

毕业设计(论文)开题报告

（适用于工科类、理科类专业）

|  |  |  |  |
| --- | --- | --- | --- |
| 课题名称 | 基于多源数据融合的台风强度和路径预报 | | |
| 副 标 题 | 无 | | |
| 学 院 | 软件学院 | | |
| 专 业 | 软件工程 | | |
| 学生姓名 | 梁厚 | 学 号 | 2051840 |
|  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 2023 | 年 | 3 | 月 | 21 | 日 |

一、毕业设计（论文）课题背景（含文献综述）

**1.1台风路径和强度预报研究现状**

台风是最危险的自然灾害之一，它们总是发生在西太平洋和西南太平洋，每年都会给太平洋沿岸带来经济和人身安全威胁。因此，相关领域的许多学者致力于寻找一种有效的方法来分析和预测台风路径，以防止灾害。以往的台风路径和强度预报有几种流行的技术类型，包括：① 动态建模方法，也称为数值预报方法（Numerical Forecast Method）。它主要使用动力学模型，利用数学方程来模拟台风运动。这些动力学方程往往过于复杂，需要超级计算机的数值方法获得近似解，因此成本较高；② 统计方法。它密切关注天气资源的统计数据而不是物理运动。这种方法的优点是比动态建模方法消耗更少的计算资源，同时可以应用于大多数国家和地区；③ 统计动态方法。这类方法结合了统计方法和动态建模方法。统计数据常常用于计算动力力学方程的更重要的初始值。预测结果主要取决于动态建模方法。

近年来，统计方法因其客观性、小计算量的特性受到广泛关注。本课题也将计划搭建统计方法相关的模型对台风的路径和强度进行预测。统计方法的关键问题在于如何定义或推导出良好的统计模型。

台风路径和强度的模型的训练常常需要用到各个气象中心提供的历史台风数据，包括二维的台风数据、卫星云图等。常用的公开数据集有，由中国气象局（CMA）维护的二维台风数据集、由欧洲中期天气预报中心（ECMWF）维护的3D台风数据集EAR-5，以及在热带气旋图像强度回归数据集网站上的台风强度卫星图像（TCIR）。CMA数据集提供了自 1949年以来 WNP 盆地中每 6 小时热带气旋的位置和强度，提供包括时间、经度、纬度、风速、中心气压的台风数据。EAR-5 数据集提供了自 1979 年以来收集的 14 种全球大气再分析数据。EAR-5的空间分辨率约为31公里，时间分辨率为每小时，数据通常以标准的气象学变量（如温度、湿度、风速等）以及衍生的气象学产品（如降水、云量、地面气压等）的形式提供。TCIR 收集了2003 年至 2017 年期间全球 1379个热带气旋的数据集，包括每 3 小时一次的卫星红外图像（IR）、水汽（WV）和被动微波降雨量（PMW）热带气旋图像，水平分辨率为 0.07度纬度/经度。

台风路径可以看作一种路径预测或轨迹预测问题，而用于路径预测的主要技术包括传统方法和深度学习方法。台风强度预测则往往使用深度学习的相关模型。以往的研究使用的方法有多元线性回归（MVLR），支持向量机（SVM），极限梯度上升（XGBoost）、包括长短期记忆（LSTM）和门控循环单元（GRU）循环神经网络（RNN），卷积神经网络（CNN），对抗神经网络（GAN）等。

因此，本文希望能够构建一个基于机器学习的多源数据融合的台风路径和强度预测的模型，对主要是太平洋地区中大气变量（台风中心经度纬度、台风中心气压、风速等）进行分析，训练一个具有良好性能的路径、强度预测模型，并在此基础上将输出的两个纬度结合起来运用台风路径相似度计算模型衡量模型性能的优劣。

**1.2本课题可能使用的深度学习模型**

1）AM-ConvGRU

AM-ConvGRU 是一种结合了注意力机制（Attention Mechanism）的卷积门控循环单元（Convolutional Gated Recurrent Unit，ConvGRU）的神经网络结构。它在 ConvGRU 的基础上引入了注意力机制，用于增强模型对输入数据的建模能力和对重要信息的关注程度。

ConvGRU 是一种改进的循环神经网络（RNN）单元，它在传统的循环神经网络基础上引入了门控机制，使得模型能够更好地捕捉序列数据中的长期依赖关系，并且由于引入了卷积操作，使得模型能够并行化处理序列数据，从而提高了训练效率。

AM-ConvGRU 在 ConvGRU 的基础上进一步引入了注意力机制。注意力机制允许模型在每个时间步根据输入数据的重要性动态地调整对应的权重，从而更好地关注和利用重要的信息。通过引入注意力机制，AM-ConvGRU 可以使得模型更加灵活地对待输入序列中的不同部分，提高了模型的表达能力和预测性能。

2）CNN + RNN

采用 CNN 和 RNN 结合的架构：卷积层首先被用来从卫星图像中自主提取关键特征。随后，采用带有卷积LSTM的RNN块来处理所选时期内的特征演变。模型的最后部分由一个卷积层和两个密集层组成，进行特征到强度的回归，预测+24小时时的Vmax/Vmmax。

目前，GoogLeNet已经提出了第四代，与残差网络融合，进一步提升了训练的速度。

3）GAN

GAN 是生成对抗网络（Generative Adversarial Network）的缩写，是一种深度学习模型，由 Ian Goodfellow 在 2014 年提出。GAN 由两个主要组件组成：生成器（Generator）和判别器（Discriminator）。GAN 的核心思想是通过训练生成器和判别器两个网络来实现。在训练过程中，生成器试图生成尽可能逼真的合成数据，而判别器试图正确地区分真实数据和合成数据。这种竞争驱动的训练过程最终会导致生成器生成的数据越来越逼真。

二、毕业设计（论文）方案介绍（主要内容）

**2.1课题拟研究的内容**

本课题旨在研究太平洋地区的台风路径和强度，拟使用深度学习方法来对台风路径和强度进行预测。

**2.2数据集下载和数据预处理的具体步骤**

1）下载源数据

下载 CMA 数据集、EAR5数据集和TCIR数据集。具体的数据集需要根据最终模型的需求进行调整。

2）对数据进行预处理，可能选择的数据预处理的方法

训练数据的使用拟采用卫星云图与二维台风数据相结合的方法。图像数据作为主要的训练数据，从图像中提取相应的特征。二维的台风数据作为辅助特征输入。

拟参照参考文献[1]，使用 CLIPER 方法对CMA数据集的二维的台风路径数据进行预处理，将经度、纬度、风速、中心气压等初始的特征组合扩展为多个特征（feature）。

对于EAR-5 数据集，本课题计划根据台风发生的区域和影响半径，在再分析数据集中截取一定范围内一定层次内一定变量类型的数据作为初始再分析数据集。

拟参照参考文献[16]，将初始的台风大气卫星图像裁剪为多个剪辑，例如将大小为250\*238 的训练数据裁剪成32\*32的剪辑，这样大量的训练剪辑可以防止过拟合。此外，使用裁剪的部分可以减少在一个训练步骤中消耗的内存，同时专注于小部分能够使模型有效地学习图像的细节。

具体使用的训练数据需根据模型的具体架构进行选择和修改。

**2.3模型训练的主要步骤**

模型训练主要包括如下步骤：

a.将数据集划分为训练、验证和测试 3 个部分。将当前的卫星云图和一些有关台风的数值（经度、纬度、风速、中心气压等）作为输入，希望得到未来一定时间内（6h 或者 12h）的台风路径和强度（一个数值关于时间的序列）；

b.搭建模型。设计合理的模型结构，并选择适当的超参数。考虑到台风路径和强度预测是一个复杂的时空序列预测问题，可能需要使用适应于序列数据的深度学习模型，如循环神经网络（RNN）或者生成对抗网络（RAN）。同时，模型的结构需要能够有效地捕捉台风路径和强度之间的时空关系；

c.训练模型。利用训练数据对模型进行训练，不断调整模型参数以最小化预测误差。在训练过程中，可以采用合适的损失函数和优化算法，如均方误差损失函数和随机梯度下降算法；

d.验证。训练完成后，利用验证集对模型进行验证，并分析预测结果与实际观测之间的误差。根据分析结果，制定模型改进方案，例如调整模型结构、优化超参数、增加更多的训练数据等，以提高模型的准确性和泛化能力；

e.对模型的性能进行测试和评估。最后，对经过验证和改进的模型进行性能测试和评估。可以使用测试集数据进行模型评估，考察模型在未见过的数据上的预测能力，并对模型的整体性能进行评价和总结

**2.4模型性能评估指标的选择**

由于模型的输出是预测的台风的路径和强度，本课题计划整合两个预测数据对台风的相似性进行计算。参考文献 [3]研究的主要工作是将原始台风数据进行序列化，基于DTW算法计算了两个台风之间的过渡相似性距离，并首次提出了一种新的度量标准，通过修正的双曲正切函数对相应的相似性距离进行归一化，从而量化了相似度程度。研究中对台风相似程度的量化涉及经过时间（h）、中心气旋的经度和纬度（无单位）、中心气压（hpa）、风速等级（km/h），但是本课题拟训练的模型的输出仅有台风路径（经度和纬度）、强度两个指标，故需将原文献用于量化的变量的个数进行一定程度的化简。最终得到预测的台风和真实数据之间的相似程度，从而评价模型的性能与表现。

三、毕业设计（论文）的主要参考文献

[1] Ying M, Zhang W, Yu H, et al. An overview of the China Meteorological Administration tropical cyclone database[J]. Journal of Atmospheric and Oceanic Technology, 2014, 31(2): 287-301.

[2] Xu G, Xian D, Fournier-Viger P, et al. AM-ConvGRU: a spatio-temporal model for typhoon path prediction[J]. Neural Computing and Applications, 2022, 34(8): 5905-5921.

[3] Di Y, Lu M, Chen M, et al. A quantitative method for the similarity assessment of typhoon tracks[J]. Natural Hazards, 2022: 1-16.

[4] Qin W, Tang J, Lu C, et al. A typhoon trajectory prediction model based on multimodal and multitask learning[J]. Applied Soft Computing, 2022, 122: 108804.

[5] Hou H, Yu S, Wang H, et al. A hybrid prediction model for damage warning of power transmission line under typhoon disaster[J]. IEEE Access, 2020, 8: 85038-85050.

[6] Gao S, Zhao P, Pan B, et al. A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network[J]. Acta Oceanologica Sinica, 2018, 37: 8-12.

[7] Yang Y, Wang J. An integrated decision method for prediction of tropical cyclone movement by using genetic algorithm[J]. Science in China Series D: Earth Sciences, 2005, 48: 429-440.

[8] Chang C C, Yang S C, Keppenne C. Applications of the mean recentering scheme to improve typhoon track prediction: A case study of Typhoon Nanmadol (2011)[J]. Journal of the Meteorological Society of Japan. Ser. II, 2014, 92(6): 559-584.

[9] Wei C C. Collapse warning system using LSTM neural networks for construction disaster prevention in extreme wind weather[J]. Journal of Civil Engineering and Management, 2021, 27(4): 230-245.

[10] Li S, Lu L, Hu W, et al. Prediction Algorithm of Wind Waterlogging Disaster in Distribution Network Based on Multi-Source Data Fusion[J]. Mathematical Problems in Engineering, 2022, 2022.

[11] Iwasaki T, Nakano H, Sugi M. The performance of a typhoon track prediction model with cumulus parameterization[J]. Journal of the Meteorological Society of Japan. Ser. II, 1987, 65(4): 555-570.

[12] Wang C, Li X, Zheng G. Tropical cyclone intensity forecasting using model knowledge guided deep learning model[J]. Environmental Research Letters, 2024, 19(2): 024006.

[13] Giffard-Roisin S, Yang M, Charpiat G, et al. Tropical cyclone track forecasting using fused deep learning from aligned reanalysis data[J]. Frontiers in big Data, 2020, 3: 1.

[14] 周冠博, 钱奇峰, 吕心艳, 等. 人工智能新技术在国家气象中心台风业务中的 应用探索[J]. Journal of Tropical Meteorology (1004-4965), 2022, 38(4).

[15] 周冠博, 钱奇峰, 吕心艳. 人工智能在台风监测和预报中的探索与展望[J]. 气象研究与应用, 2022, 43(2): 1-8.

[16] Rüttgers M, Jeon S, Lee S, et al. Prediction of typhoon track and intensity using a generative adversarial network with observational and meteorological data[J]. IEEE Access, 2022, 10: 48434-48446.

[17] Chen B F, Kuo Y T, Huang T S. A deep learning ensemble approach for predicting tropical cyclone rapid intensification[J]. Atmospheric Science Letters, 2023, 24(5): e1151.

四、审核意见

|  |
| --- |
| 指导教师审核意见：（针对选题的价值及可行性作出具体评价）  选题兼具学术研究价值和工程应用价值，拟采取的技术路线切实可行，同意开题。  指导教师签名 穆斌  2024 年 03 月 25 日 |
| 专业审核意见：  该项毕业设计选题合适，方案合理。教研室审核通过。  负责人签名 范鸿飞  2024 年 03 月 26 日 |