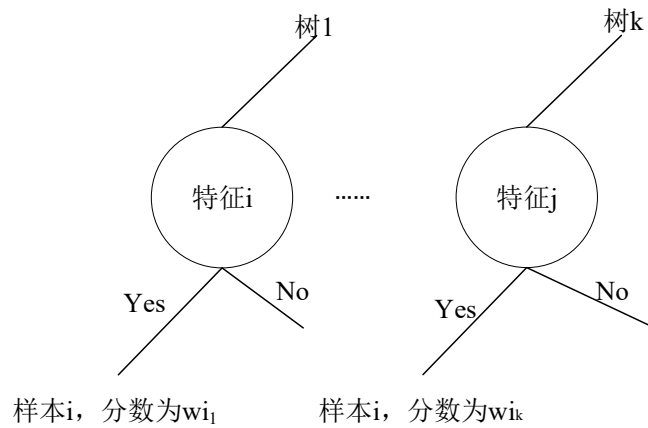


图 2.1 样本预测值图解

Fig. 2.1 Sample predictive value diagram

为了学习出公式（2.1）中的函数，XGBoost中构造正则目标函数（损失函数） $\mathcal{L}(\phi)$ 如公式（2.3）所示。

$$\mathcal{L}(\phi) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \Omega(f_k) \quad (2.3)$$

公式(2.3)中, l 为连续凸函数, 用来度量预测值 \hat{y}_i 与实际值 y_i 之间的差距; $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\|\omega\|^2$ 为惩罚项, 用于控制模型复杂度, 防止模型过拟合。

公式(2.3)中的参数无法通过传统欧氏空间中的方法进行估计, 因此采用加法方式, 通过公式(2.1)将预测值进行递推展开。具体而言, 设 \hat{y}_i^t 为第 t 次迭代中第 i 个样本点的预测值, 损失函数 \mathcal{L}^t 的计算如公式(2.4)所示。

$$\mathcal{L}^t = \sum_{i=1}^n l(y_i, \hat{y}_i^{t-1} + f_t(x_i)) + \Omega(f_t) \quad (2.4)$$

同时, XGBoost对目标函数进行了二阶泰勒展开, 加快了收敛速度, 提高了收敛精度。先进行二阶泰勒展开然后去掉常数项, 则公式(2.4)可改写为公式(2.5)。

$$\mathcal{L}^t \approx \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{t-1}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad (2.5)$$

公式(2.5)中, $g_i = \partial_{\hat{y}_i^{t-1}} l(y_i, \hat{y}_i^{t-1})$; $h_i = \partial_{\hat{y}_i^{t-1}}^2 l(y_i, \hat{y}_i^{t-1})$ 分别为一阶, 二阶梯度估计量。公式(2.5)去掉常数项后结果如公式(2.6)所示。

$$\tilde{\mathcal{L}}^t \approx \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad (2.6)$$

定义 $I_j = \{i | q(x_i) = j\}$ 为叶子节点 j 上的样本点集合, 带入正则化项, 公式(2.6)可改写为公式(2.7)。

$$\begin{aligned} \tilde{\mathcal{L}}^t &= \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \\ &= \sum_{i=1}^n \left[\left(\sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] + \gamma T \end{aligned} \quad (2.7)$$

考虑一个树结构为 $q(x)$, 公式(2.7)对 ω_j 求导可得极值 ω_j^* , 极值 ω_j^* 的计算如公式(2.8)所示。

$$\omega_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (2.8)$$

对应的损失函数值 $\tilde{\mathcal{L}}(q)$, 也是用于度量 $q(x)$ 好坏的分数函数, 它的计算如公式(2.9)所示。

$$\tilde{\mathcal{L}}(q) = - \frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (2.9)$$

分数函数值 $\tilde{\mathcal{L}}(q)$ 越小代表当前树的结构越好, 通过贪婪算法从一个叶节点开始迭代地向树结构添加分支, 来寻求最优的树结构 q 。在回归树中, 一般都采用二分裂的方式, 对将当前某叶节点向下进一步分为左右两个叶节点。设分裂后左右节点所包含的样本点分别为 I_L 和 I_R , $I = I_L \cup I_R$, 则根据公式(2.8)可得到分裂后的损失减少为 \mathcal{L}_{split} 。

$$\mathcal{L}_{split} = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma \quad (2.10)$$

由此便可判断当前树结构是否需要进一步分裂。

XGBoost不同于其他Boosting算法的功能是，在行抽样的基础上（即随机抽取一定比例的样本点进行当次学习），XGBoost进一步添加了列抽样，即在每次迭代学习中，考虑了对样本的特征进行抽样。这种方式能够进一步防止过拟合，且提高了计算速度。

2.2.2 AdaBoost

AdaBoost算法^[39]基本思想：初始时给每个训练样本设置相同的权重，然后使用一个弱分类器对训练集进行多次迭代训练，每次会根据训练结果为训练样本重新赋予权重，然后继续训练。为了提高正确率，算法在每一次训练后，会对不能正确分类的样本赋予较大的权重，对正确分类的样本赋予较小的权重，这样在下次迭代训练时重点学习那些不能正确分类的样本。每次迭代后都会得到一个预测模型，每个预测模型根据其预测准确率也赋予一个权重。迭代结束后，产生一个预测模型序列。最后对预测模型采用加权平均法得到最终的预测模型。

AdaBoost算法更重视预测性能好的预测器和预测误差较大的样本，然后降低训练效果差的预测器的权重，提高训练效果强的预测器的权重；它通过综合多个弱预测算法构建高精度算法，也提高了泛化能力。注意，每个样本在算法初始化时有相同的权重。具体步骤如下：

(1) 初始化权值 $X_t(i)$ 如公式（2.11）所示。

$$X_t(i) = \frac{1}{n} \quad (2.11)$$

训练弱预测器 $z_t(x)$ ，公式（2.11）中 $X_t(i)$ 表示在第 t 次迭代中样本的权重， n 为样本总数。

(2) 计算预测器 $z_t(x)$ 在样本下的误差 ε_t 如公式（2.12）所示。

$$\varepsilon_t = \frac{1}{n} \sum_{i=1}^n \varepsilon_i \quad (2.12)$$

(3) 更新样本权重 $X_t(i)$ ，更新的计算如公式（2.13）所示。

$$X_t(i) = X_{t-1}(i) \frac{\beta_t^{\varepsilon_i}}{Q_t} \quad (2.13)$$

分类器权重 $X_t(i)$ 计算如公式（2.14）所示。

$$W_t = \ln\left(\frac{1}{\beta_t}\right) \quad (2.14)$$

公式(2.14)中: $\beta_t = \varepsilon_t / (1 - \varepsilon_t)$, Q_t 是归一化因子。

(4) 重复迭代, 直到迭代次数为 M 结束。

(5) 得到强预测器 $H(x)$:

$$H(x) = \sum_{t=1}^M W_t h_t(x) \quad (2.15)$$

2.2.3 CatBoost

Boosting策略的另一种代表是CatBoost^[40], 它和XGBoost都属于GBDT类的算法。CatBoost在GBDT的基础上主要做了两点改进: 处理标称属性和解决预测偏移的问题, 从而减少过拟合的发生。GBDT算法是通过一组分类器的串行迭代, 最终得到一个强学习器, 以此来进行更高精度的分类。它使用了前向分布算法, 弱学习器使用分类回归树(CART)。

假设强学习器 $F^{t-1}(x)$ 是前一轮迭代得到的, 其损失函数是 $L(y, F^{t-1}(x))$, 则本轮迭代的目的是找到一个CART回归树模型的弱学习器 h^t , 让本轮的损失函数最小。公式(2.16)表示的是本轮迭代的目标函数 h^t 。

$$h^t = \arg \min_{h \in H} EL((y, F^{t-1}(x) + h(x))) \quad (2.16)$$

GBDT使用损失函数的负梯度来拟合每一轮的损失的近似值, 公式(2.17)中 $g^t(x, y)$ 表示的是上述梯度。

$$g^t(x, y) = \frac{\partial L(y, s)}{\partial s} \Big|_{s=F^{t-1}(x)} \quad (2.17)$$

通常用公式(2.18)近似拟合 h^t 。

$$h^t = \arg \min_{h \in H} E(-g^t(x, y) - h(x))^2 \quad (2.18)$$

最终的到本轮的强学习器 $F^t(x)$ 如公式(2.19)所示:

$$F^t(x) = F^{t-1}(x) + h^t \quad (2.19)$$

CatBoost的改进: 标称属性的一般处理方法是独热编码, 但是会出现过拟合的问题, CatBoost在处理标称属性时使用了更有效的策略, 可以减少过拟合的发生。为训练集生成一个随机序列, 假设原来的顺序是 $\sigma = (\sigma_1, \dots, \sigma_n)$ 。从 σ_1 到 σ_n 一次遍历随机序列, 用遍历到的前 p 个记录计算标称特征的数值。 $\sigma_{p,k}$ 可用公式(2.20)的 $\sigma_{p,k}$ 替换。

$$\sigma_{p,k} = \frac{\sum_{j=1}^p [x_{j,k} = x_{i,k}] \cdot Y_i + a \cdot P}{\sum_{j=1}^n [x_{j,k} = x_{i,k}] + a} \quad (2.20)$$

这里添加了一个先验值 P 和参数 $a>0$ 。这是一种常见做法，它有助于减少从低频类别中获得的噪音。

传统算法中梯度估计方法被CatBoost通过采用排序提升（Ordered Boosting）的方式替代，从而减轻梯度估计的偏差，提高模型的泛化能力。

2.3 卷积神经网络

近几年，作为一种高效识别的方法，卷积神经网络发展较快，且引起广泛重视。卷积神经网络是深度学习代表算法之一，它是一种前馈神经网络，具备深度结构且包含卷积计算。卷积神经网络具有表征学习能力，能够按其阶层结构对输入信息进行平移不变分类，因此也被称为“平移不变人工神经网络”。其结构主要包括卷积层、池化层、全连接层。在此用三维张量 $x^l \in \mathbb{R}^{H^l \times W^l \times D^l}$ 表示卷积神经网络第 l 层的输入，用三元组 (i^l, j^l, d^l) 来指示该张量对应第 i^l 行，第 j^l 列，第 d^l 通道（Channel）位置的元素，其中 $0 \leq i^l < H^l, 0 \leq j^l < W^l, 0 \leq d^l < D^l$ 。卷积神经网络第 l 层输入 x^l 示意图如图2.2所示。

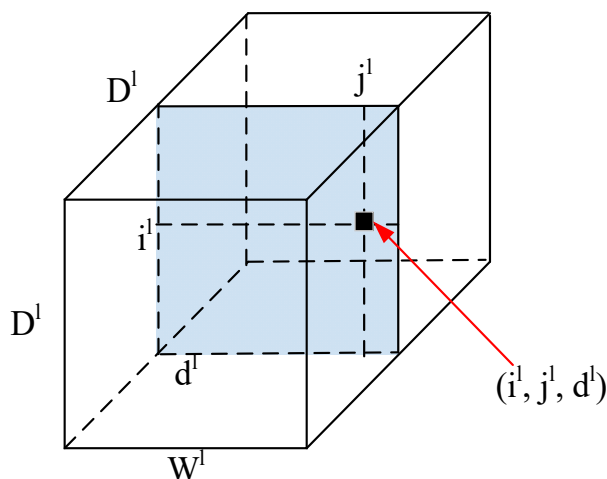


图 2.2 卷积神经网络第 l 层输入 x^l 示意图

Fig. 2.2 Schematic diagram of x^l input at layer l of convolutional neural network

2.3.1 卷积层

完成卷积神经网络中的“卷积”基础操作的层被称为卷积层（Convolution Layer）。下面以 $d^l = 1$ （通道数为1）的情形为例介绍二维场景的卷积操作。

假设输入图像（输入数据）为如图2.3（b）的 5×5 矩阵，其对应的卷积核（亦称卷积参数，Convolution Kernel或Convolution Filter）为一个 3×3 的矩阵。同时，这里设定每做一次卷积，卷积核移动一个像素位置，即卷积步长（Stride）为1。

图2.3为二维场景下的卷积核与输入数据。图2.3（a）为一个 3×3 的卷积核，图2.3（b）为 5×5 的输入数据。

| | | |
|---|---|---|
| 1 | 0 | 1 |
| 0 | 1 | 0 |
| 1 | 0 | 1 |

(a) 卷积核

| | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 6 | 7 | 8 | 9 | 0 |
| 9 | 8 | 7 | 6 | 5 |
| 4 | 3 | 2 | 1 | 0 |
| 1 | 2 | 3 | 4 | 5 |

(b) 输入数据

图 2.3 卷积核和输入数据

Fig. 2.3 Convolution kernel and input data

图2.4展示了卷积的详细过程。从图像(0, 0)像素开始第一次卷积操作，操作过程：卷积核中参数与对应位置图像像素逐位相乘，得出的结果全部累加求和，这个和就是这次卷积的结果，即 $1 \times 1 + 2 \times 0 + 3 \times 1 + 6 \times 0 + 7 \times 1 + 8 \times 0 + 9 \times 1 + 8 \times 0 + 7 \times 1 = 1 + 3 + 7 + 9 + 7 = 27$ 如图2.4 (a) 所示。卷积步长为1时，如图2.4 (b) 至图2.4 (d) 所示，卷积核按照步长大小在输入图像上从左至右自上而下依次将卷积操作进行下去，最终输出 3×3 大小的卷积特征，同时该结果将作为下一层操作的输入。

| | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 6 | 7 | 8 | 9 | 0 |
| 9 | 8 | 7 | 6 | 5 |
| 4 | 3 | 2 | 1 | 0 |
| 1 | 2 | 3 | 4 | 5 |

第1次卷积操作

| | | |
|----|--|--|
| 27 | | |
| | | |
| | | |

卷积后结果 (卷积特征)

(a) 第一次卷积操作及得到的特征

| | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 6 | 7 | 8 | 9 | 0 |
| 9 | 8 | 7 | 6 | 5 |
| 4 | 3 | 2 | 1 | 0 |
| 1 | 2 | 3 | 4 | 5 |

第2次卷积操作

| | | |
|----|----|--|
| 27 | 28 | |
| | | |
| | | |

卷积后结果 (卷积特征)

(b) 第二次卷积操作及得到的特征

| | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 6 | 7 | 8 | 9 | 0 |
| 9 | 8 | 7 | 6 | 5 |
| 4 | 3 | 2 | 1 | 0 |
| 1 | 2 | 3 | 4 | 5 |

第3次卷积操作

| | | |
|----|----|----|
| 27 | 28 | 29 |
| | | |
| | | |

卷积后结果 (卷积特征)

(c) 第三次卷积操作及得到的特征

| | | | | |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |
| 6 | 7 | 8 | 9 | 0 |
| 9 | 8 | 7 | 6 | 5 |
| 4 | 3 | 2 | 1 | 0 |
| 1 | 2 | 3 | 4 | 5 |

第9次卷积操作

| | | |
|----|----|----|
| 27 | 28 | 29 |
| 28 | 27 | 16 |
| 23 | 22 | 21 |

卷积后结果 (卷积特征)

(d) 第九次卷积操作及得到的特征

图 2.4 卷积操作示例

Fig. 2.4 Example of convolution operation

与之类似，若三维情形下的卷积层 l 的输入张量为 $x^l \in \mathbb{R}^{H^l \times W^l \times D^l}$ ，该层卷积核为 $f^l \in \mathbb{R}^{H \times W \times D^l}$ 。三维输入时卷积操作实际只是将二维卷积扩展到了对应位置的所有通道上（即 D^l ），最终将一次卷积所处理的所有 HWD^l 个元素求和作为该位置卷积结果。三维

场景下的卷积核与输出结果如图2.5所示。图2.5 (a) 卷积核大小为 $4 \times 4 \times 3$ ，图2.5 (b) 是在该位置卷积操作后得到的 $1 \times 1 \times 1$ 的输出结果。

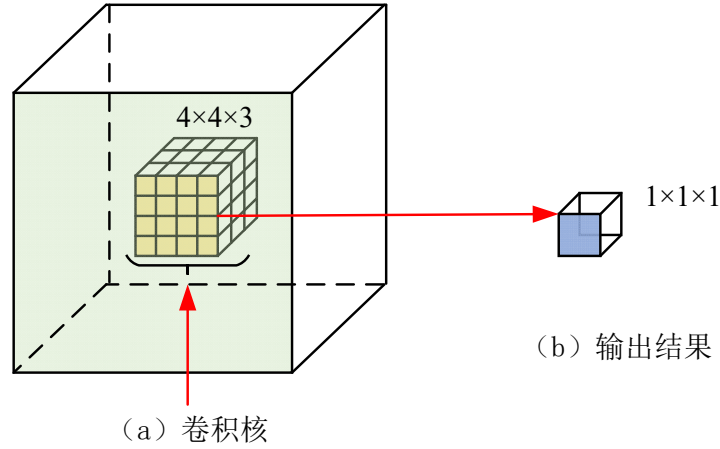


图 2.5 三维场景下的卷积核与输出结果

Fig. 2.5 Convolution kernel and output in 3d scene

进一步地，若类似 f^l 这样的卷积核有 D 个，则在同一个位置上可得到 $1 \times 1 \times 1 \times D$ 维度的卷积输出，而 D 即为第 $l+1$ 层特征 x^{l+1} 的通道数 D^{l+1} 。

三维卷积操作的计算过程如公式 (2.21) 所示。

$$y_{i,j,d}^{l+1} = \sum_{i=0}^H \sum_{j=0}^W \sum_{d^l=0}^{D^l} f_{i,j,d^l,d} \times x_{i^{l+1}+i,j^{l+1}+j,d^l}^l \quad (2.21)$$

其中 (i^{l+1}, j^{l+1}) 为卷积结果的位置坐标，它应满足公式 (2.22) 和 (2.23)。

$$0 \leq x^{l+1} < H^l - H + 1 = H^{l+1} \quad (2.22)$$

$$0 \leq j^{l+1} < W^l - W + 1 = W^{l+1} \quad (2.23)$$

需指出的是，公式 (2.21) 中的 $f_{i,j,d^l,d}$ 可视为学习到的权重 (Weight)，可以发现该项权重对不同位置的所有输入都是相同的，这便是卷积层“权值共享” (Weight Sharing) 特性。除此之外，通常还会在 $y_{i,j,d}^{l+1}$ ， j^{l+1} ， d 上加入偏置项 (Bias Term) b_d 。在误差反向传播时可针对该层权重和偏置项分别设置随机梯度下降的学习率。当然根据实际问题需要，也可以将某层偏置项设置为全0，或将学习率设置为0，以起到固定该层偏置或权重的作用。此外，卷积操作中有两个重要的超参数 (Hyper Parameters)：卷积核大小 (Filter Size) 和卷积步长 (Stride)。

2.3.2 池化层

接上节，第 l 层池化核可表示为 $p^l \in \mathbb{R}^{H \times W \times D^l}$ 。平均值 (最大值) 池化在每次操作时，将池化核覆盖区域中所有值的平均值 (最大值) 作为池化结果，平均值池化计算过程见公式 (2.24)，最大值池化计算过程见公式 (2.25)。

$$Avg: y_i^{l+1}, j^{l+1}, d = \frac{1}{HW} \sum_{0 \leq i < H, 0 \leq j < W} x_{i^{l+1} \times H + i, j^{l+1} \times W + j, d^l} \quad (2.24)$$

$$Max: y_i^{l+1}, j^{l+1}, d = \max_{0 \leq i < H, 0 \leq j < W} x_{i^{l+1} \times H + i, j^{l+1} \times W + j, d^l} \quad (2.25)$$

其中, $0 \leq i^{l+1} < H^{l+1}, 0 \leq j^{l+1} < W^{l+1}, 0 \leq d < D^{l+1} = D^l$ 。图2.6是最大值池化操作示例, 池化核大小为 2×2 、步长为1。

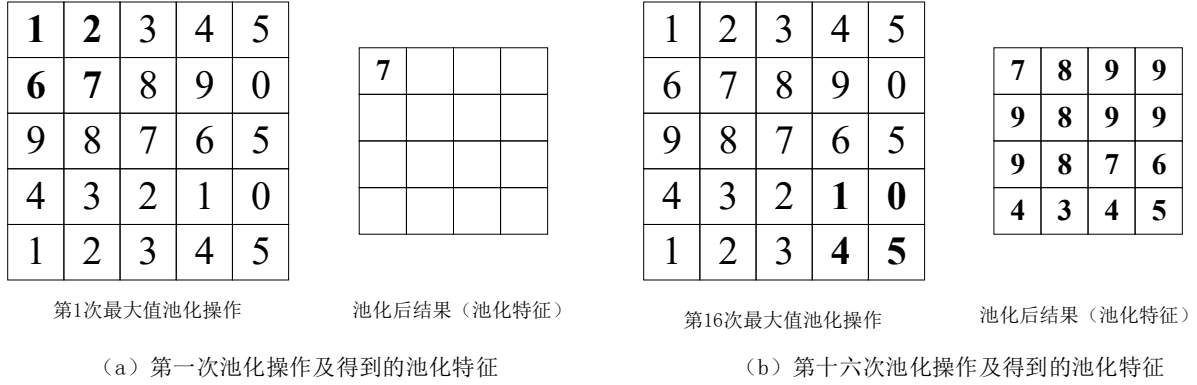


图 2.6 最大值池化操作示例

Fig. 2.6 Example of maximum pool operation

除了最常用的上述两种池化操作外, 随机池化则介于二者之间。随机池化操作非常简单, 只需对输入数据中的元素按照一定概率值大小随机选择, 这和最大值池化完全不同, 因最大池化只取覆盖区域中的最大值。从随机池化来看, 元素值小的相应被选中的概率也小, 元素值大的被选中的概率大。可以说, 从整体上来看, 随机池化与平均池化近似; 在局部意义上, 是采用最大池化的准则。

2.3.3 全连接层

卷积神经网络中的全连接层^[41]具有“分类器”的作用。卷积层、池化层等操作可以将原始数据映射到隐层空间, 而全连接层会将已经学习到的特征表示映射到样本的标记空间。在实际使用中, 卷积操作也可实现全连接层的作用, 具体操作为: 卷积核为 1×1 的卷积可替代前层是全连接的全连接层; 而卷积核为 $h \times w$ 的全局卷积可替代前层是卷积层的全连接层, h 和 w 分别为前层卷积输出结果的高和宽。

2.3.4 目标函数

将学习到的特征映射到样本的标记空间, 这是全连接层的作用, 而衡量该预测值与真实样本标记之间的误差是目标函数的作用。在当下的卷积神经网络中, 对于回归问题最为常用的目标函数 ℓ_2 损失函数, 对于分类问题最为常用的目标函数是交叉熵损失函数^[42]。同时, 越来越多针对不同问题特性的目标函数被提出供选择。

2.4 循环神经网络

循环神经网络^[43]是根据人的认知是基于过往的经验和记忆这一观点提出的，也就是一个序列的前面的输出会影响当前的输出。具体表现形式为网络在计算当前输出时，会使用前面的信息，即隐藏层之间的节点不再无连接而是有连接的，并且隐藏层的输入不仅包括输入层的输出还包括上一时刻隐藏层的输出。下面将介绍几种常用的循环神经网络。

2.4.1 RNN 模型

RNN 模型有多种，这里介绍最主流的 RNN 模型结构如图 2.7。

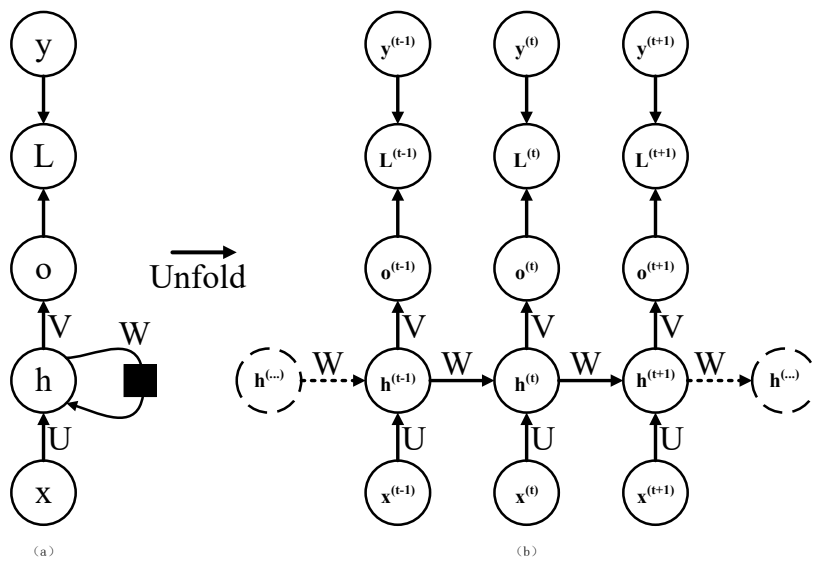


图 2.7 RNN 模型

Fig. 2.7 RNN model

图2.7（a）是没有按时间展开的RNN模型图，图2.7（b）是按时间序列展开的RNN模型图，描述了在序列索引号 t 附近RNN的模型。其中：

1) $x^{(t)}$ 表示在序列索引号为 t 时，训练样本的输入。同样的， $x^{(t-1)}$ 和 $x^{(t+1)}$ 代表在序列索引号 $t-1$ 和 $t+1$ 时训练样本的输入。

2) $h^{(t)}$ 表示在序列索引号为 t 时，模型的隐藏状态。 $h^{(t)}$ 由 $x^{(t)}$ 和 $x^{(t-1)}$ 共同决定。

3) $o^{(t)}$ 表示在序列索引号为 t 时，模型的输出。 $o^{(t)}$ 只由模型当前的隐藏状态 $h^{(t)}$ 决定。

4) $L^{(t)}$ 表示在序列索引号为 t 时，模型的损失函数。

5) $y^{(t)}$ 表示在序列索引号为 t 时，训练样本序列的真实输出。

6) U , W , V 代表三个矩阵，它们是模型的线性关系参数矩阵。与DNN不同的是，在整个RNN网络中它们三个是共享的。也正因为是共享的，所以体现了RNN的模型的“循环反馈”的思想。

如果我们略去 $o^{(t)}$, $L^{(t)}$, $y^{(t)}$, 则RNN模型可简化为图2.8:

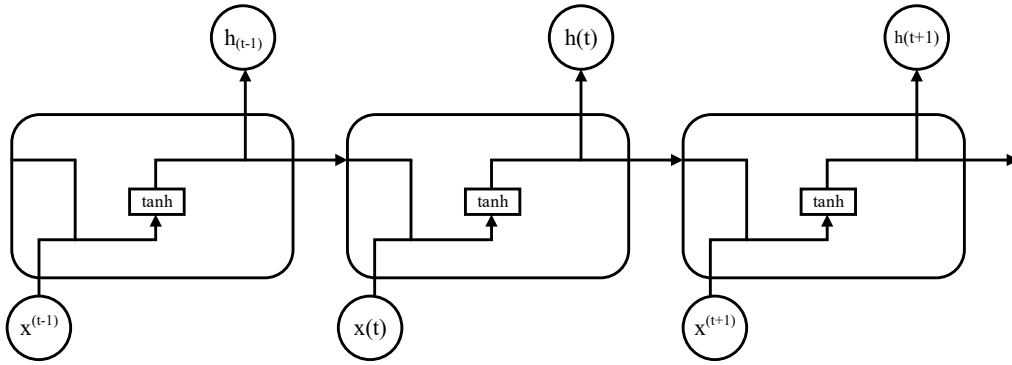


图 2.8 RNN 简化模型

Fig. 2.8 RNN simplifies the model

2.4.2 LSTM 模型

近年来, 各种 RNN 的变种被提出来, 在机器翻译、语音识别以及时间序列预测等领域取得了令人瞩目的成就。传统的 RNN 模型受到了梯度消失的困扰, 导致网络无法建模长距离的依赖, LSTM (Long Short Term Memory) 通过在传统的 RNN 网络中增加了记忆单元和 3 个控制门, 从而解决了 RNN 的梯度消失问题, 使得网络可以建模长距离的依赖。

LSTM 模型单元如图 2.9 所示。

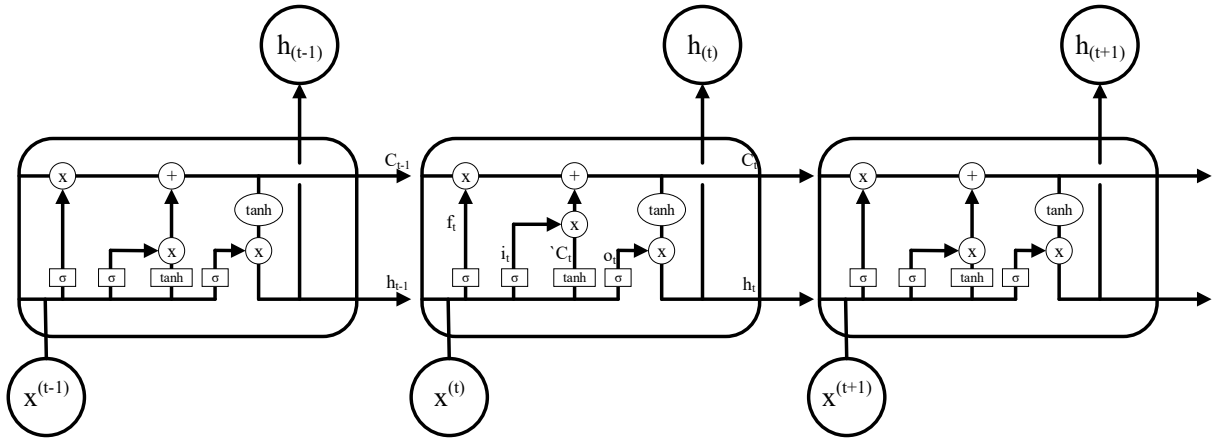


图 2.9 LSTM 模型

Fig. 2.9 LSTM model

LSTM 可以由公式 (2.26) ~ (2.31) 表示:

$$i_t = \text{sigmoid}(W_i h_{t-1} + U_i x_t + b_i) \quad (2.26)$$

$$f_t = \text{sigmoid}(W_f h_{t-1} + U_f x_t + b_f) \quad (2.27)$$

$$o_t = \text{sigmoid}(W_o h_{t-1} + U_o x_t + b_o) \quad (2.28)$$

$$\tilde{c}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c) \quad (2.29)$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (2.30)$$

$$h_t = o_t * \tanh(c_t) \quad (2.31)$$

在公式 (2.26) ~ (2.31) 中由符号 W_i 、 W_o 、 W_f 、 W_c 、 U_i 、 U_o 、 U_f 、 U_c 表示的矩阵以及由符号 b_i 、 b_f 、 b_o 、 b_c 表示的向量是 LSTM 网络所需要训练的参数，向量 h_{t-1} 表示 $t-1$ 时刻 LSTM 的隐藏层输出向量， x_t 表示第 t 时刻的输入向量。

2.4.3 GRU 模型

2014 年提出 GRU，它是基于 LSTM 改进而来，它使用一个单一的更新门来替代忘记门和输入门，同时合并了数据单元状态和隐藏状态，使得模型结构比之于 LSTM 更为简单，其单元内部结构如图 2.10。

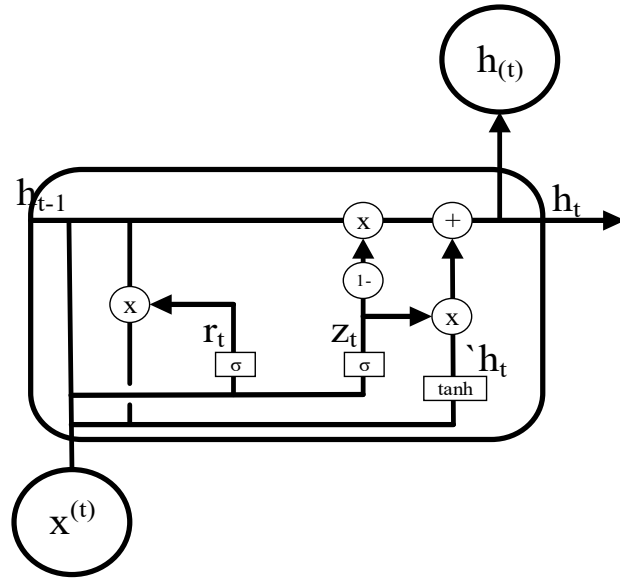


图 2.10 GRU 的一个单元结构图

Fig. 2.10 The framework of a unit of GRU

其各个部分满足关系式如下：（图中 \tilde{h}_t 即公式中 \tilde{h}_t ）

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (2.32)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (2.33)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \quad (2.34)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (2.35)$$

2.5 注意力机制

注意力机制（Attention Mechanism）^[44]可以为模型的不同部分赋予不同的权重，使得模型能够更关注那些与任务更相关的部分，从而提高了模型的建模质量。近年来，大量的基于注意力机制的模型被提出来，它们在机器翻译^[45]、文档分类^[46]、关系分类^[47]以及图像字幕生成^[48]等任务中取得了很好的成绩。

注意力机制可以看成是利用目标中的某个元素作为Query和由原始数据构成的一系列<Key, Value>对一起作用得到一个新的向量。首先通过Query与所有<Key, Value>对中的Key分别利用相似性函数Similarity（见图2.11）求一个相似性分数。然后再对由这些相似性分数构成的向量使用softmax函数将相似性分数归一化，接着使用归一化后的分数作为Value的权重，对所有Value进行加权求和得到最后的目标向量，该过程如图2.11所示。

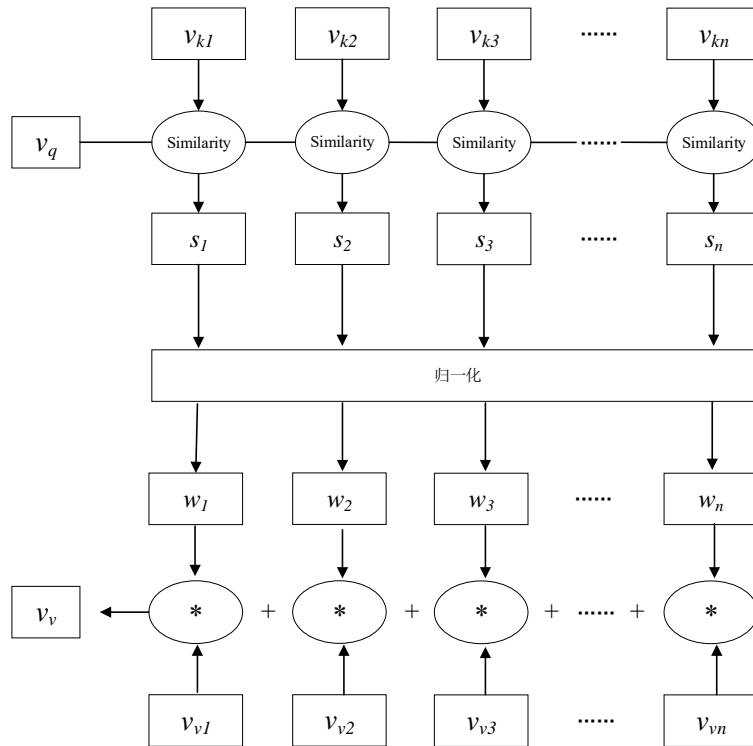


图 2.11 注意力机制的计算过程

Fig. 2.11 The Calculation process of attention mechanism

在图2.11中，向量 v_q 表示对应的Query，向量 v_{ki} 与向量 v_{vi} 是<Key, Value>对中的Key和Value的向量表示。公式（2.36）中列出了一些常见的一些用于求Key与Query的相似性分数的函数，其中 v_q 与 v_k 分别是Query和Key的向量表示，其他变量均是学习参数。

$$f(v_q, v_k) = \begin{cases} v_q^T v_k & \text{dot} \\ v_q^T W_a v_k & \text{general} \\ W_a[v_q; v_k] & \text{concat} \\ v_a^T \tanh(W_a v_q + U_a v_k) & \text{preceptran} \end{cases} \quad (2.36)$$

2.6 本章小结

本章主要介绍了包括了时间序列预测技术、Boosting、卷积神经网络和循环神经网络以及注意力机制等相关理论知识。在后续的章节将基于本章介绍的这些技术和方法进行改进或扩展，从而完成对海表温预测的研究。

第3章 单点海表温预测

针对海洋单观测点（单点）海表温的预测问题，本章首先介绍基于 XGBoost 建模的方法，然后提出基于 XGBoost 结合 LSTM 的预测模型，最后提出基于 GRU 和 CNN-1D 结合注意力机制的预测模型。

3.1 问题提出

近年来全球气候变暖逐渐加剧，而海洋是地球上最大的储热体，它的温度变化也从侧面反映出全球气候的变化。海表温是海洋重要物理特征，也是影响海洋和大气运动的重要因素。海表温的预报在气候变化、海洋热量存储、生态环境、国防军事和水产养殖等方面具有重要作用。针对海表温的预报传统上是通过统计学来进行建模预测，但在海洋温度场这样的混沌系统中，要减小预测误差，就需要建立更加复杂的物理方程，但目前并没有掌握温度场的所有规律，也无法从中总结出更适合的物理公式。所以现在尝试把温度数据单独考虑，无需了解影响海温变化的因素和机制，而是直接从所有数据中去发掘潜在的变化规律，所以相对于要建立复杂物理学公式的传统预测来说要简单许多。传统的时间序列预测技术主要分为两类，一类是根据历史数据特征预测未来数据的变化特征，是一种确定型的预测；另一类将预测变量看做随机过程，建立模型进而预测，是随机型时间序列预测方法。传统时间序列预测技术主要有自回归模型（AR）、移动平均模型（MA）、自回归移动平均模型（ARMA）以及这些模型的变异模型组成。虽然传统的时间序列预测方法理论成熟、使用简单，但不适合处理非线性数据。海表温数据是非线性的，但传统非线性模型在海表温预测上表现不佳，且只能进行单步预测。基于机器学习的预测方法虽然在处理非线性数据上的表现比传统时间序列方法更好，但仍然是单步预测。基于深度学习的预测方法可以进行多步预测，所以本文使用深度学习的方法来解决多步预测的问题。

由于海表面是一个平面，直接对平面进行研究较为困难，所以选取平面上某点进行研究，通过某点历史数据，提取出更有效的数据特征进行建模，从一个点的研究开始，后期再扩展到一个面的分析和预测。针对单点海表温预测问题，本文将其分为短期、中期、长期三类，对每一类提出适合的模型，本章总体研究框架如图 3.1 所示。本章首先介绍数据的选取和处理，然后依次介绍基于 XGBoost 的短期海表温预测模型、基于 XGBoost 结合 LSTM 的中期海表温预测、基于 GRU 和 CNN-1D 结合注意力机制的长期海表温预测。

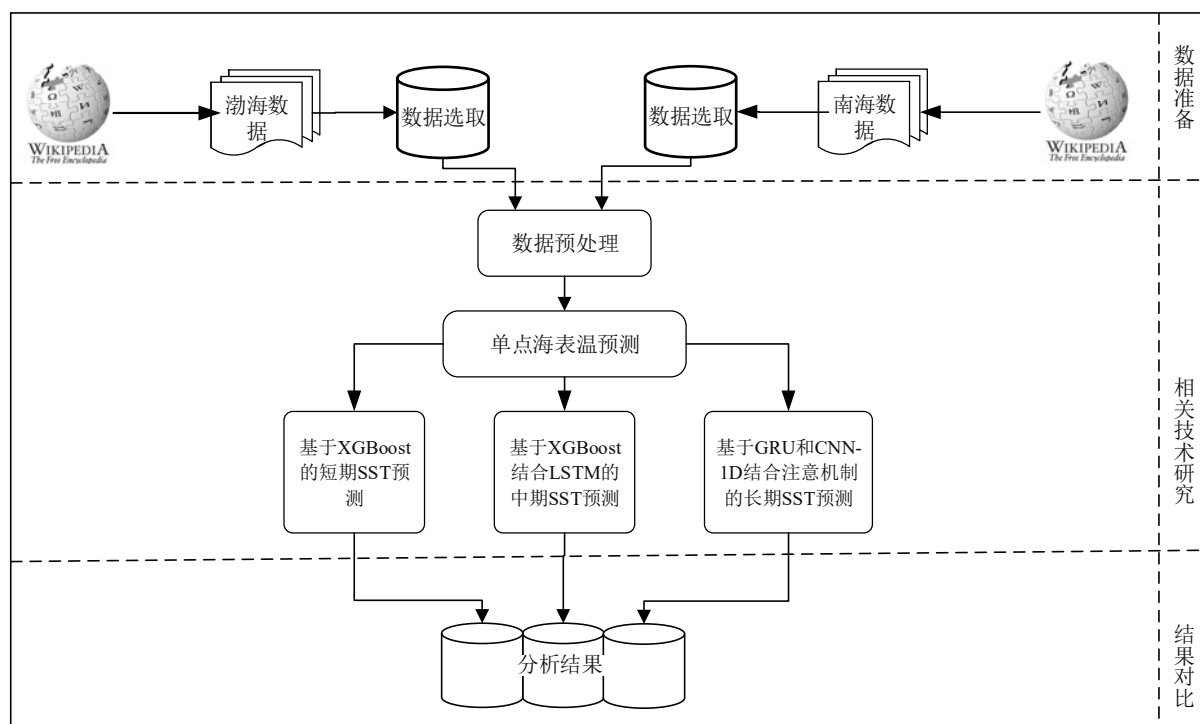


图 3.1 总体研究框架

Fig. 3.1 The overall research framework

3.2 数据选取及预处理

本章研究的是单点海表温预测。所使用的数据集是 NOAA/OAR/ESRL PSD 提供的高分辨率海温数据，该数据来自美国科罗拉多州波尔德市的 NOAA/OAR/ESRL PSD 的网站 <http://www.esrl.noaa.gov/psd/>。这个数据集包含 SST 日平均值、SST 周平均值、SST 月平均值，时间从 1981 年 9 月 1 日开始至今，覆盖全球海洋范围 98.875S~98.875N，0.125E~359.875E，它是 $0.25^{\circ}\text{C} \times 0.25^{\circ}\text{C}$ 的网格形式。本文所有实验所采用的数据是从以上数据中选取的我国渤海和南海 SST 日平均数据。时间范围从 1982 年 1 月 1 日到 2018 年 12 月 31 日，总共 13514 天。本文将 2018 年的数据作为测试集，1982~2017 年的数据作为训练集。

3.2.1 数据选取

渤海数据选取经纬度范围为：37.07N~41N，117.35E~121.10E；南海数据选取经纬度范围 4N~21N，105E~118E。

渤海区域是一个 16×15 的区域，其中包含陆地和海洋，海洋部分在图 3.2 中使用“有效”来标注，每个小的区域都有自己的标号，其中绿色的小区域就是选中用来测试的区域，周围蓝色小区域是选取的邻域。图 3.2 最右侧的一列数据和最下侧的一列数据分别是纬度和经度，该数据分布图和实际地区是上下颠倒的。

渤海观测点的选取如图 3.2 所示。

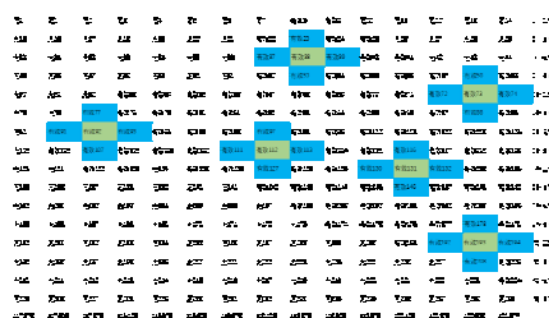


图 3.2 渤海观测点选取

Fig. 3.2 BoHai observation point selection

南海观测点的选取和渤海类似，选取的 6 个观测点如图 3.3 所示。

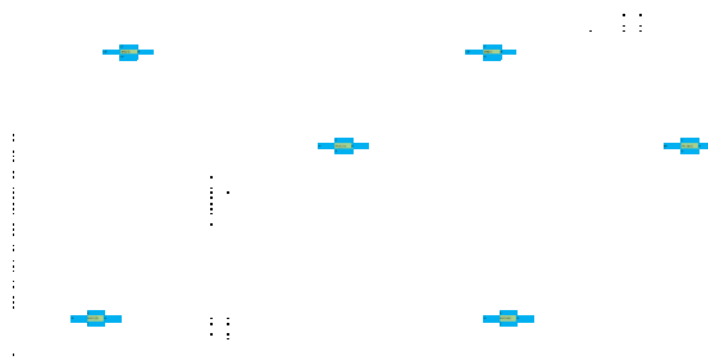


图 3.3 南海观测点选取

Fig. 3.3 NanHai observation point selection

图 3.3 中标有“F”的表示海面，与渤海数据图一样，该图和真实相比，也是上下颠倒的。

3.2.2 数据预处理

由于从网站上获取的数据是标准的.nc 格式，无法直接输入到模型中进行训练和预测，所以本文对该数据集进行了一些预处理操作，主要包括：

文件格式转换。后缀名.nc 的文件又名 NetCDF 文件，NetCDF 全称为 Network Common Data Format（网络通用数据格式）。它适于网络共享的数据描述和编码标准，属于面向数据型数据。但正是它的这种独特的格式，并不适合直接拿来训练和预测，所以本文通过 netCDF4 工具包，将其中的数据提取出来存储为常用的.csv 格式，方便后续的数据处理。

数据平滑处理。分别从渤海、南海中选取 6 个观测点，画出所选观测点在 2018 年的日平均海表温变化趋势图。图 3.4 中可看出渤海数据的变化先降后升再降，数据的变化较大，尤其是七八月份；南海数据相对渤海数据来说，比较平稳。对模型而言，这样

的数据很难获得更好的拟合效果，所以将每个观测点的历史数据使用公式（3.1）进行平滑处理，处理后的效果如图 3.4 中的红线所示，可看出数据比之前平滑了很多，有利于后期的训练和预测。

$$\log 1p = \ln(1 + x) \quad (3.1)$$

实际数据中存在负数，由于最小的负数仍然大于-5，所以这里在进行数据平滑处理之前，先将所有数据加 5，消除负数，然后再进行 $\log 1p$ 操作。

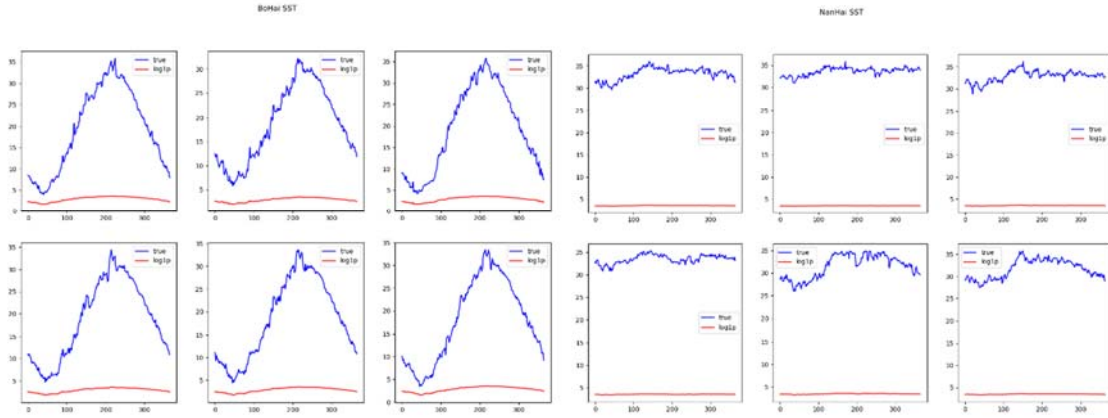


图 3.4 数据平滑处理对比

Fig. 3.4 Data smoothing processing contrast

数据标准化。本文所有标准化操作，都是采用最大最小标准化，将所有训练集数据标准化到[0,1]范围内，最大最小标准化计算见公式（3.2）。

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3.2)$$

其中 x_{\max} 为样本数据最大值， x_{\min} 为样本数据最小值。 x^* 是标准化后的数据。

滑动窗口法拆分数据。本文是以历史海表温度来预测未来海表温度，属于自回归。这里通过滑动窗口法将自回归转换成监督学习，即将前几天历史数据作为特征，而后几天数据作为标签。

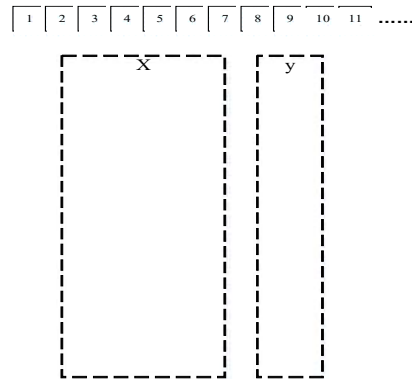


图 3.5 滑动窗口法拆分数据

Fig. 3.5 Split data by sliding window method

邻域点数据处理。在以往的研究中，都是利用某观测点历史数据来预测某观测点未来数据。但针对于海表温来说，某观测点和其邻域点同属于一个温度场，所以它们各自的海表温是相互影响的。单纯以某观测点历史海表温来预测其未来的海表温，这种做法是欠考虑的，所以本文在单点海表温预测时，将邻域点的影响考虑在内。具体做法是：将某观测点的邻域点的历史海表温数据作为新特征，输入到 XGBoost 模型中。数据处理过程如图 3.6 所示。本文选择标号 1（左）、2（下）、3（右）、4（上）四个观测点作为标号 0 的邻域，假设现在要预测 $t+1$ 时刻的标号 0 的数据，那么本文使用的数据是标号 0 的历史数据和 t 时刻邻域点的数据， $X_t = \{0_{t-3}, 0_{t-2}, 0_{t-1}, 0_t, 1_t, 2_t, 3_t, 4_t\}$, $Y_t = 0_{t+1}$ 。

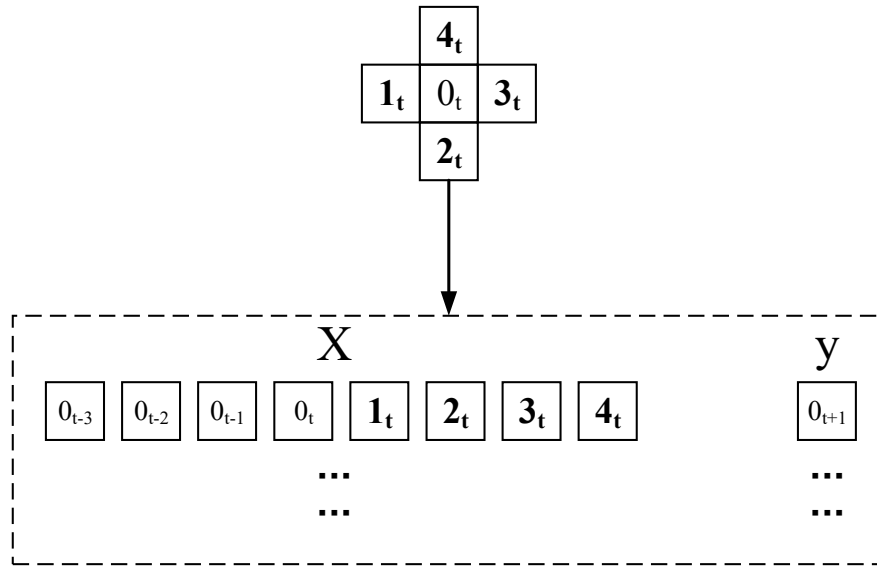


图 3.6 邻域点数据处理

Fig. 3.6 Neighborhood point data processing

3.2.3 时间特征的选择

针对于单点海表温预测，由于渤海海表温受季节影响较大，所以考虑添加时间特征，包括年、季度、月、一年中的第几周、一周中的第几天、一个月的第几天和季节这七个时间特征，通过皮尔森相关系数法来选择出合适的时间特征。皮尔森相关系数是量化特征和响应变量间相关性的方法之一，取值区间为 $[-1, 1]$ ，越接近-1 或 1 表示特征和变量的相关性越大，越接近 0 表示特征和变量的相关性越小。皮尔森相关系数的公式(3.3)：

$$\rho(x, y) = \frac{cov(x, y)}{\sigma_x \sigma_y} \quad (3.3)$$

公式(3.3)中， $cov(x, y)$ 表示数值序列 x 与数值序列 y 的协方差， σ_x 表示数值序列 x 的标准差， σ_y 表示数值序列 y 的标准差。对渤海和南海选出的观测点的数据分别计算七个时间特征和数据的相关性，结果如表 3.1、表 3.2 所示。

表 3.1 时间特征和渤海数据的皮尔森相关系数

Table 3.1 Pearson correlation coefficient between time characteristics and BoHai data

| 时间特征 | 38 | 73 | 92 | 112 | 131 | 193 |
|-------------------|--------------|--------------|--------------|--------------|--------------|--------------|
| year | 0.041 | 0.032 | 0.044 | 0.035 | 0.033 | 0.029 |
| quarter | 0.51 | 0.59 | 0.5 | 0.59 | 0.59 | 0.58 |
| month | 0.49 | 0.57 | 0.48 | 0.57 | 0.57 | 0.57 |
| weekOfYear | 0.49 | 0.57 | 0.48 | 0.56 | 0.56 | 0.56 |
| dayOfWeek | -0.00023 | -0.00072 | -0.00027 | -0.00036 | -0.00071 | -0.00061 |
| dayOfMonth | 0.01 | 0.0086 | 0.0084 | 0.0092 | 0.009 | 0.0096 |
| season | -0.28 | -0.14 | -0.29 | -0.16 | -0.15 | -0.16 |

表 3.2 时间特征和南海数据的皮尔森相关系数

Table 3.2 Pearson correlation coefficient between time characteristics and NanHai data

| 时间特征 | 38 | 73 | 92 | 112 | 131 | 193 |
|-------------------|--------------|--------------|--------------|--------------|-----------------------|--------------|
| year | 0.075 | 0.18 | 0.12 | 0.16 | 0.089 | 0.083 |
| quarter | 0.27 | 0.39 | 0.29 | 0.23 | 0.5 | 0.4 |
| month | 0.26 | 0.38 | 0.28 | 0.22 | 0.49 | 0.39 |
| weekOfYear | 0.25 | 0.37 | 0.27 | 0.22 | 0.48 | 0.39 |
| dayOfWeek | 0.0032 | -0.0024 | 0.0013 | 0.0028 | -7.5×10^{-6} | 0.00054 |
| dayOfMonth | 0.01 | 0.015 | -0.00066 | 0.015 | 0.0073 | 0.014 |
| season | -0.59 | -0.38 | -0.52 | -0.43 | -0.28 | -0.45 |

由表 3.1、表 3.2 可看出,无论是渤海还是南海,在 year、quarter、month、weekOfYear、dayOfWeek、dayOfMonth、season 这 7 个时间特征中,皮尔森相关系数较大的是 quarter、month、weekOfYear、season 四个特征。根据计算结果,从 7 个特征中选择相关性相对较大的特征,本文选择时间特征包括 quarter、month、weekOfYear、season。

3.3 基于 XGBoost 的短期海表温预测

在过去的预测问题中,SVM 的一些特性:基于小样本统计理论基础,泛化能力更强;使用空间核函数,避开高维空间复杂性等,使得 SVM 应用广泛。但是,对于一个实际问题而言,如何根据实际的数据模型选择合适的核函数从而构造 SVM 算法是一个难点。目前比较成熟的核函数及其参数的选择都是人为的、根据经验来选取的,带有一定的随意性,至今没有好的方法来解决核函数的选取问题。本文是通过大量实验来选取一个预测误差较小的核函数作为模型的核函数,将 SVM 模型预测算法作为本文的一个对照。

XGBoost 模型在近年的比赛中出现频次较高，且效果较好，引起了广泛关注。XGBoost 算法在其目标函数中显示地添加了正则化项，减少了过拟合现象地发生；在损失计算过程中，不仅使用到一阶导数，还用到了二阶导数，这样计算的损失更加精确，减少了预测误差；还可以自定义损失，使用非常方便；在特征粒度上实现了并行优化提高了计算效率。但 XGBoost 模型在海表温预测领域应用较少，考虑到 XGBoost 这些特性，在进行多次实验后，本文提出基于 XGBoost 的短期海表温预测模型。

XGBoost 属于 Boosting 算法的一种，其他的 Boosting 算法还有 AdaBoost 和 CatBoost 等，本文将 AdaBoost 和 CatBoost 也作为对照。

针对于短期海表温预测实验，本文会分为两部分，第一部分验证基于 XGBoost 的短期海表温预测模型的有效性，具体做法：仅仅使用某观测点数据来预测某观测点的未来数据，然后输入 XGBoost 模型进行训练预测。第二部分验证邻域点对预测结果的影响作用，具体做法：数据处理上考虑“邻域点”的影响（见图 3.6），然后将处理后的数据输入 XGBoost 进行训练预测。

3.4 基于 XGBoost 结合 LSTM 的中期海表温预测

利用 XGBoost 模型的并行优化、损失精确、防止过拟合等优点，实现了对短期海表温的准确预测，但是原生的 XGBoost 模型只能进行单步输出，无法进行多步输出，这使得 XGBoost 无法完成中期海表温预测的任务。

循环神经网络中的 LSTM 单元可以进行多步输出，如图 3.7 所示：

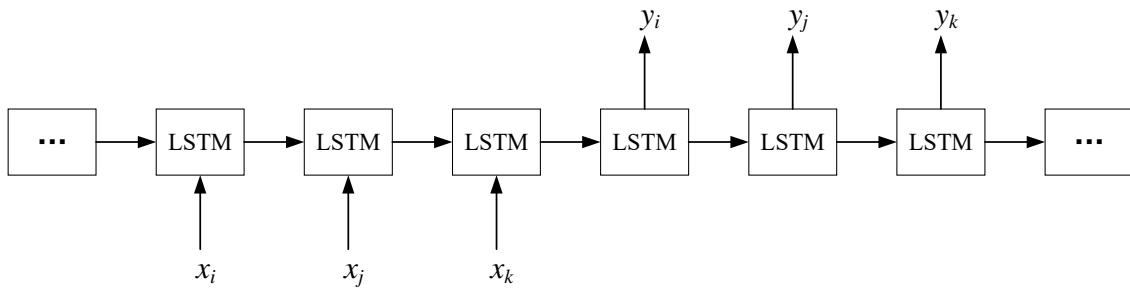


图 3.7 LSTM 多步输出

Fig. 3.7 LSTM multi-step output

LSTM 除了可以进行多步输出，相比 RNN 来说，LSTM 添加了三个门，这种独特的内部结构，使得 LSTM 可以学习更长序列的时序特征。LSTM 模型单元结构如图 3.8。

LSTM 可以由公式 (3.4) ~ (3.9) 表示：

$$i_t = \text{sigmoid}(W_i h_{t-1} + U_i x_t + b_i) \quad (3.4)$$

$$f_t = \text{sigmoid}(W_f h_{t-1} + U_f x_t + b_f) \quad (3.5)$$

$$o_t = \text{sigmoid}(W_o h_{t-1} + U_o x_t + b_o) \quad (3.6)$$

$$\tilde{c}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c) \quad (3.7)$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (3.8)$$

$$h_t = o_t * \tanh(c_t) \quad (3.9)$$

在公式 (3.4) ~ (3.9) 中由符号 W_i 、 W_o 、 W_f 、 W_c 、 U_i 、 U_o 、 U_f 、 U_c 表示的矩阵以及由符号 b_i 、 b_f 、 b_o 、 b_c 表示的向量是 LSTM 网络所需要训练的参数，向量 h_{t-1} 表示 $t-1$ 时刻 LSTM 的隐藏层输出向量， x_t 表示第 t 时刻的输入向量。

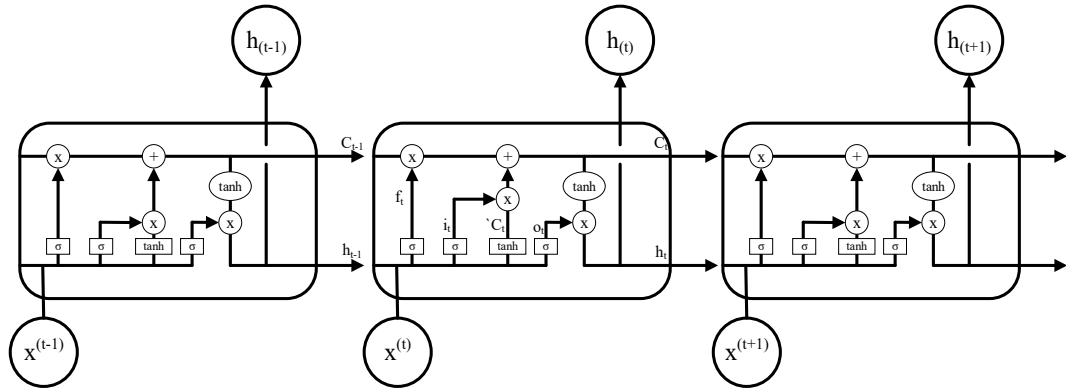


图 3.8 LSTM 模型

Fig. 3.8 LSTM model

本文将 XGBoost 和 LSTM 结合起来，通过 XGBoost 提取数据和时间特征的关系，使用 LSTM 学习数据的时序特征，并进行多步输出，从而减小多步长的情况下的预测误差。基于 XGBoost 结合 LSTM 的单点中期海表温预测模型处理流程如图 3.9 所示。

在图 3.9 中，首先提取出某观测点海表温数据 X_i ， $i \in \{1, 2, 3, 4, 5, 6\}$ ，这里的 1, 2, 3, 4, 5, 6 对于渤海观测点来说，分别对应 38, 73, 92, 112, 131, 193 这 6 个观测点；对于南海观测点来说，分别对应 375, 397, 1324, 1345, 3025, 3050 这 6 个观测点。本文将第 i 个观测点的上、下、左、右四个观测点作为某观测点的邻域点，对应的数据表示为 $X_{near} = \{X_i^{up}, X_i^{down}, X_i^{left}, X_i^{right}\}$ 。将第 i 个观测点的数据 X_i 分为两部分使用，一部分是第 i 个观测点历史数据 X_i 和对应的时间特征数据 $T_i = \{\text{quarter}, \text{weekOfYear}, \text{month}, \text{season}\}$ ，这里的时间特征指的是某观测点海表温数据的 quarter、month、weekOfYear、season 这 4 个特征。在获得了观测点历史数据 X_i 和对应时间特征数据 T_i 后，设置训练集 X_{train} 、 Y_{train} 并使用 XGBoost 模型进行训练得到模型 xgb -

model, 最后将 X_{train} 输入到 $xgb-model$ 中得到结果 XT_i , XT_i 即通过 XGBoost 模型提取的观测点海表温 X_i 和时间特征 T_i 的关系。另一部分是第 i 个观测点数据 X_i 和其邻域点的历史数据 X_{near} , 本文把邻域点的数据作为特征, 然后整合 XGBoost 的输出数据 XT_i 生成一个新的数据集 $D_i = \{X_i, X_{near}, XT_i\}$, 表示第 i 个观测点的新数据集。这部分的数据切分是将 2018 年的数据作为预测集, 2018 年之前的数据作为训练集, 输入到 LSTM 网络中, 最后通过前馈神经网络得到最终第 i 个观测点预测结果 y_i 。

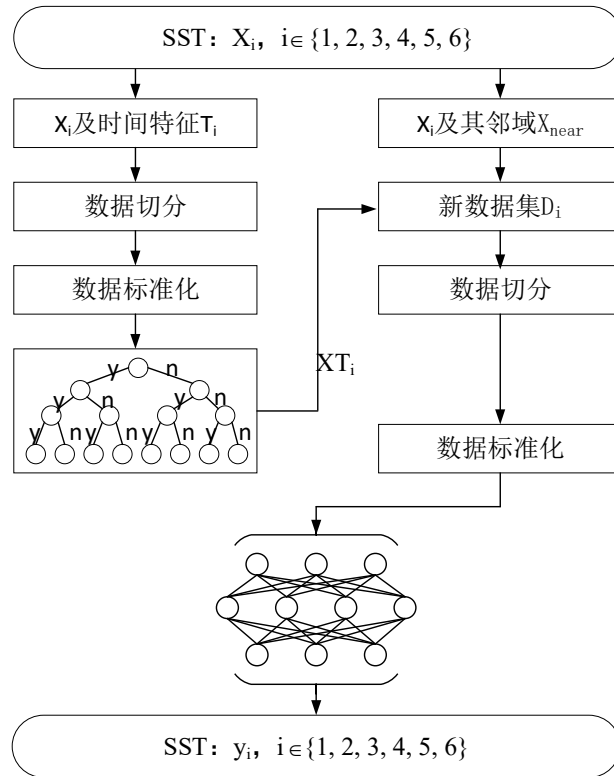


图 3.9 XGB-LSTM 处理流程

Fig. 3.9 XGB-LSTM processing flow

3.5 基于 GRU 和 CNN-1D 结合注意力机制的长期海表温预测

在上节提出的基于 XGBoost 结合 LSTM 的中期海表温预测模型中, 可以实现多步长的预测任务, 且实验数据表明, 其预测误差较小。但进一步实验后发现, 随着预测步长的增加, 每步长预测结果的误差越来越大, 在长期海表温预测中尤其明显。除此之外, 随着预测步长的增加, 所需的历史数据迅速增加, 但是, 历史数据对未来数据的影响力并不完全相同。一般来说, 时间距离越近, 影响越大。若直接将数据输入模型中, 这会增大模型的复杂度。所以如何从长时间序列中选取更有效的数据也是亟待解决的问题。

卷积神经网络在近几年的图像处理中发挥了巨大作用, 尤其是二维卷积 (CNN-2D) 在图像数据特征的提取和数据压缩上表现优异。针对单点海表温预测, CNN-2D 并不适用, 但一维卷积 (CNN-1D) 具备特征提取和数据压缩的优点, 且在音频生成和机器翻

译领域取得了巨大成功，对于时间序列预测任务，CNN-1D 同样适用。所以本文使用 CNN-1D 来从长时间序列中提取更有效的特征的同时压缩数据(降低一维输入的长度)。一维卷积的工作原理如图 3.10 所示。

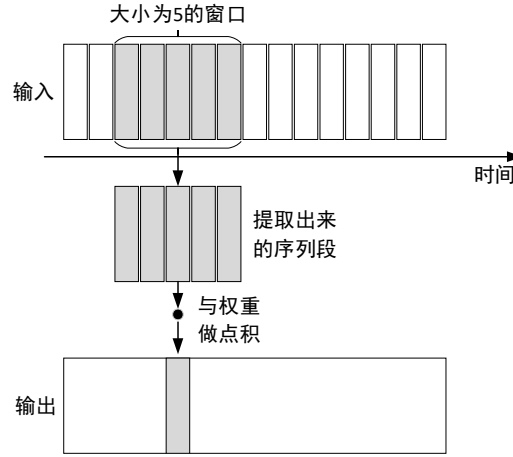


图 3.10 CNN-1D 工作原理

Fig. 3.10 CNN-1D works

注意力机制最初来源于对人类视觉的研究，人类会选择性的关注信息的某一部分，同时忽略其他。近些年来，注意力机制在机器翻译、图像生成等领域获得巨大成功。总体上看，注意力机制的原理就是：计算输入序列与输出序列的匹配程度，匹配度高也就是注意力集中点，相对的得分越高，权重越大，最后加权求和。注意力机制可以应用到 GRU 输出之后，对 GRU 的输出进行修正，减小长期预测的误差。GRU 内部结构如图 3.11 所示。

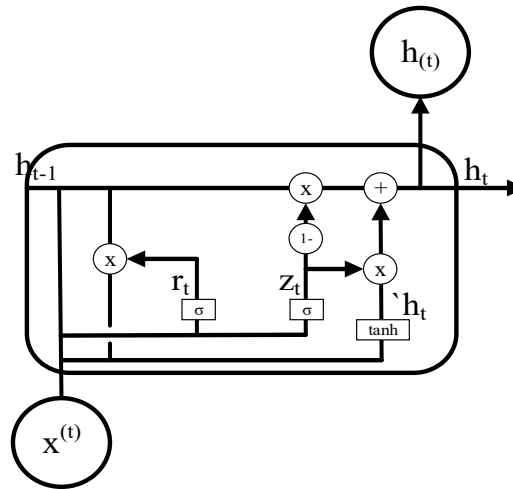


图 3.11 GRU 的一个单元结构图

Fig. 3.11 The framework of a unit of GRU

其各个部分满足关系式如下：（图 3.11 中 h_t 即公式中 \tilde{h}_t ）

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (3.10)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (3.11)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \quad (3.12)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (3.13)$$

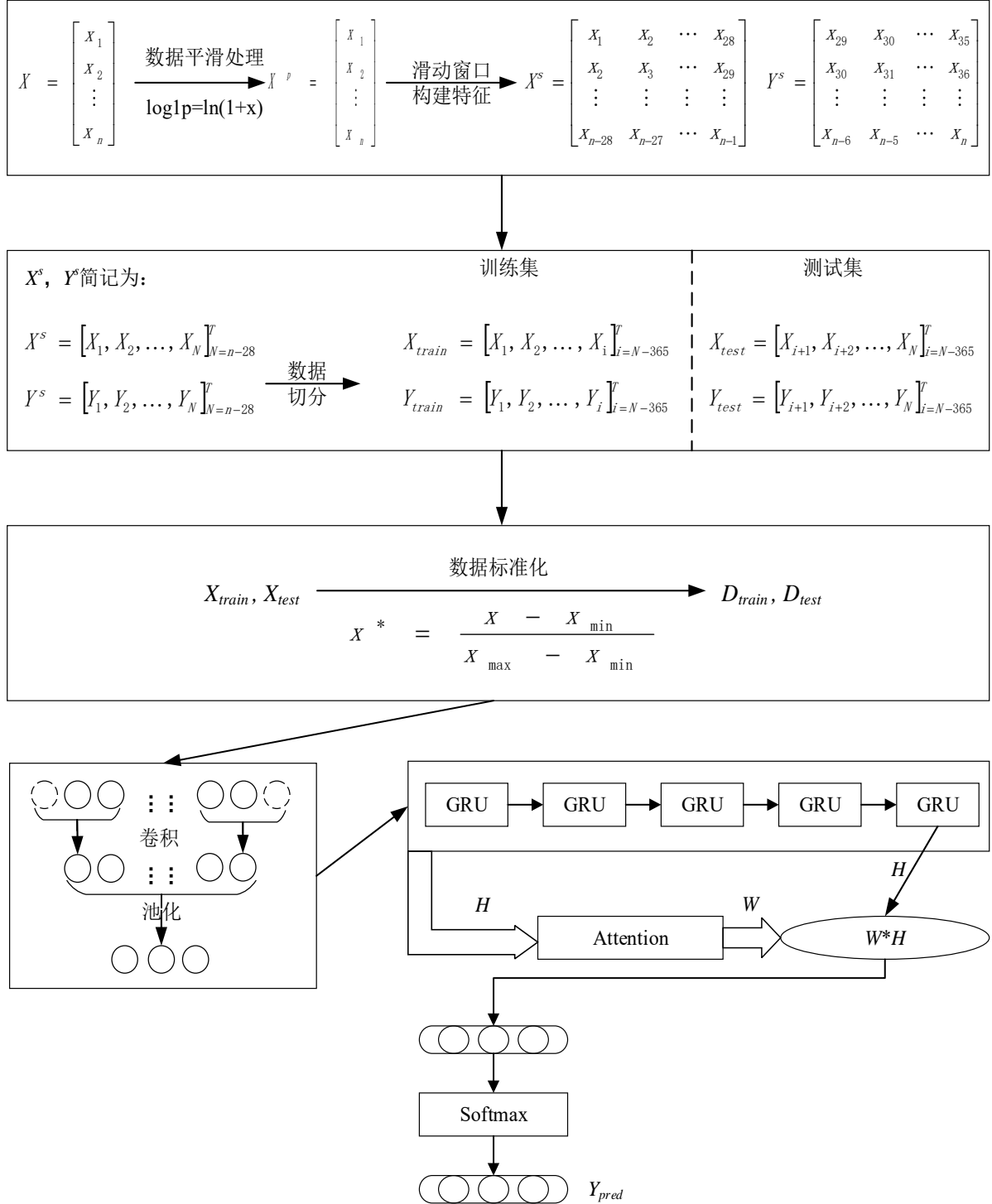


图 3.12 海表温预测流程
Fig. 3.12 SST prediction process

为此，在实验的基础上，本文提出基于 GRU 和 CNN-1D 结合注意力机制的长期海表温预测模型。本文使用一维卷积对历史数据进行特征的筛选，提取特征的同时压缩数据；在 GRU 层后加入注意力机制，使得模型有能力有区别的对待输入中的不同部分。

算法 3.1: 基于 GRU 和 CNN-1D 结合注意力机制的海表温预测算法

输入: 海表温单观测点数据集 X ，训练轮数 E

输出: 海表温预测值数据集 D'

BEGIN

1. 将数据集 $X = [x_1, x_2, \dots, x_n]^T$ 使用 $\ln(1+x)$ 进行平滑处理，处理结果记为 X^p ；
2. 采用滑动窗口法对平滑后的数据 X^p 构建特征 X^s 和对应值 Y^s ；
3. 将数据 X^s 和 Y^s 切分为训练集 X_{train} 、 Y_{train} ，测试集 X_{test} 、 Y_{test} ；
4. 将 X_{train} 和 X_{test} 进行最大最小标准化，结果记为 D_{train} 和 D_{test} ；
5. 使用 D_{train} 和 Y_{train} 训练模型；
6. **For** 第 j 轮 E_j

For 数据集 D_{train} 中的第 i 批数据 D_i ；

将 D_i 输入 CNN-1D 提取数据特征，获得向量 S_i ；

将向量 S_i 输入 GRU，结果记为 H_i ；

将数据 H_i 通过注意力机制求出权重向量 W_i ；

将 H_i 和权重向量 W_i 的乘积通过全连接层映射为最终的输出 D'_i ；

EndFor

EndFor

7. 将测试集 D_{test} 输入到训练好的模型中；
8. **return** 海表温预测值数据集 D' ；

END

长期海表温预测模型流程如图 3.12 所示。图 3.12 是基于 GRU 和 CNN-1D 结合注意力机制的 SST 模型处理流程。数据 X 是单个观测点的历史数据，使用公式 (3.1) 对 X 进行平滑处理，结果记为 X^p 。数据平滑处理使得数据的分布比原数据更加平稳，平滑处理后的数据与原数据对比如图 3.4。采用滑动窗口法构建数据特征 X^s 和对应值 Y^s 。本文将 1982~2017 年的所有数据作为训练集并记为 X_{train} 、 Y_{train} ，将 2018 年的数据作为测试集并记为 X_{test} 、 Y_{test} 。由于数据在不同维度上的特征的尺度并不完全一致，所以要对数据进行标准化，消除特征之间的差异。本文采用最大最小标准化方式对特征进行标准化，标准化的结果记为 D_{train} 、 D_{test} 。数据全部标准化后，将训练集 D_{train} 、 Y_{train} 输入到构建好的网络模型中。随着预测步长的不断增长，所需的历史序列越长。这里先使用 CNN-1D 对历史数据进行特征的提取，减少模型输入的维度，降低模型复杂度，提高训

练速度。本文经过多次实验,发现使用两层一维卷积的效果更好。将一维卷积后的数据输入到 GRU 网络,提取出序列的时序特征并训练预测。本文使用一个全连接层来实现注意力机制:将 GRU 网络的输出 H ,输入到全连接层中,通过反向传播,最终求出对应的 GRU 网络的输出向量的权重 W 。然后将 GRU 网络的输出向量 H 乘以对应权重 W 得出加权后的向量,再经过前馈神经网络输出最终预测结果。算法 3.1 详细说明了本章节基于 GRU 和 CNN-1D 结合注意力机制的海表温预测模型的训练预测过程。

3.6 本章小结

本章提出了单点海表温预测问题。首先根据预测步长的不同,将预测分为短期(单步长)、中期(三步长)、长期(七步长)三种预测,每种预测根据步长的不同采取不同的预测方法。

针对于单步长的短期预测问题,首先提出基于 XGBoost 模型的短期海表温预测。通过实验比较 XGBoost 模型、SVM 模型、AdaBoost 模型、CatBoost 模型在短期海表温预测的表现,验证基于 XGBoost 的短期海表温预测效果要好于其他模型。

针对中期预测,提出基于 XGBoost 结合 LSTM 的中期预测模型。由于 XGBoost 模型无法进行多步长的预测,而 LSTM 可以进行多步长的预测,且 LSTM 模型可以学习序列的历史数据的关系,所以针对中期海表温预测,提出基于 XGBoost 结合 LSTM 的中期海表温预测。与单纯使用 LSTM 模型来进行多步预测相比,基于 XGBoost 结合 LSTM 的预测模型预测结果更好。

针对长期预测,提出基于 GRU 和 CNN-1D 结合注意力机制的长期海表温预测模型。随着预测步长的进一步增加,针对于每次的预测所需要的历史数据也越来越大,无论是 XGBoost 还是 LSTM,都无法进一步学习到长序列的特征,所以预测的误差越来越大。而一维卷积非常适合进行时间序列特征的提取,所以使用一维卷积来提取历史数据特征。而前些天的数据对未来数据的影响并不是完全相同的,所以在 GRU 层后添加注意力机制,来修正 GRU 层的输出,进而获得更好的预测结果。

第 4 章 局域海表温预测

针对局域海表温预测问题,本章按预测步长的不同,将预测分为短期预测(单步长)、中期预测(3 步长)、长期预测(7 步长)三种。针对短期、中期预测,采用基于 LSTM 结合二维卷积(CNN-2D)的预测模型;针对长期预测,提出基于 CNN-2D 和 LSTM 融合注意力机制的长期预测模型。

4.1 问题提出

海表温时间序列预测,主要通过历史时间序列数据来预测未来一天或几天的海表温数据。之前的方法都是针对单点的海表温进行预测,主要是通过获取历史时间段内的数据和未来数据的关系来预测海表温。但这只是考虑了时间因素,然而海平面上的点的温度,也会受到周围点的影响。传统的时间序列预测技术主要分为两类,一类是根据历史数据特征预测未来数据的变化特征,是一种确定型的预测;另一类将预测变量看作随机过程,建立模型进而预测,是随机型时间序列预测方法。传统时间序列预测技术主要有自回归模型(AR)、移动平均模型(MA)、自回归移动平均模型(ARMA)以及这些模型的变异模型组成。虽然传统的时间序列预测方法理论成熟、使用简单,但不适合处理非线性数据且无法进行多步预测,针对海域海表温的计算需要逐点进行计算,难以考虑点与点之间的影响关系。海表温数据是非线性的,但传统非线性模型在海表温预测上表现不佳且只能进行单步预测。基于机器学习的预测方法虽然在处理非线性数据上的表现比传统时间序列方法更好,但仍然是单步预测,针对海域海表温的计算需要逐点进行计算,难以考虑点与点之间的影响关系。虽然单点海表温时间序列预测,可以把周围点的历史数据作为一个特征,但还是针对单观测点的预测,也就是说如果想预测多个点且取得较好的预测效果,那么每个点必须都得有自己得模型。现在深度学习中的卷积神经网络可以提取这种横向的空间特征从而解决这个问题。二维卷积在图像领域取得了众多成果,所以本文将海平面上的点看作一张通道数为 1 的图像,使用卷积神经网络提取点与点之间影响关系的同时压缩输入数据特征数量,降低模型复杂度。考虑到循环神经网络既可以学习数据的时序特征又可以进行多步输出,所以本文使用循环神经网络来解决多步预测的问题。

针对局域海表温预测问题,本文将其分为短期、中期、长期三类,对每一类提出适合的模型,本章总体研究框架如图 4.1 所示。本章首先介绍数据的选取和处理,然后依

次介绍基于 LSTM 结合 CNN-2D 的短中期海表温预测模型、基于 CNN-2D 和 LSTM 融合注意力机制的长期海表温预测模型。

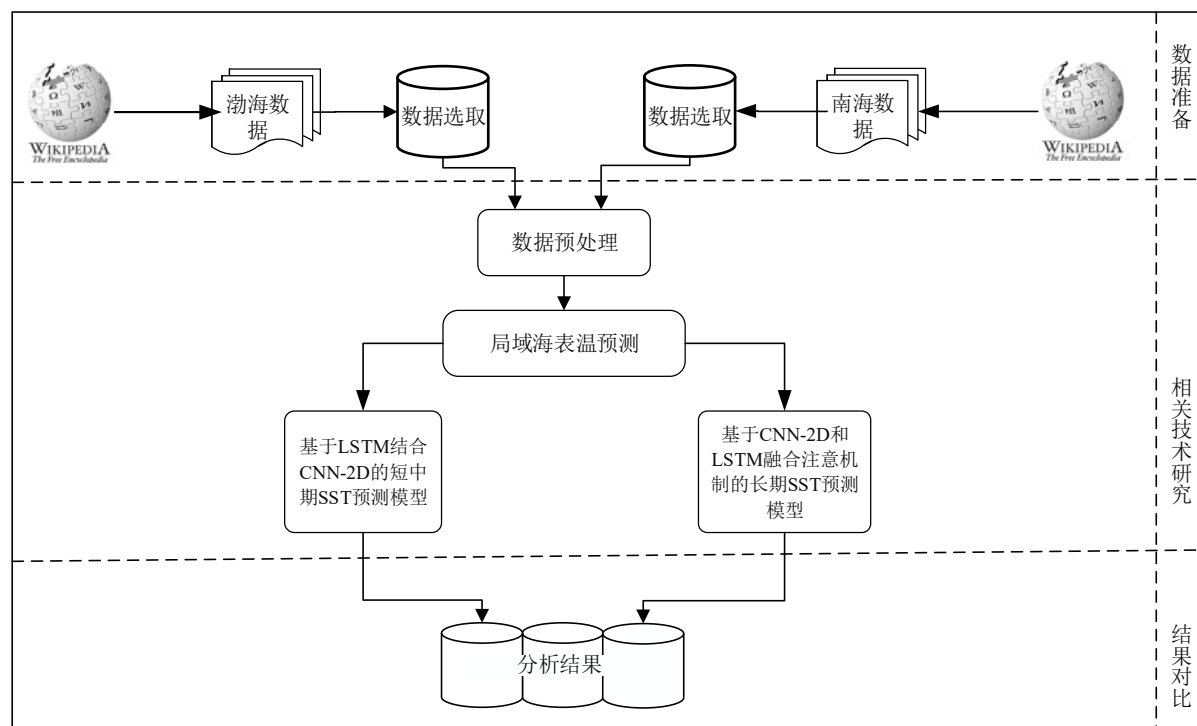


图 4.1 总体研究框架

Fig. 4.1 The overall research framework

4.2 数据选取及预处理

本章研究的是局域海表温预测。所使用的数据集是 NOAA/OAR/ESRL PSD 提供的高分辨率海温数据，该数据来自美国科罗拉多州波尔德市的 NOAA/OAR/ESRL PSD 的网站 <http://www.esrl.noaa.gov/psd/>。这个数据集包含 SST 日平均值、SST 周平均值、SST 月平均值，时间从 1981 年 9 月 1 日开始至今，覆盖全球海洋范围 98.875S~98.875N，0.125E~359.875E，它是 $0.25^{\circ}\text{C} \times 0.25^{\circ}\text{C}$ 的网格形式。本文所有实验所采用的数据是从以上数据中选取的我国渤海和南海 SST 日平均数据。时间范围从 1982 年 1 月 1 日到 2018 年 12 月 31 日，总共 13514 天。渤海数据选取经纬度范围为：37.07N~41N，117.35E~121.10E；南海数据选取经纬度范围 4N~21N，105E~118E。

4.2.1 数据选取

首先是选取局域海面，针对渤海数据集，渤海数据集每天的海表温是一张 16×15 的二维表。在选取数据时，尽量避开陆地减少无效数据。所选取海面如图 4.2 中的橙色部分所示。图 4.2、图 4.3 中观测点的标号都是从 0 开始，按行优先的原则依次递增。

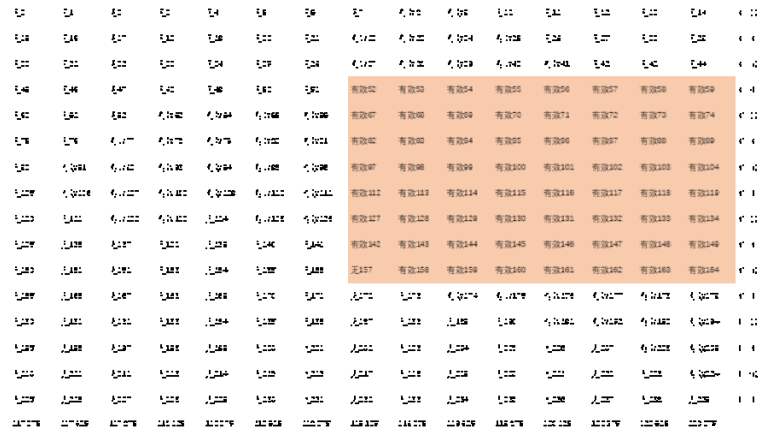


图 4.2 渤海数据区域选取

Fig. 4.2 BoHai data area selection

图 4.2 是渤海地区 2018 年某天的海表温图，“无”表示陆地区域，“有效”表示海域，右侧是北纬刻度，下侧是东经刻度，其中橙色部分为选取的海域 8×8 ，每块海域都有自己的标号。对于南海数据的选取如图 4.3 中橙色部分所示，左上角橙色部分是选取的 8×8 区域。

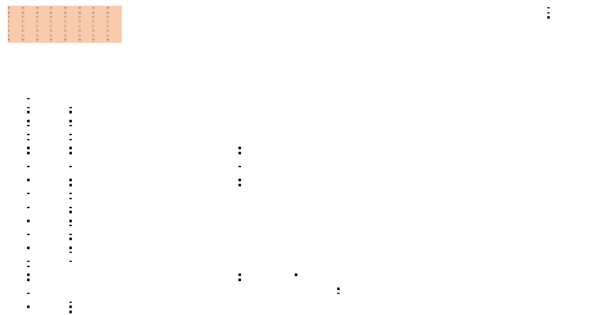


图 4.3 南海数据区域选取

Fig. 4.3 NanHai data area selection

4.2.2 数据预处理

由于从网站上获取的数据是标准的.nc 格式，无法直接输入到模型中进行训练和预测，所以本文对该数据集进行了一些预处理操作，主要包括：

文件格式转换。后缀名.nc 的文件又名 NetCDF 文件，NetCDF 全称为 Network Common Data Format（网络通用数据格式）。它适于网络共享的数据描述和编码标准，属于面向数据型数据。但正是它的这种独特的格式，并不适合直接拿来训练和预测，所以本文通过 netCDF4 工具包，将其中的数据提取出来存储为常用的.csv 格式，方便后续的数据处理。

数据平滑处理。选取的渤海、南海数据集，分别从渤海、南海中选取 8×8 大小的区域。分别从渤海、南海随机选取 6 个观测点，画出这些点在 2018 年的日平均海表温变

化趋势图。从图中可看出数据的变化先降后升再降，数据的变化较大，尤其是七八月份，温度变化大。对模型而言，这样的数据很难获得更好的拟合效果，所以将每个点的历史数据使用如下函数进行平滑处理，处理后的效果如图中的红线所示，可看出数据比之前平滑了很多，有利于后期的训练和预测。这里进行平滑处理使用的是如下公式：

$$\log 1p = \ln(1 + x) \quad (4.1)$$

实际数据中存在负数，由于最小的负数仍然大于-5，所以这里在进行数据平滑处理之前，先将所有数据加 5，消除负数，然后再进行 $\log 1p$ 操作。

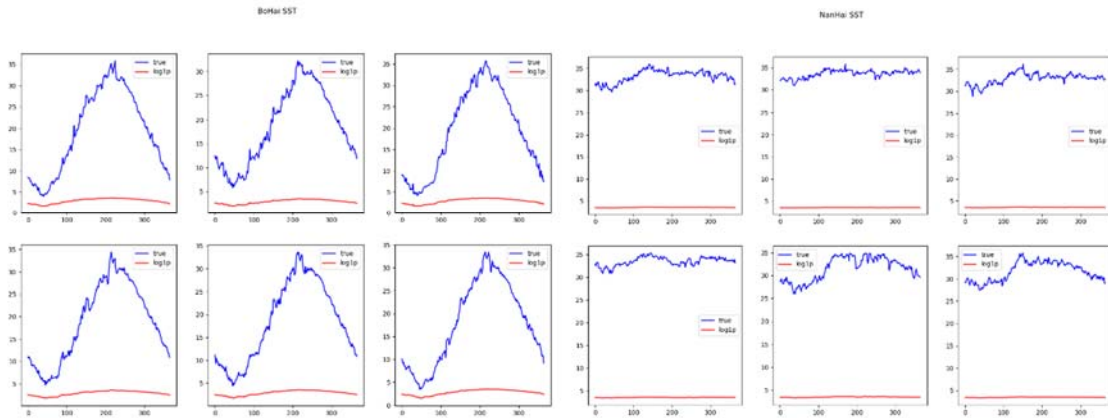


图 4.4 海表温平滑处理效果对比图

Fig. 4.4 Comparison chart of SST smoothing effect

由图 4.4 可以看出，渤海海域海表温随时间变化较大，而南海海域海表温相对来说变化较小。

数据标准化。本文所有标准化操作，都是采用最大最小标准化，将所有训练集数据标准化到[0,1]范围内，计算过程如公式（4.2）所示。

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4.2)$$

其中 x_{max} 为样本数据最大值， x_{min} 为样本数据最小值。 x^* 是标准化后的数据。

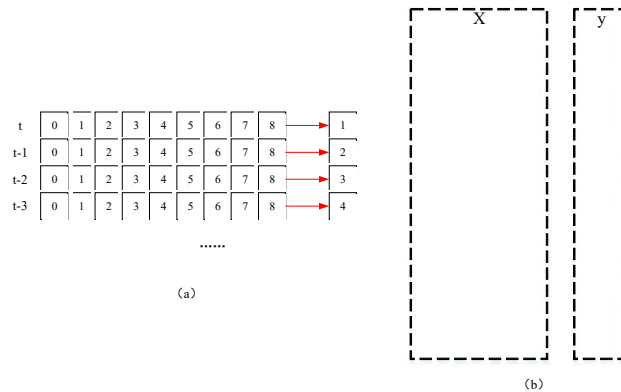


图 4.5 滑动窗口法拆分数据

Fig. 4.5 Split data by sliding window method

滑动窗口法拆分数据。本文是以历史海表温度来预测未来海表温度，属于自回归。这里通过滑动窗口法（图 4.5）将自回归转换成监督学习，即将前几天历史数据作为特征，而后几天数据作为标签。图 4.5（a）中橙色 0~8 表示观测点标号，为了方便说明将第 t 天的数据记为标号为 1 的绿色块儿，第 $t-1$ 天的数据记为标号为 2 的绿色块儿，以此类推。

4.3 基于 LSTM 结合 CNN-2D 的短中期海表温预测

对于局域海表温预测问题，本文采用二维卷积来提取海平面点和点的温度之间的空间关系，使用 LSTM 网络提取每个点历史数据和未来数据的关系，最后通过全连接层输出最后的预测结果，基于 LSTM 结合 CNN-2D 的短中期海表温预测模型如图 4.6。

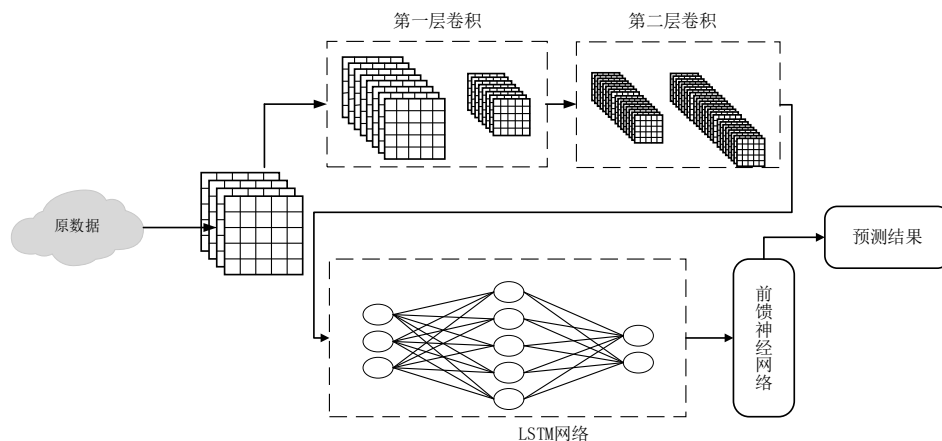


图 4.6 基于 LSTM 结合 CNN 的模型

Fig. 4.6 LSTM combines the model of CNN

第一步如图 4.7（b）所示将原数据转换成.csv 格式数据 D ，图 4.7（a）中的 0~8 表示观测点标号， t 、 $t-1$ 、 $t-2$ 分表代表第 t 天、第 $t-1$ 天、第 $t-2$ 天。使用公式（4.1）进行平滑处理，接着使用滑动窗口法拆分数据，再对数据使用公式（4.2）进行最大最小标准化。第二步将 1982~2017 年的数据作为训练集，2018 年的数据作为测试集。第三步使用二维卷积提取平面上点和点之前的空间关系。第四步使用 LSTM 网络提取数据的时序关系。最后通过全连接层（前馈神经网络）将 LSTM 网络的输出结果映射为最终的预测结果，算法 4.1 详细说明了基于 LSTM 结合 CNN-2D 的训练预测过程。

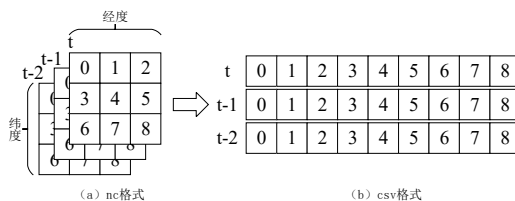


图 4.7 数据转换

Fig. 4.7 The data conversion

算法 4.1: 基于 LSTM 结合 CNN-2D 的短中期海表温预测算法

输入: 海表温数据集 D , 轮数 E

输出: 海表温预测值数据集 D'

BEGIN

1. 将数据集 D 使用函数 $\ln(1+x)$ 处理后的结果记为 X^p ;
2. 使用滑动窗口法将数据 X^p 处理为特征 X , 特征对应值 Y ;
3. 将数据 X 、 Y 划分为训练集 X_{train} 、 Y_{train} , 测试集 X_{test} 、 Y_{test} ;
4. 对特征 X_{train} 、 X_{test} 进行最大最小标准化结果分别记为 S_{train} 、 S_{test} ;
5. 使用 S_{train} 、 Y_{train} 训练模型;
6. **For** 第 j 轮 E_j

For 数据集 S_{train} 中的第 i 批数据 B_i

将 B_i 输入 CNN-2D, 提取数据空间特征 M_i ;

将 M_i 输入 LSTM 提取时序特征 T_i ;

通过全连接层将 T_i 映射为最终的预测输出 D'_i ;

EndFor

EndFor

7. 将测试集 S_{test} 输入到训练好的模型中;
8. **return** 海表温预测值数据集 D' ;

END

4.4 基于 CNN-2D 和 LSTM 融合注意力机制的长期海表温预测

上节使用二维卷积提取数据空间特征, LSTM 网络提取时序特征, 将空间和时间维度都考虑到了, 但随着预测步长的增加, 误差逐渐增大。因为使用二维卷积主要用来提取数据的空间特征, 所以步长的影响较小。而 LSTM 网络提取的是数据的时序特征, 随着步长的增加, 训练时需要的步长增加了很多, 由于时间跨度变大, 所以 LSTM 网络学习时序特征的效果变差, 然后在 LSTM 网络后加入注意力机制, 进一步减少预测误差。

如图 4.8, 首先对原数据进行数据选取, 选取渤海观测区域和南海观测区域的历史数据。由于温度的变化并不平稳, 所以使用公式 (4.1) 对观测区域的历史数据 X 进行平滑处理, 使得数据的分布比原数据更加平稳。由于数据在不同维度上的特征的尺度并不完全一致, 所以要对数据进行标准化, 消除特征之间的差异。数据全部标准化后, 将 1982~2017 年的所有数据作为训练集 X_{train} , Y_{train} , 2018 年的数据作为测试集 X_{test} , Y_{test} 。使用 CNN-2D 对每天的海域数据进行卷积运算, 压缩数据维度的同时提取出观测区域上点和点之间海表温的空间特征, 然后将带有空间特征的数据输入到 LSTM 网络

中，通过 LSTM 网络学习数据的时序特征。使用注意力机制来修正 LSTM 网络的输出，最终通过全连接层获得预测结果。

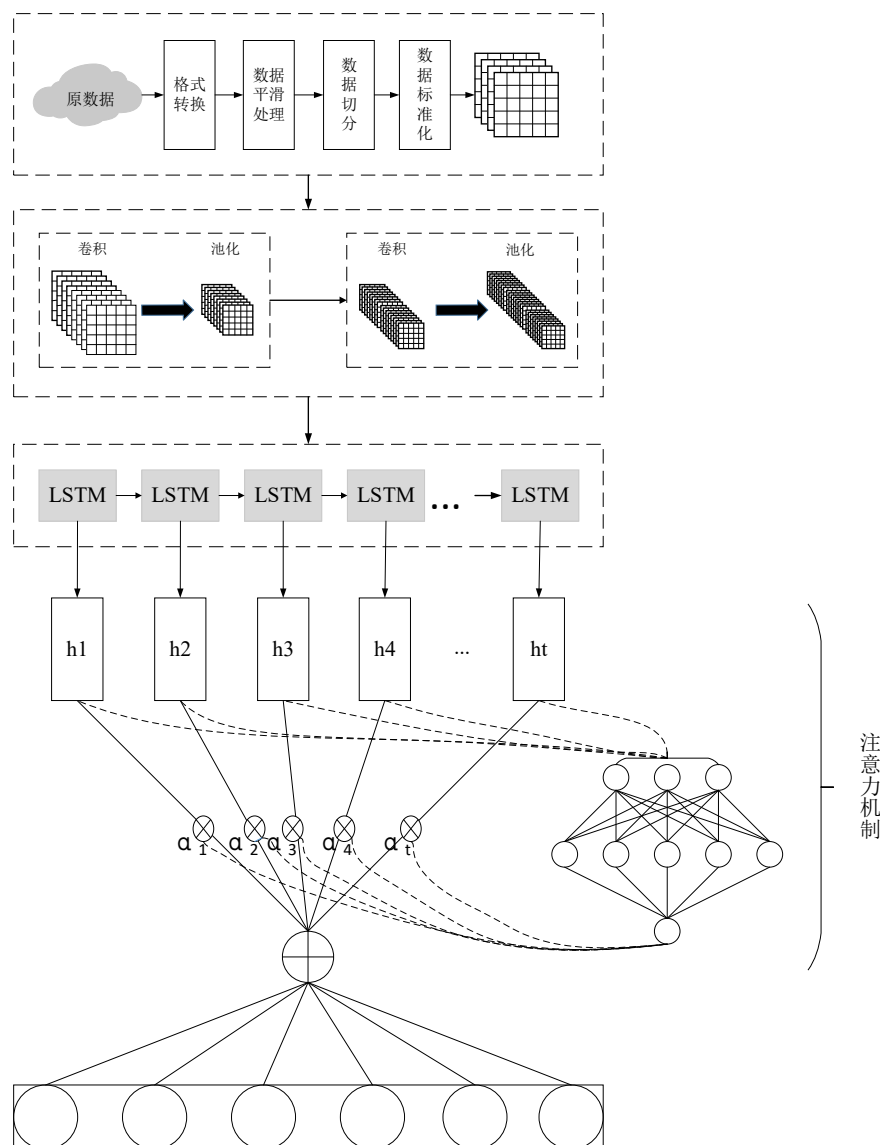


图 4.8 局域海表温预测流程

Fig. 4.8 Model based on LSTM combined with attention mechanism

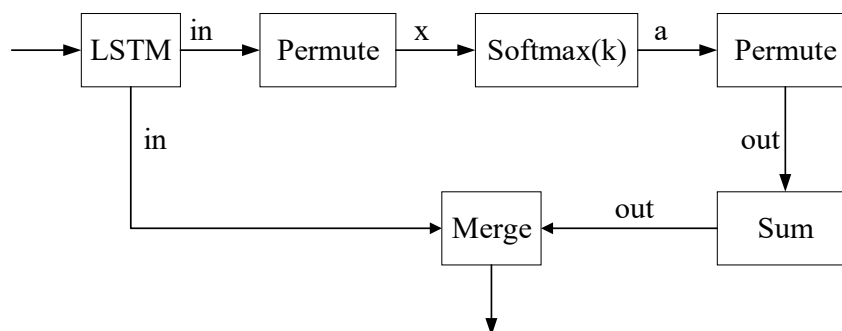


图 4.9 自定义 Keras 层

Fig. 4.9 Custom Keras layer

图 4.9 是自定义 Keras 层, $k = \tanh(x \cdot W + b)$, W 为权重矩阵, b 为偏置矩阵。本文是自定义一个 Keras 层(AttentionLayer)实现注意力机制(图 4.9), 通过 AttentionLayer 来获得 LSTM 每次迭代输出向量的权重。将 LSTM 网络每次迭代的输出与其对应权重相乘并求和, 得出加权向量, 然后通过全连接层映射为最终的输出结果。算法 4.2 详细说明了本章节基于 CNN-2D 和 LSTM 融合注意力机制的海表温预测模型的训练预测过程。

算法 4.2: 基于 CNN-2D 和 LSTM 融合注意力机制的海表温预测算法

输入: 海表温数据集 D , 轮数 E

输出: 海表温预测值数据集 D'

BEGIN

1. 将数据集 D 使用函数 $\ln(1+x)$ 处理后的结果记为 X^p ;
2. 使用滑动窗口法将数据 X^p 处理为特征 X , 特征对应值 Y ;
3. 将数据 X 、 Y 划分为训练集 X_{train} 、 Y_{train} , 测试集 X_{test} 、 Y_{test} ;
4. 对特征 X_{train} 、 X_{test} 进行最大最小标准化结果分别记为 S_{train} 、 S_{test} ;
5. 使用 S_{train} 、 Y_{train} 训练模型;
6. **For** 第 j 轮 E_j

For 数据集 S_{train} 中的第 i 批数据 B_i

将 B_i 输入 CNN-2D, 提取数据空间特征 M_i ;

将 M_i 输入 LSTM 提取 LSTM 每步的输出 T_i ;

将数据 T_i 通过注意力机制求出权重向量 α ;

再求出数据 T_i 和权重向量 α 的乘积之和 SUM;

通过全连接层将 SUM 映射为最终的预测输出 D'_i ;

EndFor

EndFor

7. 将测试集 S_{test} 输入到训练好的模型中;
8. **return** 海表温预测值数据集 D' ;

END

4.5 本章小结

本章提出了局域海表温预测。首先将区域某一天的数据看作一张只有 1 一个通道的二维图像, 针对短期、中期预测, 提出了基于 LSTM 结合二维卷积的中短期局域海表温预测模型。由于某点温度数据和其周围邻域点温度数据是存在某种关系的, 所以将区域看作为二维图像, 使用二维卷积来提取数据点间的空间相互关系; 然后将提取了空间相互关系的数据通过 LSTM 提取数据的时序关系, 获得最终的预测结果。

针对长期预测, 提出基于 CNN-2D 和 LSTM 融合注意力机制的长期预测模型。随着预测步长的进一步增加, 针对于每次的预测所需要的历史数据也越来越大, 如果将数据直接输入到 LSTM 网络, 会增大网络的复杂度, 且数据的重要程度不同, 不应该同等对待。所以使用二维卷积提出更重要的特征, 减少了输入数据量的同时也提取了点和点之间的空间特征。然后在使用 LSTM 学习数据的时序特征。但随着预测步长的增加, 每步长的误差越来越大, 为此, 在 LSTM 层后添加注意力机制, 对 LSTM 步的输出添加合适的权重, 来修正 LSTM 的输出, 进而获得更好的预测结果。

第 5 章 实验与分析

在海洋领域中，海洋表面温度（海表温，SST）是影响海洋气候的重要因素，海表温预测是海洋预报的关键环节。海表面单个观测点的历史温度是典型的长时间序列数据，而且海表温数据量太过庞大，传统方法无法适用于处理大量观测点的海表温数据。目前对海表温的预测，主要是针对单观测点（单点）、短期的预测。而本文根据预测步长的不同，将海表温预测进行分类。针对每类步长下的预测，提出适合的模型，使得预测更加有针对性。

5.1 实验环境

本章实验所基于的软硬件环境如表 5.1 所示。

表 5.1 软硬件环境

Table 5.1 Software and hardware environment

| | | |
|----|------|-----------------------------------|
| 硬件 | CPU | Inter Core i7-9750H CPU @ 2.60GHz |
| | GPU | GeForce GTX 1650 4.00GB |
| | 内存 | 16.00GB |
| | 硬盘 | 512G SSD |
| 软件 | 操作系统 | Windows 10 |
| | 开发语言 | Python 3.6.8 |
| | 开发框架 | Keras 2.2.4 |
| | 开发工具 | PyCharm 2019.2.3 Professional |

5.2 评估标准

本文实验所采取的评估标准是均方误差（MSE）、均方根误差（RMSE）、平均绝对误差（MAE）和确定系数（ R^2 ）。假设预测值： $\hat{y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}$ ，真实值： $y = \{y_1, y_2, \dots, y_n\}$ ，四种标准的计算如公式（5.1）~公式（5.4）所示。

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (5.1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (5.2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (5.3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y} - y_i)^2} \quad (5.4)$$

公式 (5.1) 是均方误差 MSE 的计算公式, 其范围 $[0, +\infty)$ 。若 MSE 为 0, 则表示模型的预测结果和真实结果完全一致, 说明该模型的拟合效果非常好; MSE 越大, 表示预测值和真实值相距越远, 即模型的拟合效果越差; MSE 越小, 表示预测值和真实值相距越近, 即模型的拟合效果越好。MSE 是带有平方形式方便求导, 常用来作回归的损失函数, 而本文中也将 MSE 作为模型的评估标准之一。

公式 (5.2) 是均方根误差 RMSE 的计算公式, 其范围 $[0, +\infty)$ 。若 RMSE 为 0, 则表示模型的预测结果和真实结果完全一致, 说明该模型的拟合效果非常好; RMSE 越大, 表示预测值和真实值相距越远, 即模型的拟合效果越差; RMSE 越小, 表示预测值和真实值相距越近, 即模型的拟合效果越好。RMSE 是 MSE 的开方, 常作为模型预测结果的衡量标准, 本文将 RMSE 作为模型的评估标准之一, RMSE 可以将误差放大从而更容易区分出模型的差异。

公式 (5.3) 是平均绝对误差 MAE 的计算公式, 其范围 $[0, +\infty)$ 。若 MAE 为 0, 则表示模型的预测结果和真实结果完全一致, 说明该模型的拟合效果非常好; MAE 越大, 表示预测值和真实值相距越远, 即模型的拟合效果越差; MAE 越小, 表示预测值和真实值相距越近, 即模型的拟合效果越好。

公式(5.4)是确定系数 R^2 的计算公式, 正常取值范围为 $[0, 1]$, 理论取值范围为 $(-\infty, 1]$, 确定系数反映了 y 的波动有多少百分比能被 x 的波动所描述。其值越靠近 0, 表示自变量对因变量的解释程度越低, 对模型而言, 拟合效果越差; 越靠近 1, 表示自变量对因变量的解释程度越高, 对模型而言, 拟合效果越好。本文只是将 R^2 作为一个参考, 并不起决定性作用。

5.3 单点海表温预测

本小节将介绍关于单点海表温预测的相关实验。本节的实验是采用基于 Python 的深度学习框架 Keras 和 sklearn 机器学习库进行实现的, 所有的网络权重参数均采用 Keras 默认的初始化器进行初始化。实验的数据集使用渤海和南海日平均海表温数据, 渤海数据选取经纬度范围为: 37.07N~41N, 117.35E~121.10E; 南海数据选取经纬度范围 4N~21N, 105E~118E。时间从 1982 年 1 月 1 日到 2018 年 12 月 31 日, 总共 13514 条; 训练集时

间范围是 1982 年 1 月 1 日到 2017 年 12 月 31 日，总共 13149 条；测试集时间范围是 2018 年 1 月 1 日到 2018 年 12 月 31 日，总共 365 条数据。为了减少随机因素的影响，本节所有结果都是将模型运行 3 次，然后取这些结果的平均值。作为最终结果。

5.3.1 实验相关设置

基于 XGBoost 模型的短期海表温预测，XGBoost 模型的最大迭代次数设置为 1100，学习率设置为 0.15，其余参数按默认设置。SVR、XGBoost、AdaBoost、CatBoost 实验的滑动窗口大小均设置为 4，预测步长为 1。

基于 XGBoost 结合 LSTM 的中期海表温预测，其中 XGBoost 的最大迭代次数设置为 500，学习率设置为 0.15；LSTM 设置为 1 个隐藏层，9 个单元，激活函数 RELU 循环次数设置为 100，批大小设置为 20，滑动窗口设置为 12，损失函数设置为 MSE，优化器设置为 Adam；Dense 设置为 1 层，3 个单元即预测步长为 3。对比实验为 LSTM、GRU 中期海表温预测，参数设置同上。中期海表温预测的滑动窗口大小设置为 12。

基于 GRU 和一维卷积（CNN-1D）结合注意力机制的长期海表温预测，实验设置了 2 层卷积：第 1 个卷积层设置 64 个过滤器，卷积核大小 5×1 ，填充方式 SAME，激活函数 RELU；第 1 个池化层大小设置为 2×1 ；第 2 个卷积层设置 128 个过滤器，卷积核大小设置为 3×1 ，填充方式 SAME，激活函数 RELU；第 2 个池化层大小设置为 2×1 。隐藏层设置为 1 层，15 个隐藏单元，激活函数 RELU。本文采用 Dense 层来实现注意力机制，其单元个数设置为隐藏层输出的第二个维度，激活函数设置为 Sigmoid。最后的 Dense 层设置单元数为 7，即预测步长为 7，激活函数为 Sigmoid。模型的损失函数设置为 MSE，优化器设置为；批大小设置为 20，轮数设置为 100，其他参数均按默认设置。长期海表温预测的滑动窗口大小设置为 28。

5.3.2 实验结果及分析

本节将根据预测步长的不同分为短期、中期、长期三类实验，每类实验都会用渤海、南海两个数据集的数据进行测试。

(1) 针对短期海表温预测，本文提出基于 XGBoost 的短期海表温预测模型，并进行相关实验。数据：渤海 38、73、92、112、131、193 六个观测点的数据，南海 375、397、1324、1345、3025、3050 六个观测点的数据。滑动窗口设置为 4，预测步长为 1。针对每个观测点，本文均计算其预测结果的均方误差、均方根误差、平均绝对误差和确定系数，通过对比图、表格等形式展示结果。最后会计算渤海 6 个观测点和南海 6 个观测点的预测结果的平均值，使用平均值代表模型在渤海和南海的预测值。

图 5.1 是模型在渤海 6 个观测点的预测误差对比。在渤海海域所选取的 6 个观测点上, SVR、XGBoost、AdaBoost、CatBoost 四个模型的 R^2 在 6 个位置点的数值相差不大, 但 XGBoost 模型的 MSE、RMSE、MAE 均小于其他模型, 说明 XGBoost 模型在渤海数据集上的预测误差更小, 模型的拟合效果要好于其他模型。

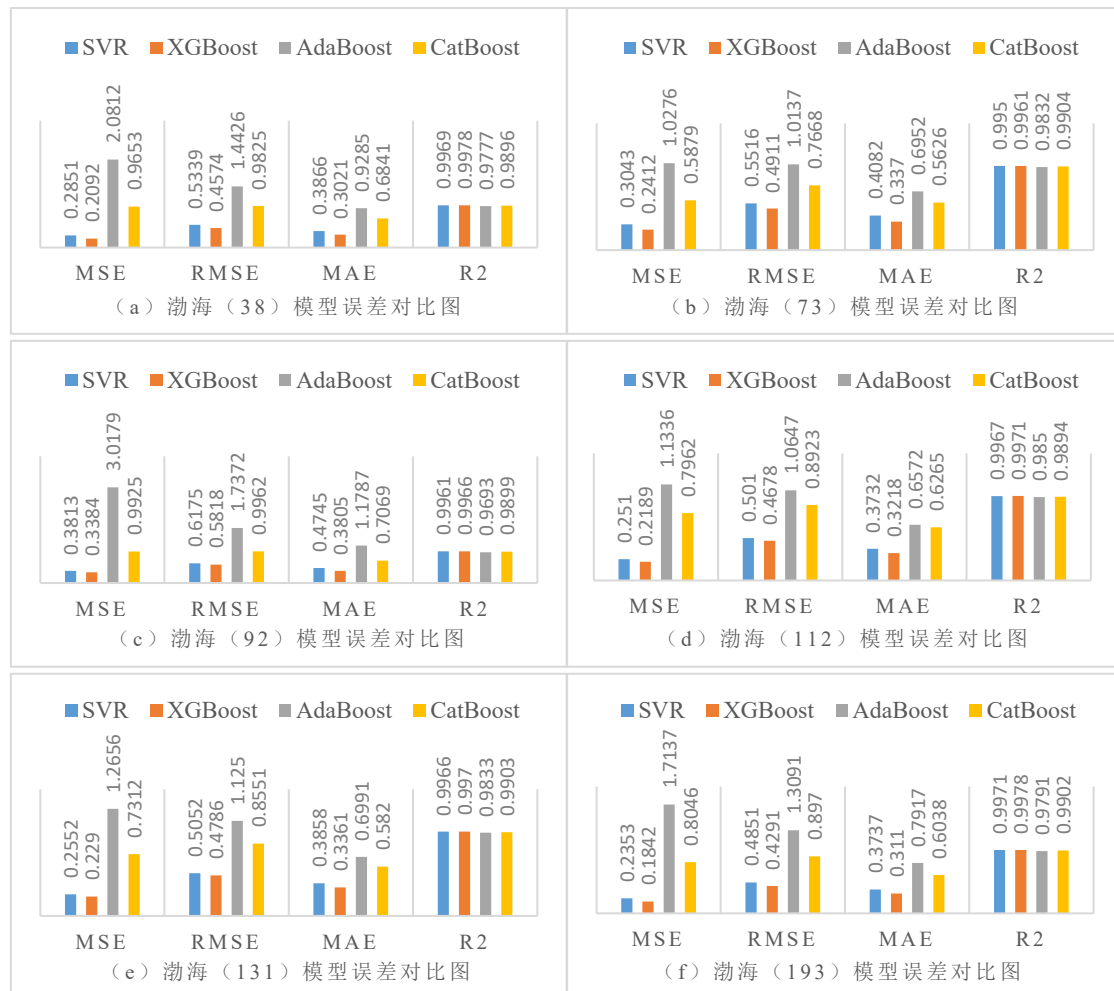


图 5.1 渤海 6 个观测点误差对比图

Fig. 5.1 BoHai 6 position error contrast chart

图 5.2 是模型在南海 6 个观测点的预测误差对比。在南海海域所选取的 6 个观测点上, XGBoost、AdaBoost、CatBoost 三个模型的 R^2 相差不大, 但 SVR 模型的 R^2 明显小于其他模型, 在个别点出现了负值, 表明 SVR 模型在南海数据的拟合能力很差; XGBoost 模型的 MSE、RMSE、MAE 均小于其他模型, 说明 XGBoost 模型在南海数据集上的预测效果要好于其他模型。

将渤海 6 个观测点的误差、南海 6 个观测点的误差分别取平均值, 每个模型在渤海数据的误差如表 5.2 所示, 在南海数据的误差如表 5.3 所示。表中 AVG_MSE、AVG_RMSE、AVG_MAE、AVG_R2 分别是是六个点的 MSE、RMSE、MAE、 R^2 的平均值。

表 5.2 渤海 6 个位置点预测误差平均值

Table 5.2 The mean value of 6 position points in BoHai sea is predicted

| 模型 | AVG_MSE | AVG_RMSE | AVG_MAE | AVG_R2 |
|----------------|---------------|---------------|---------------|---------------|
| SVR | 0.2854 | 0.5324 | 0.4003 | 0.9964 |
| XGBoost | 0.2368 | 0.4843 | 0.3314 | 0.9971 |
| AdaBoost | 1.7066 | 1.2812 | 0.8251 | 0.9796 |
| CatBoost | 0.8130 | 0.8983 | 0.6277 | 0.9900 |

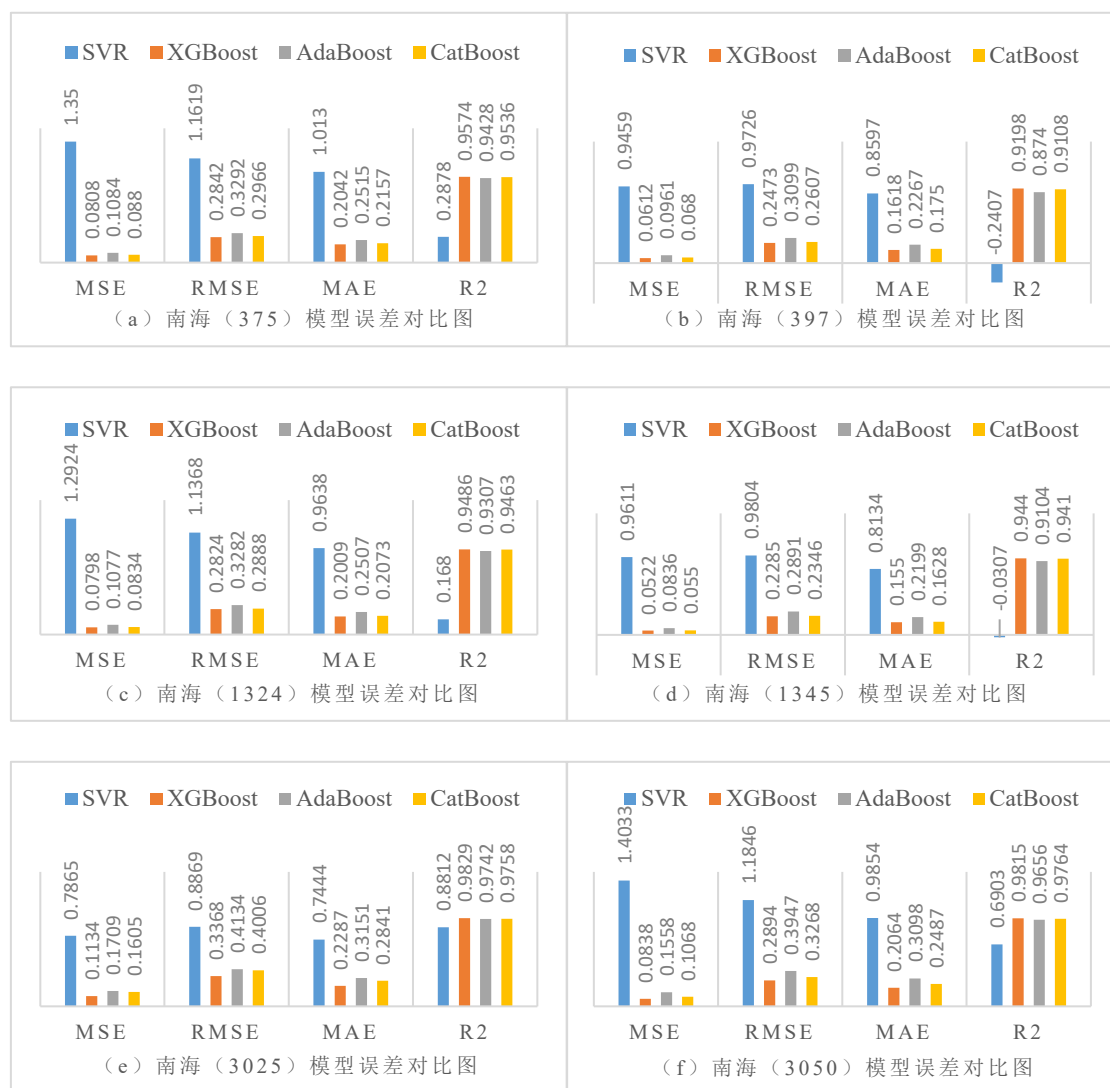


图 5.2 南海 6 个观测点误差对比图

Fig. 5.2 NanHai 6 position error contrast chart

从表 5.2、表 5.3 上可看出，无论是渤海海域还是南海海域，XGBoost 模型的 MSE、RMSE、MAE 的平均值均小于其他模型，表明 XGBoost 模型在短期预测的误差更小，模型的拟合能力更好；而 XGBoost 的 R^2 大于其他模型，这也从侧面说明 XGBoost 模型的拟合能力更强。

表 5.3 南海 6 个位置点预测误差平均值

Table 5.3 The mean value of 6 position points in NanHai sea is predicted

| 模型 | <i>AVG_MSE</i> | <i>AVG_RMSE</i> | <i>AVG_MAE</i> | <i>AVG_R2</i> |
|----------|----------------|-----------------|----------------|---------------|
| SVR | 1.1232 | 1.0539 | 0.8966 | 0.2927 |
| XGBoost | 0.0785 | 0.2781 | 0.1928 | 0.9557 |
| AdaBoost | 0.1204 | 0.3441 | 0.2623 | 0.9330 |
| CatBoost | 0.0936 | 0.3014 | 0.2156 | 0.9507 |

以上的短期海表温预测，只是使用历史数据来预测未来数据，并未考虑“邻域点”的影响。下面进行考虑“邻域点”的短期海表温预测实验，实验结果如下：

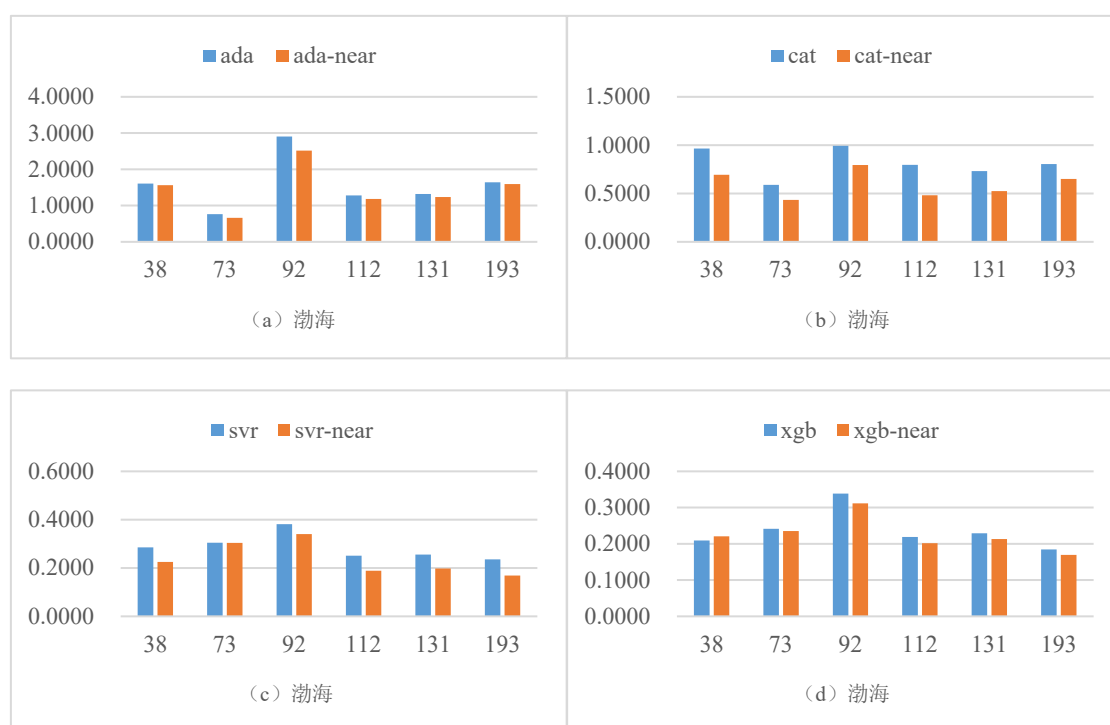


图 5.3 渤海 6 个观测点 MSE 对比图

Fig. 5.3 BoHai 6 position MSE contrast chart

图 5.3 是考虑“邻域点”的短期海表温预测的渤海 MSE 对比图，横坐标是渤海 6 个观测点标号，纵坐标是 MSE。很明显，无论是哪种模型，在考虑“邻域点”之后，其 MSE 比原来都减小了。

图 5.4 是考虑“邻域点”的短期海表温预测的南海 MSE 对比图，横坐标是南海 6 个观测点标号，纵坐标是 MSE。在图 5.4 (a) 中，6 个观测点中有 4 个位置在考虑“邻域点”后，MSE 更小；图 5.4 (b) 中，6 个观测点有 5 个观测点在考虑“邻域点”后，MSE 更小。图 5.4 (c)、图 5.4 (d) 中 6 个观测点有 3 个观测点在考虑“邻域点”后，MSE 更小。由此来看在南海海域，考虑“邻域点”的模型比之前的表现更好。

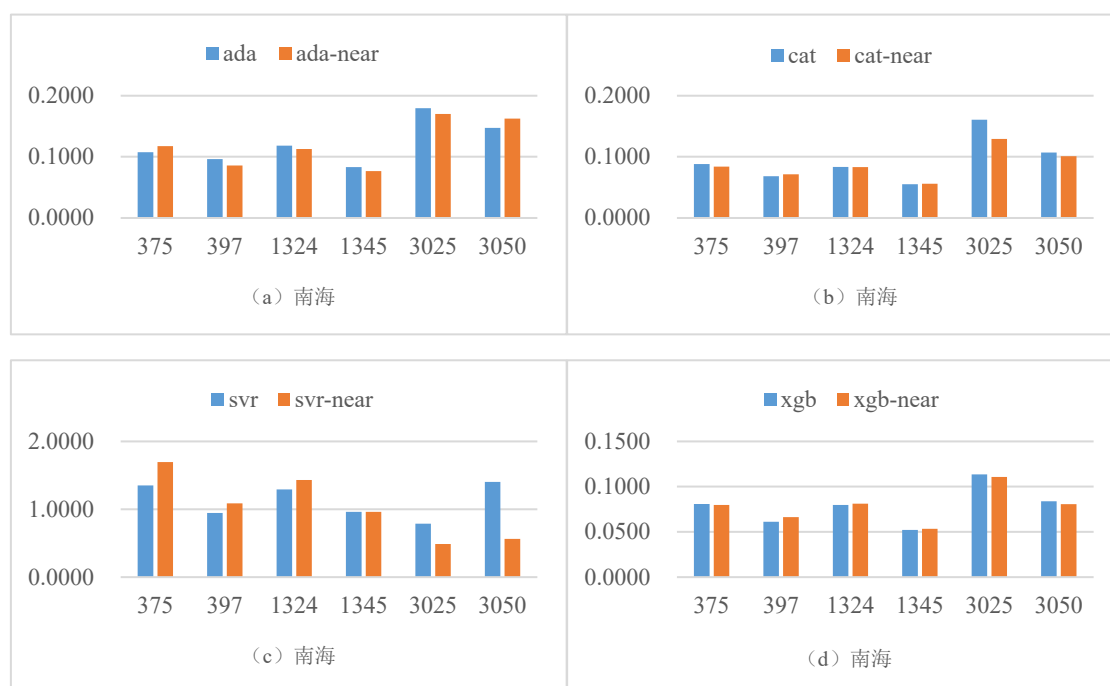


图 5.4 南海 6 个观测点 MSE 对比图

Fig. 5.4 NanHai 6 position MSE contrast chart

表 5.4、表 5.5 分别是渤海 6 个观测点、南海 6 个观测点的 MSE。表中 ada、cat、svr、xgb 分别是未考虑“邻域点”的 AdaBoost、CatBoost、SVR、XGBoost 模型的 MSE；ada-near、cat-near、svr-near、xgb-near 分别是考虑了“邻域点”的 AdaBoost、CatBoost、SVR、XGBoost 模型的 MSE。这里我们看 6 个观测点 MSE 的平均值和中位数，在渤海数据集上，xgb-near 的平均值最小，svr-near 的中位数最小；在南海数据集上，xgb-near 的平均值和中位数均最小。整体来看，考虑了“邻域点”的 XGBoost 的模型的误差更小，模型的拟合能力更强。

表 5.4 渤海 6 个观测点 MSE

Table 5.4 MSE of 6 position points in BoHai

| 位置点 | <i>ada</i> | <i>cat</i> | <i>svr</i> | <i>xgb</i> | <i>ada-n</i> | <i>cat-n</i> | <i>svr-n</i> | <i>xgb-n</i> |
|-----|------------|------------|------------|------------|--------------|--------------|---------------|---------------|
| 38 | 1.6045 | 0.9653 | 0.2851 | 0.2092 | 1.5612 | 0.6938 | 0.2251 | 0.2207 |
| 73 | 0.7595 | 0.5879 | 0.3043 | 0.2412 | 0.6576 | 0.4322 | 0.3034 | 0.2353 |
| 92 | 2.9034 | 0.9925 | 0.3813 | 0.3384 | 2.5134 | 0.7942 | 0.3404 | 0.3118 |
| 112 | 1.2771 | 0.7962 | 0.2510 | 0.2189 | 1.1800 | 0.4807 | 0.1882 | 0.2014 |
| 131 | 1.3167 | 0.7312 | 0.2552 | 0.2290 | 1.2351 | 0.5240 | 0.1980 | 0.2133 |
| 193 | 1.6395 | 0.8046 | 0.2353 | 0.1842 | 1.5902 | 0.6495 | 0.1686 | 0.1694 |
| avg | 1.5835 | 0.8130 | 0.2854 | 0.2368 | 1.4563 | 0.5957 | 0.2373 | 0.2253 |
| med | 1.4606 | 0.8004 | 0.2702 | 0.2240 | 1.3982 | 0.5868 | 0.2116 | 0.2170 |

表 5.5 南海 6 个观测点 MSE

Table 5.5 MSE of 6 position points in NanHai

| 位置点 | <i>ada</i> | <i>cat</i> | <i>svr</i> | <i>xgb</i> | <i>ada-n</i> | <i>cat-n</i> | <i>svr-n</i> | <i>xgb-n</i> |
|------|------------|------------|------------|------------|--------------|--------------|--------------|---------------|
| 375 | 0.1073 | 0.0880 | 1.3500 | 0.0808 | 0.1173 | 0.0838 | 1.6935 | 0.0797 |
| 397 | 0.0960 | 0.0680 | 0.9459 | 0.0612 | 0.0857 | 0.0712 | 1.0854 | 0.0662 |
| 1324 | 0.1182 | 0.0834 | 1.2924 | 0.0798 | 0.1126 | 0.0831 | 1.4293 | 0.0811 |
| 1345 | 0.0829 | 0.0550 | 0.9611 | 0.0522 | 0.0764 | 0.0558 | 0.9611 | 0.0534 |
| 3025 | 0.1795 | 0.1605 | 0.7865 | 0.1134 | 0.1700 | 0.1291 | 0.4886 | 0.1107 |
| 3050 | 0.1473 | 0.1068 | 1.4033 | 0.0838 | 0.1623 | 0.1009 | 0.5636 | 0.0806 |
| avg | 0.1219 | 0.0936 | 1.1232 | 0.0785 | 0.1207 | 0.0873 | 1.0369 | 0.0786 |
| med | 0.1128 | 0.0857 | 1.1268 | 0.0803 | 0.1150 | 0.0835 | 1.0233 | 0.0802 |

(2) 针对于中期海表温预测，本文提出使用基于 XGBoost 结合 LSTM 的中期海表温预测模型，实验结果如图 5.5。

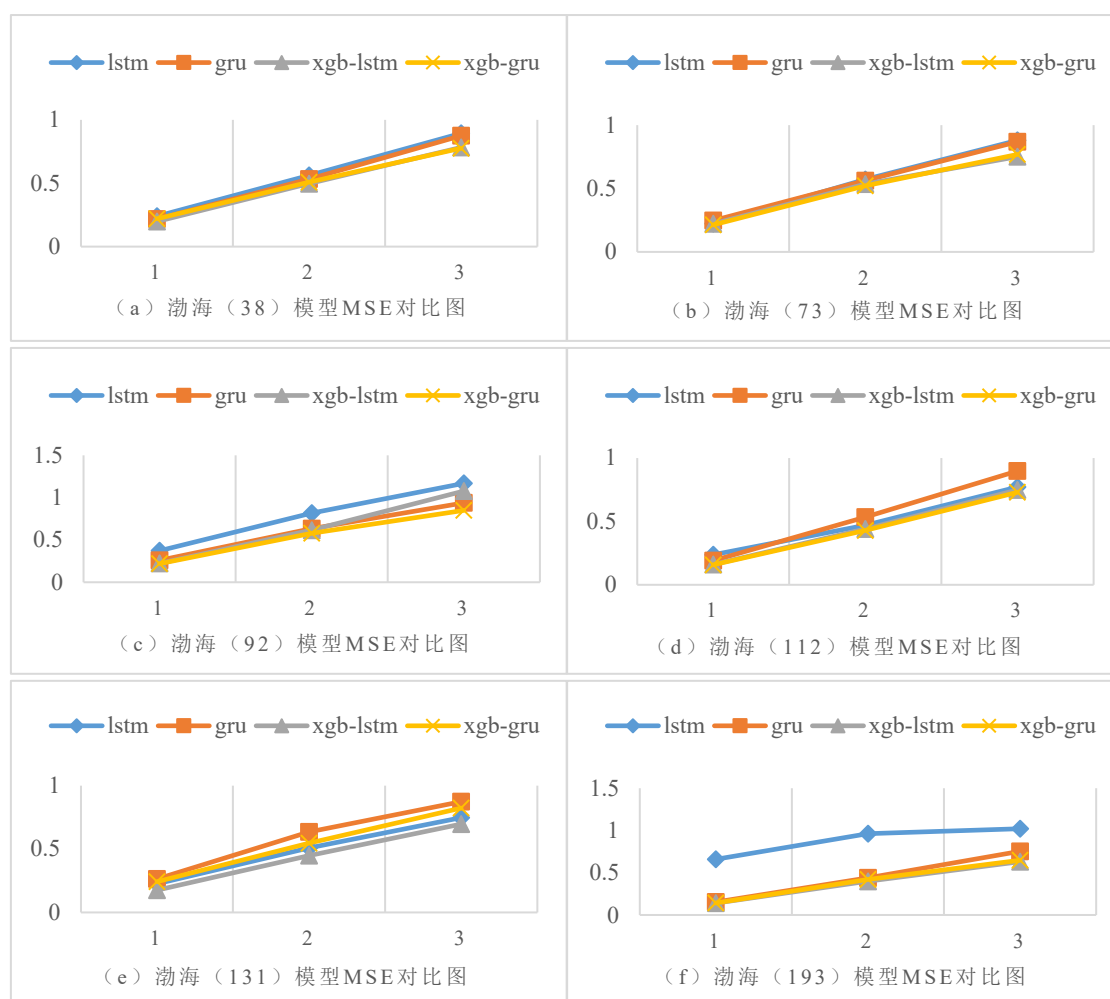


图 5.5 渤海 6 个观测点 MSE 对比图

Fig. 5.5 BoHai 6 position MSE contrast chart

图 5.5 中横轴代表预测的第几步，也就是第几天的预测，纵轴代表 MSE 的值。由图 5.5 可看出，在渤海海域选定的 6 个观测点中，38、73、112、131、193 这 5 个观测点上的预测误差最小的是 XGB-LSTM 模型。在 92 观测点，XGB-LSTM 模型在前两天的预测误差小于其他模型，第三天的预测误差大于 GRU、XGB-GRU 模型，但小于 LSTM 模型。从曲线的走向可以看出，随着预测步长的增加，预测的误差呈增大的趋势，但 XGB-LSTM 模型的误差增长比其他模型要慢。整体来看 XGB-LSTM 模型在渤海数据集上的拟合能力比其他模型更好。

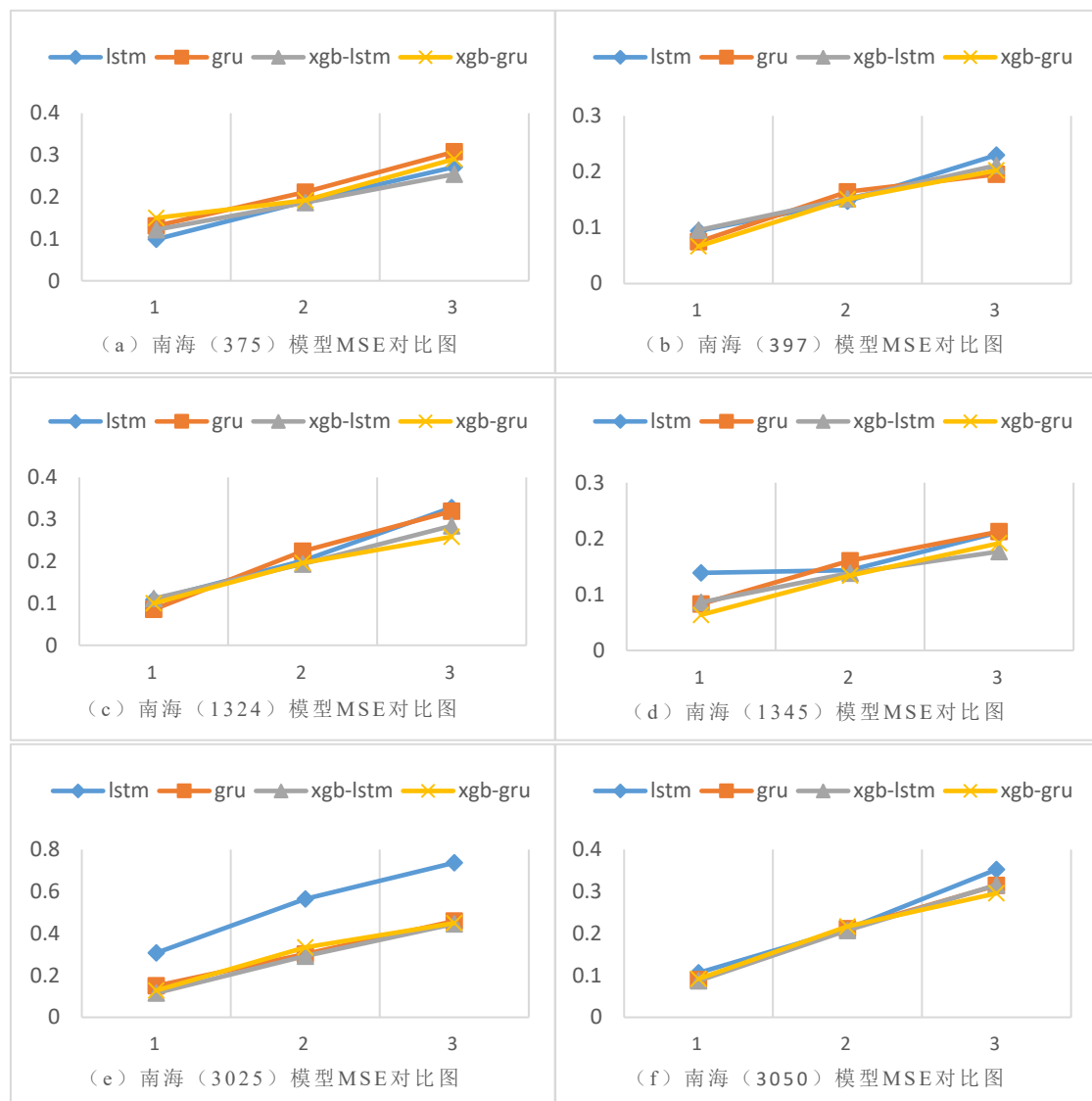


图 5.6 南海 6 个位置 MSE 对比图

Fig. 5.6 NanHai 6 position MSE contrast chart

图 5.6 中横轴代表预测的第几步，也就是第几天的预测，纵轴代表 MSE 的值。由图 5.6 可看出，在南海数据集，整体上可以看出 XGB-LSTM 模型的预测误差小于其他模型。只是在点 1345 上预测步长 1 的 MSE 和 GRU 模型的 MSE 相差不大，但比 LSTM 模型

的 MSE 要小。整体上看,随着预测步长的增加,所有模型的预测误差逐渐变大,LSTM 模型的波动较大。

表 5.6 渤海 6 个观测点 MSE 平均值

Table 5.6 MSE mean value of 6 positions in BoHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>XGB-LSTM</i> | <i>XGB-GRU</i> |
|----|-------------|------------|-----------------|----------------|
| 1 | 0.3263 | 0.2221 | 0.1869 | 0.1990 |
| 2 | 0.6481 | 0.5561 | 0.4889 | 0.5003 |
| 3 | 0.9140 | 0.8678 | 0.7816 | 0.7656 |

表 5.7 渤海 6 个观测点 RMSE 平均值

Table 5.7 RMSE mean value of 6 positions in BoHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>XGB-LSTM</i> | <i>XGB-GRU</i> |
|----|-------------|------------|-----------------|----------------|
| 1 | 0.5578 | 0.4693 | 0.4309 | 0.4442 |
| 2 | 0.7979 | 0.7444 | 0.6974 | 0.7061 |
| 3 | 0.9531 | 0.9310 | 0.8808 | 0.8742 |

表 5.8 渤海 6 个观测点 MAE 平均值

Table 5.8 MAE mean value of 6 positions in BoHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>XGB-LSTM</i> | <i>XGB-GRU</i> |
|----|-------------|------------|-----------------|----------------|
| 1 | 0.4142 | 0.3325 | 0.3011 | 0.3074 |
| 2 | 0.5984 | 0.5449 | 0.5104 | 0.5127 |
| 3 | 0.7236 | 0.6883 | 0.6478 | 0.6417 |

表 5.9 渤海 6 个观测点 R^2 平均值

Table 5.9 R^2 mean value of 6 positions in BoHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>XGB-LSTM</i> | <i>XGB-GRU</i> |
|----|-------------|------------|-----------------|----------------|
| 1 | 0.9960 | 0.9972 | 0.9977 | 0.9975 |
| 2 | 0.9920 | 0.9930 | 0.9939 | 0.9937 |
| 3 | 0.9886 | 0.9891 | 0.9903 | 0.9904 |

表 5.6~表 5.9,在渤海数据集上,XGB-LSTM 模型的平均 MSE、平均 RMSE 是最小的,平均 MAE 在 1、2 步时是最小的,在第 3 步时 XGB-LSTM 的平均 MAE 是小于 LSTM、GRU 模型的,而 R^2 在 1、2 步时是最大的,第三步时,XGB-LSTM 的平均 R^2 是大于 LSTM、GRU 模型的。所以,综合来看 XGB-LSTM 在渤海数据集上的表现优于其他模型。

表 5.10~表 5.13，在南海数据集上，XGB-LSTM 模型的平均 MSE 在第 2 步长时是最小的，在第 1、3 步长时是小于 LSTM、GRU 的；而它的 RMSE、MAE、 R^2 都优于 LSTM、GRU。虽然 XGB-LSTM 并不是在每一步上表现最优，但综合来看 XGB-LSTM 在南海数据集上的表现优于其他模型。

表 5.10 南海 6 个观测点 MSE 平均值
Table 5.10 MSE mean value of 6 positions in NanHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>XGB-LSTM</i> | <i>XGB-GRU</i> |
|----|-------------|------------|-----------------|----------------|
| 1 | 0.1423 | 0.1024 | 0.1030 | 0.1000 |
| 2 | 0.2426 | 0.2121 | 0.1948 | 0.2037 |
| 3 | 0.3549 | 0.3013 | 0.2814 | 0.2813 |

表 5.11 南海 6 个观测点 RMSE 平均值
Table 5.11 RMSE mean value of 6 positions in NanHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>XGB-LSTM</i> | <i>XGB-GRU</i> |
|----|-------------|------------|-----------------|----------------|
| 1 | 0.3673 | 0.3172 | 0.3201 | 0.3124 |
| 2 | 0.4760 | 0.4579 | 0.4381 | 0.4462 |
| 3 | 0.5808 | 0.5435 | 0.5246 | 0.5249 |

表 5.12 南海 6 个观测点 MAE 平均值
Table 5.12 MAE mean value of 6 positions in NanHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>XGB-LSTM</i> | <i>XGB-GRU</i> |
|----|-------------|------------|-----------------|----------------|
| 1 | 0.2934 | 0.2365 | 0.2415 | 0.2309 |
| 2 | 0.3708 | 0.3478 | 0.3300 | 0.3436 |
| 3 | 0.4588 | 0.4189 | 0.4039 | 0.4079 |

表 5.13 南海 6 个观测点 R^2 平均值
Table 5.13 R^2 mean value of 6 positions in NanHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>XGB-LSTM</i> | <i>XGB-GRU</i> |
|----|-------------|------------|-----------------|----------------|
| 1 | 0.9230 | 0.9414 | 0.9354 | 0.9439 |
| 2 | 0.8823 | 0.8779 | 0.8903 | 0.8892 |
| 3 | 0.8214 | 0.8349 | 0.8549 | 0.8461 |

(3) 针对于长期海表温预测，本文提出使用基于 GRU 和 CNN-1D 结合注意力机制的预测模型。依然选择渤海、南海数据，仍然选择上文使用的 6 个观测点进行实验，实验结果如图 5.7。

表 5.14 渤海 6 个观测点 MSE 平均值

Table 5.14 MSE mean value of 6 positions in BoHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>C-LSTM-ATT</i> | <i>C-GRU-ATT</i> |
|----|-------------|---------------|-------------------|------------------|
| 1 | 0.3148 | 0.2248 | 0.2864 | 0.2461 |
| 2 | 0.6107 | 0.4957 | 0.5391 | 0.4842 |
| 3 | 0.8076 | 0.7268 | 0.8037 | 0.7063 |
| 4 | 1.0366 | 0.9148 | 0.8971 | 0.8904 |
| 5 | 1.2559 | 1.1115 | 1.0228 | 1.0387 |
| 6 | 1.4935 | 1.3033 | 1.1665 | 1.1830 |
| 7 | 1.6552 | 1.4334 | 1.3413 | 1.3147 |

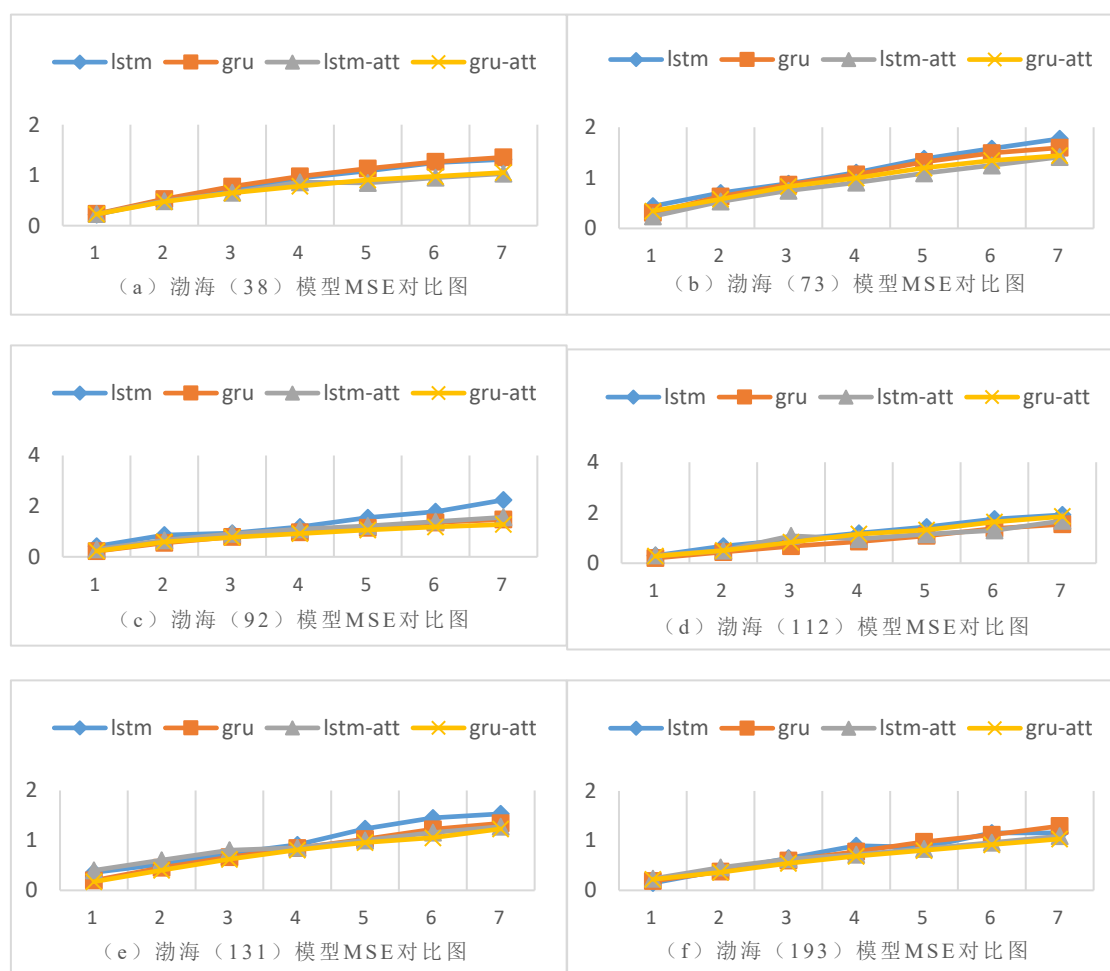


图 5.7 渤海 6 个观测点 MSE 对比图

Fig. 5.7 BoHai 6 position MSE contrast chart

图 5.7 中 *gru-att* 即本文提出的基于 GRU 和 CNN-1D 结合注意力机制的预测模型，同理 *lstm-att* 是 LSTM 和 CNN-1D 结合注意力机制的预测模型。从以上 6 个观测点的 MSE 走势可看出，无论哪个模型，随着预测步长的增加，MSE 逐渐增大。每个观测点

上结合了注意力机制的模型随步长的增加, MSE 增长更为缓慢, 说明注意力机制在长期预测中起到了作用。与此同时, 本文同时测试了 GRU 结合一维卷积进行预测, 但预测效果和 GRU 相差不大, 故此处并未将其作为一个对比实验。再添加了注意力机制后, 每个模型预测步长前 3 步的 MSE 相差不大, 从第 4 步开始, LSTM 模型和 GRU 模型的 MSE 明显比 LSTM-ATT 模型和 GRU-ATT 模型的 MSE 要大。在点 112 上, 四个模型的表现现在每个点相差不大。

图 5.8, 从南海区域 6 个观测点的实验结果可看出, 随预测步长的增加, MSE 整体呈增加趋势。结合注意力机制的模型的表现更加稳定些, 在 375、397、1324 三个位置点, 结合注意力机制的模型误差更小, 在其余三点上, 这四个模型的表现趋于一致。在点 397 上, GRU 模型 LSTM-ATT 模型的波动幅度大, 而 GRU-ATT 模型表现平稳。在点 375、1324、3025 上 LSTM 模型的均有小幅波动, 而添加注意力机制的 LSTM-ATT 模型、GRU-ATT 模型表现较为平稳。

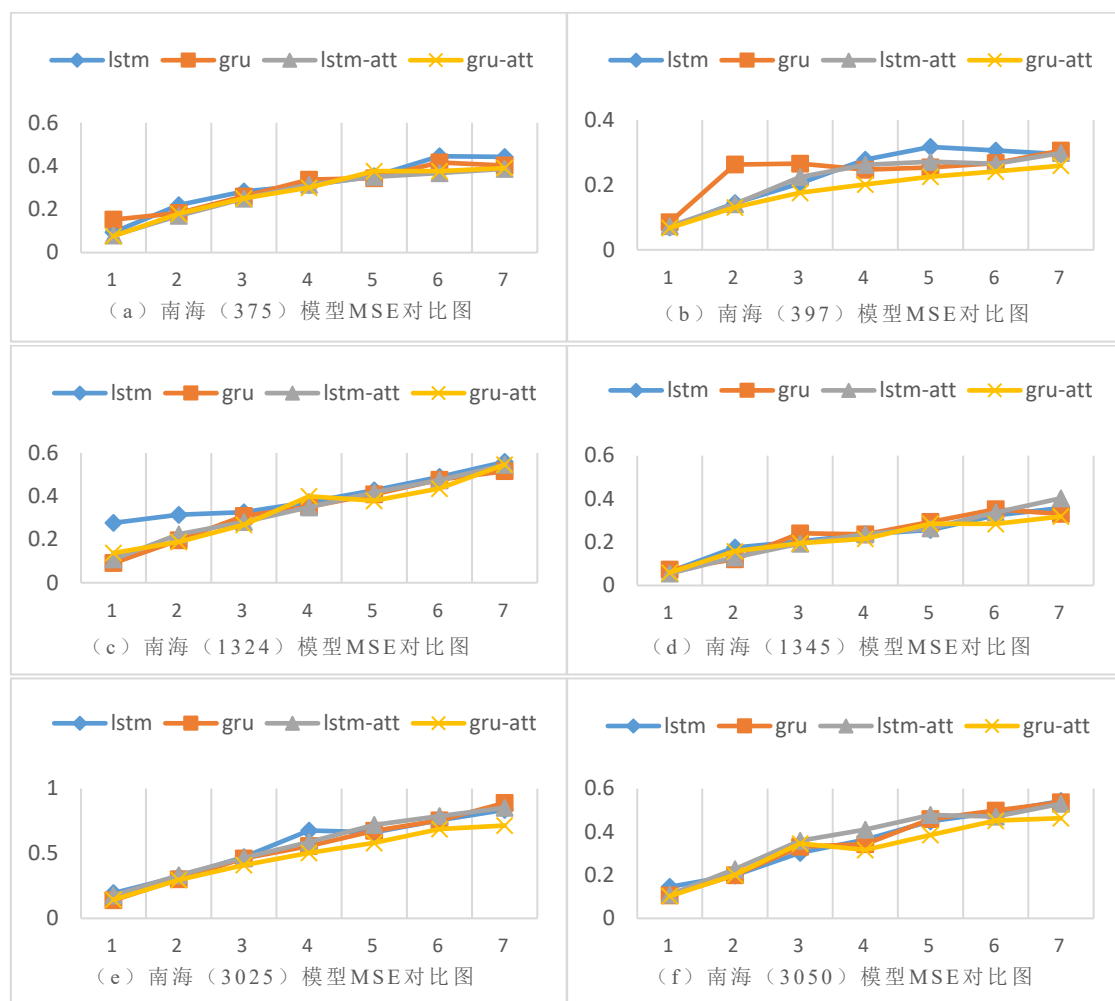


图 5.8 南海 6 个观测点 MSE 对比图

Fig. 5.8 NanHai 6 position MSE contrast chart

表 5.15 渤海 6 个观测点 RMSE 平均值
Table 5.15 RMSE mean value of 6 positions in BoHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>C-LSTM-ATT</i> | <i>C-GRU-ATT</i> |
|----|-------------|---------------|-------------------|------------------|
| 1 | 0.5524 | 0.4724 | 0.5324 | 0.4932 |
| 2 | 0.7756 | 0.7017 | 0.7329 | 0.6933 |
| 3 | 0.8965 | 0.8509 | 0.8924 | 0.8379 |
| 4 | 1.0164 | 0.9552 | 0.9453 | 0.9402 |
| 5 | 1.1158 | 1.0531 | 1.0088 | 1.0157 |
| 6 | 1.2182 | 1.1404 | 1.0773 | 1.0822 |
| 7 | 1.2786 | 1.1964 | 1.1536 | 1.1407 |

表 5.16 渤海 6 个观测点 MAE 平均值
Table 5.16 MAE mean value of 6 positions in BoHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>C-LSTM-ATT</i> | <i>C-GRU-ATT</i> |
|----|-------------|---------------|-------------------|------------------|
| 1 | 0.4240 | 0.3374 | 0.3827 | 0.3669 |
| 2 | 0.5875 | 0.5170 | 0.5451 | 0.5085 |
| 3 | 0.6744 | 0.6236 | 0.6716 | 0.6193 |
| 4 | 0.7550 | 0.6959 | 0.7040 | 0.6888 |
| 5 | 0.8244 | 0.7679 | 0.7520 | 0.7429 |
| 6 | 0.9058 | 0.8373 | 0.8190 | 0.7955 |
| 7 | 0.9513 | 0.8825 | 0.8771 | 0.8467 |

表 5.17 渤海 6 个观测点 R^2 平均值
Table 5.17 R^2 mean value of 6 positions in BoHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>C-LSTM-ATT</i> | <i>C-GRU-ATT</i> |
|----|-------------|---------------|-------------------|------------------|
| 1 | 0.9959 | 0.9971 | 0.9964 | 0.9968 |
| 2 | 0.9923 | 0.9937 | 0.9932 | 0.9939 |
| 3 | 0.9897 | 0.9908 | 0.9898 | 0.9910 |
| 4 | 0.9869 | 0.9884 | 0.9887 | 0.9886 |
| 5 | 0.9841 | 0.9858 | 0.9870 | 0.9867 |
| 6 | 0.9811 | 0.9834 | 0.9852 | 0.9848 |
| 7 | 0.9791 | 0.9818 | 0.9830 | 0.9831 |

表 5.14~表 5.17 中, C-LSTM-ATT 表示基于 LSTM 和 CNN-1D 结合注意力机制的预测模型, C-GRU-ATT 表示基于 GRU 和 CNN-1D 结合注意力机制的预测模型。渤海 6

个观测点，C-GRU-ATT 模型在步长 2、3、4、7 时平均 MSE、平均 RMSE 均最小，C-LSTM-ATT 模型平均 MSE、平均 RMSE 在步长 5、6 时最小。平均 MAE 中，C-GRU-ATT 表现最好。平均 R^2 上，结合的模型表现优于单个的模型。GRU 模型在步长为 1 时比其他模型的表现要好，然而其他步长下，LSTM 模型和 GRU 模型的表现较差，而 C-LSTM-ATT 模型和 C-GRU-ATT 模型的表现较好。C-LSTM-ATT 模型和 C-GRU-ATT 模型相比，C-LSTM-ATT 在步长 4、5、6 时表现较好，而 C-GRU-ATT 模型在步长为 2、3、7 时表现较好。所以，添加注意力机制的模型的表现优于没有添加注意力机制的模型。说明注意力机制起到了减少误差的作用。整体上来看，C-GRU-ATT 模型的表现要优于其他模型。

表 5.18~表 5.21 是南海 6 个观测点每步长 MSE、RMSE、MAE、 R^2 的平均值（小数点后保留四位）。

表 5.18 南海 6 个观测点 MSE 平均值
Table 5.18 MSE mean value of 6 positions in NanHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>C-LSTM-ATT</i> | <i>C-GRU-ATT</i> |
|----|-------------|------------|-------------------|------------------|
| 1 | 0.1406 | 0.1078 | 0.1003 | 0.0979 |
| 2 | 0.2286 | 0.2104 | 0.2037 | 0.1921 |
| 3 | 0.2983 | 0.3109 | 0.2952 | 0.2740 |
| 4 | 0.3726 | 0.3448 | 0.3596 | 0.3225 |
| 5 | 0.4106 | 0.4051 | 0.4165 | 0.3719 |
| 6 | 0.4680 | 0.4604 | 0.4500 | 0.4125 |
| 7 | 0.5051 | 0.4963 | 0.5024 | 0.4485 |

表 5.19 南海 6 个观测点 RMSE 平均值
Table 5.19 RMSE mean value of 6 positions in NanHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>C-LSTM-ATT</i> | <i>C-GRU-ATT</i> |
|----|-------------|------------|-------------------|------------------|
| 1 | 0.3617 | 0.3254 | 0.3111 | 0.3083 |
| 2 | 0.4731 | 0.4541 | 0.4454 | 0.4346 |
| 3 | 0.5403 | 0.5540 | 0.5372 | 0.5177 |
| 4 | 0.6009 | 0.5810 | 0.5925 | 0.5606 |
| 5 | 0.6333 | 0.6280 | 0.6351 | 0.6033 |
| 6 | 0.6762 | 0.6697 | 0.6604 | 0.6330 |
| 7 | 0.7006 | 0.6924 | 0.6987 | 0.6605 |

表 5.20 南海 6 个观测点 MAE 平均值
Table 5.22 MAE mean value of 6 positions in NanHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>C-LSTM-ATT</i> | <i>C-GRU-ATT</i> |
|----|-------------|------------|-------------------|------------------|
| 1 | 0.2847 | 0.2484 | 0.2288 | 0.2341 |
| 2 | 0.3679 | 0.3487 | 0.3400 | 0.3360 |
| 3 | 0.4243 | 0.4407 | 0.4174 | 0.4009 |
| 4 | 0.4785 | 0.4562 | 0.4668 | 0.4333 |
| 5 | 0.5010 | 0.4950 | 0.4980 | 0.4695 |
| 6 | 0.5360 | 0.5333 | 0.5158 | 0.4911 |
| 7 | 0.5524 | 0.5501 | 0.5462 | 0.5158 |

表 5.21 南海 6 个观测点 R^2 平均值
Table 5.21 R^2 mean value of 6 positions in NanHai sea

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>C-LSTM-ATT</i> | <i>C-GRU-ATT</i> |
|----|-------------|------------|-------------------|------------------|
| 1 | 0.9266 | 0.9388 | 0.9479 | 0.9466 |
| 2 | 0.8707 | 0.8710 | 0.8920 | 0.8938 |
| 3 | 0.8380 | 0.8214 | 0.8413 | 0.8546 |
| 4 | 0.8012 | 0.8117 | 0.8073 | 0.8255 |
| 5 | 0.7763 | 0.7854 | 0.7836 | 0.7984 |
| 6 | 0.7475 | 0.7545 | 0.7624 | 0.7835 |
| 7 | 0.7319 | 0.7414 | 0.7294 | 0.7585 |

表 5.18~表 5.21 中，C-LSTM-ATT 表示基于 LSTM 和 CNN-1D 结合注意力机制的预测模型，C-GRU-ATT 表示基于 GRU 和 CNN-1D 结合注意力机制的预测模型。南海 6 个观测点平均 MSE、平均 RMSE、平均 MAE、平均 R^2 ，C-GRU-ATT 表现明显优于其他模型。综上所述，针对长期海表温预测，C-GRU-ATT 模型的表现更好。

5.4 局域海表温预测

5.4.1 实验相关设置

本章的实验是采用基于 Python 的深度学习框架 Keras 和 sklearn 机器学习库进行实现的，所有的网络权重参数均采用 Keras 默认的初始化器进行初始化。实验的数据集使用渤海和南海日平均海表温数据，渤海数据选取经纬度范围为：37.07N~41N，117.35E~121.10E；南海数据选取经纬度范围 4N~21N，105E~118E。时间从 1982 年 1 月 1 日到 2018 年 12 月 31 日，总共 13514 条；训练集时间范围是 1982 年 1 月 1 日到 2017 年 12 月 31 日，总共 13149 条；测试集时间范围是 2018 年 1 月 1 日到 2018 年 12 月 31

日，总共 365 条数据。为了减少随机因素的影响，本节所有结果都是将模型运行 3 次，然后取这些结果的平均值。作为最终结果。

短期预测 CNN-LSTM 模型设置：第 1 层卷积设置 8 个过滤器，卷积核大小 5×5 ，移动步长 1，填充方式 SAME，激活函数 RELU；第 1 层池化大小设置为 2×2 ，填充方式 SAME。第 2 层卷积设置 16 个过滤器，卷积核大小设置为 3×3 ，移动步长 1，填充方式 SAME，激活函数 RELU；第 2 层池化层大小设置为 2×2 ，填充方式 SAME；1 层隐藏层，隐藏层单元设置为 5，激活函数 RELU；损失函数设置为 MSE，优化器设置为 Adam；训练轮数设置为 100，批大小设置为 20；滑动窗口设置为 4，预测步长设置为 1。

中期预测 CNN-LSTM 模型设置：第 1 层卷积设置 32 个过滤器，卷积核大小 5×5 ，移动步长 1，填充方式 SAME，激活函数 RELU；第 1 层池化大小设置为 2×2 ，填充方式 SAME。第 2 层卷积设置 64 个过滤器，卷积核大小设置为 3×3 ，移动步长 1，填充方式 SAME，激活函数 RELU；第 2 层池化层大小为 2×2 ，填充方式 SAME；1 层隐藏层，隐藏层单元设置为 15，激活函数 RELU；损失函数设置为 MSE，优化器设置为 Adam；训练轮数设置为 100，批大小设置为 20；滑动窗口设置为 12，预测步长设置为 3。

长期预测 CNN-LSTM-ATT 模型设置：第 1 层卷积设置 32 个过滤器，卷积核大小 5×5 ，移动步长 1，填充方式 SAME，激活函数 RELU；第 1 层池化大小设置为 2×2 ，填充方式 SAME。第 2 层卷积设置 64 个过滤器，卷积核大小设置为 3×3 ，移动步长设置为 1，填充方式 SAME，激活函数 RELU；第 2 层池化层大小设置为 2×2 ，填充方式 SAME；Droupout 层 rate 设置为 0.25；1 层隐藏层，隐藏层单元设置为 15，激活函数 RELU，return_sequences 设置为 TRUE；损失函数设置为 MSE，优化器设置为 Adam；训练轮数设置为 100，批大小设置为 20；滑动窗口设置为 28，预测步长设置为 7。

5.4.2 实验结果及分析

(1) 针对于短期局域海表温预测结果如图 5.9。

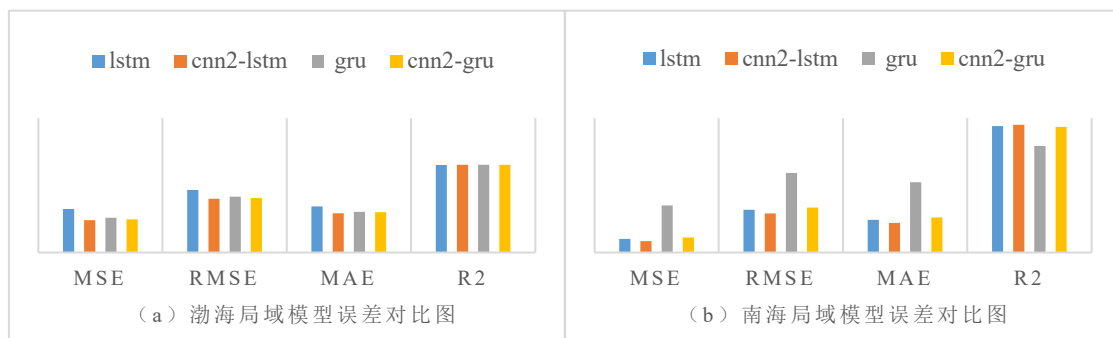


图 5.9 预测误差对比

Fig. 5.9 Comparison of prediction errors

图 5.9（1）是渤海局域模型误差对比图，其中 cnn2-lstm 表示基于 LSTM 结合二维卷积的预测模型，cnn2-gru 表示基于 GRU 结合二维卷积的预测模型。图中 cnn2-lstm 模型在渤海局部海域的 MSE、RMSE、MAE 均小于其他模型，表示 cnn2-lstm 模型的预测误差更小，模型的拟合能力更强；从 R^2 上看，四种模型相差不大，cnn2-lstm 的得分最高，这也从侧面表明模型的拟合能力更好。

图 5.9（2）是南海局域模型误差对比图，cnn2-lstm 模型的 MSE、RMSE、MAE 的值均小于其他模型， R^2 的得分也高于其他模型，表明 cnn2-lstm 模型在南海局域海域的表现更好。

由上可以看出，在短期海表温预测上，cnn2-lstm 模型无论在渤海海域还是南海海域，其拟合效果优于其他模型。

（2）针对于中期局域海表温预测实验结果如图 5.10。

图 5.10（1）是四种模型在渤海海域的 MSE 对比图，图 5.10（2）是四种模型在南海海域的 MSE 对比图。无论是渤海海域还是南海海域，可以很明显看出，随着预测步长的增加，每个模型预测的 MSE 都是呈上升的趋势；而四个模型中 cnn2-lstm、cnn2-gru 两个模型的 MSE 要小于其他模型，说明 cnn2-lstm、cnn2-gru 的拟合效果更好些。

表 5.22 渤海局域模型误差对比

Table 5.22 Error comparison of BoHai regional model

| 模型 | 步长 | RMSE | MAE | R^2 |
|-----------|----|---------------|---------------|---------------|
| LSTM | 1 | 0.6665 | 0.5030 | 0.9783 |
| | 2 | 0.7472 | 0.5567 | 0.9767 |
| | 3 | 0.8516 | 0.6250 | 0.9743 |
| GRU | 1 | 0.6095 | 0.4489 | 0.9792 |
| | 2 | 0.7414 | 0.5465 | 0.9767 |
| | 3 | 0.8631 | 0.6309 | 0.9741 |
| CNN2-LSTM | 1 | 0.6064 | 0.4454 | 0.9792 |
| | 2 | 0.7292 | 0.5385 | 0.9769 |
| | 3 | 0.8291 | 0.6105 | 0.9748 |
| CNN2-GRU | 1 | 0.6077 | 0.4449 | 0.9792 |
| | 2 | 0.7453 | 0.5517 | 0.9767 |
| | 3 | 0.8360 | 0.6159 | 0.9747 |

从表 5.22 可看出随着预测步长的增加，RMSE、MAE 逐渐增大， R^2 逐渐减小。在每一步长中，CNN2-LSTM 模型的表现较好，RMSE 在每个步长中均最小，MAE 在第 2、

3 步长时最小, R^2 的得分在每个步长中最高。综合来看, 在中期海表温预测上, CNN2-LSTM 模型的拟合效果更好。

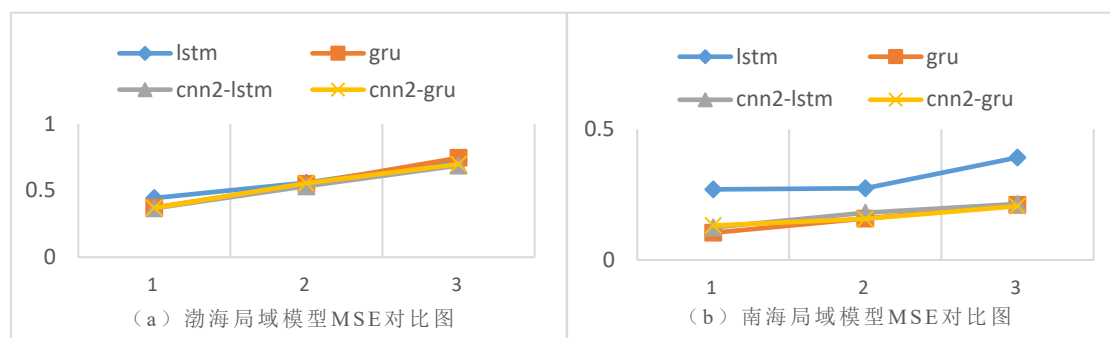


图 5.10 预测误差对比

Fig. 5.10 Comparison of prediction errors

表 5.23 南海局域模型误差对比

Table 5.23 Error comparison of NanHai regional model

| 模型 | 步长 | $RMSE$ | MAE | R^2 |
|-----------|----|---------------|---------------|---------------|
| LSTM | 1 | 0.5193 | 0.4483 | 0.8402 |
| | 2 | 0.5235 | 0.4348 | 0.8402 |
| | 3 | 0.6257 | 0.5250 | 0.7710 |
| GRU | 1 | 0.3225 | 0.2452 | 0.9395 |
| | 2 | 0.3975 | 0.3081 | 0.9080 |
| | 3 | 0.4589 | 0.3583 | 0.8770 |
| CNN2-LSTM | 1 | 0.3523 | 0.2781 | 0.9266 |
| | 2 | 0.4247 | 0.3339 | 0.8946 |
| | 3 | 0.4630 | 0.3650 | 0.8749 |
| CNN2-GRU | 1 | 0.3628 | 0.2911 | 0.9223 |
| | 2 | 0.3959 | 0.3105 | 0.9074 |
| | 3 | 0.4531 | 0.3585 | 0.8795 |

从表 5.23 可看出, 在南海区域的数据集上, GRU、CNN2-GRU 模型的表现更好些, 而 CNN2-LSTM 模型的表现是优于 LSTM 的表现的。

综合来看, 在中期海表温预测上, 四个模型中, 结合 CNN 的模型的拟合效果更好些。

(3) 针对长期海表温预测实验结果如图 5.11。

图 5.11, 随着预测步长的增加, 误差逐渐增大。在渤海数据集上, LSTM 模型、GRU 模型、C2-LSTM-ATT 模型和 C2-GRU-ATT 模型的趋势几乎一样, 且在每步长的表现相

差不多，但 C2-LSTM-ATT 模型 MSE 更小些。在南海数据集上，LSTM 模型的波动最大，C2-LSTM-ATT 模型的波动其次，C2-GRU-ATT 模型表现较好。LSTM 模型、GRU 模型、C2-LSTM-ATT 模型和 C2-GRU-ATT 模型的 RMSE 变化趋势和 MSE 一致。在渤海，四种模型的 MAE 中，C2-LSTM-ATT 模型在每一步的 MAE 更小，但四种模型的变化趋势基本一致，呈随步长增加的状态。

综合看来，四种模型在每种步长下的误差差别不大，但 C2-LSTM-ATT、C2-GRU-ATT 模型的表现比 LSTM 模型和 GRU 模型的表现更加稳定，在 MSE、RMSE、MAE 上比其他模型的误差小一些。

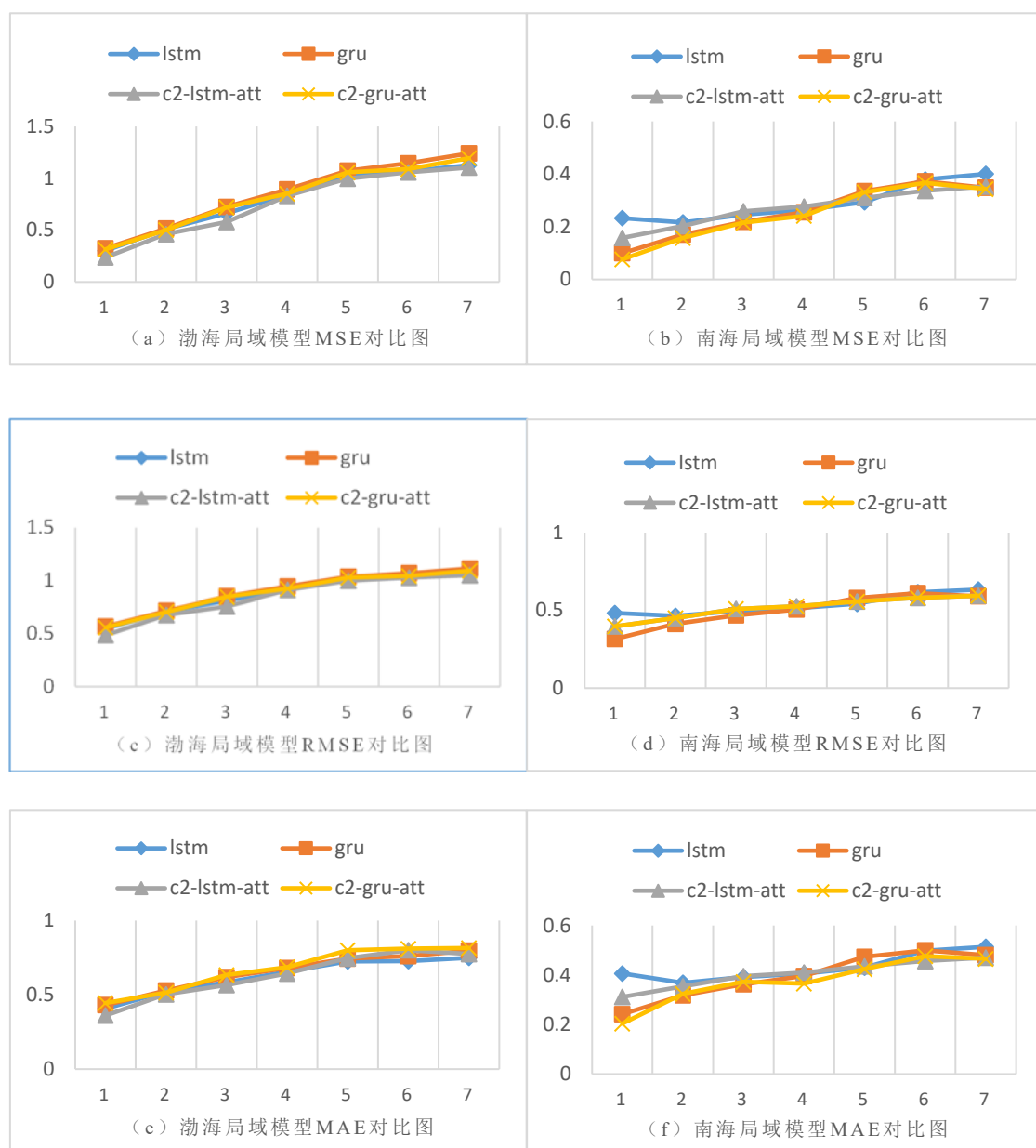


图 5.11 预测误差对比

Fig. 5.11 Comparison of prediction errors

表 5.24 渤海局域 R^2
Table 5.24 BoHai area R^2

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>C2-LSTM-ATT</i> | <i>C2-GRU-ATT</i> |
|----|-------------|------------|--------------------|-------------------|
| 1 | 0.9801 | 0.9798 | 0.9971 | 0.9934 |
| 2 | 0.9773 | 0.9771 | 0.9944 | 0.9902 |
| 3 | 0.9752 | 0.9743 | 0.9906 | 0.9822 |
| 4 | 0.9726 | 0.9719 | 0.9890 | 0.9805 |
| 5 | 0.9699 | 0.9694 | 0.9837 | 0.9802 |
| 6 | 0.9695 | 0.9684 | 0.9892 | 0.8714 |
| 7 | 0.9686 | 0.9671 | 0.9820 | 0.8354 |

表 5.25 南海局域 R^2
Table 5.25 NanHai area R^2

| 步长 | <i>LSTM</i> | <i>GRU</i> | <i>C2-LSTM-ATT</i> | <i>C2-GRU-ATT</i> |
|----|-------------|---------------|--------------------|-------------------|
| 1 | 0.8691 | 0.9429 | 0.9086 | 0.9944 |
| 2 | 0.8774 | 0.9026 | 0.8836 | 0.9607 |
| 3 | 0.8588 | 0.8739 | 0.8504 | 0.9504 |
| 4 | 0.8475 | 0.8527 | 0.8393 | 0.8313 |
| 5 | 0.8295 | 0.8077 | 0.8298 | 0.8067 |
| 6 | 0.7806 | 0.7859 | 0.8038 | 0.7689 |
| 7 | 0.7672 | 0.7974 | 0.7941 | 0.6853 |

从表 5.24、表 5.25 可观察到，本文提出的模型 C2-LSTM-ATT 的 R^2 得分更高，表明该模型的表现更好，这表明通过使用二维卷积来提取数据空间特征，然后使用注意力机制来修正预测结果这种方式能够提高模型的拟合能力。

5.5 本章小结

本章是验证提出的模型的有效性。首先是进行单点海表温预测的实验验证，然后是进行局域海表温的实验验证。

针对单点海表温预测模型，首先验证基于 XGBoost 的短期海表温预测模型。该模型在数据处理上考虑“邻域点”对预测点的影响，利用 XGBoost 模型的并行优化、损失精确、防止过拟合等优点，实现了对短期海表温的准确预测。实验中在未考虑“邻域点”的影响因素时，XGBoost 模型的表现优于对比模型；在考虑“邻域点”后，每种模型的表现均优于原来的表现，而 XGBoost 的预测结果更小，说明在短期海表温预测中，XGBoost 的效果更好，同时说明了，“邻域点”的影响不可忽视。然后验证基于 XGBoost

结合 LSTM 的中期海表温预测模型，该模型利用 XGBoost 来提取观测点数据的时间特征，并结合周围邻域点的历史数据，作为 LSTM 模型的输入，充分利用 LSTM 模型具有提取长短期特征和多步输出的优势实现了海表温的中期预测。实验结果表明，基于 XGBoost 结合 LSTM 的中期海表温模型的误差更小。最后验证基于 GRU 和一维卷积（CNN-1D）结合注意力机制的预测模型。该模型首先采用 CNN-1D 对历史数据进行卷积操作，提取数据特征并实现对历史数据的压缩，从而降低后续处理的复杂度；然后再使用 GRU 模型进行训练，提取序列的时序特征；最后采用注意力机制，为 GRU 每一步的输出设定一个合理的权重，修正 GRU 输出结果。实验表明，该模型的拟合效果优于现有模型。

针对局域海表温预测模型，所提模型的预测误差小于现有模型的预测误差，尤其是长期海表温预测模型的表现更加优异。

第6章 总结与展望

海表温 (SST) 是海洋环境的重要因子, 它不仅在气候、海气相互作用等研究中占有重要位置, 而且在近几年提出的海洋生态环境保护及预报的研究中也有重要研究价值。对海表温的研究, 尤其是我国临海海域的 SST 研究, 将为发展我国海洋渔业生产和开发海洋生物资源以及海洋生态保护及其动态监测提供极为重要的辅助作用。为此, 本文基于海表温大数据, 结合深度学习技术对海表温预测技术进行了深入研究, 本文将海表温预测问题划分为单点海表温预测和局域海表温预测两大类, 每一类又根据预测步长分别提出了短期、中期、长期三类预测模型, 通过实验验证了所提出模型的有效性。

6.1 总结

围绕海表温预测技术的研究, 本文完成了如下工作:

(1) 提出了一种基于 XGBoost 的单点短期海表温预测模型。该模型在数据处理上考虑“邻域点”对预测点的影响, 利用 XGBoost 模型的并行优化、损失精确、防止过拟合等优点, 实现了对短期海表温的准确预测; 在此基础上, 提出了基于 XGBoost 结合 LSTM 的中期海表温预测模型, 该模型利用 XGBoost 来提取观测点数据的时间特征, 并结合周围邻域点的历史数据, 作为 LSTM 模型的输入, 充分利用 LSTM 模型具有提取长短期特征和多步输出的优势实现了海表温的中期预测。

(2) 针对单点长期海表温预测, 提出了基于 GRU 和一维卷积 (CNN-1D) 结合注意力机制的预测模型。首先采用 CNN-1D 对历史数据进行卷积操作, 提取数据特征并实现对历史数据的压缩, 从而降低后续处理的复杂度; 然后再使用 GRU 模型进行训练, 提取序列的时序特征; 最后采用注意力机制, 为 GRU 的输出向量设定一个合理的权重, 利用所求权重, 对 GRU 输出结果进行加权求和, 最后通过全连接层将数据映射为最终的预测结果。注意力机制在单点长期海表温预测中, 起到修正 GRU 输出结果的作用, 从而进一步提高长期预测的精度。

(3) 针对局域海表温预测问题, 提出基于二维卷积 (CNN-2D) 和 LSTM 融合注意力机制的长期预测模型。首先将局域同一时刻的数据看作是一张通道数为1的二维图像。针对短期、中期海表温预测, 采用 CNN-2D 结合 LSTM 进行预测。CNN-2D 来提取观测点与观测点之间数据的空间特征, 然后使用 LSTM 学习数据的时序特征, 最后进行训练预测; 针对长期预测, 提出基于 CNN-2D 和 LSTM 融合注意力机制的模型, 学习到数据的时空特征以后, 通过注意力机制来对结果进行修正, 从而减小预测的误差。

(4) 分别选取渤海海域和南海海域的观测点的温度数据，从模型的预测精度和泛化能力等方面对提出的上述几种模型进行了实验验证。实验结果表明本文提出的几种预测模型具有较小的预测误差，特别是中长期预测模型的精度均优于现有的预测模型。

6.2 进一步的工作

从目前的实验结果表现来看，基于循环神经网络的海表温预测模型在海表温预测问题上有一定的优势，但是本文的工作仍存在着较大的改进空间：

(1) 数据处理上的改进

本文主要是使用历史温度来预测未来温度，属于自回归范畴。训练特征相对来说比较少，应该再加入例如洋流、太阳辐射等数据，然而各种数据在地区或经纬度上并不统一，数据的选取有一定难度。对于邻域点的处理，由于经纬度划分的区域对应到地球表面上，其实两个点的距离相距很大，点的温度间的影响情况无法确定，应该添加一个对应的“影响因子”，更好的刻画点的温度间的影响。

(2) 注意力机制

本文添加注意力机制是为了缓和预测步长增大时误差增大这个问题。但是注意力机制本身的实现有多种方式，具体如何做才能起到更好的效果，这个是一个值得研究的方向。

(3) 神经网络训练时间较长。

如何选择恰当的网络参数，使得误差不增大的同时减少网络的训练时间，这是一个值得研究的方向。

(4) 数据选取

本文使用的是日平均海表温数据，时间分辨率较低，今后可以进一步提高时间分辨率，采用“小时”单位的数据，以便更好反应和发现海表温变化规律。

希望在未来工作中能从以上角度研究，对现有的模型进行进一步的更新。

参考文献

- [1] Cheng L, Zhu J, Abraham J. 2018 Continues Record Global Ocean Warming[J]. *Advances in Atmospheric Sciences*, 2019, 36(3): 249-252.
- [2] Tao S S, Dong S, Wang Z F. Intensity Division of the Sea Ice Zones in China[J]. *Cold Regions Science and Technology*, 2018, 151: 179-187.
- [3] Chan F T, Stanislawczyk K, Sneekes A C. Climate Change Opens New Frontiers for Marine Species in the Arctic: Current Trends And Future Invasion Risks[J]. *Global Change Biology*, 2019, 25(1): 25-38.
- [4] Wijaya P, Putra T, Hidayat F. Suitability Analysis For Scuba Diving To Develop Marine Tourism At Saebus Island, East Java, Indonesia[C]. In *Proceedings of the 2018 E3S Web of Conferences*. EDP Sciences, 2018, 31: 08020.
- [5] Solanki H U, Bhatpuria D, Chauhan P. Integrative analysis of AltiKa-SSHa, MODIS-SST, and OCM-chlorophyll Signatures for Fisheries Applications[J]. *Marine Geodesy*, 2015, 38(sup1): 672-683.
- [6] 阳宁, 王英杰. 海底矿产资源开采技术研究动态与前景分析[J]. *矿业装备*, 2012 (1): 54-57.
- [7] Yu Y, Zhang H R, Jin J. Trends of Sea Surface Temperature and Sea Surface Temperature Fronts in the South China Sea During 2003–2017[J]. *Acta Oceanologica Sinica*, 2019, 38(4): 106-115.
- [8] Newman M, Alexander M A, Scott J D. An Empirical Model of Tropical Ocean Dynamics[J]. *Climate Dynamics*, 2011, 37(9-10): 1823.
- [9] Hui L, Shu-Kai Y, Yang Z. Investigation on Method for Estimating Sea Wave Power Spectrum Based on ARMA Model[J]. *Ocean Technology*, 2008: 96-98.
- [10] Duan W, Huang L, Han Y. A Hybrid EMD-AR Model for Nonlinear and Non-stationary Wave Forecasting[J]. *Journal of Zhejiang University-SCIENCE A*, 2016, 17(2): 115-129.
- [11] Shan L, Xing-Rong C, Yi C. Typhoon Generation Frequency Prediction Based on ARIMA Model[J]. *Marine Forecasts*, 2016: 53-58.
- [12] Ying Z, Yan-Chun T, Fa-Ding P. Studyon Time Series Prediction Model of Sea Surface Temperature Based on Ensemble Empirical Mode Decomposition and Autoregressive Integrated MovingAverage[J]. *Journal of Marine Sciences*, 2019: 9-14.

- [13] 朱坚真, 刘汉斌. 中国水产品价格波动分析——基于 ARCH 类模型[J]. 南方农村, 2012, 28(6):66-69.
- [14] 苏治, 方彤, 马景义. 一类包含不同权重函数的混频 GARCH 族模型及其应用研究[J]. 数量经济技术经济研究, 2018 (10): 8.
- [15] 周达军. 我国经济波动对增长的负面效应的实证分析[J]. 经济管理, 2007(14):77-82.
- [16] Yao P, Shen C, Lv X. A TVaR-EGARCH-POT Based Market Financing Risk Evaluation of the New OTC (Over the Counter) Market[C]. In Proceedings of the 2nd International Conference on Data Science and Business Analytics (ICDSBA), 2018: 334-339.
- [17] Thorlie M A, Song L, Amin M. Modeling and Forecasting of Stock Index Volatility with APARCH Models under Ordered Restriction[J]. Statistica Neerlandica, 2015, 69(3): 329-356.
- [18] Chaki S, Bathe R N, Ghosal S. Multi-objective Optimisation of Pulsed Nd: YAG Laser Cutting Process Using Integrated ANN-NSGAI Model[J]. Journal of Intelligent Manufacturing, 2018, 29(1): 175-190.
- [19] Vedrtnam A, Singh G, Kumar A. Optimizing Submerged Arc Welding Using Response Surface Methodology, Regression Analysis, and Genetic Algorithm[J]. Defence Technology, 2018, 14(3): 204-212.
- [20] Rafiei M, Niknam T, Aghaei J. Probabilistic Load Forecasting Using an Improved Wavelet Neural Network Trained by Generalized Extreme Learning Machine[J]. IEEE Transactions on Smart Grid, 2018, 9(6): 6961-6971.
- [21] Li X, Liu C F, Zhu X H. Integrated Assessment of Sea Water Quality Based on BP Artificial Neural Network[J]. Marine Science Bulletin, 2010, 29(2): 225-230.
- [22] HAN Z, ZHAO N. Seawater Temperature Model from Argo Data by LM-BP Neural Network in Northwest Pacific Ocean[J]. Marine Environmental Science, 2012 (4): 20.
- [23] Qiang L I, Xue-Hua Z. Application of BP Neural Network to the Prediction of loss on Ignition in Marine Sediments Based on Genetic Algorithm[J]. Metallurgical Analysis, 2019: 25-30.
- [24] ZHANG Y, GAO Q. Prediction Method of Chlorophyll-a Concentration in Seawater Based on Extreme Learning Machine Regression[J]. Marine Environmental Science, 2015 (1): 19.
- [25] Xiangpeng C, Hongqiong H. Ship Traffic Flow Prediction Model Based on GA-ELM

- Algorithm[J]. Microcomputer & Its Applications, 2017 (9): 5.
- [26] WUZhongqiang, Zhihua M, WANGZheng. Research on Remote Sensing Inversion of Shallow Water Depth Based on Extreme Learning Machine[J]. Hydrographic Surveying and Charting, 2019: 13-15.
- [27] Zhang Q, Wang H, Dong J. Prediction of Sea Surface Temperature Using Long Short-term Memory[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(10): 1745-1749.
- [28] Xingjian G U, Lu Z, Ming J. Research on Short-Term Prediction of Typical Trial Sea Environment in China Based on LSTM Neural Network[J]. Shipbuilding of China, 2017: 101-107.
- [29] Yang Y, Dong J, Sun X. A CFCC-LSTM Model for Sea Surface Temperature Prediction[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 15(2): 207-211.
- [30] 朱贵重, 胡松. 基于 LSTM-RNN 的海水表面温度模型研究[J]. 应用海洋学学报, 2019 (2019 年 02): 191-197.
- [31] 张莉, 罗鹏. 指数平滑模型在日本珍珠产量预测中的应用[J]. 海洋科学, 2009, 33(10): 59-62.
- [32] Ahmad N F Y, Khalid K, Rusiman M S. Analysis Monthly Import of Palm Oil Products Using Box-Jenkins Model[C]. In Proceedings of the 2018 Journal of Physics: Conference Series. IOP Publishing, 2018: 012012.
- [33] Allen T, Prospero P, Cogill B. A Delphi Approach to Develop Sustainable Food System Metrics[J]. Social Indicators Research, 2019, 141(3): 1307-1339.
- [34] Abbar S, Castillo C, Sanfilippo A. To Post or Not to Post: Using Online Trends to Predict Popularity of Offline Content[C]. In Proceedings of the 29th on Hypertext and Social Media (HT), 2018: 215-219.
- [35] Freund Y. Boosting a Weak Learning Algorithm by Majority[J]. Information and Computation, 1995, 121(2): 256-285.
- [36] Chen T, He T, Benesty M. Xgboost: Extreme Gradient Boosting[J]. R Package Version 0.4-2, 2015: 1-4.
- [37] Huang Z, Yi K. Communication-Efficient Weighted Sampling and Quantile Summary for GBDT[J]. arXiv preprint arXiv:1909.07633, 2019.
- [38] Choubin B, Zehtabian G, Azareh A. Precipitation Forecasting Using Classification and Regression Trees (CART) Model: A Comparative Study of Different Approaches[J].

- Environmental Earth Sciences, 2018, 77(8): 314.
- [39] Prabhakar S K, Rajaguru H. Adaboost Classifier with Dimensionality Reduction Techniques for Epilepsy Classification from EEG[M]. Precision Medicine Powered by pHHealth and Connected Health. Springer, Singapore, 2018: 185-189.
- [40] Dorogush A V, Ershov V, Gulin A. CatBoost: Gradient Boosting with Categorical Features Support[J]. arXiv preprint arXiv:1810.11363, 2018.
- [41] Zhang Q, Wang H, Dong J. Prediction of Sea Surface Temperature Using Long Short-term Memory[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(10): 1745-1749.
- [42] Guerrero-Pena F A, Fernandez P D M, Ren T I. Multiclass Weighted Loss for Instance Segmentation of Cluttered Cells[C]. In Proceedings of the 25th IEEE International Conference on Image Processing (ICIP), 2018: 2451-2455.
- [43] Li S, Li W, Cook C. Independently Recurrent Neural Network (Indrnn): Building a Longer and Deeper Rnn[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 5457-5466.
- [44] Peng Y, Tian S, Yu L. A Joint Approach to Detect Malicious URL Based on Attention Mechanism[J]. International Journal of Computational Intelligence and Applications, 2019, 18(03): 1950021.
- [45] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. arXiv preprint arXiv:1409.0473, 2014.
- [46] Chen H, Sun M, Tu C. Neural Sentiment Classification with User and Product Attention[C]. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2016: 1650-1659.
- [47] Wang L, Cao Z, De Melo G. Relation Classification via Multi-level Attention Cnns[C]. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL), 2016: 1298-1307.
- [48] Xu K, Ba J, Kiros R. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention[C]. In Proceedings of the 32nd International Conference on Machine Learning (ICML), 2015: 2048-2057.

致 谢

时光如白驹过隙，不经意间两年半的研究生生活就要结束了。在研究生这两年半的时光里，我的计算机理论知识得到了极大的丰富，在专业知识领域得到了长足的进步，在完成一些项目和实验的同时，我的实践能力也得到了极大进步。在那一次次寻找 bug 的过程中，在那一次次实验结果分析的过程中，我对于问题的分析能力提高了很多，看待问题也更加全面了，对于计算机的知识有了更加系统全面的认识。在研究生生活中，我不仅仅收获了专业技能，更有幸能结识一些有着共同追求的朋友和一些令人尊敬的师长。在他们潜移默化的影响下，我改变了自身的一些不足，形成了正确的人生观和价值观。借此毕业的机会，我向我的老师、同学、朋友们以及家人奉上我最由衷的感谢！

感谢我的导师乔百友老师在科研和论文上对我的指导，乔老师言传身教，用博学的知识和严谨的态度让我受益良多。乔老师在两年半的时间培养我们、指导我们、督促我们，每一次周会都会悉心为每一位同学考虑，严谨治学，同时还在生活上关心我们，让我不仅明白了在研究生阶段如何去培养自己的能力，如何步入工作和社会，同时还明白了许多为人处世的道理。感恩乔老师的悉心培养，在以后人生中，我也会让自己发光发热去给别人带来光明和温暖。

在此，我还要感谢实验室的韩东红老师、王波涛老师、吴刚老师、刘辉林老师，他们敬业的工作态度让我受益匪浅，平时从他们的言谈举止中学到了很多。还要感谢计算机科学系的各位老师，感谢老师们在相关课程中给予我的指导，在生活中给予的帮助与关怀。

感谢实验室的每位成员：马勇志、郭庆杰、郑皓文、吴忠强、张婧茹、汪锐、姚来刚、王昊东、霍玉起、康健、杨璐、胡兵和已经毕业的师兄师姐：吕鑫、陈薪、闫颖楠。感谢他们在学习和生活上对我的建议和帮助。

感谢我的家人，感谢他们对我不辞辛苦的付出，他们的爱是我一路向前的最大动力。

感谢我的室友，郭旭炆、尹少芊、姚佑博，感谢他们在我生活中给予的帮助。在生活上相互关心。

最后感谢所有在我攻读硕士期间帮助过我、关心过我的人们。最后祝愿我的母校东北大学能够披荆斩棘，再创辉煌！祝愿计算机科学与工程学院越来越好！

攻硕期间参与项目及获奖情况

参加项目

1. 国家重点研发计划“海洋环境安全保障”重点专项“海洋大数据分析预报技术研发”
(项目编号: 2016YFC1401900), 2018-03 至 2018-12

获奖情况

1. 获得 2017-2018 年度东北大学硕士研究生二等奖学金
2. 获得 2018-2019 年度东北大学硕士研究生二等奖学金
3. 获得 2019-2020 年度东北大学硕士研究生二等奖学金