

分类号 TP399

学号 16063139

UDC 004.9

密级 公开

专业学位硕士学位论文

基于深度学习的台风预测关键技术研究

硕士生姓名 陈睿

专业学位类别 工程硕士

专业学位领域 计算机技术

指导教师 张卫民 研究员

协助指导教师 汪祥 助理研究员

国防科技大学研究生院

二〇一八年十二月

Research on The Key Technology of Typhoon Prediction Based on Deep Learning

Candidate: **RUI CHEN**

Supervisor: **Prof. WEIMIN ZHANG**

Associate Supervisor: **XIANG WANG**

A dissertation

Submitted in partial fulfillment of the requirements

for the degree of **Master of Engineering**

in **Computer Technology**

Graduate School of National University of Defense Technology

Changsha, Hunan, P. R. China

December, 2018

独创性声明

本人声明所呈交的学位论文是我本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表和撰写过的研究成果，也不包含为获得国防科技大学或其它教育机构的学位或证书而使用过的材料。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。

学位论文题目： 基于深度学习的台风预测关键技术研究

学位论文作者签名： 陈睿 日期： 2018 年 10 月 24 日

学位论文版权使用授权书

本人完全了解国防科技大学有关保留、使用学位论文的规定。本人授权国防科技大学可以保留并向国家有关部门或机构送交论文的复印件和电子文档，允许论文被查阅和借阅；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，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

(保密学位论文在解密后适用本授权书。)

学位论文题目： 基于深度学习的台风预测关键技术研究

学位论文作者签名： 陈睿 日期： 2018 年 10 月 24 日

作者指导教师签名： 何 日期： 2018 年 10 月 24 日

目 录

摘 要	i
ABSTRACT	ii
第一章 绪论	1
1.1 研究背景及意义	1
1.2 本文工作	9
1.3 论文结构	9
第二章 相关研究方法	12
2.1 相关台风预测方法	12
2.1.1 数值预报方法	12
2.1.2 统计预报方法	14
2.1.3 机器学习方法	19
2.2 相关深度学习网络	23
2.2.1 前馈神经网络	24
2.2.2 卷积神经网络 (CNN)	28
2.2.3 长短期记忆神经网络 (LSTM)	30
第三章 基于深度学习的台风时空深度混合预测模型	33
3.1 引言	33
3.2 问题定义	33
3.3 模型框架	35
3.3.1 海表变量 X^S 二维空间关系的特征学习	36
3.3.2 大气变量 X^P 三维空间关系的特征学习	38
3.3.3 台风变化过程中时序关系的特征学习	40
3.4 实现方法	41
3.4.1 实现平台	41
3.4.2 实现流程	42
第四章 基于时空深度混合预测模型的台风形成及强度预测	47
4.1 台风数据集和气象海洋再分析数据集的获取	47
4.2 基于台风时空深度混合预测模型的台风形成预测	49
4.2.1 实验数据集的构造与预处理	49
4.2.2 评价指标与损失函数	50

4.2.3	网络设置	50
4.2.4	对比方法	52
4.2.5	结果分析	53
4.2.6	参数分析	54
4.3	基于台风时空深度混合预测模型的台风强度预测	57
4.3.1	实验数据集的构造与预处理	57
4.3.2	评价指标与损失函数	58
4.3.3	网络设置	58
4.3.4	对比方法	59
4.3.5	结果分析	59
4.4	模型预测实验总结	60
第五章	总结与展望	62
5.1	研究工作总结	62
5.2	未来工作展望	63
	致谢	64
	参考文献	67
	作者在学期间取得的学术成果	73

表 目 录

表 1.1	热带气旋等级划分及对应名称	3
表 2.1	STPIS 的潜在静态预测因子	15
表 2.2	STPIS 的潜在综合预测因子	17
表 3.1	台风预测分类混淆矩阵	45
表 4.1	台风和热带风暴样本数量	47
表 4.2	台风形成预测实验结果对比	53
表 4.3	台风强度预测误差 (kt) 对比	59

图 目 录

图 1.1	台风示意图 (wikipedia)	1
图 1.2	1985-2005 年间热带气旋全球分布图 (wikipedia)	2
图 1.3	台风结构图 (wikipedia)	2
图 1.4	热带气旋生成预报概率图 (NOAA)	5
图 1.5	近年来 NHC 对于北大西洋台风的预报误差趋势图	6
图 1.6	论文的总体框架	10
图 2.1	神经元模型示意图	23
图 2.2	典型神经元激活函数	24
图 2.3	多层前馈神经网络示意图	25
图 2.4	RNN 典型结构示意图	30
图 2.5	LSTM 典型结构示意图 (Christopher Olah)	32
图 3.1	台风预测相关变量的时空结构示意图	34
图 3.2	台风时空深度混合模型框架图	35
图 3.3	用于海表变量空间特征提取的二维卷积层示意图	36
图 3.4	海表温度分布示意图	37
图 3.5	2DCNN 用于海表变量二维空间关系的特征学习示意图	38
图 3.6	用于大气变量空间特征提取的三维卷积层示意图	39
图 3.7	3DCNN 用于大气变量三维空间关系的特征学习示意图	39
图 3.8	台风最佳路径示意图	40
图 3.9	LSTM 用于台风变化过程中时序关系的特征学习示意图	41
图 4.1	模型网络设置及流程图示例	51
图 4.2	Acc 以及 Auc 随学习率 (lr) 的变化	54
图 4.3	Auc 值在不同的学习率下随 epoch 的变化	55
图 4.4	Auc 及 Acc 值随中心周边环境变量经纬度区域范围的变化	55
图 4.5	西太平洋 Acc 及 Auc 在不同 LSTM 时间步下随最佳区域范围的变化	56
图 4.6	东太平洋 Acc 及 Auc 在不同 LSTM 时间步下随最佳区域范围的变化	56
图 4.7	北大西洋 Acc 及 Auc 在不同 LSTM 时间步下随最佳区域范围的变化	57

摘要

台风是一种极端天气事件，每年夏季会对沿海地区城市经济造成重大损失，预测台风的形成和强度对台风灾害的预警至关重要。传统的基于热动力方程的数值预报方法和基于经验关系的统计预报方法现如今仍难以准确地预测台风强度。此外，一些研究虽然也尝试用机器学习方法进行预测来提高准确率，但通常没有考虑台风相关变量之间的时空关系。本文提出了一种基于深度学习的台风时空深度混合预测模型，可以弥补现有方法的不足之处。本文模型引入了三维卷积神经网络 (3DCNN) 和二维卷积神经网络 (2DCNN) 来学习台风相关大气海洋变量之间的空间关系，利用 LSTM 来学习台风变化过程中的时序关系。针对西太平洋 (WP)、东太平洋 (EP)、北大西洋 (NA) 三个海域的大量台风形成与强度预测实验表明，本文模型优于现有官方组织使用的诸多数值预报方法、统计预报方法和机器学习方法。台风 24h 形成预测实验中，本文模型能达到的最高准确率为 85.2%，Auc 值 (ROC 曲线下的面积) 为 92.2%，24h 强度预测实验中，最小误差为 7.4kt (1kt=0.512m/s)。并且在参数分析时发现，模型训练时的超参数如学习速率最好设置为 10^{-4} 的量级，并在 25 ~ 35 个 epoch 后停止模型的训练。构造输入数据集时，台风中心周围经纬度区域范围最好确定在 $9^{\circ} \times 9^{\circ} \sim 13^{\circ} \times 13^{\circ}$ 内，且 LSTM 时间步确定为 2 时能获得最好的预测效果。

关键词: 台风；强度；预测；时空；深度学习；时空深度混合预测模型

ABSTRACT

Typhoon is an extreme weather event that can cause heavy damages to urban economy in coastal areas. It is very important to predict the formation and intensity of typhoon to give early warnings for the typhoon disasters. Traditional numerical forecast models based on thermodynamic equations and statistical models based on empirical relationships are still hard to predict intensity of typhoon accurately. Some researches have tried using to use machine leaning methods to predict the formation and intensity of typhoon, but they did not consider the spatial and temporal relationships among typhoon related variables. Here we propose a deep hybrid spatio-temporal model to make up for the shortcomings of existing methods. Our model introduces 3D convolutional neural networks (3DCNN) and 2D convolutional neural networks (2DCNN) to learn the spatio-temporal correlation of atmospheric and ocean variables. We utilize LSTM to learn the temporal sequence relations in the path of typhoon. Extensive experiments based on the Western Pacific (WP), East Pacific (EP), and North Atlantic (NA) show that our model is better than existing methods, including numerical forecast models used by many official organizations, statistical forecast methods and machine leaning based methods. In the experiments of 24h typhoon formation, the accuracy of our model can reach 85.2%, the Auc is 92.2 %, and the intensity prediction error is 7.4kt. At the same time, it is found that the super-parameters such as the learning rate during model training are preferably set to the order of 10^{-4} , and the training can be stopped after 25 ~35 epochs. The best latitude and longitude range around the typhoon center of input data set is found in range of $9^{\circ} \times 9^{\circ} \sim 13^{\circ} \times 13^{\circ}$, and the best predictive effect is obtained when the time step (LSTM time step) is 2.

Key Words: typhoon; intensity; forecasting; spatio-temporal; deep-learning; deep hybrid spatio-temporal model

第一章 绪论

1.1 研究背景及意义

热带气旋是由温暖海洋热能驱动的天气尺度风暴，含一个无云区的中心，中心上空有一个暖核，围绕中心的气流呈逆时针方向旋转，如图 1.1。在这个涡旋中，最低气压出现在中心，最大风速出现在中心附近，通常会达到或超过 34kt (10.8m/s)。根据世界气象组织定义，在西北太平洋区域，当热带气旋中心附近最大风速达到 64kt (32.7m/s) 时则称为台风。这样的涡旋在大西洋和东太平洋称之为飓风，在北印度洋称为强气旋性风暴^[1]。



图 1.1 台风示意图 (wikipedia)

台风乃至热带气旋最常形成于 5°N 以北的西北太平洋热带洋面上，但全球除东南太平洋和南大西洋外各大洋区的其他热带海域也均有热带气旋发生和活动，其中以北大西洋、东太平洋、北印度洋最为频繁。图 1.2 为 1985 ~ 2005 间热带气旋的全球分布图。

台风作为一种大气涡旋。其大小各异，大台风的直径上千千米，小台风二三百米，但结构特征却基本一致。台风的宏观结构大致分为六部分，分别是台风眼、眼外云墙、外包区、外围区、流入层、流出层。其中最为显著的特征是台风眼，它位于涡旋之中，眼内为下沉气流，静风，无云，亦为晴空区。有的台风有两个眼，称为同心双眼，也有两眼并列的台风，但极为罕见。台风眼外的云墙，是由于强烈上升运动而造成的深厚积云对流，台风最强的强风和暴雨也出现于此。台风眼和眼壁合称为内核区，内核区直径小的可为二三十千米，而大的可扩大至为上百

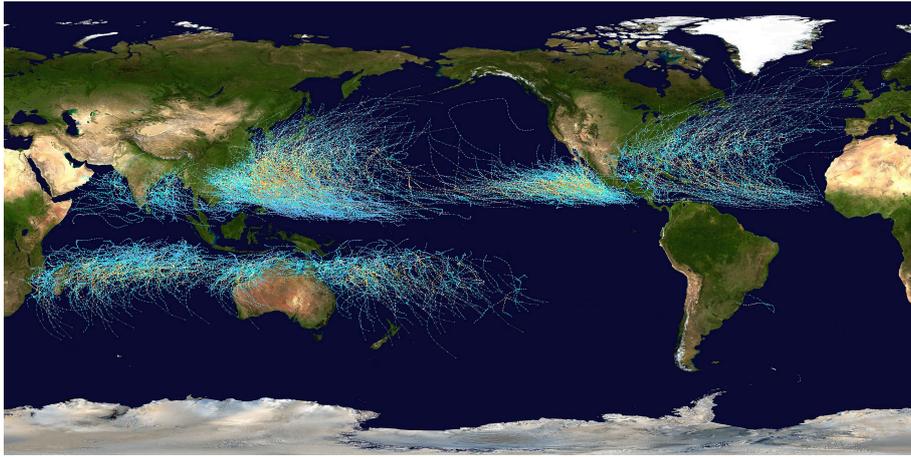


图 1.2 1985-2005 年间热带气旋全球分布图 (wikipedia)

千米。外包区紧挨着内核区，风速比内核区稍弱，螺旋雨带的主体在此区中。外围区紧挨着外包区，水平尺度是外包区的几倍，风速在该区虽明显减弱，但破坏力依然很强。此外，台风在低气压阶段，四周的空气是向中心流入的，从而形成了流入层。随着热带气旋的增强，切向风增大，流入现象也会减弱。流出层位于台风的顶部，对流层顶之上。该层气流在北半球呈反气旋式向外流出，使台风中心气压急剧下降，从而使得台风开始加强^[2]。图 1.3 可以清晰地展现台风的结构。

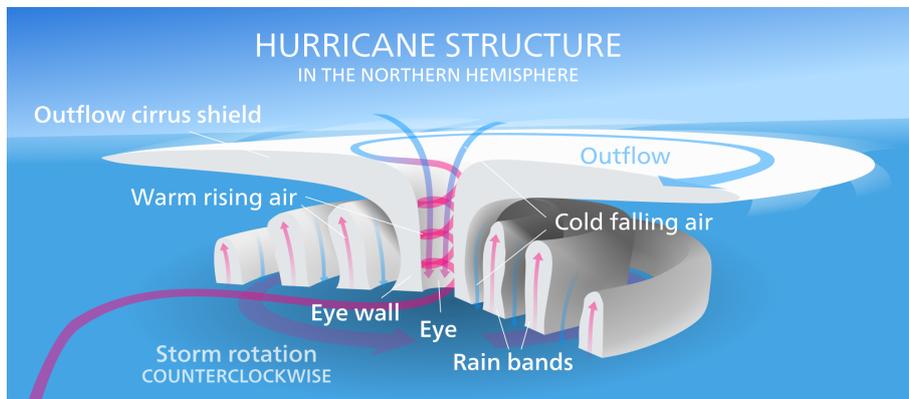


图 1.3 台风结构图 (wikipedia)

通常，在进行台风预警时需要对台风的破坏性进行评估，最常用的一个衡量指标为强度。通常，台风强度可由三个要素来表示，一是台风内核的强度，即台风中心海平面的最低气压或台风中心附近最大风速。二是台风外围区平均风速的大小。三是尺度和大小，即台风最外一圈闭合等压线的平均直径。其中，中心附近最大风速一般作为强度的主要参数^[3]。因此，本文中所定义的强度即为台风中心附近最大风速。

根据我国气象局 (China Meteorological Administration, CMC) 对于西北太平洋热带气旋强度的划分，热带气旋不同等级分别对应不同的名称，如表 1.1 所示。因

此，严格意义上，无论是台风还是飓风，均是热带气旋的一种等级划分，本质上仍为热带气旋。本文把所有全球大洋强度大于 64kt 的热带气旋统称为“台风”，强度大于 34kt 的热带气旋称之为“热带风暴”。

表 1.1 热带气旋等级划分及对应名称

热带气旋等级名称	强度
热带低压	17 kt ~ 33 kt
热带风暴	34 kt ~ 63 kt
台风	64 kt ~ 84 kt
强台风	84 kt ~ 104kt
超强台风	≥ 105kt

台风具有极强的破坏力，通常在夏秋季节严重威胁我国沿海，尤其是华东、华南地区甚至部分内陆省份。我国遭受台风灾害有次数多、季节性强、受灾程度重、影响范围广等特点。根据《中国气象灾害年鉴》^[4]统计，平均每年有 7 ~ 8 个台风登陆我国。台风季节内约半个月便有台风登陆，严重的台风灾害主要集中在 7 月至 9 月，受灾地区主要为台湾、海南、广东、福建等众多省沿海地区。此外，少数登陆浙江、福建后向北移入黄渤海的台风同样也会对上海、江苏、河北等省份造成破坏性损害。例如，于 2018 年 9 月 7 日 20 时在西北太平洋洋面上生成的 2018 年 22 号超强台风“山竹”。截至 2018 年 9 月 18 日 17 时，已造成的人员损失包括广东、广西、海南、贵州、湖南等 5 个省（区）近 300 万人受到影响，5 人死亡，1 人失踪，以及 160.1 万人紧急转移。经济损失包括 5 省（区）1200 余间房屋倒塌，超过 174.4 千公顷农作物受灾，且至少 52 亿元的直接经济损失。

台风发生时，人们一般能直接感受到的是狂风暴雨，而其造成的恶劣灾害事件可分为直接灾害和间接灾害两种。直接灾害主要是直接由台风的狂风暴雨而带来的风灾和洪涝灾害，而间接灾害主要是由暴雨间接引发的山体滑坡和泥石流等地质灾害，以及由大风引起的沿海风暴潮灾害^[1]。

直接造成的灾害如台风大风导致的风灾，主要表现为（1）海上航行受阻、海上石油勘探和平台生产设施、船舶、码头装卸运输设施等受损；（2）沿海渔业和海洋养殖业设置受损而停业；（3）农作物、果树等经济作物受强风损毁而减产甚至绝收；（4）城市建筑工程设施被摧毁、城市交通运输受阻，事故频发。台风暴雨灾害主要是由台风强降雨引发的灾害。其主要表现为（1）城市的严重积水，（2）交通运输受阻，城市经济活动和居民生活严重受影响。台风洪涝灾害主要是在沿海或者内陆丘陵地区累积雨量很大却无处排放的情况下的引发的大范围灾害。主

要表现为 (1) 大面积农田被淹没, 导致农作物减产甚至绝收; (2) 堤岸、路基、通信和输电网设施被冲毁, 导致大面积停电 (3) 房屋、禽畜棚被损毁或者浸坏, 导致人畜伤亡, 从而引发流行疾病和瘟疫。

间接造成的灾害如台风暴雨引发的地质灾害通常发生在地质状况较为脆弱的地区。其中泥石流是山谷中含有大量泥沙和石块的洪流, 主要表现为突然爆发, 浑浊的泥石流体会沿陡峻的山沟顺势流下, 在宽阔的堆积区堆积, 从而对山区人民财产的生命安全造成了严重的危害。山体滑坡是构成斜坡的岩土体在重力作用下失稳, 沿着坡体内部的一个 (或几个) 松软脆弱面 (带) 发生剪切而产生整体性下滑的现象。是山区水库、铁路、公路及依山而建的居民等建筑设施常遇到的一种地质灾害。主要表现为河道堵塞、公路阻断、公车毁坏等。而风暴潮是指极端天气作用下所引发的海面异常增高的现象, 若同时处于天文潮的高潮阶段, 将导致潮位快速升高。主要表现为 (1) 水位暴涨、堤岸绝口、人畜伤亡; (2) 海水倒灌、地下水遭受入侵、水体和土壤盐度升高、植物生存受威胁。

上述所有的灾害中, 台风风灾往往将与其他台风灾害同时出现, 如狂风伴随着巨浪对船舶的损害、狂风激起风暴潮冲毁海塘、堤坝或者导致海水漫滩、狂风夹带暴雨引发的洪水灾害、暴雨情况下强风拔树可触发山体滑坡和泥石流灾害等次生地质灾害。因而, 台风的风力强度和持续时间是致灾的重要气象因素之一。除此之外, 灾害的发生以及灾情严重程度还取决于台风预警、预报的准确率, 受灾体状况, 包括与之密切相关的社会环境和经济发展水平, 以及防台抗灾措施的及时性和有效性。对于气象领域的研究人员和业务人员来说, 为了减轻台风灾害带来的损失, 主要关注部分还是在于提供更为准确的台风预警、预报结果, 为接下来的防灾减灾措施做指引。

确定台风中心位置和强度是制作台风预报和发布台风预警的第一步, 因为台风预报的质量依赖于台风初始定位和定强的精度。台风监测是指利用各种手段对台风移动过程中的物理现象以及气象要素进行观测。台风监测所获取的气象记录、资料是进行台风预警、科学研究和防灾减灾、决策服务的基础。目前对于台风监测主要包括: 地面探测、高空探测、雷达观测、以及遥感探测等。在台风监测和预警的实际业务中, 确定台风中心位置和强度的主要依据是气象卫星、地面基准雷达和地面气象观测, 少数国家则应用飞机探测作为一种重要的辅助监测手段^[5]。气象卫星图像是确定台风中心位置最常用的资料, 然后使用 Dvorak 方法^[6], 并结合透明螺旋线板和主观解释, 通过确定适当的云系中心的位置来定位的。第一步为根据云型特征确定台风中心, 第二步根据云图上的地标特征进行网格误差的校正, 第三步进行斜视误差的校正, 第四步再利用台风的前期位置活动强度变化以及路径与台风位置的相关来检验所确定的位置是否合理, 最后定出台风中心位置。

至于台风强度的确定，同样无法直接测量。目前世界各国（包括中国、日本、美国等）主要仍主要采用 Dvorak 研发的 Dvorak 分析技术，根据静止气象卫星在红外和可见光波段观测的台风云型特征及其变化估计台风的强度，该技术已成为在缺少飞机探测地区监测台风的世界标准，也是进行强度预报最通用的方法。由于在卫星云图上，需要通过台风云系的多种结构特征才能真实反映强度。这些特征包括：台风中心大小、强对流区范围以及外围螺旋雨带等。可以通过对卫星云图中这些特征的分析判断，给出用于表征台风强度的现实指数，然后根据观测现实强度指数与中心最大风速的经验关系，进而得到台风中心最大风速，即台风的强度。

对于台风预警和防台减灾而言，首先需要确定的是海域内是否有热带气旋生成，且是否会形成为台风。因而热带气旋的生成或台风的形成预报是首要需要考虑的问题。其次，需要知道的是台风未来途径的区域，而这主要取决于预报台风的移动路径（一般用 3 或 6 小时间隔的台风中心位置来表达），因此台风路径的预报也是需要重点考虑的问题。进而为了确定台风对某海域或某地的影响程度，需要关注台风的强度，而风雨预报也是在台风强度的基础上，综合考虑未来天气形势的变化而进行的，因此，台风的强度的预报是需要考虑的关键问题。再之后，则需重点关注的是台风的风雨预报、间接引发的地质灾害预报以及风暴潮预报等。

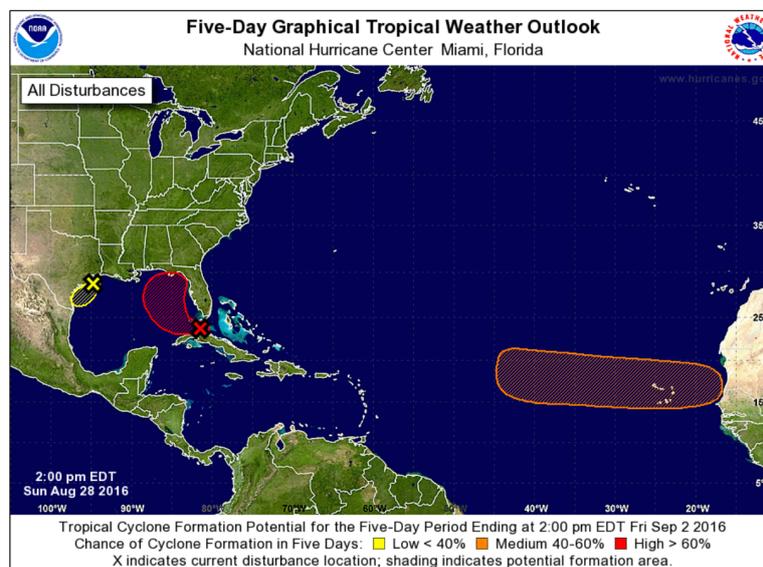


图 1.4 热带气旋生成预报概率图 (NOAA)

就台风形成预报而言，国内外一般的预报方法是在热带气旋生成后，通过预报其强度来预报其是否会有台风形成。但已有的研究关注点普遍在于热带气旋的生成问题上，比如生成机理的研究上，通常认为热带气旋的生成是需要预先存在的扰动，暖而深的海洋混合层，大于平均的低层涡度和中层湿度，不稳定的大气

层，弱的水平风垂直切变以及科氏力作用的存在。业务预报上，国外相对来说有较为成熟的生成预报模式^[7]，如 NOAA 采用的全球模型 GFS 和 UKMET 及其集合模式是预测热带气旋发生的主要工具^[8]，可以预报特定海域热带气旋的生成概率，图 1.4 所示，而国内暂时没有。除开数值模式，预测热带气旋生成的方法从以前的统计方法，到现在与机器学习方法融合的新型方法，使得热带气旋生成的预测方法更加丰富，效果相比于以前有了较大的提升。研究学者们会在洋面上已有热带扰动的基础上，通过考虑周围的大气环境条件建立模型来预测是否会有热带气旋生成。其中一部分研究的目的是为了研究更为有效预测方法，另一部分研究的目的是为了开发为业务预报服务的热带气旋生成预报系统。此外，还有部分研究会利用大尺度环境场对热带气旋的活动进行季节性预报，主要通过定义热带气旋的生成指数 (GPI)^[9]，来描述热带气旋生成可能性大小的空间分布。

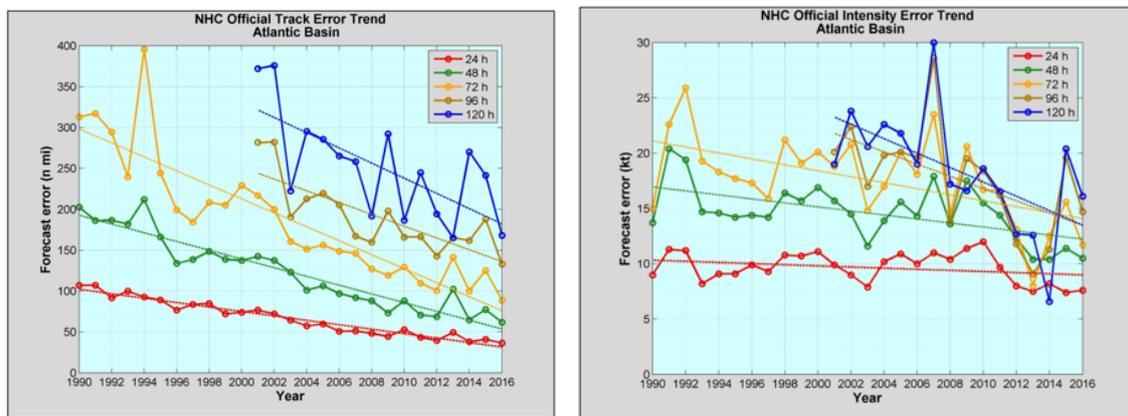


图 1.5 近年来 NHC 对于北大西洋台风的预报误差趋势图

台风路径最初大多是建立在预报经验的基础上，传统的路径业务预报方法包括天气气候学预报方法、环境引导气流预报方法。近三十年来，出现了一些客观预报方法，包括统计预报方法、统计动力预报法、模式输出统计释用法、神经网络法等。大多采用的是历史台风资料作为统计样本，在对影响台风移动的大尺度环境场气压系统（如副热带高压、高空长波槽、高空切断冷涡等）、环境场引导气流、海洋要素（海温、洋流等）、下垫面状态（如地形、地温等）作相关分析基础上，选择具有天气学、大气热力动力学等物理意义作为预报因子，一般以特定时效（如 12、24、36、48 小时）的经向和纬向移动距离作为预报量，采用各种数理统计方法对历史样本研究建立台风路径预报模型来进行预报。而数值预报模式仍然是现阶段常用的一种预报方法，台风路径数值预报模式系统是在数值天气预报模式上研发的专业应用的数值预报系统，在模式中反映了台风独有的参数化方案（如积云参数化，边界层参数化，海气耦合）和针对台风建立相应的物理过程。由

于台风数值预报模式的初始场含有观测误差和分析误差，且大气运动本身存在不确定性，所以数值预报同时受初始场误差和观测误差的影响，这将使最终的预报结果偏离实际预测结果^[10]。台风强度预报和台风路径预报有很多相似之处，但是由于台风的发展过程中存在迅速增强等现象，会使得台风强度预测更具有挑战性。同样业务预报中的预报方法和路径预报类似，但是比较常用的是统计法、统计动力法、以及数值天气预报方法。近年来统计预报方法和统计动力预报方法表现了最优的预报效果。其中统计预报方法除考虑气候相关变量外，还引入了当前和前期大气环境变量、海面变量以及卫星图像特征建立预报模型。而统计动力预报方法是以数值天气预报的预报产品为依托，通过考虑未来大气环境和海洋状况的变化来建立预报模型。总的来说，可以发现预报的总体趋势是，路径误差在逐年降低，而强度误差基本无太大的变化^[11]，如图1.5以美国飓风中心（NHC）为例。

机器学习是人工智能研究发展到一定阶段的必然产物，是实现人工智能的途径之一。由于人工智能系统需要自己具备获取知识的能力，即从原始数据中捕捉规律、学习模型的能力，为了明确定义这种能力，从而称之为“机器学习”。二十世纪八十年代开始把机器学习^[12]研究划分为“通过观察和发现学习”，“在问题求解和规划中学习”，“从样本中学习”等类别。其中被研究、应用范围最广的是“从样本中学习”，包含了监督学习、无监督学习等。“从样本中学习”的一大主流是符号主义学习，以决策树为例。典型决策树模型基于信息论理论，以最小化信息熵为目标，直接模拟了人类判定事物的树形流程。另一主流技术是基于神经网络的连接主义学习，不同于符号主义学习能够产生的明确概念，连接主义学习通常产生的是“黑箱模型”。二十世纪九十年代中期，以统计学习理论为直接支撑的“统计学习”技术逐渐占据主要地位，代表技术是支持向量机以及更为一般的核方法。到了二十一世纪初，连接主义学习又掀起了名为“深度学习”的热潮，“深度学习”可以简单理解为“很多层”的神经网络。近年来，深度学习在语音、图像等复杂对象的应用中，均取得了优越效果。虽然以往传统的机器学习技术也能在应用中要取得良好的性能，但是相对来说对于使用者有较高的要求。而深度学习虽然模型复杂度高，参数多，但是在使用方面降低了门槛，通常只要设计好模型，调节好参数，就能取得理想的效果，因此，深度学习为机器学习技术走向工程实践带来了便利。此外，随着“大数据时代”的到来，数据储量和计算能力均有了快速发展，深度学习模型复杂度高、参数量过大的缺点也逐渐得以淡化，从而深度学习应用的领域越来越广。

过去几十年中，人类在收集、存储、处理数据方面取得了巨大的进步。虽然数据量急剧爆炸，但能有效分析大规模数据的算法却极度缺乏，深度学习乃至机器学习的出现则恰恰顺应用了现阶段的急切需求。机器学习不仅在计算机视觉、

自然语言处理等“计算机应用技术”领域大放异彩，还为许多交叉学科提供了重要的技术支撑。例如“生物信息学”尝试用信息技术来研究生命现象和规律，其研究过程主要包含数据的获取、数据的管理、数据的分析、以及仿真实验等环节，其中最为关键的数据分析环节，机器学习在其中起到了不可忽视的作用。此外，随着科学研究的基本手段从传统的“理论+实验”走向现在的“理论+实验+计算”，乃至“数据科学”的出现，机器学习的重要性日趋显著，因为“计算”的目的往往是数据分析，数据科学的核心也是通过分析数据来获得价值。如今，机器学习已经与普通人的生活密切相关。例如在商业营销中，可以通过机器学习对客户基本信息、销售数据进行分析，从而制定针对用户群设计特殊的营销策略，进而帮助商家优化库存、降低成本、提高利润。此外，在能源勘探、环境监测、天气预报等方面，可以通过机器学习对卫星和监测数据以及历史数据进行分析，从而提高预报和检测的准确性。

为了提高台风预测的准确性，同样出现了越来越多的学者尝试将机器学习应用于台风预测相关领域。一些学者尝试利用深度学习里的前馈神经网络量化台风引起的海面温度冷却，通过增加在预测模型中海表温度冷却的作用^[13]来提高台风预测的准确率。一些学者利用深度学习里的卷积神经网络（Convolutional Neural Network, CNN）提取卫星遥感图片里的特征，从而预测台风发生时的多种天气指标，如降水、气压、温度和风速等^[14]。一些学者为了准确预测表示台风发生时的风险概况，利用机器学习的主成分分析（Principal Component Analysis, PCA）预测相关参数的分布，发掘机器学习算法作为当前数值模式中参数方法的替代方案的潜力^[15]，来确定风场可变性的全部范围。一些学者利用机器学习中的支持向量回归（Support Vector Regression, SVR）模型作为预测方法对台风的季节性进行预测^[16]，从而为准确获得台风的发生频率、空间分布以及季节性的强度变化。还有一些学者利用深度学习从大量的卫星云图中研究其台风“胚胎”的可预测性^[17]，或者对现有的预报方法进行误差订正^[18]。然而，目前被研究的最多的还是机器学习或者深度学习在台风路径和强度研究中的应用^[19-23]。对于路径研究，主要研究的是台风中心的定位和中心位置以及轨迹的的预测^[19]。对于强度研究，主要研究的是台风强度的估计，台风的强度分类，以及强度的预测^[14]。

由于现阶段台风的形成预测问题基本被忽视，而路径预测在有明显的提升的同时强度预测却没有太大的改善，从而重点关注于台风的形成预测和强度预测问题。本文的目的是在以往台风预测相关研究的基础上，利用深度学习自动提取特征并且处理时空相关性的优势，提出一个能够有效提高台风预测的效果且减少实施的代价的台风预测模型。并以台风形成和台风强度实际预测为例，证明能够提供更为准确的台风预报和预警，为防灾减灾做贡献。

1.2 本文工作

本文工作及主要贡献包括以下三个部分内容：

(1) 在现有数值预报、统计学习以及机器学习台风预测方法的基础上，为弥补其计算代价过大、理论机制不明确、未考虑预测变量的复杂关系等不足之处，结合台风数据和环境大气海洋数据的时空特性，将台风预测问题具化定义为了一个时空变量序列预测的问题。

(2) 针对被重新定义的台风预测问题，结合深度学习的优势，提出了一个基于深度学习的解决方法，即本文的台风时空深度混合预测模型。其主要思想是利用历史大气海洋的空间变量数据以及台风最佳路径资料，通过 3DCNN (3 Dimesional Convolutional Neural Network) 进行大气变量的三维空间特征学习，2DCNN (2 Dimesional Convolutional Neural Network) 进行海表变量的二维空间特征学习，LSTM (Long Short Term Network) 进行台风从形成之前到形成之后的时序特征学习，从而达到预测台风的形成以及形成后强度、路径等参量的目的。该模型可以根据不同的预测目标，选择不同的输入、输出（标签）数据集进行模型的训练学习，进而预测不同的对象。其优点是可扩展性强，计算代价小，且易于实现。

(3) 具体应用是从台风预测中的重点以及难点问题 - 台风形成及台风强度预测出发，根据本文提出的台风时空深度混合模型，进行了台风形成与台风强度的预测实验分析以及模型训练过程的重要参数分析。实验过程中工作主要集中于实验数据集的构造、模型中网络的设计与实现，参数的调整以及和与其它方法的对比分析。本文给出了利用 IBTrACS 热带气旋最佳路径数据和欧洲中期天气预报中心 ECMWF (European Centre for Medium-Range Weather Forecasts) python API 获取台风样本空间大气海洋环境变量数据，以及利用 MATLAB 进行台风样本输入数据集构造的具体步骤。采用以 TensorFlow 作为后端的 Keras 深度学习框架对模型进行实现与测试，再以手动调参和调用 Sklearn 的 GridSearchCv 模块搜索最优参数的方法，确定了模型的最优超参数和输入数据规模的选择。并且在与现有方法进行了对比分析后发现，本文所提出的基于深度学习的台风预测关键技术 - 台风时空深度预测模型具有明显的优势。

1.3 论文结构

本文旨在现有的统计预报和机器学习等台风相关预测方法的基础上，结合深度学习中卷积神经网络获取空间特征和长短期记忆神经网络获取时序特征的能力，充分学习台风活动过程的时空关系，建立台风时空深度混合预测模型来对台风的各种关键问题进行预测。

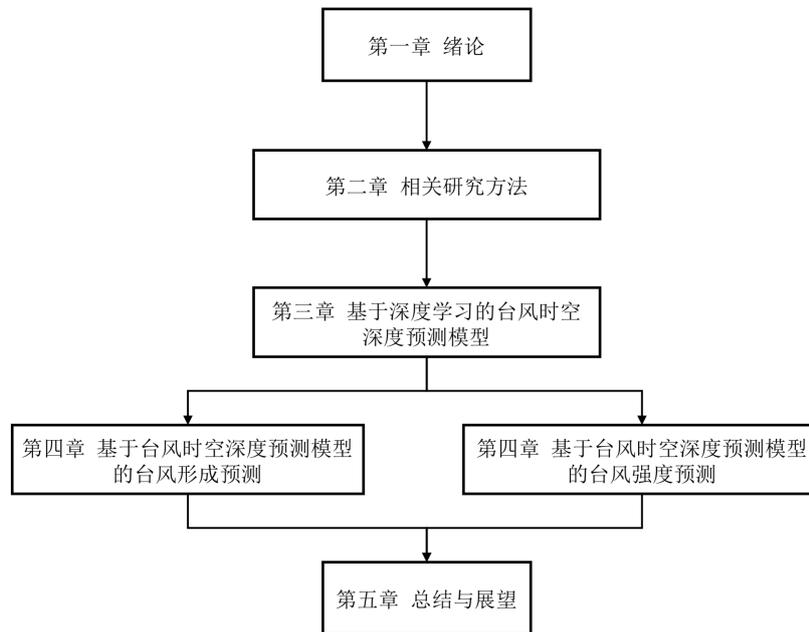


图 1.6 论文的总框架

本文共分为五章，其组织如下：

第一章为绪论部分，主要介绍课题的研究背景及意义。首先从台风的定义，台风的的活动范围，台风的的结构，台风的破坏性指标等方面给出了对台风的基本认识。然后简单介绍了台风对我国造成的威胁以及已酿成的损失，以 2018 年 22 号台风为例，重点介绍了台风的破坏性主要的表现方面。根据其危害性提出了台风预警以及预报的重要性，进一步给出了实现预警和预报的预测方法的现状，分析了其存在的不足之处且可以改进的部分。随后引出了机器学习、深度学习的概念以及在多个领域的成功实践，阐明了它在数据分析和预测问题上的优势。最后整理了近年来利用机器学习进行台风相关问题的部分研究，证明了本课题方向的正确性，从而提出了本文的研究目的，并解释了其重要的研究意义。

第二章为相关研究方法部分，主要分为对相关台风预测方法的介绍和相关深度学习理论和实现方法的介绍。首先对现有的台风预测方法进行了深入的调研，大致将其分为了三类，即数值预报方法、统计预报方法以及机器学习方法展开介绍。分别从理论、研究现状、优缺点等方面出发针对上述三种预测方法进行了具体的分析总结。随后给出了深度学习的基本概念、并重点介绍了与本文密切相关

的前馈神经网络以及的卷积神经网络 (CNN) 和长短期记忆神经网络 (LSTM) 的原理及参数的更新方式等基础理论, 为模型的提出做理论铺垫。

第三章为核心部分, 即基于深度学习的台风时空深度模型的提出、设计与实现部分, 主要介绍了研究问题的定义以及解决该问题的具体手段。首先根据数据的特性以及从已有方法中吸取到的相关经验, 对研究问题给出了具化的定义, 且是深度学习能够解决的问题。然后针对问题给出了模型的总体框架, 为对其进行概括性的总结描述。随后给出了设计该模型思路, 包括设计的合理性等方面。最后介绍了多种可以实现的深度学习框架, 并且给出了本文实现该模型以及调参的方法。

第四章为应用部分, 主要针对台风预测两大重点以及难点问题 - 台风形成及强度预测进行了预测实验。该部分首先介绍了实验所需的公用数据集的获取方法, 然后分别从形成预测和强度预测两方面进行了实验的设计与实施。以台风形成预测为例, 实验数据集的构造及预处理是实施实验的第一步, 选定模型的评价指标与损失函数, 以及设置模型内部的网络是第二步。分析相关的预测方法并以相同的数据进行实验测试是第三步。对比现有的方法随后进行预测结果的分析以及重要参数的分析是第四步。

第五章为总结展望部分, 主要针对本文对于台风预测所定义的研究问题, 提供的解决途径, 以及实验论证部分进行了简要的总结, 并对后续工作进行了展望。

第二章 相关研究方法

2.1 相关台风预测方法

2.1.1 数值预报方法

台风数值预报一般是通过对一定海域内的气象海洋环境条件，如副热带高压、水汽输送、海温等大尺度环境场进行预报后，再对台风路径和台风强度等台风要素进行推断预报。其中路径预报一般只需要数值模式能够对大尺度环境场进行合理的描述，即可对路径做出较为准确的预报。而强度预报不仅需要数值模式对大尺度环境场进行描述，还需了解内核热动力过程与环境场的关系，对台风涡旋结构、多尺度物理过程等多方面进行合理的定义。因此，根据数值预报模式准确预报台风强度的难度比路径要大。若要对台风的数值预报过程进行描述，需要从模式初始化、物理过程参数化这两方面展开^[24]。

模式初始化是进行台风数值预报的第一步。模式初始化是根据观测和背景场资料，在热动力约束下生成模式初始场的过程。早期由于缺乏海上观测资料，台风模式初始化方法通常是构造少量的 **Bogus** 涡旋，然后通过插值或变分同化方法融合到模式初始分析场中^[25]。上述方法虽然对于预报能力在一定程度上有提升，但由于考虑了许多假设以及经验关系，所以仍然会导致台风预报误差过大。近年来，国内外均逐渐尝试将多种卫星遥感、雷达、飞机实测资料与同化方法结合，应用于台风模式的初始化及预报。**Ma and Tan(2010)**^[26] 提出了一种典型的模式初始化方法——**VIRV** 法，它不仅避免了传统三维变分同化 (**3DVAR**) 方法不适用台风的问题，融合了四维变分同化 (**4DVAR**) 的优点，而且考虑了边界层对台风对流及其发展的重要性，在模式初始场中引入了与实际观测相符的台风结构。**VIRV** 的过程如下：

- (1) 在 **QuickSCAT** 资料的中纬度区域的每个格点，利用边界层模式估计边界层顶的气压梯度，然后进行梯度订正；
- (2) 在热带地区，根据混合层边界层模式估计边界层顶的气压梯度；
- (3) 通过最小二乘法由气压梯度拟合最佳相对气压场；
- (4) 根据 (3) 得到的相对气压和浮标站观测和数值模式背景场分析得出的参考气压计算海平面气压；
- (5) 将海平面气压同化进中尺度数值模式；

过程中选择的反演中纬度海平面气压的方法是由 **Patoux**^[27] 等提出。假设边界层顶存在地转平衡，那么边界层顶的地转风 (G) 可根据 (2.1) 中由海面风 (U_h) 计算得出：

$$\frac{U_h}{G} = \frac{V_T}{G} + \frac{1}{G} \frac{\partial V_2}{\partial \xi} + \cos \alpha + \sin \alpha \quad (2.1)$$

式中 V_T 为热成风, V_2 为次级涡旋扰动, α 为边界层内风相对于垂直方向的旋转角度。由于假设存在地转平衡, 所以需要再作梯度风订正。

而热带地区风压场不再满足地转平衡, 于是 Stevens et al(2002) [28] 通过使用混合层模式来对气压场进行估计, 可以得到如下方程:

$$f \mathbf{k} \times \mathbf{U} + \frac{1}{\rho_0} \nabla P = \frac{\boldsymbol{\tau}(h) - \boldsymbol{\tau}(0)}{h} \quad (2.2)$$

其中, \mathbf{U} 为边界层平均风速, P 分别为海平面气压, $\boldsymbol{\tau}(0)$ 为地表湍流应力, $\boldsymbol{\tau}(h)$ 为边界层顶的湍流应力。

气压场 p 可由下式得出

$$|\mathbf{H}^T \mathbf{H}x - \mathbf{H}^T y| \equiv 0 \quad (2.3)$$

$$\text{其中算子 } \mathbf{H} = \begin{vmatrix} \frac{1}{\alpha \cos \varphi} & \frac{\partial}{\partial \lambda} \\ \frac{1}{a} & \frac{\partial}{\partial \varphi} \end{vmatrix},$$

且 a 为地球半径, λ 为经度, φ 为纬度。

由 (2.1 ~ 2.3) 式结合海上的浮标站观测和数值模式背景场分析得出的参考气压值, 可以得到中纬度地区和热带地区的气压场分布。

在反演出海面气压场后还需将其同化进入台风模式的初始场中。典型同化方法包括包括最优差值、三维变分同化、四维变分同化、集合卡尔曼滤波、粒子滤波等。Ma and Tan(2010)[26] 选择了较为成熟的三维变分同化, 通过迭代最小化代价函数 $J(x)$ 来得到最接近观测的最优分析场 x , 其代价函数由观测误差和背景场误差构成, 具体定义为

$$J(x) = J^b + J^o = \frac{1}{2}(x - x^b)^T B^{-1}(x - x^b) + \frac{1}{2}(y - y^o)(E + F)^{-1}(y - y^o) \quad (2.4)$$

上式对每个格点的误差分析, 使用了背景误差, 观测误差和代表性误差的加权, 它们的误差协方差矩阵分别为 \mathbf{B} 、 \mathbf{E} 、和 \mathbf{F} 。

对流参数化是台风数值预报的第二步。在数值模式中, 对流参数化方案是为了反映次网格尺度对流作用而设计的, 其对于描述台风对流发展至关重要。目前, 台风模式研究和业务应用中广泛应用的是 Kain-Fritsch (KF93) [27] 对流参数化方案。KF93 方案为估计对流的云发展过程的不稳定能量及区域而采用的方法是拉格朗日气块法。具体细节: 首先比较在气块抬升凝结高度 (LCL) 的气块温度与大气温度, 来寻找上拽气流的源地。由于一般对流发生前, 气块温度比环境温度低, 气块难以抬升, 因此, 需要在参数化方案中定义一个温度初始扰动 δT_w 附加至气

块现有的温度上，以强迫气块抬升。具体如下：

$$\delta T_w = c_1 w_G^{1/3} \quad (2.5)$$

其中 $c_1 C^\circ s^{1/3}$; w_G 为抬升凝结高度的垂直速度 ($cm \cdot s^{-1}$)。

当 $T_{LCL} + \delta T_w \leq T_{ENV}$ 时，则不满足触发对流的条件，须在更高一层的模式层中判断是否为上拽气流的源地层，依次类推，直至找到源地层为止。相反，如经附加温度扰动的气块温度高于环境温度，则已满足触发对流的条件。显然，定义温度扰动 (δT_{vv}) 对于气块是否上升以及是否触发对流非常重要。

2.1.2 统计预报方法

一般来说，通过台风数值预报模式能对台风路径进行较为准确的预报，而台风强度预报的准确性不高。这是由于热带气旋强化过程的复杂性，使得强度预测技巧相对较低，因此强度预测也更加困难。此外，目前的数值模型无法以实时和可操作的方式明确解决对流问题，因此需要用其他强度预测方法来克服现有的数值模型的缺陷。通常，当无法通过机理分析建立模型时，可以采取收集大量数据的方法，通过对数据的统计分析建立模型。因为在大量的观察下，数据会表现出一定的规律性，这种规律可以通过函数关系式表示，函数关系式则称为回归方程，统计回归模型确定的是变量之间的相关关系。但是和台风相关的很多数据是无法通过实测得来的，所以仍然需要利用数值模式的产品。理想的替代方法是利用统计模型和数值模型的优势，将统计方法与数值天气预报的环境预测结合起来，这种方法通常被称为统计学 - 动力学方法。其中比较典型的是台风强度统计预测模式 (Statistical Typhoon Intensity Prediction Scheme, STIPS) [29]，为美国联合台风预警中心 (Joint Typhoon Warning Center, JTWC) 提供为期 5 天 12 小时间隔的热带气旋强度指导。STIPS 是基于统计 - 动力的框架的一个多元线性回归模型，需要利用大量从美国海军实用全球大气预测系统 (Navy Operational Global Analysis and Prediction System, NOGAPS) 获取到的环境信息和 JTWC 的历史最佳路径数据来构建模型，其预测因子则选择和台风相关的众多变量。相似的还有飓风统计预测模式 (The Statistical Hurricane Prediction Scheme, SHIPS) [30-32]，该模型早已经开发出来供北大西洋和东太平洋使用。在大西洋和东太平洋地区的预测指南中，SHIPS 模型在大西洋和东太平洋的热带气旋预报中产生了较小的错误。在北太平洋西部地区，联合台风警报中心 (JTWC) 通常会在 72 小时内进行强度预测，2003 年后开始进行 120 小时的预报。例如 2017 年 24、48、72、96 和 120h 的预测误差分别为 10.7、16.1、20.0、22.4 和 19.04kt，而根据 120 小时的台风强度统计预报所作的预测 12.4、19.7、24.7、26.2 和 26.7kt^[3]。

表 2.1 STIPS 的潜在静态预测因子

预报因子	描述
VMAX	初始强度
VMAX ²	初始强度的平方
DVMX	12 小时强度的改变
JDAY	日期的年日绝对值 -248
SPD	风暴移动速度

本部分主要以 STIPS 为例，介绍台风统计预报方法。重点关注潜在预测因子的选择、统计方法的实施步骤。STIPS 是一个多元线性回归模型，因变量为风暴从初始预报时间 (DELV) 开始，在登陆前每 12h 间隔的强度变化。自变量（潜在的预测因子）为文献中与热带气旋强度变化有关的参数。在构建好模型后通过预测热带气候强度的变化评估模型潜在预测因子的综合统计能力。由于要预测 12-120 小时的状况，以 12 小时为时间间隔，因而这就导致了 10 个预测方程。这样产生的方程可以预测和环境气候变化趋势有关的强度变化，但不能预测登陆后的强度变化。

STIPS 开发中能使用的潜在预测因子可以分为两类：(1) 与气候学、持久性和强度趋势有关的预测因子”静态预测因子”。(2) 与当前和未来大气环境和海表温度 (Sea Surface Temperature, SST) 条件有关的因素”时间依赖性预测因子”。

潜在的静态预测因子是初始化时的日期、当前强度和 12h 内的强度变化、运动和位置。STIPS 中使用的静态预测因子是在 ST5D^[33] 模型中提出的，如表 2.1 所示。

潜在的”时间依赖性预测因子”首先从最大潜在强度 (MPI) 开始。MPI 作为 SST 的函数，可以表示热带气旋强度能够达到的最大上限。Emanuel (1988)^[34] 表示 MPI 可以从理论上估计，不过为了发展 STIPS，DeMaria and Kaplan (1994)^[35] 给出了一个经验方程，如方程 2.6 所示。其系数分别为 $A=38.21\text{kt}$ ， $B=170.72\text{kt}$ ， $C = 0.1909^\circ\text{C}^{-1}$ 和 $T_0 = 30.0^\circ\text{C}$ ：

$$MPI = A + Be^{C(T-T_0)} \quad (2.6)$$

在 STIPS 的方程中，MPI 最大允许值为 185kt。

此外，环境相对湿度 (RH) 的变化会通过不饱和空气的夹带影响对流浮力。在热带气旋中，由于对流利用的势能相对较小，在中心附近降低到几乎为零^[36]。因此，中间层的相对湿度应该相对较大，从而可以减少干燥空气进入积云对流。而对流是热带气旋能量的直接来源，中层和上层的 RH 变化会直接影响热带气旋强化率，所以需要根据环境中上层水分检查对热带气旋强度的潜在影响。这里计算了距离气旋中心 200-800 公里区域范围的大气层 850-700hpa 的 RH(RHLO) 以及

500-300hpa 的 RH(RHHI), 这个区域平均值用来估计环境要素的平均值。在 200hpa 时, 考虑了纬向风 (U_{200})、温度 (T_{200})、散度 (δ_{200}) 和相对涡度 ($REFC$)。纬向风和温度同样取 200-800 公里区域范围内的平均值, 而散度值取更大的区域范围 1000 公里的平均值, $REFC$ 取 600 公里区域范围的平均值:

$$REFC = -r^{-2} \frac{\partial}{\partial r} (r^2 \overline{U'_L V'_L}) \quad (2.7)$$

其中 U 是径向风, V 是切向风, r 是半径, $\frac{\partial}{\partial r}$ 表示偏离方位平均值, 下标 L 表示计算是跟踪风暴风暴进行的。

除了上述这些预测因子, 还考虑了 200hpa 的一些预测因子, 1000km 区域范围内 850hPa 的涡度平均值 (ξ_{850}), 200-800hpa 间的垂直风切变值。计算垂直风切变与计算相对湿度时一样, 风暴的核心区域被移除, 以用于测量周围的垂直风切变。其传统的垂直风切变计算方法只是简单地使用不同压力层之间的标量差。采用这种方法, 可以得到两个预测因子: 200-850hpa 风差 (SHRD) 和 500-850hPa 风差 (USHRD)。除了 SHRD 和 SHRS 之外, 还创建了这些压力层的纬向风切变 (USHRD, USHRS)。作为垂直风切变传统测量方法的一种替代方法, 可以计算一般的垂直风切变。另一种方式是计算某一点的广义剪切, 如 2.8 所示。由于高纬度地区的风暴对垂直风切变的敏感性较低, 因此也测试了 SHRG 这两个因子, 此外还有 SHRD 和 SHRG 乘以路径变化过程中的纬度值。如表 2.2 所示, 增加这些预测因子, 可以得出 15 个可用于 STIPS 测试的天气尺度预测因子。

$$SHRG = 4.0 \times \sum_{p=850}^{p=200} w_p \sqrt{(u_p - \bar{u})^2 + (v_p - \bar{v})^2} \quad (2.8)$$

其中

$\bar{u} = \sum_{p=850}^{p=200} w_p u$ 是深层纬向风; $\bar{v} = \sum_{p=850}^{p=200} w_p v$ 是深层经向风; 且 w_p 是权重。

在构建多元回归模型时, 通常选择预报因子的一种方法是根据它们相互组合的能力来进行预报, 其实现方法可称为逐步预测法。逐步预测法第一步是根据每个预测时长来从预测因子集中选择因子, 然后对单个预测因子进行标准的 F 检验。在最初模型里, 单个预测因子要求有 99% 的统计意义。一旦在模型加入其他因子导致结果小于 98% 后时, 就能去掉这个新加的因子。由于模型构建时使用两种不同的方法来测量垂直风切变, 即 SHRG 项和两层标量差 SHRD 和 SHRS, 从而建立了两个预测因子集。这两个预报因子集除开与表 2.1 和表 2.2 所示垂直风切变因子有不同的处理外, 其他是相同的。逐步预测法在每个不同预测时长均可识别其重要的预测因子, 这些因子有时在预测时会不稳定。为了避免这个问题, 任何预测时长采用逐步选择法所选出的一组预测因子, 从而为每个预测时长创建一个单一的多元回归模型。

表 2.2 STPIS 的潜在综合预测因子

预报因子	描述
MPI	最大潜在强度
MPI^2	最大潜在强度的平方
MPI * VMAX	最大潜在强度 * 初始强度
RHLO	850-700hPa 区域 (200-800 km) 平均相对湿度
RHHI	500-300hPa 区域 (200-800 km) 平均相对湿度
U_{200}	200hpa 区域 (200-800 km) 平均纬向风
T_{200}	200hpa 区域 (200-800 km) 平均温度
δ_{200}	200hpa 区域 (0-1000 km) 平均散度
REFC	600km 内的相对涡通量散度
SHRG	200-850hPa 垂直风切变
SHRS	500-850hPa 区域 (200-800 km) 平均风切变
SHRD	200-850hPa 区域 (200-800 km) 平均风切变
USHRS	500-850hPa 区域 (200-800 km) 平均纬向风切变
USHRD	200-850hPa 区域 (200-800 km) 平均纬向风切变
SHRD * SIN(LAT)	200-850hPa 区域平均风切变 * 纬度的 sin 值
SHRG * SIN(LAT)	200-850hPa 垂直风切变 * 纬度的 sin 值
ξ_{850}	850hPa 区域 (0-1000km) 平均相对涡度

统计方法在热带气旋生成和强度预报中的其他应用如下：

热带气旋 (Tropical cyclone, TC) 生成预测已经在以往的研究中进行了不同时间尺度的预测，包括长期预测 (Emanuel 2013; Murakami 2010)，以及短期预测 (Halperin et al. 2013; Shen et al. 2010)。长期预测通常是利用大尺度环境场信息，建立环境因子与热带气旋活频数的统计关系，构建出热带气旋潜在生成指数 (Genesis Potential Index, GPI)，以此进行热带气旋活频数的季节预报。GPI 是热带气旋生成可能性大小的空间分布函数，通常因子的选择对 GPI 的计算起到了关键作用。Gray (1968) [37] 提出了基于低层涡度、中层湿度、对流不稳定性、海洋混合层厚度、水平风垂直切变、科氏参数这六个驱动因素乘积的季节性指数。在 Gray 的基础上，Emanuel (2004) [38] 提出了一个新的热带气旋潜在生成指数 (Genesis Potential Index, GPI)，该指数涉及了最大潜在强度、绝对涡度、700hPa 相对湿度、850 和 200hPa 的风切变等相关大尺度环境因子。Camargo (2007) [39] 根据 ENSO 对热带气旋活动的影响，认为低空涡度 (850 hPa)，600 hPa 相对湿度，850 ~ 200 hPa 垂直风切变强度和潜在强度 (PI) 等 4 个因子对成因潜势指数有贡献，于是建立新的 GPI。由于大多数 GPI 是针对全球热带气旋构建的，对于区域性热带气旋生成的频数和位置的模拟不太理想。于是赵军平等 (2012) [40] 在南海选取 850

hPa 绝对涡度、200hPa 和 850hPa 间的垂直风切变、500 hPa 的垂直速度、最大可能强度进行了改进。并在 El Nino 和 La Nina 事件中进行了验证，得到了很好的模拟效果。张敏 (2016) [41] 构建了一个新的 TC 生成指数 (GPIocean)，不仅考虑了必要大气因素，还考虑了海洋因素，能较好地刻画 TC 生成的长期（气候态）变化，其中包括 100h Pa 绝对涡度，净海表面长波辐射，上混合层的平均海洋温度以及 26°C 等温层的厚度。对于短期 TC 生成的预测问题，通常情况下是考虑热带洋面上已存在有热带扰动、热带云团，这些小涡旋被称之为 TC 生成前的“胚胎”，然后判断这些胚胎是否会发展成为 TC。对于“胚胎”是否存在的判断，一部分研究是根据大气变量的状态（风场和涡度场等）确认是否有热带扰动的存在 (S.Peng,Bing Fu,2010) [42, 43]，另一部分研究是根据卫星云图的亮温数据，设立亮温阈值，挑选出满足条件的热带云团 (Hennon,2003,2011)[44, 45]。现有预测研究中，短期预测的时次一般为 6h、12h、18h、24h、36h、48h、60h、72h、84h、96h、120h，预测方法大多基于统计预测模型，而预测因子则从 Gray 提出的 6 个 TC 生成的要素—涡度、垂直风切变、湿度、SST、纬度、大气不稳定性，然后加入海表因子与大气诊断量等不断进行改进。上述季节预测因子对于 TC 短期生成也起到了辅助决策作用。Hennon (2003) [44] 利用判别分析法对热带云团进行分类以预测 TC 是否会生成，取得了一定的预测效果，并在 2005 年改进了模型，进一步提升了预测的性能。Schumacher(2009)[46] 通过将研究区域划分为的子区域，根据子区域内的环境条件利用判别分析筛选出不利于 TC 生成的点，然后对 TC 进行 24h 预测，并生成了可以进行 6h 实时更新的 TC 生成预测产品，输入定义的相关参数值即可预测台风生成的概率。Chand 和 Walsh (2011) [47] 选择针对特定海域的预测因子使用概率回归模型来预测南太平洋（斐济，萨摩亚和汤加）地区 48h 的 TC 发生。N jaswal(2011)[48] 年提出了一种新方法，建立 24h-96h 发展为 TC 系统的 QuikSCAT 风场数据库，然后根据当前风场与数据库的匹配度是否达到 TC 生成的阈值来预测是否有 TC 生成，并在 2013 年利用 OASCAT 风场进行了数据库的更新。

对于台风强度预报，较早的统计预报方法是 Jarvinen adn Neumann (1979) [49] 研制的统计飓风强度预报方法 (Statistical Hurricane Intensity Forecast, SHIFOR)，该方法给予气候持续性因子（包括 TC 的当前位置和强度以及过去 12 的变化趋势等）进行 3 天的预报，适用于大西洋和东北太平洋海域。基于类似的预报因子，Elsberry et al (1988) [50] 提出了适用西北太平洋海域的 TC 强度统计预报方法 (Statistical Typhoon Intensity Forecast, STIFOR)。Chu (1994) [51] 年将该方法作了进一步改进，目前这两个方法可以提前 5 天进行预报。钮学新 (1984) [52] 建立了我国首个 TC 强度气候持续性预报方法，但仅适用于近海海域。目前在业务中使用的方法可以提前 3 天进行预报，适用于整个西北太平洋海域。另一类统计预报方

法除了考虑气候持续态因子外，还引入了当前和前期大气环境变量、海表面温度以及卫星图像建立预报模型。将天气因子用于建立台风强度预测模型的早期工作是始于 20 世纪 80 年代，主要的影响因子包括环境垂直切变、副热带高压、高空槽等。有研究从动力-热力学方程出发，得到反映台风强度变化的物理因子，然后通过建立这些因子与台风强度的统计关系进行预报^[29]。基于海表面温度计算得到的最大可能强度以及台风发展潜力对未来台风的强度变化有很好的预示作用^[31, 32]。除此之外，云顶亮温、红外云型非对称结构特征因子的引入可在一定程度上改进台风强度统计预报。除了多元线性回归外，还可以使用逐步回归方法建立预报模型的有效技术，如耦合飓风强度预测方案 CHIPS(Emanuel, 2004)^[38]、新型海陆网络模型 ICON(Grandau and Engel, 2004)^[53]，ST5D(Knafl et al, 2003)^[33]。台风强度的统计动力预报方法是基于数值天气预报模式，通过考虑未来大气环境和海洋状况的变化来建立预报模型。此类方法一般假定模式可以准确地预知未来大环境的变化，而且能够准确预报台风路径。最具有代表性的是美国国家飓风中心业务运行的飓风强度统计预报模型 (Statistical Hurricane Intensity Prediction Scheme, SHIPS)^[32]。该方法的回归系数在每年飓风季节结束后都要用最新的样本进行更新，并继续在业务使用的过程中得到不断改进。(Demaria 1994; 1997; 2005)^[31-33]。

2.1.3 机器学习方法

近年来，机器学习在众多领域内都取得了良好的效果，因此，也有许多学者尝试将机器学习应用于台风领域。这些方法能在一定程度上弥补以上数值预报方法和统计预报方法的不足。

本部分以机器学习在 TC 生成预测上的研究方法为例。该方法是为进行热带气旋生成预测中热带云团 (Tropical Cloud Clusters, TCC) 的预测分类提供方法。其主要思想是利用 IBTrACs 1979 至 2014 的热带气旋数据和 ERA-Interim 不同气压水平下的相应的局部环境状态数据，基于互信息，初步选取有利于 TC 生成的变量，并以此构成需要利用 Peter-Clark (PC) 算法进行概率图模型结构学习的节点集。在利用结构学习考虑到与其他变量的相互关系时，确定了对 TC 成因影响最大的变量，并在依赖性测试中对变量的最大观测 p 值进行排序后，利用 Logistic 模型进行预测建模。该方法采用概率图模型方法对局部环境数据进行了研究。概率图模型由代表随机变量的节点组成，并描述图中所有节点的联合分布。这个概率分布可以被称为图中连接节点子集上的某些函数的产物。图形模型还描述了节点之间的统计依赖结构，使图形模型结构学习成为发现因果关系的途径。通过确定图中边的方向可以从而获得因果关系。因此结构学习不仅可以考虑变量之间的关系，还可以考虑到变量集合之间的相互作用。Koller and Friedman (2009)^[54] 在对概率图

模型进行了详细描述后，利用概率图模型进行变量选择 (通常被称为特征选择)，这早已经在其他气候或海洋气候得到了实践 (Abramson et al., 1996; Ebert-Uphoff and Deng, 2012) [55,56]。对于该方法，使用的有向无环图 (DAG) 是结构化的，而节点使用的是输入数据集中的变量以及一个来编码是否形成了一个 TC 的二进制节点，后者被称为 TC 的起源节点。为了探索变量与 TC 起源节点的因果结构，理论上，TC 生成节点上的马尔科夫链应该给出相关的完整变量集。然而，首先对完整图进行了深入学习，其中使用了图的其余结构 (不包括 TC 起源节点上的马尔科夫链)。最终图也可以看作是一个时空随机场。也就是说，所有节点都在空间上指定气候状态，除了对于 TC 起源节点的规定，它指定了是否会在以后发生。此外，气候状态可以在每个时间点进行更新，从而导致在选定的局部环境中对于 TC 起源的不同概率观点。

该方法需要基于结合贪婪的搜索和约束算法的得分算法对图结构进行学习。通过对各种算法的探索，选择了基于约束的 Peter-Clark (PC) 算法 (Spirtes et al. 2000) [57] 来进行图形模型结构学习。这个算法具有理论背景，因为满足了四个假设。这些假设是 (1) 选定的变量形成了一个因果集，(2) 每个变量之间的因果关系对每个记录都有相同的因果关系，(3) 观测到的变量的联合分布符合因果结构的一个 DAG，(4) 该算法所要求的统计决策对于集合点来说是正确的。实践中往往无法满足四个假设中的一些假设。例如，对于连续变量的概率图模型结构学习的条件独立测试通常用于测试遵循多元正态分布的相关可变性之间的线性相关性。这些分布和线性假设似乎不适合非高斯分布的气候变量，此外，气候过程之间的关系可能是非线性的。为了解决上述问题，将非参数独立性测试应用于离散化数据，并使用 G2 测试在剩余节点的一个 (可能是空的) 子集之间进行了独立性测试。

图结构的组合空间随着节点数的增加呈超级指数增长 (Robinson, 1977) [58]。因此即使是 PC 算法也不能评估每一个有向无环图，因为节点数量太大，无法成功完成算法的全部运行，这也是对离散变量而不是连续变量使用独立测试的结果。所以，在学习图形模型结构之前，需基于每个节点与 TC 起源节点之间的强度进行初始选择。然后进行基于互信息的变量评估，其中两个随机变量的互信息是变量间相互依赖性的量度。在消除变量的同时，还要确保剩余的变量能够给出不同的气候状态表示，这样节点才仍可以提供一个完整的结构。此外，图形化模型的学习也面临挑战，根据多个节点的等级组合对可用数据进行配对，往往会导致许多分组中没有或很少的成因观测，这意味着在对多个节点进行条件作用时，可能无法可靠地对涉及的离散变量进行可靠的测试。由此产生的虚假否定可能最终确定一个网络结构，而在这种结构中，已经消除了已知对 TC 成因有很大影响的变量。因此这就违反了上面提到的 PC 算法的要求，即 (4) 中所要求的统计决策应该对

所有集合点是正确的。研究人员发现，存在解决方案可以避免这些问题，因为依赖性测试对于只包含两个节点的条件集来说是成功的。因此，该方法采用的图形模型结构学习方法如下：

该算法是从全连通网络开始的。在第一个回合中，基于对这些边的不明确的条件独立性测试，考虑重新移动边。然后进行两轮条件独立测试，测试每对节点之间的条件独立分别给出一个和两个其他节点。随后，停止结构学习算法。网络的目的是找到与 TC 起源节点有最强连接的节点，以便进行变量排序。对于一个由 25 个节点和一个固定时间的网络，使用这种方法总共进行了约 17 万次独立测试。变量根据所有 (条件) 独立性测试中的最大观测 p 值，将变量与 TC 起源节点进行排序。该最大观测 p 值在变量 i 和 TC 起源节点之间，可以 2.12 用来表示：

$$\xi_i = \max_{\forall S \in \Phi_i} Pr(U_{i|S} > g_{i|S}) \quad (2.9)$$

用 G2 统计对 ξ_i 和 TC 生成节点 s 的相关性测试的值和 $g_{i|s}$ 相比； Φ_i 是通过 PC 算法对 ξ_i 和 TC 生成节点考虑的所有 0、1、2 个节点集合。最后， $U_{i|S}$ 是一个随机变量，具有 χ^2 分布和自由度的随机变量。最大观测 p 值越低表明该节点与 TC 生成的联系越强。利用 Le et al(2016)^[59] 对 PC 算法进行并行处理，在 R 语言中进行了概率图模型结构学习的实验，最后的结构学习是使用 R 包中的 PC-stable (Kalisch et al, 2012)^[60]。

在进行完变量的确定后，采用逻辑斯蒂回归^[61]进行预测。逻辑斯蒂回归模型 (Logistic Regression, LR) 作为机器学习中一种常用的分类模型，可由条件概率 $P(Y|X)$ 表示。随机变量 X 取值为预测变量，随机变量 Y 为预测目标，取值为生成 ($Y = 1$) 或不生成 ($Y = 0$)。对于给定的输入样本 x ，按照 2.13 可以求得 $P(Y = 1|x)$ 和 $P(Y = 0|x)$ ，LR 通过比较两个条件概率值的大小，将输入样本 x 分到概率值较大的那一类。有时可将权值向量与偏置项合并，仍记作 w ， x ，即 $w = (w^1, w^2, \dots, w^n, b)^T$ ， $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)}, 1)^T$ ，这时，LR 如下：

$$\begin{aligned} P(Y = 1|x) &= \frac{e^{w \cdot x + b}}{1 + e^{w \cdot x}} \\ P(Y = 0|x) &= \frac{1}{1 + e^{w \cdot x}} \end{aligned} \quad (2.10)$$

在模型训练时，对于给定的变量数据集 $T = (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ ，其中， $x_i \in \mathbf{R}^n$ ， $y_i \in (0, 1)$ ，可以使用极大似然法来估计模型的参数，从而得到训练好的 LR 模型。

假设： $P(Y = 1|x) = \pi(x)$ ， $P(Y = 0|x) = 1 - \pi(x)$

似然函数则为

$$\prod_{i=1}^N [\pi(x_i)]^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (2.11)$$

对数似然函数为

$$\begin{aligned} L(w) &= \sum_{i=1}^N N[y_i \log \pi(x_i) + (1 - y_i) \log(1 - \pi(x_i))] \\ &= \sum_{i=1}^N N[y_i \log \frac{\pi(x_i)}{1 - \pi(x_i)} + \log(1 - \pi(x_i))] \\ &= \sum_{i=1}^N N[y_i(w \cdot x_i) - \log(1 + e^{w \cdot x_i})] \end{aligned} \quad (2.12)$$

对 $L(w)$ 求极大值, 得到 w 的估计值。假设 w 的极大似然估计值 \hat{w} , 那么学到的 LR 模型为

$$\begin{aligned} P(Y = 1|x) &= \frac{e^{\hat{w} \cdot x + b}}{1 + e^{\hat{w} \cdot x + b}} \\ P(Y = 0|x) &= \frac{1}{1 + e^{\hat{w} \cdot x + b}} \end{aligned} \quad (2.13)$$

由此训练出的模型可以用于热带气旋生成预测的热带云团预测分类。

其他机器学习方法在台风领域的应用如下:

姚才 (2007) [62] 针对南海的热带气旋样本, 采用遗传算法与神经网络相结合的方法进行了热带气旋强度预报建模研究, 并根据相同的热带气旋个例与气候持续法热带气旋强度预报方法进行对比分析。试验预报结果表明遗传 - 神经网络方法具有更好的预报能力。郑颖青 (2013) [63] 根据最佳台风路径资料的 TC 位置、强度、路径长度和方向等主要信息的特征参数, 利用 k- 均值聚类法建立了西北太平洋 TC 路径的客观分类方法, 并对分类得到的 7 个路径类型 TC 在生命史、能量、活动季节和登陆概率等方面进行了对比分析, 从而得出各类 TC 的这些特征量确实具有明显的差异。Zhang et al (2013) [64] 将 TC 强度的 24 小时变化 (即强化和弱化) 视为二元分类问题, 然后应用 C4.5 算法对西北太平洋热带气旋 (TC) 强度变化进行了分类, 达到了较高的准确率。此研究的结果可用 TC 强化的理论来解释, 因此对于 TC 强度变化过程有一些指导意义, 并有可能改善 TC 强度的预测。在 2015 年对西北太平洋热带扰动的分类预测使用了决策树算法, 不仅得到了预测的效果, 而且确认了 TC 生成的变量阈值条。Wijnands (2016) [65] 基于互信息对 TC 生成的变量进行初始选择, 然后通过结构学习识别对 TC 成因影响最大的变量, 并同时考虑了与其他变量的相互关系。再根据 TC 变量与 TC 发生节点的所有 (条件) 独立测试中观察到的最大 p 值对变量进行排序后, 选取最大 p 值的前 5 个

点作为预测变量，最后利用逻辑回归进行预测建模确认了排名靠前的变量的卓越表现。此研究利用机器学习及统计方法建立了一套有效的变量筛选模型，对于 TC 的预测研究有很重要的意义。Chaudhuri (2017) [66] 引入蚁群优化 (ACO) 技术形式的群体智能来计算热带气旋轨迹的信息增益，然后通过神经网络预测北印度洋孟加拉湾周围发生的气旋的最大持续风速。结果表明，6h、12h、18h 和 24 h 热带气旋最大持续风速预报误差分别为 2.6m/s、2.9m/s、3.1m/s 和 4.8m/s。其结果与现有的动力、统计和自适应模型进行比较后发现有了较大的提升，而且很好地验证了观测。Li Yun (2017) [67] 利用深度学习的方法预测 TC 快速增强的条件。实验表明，长期短期记忆 (LSTM) 网络可通过利用过去状态预测 TC 快速增强的能力，这对 TC 快速增强的研究有很重要的意义。

2.2 相关深度学习网络

深度学习可以简单认为是层数更深的神经网络模型，而神经网络，则可类比为由适应性的神经元通过并行互联而构建组织成的生物网络。神经网络的目标是在数值计算上近似拟合某个函数 f 。例如分类器， $y = f(x)$ 将输入数据 x 预测分类为某个类别 y 。网络定义了映射 $y = f(x; \theta)$ ，并通过训练学习到参数 θ 的最优值，使它尽可能接近最优函数。神经网络的最基本的成分是人神经元模型，它的原理是模拟生物神经元。人工神经元在接收到来自 n 个神经元传递过来的信号后，将这些信号作为当前神经元的输入并通过带权重的连接后进行传递，并通过将总输入值将与当前神经元的阈值进行比较后传递给“激活函数”来产生输出。神经元模型中 x_i 为来自第 i 个神经元的输入， w_i 表示第 i 个神经元的连接权重， θ 为当前神经元的阈值输出 $y = f(\sum_{i=1}^n w_i x_i - \theta)$ ，如图 2.1。

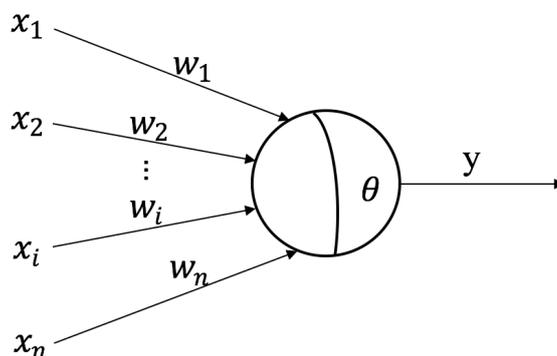


图 2.1 神经元模型示意图

f 表示“激活函数”。激活函数将输入值映射为输出值 0 或 1，典型的激活函数有 Sigmoid 函数、Tanh 函数、Relu 函数、Leaky Relu 函数，如图 2.2。具体函数形式表示如下：

$$\begin{aligned} \text{sigmoid}(x) &= \frac{1}{1 + e^{-x}} \\ \text{tanh}(x) &= \frac{e^x - e^{-x}}{e^x + e^{-x}} \\ \text{relu}(x) &= \begin{cases} x & x > 0 \\ 0 & \leq 0 \end{cases} \\ \text{leaky relu}(x) &= \begin{cases} x & x > 0 \\ \lambda x & x \leq 0 \end{cases} \end{aligned} \quad (2.14)$$

神经网络则是把许多个这样的神经元按一定的层次机构连接起来，可以理解为神经网络是包含了许多参数的数学模型，这个模型是若干个函数，例如 $y_j = f(\sum_{i=1}^n w_i x_i - \theta_j)$ 相互嵌套而得。

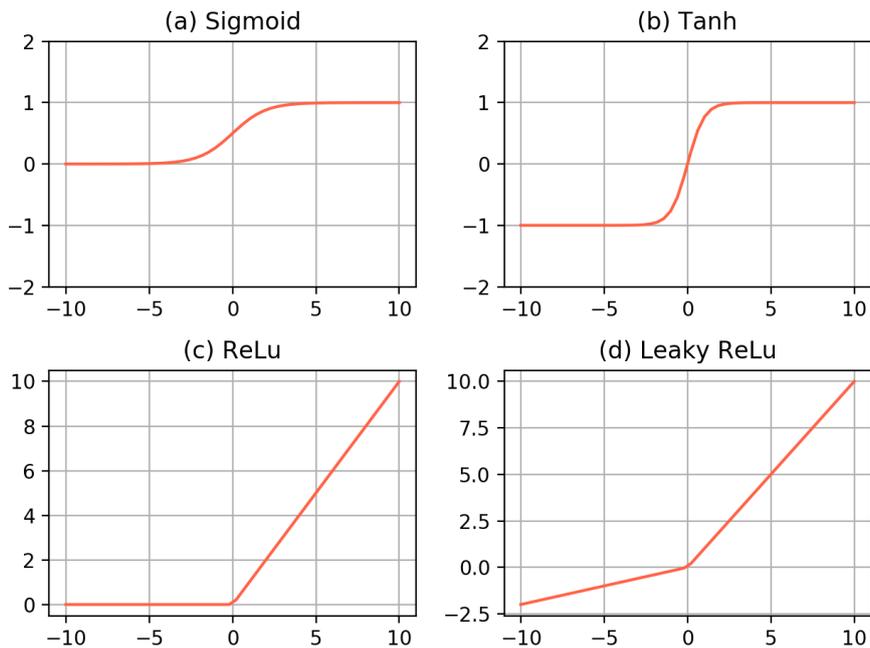


图 2.2 典型神经元激活函数

2.2.1 前馈神经网络

最早的神经网络模型为感知机模型，感知机是由输入层和输出层这两层神经元组成的网络，输入层含两个神经元，输出层只含单个神经元。该模型只能应用

于完成简单的线性可分任务，甚至解决不了非线性问题，如异或问题。要解决非线性可分问题则需尝试构建多层神经元，网络中输入层与输出层之间的所有层都被称为隐藏层。常见的神经网络是如图 2.3 所示，每一层中的神经元都与下一层的神经元互相连接，这些神经元之间不存在同层之间的连接，更不存在跨层的互连，这种神经网络称为“多层前馈神经网络”，也称为“多层感知机”。多层网络的学习能力相较于单层感知机得到了很大地提升，欲训练多层神经网络，需要强大的学习算法，误差反向传播算法（BP）算法是目前最成功也是最常用的模型学习算法。不仅可以用于训练多层前馈神经网络模型，还可以用于训练循环神经网络。

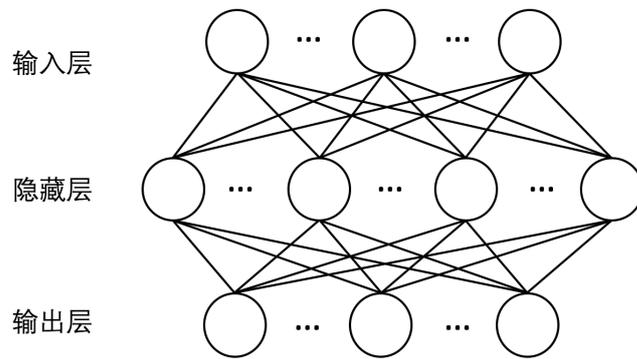


图 2.3 多层前馈神经网络示意图

BP 算法^[68]具体如下：

给定训练集 $D = (x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$, $x_i \in \mathbb{R}^p$, $y_i \in \mathbb{R}^q$, 即有 m 个样本，每个输入样本有 p 个属性，每个样本输出为 q 维向量。因此，给定输入层含 p 个输入神经元，输出层含 q 个输出神经元，且含 t 个隐藏层神经元的多层前馈网络结构。其中输出层第 j 个神经元的阈值用 θ_j 表示，隐藏层第 h 个神经元的阈值用 γ_h 来表示。输入层第 i 个神经元与隐藏层第 h 个神经元之间的连接权为 v_{ih} ，隐藏层第 h 个神经元与输出层第 j 个神经元之间的连接权为 w_{hj} ，令第 h 个神经元接收到的输入为 $\alpha_h = \sum_{i=1}^p v_{ih}x_i$ ，输出层第 j 个神经元接收到的输入为 $\beta_j = \sum_{h=1}^t w_{hj}b_h$ ，其中 b_h 为隐藏层第 h 个神经元的输出。假设隐藏层与输出层神经元均使用图 2.2 中的 Sigmoid 函数。

对训练集样本 (x_k, y_k) ，假定神经网络的输出为 $\hat{y}_k = (\hat{y}_1^k, \hat{y}_2^k, \dots, \hat{y}_q^k)$ ，即

$$\hat{y}_j^k = f(\beta_j - \theta_j) \quad (2.15)$$

则网络在 (x_k, y_k) 上的均方误差为

$$E_k = \frac{1}{2} \sum_{j=1}^q (\hat{y}_j^k - y_j^k)^2 \quad (2.16)$$

图 2.3 网络中共有 $(p+q+1)t+l$ 个参数需要进行选择设置, 包括输入层到隐藏层的 $p \times t$ 个权值参数, 隐藏层到输出层的 $t \times q$ 个权值参数、 t 个隐藏层神经元的常量阈值、 q 个输出层神经元的常量阈值。而参数的更新采用迭代的思想, 任意参数 v 的更新估计为

$$v \leftarrow v + \Delta v \quad (2.17)$$

具体推导以隐藏层到输出层的权重参数 w_{hj} 为例:

BP 算法基于梯度下降算法, 根据目标的负梯度方向调整参数, 假设误差为 E_k , 学习率为 η , 有

$$\Delta w_{hj} = -\eta \frac{\partial E_k}{\partial w_{hj}} \quad (2.18)$$

根据“链式法则”, w_{ij} 首先直接影响输出层第 j 个神经元的输入值 β_j , 再间接影响其输出值 \hat{y}_j^k , 然后影响随着网络被传播到 E_k , 因而

$$\frac{\partial E_k}{\partial w_{hj}} = \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial w_{hj}} \quad (2.19)$$

根据 β 的定义, 有

$$\frac{\partial \beta_j}{\partial w_{hj}} = b_h \quad (2.20)$$

对 sigmoid 函数求导, 有

$$f'(x) = f(x)(1 - f(x)) \quad (2.21)$$

根据式 (2.2) 和式 (2.1), 有

$$\begin{aligned} g_j &= -\frac{\partial E_k}{\partial w_{hj}} = \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \\ &= -(\hat{y}_j^k - y_j^k) f'(\beta_j - \theta_j) \\ &= \hat{y}_j^k (1 - \hat{y}_j^k) (y_j^k - \hat{y}_j^k) \end{aligned} \quad (2.22)$$

将 (2.8) 代入 (2.6) 和 (2.5), 再代入 (2.4), 即可得到 BP 算法对于 w_{hj} 的更新公式

$$\Delta w_{hj} = \eta g_j b_h \quad (2.23)$$

同理可得

$$\Delta \theta_j = -\eta g_j \quad (2.24)$$

$$\Delta v_i h = \eta e_h x_i \quad (2.25)$$

$$\Delta \gamma_h = -\eta e_h \quad (2.26)$$

式 (2.11) 和 (2.12) 中

$$\begin{aligned} e_h &= -\frac{\partial E_k}{\partial b_h} \cdot \frac{\partial b_h}{\partial \alpha_h} \\ &= -\sum_{j=1}^l \frac{\partial E_k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} f'(\alpha_h - \gamma_x) \\ &= \sum_{j=1}^l w_{hj} g_j f'(\alpha_h - \gamma_h) \\ &= b_h(1 - b_h) \sum_{j=1}^l w_{hj} g_j \end{aligned} \quad (2.27)$$

学习率 $\eta \in (0, 1)$ 代表算法在每一轮迭代中更新的步长, 若太大则容易越过最优值导致振荡, 若太小则会使得网络的收敛速度过慢。

Hornik et al 证明, 前馈神经网络只需要一个包含足够多神经元的隐藏层就能以任意精度逼近任意复杂度的连续函数。然而, 如何设置隐藏层神经元的个数仍然是至今为止未解决的问题, 实际应用中通常靠“试错法”来进行调整。不过, 正是由于其强大的表示能力, BP 神经网络经常会“过拟合”, 其训练误差持续降低, 但测试误差却上升。缓解网络过拟合问题一般有两种方法, 一是“早停”, 即若训练集的误差减小但验证集误差却增大, 则停止训练, 并确定能得到最小验证集误差的连接权和阈值。二是“正则化”, 其主要思想是在目标误差函数中增加一个近似度量网络结构复杂度的项, 例如权重参数与阈值的平方和, 本部分参考《机器学习》[12] 以及《Deep Learning》[69]。

2.2.2 卷积神经网络 (CNN)

卷积神经网络 (Convolutional Neural Networks, CNN) 是一种典型的前馈神经网络。在结构上有三个特性, 分别是局部连接, 共享权重以及 (时间或空间上的) 次采样。这些结构特性使得卷积神经网络具有一定程度上的平移、缩放和扭曲不变性 (LeCun et al, 1998)^[70]。

卷积是分析数学中一种重要的运算。这里只考虑离散序列的情况。一维卷积经常用于信号处理中。假设输入信号为序列 x_t , 其中 $t = 1, \dots, n$ 和滤波器 f_k , 其中 $k = 1, \dots, m$, 且 $m \ll n$ 。卷积后的输出为:

$$y_t = \sum_{k=1}^m f_k x_{t-k+1} \quad (2.28)$$

当滤波器 $f_k = \frac{1}{m}$ 时, 卷积的意义是信号序列的移动平均。卷积的结果按不同的输出长度可分为两类: 一类是宽卷积, 输出长度 $n + m - 1$, 对于不在 $[1, n]$ 范围之外的 x_t 用零补齐 (zero-padding)。一类是窄卷积, 输出长度 $n - m + 1$, 不补零。

二维卷积经常用于图像处理。给定一个图像矩阵 X_{ij} , 其中 $1 \leq i \leq M, 1 \leq j \leq N$ 和滤波器 f_{ij} , 其中 $1 \leq i \leq m, 1 \leq j \leq n$, 一般 $m \ll M, n \ll N$ 。卷积的输出为:

$$y_{ij} = \sum_{u=1}^m \sum_{v=1}^n f_{uv} x_{i-u+1, j-v+1} \quad (2.29)$$

在全连接网络中, 如果第 l 层有 n^l 个神经元, 第 $l-1$ 层有 $n^{(l-1)}$ 个神经元, 连接边有 $n^l \times n^{(l-1)}$ 个, 也就是权重矩阵有 $n^l \times n^{(l-1)}$ 个参数。当 m 和 n 都很大时, 权重参数会非常多。而如果采用卷积层代替全连接, 网络中第 l 层的每一个神经元都只和第 $l-1$ 层的一个局部窗口内的神经元相互连接, 组成的局部的连接网络会减少大量的权重参数。第 l 层的第 i 个神经元的输入定义为:

$$\begin{aligned} a_i^{(l)} &= f_l \left(\sum_{j=1}^m W_j^{(l)\top} a_{i-j+m}^{(l-1)} + b_i^{(l)} \right) \\ &= f_l (W^{(l)\top} a_{i+m-1:i}^{(l-1)} + b_i^{(l)}) \end{aligned} \quad (2.30)$$

其中 $W^{(l)} \in \mathbb{R}^m$ 为 m 维卷积核, $a_{i+m-1:i}^{(l-1)} = [a_{i+m-1}^{(l-1)}, \dots, a_i^{(l-1)}]^\top$

上述公式可以写为:

$$a^{(l)} = f_l (W^{(l)} \otimes a^{(l-1)} + b^{(l)}) \quad (2.31)$$

式中 \otimes 表示卷积运算

从上式可以看出, $W^{(l)}$ 对于全部神经元都保持相同, 从而体现了卷积网络权重共享的结构特性。

通常情况下，图像数据是以二维矩阵的数值形式表示，为了提取特征，仍然以矩阵的形式输入到神经网络中，因此需要二维卷积来进行处理。假设 $X^{(l)} \in \mathbb{R}^{w_l \times h_l}$ 和 $X^{(l-1)} \in \mathbb{R}^{w_{l-1} \times h_{l-1}}$ 分别是第 l 层和第 $l-1$ 层的神经元值。 $X^{(l)}$ 的每一个元素为：

$$X_{s,t}^{(l)} = f_l \left(\sum_{i=1}^u \sum_{j=1}^v W_{i,j}^{(l)} \cdot X_{s-i+u, t-j+v}^{(l-1)} + b^{(l)} \right) \quad (2.32)$$

其中 $W^{(l)} \in \mathbb{R}^{u \times v}$ 为二维卷积核， $b^{(l)}$ 为第 l 层的偏置。第 l 层的神经元个数为 $w_l \times h_l$ ，并且 $w_l = w_{(l-1)-u+1}$ ， $h_l = h_{(l-1)v+1}$ 。上式可以写为：

$$X^{(l)} = f_l(W^{(l)} \otimes X^{(l-1)} + b^{(l)}) \quad (2.33)$$

为了增强卷积层的表达能力，可以通过使用 K 个不同的卷积核来获得 K 组输出。如果把卷积核看成是一个特征提取器，那么每一组输出都可以看成是输入的图像数据经过某个特征抽取所得到的特征，从而每组输出也称为一组特征映射。为了不失一般性，若第 $l-1$ 层的特征映射组数为 n_{l-1} ，每组特征映射的大小为 $m_{(l-1)} = w_{(l-1)} \times h_{(l-1)}$ 。则第 $(l-1)$ 层的总神经元数为 $n_{l-1} \times m_{l-1}$ 。第 l 层的特征映射组数为 n_l 。若第 l 层的每一组特征映射 $X^{(l,k)}$ 的输入为第 $l-1$ 层的所有组特征映射。第 l 层的第 k 组特征映射 $X^{(l,k)}$ 为：

$$X^{(l,k)} = f_l \left(\sum_{p=1}^{n_{l-1}} (W^{(l,k,p)} \otimes X^{(l-1,p)}) + b^{(l,k)} \right) \quad (2.34)$$

其中， $W^{(l,k,p)}$ 表示第 $l-1$ 层的第 p 组特征向量到第 l 层的第 k 组特征映射所需的卷积核。

卷积层虽然相比于全连接网络能在一定程度上大幅减小连接的数目，但是每个特征映射的神经元个数并没有明显减少。假若特征图直接与分类器相连，那么输入分类器的维数依然会很高，从而导致过拟合问题。为了解决该问题，在卷积神经网络中一般会在卷积层之后进行池化操作（子采样），通过添加子采样层可以显著降低特征数据的维数，从而避免过拟合。

对于特征映射 $X^{(l)}$ ，可以将划分为多个子区域 R_k ，其中 $k = 1, \dots, K$ ，那么子采样函数 $down(\cdot)$ 可以定义为：

$$X_k^{(l+1)} = f(Z_k^{(l+1)}) = f(W^{(l+1)} \cdot down(R_k) + b^{(l+1)}) \quad (2.35)$$

其中 $W^{(l+1)}$ 和 $b^{(l+1)}$ 分别是可训练的权重和偏置参数。

其中 $down(X^l)$ 为子采样后的特征映射。子采样函数 $down(\cdot)$ 一般取区域的最大值 (*Maximum Pooling*) 或平均值 (*Average Pooling*)。

$$\begin{aligned} pool_{max}(R_k) &= \max_{i \in R_k} a_i \\ pool_{avg}(R_k) &= \frac{1}{|R_k|} \sum_{i \in R_k} a_i \end{aligned} \quad (2.36)$$

2.2.3 长短期记忆神经网络 (LSTM)

循环神经网络 (**Recurrent neural network, RNN**)^[71] 是一种不同于前馈神经网络和卷积神经网络的神经网络结构。其网络隐藏层之间的节点均有连接，隐藏层不仅包括输入层的信息，还包括上一时刻隐藏层的输出信息。其主要用途是能够通过挖掘数据中的时序信息来处理和预测序列数据，以及充分利用深度表达语义信息能力，使之在语音识别、机器翻译等方面取得重大突破。

正如图 2.4 的单个结构所示，在每个时刻 t ，循环神经网络会针对该时刻的输入结合当前模型的状态给出一个输出，并更新模型状态。网络的主体的输入除了来自输入层的 x_t ，还有上一个时刻的隐藏层状态 h_{t-1} ，在每一时刻，循环神经网络的主体在读取了 x_t 和 h_{t-1} 之后会生成新的隐藏层状态 h_t ，并产生本时刻的输出 o_t 。由于模块中的运算和变量在不同时刻是相同的，因此循环神经网络可以直观上被看作是同一网络的多次复制。正如卷积神经网络在不同的空间位置共享参数，循环神经网络在不同时刻共享参数，从而能够使用有限的参数处理任意长度的序列。

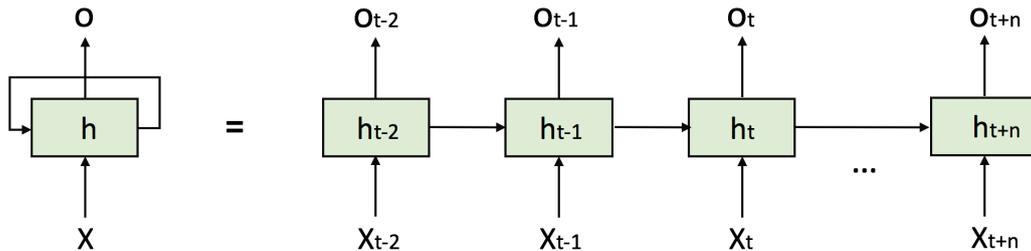


图 2.4 RNN 典型结构示意图

为了进一步阐明网络的结构，可作如下分解：

由于隐藏层状态依赖于上一时刻的隐藏层状态，因此可以认为 h_t 是一状态序列，且

$$h_t = f(h_{t-1}, \theta) \quad (2.37)$$

其中 f 为激活函数， θ 为参数，两者均与时间无关。

进一步可展开为

$$\begin{aligned}
 h_t &= f(h_{t-1}, \theta) \\
 &= f(f(h_{t-2}, \theta), \theta) \\
 &= \dots \\
 &= f(f((h_1, \theta), \theta), \dots, \theta)
 \end{aligned} \tag{2.38}$$

由于隐藏层状态不仅与上一时刻的状态有关，仍受输入信息 x_t 的影响，因此

$$h_t = f(h_{t-1}, x_t, \theta) \tag{2.39}$$

完整展开形式为

$$\begin{aligned}
 h_t &= f(h_{t-1}, x_t, \theta) \\
 &= f(f((h_{t-2}, x_t, \theta), \theta)) \\
 &= \dots \\
 &= f(f((h_1, x_t, \theta), \theta), \dots, \theta)
 \end{aligned} \tag{2.40}$$

加入权重信息后

$$h_t = f(W_h h_{t-1} + W_x x_t + b_t) \tag{2.41}$$

最后循环神经网络输出层的结果为

$$o_t = g(W_o h_t) \tag{2.42}$$

其中 g 为激活函数。

长短期记忆神经网络 (Long short-term memory, LSTM) 是由 Hochreiter and Schmidhuber (1997)^[72] 提出，并由被 Alex Graves(2013)^[73] 进行了改良和推广的一种典型的循环神经网络结构。LSTM 可保留误差，用于沿时间和层进行反向传递，并且能将误差保持在更为恒定的水平，让循环网络能够进行许多个时间步的学习，从而建立远距离因果联系的通道。该网络相比于循环神经网络的优势是能够处理长期依赖的问题，且能够解决由于网络太深，网络权值更新不稳定而造成的梯度消失问题。LSTM 同 CNN 一样，也是前馈神经网络的一种。

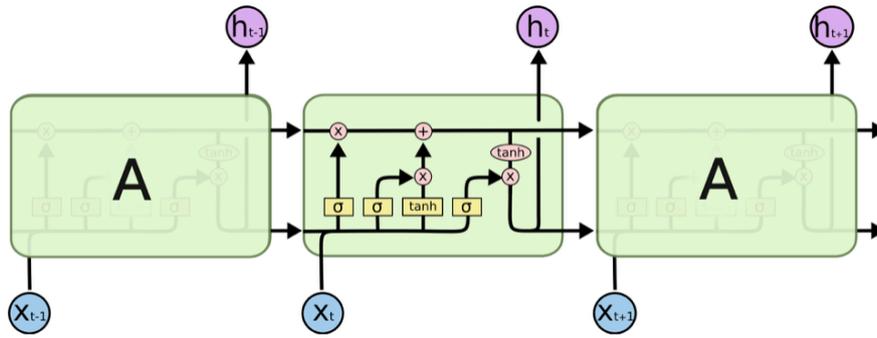


图 2.5 LSTM 典型结构示意图 (Christopher Olah)

如图 2.5, LSTM 通过三个门来控制信息的输入、更新、输出, 分别是忘记门、输入门、输出门。

LSTM 流程的第一步是决定从 *cell* 中扔掉什么信息, 需要对上一步 *cell* 中的信息进行选择性的遗忘。实现方式: 通过 *sigmoid* 层实现的“遗忘门”。以上一步的 h_{t-1} , 和这一步的 x_t 作为输入, 然后为 C_{t-1} 里的每个数字输出一个 $0 \sim 1$ 间的值, 表示保留多少信息 (1 代表完全保留, 0 表示完全舍弃), 然后与 C_{t-1} 相乘。

第二步是决定在 *cell* 里存什么。将新的信息选择性地记录到 *cell* 中。实现方式: 包含两部分, (1) *sigmoid* 层 (“输入门层”) 决定我们要更新什么值; (2) *tanh* 层创建一个候选值向量 \tilde{C}_t 将会被增加到 *cell* 中。将会在下一步把这两个结合起来更新 *cell*。

第三步是更新 *cell*, 更新旧的 *cell* 实现方式: $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$, f_t 表示保留上一次的多少信息, i_t 表示更新哪些值, \tilde{C}_t 表示新的候选值。候选值被要更新多少 (即 i_t) 放缩。这一步真正实现了移除哪些旧的信息, 增加哪些新信息。

最后, 决定什么信息需要被输出。实现方式: (1) 我们通过 *sigmoid* 层 (“输出层门”) 来决定输出新的细胞状态的哪些部分; (2) 然后将细胞状态通过 *tanh* 层 (使值在 $-1 \sim 1$ 之间), 然后与 *sigmoid* 层的输出相乘。因此 LSTM 能够输出想输出的部分。

第三章 基于深度学习的台风时空深度混合预测模型

3.1 引言

台风从生成到发展的过程是一个时空变化的过程，影响台风变化甚至是台风周围的气象海洋要素也是时空相关的变量。然而，现有的统计预报方法以及机器学习方法很少将这种时空关系考虑进来，通常，预报因子都只是作为独立变量进行预测，因此，可能是造成是预报效果不理想的原因之一。数值预报方法根据流体动力热力学方程考虑了这种时空关系，但是由于数值模式的计算复杂太大，且各种和台风相关的机制还不够完全，因此，虽然作为近些年来主流预报方法，但是仍然存在很大的缺陷。本文旨在利用现有的统计预报和机器学习等台风相关预测方法的基础上，结合深度学习中卷积神经网络获取空间特征和长短期记忆神经网络获取时序特征的能力，充分学习台风活动过程的时间关系，建立台风时空深度混合预测模型来对台风的各种关键问题进行预测。本章则主要介绍如何把台风预测问题转换为时空序列预测问题，然后提出了解决这个问题的方法，即基于深度学习的台风时空混合预测模型，并详细给出了设计该模型思路。

3.2 问题定义

台风作为一种极端天气事件，其形成到形成后的各种活动都与大气和海洋息息相关。大气海洋中常定义的一些变量如涡度、湿度等能够反映出台风的一些特性。而台风自身的特性也是常常需要通过大气变量来定义，比如本文中台风强度定义为中心附近的最大风速。因此大气海洋变量与台风特性之间存在某种潜在的对应关系。同时，随着台风的发生与发展中，这些大气海洋变量的变化以及台风自身特性的变化是同样存在内在规律，但由于现阶段对这种规律把握地不准确，造成了预测或者是预报的不准确性。多年来，台风对各个沿海国家的影响甚大，因此一直是为人们所关注的对象，历史资料也日益积累。虽然和台风相关的大气和海洋变量难以通过实际观测获取，但由于现有观测手段的不断发展，各种卫星遥感资料，再分析资料十分丰富。台风预测方法的研究由此有了强有力的数据支撑。深度学习乃至机器学习最大的优势就是能够从历史数据中发现规律用以预测未来，因此相关成熟的技术给台风预测方法的研究提供了重要途径。

本文希望通过历史台风资料，大气数据和海洋数据构建台风的预测模型，以能对未来台风的的活动进行预测，比如是否形成、路径变化、强度变化、风雨状况等。在实际观测中，台风形成前的热带气旋等级为热带风暴。由于从热带风暴发展到台风的过程是一个空间和时间变化的过程，台风之后的发展消亡也是空间和

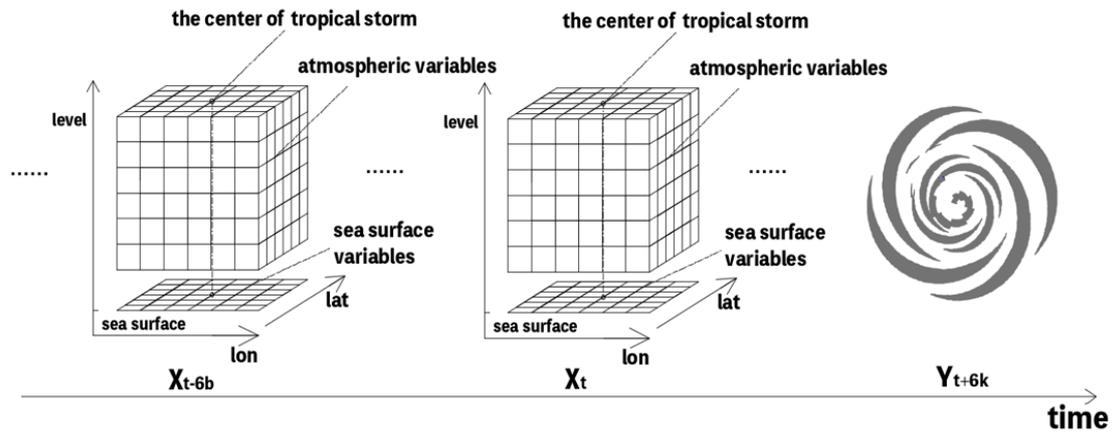


图 3.1 台风预测相关变量的时空结构示意图

时间变化的过程。在不考虑空间信息的情况下，热带风暴和台风的路径是时间序列，为分析获取其时序关系，需要采用多个时刻的大气和海洋数据来进行预测。但是热带风暴或者是台风中心的周围环境变量在任意时刻都是空间场，大气变量在空间中的三个维度包括经度、纬度、层。而与台风紧密相关的海面变量在空间中的维度只包括经度和纬度。如何从历史数据中获取台风形成之前与之后时空变量的规律来对未来的台风活动进行预测？这是本文需要考虑的问题。本文由此可以将台风预测问题定义为时空序列预测问题。

为了能通过深度学习的手段来解决台风预测这类时空序列预测问题，需要在已划分为空间网格点的区域中选择相应的数据。无论数据类型是再分析资料或者是卫星遥感的反演数据，选择的单层大气层或海表层的的相关变量数据都可以用二维网格图表示。给定热带风暴或台风中心周围的二维网格图有 $M \times M$ 个网格点。大气压力层的数量是 L ，大气变量的数量是 R ，海面变量的数量是 O 。因此，热带风暴中心或台风中心 X^P 周围的大气变量是 $M \times M \times L \times R$ 规格的网格数据集。热带风暴中心周围的海面变量 X^S 是 $M \times M \times O$ 规格的网格数据集。包含大气变量和海洋变量的空间环境变量由此可以表示为 $X = [X^P, X^S]$ 。定义预测当前时刻是 t ，预测当前时刻的环境变量是 X_t 。 Y 是预测对象，可以表示台风是否形成，路径上的中心经纬度位置，强度或者风雨量等， k 是预测时间步， b 是在预测当前时刻之前应该考虑的回顾时间步，每个时间步长为 6 小时。

因此，问题可以进一步被定义为给定预测当前时刻和回顾时刻风暴中心或者是台风中心周围环境变量网格数据集，预测台风在未来时刻的状态，如是否形成，形成后的强度、路径以及风雨等状态量。即给定时空变量序列 $X_{t-6b}(b = 0, 1, 2, 3)$ ，预测台风的状态量 Y_{t+6k} 。

简单表述如下:

$$Y_{t+6k} = f(X_t, X_{t-6b}), k, b = 0, 1, 2, 3... \quad (3.1)$$

Y_{t+6k} 是从当前预测时刻 t 开始 $6k$ 小时后的预测对象, f 是需要通过历史数据学习的模型, X_t 是当前预测时刻热带风暴或者台风中心周围环境变量的网格数据集, X_{t-6b} 是预测当前时刻之前 $6b$ 小时环境变量的网格数据集。图 3.1 表示台风预测相关时空变量结构。

3.3 模型框架

该模型的总体框架如图 3.2 所示。本文提出了台风时空预测模型来学习与台风相关的大气和海洋变量的时空相关性。其主要思想是使用 3DCNN 来学习大气变量的三维空间关系, 2DCNN 来学习海面变量的二维空间关系以及 LSTM 来学习台风变化过程中变量的时序关系。

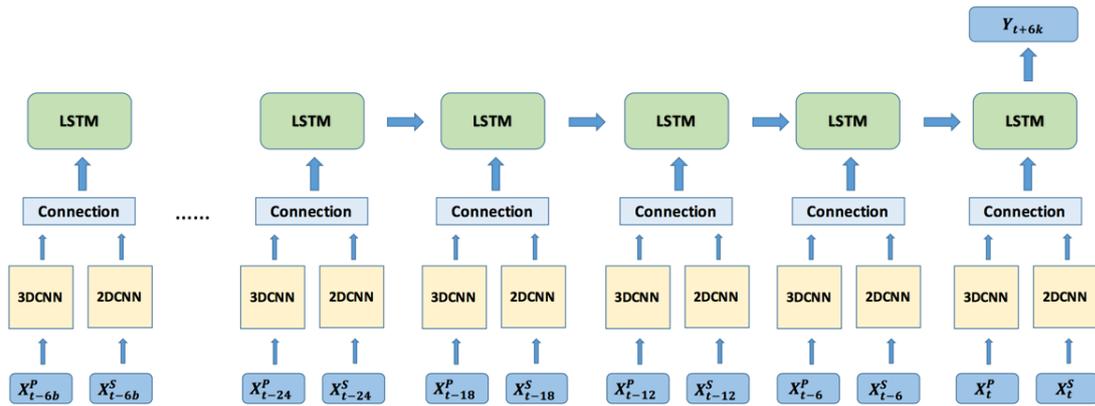


图 3.2 台风时空深度混合模型框架图

任意时刻热带风暴中心或台风周围的大气变量 X^P 都是 $M \times M \times L \times N$ 规格的网格数据集, 亦可作为 3DCNN 输入的 4D 张量。任意时刻热带风暴中心或台风周围中心周围的海面变量 X^S 都是 $M \times M \times O$ 规格的网格数据集, 亦可作为 2DCNN 输入的 3D 张量。该模型在“connection”部分拼接 3DCNN 和 2DCNN 的输出向量, 其长度为 K , 并作为 LSTM 的输入。

因此, 台风时空深度混合预测模型可以表述为:

$$Y_{t+6k} = LSTM(3DCNN(X_{t-6b}^P), 2DCNN(X_{t-6b}^S)) \quad (3.2)$$

其中 $k = 1, 2, 3...$ 且 $b = 0, 1, 2, 3...$

由于根据不同的预测问题，模型损失函数有不同的定义。对于神经网络模型，其代价函数仍然很大程度上与其他的参数模型相近，如线性模型的损失函数。大多数情况下，参数模型定义了一个分布 $p(Y|X; \theta)$ ，并且简单地使用最大似然原理，即意味着是将训练数据和模型预测间的交叉熵作为代价函数。有时则使用更简单的方法，不是预测 Y 的完整的概率分布，而是在给定 X 的条件下 Y 的某种统计量，比如均方根误差和平均绝对误差。因此本文模型的整体损失函数可以表示为：

$$J(\theta) = \frac{1}{N} \frac{1}{T} E_{X, Y \sim \hat{p}_{data}} \|Y - f(X; \theta)\|^2 \quad (3.3)$$

其中， N 表示样本数， T 表示 LSTM 的时间步数， E 表示模型预测值与真实值之间差距的衡量标准， Y 表示真实值， X 表示输入值， f 表示模型， θ 表示模型需要学习的参数。

通常，设计网络模型时需考虑损失函数的梯度足够大且具有足够的可预测性，这样才能为训练出学习算法提供一个好的指引。用于实现最大似然估计的交叉熵函数可以处理离散型变量，因此可以用于台风预测中所定义的分类问题。而均方根误差和平均绝对误差虽然在使用基于梯度的优化方法时虽然往往不如交叉熵函数，但是通常可以适用于台风预测中所定义的回归问题。

3.3.1 海表变量 X^S 二维空间关系的特征学习

海面变量在空间上是一张二维网格图。为了了解格点海面变量之间的关系，模型使用含 2D 卷积核的 CNN，即 2DCNN。通常 CNN 包括输入层、卷积层、池化层、全连接层和激活层以及输出层。其中，卷积层是最为关键的一部分。在海面变量输入模型后，二维卷积层可以从中学习出海面变量的空间特征。如图 3.3 所示，它通过从前一个特征图的局部区域提取特征，然后传递给激活函数进行表示。

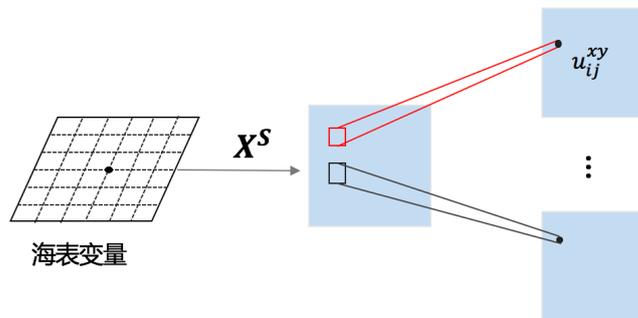


图 3.3 用于海表变量空间特征提取的二维卷积层示意图

第 i 层第 j 个特征图中 (x, y) 位置处的特征值记为 u_{ij}^{xy} ，具体公式为

$$u_{ij}^{xy} = f\left(\sum_{k=0}^{P_{i-1}} \sum_{p=0}^{Q_{i-1}} \sum_{q=0}^{Q_{i-1}} w_{ijk}^{pq} u_{(i-1)k}^{(x+p)(y+q)} + b_{ij}\right) \quad (3.4)$$

其中 f 是激活函数， k 是连接到当前特征图的 $i-1$ 层特征图的索引， w_{ijk}^{pq} 是连接 k 层特征图 (i, j) 位置处卷积核内的 (p, q) 处的值， P_i 和 Q_i 是内核的高度和宽度， b_{ij} 是这个特征图的偏置项。

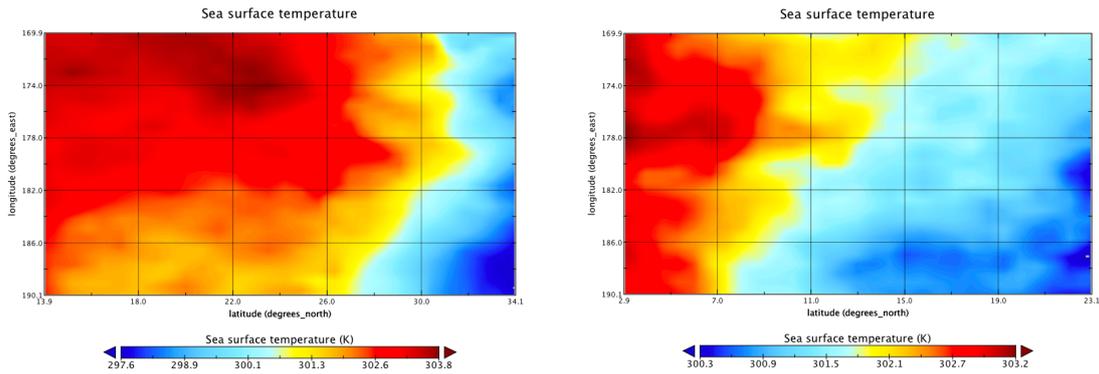


图 3.4 海表温度分布示意图

台风发生时与海洋的交互作用也异常的明显，海洋的反馈对于台风后续的发展尤为重要。比如 Ekman 输送，夹吸，内部震荡，拖拽及交换系数，海浪的影响，海表温度的改变等，由此可以定义出一些与台风相关的海表变量。其中一些相关因素，如海表温度可以通过现成的数据直接定义，另一些无法用现有数据定义的相关因素，如海浪的影响，可以通过新的量来重新定义。此处以海表温度 (SST) 为例，由于大多数与台风相关的海面变量场可能无明显的结构特征，如图 3.4 所示，因此需要设置多个卷积核多层网络来对潜在的复杂特征进行提取。每一层网络可以设置多个卷积核，来提取不同类型的特征。同样可以设置多层卷积层，来可以提取不同层的特征。然后将提取到的特征进行降维和全连接后作为海表变量的空间特征，如图 3.5 所示。以一个样本为例，该 2DCNN 网络结构的输入为 $M \times M$ 个网格位置的 N 个海表变量，通过第一个卷积层的多个卷积核进行卷积操作后，生成了多个特征图，然后对特征图进行降采样操作，得到了同样数量的子特征图，在实现中，利用多个卷积层来对海表变量之间的复杂的抽象特征进行提取，最后对所有子特征图进行全连接展开后，降维为海表特征变量，此即为 2DCNN 用于海表变量二维空间关系特征学习的过程。

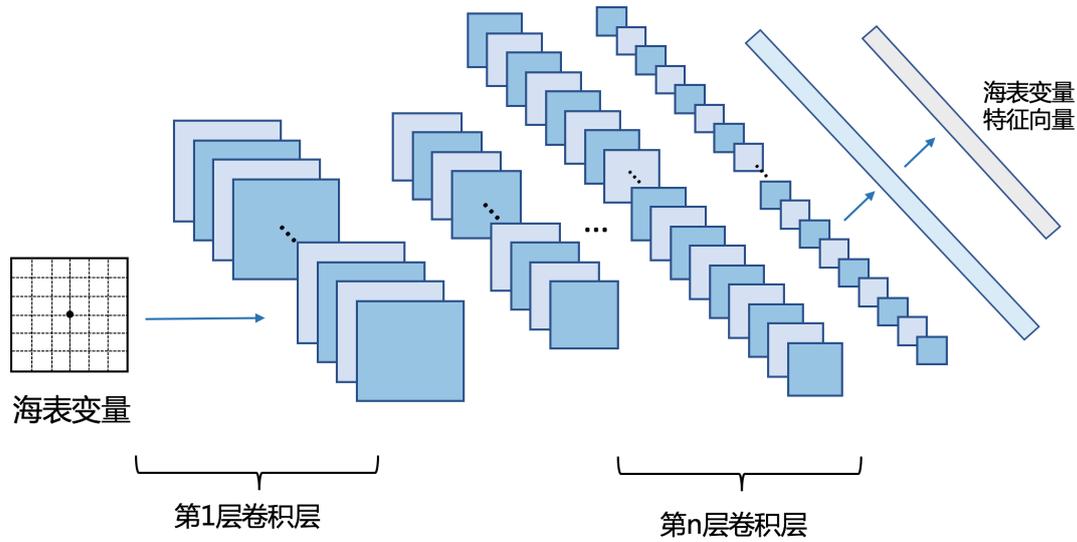


图 3.5 2DCNN 用于海表变量二维空间关系的特征学习示意图

3.3.2 大气变量 X^P 三维空间关系的特征学习

台风是一种强对流天气，其间大气存在强烈的垂直运动，因此大气变量不仅是三维空间变量，而且具有很强的相关性。常与台风相关的大气基本变量有风的 u 分量 (u)、风的 v 分量 (v)、温度 (t)、相对湿度 (rh) 和位势高度 (z)。同样可以通过已有数据来定义一些与台风的相关变量，如位涡，垂向速度，涡度 (VOR)，散度 (DIV)。而不能通过已有数据直接定义的有垂直风切变 ($VSHEAR$)、垂直不稳定度 ($THEDV$)、日生成潜力 (DGP)、低层环流 ($CIRC$)、最大潜在强度 (MPI)。

其中，最常用的涡度和散度可作如下表示 (式中 \vec{V} 为风矢量)：

$$\begin{aligned} VOR &= \nabla \times \vec{V} \\ DIV &= \nabla \cdot \vec{V} \end{aligned} \quad (3.5)$$

此外，可通过压力层数据定义一些相关量：

$$\begin{aligned} VSHEAR &= \sqrt{(\overline{u_{850}} - \overline{u_{200}})^2 + (\overline{v_{850}} - \overline{v_{200}})^2} \\ THEDV &= \sum_{p=850hpa}^{200hpa} wt(p)[\overline{\theta_E(850)} - \theta_E^*(p)], \theta_E \text{ 为 } 850hpa \text{ 位温} \\ DGP &= \overline{\xi_{900}} - \overline{\xi_{200}}, \xi_{900} \text{ 和 } \xi_{200} \text{ 为 } 900hPa \text{ 和 } 200hPa \text{ 压力层的相对涡度值} \\ CIRC &= \oint_P u \cdot dl \\ MPI &= 66.5 + 108.5e^{(0.8313(SST-30))} \end{aligned} \quad (3.6)$$

由于现在对台风的机制理解还不够到位，传统的手动特征变量可能与之相关，却不能够保证是真正有效的。受涡度、风切变等这些传统物理特征的启发，加之为了充分利用深度学习自动提取特征的能力，模型同样使用 CNN 对与台风相关的三维大气变量进行更复杂的特征学习。CNN 可以通过 3D 卷积核，即本文的 3DCNN，学习网格化大气变量各层各类型之间的相关性。如图 3.6 所示，3D 卷积核可以设置不同规格的卷积核尺寸，它将前一个特征图的相连几层的局部区域特征提取后映射到下一层子特征图的某一对应位置，再通过多个卷积核来生成多个系列的子特征图来学习整体与台风相关大气变量的空间特征。

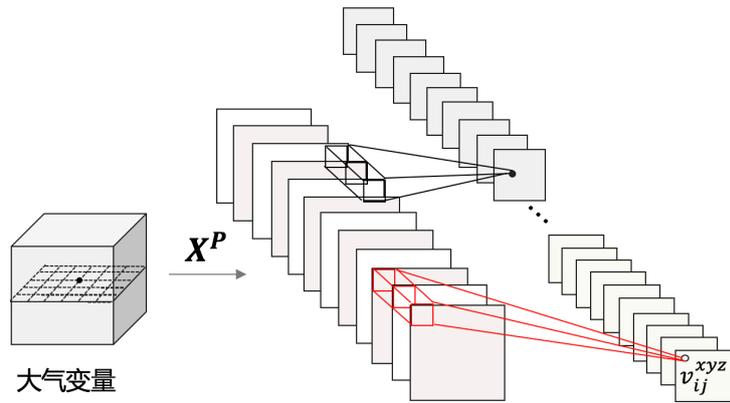


图 3.6 用于大气变量空间特征提取的三维卷积层示意图

第 i 层第 j 个特征图上位置 (x, y, z) 的值表示如下：

$$u_{ij}^{xyz} = f\left(\sum_{k=0}^{P_{i-1}} \sum_{p=0}^{Q_{i-1}} \sum_{q=0}^{R_{i-1}} \sum_{r=0}^{R_{i-1}} w_{ijk}^{pqr} u_{(i-1)k}^{(x+p)(y+q)(z+r)} + b_{ij}\right) \quad (3.7)$$

其中 R_i 是 3D 卷积核时间维度的值， w_{ijk}^{pqr} 连接 k 层特征图 (i, j) 位置处卷积核内的 (p, q, r) 处的值，例如偏差 b_{ij} 为偏置项。

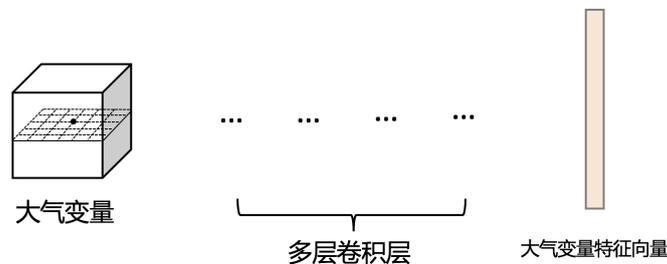


图 3.7 3DCNN 用于大气变量三维空间关系的特征学习示意图

台风中心周围的大气变量各网格点各层之间同样存在各种复杂抽象的特征关系，因此仍然需要设置多个卷积核和多个卷积层来提取复杂的特征以进行预测，如图 3.7 所示。以一个样本为例，该 3DCNN 网络结构的输入为 L 层 $M \times M$ 个网格位置的 R 个海表变量，同样通过第一个卷积层的多个卷积核进行卷积操作后，生成了多个特征图，然后对特征图进行降采样操作，得到了同样数量的子特征图，再利用多个卷积层来对大气变量之间的复杂的抽象特征进行提取，最后对所有子特征图进行全连接展开后，降维为大气变量特征向量，此即为 3DCNN 用于大气变量三维空间关系特征学习的过程。

3.3.3 台风变化过程中时序关系的特征学习

台风从热带风暴形成到台风形成再到之后的发展是一个时空变化的过程，以图 3.8 中 201601 号台风 Nerpatak 最佳路径为例，绿色点为热带气旋的生成时刻，蓝色点为热带风暴的生成时刻，橙色点为台风的形成时刻，黄色点为该台风的消亡时刻，其他所有点为台风发生发展过程中的其他时刻。由于在台风活动过程中，各个台风中心周围的大气海洋环境变量存在一定的时序关系，因此，可以通过多个时刻的大气海表变量空间特征来预测未来时刻的台风状态。

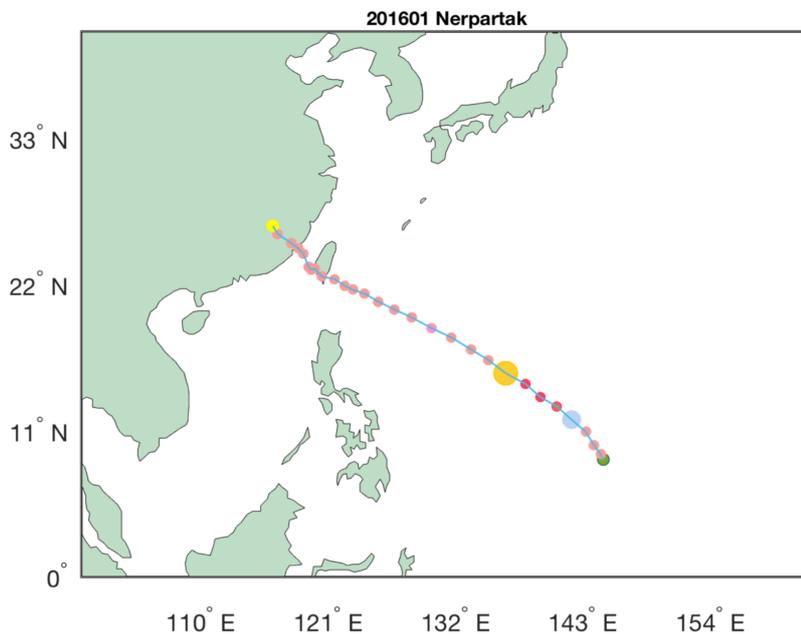


图 3.8 台风最佳路径示意图

LSTM 是递归神经网络 (RNN) 的改进模型。它可以将误差保持在一个恒定的水平，并增强模型的稳健性。当前时刻预测模型中 LSTM 部分的输入是 X_t ，由于台风预测中，间隔时间是 6h，因此上一个时刻的隐藏层状态是 h_{t-6} 和上一时刻

的记忆状态是 C_{t-6} 。LSTM 通过四个门遗忘并记住输入的信息。遗忘门决定应从上一个时刻的 C_{t-6} 中丢弃哪些信息，输入门决定应在 LSTM cell 的记忆中存储哪些新信息，输出门决定应从 C_t 中选择何种信息可以作为输出传递到下一个 LSTM 单元。四个门均由当前时刻的输入信息 X_t 和最后一个隐藏层状态 h_{t-6} 特征决定。由于台风活动过程中，当前的发展可能不仅仅只与前一个时刻的大气海洋变量相关，而是与前多个时刻的变量相关，因此，在预测时同样考虑使用多个时刻的大气海洋变量特征作为预测未来台风状态的输入。

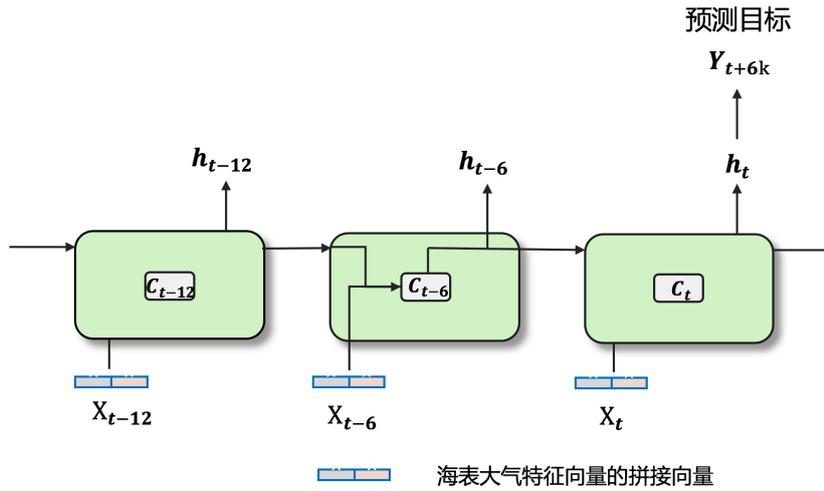


图 3.9 LSTM 用于台风变化过程中时序关系的特征学习示意图

可总结为：

$$\begin{aligned}
 Y_{t+6k} &= \sigma(W h_t + b) \\
 (h_t, C_t) &= LSTM([h_{t-6}, X_t], C_{t-6})
 \end{aligned}
 \tag{3.8}$$

其中 W 是权重矩阵， b 是偏置项， σ 表示 ‘sigmoid’ 函数。

如图 3.9，时空深度混合模型将 3DCNN 学习到的大气变量特征和 2DCNN 学习到的海表变量特征作为 LSTM 的输入向量，然后通过多个时间步学习时序关系来获得输出 Y_{t+6k} ，即该模型的预测对象。

3.4 实现方法

3.4.1 实现平台

深度学习模型可以用多种框架实现，包括 TensorFlow、Theano、Keras、Lasagne、Caffe、DSSTNE、Torch、MXNet、DL4J、Cognitive Toolkit 这十种框架。本文中的模型是使用 TensorFlow 作为后端的 Keras 搭建和实现的。硬件方面采用两块 NVIDIA

GTX 1080Ti GPU 对模型进行训练。为了构建该模型，需要使用 keras 库中的使用 Conv3D、Conv2D、LSTM、Dense、Merge 和 TimeDistributed 层。TimeDistributed 是一个层包装器，它将层应用于输入的每个时间切片上。该模型在每个时间步长的空间环境场中都需要使用卷积层，因此必须忽略时间的影响。Merge 层可以将多个张量组合成一个张量，这用于 3DCNN 和 2DCNN 的输出向量的合并。由于某些网格上存在缺失值，且不同变量之间存在较大差异。因此，需要适当的数据预处理。数据预处理包括填充丢失的值以及对数据进行规划化操作，这可以加速模型。首先需要确定模型的网络的层数、卷积核大小、卷积核数目、时间步大小以及神经元个数等参数。在完成模型的网络结构的设计后，需要对学习率(lr)、epoch、batchsize 等超参数进行调整，以得到最优的预测结果。调整超参数的方法包括手动调整、使用超参数优化算法、网格搜索、随机搜索等。本文模型使用手动调参、以及 sklearn 的 GridSearchCV 模块进行网格搜索来寻找最优超参数。

3.4.2 实现流程

训练时空深度混合模型时可采用全球台风官方预警机构认证的历史热带气旋最佳路径数据集以及再分析或其他卫星遥感空间变量数据集。首先对热带气旋最佳路径数据集的全部热带气旋路径记录进行分类，分为热带风暴数据集和台风数据集。然后根据热带风暴和台风数据集每条记录的中心位置来获取相对应的中心周围环境变量数据，以预测台风是否形成、台风路径、台风强度等三种情况下的输入数据集，即回顾时刻热带气旋中心周围的环境空间数据变量，并根据根据热带风暴和台风数据集每条记录的位置、强度构造输出数据。通过使用 TensorFlow 作为后端的 Keras 实现 3DCNN 处理大气空间数据变量，使用 2DCNN 处理海洋空间数据变量，使用 LSTM 处理热带气旋发生发展过程的时序关系，来搭建时空深度混合模型并评价模型的有效性。

具体步骤如下：

第一步，获取热带气旋最佳路径数据集，选择历史样本。匹配样本相对应的空间变量数据，构建训练拟合模型所需的数据集，方法是：

(1) 从官方网站下载近年来的历史热带气旋最佳路径数据集。选择热带气旋最佳路径数据集中按热带气旋生成海区进行划分的区域的热带气旋最佳路径数据集（元素为热带气旋最佳路径）。每一个区域热带气旋最佳路径数据集中包括不同年份的热带气旋最佳路径记录，每一年又包括不同条数的热带气旋最佳路径记录。单条热带气旋最佳路径记录含该热带气旋从形成到消亡的全部时间、中心经纬度位置以及强度信息。

(2) 令某一固定区域为区域 A ， A 的热带气旋最佳路径数据集为 Q ，逐一检

查 Q 中单条热带气旋最佳路径上的最大强度是否达到 64kt，将最大强度达到 64kt 的单条热带气旋记录放到台风数据集 Q_1 ，将最大强度达到 34kt 未达到 64kt 的热带气旋单条记录放到热带风暴数据集 Q_2 ，由此构造出台风数据集 Q_1 ，热带风暴数据集 Q_2 ，其余不满足条件的记录均舍去。若需训练其他区域的预测模型，重复上述步骤。

(3) 根据预测目标的不同，需要构建不同的台风模型，由此选择模型训练样本的时间、中心经纬度位置或强度，方法是：

(i) 若需要构建台风生成模型时，则选择台风是否形成预测模型的样本，选定预测时间步 k ，回归时间步长 b ，记录 Q_1 、 Q_2 中各条热带气旋最佳路径记录上的起报时刻、回顾时刻的时间、中心经纬度位置。

(ii) 若需要构建台风路径预测模型时，则选择预测台风路径预测模型的样本，记录 Q_1 中各条热带气旋最佳路径记录上的所有时间、中心经纬度位置。

(iii) 若需要构建台风强度预测模型时，则选择预测台风强度预测模型的训练样本，记录 Q_2 中各条热带气旋最佳路径记录上的所有时间、中心经纬度位置以及强度。

(4) 根据 (3) 中选择的模型样本的时间、中心经纬度位置，匹配相对应的大气和海洋变量组织成空间数据变量。方法是：

选择下载的变量空间范围为风暴或台风中心周围经纬度区域范围 $m^\circ \times m^\circ$ ，分辨率为 $u^\circ \times u^\circ$ ，因此区域内的格点数为 $\frac{m}{u} \times \frac{m}{u}$ 。大气空间变量所在层为 $l_1, l_2, \dots, l_i, \dots, l_R$ ， R 为大气变量层数。大气变量为 $p_1, p_2, \dots, p_i, \dots, p_I$ ， I 为大气变量数。海洋空间变量所在层为 $f_1, f_2, \dots, f_i, \dots, f_W$ ， W 为海洋变量层，海洋表变量为 $s_1, s_2, \dots, s_i, \dots, s_J$ ， J 为海洋变量数。样本匹配空间变量数据后的形式为 $X_{t-b}, \dots, X_{t-12}, \dots, X_{t-6}, X_t$ ， $X = [X^P, X^S]$ ，其中 X^P 为大气变量数据， X^S 为海表变量数据。

(5) 根据 (4) 中所获取的空间变量数据针对不同的预测模型构建不同的数据集，方法是：

(i) 在构建台风是否形成预测模型所需的数据集 R_1 时，将步骤 (4) 中 (i) 所得的数据作为输入数据集，然后把步骤 (2) 所得的台风数据集 Q_1 中的所有样本标记为 1，热带风暴数据集 Q_2 的所有样本标记为 0，将此标签作为输出数据集。

(ii) 在构建台风路径预测模型所需的数据集 R_2 时，获取步骤 (4) 中 (i) 所得的数据后，选定预测时间步 k ，回归时间步长 b ，对 Q_1 中所有样本的热带最佳路径记录和对应的再分析数据进行错位处理，使得单个样本拆解为多个训练样本。其中错位处理后的空间变量数据作为输入数据集，错位处理后的中心经纬度值作为输出数据集。

(iii) 在构建台风强度预测模型所需的数据集 R_3 时, 同样获取步骤 (4) 中 (i) 所得的数据后, 选定预测时间步 k , 回归时间步长 b , 对 Q_1 中所有样本的热带最佳路径记录和对应的空间变量数据进行错位处理, 使得单个样本拆解为多个训练样本。其中错位处理后的再分析数据作为输入数据集, 错位处理后的强度值作为输出数据集。

(6) 此步骤对 (5) 中的 R_1, R_2, R_3 数据集进行预处理。首先将 R_1, R_2, R_3 数据集集中的 Nan 和 inf 值填 0, 然后对整个数据集进行归一化, 方法如下:

$$R = \frac{R_{min}}{R_{max} - R_{min}} \quad (3.9)$$

其中 R 可以表示 R_1, R_2, R_3 中任何一个数据集, R_{max} 表示该数据集的最大值, R_{min} 表示该数据集的最小值。

第二步, 构建时空深度混合模型, 方法是:

(1) 搭建深度学习环境。在含两块 NAVID 1080Ti 型号 GPU 的小型服务器上安装基于 Python 的深度学习库 Keras, 并使用 Goole 深度学习框架 TensorFlow 作为后端。

(2) 基于 Keras 构建时空深度混合模型, 方法是:

(i) 构建模型中学习大气变量空间特征的 3DCNN 模块。3DCNN 模块中网络设置选用 keras 的输入层, 三维卷积层 (Conv3D 层), 批正则化层 (BatchNormalization 层), 平铺层 (flatten 层), 全连接层 (Dense 层) 以及输出层。设置 m_1 层 Conv3D 层, m_1 层 BatchNormalization 层, 1 层 flatten 层, 1 层 Dense 层。每层 Conv3D 含 n_1 个的 3D 卷积核 (滤波器), 卷积核的尺寸为 $i_1 \times j_1 \times k_1$, 步长为 $a_1 \times b_1 \times c_1$, 每层 Conv3D 后添加一层 BatchNormalization 层, 且选择的激活函数为 'relu' 函数, 具体形式为 $relu(x) = \max(0, x)$, Dense 层的神经元数目设置为 d_1 个, 其中参数 $m_1, n_1, i_1, j_1, k_1, a_1, b_1, c_1, d_1$ 均为正整数。

(ii) 构建模型中学习海洋变量空间特征的 2DCNN 模块。3DCNN 模块中网络设置选用 Keras 的输入层, 二维卷积层 (Conv2D 层), 批正则化 (BatchNormalization 层), 平铺层 (flatten 层) 和全连接层 (Dense 层) 以及输出层。设置 m_2 层 Conv2D 层, m_2 层 BatchNormalization 层, 1 层 flatten 层和 1 层 Dense 层。每层 Conv2D 含 n_2 个的 2D 卷积核, 卷积核的尺寸为 $i_2 \times j_2$, 步长为 $a_2 \times b_2$, 每层 Conv2D 后加一层 BatchNormalization 层, 且其激活函数为 'relu' 函数。输出 Dense 层的神经元数目为 d_2 个, 其中 $m_2, n_2, i_2, j_2, a_2, b_2, d_2$ 均为正整数。

(iii) 构建模型中学习时序特征的 LSTM 模块。LSTM 模块中网络设置选用 Keras 的 LSTM 层。设置 LSTM 的神经元个数为 d_3 个, 激活函数选用 'relu' 函数。

(3) 将 3DCNN 模块, 2DCNN 模块, LSTM 模块用 Keras 的 Sequential 顺序模

型连接，并添加一层输出层，设置输出层神经元个数为 1，激活函数选用 *sigmoid* 函数，其具体形式为 $\sigma(x) = \frac{1}{1+e^{(-x)}}$ ，连接后的模型即为本文的时空深度混合模型。

其中，3DCNN 模块的输入为 R_1, R_2, R_3 数据集中的大气再分析数据 X^P ，输出为大气空间特征向量 O^P ，其向量长度为 Dense 层的神经元数目 d_1 。2DCNN 模块的输入为 R_1, R_2, R_3 数据集中的海表再分析数据 X^S ，输出为大气空间特征向量 O^S ，其向量长度为 Dense 层的神经元数目 d_2 。LSTM 模块的输入为 3DCNN 模块和 2DCNN 模块输出向量的合并向量 $O = [O^P, O^S]$ ，该向量是由 Keras 的合并层 (Merge 层) 合并得来，其长度为 $d_1 + d_2$ 。

第三步，训练拟合台风是否形成预测模型、台风路径预测模型、台风强度预测模型，方法是：

(1) 将 R_1, R_2, R_3 按一定比例划分为训练集和测试集。训练集用来训练拟合模型，测试集用于测试模型的有效性。

(2) 训练拟合台风是否形成预测模型，方法如下：

(i) 将 R_1 的训练集中的 $X_{t-b}, \dots, X_{t-12}, X_{t-6}, X_t$ 作为输入数据， Y_{t+6k} 作为台风是否形成的输出标签数据，当 $Y_{t+6k} = 1$ 时代表台风形成， $Y_{t+6k} = 0$ 时代表台风不形成。

(ii) 选用交叉熵函数作为模型的损失函数：

$$H = -\frac{1}{N_1} \sum_{i=1}^{N_1} Y_i \times \log(\hat{Y}_i) \quad (3.10)$$

其中， N_1 为 R_1 训练集中热带风暴的样本数， Y_i 为真实的标签数据概率分布，亦为 Y_{t+6k} ， \hat{Y}_i 为模型预测的标签数据概率分布。

(iii) 选用准确率 (ACC) 和 AUC 值作为模型好坏的评价指标：

表 3.1 台风预测分类混淆矩阵

实际观测	预测	
	台风	热带风暴
台风	TP	FN
热带风暴	FP	TN

ACC 是被正确预测的样本占全部样本的比例，可以定义为：

$$ACC = \frac{TP + TN}{TP + FN + FP + TN} \quad (3.11)$$

Auc 是 ROC 曲线下的面积，ROC 曲线的横坐标为假阳性率 ($FPR = \frac{FP}{FP+TN}$)，纵坐标为真阳性率 ($TPR = \frac{TP}{TP+FN}$)。Auc 值的定义如下：

$$AUC = \int ROC \quad (3.12)$$

(iii) 调用 Keras 的模型拟合函数，调整模型的学习率、迭代次数、优化算法，网络层数以及神经元个数。观察训练过程中训练集的最小损失函数值，并将多次训练完成后评价指标输出值最大且趋于稳定的模型作为台风是否形成预测模型，并记录其测试集能达到的最高预测效果。

(2) 训练拟合台风强度或路径预测模型，方法如下：

(i) 将 R_1 的训练集中的 $X_{t-b}, \dots, X_{t-12}, X_{t-6}, X_t$ 作为输入数据， Y_{t+6k} 作为台风是否形成的输出标签数据，当预测强度时 $Y_{t+6k} = \text{强度值}$ ，当预测路径时 $Y_{t+6k} = \text{经纬度位置值}$ 。

(ii) 选用均方根误差函数作为模型的损失函数：

$$J = \sqrt{\frac{1}{N} \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T (Y_{t,i} - \hat{Y}_{t,i})^2} \quad (3.13)$$

$Y_{t,i}$ 是第 i 个样本在 $t(t \in T)$ 时刻的真实观测值， $\hat{Y}_{t,i}$ 是第 i 个样本在 t 时刻的预测值。

(iii) 选用绝对误差函数作为模型作为评价指标：

$$MAE = \frac{1}{N} \sum_{i=1}^N |(Y_i - \hat{Y}_i)| \quad (3.14)$$

其中 N 是样本的数量， \hat{Y}_i 表示预测的强度值或经纬度位置值， Y_i 表示真实的强度值或经纬度位置值。

(iii) 调用 Keras 的模型拟合函数，调整模型的学习率、迭代次数、优化算法，网络层数以及神经元个数。选择将多次训练完成后损失函数值最小，评价指标输出值最小且趋于稳定的模型作为台风强度预测模型或路径预测模型，并记录其测试集能达到的最小预测误差。

第四章 基于时空深度混合预测模型的台风形成及强度预测

4.1 台风数据集和气象海洋再分析数据集的获取

本文使用世界气象组织 (World Meteorological Organization, WMO) 版本的国际气候管理最佳轨道档案 (International Best Track Archive for Climate Stewardship, IBTrACS) 全球热带气旋最佳轨道数据集和 ERA-Interim 再分析数据集。全球热带气旋最佳轨迹数据集按生成海域或者年份进行划分。本文选用按生成海域进行划分的热带气旋最佳路径数据集。热带气旋最佳路径数据集记录是从热带风暴的形成开始, 记录的时间间隔为 6 小时。随着热带风暴的发展, 一些可能发展为台风, 另一些在中心附近达到最大风速后开始衰减随后消亡。因此根据热带气旋最佳路径数据集中心附近的最大强度, 本文将热带气旋分为两种类型: 一种是会发展成台风的热带风暴, 另一种是不会发展成台风的热带风暴。至于 ERA-Interim 再分析数据集, 自 1979 年以来, ECMWF 的全球再分析数据一直可以免费获取。再分析数据层的类型包括压力层、表面层、模式层等, 每个层都包含多分辨率和多类型变量。由于热带风暴中心周围大气和海面的剧烈变化, 本文选择高分辨率和多层再分析大气变量数据以及海表数据。

一般来说, 台风预测需要考虑不同预测时长。模型在保证一定准确率的前提下, 能够预测的时长越长, 意义就越明显。然而, 台风样本相比于其他深度学习数据样本来说比较稀少。为了保证模型具有可以操作的样本, 所有实验选的预测时长选择 24 小时。因此, 对于台风而言, 预测的当前时刻为样本数据中台风形成时刻前 24 小时。对于热带风暴, 当前时刻是在中心附近达到最大强度的时刻。为了进行预测实验, 本文不仅考虑了当前时刻的空间变量, 还考虑了回顾时刻的空间变量, 这里的回顾时刻选择的是台风形成前的 6h、12h、18h、24h。样本来自三个典型的海域, 分别是西太平洋 (WP)、东太平洋 (EP) 和北大西洋 (NA)。台风和热带风暴样本数量列于表 4.1。

关于 ERA-Interim 再分析数据集大气海洋变量的选择。因为台风或热带风暴的横向规模可以达到约 1000 公里, 影响范围选择中心周边地区的 $4^{\circ} \times 4^{\circ} \sim 20^{\circ} \times 20^{\circ}$ 区域, 该区域可覆盖整个台风区。此外, 台风是垂直方向的空间结构, 它可以分为

表 4.1 台风和热带风暴样本数量

	西太平洋 (WP)	东太平洋 (EP)	北大西洋 (NA)
台风	239	219	181
热带风暴	211	220	186

三部分: 流入层、中间层和流出层。从海面到 3 千米的高度为流入层, 3~8 千米的高度为中间层, 从 8 千米到台风顶的高度为流出层。因此根据台风可能涉及的高度选择的大气压力层为 1000/975/925/850/800/700/600/500/400/300/200/100hpa。此外, 海面对台风的活动有重要影响, 选择与台风相关的海平面变量。关于大气变量的类型, 本文选择与台风相关的基本变量, 如风的 u 分量 (u)、风的 v 分量 (v)、温度 (t)、相对湿度 (rh) 和位势高度 (z)。对于海表变量, 本文只选择最常用的海表温度 (sst)。而以往研究中发现的其他有利于热带风暴发展的物理变量, 如垂直风切变 ($Vshear$)、垂直不稳定度 ($THDEV$) 及最大潜在势能 (MPI) 等, 也均是根据基本变量计算得出。本文模型希望通过使用基本的大气和海面变量, 在保证较良好预测效果的同时简化其预测的难度。

具体数据获取步骤如下:

(1) 从美国环境信息中心 (National Centers for environmental information, NCDC) 网站 <https://www.ncdc.noaa.gov/ibtracs/> 获取经 WMO 认证的 IBTrACs 热带气旋最佳路径数据集。

(2) 选择 IBTrACs 提供的热带气旋最佳路径数据集中按热带气旋生成海区进行划分的区域的热带气旋最佳路径数据集 (元素为热带气旋最佳路径)。每一个区域热带气旋最佳路径数据集中包括不同年份的热带气旋最佳路径记录, 每一年又包括不同条数的热带气旋最佳路径记录。单条热带气旋最佳路径记录包括该热带气旋从生成到消亡的全部时间、中心经纬度位置以及强度信息等。

(3) 令某一固定区域为区域 A , A 的热带气旋最佳路径数据集为 Q , 逐一检查 Q 中单条热带气旋最佳路径上的最大强度是否达到 $64kt$, 将最大强度达到 $64kt$ 的单条热带气旋记录放到台风数据集 Q_1 , 将最大强度达到 $34kt$ 未达到 $64kt$ 的热带气旋单条记录放到热带风暴数据集 Q_2 , 由此构造出台风数据集 Q_1 , 热带风暴数据集 Q_2 , 这里 $Q_1, Q_2 \in Q$ 。

(4) 根据预测目标的不同, 需要构建不同的台风预测模型, 由此来选择 Q_1, Q_2 中相应模型训练样本的时间、中心经纬度位置或强度。

(5) 根据选定的模型样本时间、中心经纬度位置, 使用欧洲中期天气预报中心 (European Centers for Medium-Range Weather Forecasts, ECMWF) 的 python 应用程序接口从 <http://apps.ecmwf.int/datasets/data/interim-full-daily> 下载相应的大气海洋变量组织成实验数据集。变量下载的空间范围为中心周围经纬度区域范围 $20^\circ \times 20^\circ$, 分辨率为 $0.125^\circ \times 0.125^\circ$ 。大气变量为 u, v, rh, t, h , 大气变量所在层为 1000/975/925/850/800/700/600/500/400/300/200/100hpa, 海表变量为 sst 。热带风暴或台风样本匹配空间数据后形成的数据集形式为 $X_{t-b}, \dots, X_{t-12}, X_{t-6}, X_t, Y_{t+6k}$, 其中 X^P 为大气空间数据, X^S 为海表空间数据。

4.2 基于台风时空深度混合预测模型的台风形成预测

4.2.1 实验数据集的构造与预处理

4.2.1.1 数据集的构造

台风形成预测问题可以进一步被定义为：

给定时空变量序列 $X_{t-6b}, b = 0, 1, 2, 3, \dots$ ，预测台风是 ($Y_{t+6k} = 1$) 否 ($Y_{t+6k} = 0$) 形成。

这是机器学习中典型的分类问题。在定义好问题后，需要根据 4.1 节中获取到的数据进行模型输入和输出数据集的组织。对于模型的 3DCNN 部分，输入数据集中所有样本 X^P 的形式是 $N \times M \times M \times L \times R$ 。对于模型的 2DCNN 部分，输入数据集中所有样本 X^S 的形式是 $N \times M \times M \times O$ 。对于模型的 LSTM 部分，所有样本的输入形式是 $N \times T \times K$ 。其中 N 是样本数， $M \times M$ 是中心周围经纬度范围的网格数， L 是大气压力层层数， R 是大气变量数， O 是海表变量数， T 是 LSTM 的时间步， K 是 3DCNN 与 2DCNN 输出特征向量合并后的维度。本实验中，西太平洋区域 $N = 450$ ，其中台风样本数 $N_1 = 239$ ，热带风暴样本数 $N_2 = 211$ 。东太平洋区域 $N = 439$ ，其中 $N_1 = 219$ ， $N_2 = 220$ 。北大西洋区域 $N = 367$ ，其中 $N_1 = 181$ ， $N_2 = 186$ 。固定参数 $L = 12$ ， $R = 5$ ， $O = 1$ 不变， M 的选取范围为 $33 \sim 161$ ， T 的选取范围为 $1 \sim 5$ 。 $K = 200$ 。输出数据集中台风样本 Y_{t+6k} 标为 1，热带风暴样本 Y_{t+6k} 标为 0。本实验中 70% 的样本进行训练，30% 的样本进行测试。

4.2.1.2 数据集的预处理

所有的缺失值、NaN 或者 inf 值均填 0。此外由于数据的数据差异明显，因此需要进行归一化。归一化操作一方面可以提升模型的收敛速度，一方面可以提升模型的精度，还能够防止模型梯度爆炸。

整个数据集采用的归一化函数如下：

$$X = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4.1)$$

其中， X 代表数据集中的所有元素， X_{max} 为数据集的最大值， X_{min} 为数据集的最小值。

4.2.2 评价指标与损失函数

4.2.2.1 评价指标

为了评估模型的预测性能，该实验采用准确率 (Acc) 和 ROC 曲线下的面积 (Auc) 进行表示。Acc 是被正确预测的样本占全部样本的比例，根据表 3.1 可以定义为：

$$Acc = \frac{TP + TN}{TP + FN + FP + TN} \quad (4.2)$$

Auc 是 ROC 曲线下的面积，可以很好地描述模型的整体性能。ROC 曲线的横坐标为假阳性率 ($FPR = \frac{FP}{FP+TN}$)，纵坐标为真阳性率 ($TPR = \frac{TP}{TP+FN}$)。Auc 值的定义如下：

$$Auc = \int ROC \quad (4.3)$$

Auc 值是介于 0 ~ 1 之间的概率值。Auc 值越大，分类模型更有可能将正样本放在负样本前面，使这些样本能够进行更好的分类。

4.2.2.2 损失函数

对于台风形成预测问题，由于是机器学习中的分类问题，因此选择交叉熵函数作为损失函数。它可以测量预测值和实际值之间的差异，并给出两个概率分布之间的距离。定义损失函数的目的是为了优化模型，以达到更好的预测效果。

$$J = -\frac{1}{N} \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T ((1 - y_{t,i}) \times \log(1 - \hat{y}_{t,i}) + y_{t,i} \times \log(\hat{y}_{t,i})) \quad (4.4)$$

其中 $y_{t,i}$ 是第 i 个样本在 $t(t \in T)$ 时刻的真实观测值概率分布， $\hat{y}_{t,i}$ 是第 i 个样本在 t 时刻的预测值概率分布。而 $\hat{y}_{t,i} = \sigma(Wh_t)$, $h_t = \sigma[h_{t-1}, x_t]$, x_t 作为 LSTM 的输入，是 3DCNN 以及 2DCNN 输出特征的拼接，且 $x_t = [O_{3D}, O_{2D}]$, O_{3D} 是 X^P 输入 3DCNN 后的输出向量， O_{2D} 是 X^S 输入 3DCNN 后的输出向量。

4.2.3 网络设置

在该实验中，模型的 3DCNN 部分包括一个输入层，五个 Conv3D 层，一个 Flatten 层和一个 Dense 层。网络设置时第一层 Conv3D 层中卷积核大小是 $5 \times 5 \times 3$ ，步长为 $2 \times 2 \times 1$ ，其数量是 32。第二层 Conv3D 层中卷积核大小是 $5 \times 5 \times 1$ ，步长为 $2 \times 2 \times 1$ ，其数量是 64。第三层 Conv3D 层中卷积核大小是 $5 \times 1 \times 1$ ，步长为 $2 \times 1 \times 1$ ，其数量是 128。Dense 层含 100 个神经元，各层神经元的激活函数为“relu”函数，该函数能使模型的收敛速度保持在稳定状态。

模型的 2DCNN 部分包括一个输入层，三个 Conv2D 层，一个 Flatten 层和一个 Dense 层。网络设置时第一层 Conv2D 层中卷积核大小是 5×5 ，步长为 2×2 ，其数量是 32。第二层 Conv2D 层中卷积核大小是 5×1 ，步长为 2×1 ，其数量是 64。第三层 Conv2D 层中卷积核大小是 5×1 ，步长为 2×1 ，其数量是 128。Dense 层含 100 个神经元，各层神经元的激活函数为“relu”函数。

模型的 LSTM 部分，由于实验所选取的序列较短，因此只选用单层 LSTM，隐层神经元数为 100 个，激活函数为“relu”。输出层的激活函数为“sigmoid”函数。

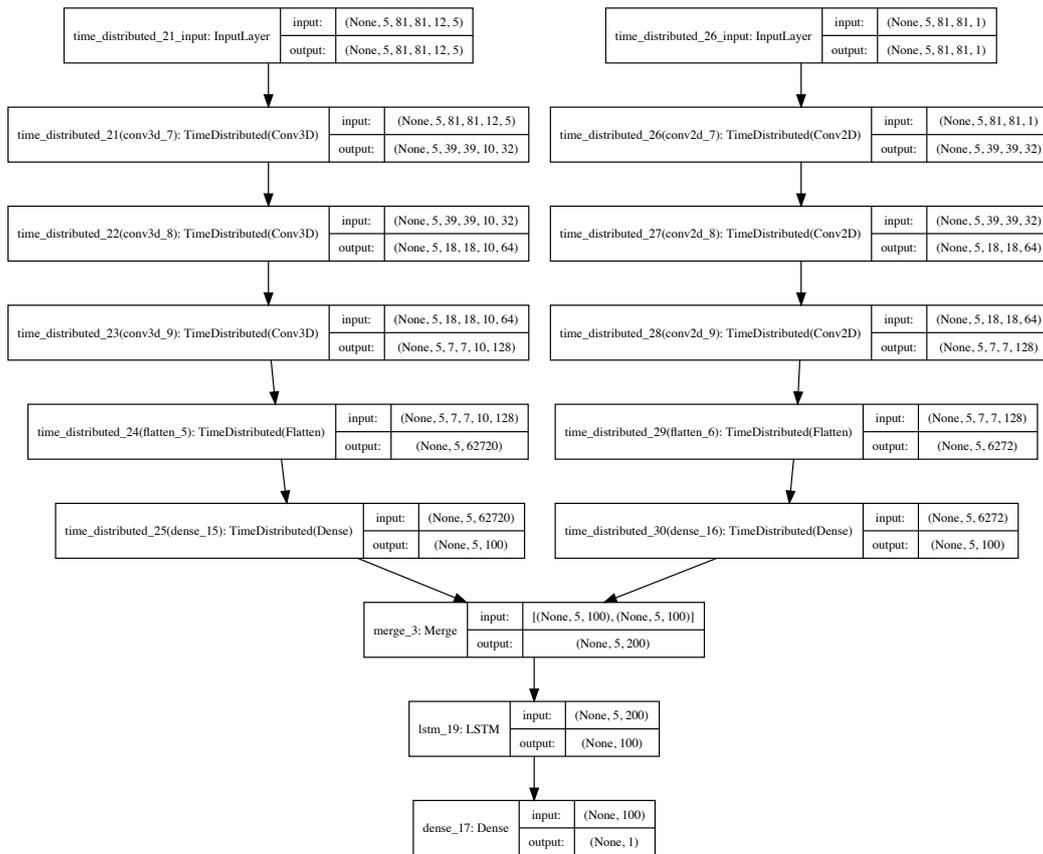


图 4.1 模型网络设置及流程图示例

以样本输入规格为 $5 \times 81 \times 81 \times 12 \times 5$ 为例，其数字按顺序分别代表 LSTM 的时间步，经度格点数，纬度格点数，压力层数，压力变量数。模型的输入输出流如图 4.1 所示，上一层的输出将作为下一层的输入。3DCNN 部分的输入为 (None, 5, 81, 81, 12, 5)，None 为未知的样本数。以一个样本为例，首先忽略时间维度 (LSTM 时间步)，即对某一个时刻的多个台风相关大气变量进行空间卷积操作。由于第一层 Conv3D 层的 3D 卷积核的尺寸为 $5 \times 5 \times 3$ ，步长为 $2 \times 2 \times 1$ ，因

此卷积后的每个特征图大小为 $39 \times 39 \times 10$ ，此外卷积核的个数为 32，所以输出含 32 个大小为 $39 \times 39 \times 10$ 的特征图。而且样本数和 LSTM 时间步不变，故第一层 Conv3D 层的输出为 (None, 5, 39, 39, 10, 32)。同理，经过第二层后，每个特征图的大小为 $18 \times 18 \times 10$ ，由于有卷积层含 64 个卷积核，输出因此为 (None, 5, 18, 18, 10, 64)。经过第三层后，每个特征图的大小为 $7 \times 7 \times 10$ ，由于卷积层含 129 个卷积核，输出因此为 (None, 5, 7, 7, 10, 128)。将经过三层卷积后的特征图集进行全连接展开后的输出为 (None, 5, 62720)，通过全连接层降维后的特征向量维度为 100。因此大气变量进入 3DCNN 卷积后得到的大气空间特征向量的长度为 100，多个样本多个 LSTM 时间步的输出为 (None, 5, 100)。2DCNN 部分同样对多个时间步的台风相关海表变量进行空间卷积操作，输入为 (None, 5, 81, 81, 1)，输出为 (None, 5, 100)。将两者输出的特征向量拼接为 200 维的向量后输入至 LSTM 中。LSTM 学习后的输出向量为 100 维，再通过全连接输出层即可得到模型的预测结果。模型所需要学习的参数总共为 7545149 个。

4.2.4 对比方法

为了证明本文所提的时空深度混合模型的有效性，将实验结果与以往热带气旋研究中三种传统机器学习算法进行对比。它们分别是逻辑回归 (Logistic Regression, LR)^[65]，线性判别分析 (Linear Discriminative Analysis, LDA)^[45] 和决策树 (Decision Tree, DT)^[41]。LR 是一种广义的线性回归分析模型，适用于分类问题。其优点是抗干扰能力强，并且可以使用 L1 和 L2 正则化的方法来避免过拟合。但缺点是容易欠拟合，从而准确度不高。LDA 既可以用来降维，又可以用来分类。其优点是在降维过程中可以使用先验知识。而缺点是不适合对非高斯分布样本进行降维，且样本分类信息依赖方差时降维效果会不好，同时可能过拟合。决策树作为一种分类模型，其优势是可读性好，理解和解释起来简单，利于人工分析。缺点是没有足够的依据来创建复杂的树，而且通常也会过拟合。这些算法以往常用于热带气旋的生成问题，但暂未发现用于台风的形成预测问题。且以往为了匹配这些算法的输入形式，普遍将大气变量和海洋变量作为独立的特征，从而无法考虑变量之间的时空关系。本文实现这三种算法的过程，首先是将所有样本展开为算法能处理的形式，并调用 Sklearn 的机器学习算法包进行预测分析。由于热带气旋生成问题的样本极不均匀，这些方法在以往的一些研究中可能获得很好的效果，但是本文样本均衡，因此结果可能出现较大偏差。此外，为了增强对比性，设置了只考虑时序关系而不考虑空间关系的对比实验，即与最原始 LSTM 模型进行了比较。LSTM 能学习变量的时序关系，但没有考虑变量的空间关系。在实现过程中考虑数据集的构建时，保持 LSTM 时间步不变，将空间多维向量展

开，然后同样基于 Tensorflow 为后端的 Keras 来进行预测分析。除此之外，本文也将与已有的 LSTM 改进方法 ConvLSTM^[74] 以及 CNN 和 LSTM 结合的其他方法，如 2DCNN+LSTM, 3DCNN+LSTM 进行对比，实验结果如表 4.2 所示。其中，ConvLSTM 是将 LSTM 中的全连接层替换为卷积层，与原始的 CNN 保持同样的输入形式。与 2DCNN+LSTM 以及 3DCNN+LSTM 对比是为了保证混合 2DCNN 和 3DCNN 的有效性，其输入形式分别对应 2DCNN 的输入形式和 3DCNN 的输入形式。

4.2.5 结果分析

在西太平洋区域，使用传统机器学习方法进行预测时，准确率 Acc 最高为 0.741，Auc 值最高为 0.774。在其他深度学习的模型中，最高准确率 Acc 为 0.812，最高 Auc 值最好的为 0.897。然而，本文的台风时空混合预测模型在 24 小时台风形成预测实验中可以达到最高准确率和最高 Auc 值，分别为 0.852 和 0.922。为了保证模型在不同区域的鲁棒性，还在东太平洋和北大西洋区域进行了实验。在东太平洋区域，采用 LDA 模型获得的准确率 Acc 最高为 0.735，Auc 值最高为 0.763，其他深度学习模型以 3DCNN+LSTM 模型获得的准确率 Acc 最高为 0.763，Auc 值最高为 0.847。而本文的台风时空混合预测模型可以达到最高准确率 0.780 和最高 Auc 值 0.858。在北大西洋区域，台风时空混合预测模型计算出最高准确率和最高 Auc 值分别为 0.759 和 0.794。根据实验结果还可以发现东太平洋和北大西洋的预测结果比西太平洋要低。一方面，这可能是样品数量的原因，另一方面，这可能是由于不同地区环境具有差异性的原因。

表 4.2 台风形成预测实验结果对比

	西太平洋 (WP)		东太平洋 (EP)		北大西洋 (NA)	
	Acc	Auc	Acc	Auc	Acc	Auc
LR	0.718	0.720	0.727	0.728	0.730	0.730
DT	0.637	0.637	0.652	0.652	0.613	0.610
LDA	0.741	0.741	0.735	0.763	0.613	0.612
LSTM	0.667	0.774	0.667	0.694	0.595	0.610
ConvLSTM	0.756	0.799	0.742	0.807	0.685	0.759
2DCNN+LSTM	0.766	0.833	0.734	0.846	0.721	0.764
3DCNN+LSTM	0.812	0.897	0.763	0.847	0.729	0.773
时空深度混合预测模型	0.852	0.922	0.780	0.858	0.759	0.794

4.2.6 参数分析

4.2.6.1 学习率

梯度下降算法是训练神经网络最常用的优化方法。参数的梯度可以通过偏导数来计算。学习率 (lr) 是控制参数更新速度的重要超参数。实验发现，模型刚开始在使用自适应优化算法如“adam”算法时会处于非收敛状态，因此需要手动调节参数。首先，固定模型框架后调整 lr 等超参数，从 0.01 降到 0.0001，下降速率是 10。实验发现，模型训练和测试损失在学习率的量级为 10^{-4} 时会稳步下降。随后在不同的数据集上进行实验验证，结果如图 4.2 所示。左子图为三个区域下 lr 随 Acc 的变化，右图为三个区域下 lr 随 Auc 的变化。根据 Acc，在西太平洋 (WP) 区域，lr 最佳选择为 0.0004。在东太平洋 (EP) 区域，最佳 lr 选择为 0.0001。北大西洋 (NA) 区域 lr 最佳选择是 0.0002。根据 Auc，在西太平洋 (WP) 区域，lr 最佳选择为 0.0003。在东太平洋 (EP) 区域，最佳 lr 选择为 0.0001。北大西洋 (NA) 区域 lr 最佳选择是 0.0003。因此,lr 的最佳选择为 10^{-4} 量级。

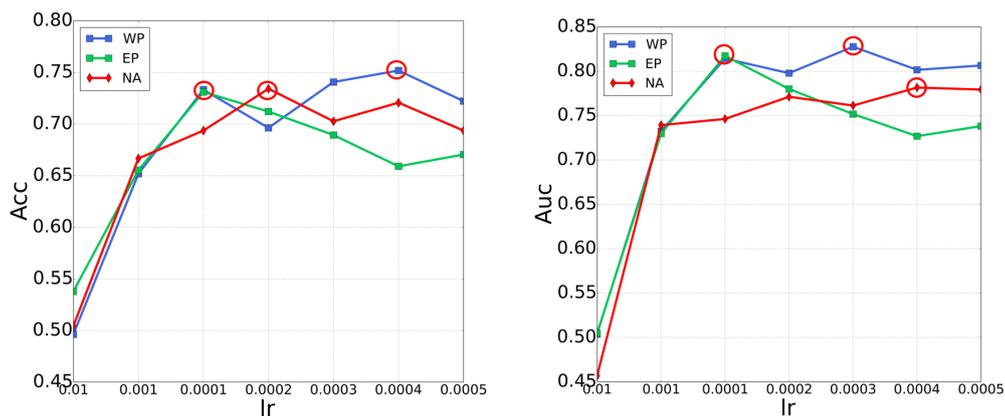


图 4.2 Acc 以及 Auc 随学习率 (lr) 的变化

4.2.6.2 epoch

在模型训练之初，为了充分训练预测模型，选择使用较大的 epoch，如 100 左右。但是实验过程中发现，epoch 过大会出现过拟合的情况，训练损失一直在减小，而测试损失在 25 ~ 35 个 epoch 之前下降，但在之后上升。这不是训练模型的目的，因为这会使得模型的泛化性能不高，即在训练完后该模型无太大的实际效用。因此为了确定台风形成预测模型的最佳 epoch，本文在不同区域运行了 25 个或 30 个 epoch 后结束模型的训练拟合。如图 4.3 所示，可以发现在西太平洋 (WP) 区域，25 个 epoch 后 AUC 处于稳定状态，在东太平洋 (EP) 区域，15 个 epoch 后 AUC 开始下降，北大西洋 (NA) 区域 AUC 在 30 个 epoch 后开始处于稳定状态。因

此, 在实验过程中, 西太平洋 (WP) 区域模型的训练选择在 25 个 epoch 后结束, 东太平洋 (EP) 区域模型的训练选择在 15 个 epoch 后结束, 北大西洋 (NA) 区域模型的训练选择在 30 个 epoch 后结束, 可以发现在 15 ~ 30 个 epoch 左右模型基本能到训练到想要的效果。

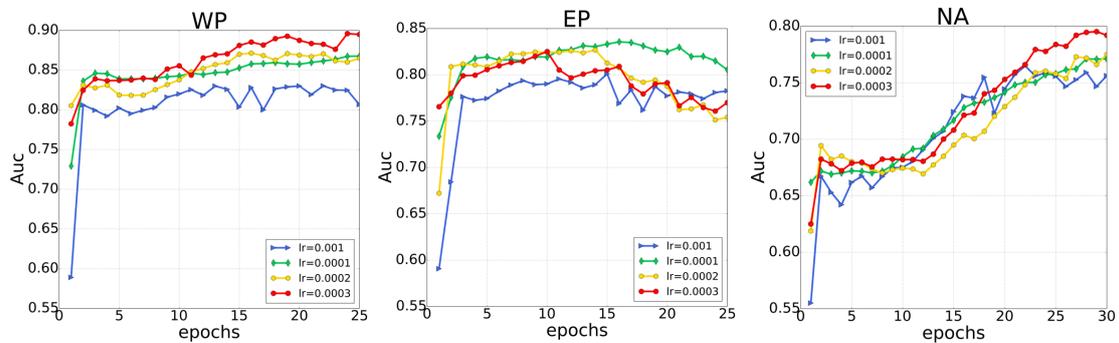


图 4.3 Auc 值在不同的学习率下随 epoch 的变化

4.2.6.3 区域范围

除了模型的超参数在模型训练拟合中起着至关重要的作用外, 输入数据集的规格对预测也有着至关重要的影响。为了确定最优的影响范围, 本文对热带风暴中心周边环境变量经纬度区域范围进行了实验, 变化范围为 $4^\circ \times 4^\circ \sim 20^\circ \times 20^\circ$, 因此网格点数为 $33 \times 33 \sim 161 \times 161$ 。图 4.4 显示了 Auc 及 Acc 值随中心周边环境变量经纬度区域范围的变化。结论是最佳区域范围是在 $9^\circ \times 9^\circ$ 以及 $13^\circ \times 13^\circ$ 之间, 因此并不是区域越大, 输入的数据越多, 模型的预测性能越好。西太平洋 (WP) 区域, 根据 Auc 值来看, 最佳区域范围是在 $12^\circ \times 12^\circ$ 之间。根据 Acc 值来看, 最佳区域范围是在 $11^\circ \times 11^\circ$ 之间。东太平洋 (EP) 区域, 根据 Auc 值来看, 最佳区域范围是在 $9^\circ \times 9^\circ$ 之间。根据 Acc 值来看, 最佳区域范围是在 $12^\circ \times 12^\circ$ 之间。北大西洋 (NA) 区域, 根据 Auc 值来看, 最佳区域范围是在 $10^\circ \times 10^\circ$ 之间。根据 Acc 值来看, 最佳区域范围是在 $13^\circ \times 13^\circ$ 之间。

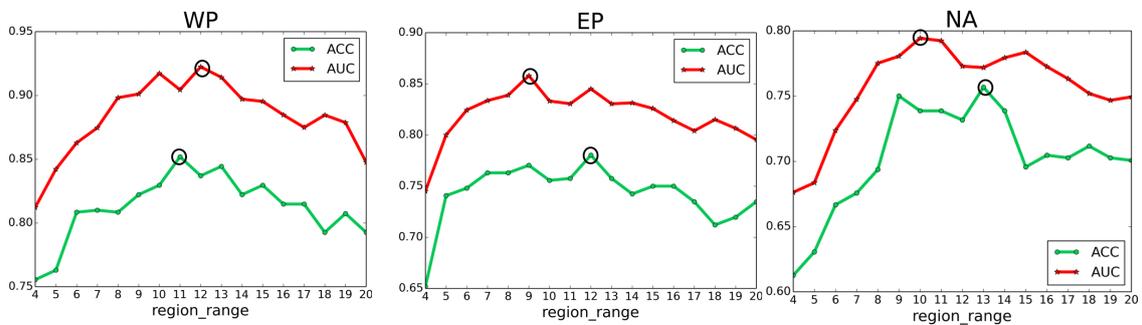


图 4.4 Auc 及 Acc 值随中心周边环境变量经纬度区域范围的变化

4.2.6.4 LSTM 时间步

时间步长是模型学习时序信息过程的一个重要参数。通常，时间步长越长，可用于预测的信息就越多。但误差的积累是否大于有效信息的积累是一个关键问题。因此，需要分别对这三个区域的数据进行测试。如图 4.5、图 4.6、图 4.7 所示，在大多数情况下“时间步 = 2”优于其他时间步。为了保证分析具有统计学意义，我们进行 t 检验进行验证，计算出 p -value。如果 p 值小于 0.05，则两种评价结果的差异具有统计学意义。 p 值越低，差异越显著。对于西太平洋区域中“时间步 = 2”和“时间步 = 1”之间的 t 检验， Acc 的 p 值为 0.021。对于西太平洋区域中“时间步 = 2”和“时间步 = 5”之间的 t 检验， Acc 的 p 值为 2.09×10^{-7} 。在东太平洋 (EP) 区域和北大西洋区域 (NA)， p 值分别是 0.002, 6.35×10^{-10} , 0.1309, 0.0147。

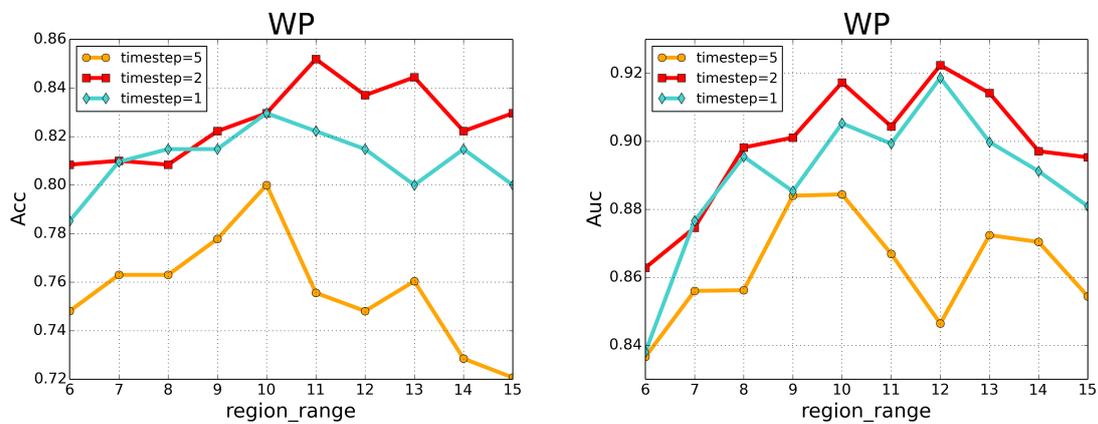


图 4.5 西太平洋 Acc 及 Auc 在不同 LSTM 时间步下随最佳区域范围的变化

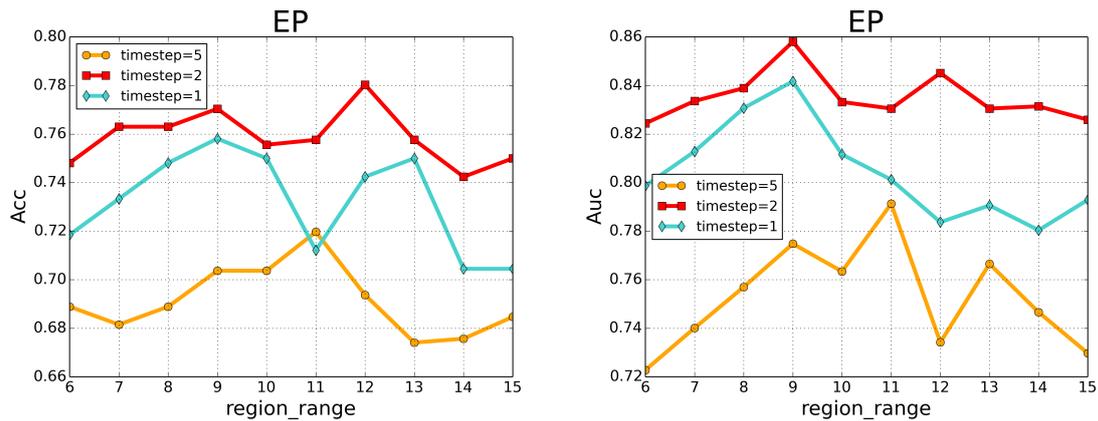


图 4.6 东太平洋 Acc 及 Auc 在不同 LSTM 时间步下随最佳区域范围的变化

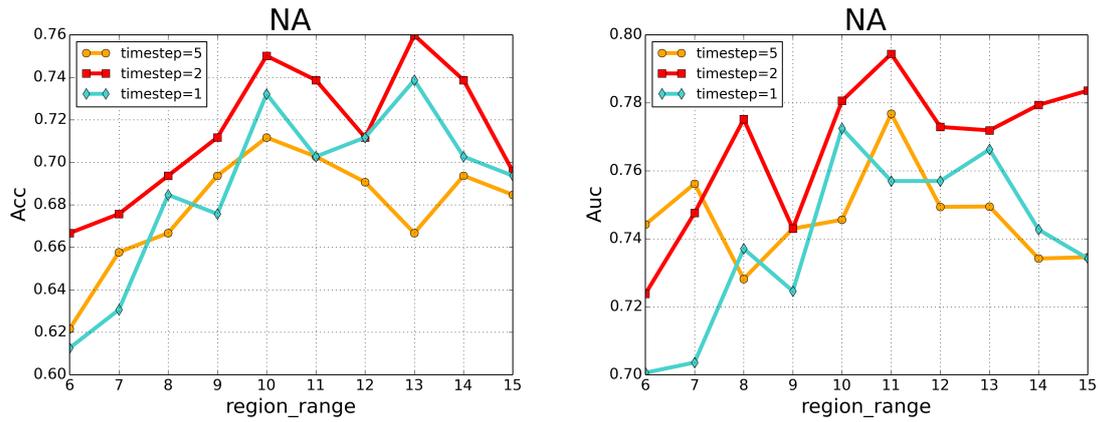


图 4.7 北大西洋 Acc 及 Auc 在不同 LSTM 时间步下随最佳区域范围的变化

4.3 基于台风时空深度混合预测模型的台风强度预测

4.3.1 实验数据集的构造与预处理

4.3.1.1 数据集的构造

台风强度预测问题可以进一步被定义为:

给定时空变量序列 $X_{t-6b}, b = 0, 1, 2, 3, \dots$, 预测台风的强度 ($Y_{t+6k} = \text{强度}$)

这是机器学习中典型的回归问题, 同样我们需要根据 4.1 节中获取到的数据进行模型输入和输出数据集的组织。对于模型的 3DCNN 部分, 所有样本 X^P 的形式是 $N \times M \times M \times L \times R$ 。对于模型的 2DCNN 部分, 所有样本 X^S 的形式是 $N \times M \times M \times O$ 。对于模型的 LSTM 部分, 所有样本的输入形式是 $N \times T \times K$ 。其中 N 是台风的样本数, 西太平洋区域 $N = 239$, 东太平洋区域 $N = 219$, 北大西洋区域 $N = 186$ 。 $M \times M$ 是中心周围经纬度范围的网格数, $M = 81$ 。 L 是大气压力层层数, $L = 12$ 。 R 是大气变量数, $R = 5$ 。 O 是海表变量数, $O = 1$ 。 T 是 LSTM 的时间步, $T = 2$ 。 K 是 3DCNN 与 2DCNN 输出特征向量合并后的维度, $K = 200$ 。以西太平洋单个台风变量序列样本为例, 需要根据时间步 T 错位构造出多个实验样本, 假设序列长度为 H , 由于回顾时间步 (LSTM 时间步) 为 T , 预测时长为 24 小时 (预测时间步为 4), 那么一个台风样本可以构造出 $H - (T + 4) + 1$ 个输入样本, 西太平洋所有的台风样本可由此构造出西太平洋台风强度预测的输入数据集。输出数据集同样按该方法进行构造, 每个输入样本对应的输出数据 (标签数据) 为单条序列起报时刻的后第 4 个强度值, 所有西太平洋的台风样本由此可以构造出西太平洋台风强度预测的输出数据集。在构造好输入数据集和输出数据集后同样 70% 的样本进行训练, 30% 的样本进行测试。

4.3.1.2 数据集的预处理

所有的缺失值、NaN 或者 inf 值均填 0，并且对所有输入变量数据和输出强度数据进行归一化操作。归一化函数如下：

$$X = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4.5)$$

其中， X 代表数据集中的所有元素， X_{max} 为数据集的最大值， X_{min} 为数据集的最小值。

4.3.2 评价指标与损失函数

4.3.2.1 评价指标

MAE 是平均绝对误差，能很好地反映预测误差。MAE 的定义如下：

$$MAE = \frac{1}{N} \sum_{i=1}^N |(Y_i - \hat{Y}_i)| \quad (4.6)$$

其中 N 是样本的数量， \hat{Y}_i 表示预测的强度值， Y_i 表示真实的强度值。

4.3.2.2 损失函数

均方根误差通常能够用于衡量预测值与真值之间的偏差，因此损失函数定义如下：

$$J = \sqrt{\frac{1}{N} \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T (y_{t,i} - \hat{y}_{t,i})^2} \quad (4.7)$$

同样， N 是样本数， $y_{t,i}$ 是第 i 个样本在 $t(t \in T)$ 时刻的真实观测值， $\hat{y}_{t,i}$ 是第 i 个样本在 t 时刻的预测值， T 为 LSTM 时间步。

4.3.3 网络设置

在该实验中，模型的 3DCNN 部分包括一个输入层，三个 Conv3D 层，一个 Flatten 层和一个 Dense 层。网络设置时第一层 Conv3D 层中卷积核大小是 $5 \times 5 \times 3$ ，步长为 $2 \times 2 \times 1$ ，其数量是 32。第二层 Conv3D 层中卷积核大小是 $5 \times 5 \times 1$ ，步长为 $2 \times 2 \times 1$ ，其数量是 64。第三层 Conv3D 层中卷积核大小是 $5 \times 1 \times 1$ ，步长为 $2 \times 1 \times 1$ ，其数量是 128，Dense 层含 100 个神经元。各层的激活函数为“relu”函数，该函数能使模型的收敛速度保持在稳定状态。

模型的 2DCNN 部分包括一个输入层，三个 Conv2D 层，一个 Flatten 层和一个 Dense 层。网络设置时第一层 Conv2D 层中卷积核大小是 5×5 ，步长为 2×2 ，其数量是 32。第二层 Conv2D 层中卷积核大小是 5×1 ，步长为 2×1 ，其数量是 64。第三

层 Conv2D 层中卷积核大小是 5×1 , 步长为 2×1 , 其数量是 128, Dense 层含 100 个神经元。各层的激活函数为 “relu” 函数。

模型的 LSTM 部分, 由于实验所选取的序列较短, 因此只选用单层 LSTM, 隐层神经元数为 100 个, 激活函数为 “relu”, 输出层的激活函数为 “sigmoid” 函数。

4.3.4 对比方法

为了保证模型在业务预测中具有可操作性, 本文对强度进行了 24 小时的预测, 并计算出了台风路径各点的强度误差。误差数据来自于 CMA 的 “西北太平洋热带气旋精度报告” [75] 和的 NHC “美国国家飓风中心预报预测报告” [76]。所选数值预测模型, 包括区域模型 “Hurricane Weather and Research Forecasting Model(HWRF)”, “Tropical Cyclone Model based on Global Regional Assimilation Prediction System(GRAPES-TCM)”, “ShangHai typhoon Model” (SHTM), 全球模型 “ECMWF global model(EMX)”, 整体模型 “Florida State Super Ensemble(FSSE)”, 和集合模型 “Global Forecast System(GFS)”。所选择的统计预测方法包括 “Climatology and Persistence model(CLIPER5)”、Partial Least Square Regression Scheme(PLS)、 “Western North Pacific Tropical Cyclone Intensity Prediction Scheme(WIPS)” 和统计动力模型 “Statistical Hurricane Intensity Prediction Scheme(SHIPS)”。本文从 CMA 和 NHC 的报告中得到了过去 5 年的强度误差, 并与过去 5 年的平均误差比较。

4.3.5 结果分析

表 4.3 台风强度预测误差 (kt) 对比

	西太平洋 (WP)		东太平洋 (EP)		北大西洋 (NA)
	Model	MAE	Model	MAE	MAE
统计模型	PLS	11.5	SHIFOR5	10.2	12.3
	WIPS	10.9	SHIPS	11.5	9.1
	GFS	12.9	GFS	10.8	11.9
数值模式	IFS	14.3	EMXI	16.8	10.6
	TCM	13.7	HWRF	11.6	9.3
时空深度混合预测模型		7.4		9.4	9.4

表 4.3 可以明显得出本文模型的预测结果明显优于西太平洋 (WP) 和东太平洋 (EP) 区域的现有方法。西太平洋 (WP) 区域, 本文模型的 MAE 是 7.4kt, 而现有模型的 MAE 最低为 10.9kt。东太平洋 (EP) 区域, 本文模型的 MAE 为 9.4kt, 而现有模型的 MAE 最低为 10.2kt。不过 MAE 在北大西洋 (NA) 区域无明显优

势。虽然本文模型在北大西洋 (NA) 区域未取得最优的预测结果, 预测误差 MAE 为 9.4kt 大于 WIPS 模型的 9.1kt, 但是仍然低于统计模型和数值模型中的大多数 MAE。总的来说, 统计模型在现有模型中有明显优势, 但与现有模型相比, 本文模型在台风强度预测效果方面有了很大的提升。

4.4 模型预测实验总结

本章的预测实验首先介绍了如何获取实验所需的台风数据集和气象海洋再分析数据集, 且给出了具体可操作的实现步骤。在调研以往相关文献的基础上, 选择了台风形成问题及强度预测问题两大难点作为切入点, 来证明本文基于时空深度混合预测模型的有效性。

在进行台风的形成预测时, 文中第一部分将台风形成预测问题定义成了深度学习可以处理的问题, 然后根据重新定义的问题构造了训练拟合模型的实验数据集。第二部分对数据集进行预处理, 并且为了提升模型的收敛速度以及保证模型的精度而进行归一化操作。第三部分选择模型的评价指标和损失函数, 采用 ACC 和 AUC 值来评价模型的优劣, 并确定了训练模型的目标函数。第四部分给出了模型内部 3DCNN、2DCNN 以及 LSTM 的具体网络设置, 并以特定数据样本为例, 对模型的输入输出流进行了解释。第五部分阐述了与本文模型对比的传统机器学习方法 LR、LDA、DT 以及原始的 LSTM 和改进的 LSTM 方法。第六部分对三大海域西太平洋 (WP)、东太平洋 (EP)、北大西洋 (NA) 的测试样本进行了预测, 发现本文的时空深度模型要优于对比的所有机器学习方法。WP 区域最高 Acc 为 0.852, 最高 AUC 为 0.922。EP 区域最高 Acc 为 0.780, 最高 AUC 为 0.858。NA 区域最高 Acc 为 0.759, 最高 AUC 为 0.794。第七部分对训练模型过程的重要参数进行了分析, 首先是模型的学习率, 发现过大模型不收敛, 过小训练时间太长, 最后发现其最佳选择为 10^{-4} 量级。随后是模型训练时的迭代次数 epoch, 发现过大模型易过拟合, 而过小则易欠拟合, 无法达到预测所需的准确度, 最后发现 epoch 在 25 ~ 35 能取得最佳预测效果。再对数据集构建时的中心周围区域范围进行了分析, 发现并非选择的经纬度区域范围越大预测效果越好, 而是在 $9^{\circ} \times 9^{\circ} \sim 13^{\circ} \times 13^{\circ}$ 之间能保证最好的预测效果。最后对模型内部和数据集构建时的 LSTM 的时间步进行了分析, 同样发现 LSTM 并非越大越好, 而是在取 2 时能达到最佳预测效果。

进行台风的强度预测时, 基本思路与台风形成预测一致。但是在实验数据集的构造与与处理、评价指标与损失函数、网络设置、对比方法等各个部分仍与台风形成预测不完全一致。其最主要的区别在于数据样本构造时, 一整条热带气旋最佳路径数据只能构造出一个台风形成预测样本, 但是可以通过错位的方法构造出

多个强度预测样本，从而扩大了样本的数量，利用模型的学习。且强度预测属于回归问题，因此在评价指标、损失函数以及网络设置时与台风形成预测存在差别。对比方法中对比的是现有官方预报机构的模型误差，因此更具有实际意义。而且实验结果发现，在 WP 和 EP 区域本文模型明显优于现有的业务预报模型，NA 区域虽却稍逊于统计模型，但是总体来说，已经优于绝大部份业务模型，说明模型具备实际应用的价值。

第五章 总结与展望

5.1 研究工作总结

本文主要研究工作总结如下：

(1) 在现有数值预报、统计学习以及机器学习台风预测方法的基础上，为弥补其计算代价过大、理论机制不明确、未考虑预测变量的复杂关系等不足之处，结合台风数据和环境大气海洋数据的时空特性，将台风预测问题具化定义为了一个时空变量序列预测的问题。

(2) 针对被重新定义的台风预测问题，联系深度学习的优势，提出了一个基于深度学习的解决方法，即本文的台风时空深度混合预测模型。其主要建模思想是利用历史大气和海洋的空间地理数据以及台风最佳路径资料，通过 3DCNN 进行大气变量的三维空间特征学习，2DCNN 进行海表变量的二维空间特征学习，LSTM 进行台风从形成之前到形成之后的时序特征学习，来达到预测台风的形成以及形成后强度、路径、风雨量变化等多种指标的目的。该模型可以根据不同的预测目标，选择不同的输入、输出（标签）数据集进行模型的训练学习，从而预测不同的对象。其优点是可扩展性强，计算代价小，且易于实现。

(3) 具体应用是从台风预测中的重点以及难点问题 - 台风形成及台风强度预测出发，根据本文提出的台风时空深度混合模型，进行了台风形成与台风强度的预测实验分析以及模型训练过程的重要参数分析。实验过程中工作主要集中于实验数据集的构造、模型中网络的设计与实现，参数的调整以及和与其它方法的对比分析。本文给出了利用 IBrATS 热带气旋最佳路径数据和 ECMWF 的 python API 对台风样本空间大气海洋环境变量数据的获取方法，以及进行台风样本输入数据集构造的具体步骤。采用了以 TensorFlow 作为后端的 Keras 深度学习框架实现测试模型，再以手动调参和调用 Sklearn 的 GridSearchCv 模块搜索最优参数的方法，确定了模型的最优超参数和输入数据规模的选择。

本文利用了西太平洋、东太平洋和北大西洋的历史台风最佳轨迹数据集和相关大气和海面环境变量进行模型训练和测试。在这三个领域的大量实验表明，本文所提出的基于深度学习的台风预测关键技术 - 时空深度预测模型优于现有的多种方法，包括许多官方机构使用的传统数值预测模型、统计预测方法和基于机器学习的方法。该模型进行台风形成预测时，准确度能达到 85.2%，Auc 值能达到 92.2%，而进行台风强度预测时，误差可以减小到 7.4kt，这使得台风的形成和强度预测有了很大的改进。此外，还对重要参数进行了分析。实验表明，模型训练时超参数如学习速率最好设置为 10^{-4} 的量级，并在 25 ~ 35 个 epoch 后停止模

型的训练，因为再加大 epoch 数会导致模型训练误差不断减小而测试误差不断升高。输入数据集构造过程中在确定台风中心周围经纬度区域范围时最好确定在 $9^{\circ} \times 9^{\circ} \sim 13^{\circ} \times 13^{\circ}$ 之间。此外，在预测 24 小时后的台风活动的实验中，回顾时间步即 LSTM 时间步在为 2 时能获得最好的预测效果。

5.2 未来工作展望

(1) 在研究问题的理解与定义方面，结合台风的生成以及变化发展理论，进行更深入的探究。比如对热带气旋生成的基本条件、大气背景场条件、触发机制进行量化定义，充分发掘台风的活动与其它现象如气候变化之间的关系，从而对问题进行深入的剖析与解释。

(2) 在历史数据的收集和获取方法方面，热带气旋最佳路径数据可以不仅仅使用 IBTrACs 数据集，大气海洋空间环境数据也可以采用多种现有的公开数据。如台风数据还可以使用模式模拟的台风数据，或者是其它官方组织颁布的数据。大气海洋空间环境数据可以选用卫星遥感、实测甚至多种资料的融合数据等。通过增大数据的优势是可以增加模型训练的有效性，但是同样需要考虑的是这样会增加数据获取、处理的难度。

(3) 在模型的设计方面，可以针对重新发现的台风相关特性结合更适宜的基础算法，或者深入研究机器学习及深度学习的其它相关理论，提出更有效的模型。

(4) 在实际预测应用方面，需要进行更长期的预测，比如 48h、...、120h，甚至是更长，这样才能更有效地指导台风预报工作，发挥台风预警的作用，增强实际应用价值。

致 谢

时光飞逝，转眼间，两年半的硕士生涯已接近尾声。回顾自己的种种经历，想表达的感谢太多，却无法一一说道。仅在此处以只言碎语记下最真挚的感谢，致给最想感谢的人。

首先，对我的导师张卫民研究员表达最深最诚挚的感恩与感激。初次与张老师相识是在硕士研究生招生宣讲之时，那时张老师给我的是高高在上，无法触及的印象。而国防科大也太遥远，遥远地只是平时听过的一次次传道，却从未想过有任何的交集。然而，正是经过了这次招生宣讲，让我下定决心报考这里，并很幸运地成为了张老师的学生，从而有机会在学校里师门里学习自己以前从未了解过的知识。所以，首先感谢张老师，是因为老师是我踏入国防科大的引路人。在进入正式的学习阶段后，面临着研究方向的选择问题，张老师多次与我讨论，寻求我自己的意愿，在我给出了自己不成熟的想法后并没有表示反对而是给予了最大的支持，鼓励我结合自己本科背景进行创新探索，并给出了可以实际着手的方向。所以，这里感谢张老师，是因为老师给予了我选择的空间，让我能够做自己想做的研究。随着研究的深入，困难也不断接踵而来，由于动手能力的薄弱以及课题突破点的难以寻找，研究工作很长一段时间没有进展，随之而来的则是心态的几近崩溃。虽然张老师在这个过程中不断给予我多方面的建议，甚至抽空单独开会帮我一起寻找解决途径，可是自己仍然进展缓慢，甚至越发糟糕，不过即使是这样老师都没有给予过一次责备，相反，是很有耐心地鼓励我让我有信心继续下去。所以，这里感谢张老师，是因为老师让我体会到了老师对学生的关爱以及学会了面对研究过程中各种困难的勇气。每每以前在实验室加班的时候，总能看见张老师办公室的灯很晚还亮着，老师有时回去的时候还会亲切地和我们交谈几句，关心一下我们的研究状况。还记得一次偶然的交流中，老师曾提及过他一年的航班次数，让我为之一惊。国防科技以及社会的进步凝结着多少研究人员的辛勤与坚守，我来这儿才真真切切感受到，而这更是从张老师身上看到的。我一直认为最好的教育莫过于言传身教，所以，这里感谢张老师，教会了我勤奋、务实、坚持、奉献。我也相信正因为老师这样，我们师门才能形成了良好的学习氛围。此外，老师也总是会支持我们外出学习，无论收获颇丰还是收效甚微，我知道老师也只是希望我们能从一次次外出经历尽可能转换为自己的知识。老师不仅在学习上给予我们指引还会关心我们的生活，经常会组织师门聚餐并和大家一起交流，讲述出差过程中的一些所见所闻，这让我感觉到了身在老师的师门真的是一件很

幸运的事情。同时也真诚地感谢师母，如果说张老师把我们当学生看待的话，师母则更像是把我们当孩子一样看待，让我感受到了浓浓的亲近感。记得第一次与师母相见是在老师家里聚餐，十几个人的饭菜全部由师母主导操办，这种家庭聚餐多多少少会为远在外地求学的学生一种家的感觉，实则难能可贵。师母在我们平时的聚餐、微信交流中也好似也是我们中的一员，完全没有任何的距离感。所以，感谢师母，是因为师母给了我们师门所有人除家人外最温暖的感觉。

真心且真诚地感谢我的协助指导老师汪祥助理研究员，老师同样也是我研究路上极为关键的人。刚入学时还未曾料想会有汪老师指导，所以，我也是我们师门中比较特殊的存在。汪老师从选题开始一直到现在都不忘给予我引导，给予我督促，因此，课题的顺利进行完全离不开汪老师的帮助。研究之初，面对完全不熟悉的领域，老师给了我许多实质性的指引，包括给我发文献，并且细到如何搜索文献，读文献，做笔记。当我表示文献晦涩难懂时会专门加班来和我一起读文献，当我表示编程能力薄弱的时候会帮我找基础技术文档，当我表示动力能力欠缺的时候会教我从装系统开始熟悉，当我表示研究中存在各种各样的问题的时候会和我一起讨论，当我需要修改论文、汇报 PPT 时会耐心细致地告诉我修改的方法并帮助我修改。最重要的是在低落的时候，一直不忘鼓励我，仍然相信我，并和我分享自己的经验。虽然现在的我可能只刚刚迈入了研究的大门，但是这中间却从老师身上学到了太多太多。所以，感谢汪老师，感谢老师教会我的一切。

感谢我的专业课任课老师殷建平老师，张春元老师，吴建平老师，祝恩老师，汪昌健老师，龚春叶老师，刘新旺老师，面对从未接触过的新领域知识，虽有难度，但幸得各位老师的指点，能让我在短时间内能尽快地消化，进而激发自己对这个领域的兴趣，成为了人生路上一个重要的转折点。同样感谢气象海洋学院的宋君强老师，赵军老师，朱敏老师，任开军老师，曹小群老师，王舒畅老师，李金才老师，赵延来老师，赵文静老师，柯朝雄老师，李小勇老师，冷洪泽老师，戴海晋老师，赵娟老师，老师们专业的研究态度和亲和的待人态度让我深度启发，认识老师们也让我倍感荣幸。还需重点感谢我的日语老师田靖老师，田老师气质优雅，但又不缺有趣的灵魂。老师在课堂上总会分享她的许多故事，让我也心存向往。在私下有任何问题同样会十分有耐心地给予我解答，让我得以在研一迷茫时知道需要坚信自己，坚定自己想走的路。

感谢师门的朱孟斌师兄，余意师姐，皇群博师兄，刘柏年师兄，邵彩霞师姐，方民权师兄，段博恒师兄，孙敬哲师兄，张泽师姐，王品强师兄，陈妍师姐，邢德师兄，邢翔师兄，朱祥茹师姐，李松师兄，王晓慧师姐，冯淼，罗藤灵，木子

尧师弟、殷何卿师弟。是师兄师姐、师弟们让我觉得师门中的我们更像是一个大家庭。从刚开始入师门时作为一个懵懵懂懂的师妹，到现在成为一个会给师弟师妹们答疑的师姐，这中间少不了大家的帮助，所以十分有幸处在一个这样和睦的师门，也十分有幸认识师门中的大家。同样感谢崔英博师兄，王恒师兄，赵成武师兄，孙迪夫师兄，张文师姐，林时剑师兄，孙嘉琪，孙祥远师弟，张家灏师弟，黄丽蓝师妹，学习生活中很多时候同样离不开大家的帮助，所以感谢有你们，让我在学校的生活变得更好。

感谢我的同学朱啸宇，感谢他在我最灰暗迷茫的时候一直不忘给我鼓励，并耐心为我分享经验，感谢我们能成为很好的朋友，互诉心事。感谢董懂，感谢她从研一开始便帮助我消灭吃不完的零食，感谢她作为朋友的每一份细致的关心，以及在平时带给我们的开心。感谢杨超，感谢他在数学建模竞赛中教会我的学习方法，并在课题实现中帮助寻求解决途径。感谢赵双，感谢她在我基本完全不懂编程的情况下，一个个字母一行行代码地帮助我打下了编程的基础。感谢我们班的冯淼，崔姝瑶，寻微微，付靖雯，崔亚文，段尧，感谢大家每年都会帮我过生日，感谢大家平时都可以互相帮助互相照料。感谢我的朋友肖仕琳，感谢她会在任何我需要陪伴的时候陪伴我。感谢黄日波，感谢他会在任何我需要帮助的时候给予我帮助。感谢廖伊，感谢和我有共同的爱好的她，让我在苦闷的时候能有发泄的途径。感谢李乐天，感谢他的乐观有趣，让我觉得生活充满了趣味。

感谢我的家人，无论我做什么都给与我最大的支持。爸爸会在我课题研究受挫而心情低落的时候，每天给我打电话，关心我的状况，并讲很多事例安慰我，而在我学业工作繁忙的时候，却不会打扰我。妈妈会在我每次回家的时候都会尽心尽力地陪着我，在我熬夜写论文的时候，会推掉自己的工作，连着给我送我几天的饭，全心全意付出她可以付出的一切。感谢我的外婆，感谢伴我成长的她依然如小时候一般地关爱我，每个周末都会打电话给我让我去她那儿吃饭，虽然大多数时候很忙都没有去，但是心底仍然很暖。感谢我家庭中的所有人，感谢我所在的不仅仅是一个温暖的小家庭，而且是一个有爱的大家庭。

感谢所有给过我帮助，给过我积极影响甚至是一个微笑的人，是你们让我觉得其实每一天都可以很美好。

最后衷心感谢百忙之中抽空评审的专家们，我知道论文还有很大的改进空间，谢谢老师们的指点。

回想硕士两年半，这是我二十三年记忆中最不后悔的时光。我知道我在进步，我知道我在朝着我自己想走的方向前行，我知道前方的道路纵有泥泞也终究会敞亮。毕业不是终点，而是下一征程的起点。未来，你好!

参考文献

- [1] 陈联寿. 台风预报及其灾害 [M]. Qi xiang chu ban she, 2012.
- [2] Wang Y, Wu C-C. Current understanding of tropical cyclone structure and intensity changes—a review [J]. *Meteorology and Atmospheric Physics*. 2004, 87 (4): 257–278.
- [3] DeMaria M, Sampson C R, Knaff J A, et al. Is tropical cyclone intensity guidance improving? [J]. *Bulletin of the American Meteorological Society*. 2014, 95 (3): 387–398.
- [4] 气象局. 中国气象灾害年鉴 [J]. 北京: 气象出版社. 2007.
- [5] Kettinger W J, Teng J T, Guha S. Business process change: a study of methodologies, techniques, and tools [J]. *MIS quarterly*. 1997: 55–80.
- [6] Velden C, Harper B, Wells F, et al. The Dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years [J]. *Bulletin of the American Meteorological Society*. 2006, 87 (9): 1195–1210.
- [7] Knabb R D, Rhome J R, Brown D P. Tropical cyclone report: Hurricane katrina, 23-30 august 2005 [M]. National Hurricane Center, 2005.
- [8] Halperin D J, Fuelberg H E, Hart R E, et al. An evaluation of tropical cyclone genesis forecasts from global numerical models [J]. *Weather and Forecasting*. 2013, 28 (6): 1423–1445.
- [9] DeMaria M, Knaff J A, Connell B H. A tropical cyclone genesis parameter for the tropical Atlantic [J]. *Weather and Forecasting*. 2001, 16 (2): 219–233.
- [10] Roy C, Kovordányi R. Tropical cyclone track forecasting techniques—A review [J]. *Atmospheric research*. 2012, 104: 40–69.
- [11] Stewart S R. National Hurricane Center Tropical Cyclone Report: Hurricane Matthew [J]. National Hurricane Center. 2017.
- [12] 周志华. 机器学习 [M]. Qing hua da xue chu ban she, 2016.
- [13] Jiang G-Q, Xu J, Wei J. A Deep Learning Algorithm of Neural Network for the Parameterization of Typhoon-Ocean Feedback in Typhoon Forecast Models [J]. *Geophysical Research Letters*. 2018, 45 (8): 3706–3716.
- [14] Hong S, Kim S, Joh M, et al. Globenet: Convolutional neural networks for typhoon eye tracking from remote sensing imagery [J]. *arXiv preprint arXiv:1708.03417*. 2017.

-
-
- [15] Loridan T, Crompton R P, Dubossarsky E. A Machine Learning Approach to Modeling Tropical Cyclone Wind Field Uncertainty [J]. *Monthly Weather Review*. 2017, 145 (8): 3203–3221.
- [16] Richman M B, Leslie L M. Adaptive machine learning approaches to seasonal prediction of tropical cyclones [J]. *Procedia Computer Science*. 2012, 12: 276–281.
- [17] Matsuoka D, Nakano M, Sugiyama D, et al. DETECTING PRECURSORS OF TROPICAL CYCLONE USING DEEP NEURAL NETWORKS [J].
- [18] Richman M B, Leslie L M, Ramsay H A, et al. Reducing Tropical Cyclone Prediction Errors Using Machine Learning Approaches [J]. *Procedia Computer Science*. 2017, 114: 314–323.
- [19] 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 (5): 8–12.
- [20] Alemany S, Beltran J, Perez A, et al. Predicting Hurricane Trajectories using a Recurrent Neural Network [J]. arXiv preprint arXiv:1802.02548. 2018.
- [21] Chaudhuri S, Dutta D, Goswami S, et al. Track and intensity forecast of tropical cyclones over the North Indian Ocean with multilayer feed forward neural nets [J]. *Meteorological Applications*. 2015, 22 (3): 563–575.
- [22] Chen B, Chen B-F, Lin H-T. Rotation-blended CNNs on a New Open Dataset for Tropical Cyclone Image-to-intensity Regression [J]. 2018.
- [23] Pradhan R, Aygun R S, Maskey M, et al. Tropical Cyclone Intensity Estimation Using a Deep Convolutional Neural Network [J]. *IEEE Transactions on Image Processing*. 2018, 27 (2): 692–702.
- [24] 马雷鸣. 国内台风数值预报模式及其关键技术研究进展 [J]. *地球物理学进展*. 2014, 29 (3): 1013–1022.
- [25] Zou X, Xiao Q. Studies on the initialization and simulation of a mature hurricane using a variational bogus data assimilation scheme [J]. *Journal of the atmospheric sciences*. 2000, 57 (6): 836–860.
- [26] TAN Z-M, et al. Tropical cyclone initialization with dynamical retrieval from a modified UWPBL model [J]. *Journal of the Meteorological Society of Japan. Ser. II*. 2010, 88 (5): 827–846.

-
-
- [27] Patoux J, Brown R A. A gradient wind correction for surface pressure fields retrieved from scatterometer winds [J]. *Journal of Applied Meteorology*. 2002, 41 (2): 133–143.
- [28] Stevens B, Duan J, McWilliams J C, et al. Entrainment, Rayleigh friction, and boundary layer winds over the tropical Pacific [J]. *Journal of climate*. 2002, 15 (1): 30–44.
- [29] Knaff J A, Sampson C R, DeMaria M. An operational statistical typhoon intensity prediction scheme for the western North Pacific [J]. *Weather and Forecasting*. 2005, 20 (4): 688–699.
- [30] Fritsch J, Chappell C. Numerical prediction of convectively driven mesoscale pressure systems. Part II. Mesoscale model [J]. *Journal of the Atmospheric Sciences*. 1980, 37 (8): 1734–1762.
- [31] DeMaria M, Kaplan J. A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin [J]. *Weather and Forecasting*. 1994, 9 (2): 209–220.
- [32] DeMaria M, Kaplan J. An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern North Pacific basins [J]. *Weather and Forecasting*. 1999, 14 (3): 326–337.
- [33] Knaff J A, DeMaria M, Sampson C R, et al. Statistical, 5-day tropical cyclone intensity forecasts derived from climatology and persistence [J]. *Weather and Forecasting*. 2003, 18 (1): 80–92.
- [34] Emanuel K A. The maximum intensity of hurricanes [J]. *Journal of the Atmospheric Sciences*. 1988, 45 (7): 1143–1155.
- [35] DeMaria M, Kaplan J. Sea surface temperature and the maximum intensity of Atlantic tropical cyclones [J]. *Journal of Climate*. 1994, 7 (9): 1324–1334.
- [36] Bogner P B, Barnes G M, Franklin J L. Conditional instability and shear for six hurricanes over the Atlantic Ocean [J]. *Weather and forecasting*. 2000, 15 (2): 192–207.
- [37] Gray W M. The formation of tropical cyclones [J]. *Meteorology and atmospheric physics*. 1998, 67 (1-4): 37–69.
- [38] Emanuel K, Nolan D S. Tropical cyclone activity and the global climate system [C]. In *26th Conference on Hurricanes and Tropical Meteorology*. 2004.
- [39] Camargo S J, Emanuel K A, Sobel A H. Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis [J]. *Journal of Climate*. 2007, 20 (19): 4819–4834.

-
-
- [40] 赵军平, 吴立广, 赵海坤. 西北太平洋热带气旋潜在生成指数的改进 [J]. 气象科学. 2012, 32 (6): 591–599.
- [41] Zhang W, Fu B, Peng M S, et al. Discriminating developing versus nondeveloping tropical disturbances in the western North Pacific through decision tree analysis [J]. *Weather and Forecasting*. 2015, 30 (2): 446–454.
- [42] Fu B, Peng M S, Li T, et al. Developing versus nondeveloping disturbances for tropical cyclone formation. Part II: Western North Pacific [J]. *Monthly weather review*. 2012, 140 (4): 1067–1080.
- [43] Peng M S, Fu B, Li T, et al. Developing versus nondeveloping disturbances for tropical cyclone formation. Part I: North Atlantic [J]. *Monthly weather review*. 2012, 140 (4): 1047–1066.
- [44] Hennon C C, Hobgood J S. Forecasting tropical cyclogenesis over the Atlantic basin using large-scale data [J]. *Monthly weather review*. 2003, 131 (12): 2927–2940.
- [45] Hennon C C, Helms C N, Knapp K R, et al. An objective algorithm for detecting and tracking tropical cloud clusters: Implications for tropical cyclogenesis prediction [J]. *Journal of Atmospheric and Oceanic Technology*. 2011, 28 (8): 1007–1018.
- [46] Schumacher A B, DeMaria M, Knaff J A. Objective estimation of the 24-h probability of tropical cyclone formation [J]. *Weather and Forecasting*. 2009, 24 (2): 456–471.
- [47] Chand S S, Walsh K J. Forecasting tropical cyclone formation in the Fiji region: A probit regression approach using Bayesian fitting [J]. *Weather and Forecasting*. 2011, 26 (2): 150–165.
- [48] Jaiswal N, Kishtawal C M. Prediction of tropical cyclogenesis using scatterometer data [J]. *IEEE Transactions on Geoscience and Remote Sensing*. 2011, 49 (12): 4904–4909.
- [49] Jarvinen B R, Neumann C J. Statistical forecasts of tropical cyclone intensity for the North Atlantic basin [J]. 1979.
- [50] Elsberry R L, Weniger E L, Meanor D H. A statistical tropical cyclone intensity forecast technique incorporating environmental wind and vertical wind shear information [J]. *Monthly weather review*. 1988, 116 (11): 2142–2154.
- [51] Chu J-H. A Regression Model for the Western North Pacific Tropical Cyclone Intensity Forecast. [R]. 1994.

-
-
- [52] 钮学新. 热带气旋强度预报的 DTS 方案 [J]. 大气科学. 1984, 8 (4): 450–456.
- [53] Grandau F J, Engel G T. 2001 Annual Tropical Cyclone Report [R]. 2001.
- [54] Koller D, Friedman N, Bach F. Probabilistic graphical models: principles and techniques [M]. MIT press, 2009.
- [55] Abramson B, Brown J, Edwards W, et al. Hailfinder: A Bayesian system for forecasting severe weather [J]. International Journal of Forecasting. 1996, 12 (1): 57–71.
- [56] Ebert-Uphoff I, Deng Y. A new type of climate network based on probabilistic graphical models: Results of boreal winter versus summer [J]. Geophysical Research Letters. 2012, 39 (19).
- [57] Spirtes P, Glymour C N, Scheines R, et al. Causation, prediction, and search [M]. MIT press, 2000.
- [58] Robinson R W. Counting unlabeled acyclic digraphs [M] // Robinson R W. Combinatorial mathematics V. Springer, 1977: 1977: 28–43.
- [59] Le T, Hoang T, Li J, et al. A fast PC algorithm for high dimensional causal discovery with multi-core PCs [J]. IEEE/ACM transactions on computational biology and bioinformatics. 2016.
- [60] Kalisch M, Mächler M, Colombo D, et al. Causal inference using graphical models with the R package pcalg [J]. Journal of Statistical Software. 2012, 47 (11): 1–26.
- [61] 李航, et al. 统计学习方法. 2012.
- [62] 姚才, 金龙, 黄明策, et al. 遗传算法与神经网络相结合的热带气旋强度预报方法试验 [J]. 海洋学报 (中文版). 2007, 4.
- [63] 郑颖青, 余锦华, 吴启树, et al. K-均值聚类法用于西北太平洋热带气旋路径分类 [J]. 热带气象学报. 2013, 29 (4): 607–615.
- [64] Zhang W, Gao S, Chen B, et al. The application of decision tree to intensity change classification of tropical cyclones in western North Pacific [J]. Geophysical Research Letters. 2013, 40 (9): 1883–1887.
- [65] Wijnands J S, Qian G, Kuleshov Y. Variable selection for tropical cyclogenesis predictive modeling [J]. Monthly Weather Review. 2016, 144 (12): 4605–4619.
- [66] Chaudhuri S, Basu D, Das D, et al. Swarm intelligence and neural nets in forecasting the maximum sustained wind speed along the track of tropical cyclones over Bay of Bengal [J]. Natural Hazards. 2017, 87 (3): 1413–1433.

- [67] Li Y, Yang R, Yang C, et al. Leveraging LSTM for rapid intensifications prediction of tropical cyclones. [J]. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences. 2017, 4.
- [68] LeCun Y, Touresky D, Hinton G, et al. A theoretical framework for back-propagation [C]. In Proceedings of the 1988 connectionist models summer school. 1988: 21–28.
- [69] Goodfellow I, Bengio Y, Courville A, et al. Deep learning [M]. MIT press Cambridge, 2016.
- [70] LeCun Y, Bengio Y, et al. Convolutional networks for images, speech, and time series [J]. The handbook of brain theory and neural networks. 1995, 3361 (10): 1995.
- [71] Elman J L. Finding structure in time [J]. Cognitive science. 1990, 14 (2): 179–211.
- [72] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural computation. 1997, 9 (8): 1735–1780.
- [73] Graves A, Mohamed A-r, Hinton G. Speech recognition with deep recurrent neural networks [C]. In Acoustics, speech and signal processing (icassp), 2013 ieee international conference on. 2013: 6645–6649.
- [74] Xingjian S, Chen Z, Wang H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting [C]. In Advances in neural information processing systems. 2015: 802–810.
- [75] Rudolph D K, Guard C P. 2016 Annual Tropical Cyclone Report [R]. 2016.
- [76] Cangialosi J P, Franklin J L. 2016 National Hurricane Center forecast verification report [J]. National Hurricane Center. 2016.

作者在学期间取得的学术成果

发表的学术论文

- [1] Rui Chen, Xiang Wang, Weimin Zhang, Xiaoyu Zhu, Aiping Li and Chao Yang. A Hybrid CNN-LSTM Model for Typhoon Formation Forecasting (MATES2018, GeoInformatica, SCI/CCF B 类期刊, IF:2.091, 已录用)

申请的专利

- [1] 汪祥, 陈睿, 张卫民, 李金才, 李小勇, 朱啸宇, 基于深度学习混合 CNN-LSTM 的台风预测方法 (申请中)

获得的奖项

- [1] 2016 年第十三届“华为杯”全国研究生数学建模竞赛三等奖
[2] 2017 年第十四届“华为杯”全国研究生数学建模竞赛二等奖
[3] 2017-2018 学年度国防科技大学计算机学院优秀学员