

学号 20154414

密级

东北大学本科毕业论文

基于深度学习的台风路径预测技术 的研究与实现

学 院 名 称：计算机科学与工程学院

专 业 名 称：计算机科学与技术

学 生 姓 名：朱强

指 导 教 师：乔百友 副教授

二〇一九年六月

郑重声明

本人呈交的学位论文，是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料真实可靠。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确的方式标明。本学位论文的知识产权归属于培养单位。

本人签名： 日期：

摘 要

家喻户晓，台风是一种破坏性极强的极端自然灾害现象，据美国海军联合台风警报中心统计，西北太平洋及中国南海海域平均每年都有好多台风生成，出现最多台风的月份是8月，其次是7月、9月、6月和5月。科学家曾计算出，一场中等强度的台风或者是高强度的台风所释放的能量相当于上百个氢弹、或10亿吨黄色炸药所释放能量的总和，可见其破坏力之大，危害程度之深。据世界气象组织的报告，全球每年死于热带风暴的人数约为2000到3000人。根据有关资料，西太平洋沿岸国家平均每年因台风造成的经济损失为40亿美元，危害巨大。我国也是一个台风灾害严重的国家，如何有效预测台风路径一直是一个重要的课题。

现有台风路径预测模型主要有三种，一种是WRF模型，该模型需要大量的人为参数的设定，并且预测误差较大；第二种方法就是利用飞机对海洋气象数据进行采样，并基于经验模型将过去的台风路径和当前的路径进行对比，并对卫星拍摄到的影像进行分析，来预测下一步的台风路径的走向；第三种预测方法主要采用统计模型。这三种模型存在着误差较大，预测精度不高的问题，为此本文提出了一种结合深度学习和多元回归技术的台风路径预测方法，该方法首先读取Matlab数据格式的温度，湿度比，压强的数据，并对数据进行预处理，去除掉无效数据；然后采用LSTM网络模型对温度，湿度比，压强进行预测，在此基础上，应用多元拟合回归技术实现了对台风路径的预测，实际数据集上的测试表明，该方法能够达到平均20km的精度。

结合本文提出的预测方法，设计实现了简单的台风路径预测系统，系统提供了数据录入，用户自定义文件输入，数据预处理，模型训练及预测，台风路径的演示等诸多功能。

关键词：台风路径预测；机器学习；深度学习；神经网络；LSTM

ABSTRACT

Typhoon is a devastating extreme natural disaster. According to the United States Navy's Joint Typhoon Warning Center, the number and month of typhoons in the Northwest Pacific Ocean and the South China Sea are many, with the most occurring in August, followed by July and September, June and May. Scientists have estimated that a moderate or high-intensity typhoon emits energy equivalent to that of hundreds of hydrogen bombs or a billion tons of yellow explosives combined, demonstrating how devastating and dangerous it can be. According to the World Meteorological Organization, about 2,000 to 3,000 people die each year from tropical storms. According to relevant information, the average annual economic loss caused by typhoons in the western Pacific coastal countries is US \$4 billion, which is extremely harmful. China is also a country with severe typhoon disasters. How to effectively forecast the track of typhoon has become an important issue.

There are mainly three kinds of typhoon track prediction models. One is WRF model, which needs a lot of artificial parameters and has a large prediction error. The second method is to use aircraft to sample the marine meteorological data, compare the past track with the current track based on the empirical model, and analyze the satellite images to predict the next track of the typhoon. The third forecasting method mainly adopts statistical model. In this paper, a method of typhoon track prediction based on depth learning and multivariate regression is proposed. Firstly, the data of temperature, humidity ratio and pressure in Matlab form are read, and then the data are preprocessed to remove the invalid data. Then the LSTM network model is used to predict the temperature, humidity ratio and pressure. On this basis, the multi-fitting regression technique is used to predict the track of typhoon. The actual test shows that the method achieves an average accuracy of 20 km.

Combined with the forecasting method proposed in this paper, a simple typhoon track forecasting system is designed and implemented. The system provides many functions such as data input, user-defined file input, data training, data preprocessing, typhoon track demonstration and so on.

Key words: Typhoon track prediction; Machine Learning; Deep Learning; Neural network; LSTM

目录

郑重声明.....	I
摘 要.....	II
ABSTRACT.....	III
1 绪论	
1.1 研究背景.....	1
1.2 国内外研究现状.....	2
1.3 本文研究内容.....	3
1.4 论文组织结构.....	5
1.5 本章小结.....	6
2 相关技术简介	
2.1 机器学习.....	7
2.2 循环神经网络.....	8
2.3LSTM 网络.....	9
2.4folium.....	10
2.5Python-tkinter.....	10
2.6 大气动力学.....	11
2.7 本章小结.....	11
3 基于深度学习的台风路径预测方法	
3.1 总体框架.....	13
3.2 基于 LSTM 的单要素预测.....	14
3.2.1 数据预处理.....	17
3.2.2 LSTM 单要素预测模型.....	17
3.2.3 模型训练和预测.....	20
3.3 基于多元回归的台风路径预测.....	21
3.4 预测结果分析.....	23
4 基于 SVM 的台风等级播报	

4.1 基于 SVM 进行等级播报.....	25
4.2 朴素贝叶斯和 SVM 对比.....	27
4.3 本章小结.....	28
5 台风路径预测系统的设计与实现	
5.1 系统需求分析.....	29
5.2 系统功能设计.....	31
5.3 系统的开发运行环境.....	32
5.4 系统的实现.....	32
5.4.1 系统的总体处理流程.....	32
5.4.2 录入数据模块实现.....	33
5.4.3 台风轨迹预测展示.....	34
5.4.4 散点图展示.....	35
5.5 系统功能测试.....	36
6 结论	
6.1 本文工作总结.....	41
6.2 系统展望.....	41
参考文献	42
致谢	46

1. 绪论

1.1 研究背景

我国是世界为数不多的上少数几个受台风灾害影响的较为严重的亚洲国家，自从建国以来，台风带来给东南沿海的损失不计其数，比如 5413 号台风-艾达，是横扫南方的一带的超强台风，市区郊区等建筑今早破坏，短时间内的强降雨可能引发城市内涝、滑坡泥石流等。台风带来的超等的特大风波雨对于交通而言，简直就是一场噩梦，不仅极大影响路面交通，航空飞行都不能幸免。平均每年有 7 个左右台风在我国上岸来袭。沿海地区从台湾、福建、浙江、上海、江苏、山东、河北、天津一直到辽宁等省（自治区、直辖市）以及我国中部地区，均可能受到台风活动的影响。台风引起房屋倒塌、玻璃碎片、漂浮物碰撞、高空坠落等可导致各种伤害，具有高导致人伤残的概率等特征。1990—2009 年，大陆地区平均每年台风造成的经济损失就达 300 多亿元，死亡人数众多。台风风速大都在 17 米/秒以上，甚至在 60 米/秒以上。据测，当风力达到 12 级时，垂直于风向平面上每平方米风压可达 230 公斤。因此台风大风及其引起的超等大波浪可以把万吨巨轮抛向半空拦腰折断，也可把巨轮推入内陆；也足以彻底摧毁甚至摧毁陆地上的建筑、桥梁、车辆等。特别是在建筑物没有被加固的地区，造成破坏更大。大风也可以把杂物吹到半空，使户外的生活环境变得非常危险。

史上最强台风号称蜂王的 TIP（泰培）可以说给我国东南沿海造成不可磨灭的灾害，泉州，厦门，等地均遭受到了台风的肆虐，便是建国以来最强的台风。不仅如此，台风还会对渔业造成难以估量的伤害，会使渔业减产。因为台风到来之际，会将海底的淤泥翻出来，将使鱼类的腮堵住，导致鱼类缺氧死亡。并且台风还会带来大规模降水造成洪涝灾害，导致江河水位迅速上涨，淹没农村，摧毁道路桥梁，除台风的大风外，台风所引发的降水也是引发造成人类的损失的重要原因。风浪潮造成灾害的总打的财产损失例子众多。如 5162 号台风在浙江象山登陆，登陆时最低气压 923hPa，最大风速 65m/s，6 级以上大风区直径 1000 余 km，是影响我国范围最大的台风之一，沿海大风风速超过 50m/s，杭州湾内最大风浪潮位瞰浦站 5.02m、乍浦 4.34m，增水位超过 2m 的 10 处，南起福建闽江口，

北至渤海湾,长达几千 km 范围内的沿海都出现增水高于 1.0m 的风浪潮,浙江有 75 个县市遭到极其惨重的损失中国也是遭受台风侵袭的主要受害亚洲国家,主要是由于海岸线过长,并且东临近太平洋^[1]。

台风后的医疗保健的防护措施和保健基础设施的破坏、自来水被污染、热泪居住的密集程度,改变(特别是拥挤的临时住所)、人口迁移、因为所居住的房屋在被彻底摧毁之后变得流离失所。以及生态环境的改变使得传染性疾病流行的可能性增加,不仅造成严重的人员伤亡和经济损失,而且给守在人带来了焦灼、郁闷、消沉创伤后应激障碍等严重的心理影响和精神伤害。可是台风路径预测困难重重,目前的方法效果均不太如意。国际上的预测方法大多数都是采用物理模型 WRF,而这种物理学,地球大气学,空气动力学可是传统的方法依赖于人为的参数但过于复杂。

1.2 国内外研究现状

近年来国内外的台风的路径预测方法大致分为 1. 经验法 2. 回归方程法 3. 动力预测系统^[2]。例如日本气象局会有多个气象采样点,对周围的情况进行实时采样,对大量的物理参数进行拟合带入模型中从而对路径进行大致的预测进而判断出风力 的衰减,登陆的时间。

热带高压对台风也会有影响。气象学相关的专家表示,如果副热带高压加强西伸至我国大陆,那么台风就有可能沿着核心一起向西登陆我国。相反,如果副热带高压西伸得没那么远,就会留给台风一个近海北上的机会,台风就不会登陆我国。因此,副热带高压的发展情况多变,也加大了台风路径预报的难度,所以对于台风路径的预报难度很大^[3]。

卫星监测也是一个行之有效的手段,中国已经建立了包括 193 部天气雷达、6 万多个自动气象站,超过 2500 个国家级观测站、及风云系列卫星等组成的综合观测系统。自 1988 以来,中国成功放飞 16 颗风云卫星,目前 9 颗在轨运行。长期以来,风云卫星为台风的监测预报提供了重要支撑。静止气象卫星重点在于台风的实时定位定强^[4],并监测台风登陆的时间地点和造成的风雨影响。极轨气象卫星具有大气三维探测能力,可以揭示台风内部热力和云雨结构,为台风路径和强度预报提供依据。在过去岁月中,中国中央气象局积极开展台风客观预报释用和融合技术的研发,突破了登陆台风精细化预报的技术瓶颈。在台风的路径集

合成员智能化应用、多源资料融合技术和降尺度技术研究等方面取得显著进展,其中台风大风半径分析及预告、台风风场适用技术、暴雨宏微观物理过程模式定量分析和应用技术等增加了我国台风风雨精细化预报技术业务空缺。构建台风监测、预报预警综合平台,显著提高了台风预报综合决策能力。1994年,中国气象局建立了第一代区域台风模式 RMTTP。2004年升级到 T213 全球模式,该模式水平分辨率为 0.5625° ,垂直分辨率为 31 层。2010年将 T213 模式升级为 T639 全球模式,水平分辨率提高到 0.28125° ,垂直分辨率提高到 60 层。2012年开发了 GRAPES_TYM 区域台风模式,该模式水平分辨率为 0.15° ,垂直分辨率为 32 层;2015年将其水平分辨率提高到 0.12° ,垂直分辨率提高到 50 层。台风预报业务的进展尤其是台风路径预报误差的降低趋势极大地促进了防台减灾工作的开展。还有一种模型叫做 WRF,是一种较为有效的台风数值模型,他是非静太力平衡的的数值模式,垂直的方向采用了 eta 坐标。在全球模式下,WRF(Weather Research And ForecastModel)不仅可以进行天气预报,还可以将台风的对大气化学,气溶胶等进行播报,主要的框架有 2 个分别是 NMM 和 ARW。NMM(Nonhydrostatic Mesoscale Model)是基于 NCEP 发展起来的一种模型,而 ARW(Advanced Research WRF)是在 NCAR 模式下发展起来。采用 Fortran90 编写,由于其强大的数据同化,先进的物理过程,在对流和中尺度的降水处理方面更有优势。而美国采用综合预报法,他是引导预报法和模拟相似法,并且也综合预报员的历史经验对大西洋和太平洋的飓风路径进行播报预测。

建国以来到 2000 年,西北太平洋共计发生热带气旋 1444 次^[5],平均 27.7 次。大多数的台风均是位于西北太平洋,一大部分在菲律宾的吕宋岛进行分流,一部分 经过我国的台湾省奔向广东,一部分北上进入日本^[6],另一部分成曲线直奔东南沿海省份福建,江苏。国内外目前许多预测机构采用物理的模型 WRF,但是该模型有需要大量的人为设置的参数,效果并不理想,本文论述的是基于深度学习的预测系统,主要采用长短期记忆神经网络(Long-Short Term Memory)^[7],分别对位势高,温度,湿度^[8]进行预测,在用机器学习的线性回归进行预测出递推方程,通过递推方程来预测出接下来的路径情况。

除此之外,对于自然灾害的综合风险评估是将灾害脆弱性评价^[9]和风险评估相结合,对于台风问题的灾害研究应该在减灾工程和非工程的措施相结合,对于台风的研究早在 20 世纪就开始了,早期的台风的形成机制,一些学者对东亚台

风形成, 路径, 结构均有研究。在 1950 年, 科学家李宪之提出了‘台风的生成的综合学说’^[10], 除此之外还有一种研究方法是台风数值模拟^[11], 这个模拟的方法是根据‘热带气旋发生发展第二类的条件不稳定理论’^[12]。在 Oogama 等诸多科学家提出的多尺度相互作用的模型使得热带气旋理论^[13]得以发展。还有一种是根据致灾因子风险的分析, 气象学要素提取。我们得到‘风眼’坐标, 结合气象站点的图, 进行叠加, 再根据各个场台风的气象站点结合国家气象数据库, 对本次台风的各个要素进行提取, 进而计算出最大的风速, 总和降雨量, 其中台风等级分为‘巨灾’、‘大灾’、‘中灾’、‘小灾’和‘微灾’。

机器学习对台风了路径的预测是一个非常火的方向, 是统计学的重要分支, 相比于传统动力学需要的参数较小, 便于分析误差^[14]。预测模型和算法采用长短期记忆神经网络 LSTM, 是一类将前面结果进行综合考虑且具有深度结构的前馈神经网络^[15] (Feedforward Neural Networks), 是深度学习的代表算法之一。另一方面, 深度神经网络算法又有着高度复杂、高度非线性、运算形式不规整的特点^[16]。神经网络的非线性拟合强, 相比于传统的统计-动力学技术, 运算资源需求相对较低^[17], 本文主要采用机器学习技术来构建南海台风路径预测模型及测试系统。

1.3 本文研究内容

本文主要提出了一种基于深度学习技术的台风移动路径预测方法, 并设计实现了一个台风移动路径预报测试系统, 对提出的方法和实现的系统进行了测试, 主要做了以下工作:

1. 对以往模型进行了对比分析。以前的 WRF 的物理模型弊端较多。这些模型依赖于发现过去台风的模型, 当前台风和过去发生过的相似程度, 导致科学家在对安德鲁这样的台风的发展方向进行为期三天的粗略预测, 它的平均误差约为 600 千米^[18]。本文对 2 种模型进行对比, 发现 RNN 的预测效果明显不如 LSTM 模型, 无论是路径预测还是对单个因素的预测均不是很理想。本文基于‘Marge’号台风, 测试了这 2 个神经网络的模型, 从两种模型的误差对比发现, LSTM 模型预测准确率明显要高于 RNN 这种传统的模型, 所以本文选用 LSTM 神经网络来预测台风移动路径。

2. 对原始数据进行了预处理。训练模型所用数据主要是来源于美国 JTWC

（美国联合台风预警中心），将 2001 年~2005 年这 5 年所有成形的台风路径数据下载后，使用 Matlab 进行初步处理，形成数据文件 ‘track.mat。还有三部分场数据，分别为海表面温度数据、位势高度数据和压强数据。共计 141 个台风的数据集分别进行预处理，包括数据噪声的去除、缺失值的填充等。

3. 提出一种结合深度学习的台风路径预测方法。初始方案采用 LSTM 模型来预测台风移动的坐标，并用实验进行了测试，结果显示预测精度不高，于是提出了一种 LSTM 结合多元回归的预测模型。该模型首先对台风发生的当前经纬度，当前海域的温度，湿度比，压强采用 LSTM 模型进行预测，在此基础上，采用多元拟合回归模型实现了对台风路径的预测，实验结果表明该方法具有较好的预测精度。

4. 台风等级播报预测方法研究。台风的等级分为 12 小时警戒线和 24 小时警戒线，通过经验公式观察发现，位于南海部分的台风大多会进入 12 小时警戒线，而西北太平洋的基本会只停留在 24 小时警戒线，这场景类似于 SVM 支持向量机对坐标系的坐标点的分类处理。本文采用 SVM 对台风的坐标点进行预测，并和朴素贝叶斯的分类效果进行了对比，结果显示基于 SVM 的分类效果明显由于朴素贝叶斯方法分类效果。

5. 结合本文提出的台风移动路径预测方法和台风等级播报预测方法，设计实现了一个简单的台风移动路径预测系统，能够直观显示台风的移动路径和相关参数，系统具有数据的加载、模型的训练与预测、结果展示等功能。

1.4 论文组织结构

本文主要采用机器学习和深度学习方法实现了对台风移动路径的预测和台风等级的预测。主要包括 LSTM、多元拟合回归、SVM 支持向量机等。论文对此进行了详细的描述，具体内容安排如下：

第一章绪论，主要介绍了课题的研究背景及国内外相关研究现状，并对本文的研究内容和内容安排进行了详细阐述。

第二章主要是相关技术简介，对于所用的深度神经网络 RNN、LSTM、机器学习算法 SVM、以及编程工具等进行了简单的介绍。

第三章基于深度学习的台风路径预测方法，重点介绍了所提方法的总体框架、数据的预处理、基于 LSTM 的单要素预测模型、基于多元回归的台风路径预测技

术，以及所提模型的预测测试结果。

第四章台风等级播报预测，根据由卫星传过来的中心台风风眼坐标，通过采用 SVM 来实现对台风等级预测播报，区分该台风到底是 24 小时警戒线还是 48 小时警戒线。

第五章系统分析和设计，重点介绍了台风路径预测测试系统的设计与实现过程，包括系统的需求分析、功能设计及相关模块的具体实现，以及相关算法及系统的测试等内容。

1.5 本章小结

本章首先介绍了课题的研究背景和意义，并对国内外研究现状进行了简单的解释和分析，之后详细描述了本文的主要研究内容，最后介绍了论文的组织结构安排。

2 相关技术简介

2.1 机器学习

机器学习广泛用于计算机视觉,自然语言处理,语音识别^[18]。所用的 Sklearn 也是广受大家喜爱的机器学习的框架, sklearn 中有封装好的数据集,封装好的各种机器学习的框架,在监督学习,无监督学习,半监督学习,在数据预测等领域效果比传统的方法好。机器学习框架集成在 sklearn 中,我们这个系统使用 keras 框架,机器学习的部分采用 sklearn 的包,编译器采用 pycharm。其中 keras 的框架是一个高度模块化的集成框架,简单易懂,易于上手。支持 GPU 和 CPU 的各种运算机制^[19]。Keras.model keras.layer,keras.initializations 等都是 keras 框架下的集成好的开发资源,本次开发采用 keras 框架,开发效率较高^[20],方便梳理逻辑,供用户使用,本次的多元回归采用的就是 sklearn 的包。开发轻松,逻辑便于理解。其中的类似于封装好的线性回归(Linear regression),SVM(支持分类向量机)。框架 sklearn 的诸多集成包使得对大量数据的处理如虎添翼^[21],面对对庞大的数据,我们的机器学习的确是一个强大的工具,尤其对于本次的气象学数据,我们将 sklearn 中的多元拟合回归预测出他的迭代方程。进而求出所有的点坐标。

机器学习运用数学知识加上强大的语言 python 在各个领域展现着强大的生命力。其中监督学习,是数据集给定的标签和预测的值进行比对,不断调整预测模型。监督学习可应用的场景很多例如垃圾邮件的分类,和对图像识别的分类,人脸识别(face recognition)。和监督学习不一样的是‘非监督学习’,非监督学习的无人工的标签,常见的应用场景比如聚类^[22]等等。介于监督学习和非监督学习的是‘半监督学习’,比如拉普拉斯支持向量机。常见的机器学习的算法大致有线性拟合回归算法,决策树学习用于数据构建决策模型^[23],更加有效的制定回归问题。编辑器选用 pycharm,它是一种高效率的可靠编辑器,是 JetBrains 提出的开发工具,具有代码自动补全,代码高亮,等多个功能。快捷键及其的丰富,大大提高了编码的效率使得开发流程简洁易懂。

深度学习是中机器学习的一个重要分支,在预测、回归、分类等领域有着诸多应用。在预测本次台风路径预测系统的时候发挥了重大的作用。深度学习主要是深度神经网络数据表征的一种学习方法^[24],早在 VGG16 提出的时候就已经证

明了, 增加网络的深度就可以来提高网络 的性能, 我们在本次的毕业设计中 也遵循了这个原理。总体来说, 深度神经网络运用了线性代数^[25], 最优化理论, 逼近论, 高等数学, 微积分, 高等代数, 离散数学, 数据结构等计算机和数学 相关的知识, 其中的正则化, 反向传播, dropout 等技术使得神经网络的容错率 和效果更好。

2.2 循环神经网络

RNN 的本质就是一个循环的。RNN 不同于普通的神经网络的地方在于层和 层之间也有权值链接, RNN 是一种传统的序列数据的构建模型, 其计算公式见 公式 (2.1) - 公式 (2.4)。

$$H_t = f(Wxh * X_t + Whh * H_t - 1) \quad (2.1)$$

$$Y_t = g(WhyH_t) \quad (2.2)$$

$$O_t = g(V \cdot S_t) \quad (2.3)$$

$$S_t = f(U \cdot X_t + W \cdot S_{t-1}) \quad (2.4)$$

下面是 RNN 的必要的推导公式, 其中 $g()$ 函数为激活函数。

$$\begin{aligned} O_t &= g(V \cdot S_t) = g(Vf(U \cdot X_t + W \cdot S_{t-1})) \\ &= g(Vf(U \cdot X_t + W \cdot f(U \cdot X_{t-1} + W \cdot S_{t-2}))) \\ &= g(Vf(U \cdot X_t + W \cdot f(U \cdot X_{t-1} + W \cdot f(U \cdot X_{t-2} + \dots)))) \end{aligned} \quad (2.5)$$

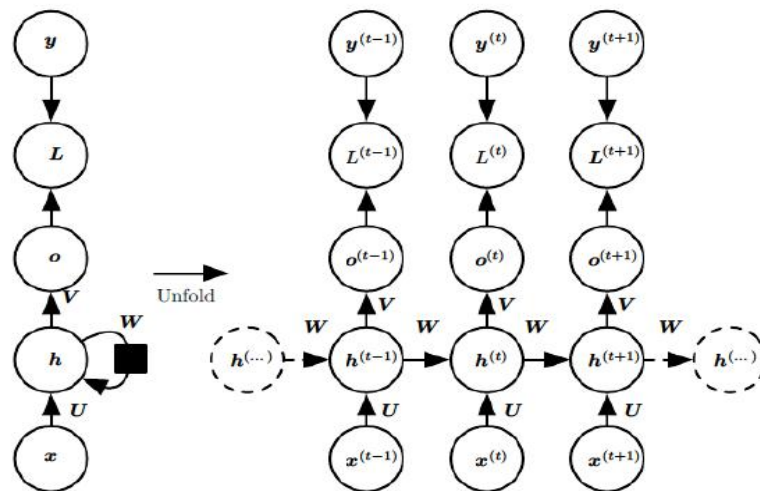


图 2.3 RNN 网络的展开

图 2.3 右侧是展开神经网络的样子, 每一个隐藏层的神经元的均有权重相连接。其中的 RNN 的训练法是 BPIT(back propagation through time)但是他的本质还是前向传播的神经网络^[26], 但还是循环神经网络需要根据时间来进行反向传播, 可

以不难看出, BITP 的本质还是前向传播^[27]。对于这种神经网络大部分采用的是 ReLU 的激活函数。

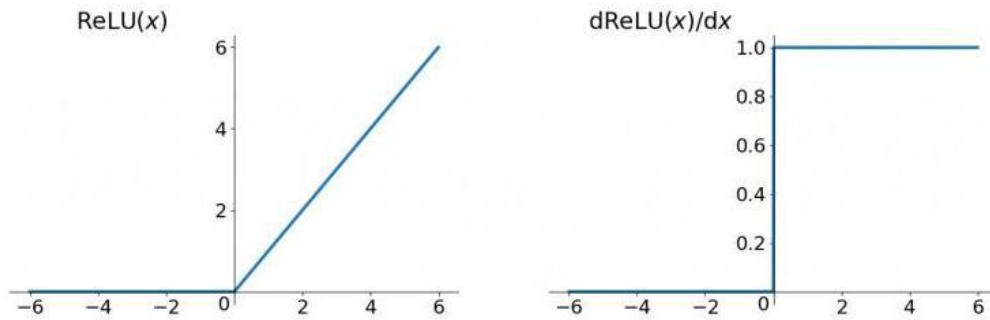


图 2.4 激活函数

RNN 的应用众多例如序列化数据的有效处理, 语音识别, 还有音乐生成, 自动写诗, DNA 遗传物质分析视频活动识别^[28], 对于句子的识别, 自然语言处理的任务名称的识别。

2.3 LSTM 网络

本次毕业设计主要采用的神经网络为 LSTM, 作为传统的 RNN 的变形^[31], 能够学习长期的以来关系, 可以长期学习依赖的信息, 在很多的记忆问题处理序列化的问题上, LSTM 上在 1997 年被提出, 这个神经网络被许多人进行改进和完善, 细胞(cell)的状态对神经网络尤为重要, 和用于解决传统的 RNN 的梯度消失和梯度爆炸的弊端, 它主要用于时间序列的预测, 其中细胞的状态就在这个链子上运行^[32, 33], 信息很容易保持不变可以将一堆时间预测相关的序列送入网络, 来达到预测的效果, 这个神经网络的删除和保留信息的功能主要是靠‘门’来实现的。‘门’主要有一个激励函数 Sigmoid 函数来实现的, 该决定的因素被称为‘忘记门’ (forget gate)^[34-36], 1 代表着保留, 0 代表着舍去, 0-1 之间代表着其中一部分舍去, 一部分保留下来。它是一种时间递归的神经网络。其中长短期记忆神经网络内部会有许多神经元, 一个循环神经网络, 可以被认为是偶同一个网络的副本, 每一个得出的消息传递给后一个神经元。

我们对于每个海域的风向影响因素, 我们均采用 LSTM 进行对各因素进行预测, 进而实现对台风路径的预测。LSTM 针对时间序列的影响因素进行预测, 所以其中的一个有效的预测场景就是航班的客流量的预测。可以对机场的客流量进行合理的预测。还有一个重要的应用场景, 就是股市的预测。比如针对沪深股市 300 最高价格, 就是依赖前 n 天的价格, 搭建好预测股市的神经网络, 来预测当

前的当天的最高价格^[40, 41]。

除此之外, LSTM 在 NLP 部分也有着广泛的应用, 分割数据集, 编译模型, 评估模型, 保存模型。还可以进行情感分析, 在我们或得一个文本数据集的时候, 可以利用 Tokenizer 进行矢量化处理, 将其送入 LSTM 中可恶意研究出本篇文本的感情的倾向。LSTM 系统在问答系统中也大放光彩, 如果采用固定长度的 LSTM, 那么可以对问题进行切片或者进行补充。

2.4 foilum

此项技术是我们图形界面以及展示台风路径的主要手段, 本次“基于深度学习的台风路径的研究”主要使用这个 python 自带的库, 将其在官网还是哪个安装之后, 我们对其进行设置使其能都对我们机器学习预测出的路径进行显示, 进而有效的得出台风坐标, 进而得到登陆点坐标。对这个系统, 我们录入数据之后, 注意是在联网的情况下自动调出这个演示系统。显示的形式是以 html 文件的形式。

2.5 Python-tkinter

近几年由于机器学习的大火, Python 语言已经一跃成为当之无愧的最火的语言。Python 语言的强大之处在于诸多的机器学习, 深度学习等框架均集成在 python 的包中, 许多的网络框架均可以调用 python 的包。我们采用这个技术进行数据录入系统的实现, 本次数据录入系统界面采用此技术编写。

为了方便观察海域的时间, 还在上面调取了系统时间, 方便用户观察当前时间和台风路径的坐标时间, 来制定自己的出海计划。本次的系统还是用了进度条功能, 显示出数据加载和训练, 方便用户进行根据自己的海域的特点进行训练达到个性化设置。针对五个因素, 经纬度, 温度, 湿度比, 压强, 我们将这些因素带入系统中, 把它们带入训练好的模型, 点击 Start to predict, 等 10 分钟, 便可以得到这个 html 文件。在此系统上鼠标点击便可标记处各个经纬度坐标, 方便用户制定自己的计划路线, 来回避台风可能的方向。

其中用到了 Matlab 相关的知识, 由于传进来的参数为 .mat 的格式, 我们又采用了 matlab 的数据处理的方法, 对数据进行了简单的预处理, 去除了 nan 等无效数据。Matlab 是一种解决数学问题功能强大的语言, matlab 是一种强大的软件。他在数学类科技应用类的数值计算数一数二。Matlab 可以进行图像处理, 信

号处理,数据分析等领域。给出的 track.mat sst.mat hgt.mat shum.mat, 其中的 track 是多维数组, 我们将这个高维数组进行分段处理, 将 nan 一个个去除并将这些 matlab 的数据中的有效信息提取出来, 组成初始向量。其中 python 中高效的将 .mat 格式数据提取出来。对于 matlab 最擅长解决的就是矩阵的处理, 拥有强大的三维运算功能。可用于可视化领域, 工程计算领域, 科学计算领域。

2.6 大气动力学

大气动力学是物理学中的流体力学的一个重要分支, 是动力气象学的一个重要分支。三圈环流(three cell circulation)并且应用了大气辐射遥感的知识, 对大气的走向, 对台风的成因进行具体推导, 得到的温度, 湿度比, 压强对台风走向矢量具有重要影响^[37-39]。

本次毕业设计先采用跨尺度, 多像简化模型等等物理学分析, 我们发现例如对台风的路径, 影响台风的走向我们采用大气动力学方程(dynamic equations of atmosphere motion)如公式(2.6)-公式(2.7)

$$gz + \ell / \Re + V^2 / 2 = const \quad (2.6)$$

$$pgz + p + pV^2 / 2 = const \quad (2.7)$$

$$z + p / p_0 g + V^2 / 2g = const \quad (2.8)$$

通过这个公式来分析气象学的诸多因素中较为主要的影响因素, 再通过地球物理学在进一步公式推导得到走向矢量的影响因素^[40-42]。

2.7 本章小结

本章主要对本文所应用的各种技术进行了简单的介绍, 并对 RNN 模型和 LSTM 模型进行了对比分析, 最终选择了更好的深度学习模型 LSTM 来进行预测, 之后简单介绍了系统的实现技术及大气动力学方程。

3 基于深度学习的台风路径预测方法

由于传统的台风路径预测主要依赖数值预测方法，其精度受到一定的限制，为此本文提出了一种基于深度学习的预测方法，本章就从总体处理框架、数据预处理、基于 LSTM 的海洋要素预测和多元回归的台风移动路径预测等几个方面对提出的方法进行详细介绍。

3.1 总体框架

本文通过对传统气象动力学预测模型的研究发现，台风移动路径和当前台风移动路径、当前的海洋温度、比湿、位势高度等海洋要素密切相关。为此提出了结合深度学习的多元回归台风路径预测方法，其总体框架如图 3.1 所示。

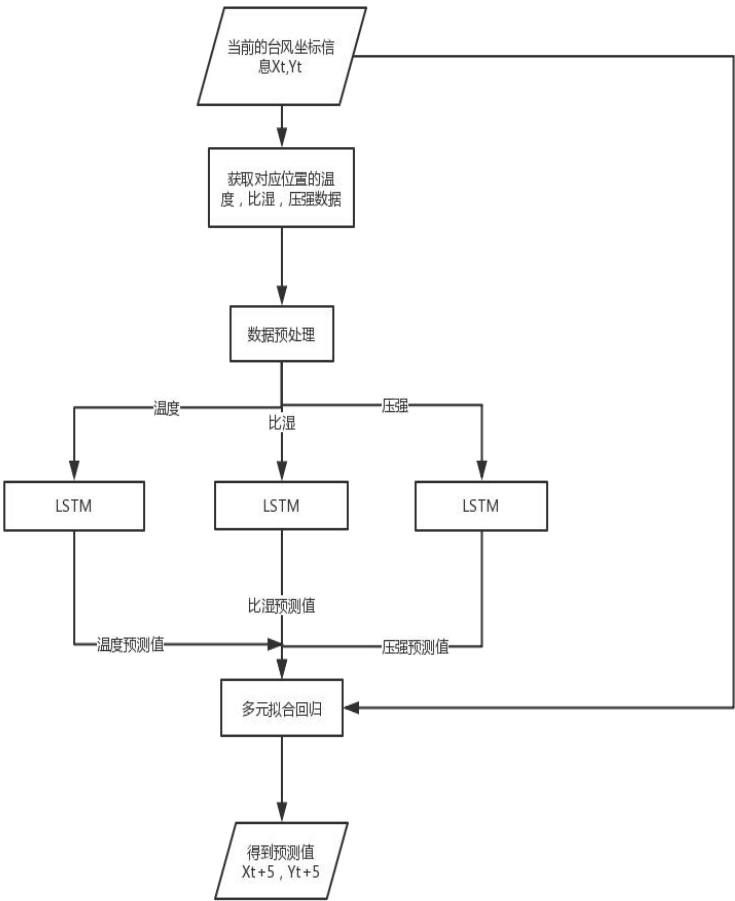


图 3.1 台风路径预测总体处理框架

从图 3.1 可以看出，当得到台风当前的位置信息后，获取相应的海洋温度、比湿、位势高度等海洋要素历史信息，采用历史海洋要素信息训练基于 LSTM 的

海洋时序要素数据预测模型。LSTM 是长短期记忆网络，是一种时间循环神经网络，非常适合对于与时间预测有关的情形。LSTM 已经在科技领域有了多种应用，如基于 LSTM 的系统可以学习翻译语言、控制机器人、图像分析、文档摘要、语音识别图像识别、手写识别、控制聊天机器人、预测疾病、点击率和股票、合成音乐等等任务^[43, 44]。利用训练好的 STML 模型得到相应的海洋温度、比湿、位势高度的预测值，然后基于这些预测值和当前台风位置信息，采用多元线性回归模型来拟合出新的台风移动路径的预测信息。

3.2 基于 LSTM 的单要素预测

3.2.1 数据预处理

在得到台风的初始位置坐标(x_0, y_0)后，获取相应位置的海洋环境要素信息，主要有海洋表面温度 SST、位势高度数据 HGT 和比湿数据，由于数据存在噪声，并且格式不满足要求，需要对数据进行预处理。首先将.mat 的数据读入，并进行预处理，将数据集中的具有无效值“nan”的数据剔除。具体数据及处理如下：

(1) 通过 matlab 将 2001 年到 2005 年南中国海的台风路径数据简单的出路一下，形成数据文件‘track.mat’，这五年的所有共计 141 场台风路径如下图 1 所示：将 2001 年~2005 年的路径数据下载下来，并通过 matlab 初步处理，形成数据文件‘track.mat’，这五年的所有(141 场)台风路径如显示如下图 3.2。

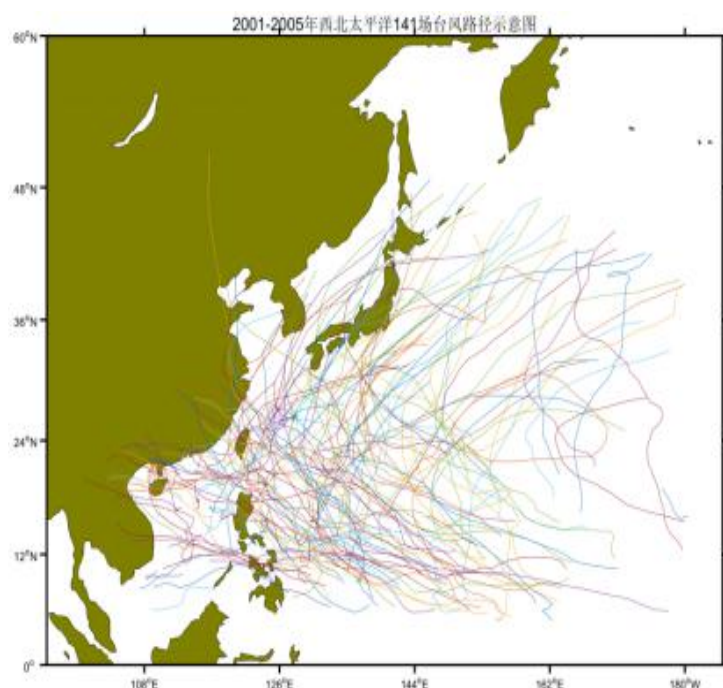


图 3.2 台风路径

对于结构体数组 `track.mat` 浮点型变量，前 141 个变量是台风的编号。文字性变量对应着 141 场台风的名字。结构数组 `track.tra` 有 141 个矩阵，4 个维度分别代表时间、维度、经度和最低气压。

(2) 海表面温度数据

SST 数据文件 “`sst.mat`” 有 4 个变量，分别是：

`time`: 2001 年 1 月 1 日 0 时至 2005 年 12 月 31 日 18 时的时间步长为 6 小时的共 7304 个时刻，是文字型数据，如 “2001010100”、“2001010106”。

`latitude`: 纬度的分布 0° ~ 60° N，步长 0.5° ，是 121×1 的数组。

`longitude`: 经度的分布 100° E~ 180° E，步长 0.5° ，是 161×1 的数组。

`sst01_05`: 海表面温度，是 $121 \times 161 \times 7304$ 的数组，对应纬度、经度和时间。

选取 2005 年 1 月 1 日 0 时，西北太平洋海表面温度作为样图如图 3.3。

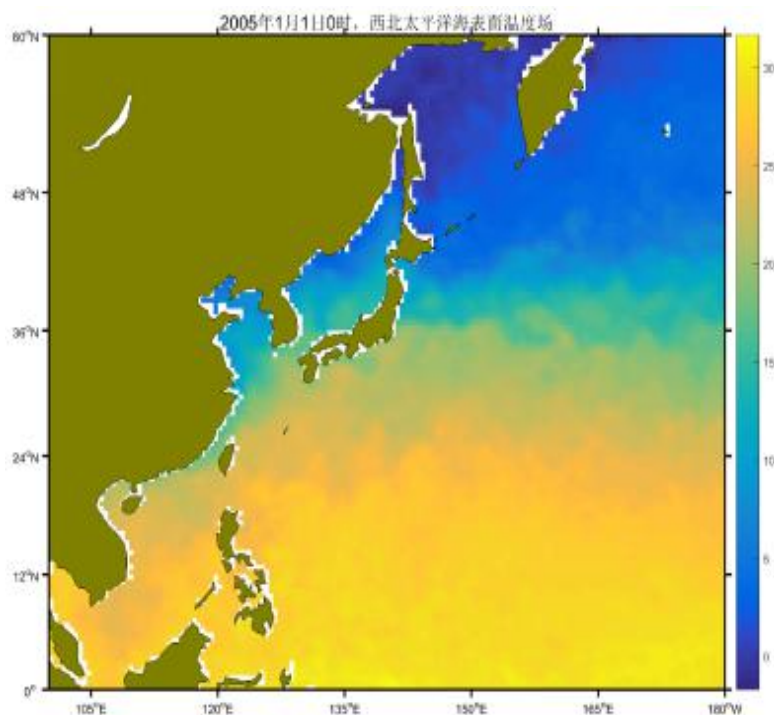


图 3.3 温度场数据

(3) 比湿数据

“`shum.mat`” 比湿数据文件变量如下：

`time`: 表示 2001 年 1 月 1 日 0 时至 2005 年 12 月 31 日 18 时的时间步长为 6 小时的共 7304 个时刻，是文字型数据，如 “2001010100”、“2001010106”。

`latitude`: 表示纬度的分布 0° ~ 60° N，步长 0.5° ，是 121×1 的数组。

`Longitude`: 表示经度的分布 100° E~ 180° E，步长 0.5° ，是 161×1 的数组

shum3: 表示 300hPa 比湿数据, 是 $121 \times 161 \times 7304$ 的数组, 对应纬度、经度和时间。

选取 2005 年 1 月 1 日 0 时, 西北太平洋 500hPa 比湿分布场作为样图。如图 3.4 所示。

(4) 位势高

位势高“hgt.mat”有 6 个变量, 分别是:

time: 表示 2001 年 1 月 1 日 0 时至 2005 年 12 月 31 日 18 时的时间步长为 6 小时的共 7304 个时刻, 是文字型数据, 如 “2001010100”、“2001010106”。

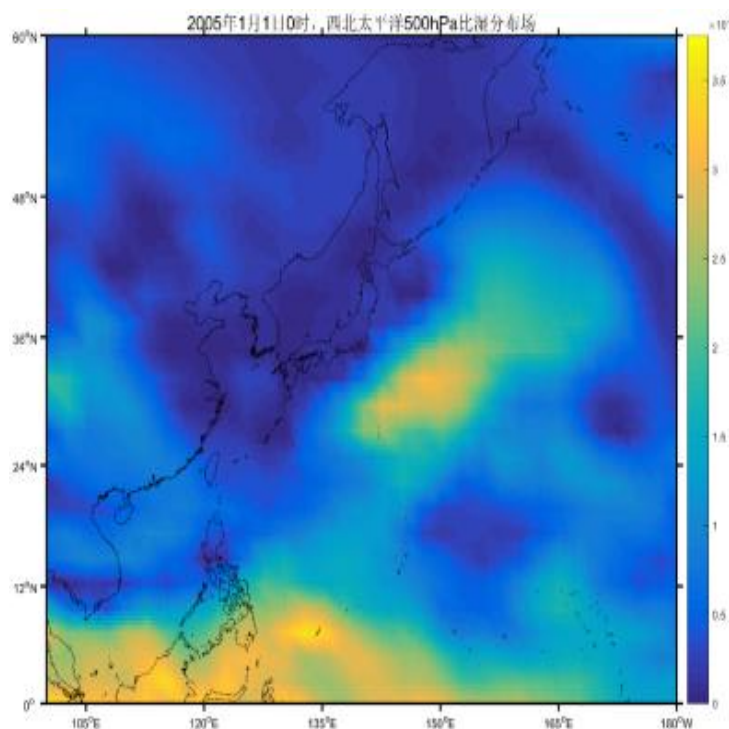


图 3.4 比湿数据场

latitude: 表示纬度的分布 $0^{\circ} \sim 60^{\circ} \text{ N}$, 步长 0.5° , 是 121×1 的数组。

longitude: 表示经度的分布 $100^{\circ} \text{ E} \sim 180^{\circ} \text{ E}$, 步长 0.5° , 是 161×1 的数组

hgt2: 表示 200hPa 位势高度数据, 是 $121 \times 161 \times 7304$ 的数组, 对应纬度、经度和时间。

选取 2005 年 1 月 1 日 0 时, 西北太平洋 500hPa 位势高度场作为样图。如下图 3.5 所示。

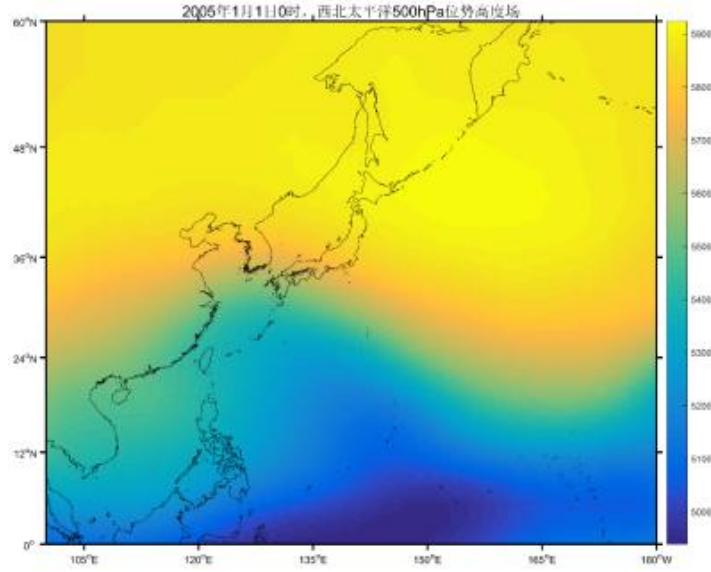


图 3.5 位势高场

在细致分析之后,对这个路径数据的处理大致过程为采用 `scipy.io` 对 `matlab` 数据集进行读入,分别将 `track.mat`、`sst.mat`、`hgt.mat`、`shum.mat` 进行读入,将所有数据中的 `NAN` 进行剔除,并使用 `Loadmat` 函数返回的是一个字典,在 140 场台风中用构造出如式 (3.1) 和式 (3.2) 所示向量,以及标签: X_t , Y_t 。

$$V01 = (x_t, sst, shum, hgt) \quad (3.1)$$

$$V02 = (y_t, sst, shum, hgt) \quad (3.2)$$

3.2.2 LSTM 单要素预测模型

经过深入研究本文采用深度神经网络 LSTM(Long Short Term Memory),对台风路径预测相关的三个影响因素海表面温度、比湿、位势高度进行预测。LSTM 是当下最流行的 RNN 网络,其模型可以表示为式 (3.3)。

$$H_t = \text{sigmoid}(W_{xh} * X_t + W_{hh} * H_{t-1}) \quad (3.3)$$

图 3.6 为 LSTM 网络的原理图。

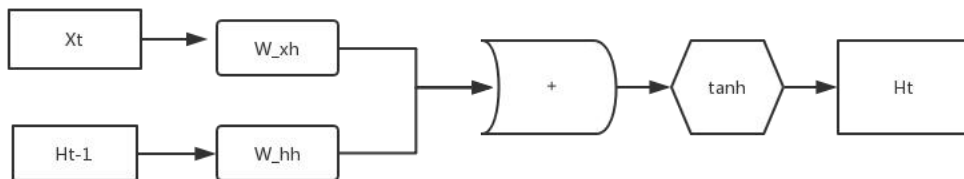


图 3.6 网络数学原理图

普通的 RNN 会因为梯度消失 (vanishing gradient) 和梯度爆炸 (gradient explosion) 而导致整个网络的崩溃, 而 LSTM 比普通的 RNN 多出来了 3 个控制器 输出控制 输入控制 和忘记控制, LSTM 中的重复模块包含的四层交互神经网络层, 如图 3.7 所示。

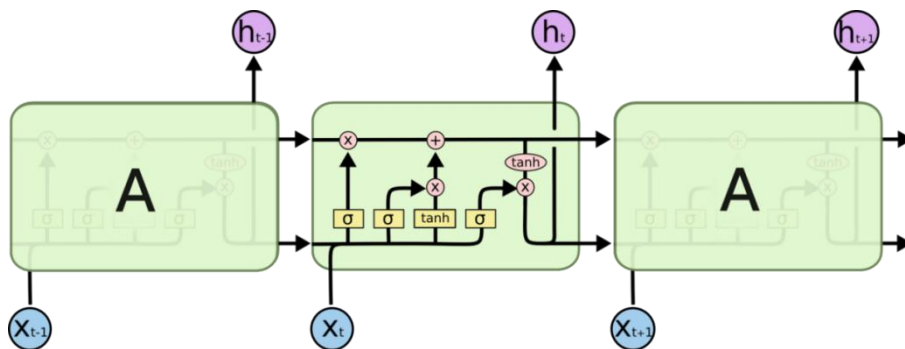


图 3.7 长短期神经网络示意图

长短期记忆神经网络是一种时间循环的神经网络, 擅长处理像时间序列, 相比于 RNN, LSTM 会记住更长的序列, 主要是它加入了一个处理器 Cell, 这个 cell 包含“输入门”, “输出门”, “忘却门”。当一个信息进入 LSTM 神经网络中的时候, 根据股则会判断其是否有用, 只有能按照算法流程的数据才会被留下, 而忘却门用于去忘却淘汰的数据, 一进二出的原理, 可以再反复运算下解决序列预测的问题。RNN 在时间序列太长的时候, 会导致训练效果很差 LSTM 会选择性的忘却一些不重要的事情, 拥有一个自我衡量的机制。

如图 3.8 所示, 每一个黄色的框框代表着一个前馈神经网络 (feedforward neural network)。LSTM cell 里面的 num_units 隐藏神经元, cell 的状态 state 是一组向量, 上一次的状态 h_{t-1} ^[45]。LSTM 会提共一种‘控制参数’的机制, 通过参数来控制什么样的信息会被传递下去, 什么信息会被舍去。

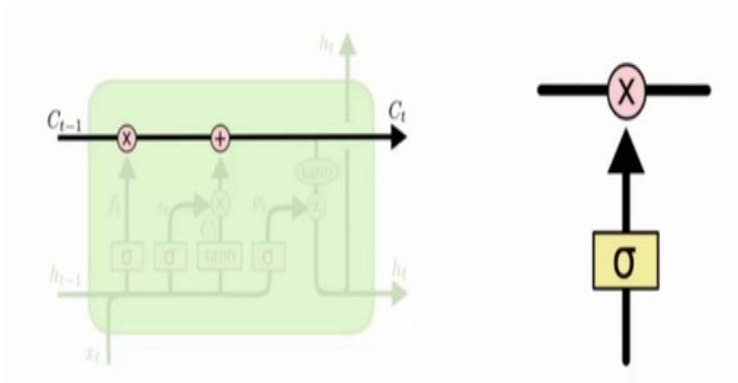


图 3.8 控制参数

如图 3.8，门是一种让信息选择通过的方法，1 代表保留，0 代表全部舍弃，其中 0-1 之间代表部分保留部分舍弃，其中 Sigmoid 函数负责来衡量输出值

(1) The Forget Gate

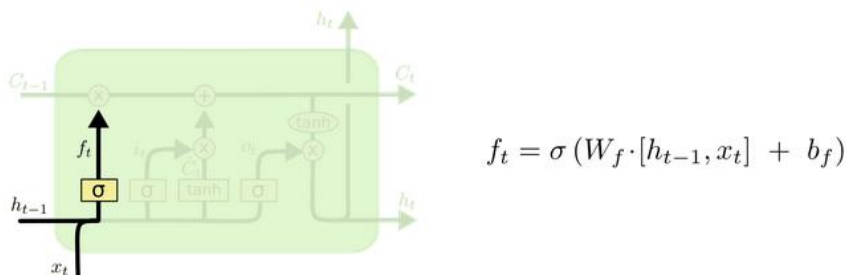


图 3.9 遗忘门

这个就是神经网络的忘却门(the forget layer),通俗的来说，神经网络为了记住更多的序列，会选择性忘却(selective forget).比如一个物体运动的视频，如果物体运动的很慢，那么只需要神经网络记住开头，中间，结尾即可。而 RNN 则相反，他会将所有信息，无论是否有用均记住。

(2) The Current State

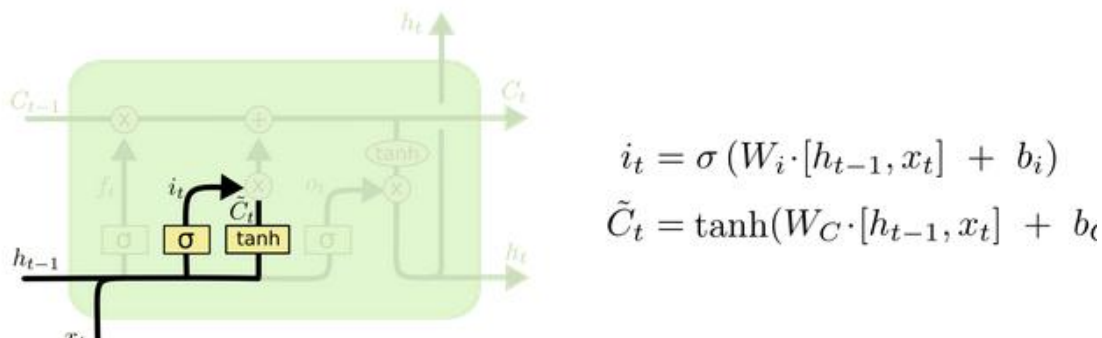


图 3.10 当前状态

这个实际上是和 ForgetGate 联合起来使用，来决定到底是将前一帧的信息传下去，还是将自己的信息传递下去，自我衡量(Measure Myself)和选择忘记(Choose to Forget)实际上是 LSTM 的最大的特点。

(3) Output Layer

运行一个 sigmoid 层，它决定要输出单元格状态的哪些部分。然后将单元格状态放入 tanh(将值推到- 1 和 1 之间)。

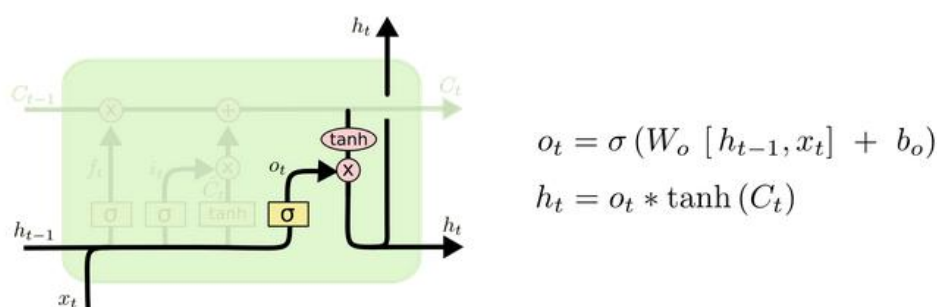


图 3.11 输出层

3.2.3 模型训练与预测

将温度 sst.mat 这个三维数组 121X161X7304 个数据输入 LSTM 神经网络，其中激活函数采用 tanh，损失函数采用均方差函数 MSE。优化器刚开始选用随机梯度下降法 SGD(Stochastic Gradient Descent)，这种优化器会将数据分批送入神经网络 NN(Neural Network)，SGD 一次只进行一次更新，没有冗余而且比较快，并且可以新增样本。但是 SGD 及有可能被困在鞍点，通过对比发现了更好的优化器 RMSprop。RMSprop 是 Geoff Hinton 提出的一种自适应学习率方法，当梯度大时，学习率变小，当梯度小时，学习率变大，以确保网络的稳定性。

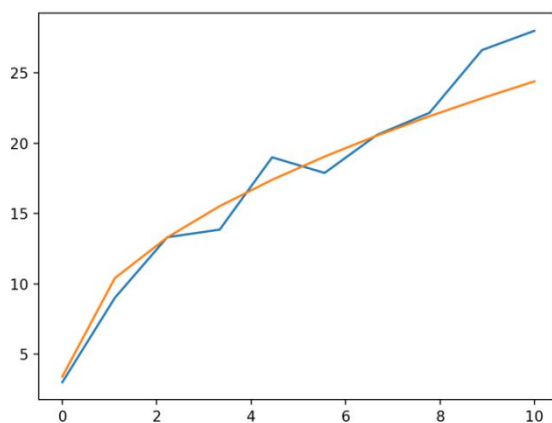


图 3.12 LSTM 温度预测效果

将这些数据送入 LSTM 神经网络中进行练习，再将训练好的模型依次保存在 save_sst_model.pkl、save_hgt_model.pkl、save_shum_model.pkl 之后，我们开始搭建网络，第一层的网络是一层 LSTM，输入的维度是 1，输出的维度是 50，第二层的网络层是 100 个神经元的 LSTM，最后一层是激励函数 tanh。对于已经训练好的模型我们将其分别保存到.pkl 文件中。再调用这个文件对已经输入的数

据进行预测,得到下一个时刻的温度湿度比和压强数据。图 3.12-图 3.14 是 LSTM 神经网络预测的效果比较图。

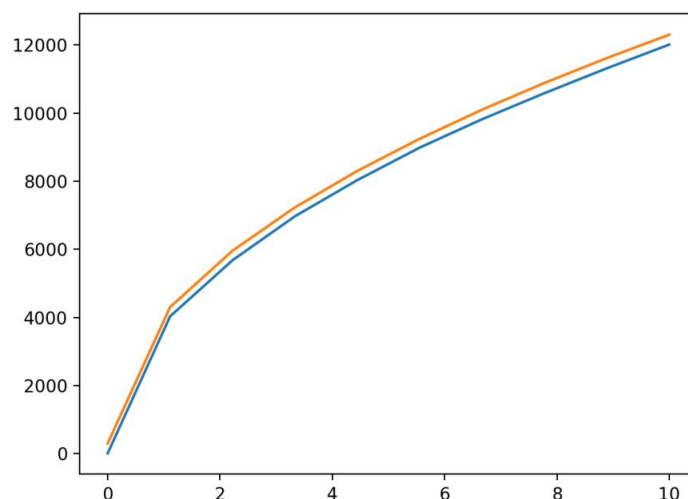


图 3.13 基于 LSTM 的位势高预测比较

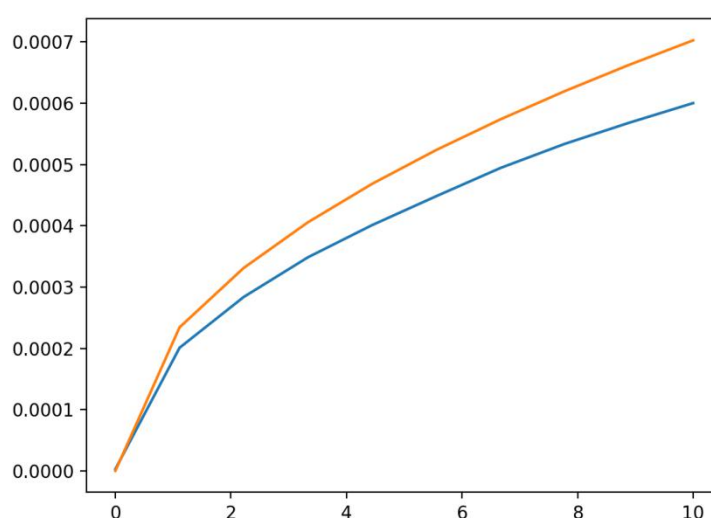


图 3.14 基于 LSTM 的比湿预测比较

3.3 基于多元回归的台风路径预测

开始的思路是将训练好的模型加上中央气象局的所有海域坐标的参数,但是一年的气象数据和海洋数据就非常庞大,并且每个海域的各个物理参数是对台风的走向有影响的,每到一片海域,台风的走向会有诸多气象学参数影响,由于数据集有限,于是放弃对于大气物理推导加上 LSTM 的模型,转而把目光投向

机器学习的算法,既然我们的影响因素众多,可以将主要的参数列出来,抛弃原来的物理学细节,大气动力学的细节,直接进行黑盒训练。在此仅仅讨论温度,湿度比,位势高对台风走向的影响。对于台风走向的因素有诸多因素,我们将大致的大气物理学的公式列出的时候发现诸多求对数的运算以及对其参数求导的物理模型,导致会有大量的运算,多维向量的运算导致大型计算机运算起来也很复杂,自然条件的多变也导致了台风的走向的不确定性。运用上面得到的数据,将其分割训练集(train-set),测试集(test-set)。调用 sklearn 的 LinearRegression 对训练集进行训练。线性回归是有监督学习的一种算法,其中代价函数就是用来描述线性回归模型与正式数据之前的差异。如果完全没有差异,则说明此线性回归模型完全描述数据之前的关系最佳拟合的线性回归模型,就需要使得对应的代价函数最小。

本文采用线性回归对大量数据进行拟合,采用 sklearn 中的 LinearRegression 求出线性迭代方程,具体预测模型如公式(3.4)和公式(3.5)所示。

$$X_{t+5} = f(x_t, sst, shum, hgt) \quad (3.4)$$

$$Y_{t+5} = g(Y_t, shum, sst, hgt) \quad (3.5)$$

进而采用图像化界面 python 的 tkinter 做出一个界面预测系统,有显示当前时间,输入初始向量 $v_0=(x_0, y_0, sst, shum, hgt)$ 进行对台风路径进行预测,最终绘制出一个经纬度草图。再将生成的经纬度数组传入 Arcgis,用这个专业的地理信息软件将其路径绘制出来,来确定登录的位置和地点。

采用递推方式进行训练,使用当前坐标对接下来的坐标进行预测,具体如公式(3.6)和(3.7)所示。

$$X_{t+5} = C_1 + C_2 * X_t + C_3 * shum + C_4 * hgt + C_5 * sst \quad (3.6)$$

$$Y_{t+5} = C_6 + C_7 * Y_t + C_8 * shum + C_9 * hgt + C_{10} * sst \quad (3.7)$$

调用 sklearn 中的 LinearRegression 进行对全体数据的拟合,其中抽象原理如公式(3.8)所示。

$$D = \{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\} \quad (3.8)$$

最后将这些数据进行拟合成如公式(3.9)所示:

$$(w, b) = \arg \min \sum_{i=1}^n (f(X_i) - y_i)^2 \quad (3.9)$$

对其进行求偏导数,得到式(3.10):

$$\partial E(w, b) / \partial w = \sum 2(x^2 * w + (b - y_i) * x) \quad (3.10)$$

令上面 2 个公式等于 0，可求出对应的 w 和 b 。也就是求出对应的 $C1, C2, C3, C4, C5, C6, C7, C8, C9, C10, C11, C12$ 等系数。将训练出来的线性系数存放到 LinearCoef.csv 的文件中。

3.5 预测结果分析

将初始的点台风风眼的坐标带入递推方程，求出下一次的坐标 X_{t+5}, Y_{t+5} 。由于此刻的位置的是五小时之后的情况，并不知晓此刻的温度，湿度比，位势高，于是将这个经纬度的前几个小时的温度，湿度比，位势高带入训练好的 LSTM 预测模型中，将训练好的模型分别保存在 save_hgt_model.pkl, save_sst_model.pkl, save_shum_model.pkl 三个文件中。使用时，将已经训练好的模型调用出来，并将该位置几个小时之前的温度、压强、湿度比数据带入这个模型，得到预测值，进而接着将新的坐标(经纬度)和新的温度，湿度比，位势高，带入这个递推方程，求出每隔 6 小时一次的坐标，然后输入 foilium，在联网的条件下，调用网络的背景图，进行对登陆点坐标进行比对。

对已经运行的模型进行测试，计算预测误差并进行对比，误差分析公式如公式 (3.11) 和式 (3.12)。

$$R = \sum_{i=1}^n Dis((X_i, Y_i), (X_i, Y_i)) / n \quad (3.11)$$

$$Dis((X_i, Y_i), (X_i, Y_i)) = \sqrt{(X_i - X_i)^2 + (Y_i - Y_i)^2} \quad (3.12)$$

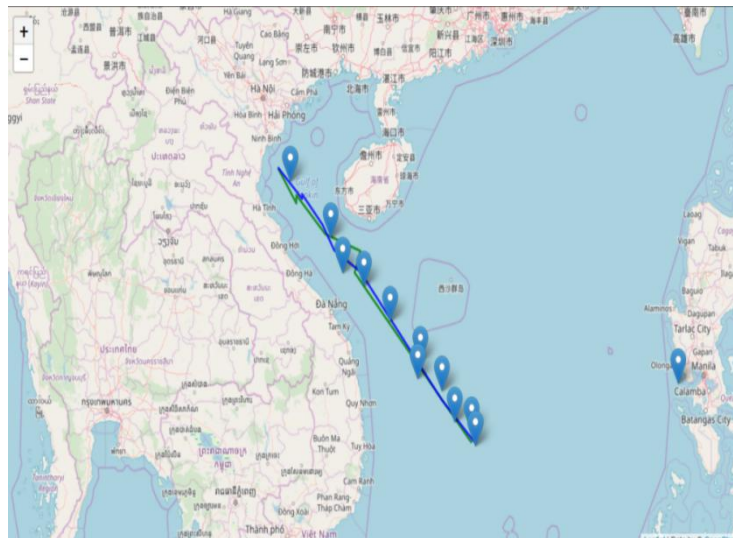


图 3.15 台风路径对比图

图 3.15 是实际台风路径和预测路径之间的对比展示,表 3.1 是不同时间下周期预测误差情况,从中可以看出预测误差不超过 40 公里,具有较好的预测精度。

表 3.1 预测误差表

台风	6 小时误差	12 小时误差	18 小时误差
chebi	15.213	15.37888	16.712
Wukong	19.213333333	19.0234	19.293
Muifa	21.09812	20.1213	21.345
Langka	12.223409	13.22213	12.442
Saomai	40.21334	39.021	39.214
Usagi	37.7212	38.214	38.213
Hato	30.00021	30.213	30.213
Rammasun	31.2131	31.234	30.6001
Mujigae	32.11123	32.1021	32.0123
Haiyan	20.1233	20.2134	20.409
Nepartak	19.213	20.34	20.1
Mirinae	20.213	20.123	20.78222
Maysak	19.0213	20.13	20.1
Soudelor	12.345	15.87	16.901
Fengshen	13.4455	14.67	15.2
Maemi	32.121214	33.112	38.21
Hagupit	45.67091	40.213	49.21
Vongfong	23.213	20.14	26.57
Phanfone	23.2131	26.12	26.14

4 基于 SVM 的台风等级播报

4.1 基于 SVM 进行等级预测

由于台风的分为‘24 小时警戒线’，‘48 小时警戒线’，其中台风 24 小时内可能影响我国内陆,并可能于 24 小时内登陆,进入 24 小时警戒线以后,气象部门会每个小时对台风进行监测,密切注视台风动向,并随时通报台风的实际情况。台风来临时,警报会在 48 小时内发布,紧急警报在 24 小时之内发布,在通报出紧急警报的时候,登陆的省份应该做出应对的方案。

台风警报根据编号热带气旋的强度、影响时间和程度可分为消息、警报和紧急警报三级。根据编号热带气旋的强度,具体名称可分为:热带风暴消息、热带风暴警报、强热带风暴消息、强热带风暴警报、强热带风暴紧急警报、台风消息、台风警报、台风紧急警报。此外,对其他一些灾害性天气根据需要也发布灾害性天气警报;如暴雨警报、强对流天气警报等。

本系统可以提前在路径画出来之前就可以预判是否台风的生成点是否可以到达‘24 小时警戒线’,‘48 小时警戒线’^[51],目前的我们的预报系统就是根据成形坐标之后的运行轨迹来判断的。我们这里采用 SVM 的分类方法可以使得台风的等级播报更加迅速。SVM 是一种对样本的进行二维分类的广义分类器,采用铰链损失函数(Hinge loss function),此 svm 分类向量机在人脸识别(face recognition)等领域有着广泛的应用。支持向量机有着广泛的应用,可以解决软间隔问题,软件个就是当我们给定的数据集(dataset)有噪音的时候我们如果还是用传统的一条线将 2 类严格划分出来,那么就严格了,我们引入松弛样因子,见公式(4.1)。之后采用拉格朗日乘数法,求解最优的解

$$Y_i(Wx_i + b) > 1 + q \quad (4.1)$$

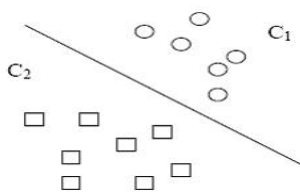


图 4.1 支持向量机

如图 4.1 也叫感知机, 如果一个数据集能用一个直线正确的分开, 那么这个数据集对于给定的数据集(dataset), 我们训练处合适的分割参数 W, b 使得我们分类的结果更加接近真实值, 属于最优化问题。常见的最优化的方法有梯度下降。Svm 最关键的思维就是引入了间隔得到的向量机(vector machine), 得到最优的分类面(optimal classification face)。通过解散复杂的拉格朗日的最优解。除此之外, 还可以解决低纬度的数据不可分的问题, 那么此时可以使用‘核变换’, 可以将低纬度的数据映射到高纬度的数据。假设有 2 个数据集合, 表示见公式 (4.2) 和公式 (4.3)。

$$x = (x_1, x_2, x_3) \quad (4.2)$$

$$y = (y_1, y_2, y_3) \quad (4.3)$$

对于上述三个维度的数据, 此时的 3D 空间已经很难对其进行线性化分了。通过对这个空间中的所有数据采用一个函数进行数据映射, 将其映射到高纬度的空间, 其中那个‘核函数’便是变量 x 变换见式 (4.4)。

$$f(x) = \exp\{-(|x - y|^2) / 2q^2\} \quad (4.4)$$

台风的等级分为 12 小时警戒线和 24 小时警戒线。通过经验公式观察发现, 位于南海部分的台风大多会进入 12 小时警戒线, 而西北太平洋的基本会只停留在 24 小时警戒线, 这个场景很类似于 SVM 支持分类向量机对坐标系的坐标点进行分类的原理, 本文将按照台风的坐标点进行分类。

初始向量是 $(x_t, y_t, shum, sst, hgt)$, 本来可以训练 5 元组, 但是仔细观察就会发现, 降维处理的方法中的纬度间接影响的是温度(sst), 越靠近赤道, 温度越高, 那么可以将 5 维的数据进行降维成 2 维的向量 (X, Y) , 对此二维的向量进行台风等级预测, 送入 SVM 分类器对其进行分类。

对于台风的等级播报, 我们首先构造数据集, 对于 track.mat 这里给定的数据, 我们手动的收集进行构造数据集, 自己上网用标签构造数据集, 划分数据的训练集和测试集针对每个台风的成形的风眼坐标对应的标签就是是 24 小时警戒线还是 48 小时警戒线, 23 小时警戒线便标签是 0, 48 小时警戒线标签是 1。如图 4.2 所示便是 24 小时警戒线的播报。

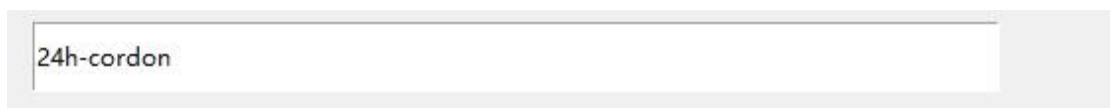


图 4.2 台风等级播报

4.2 朴素贝叶斯和 SVM 对比

在分类预测方面，朴素贝叶斯预测方法同样应用广泛。朴素贝叶斯模型(Naive Bayesian Model)的朴素(Naive)的含义是“很简单很天真”地假设样本特征彼此独立。这个假设现实中基本上不存在，但特征相关性很小的实际情况还是很多的，所以这个模型仍然能够工作得很好。朴素贝叶斯法是基于贝叶斯定理与特征条件独立假设的分类方法。最为广泛的两种分类模型是决策树模型(Decision Tree Model)和朴素贝叶斯模型。朴素贝叶斯分类器发源于古典数学理论，有着坚实的学基础，以及稳定的分类效率。同时，NBC 模型所需估计的参数很少，对缺失数据不太敏感，算法也比较简单。理论上，NBC 模型与其他分类方法相比具有最小的误差率。但是实际上并非总是如此，这是因为 NBC 模型假设属性之间相互独立，这个假设在实际应用中往往是不成立的，这给 NBC 模型的正确分类带来了一定影响。

朴素贝叶斯预测方法进行和 SVM 对比，支持向量机通过寻找分类超平面进而最大化类别间隔实现分类。相比之下，朴素贝叶斯独辟蹊径，通过考虑特征概率来预测分类。将同样的几场台风数据运用朴素贝叶斯进行预测，和 SVM 方法进行对比，凸显出 SVM 的优越性，表 4.1 给出台风等级播报的预测值和真实值对比，从中便可以发现 SVM 方法在预测台风等级上明显优于朴素贝叶斯方法。

表 4.1 等级播报误差表

台风名称	实际等级	预测台风等级(朴素贝叶斯)	预测台风等级(支持分类向量机)
Chebi	24 小时警戒线	24 小时警戒线	24 小时警戒线
WuKong	48 小时警戒线	48 小时警戒线	48 小时警戒线
Hato	48 小时警戒线	24 小时警戒线	48 小时警戒线
Phanfone	48 小时警戒线	48 小时警戒线	48 小时警戒线
Soudelor	24 小时警戒线	24 小时警戒线	24 小时警戒线
Maemi	24 小时警戒线	24 小时警戒线	24 小时警戒线
Fengshen	24 小时警戒线	24 小时警戒线	24 小时警戒线
Rammasun	24 小时警戒线	24 小时警戒线	24 小时警戒线
Hagupit	24 小时警戒线	24 小时警戒线	24 小时警戒线

续表 4.1 等级播报误差表

台风名称	实际等级	预测台风等级(朴素贝叶斯)	预测台风等级(支持分类向量机)
Mirinae	24 小时警戒线	24 小时警戒线	24 小时警戒线
Usagi	24 小时警戒线	48 小时警戒线	24 小时警戒线
Bolaven	24 小时警戒线	24 小时警戒线	24 小时警戒线
Sanba	24 小时警戒线	24 小时警戒线	24 小时警戒线
Jelawat	24 小时警戒线	24 小时警戒线	24 小时警戒线
Ewiniar	24 小时警戒线	24 小时警戒线	24 小时警戒线
Maliksi	48 小时警戒线	48 小时警戒线	48 小时警戒线
Gaemi	48 小时警戒线	48 小时警戒线	48 小时警戒线
Prapiroon	48 小时警戒线	48 小时警戒线	48 小时警戒线
Maria	48 小时警戒线	48 小时警戒线	48 小时警戒线
Son-Tinh	24 小时警戒线	48 小时警戒线	24 小时警戒线
Ampil	48 小时警戒线	48 小时警戒线	48 小时警戒线
Jongdari	24 小时警戒线	48 小时警戒线	24 小时警戒线
Shanshan	24 小时警戒线	48 小时警戒线	24 小时警戒线
Yagi	48 小时警戒线	24 小时警戒线	24 小时警戒线
Leepi	24 小时警戒线	24 小时警戒线	24 小时警戒线
Bebinca	48 小时警戒线	24 小时警戒线	48 小时警戒线
Rumbia	24 小时警戒线	24 小时警戒线	24 小时警戒线
Soulik	48 小时警戒线	24 小时警戒线	48 小时警戒线
Cimaron	48 小时警戒线	48 小时警戒线	48 小时警戒线
Jebi	48 小时警戒线	48 小时警戒线	24 小时警戒线
Mangkhut	48 小时警戒线	48 小时警戒线	48 小时警戒线

4.3 本章小结

本章主要对台风等级播报进行了简单的预测,应用了机器学习的SVM方法进行对坐标进行划分,实现了台风等级的预报,并和朴素贝叶斯方法进行了对比,从实际数据的测试结果来看,SVM方法在预测台风等级上明显优于朴素贝叶斯方法。

5 台风路径预测系统的设计与实现

5.1 系统需求分析

本问的主要目标是研究台风路径的预测方法,在此基础上设计开发一个简单台风路径预测系统,能够根据卫星传来的风眼的坐标(X_t, Y_t),从后台数据库标找出相应的温度、湿度比、位势高等影响因素,从而实现对台风路径的实施预测。系统的主要功能需求如下:

1. 数据录入子系统功能需求

- (1) 显示当前时间
- (2) 显示出台风的图标
- (3) 展示出加载 matlab 数据的路径
- (4) 经纬度
- (5) 温度
- (6) 位势高
- (7) 点击进行预测
- (8) 显示出台风的等级播报,分别出到底会是到达 24 小时警戒线还是 48

小时警戒线

2. 台风路径展示子系统功能需求

- (1) 显示出台风的坐标
- (2) 在演示模式下可以对比出实际和预测的对比路径演示图
- (3) 点击这个系统会显示出这个经纬度

非功能性需求如下:

(1) 需要对台风路径进行合理的演示,得出路径数组,最终能得到登陆点坐标,判断出登录的时间,得到的登陆坐标可以对时间,进行预测。支持上传数据集,路径的.mat 数据温度的.mat, 位势高.mat, 非功能易用性: 系统应提供简洁的界面,用户不需要进行复杂的操作即可实现相应功能。

(2) 易用性: 系统应提供简洁的界面,用户不需要进行复杂的操作即可实现相应功能。

(3) 可靠性: 对于本系统,很可靠

- (4) 可拓展性：对于本系统我留下来诸多接口，可以实现很多额外的功能，对于本次台风的
- (5) 可复用性：所有的模块均可以在其他文件中，代码均可以复用
- 系统的用例图如图 5.1 所示。

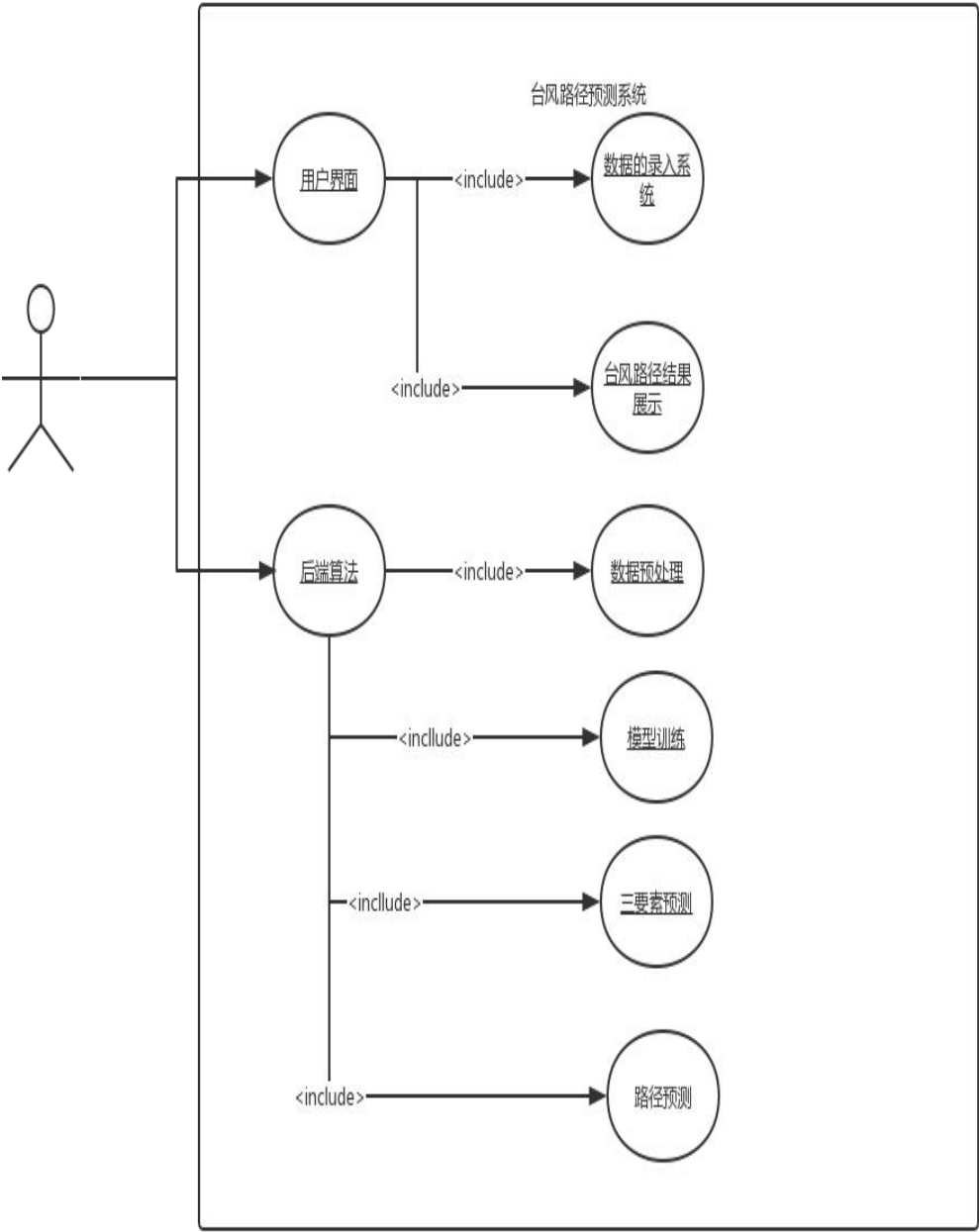


图 5.1 系统用例图

5.2 系统功能设计

如图 5.2 系统总体功能模块图。

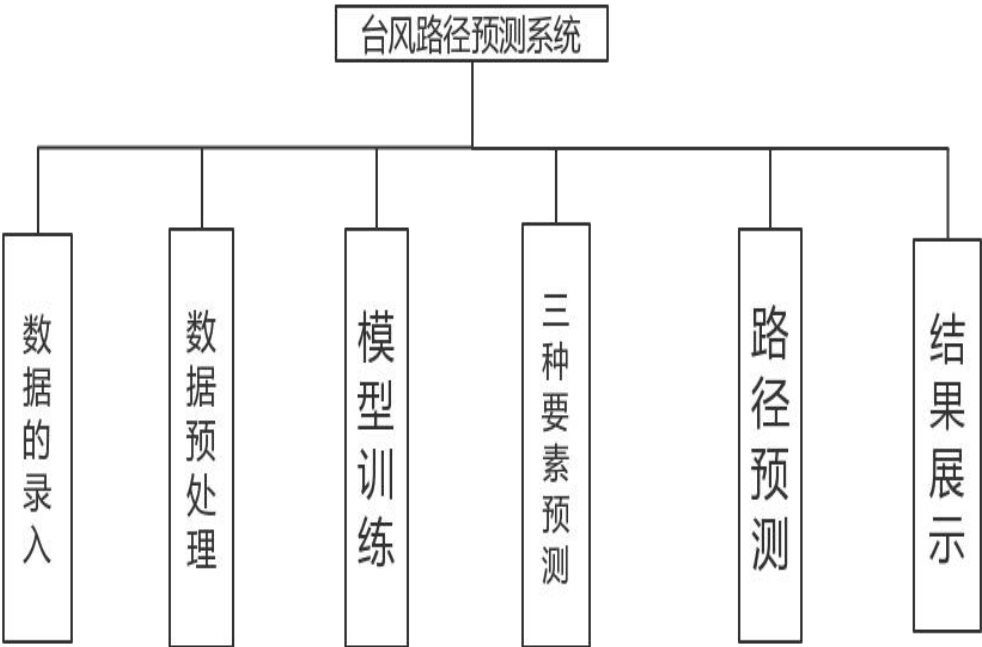


图 5.2 系统总体功能模块图

数据录入模块：该系统支持用户上传台风参数的功能。

数据预处理模块：本系统支持用户自己定义训练模型所用的数据集，用户在前端系统更改文件路径即可。系统在进行训练模型的时候，将用户给出的 141 场台风的路径数据，海洋温度数据，海洋压强数据，海洋位势高数据进行载入，通过 python 对这些气象数据进行预处理。将 nan 等无效数据、台风文字数据、无用的气象物理进行剔除

模型训练模块：针对用户上传的的对应海域的温度数据、比湿数据、位势高数据传入 LSTM 进行模型的训练。选择合适的优化器，学习率等。当训练效果较好时，将模型保存到文件中。

三要素的预测模块：在卫星监测到台风成型的风坐标的时候，将中心坐标经纬度，海域温度，海域的压强，海域的位势高传入系统。调用上个模块所保存的已经训练好的模型，进行对下一个时间戳的三要素进行预测。

路径预测模块: 用户点击预测按钮, 系统将前端的输入的初始数据传给后端, 后端通过对数据的处理, 模型的调用, 开始进行对坐标点的迭代计算, 将生成的经纬度坐标点数组送给前端的台风路径演示系统, 台风路径演示系统将这个经纬度数组连接成一条路径, 展现给用户。该系统还支持用户在系统上对台风的登陆点等关键的坐标点进行标记, 点击页面便可以显示出该点的经纬度坐标。

结果展示模块: 显示出预测后的路径之后, 后端会根据输入的初始的台风的坐标通过支持分类向量机进行分类, 返回该台风是会进入 24 小时警戒线还是 48 小时警戒线。供用户做出相应的对策。

5.3 系统的开发运行环境

硬件 (hardware) :

- (1) CPU: core i5 CPU: Intel(R) 5-3337U CPU @ 1.80GHz
- (2) 运行内存 8G
- (3) 580G 机械硬盘

软件 (software) :

- (1) Python 3.6
- (2) Pycharm 集成开发环境
- (3) Keras 深度学习框架

5.4 系统的实现

5.4.1 系统总体处理流程

用户拿到该系统的时候在对应的位置输入文件路径, 将所用到的路径数据、温度数据、位势高数据, 压强数通过前端的数据录入界面进行上传。后端的系统对传来的数据进行训练, 将训练好的模型保存到.pkl 文件中。之后将卫星监测到台风的风眼坐标的经纬度、海域温度、海域压强, 代入训练好的多元拟合回归模型中, 根据这个递推方程得到下一个时间戳的坐标, 在根据这个坐标到后台的数据库中找到对应的前 6 小时的温度, 湿度比, 位势高带入.pkl 文件已经保存好的神经网络模型中进而得到这个时间戳的温度, 湿度比, 位势高进而接着循环得到坐标点。如下图 5.3 便是系统总体的处理流程图。

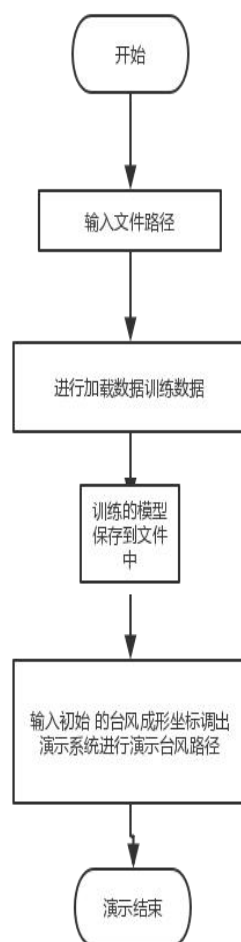


图 5.3 系统总体处理流程图

5.4.2 录入数据模块实现

先采一个时钟记录当前的时间，当台风成形时，输入当前点坐标，初始经纬度坐标，当前点的温度，湿度，位势高，点击 Start to predict 来绘制路线图，下面的显示框会显示出欧诺个户输入的坐标(经纬度)，温度，湿度比，位势高：如图所示本系统具有加载数据，训练数据，预测路径，显示当前的时间的功能。

如图 5.4 所示 2 个进度条是显示加载数据的进度，和数据训练的进度 (training)图标是台风的标志，最上面的是调用 python 自带的 tkinter 界面中的时间，我们调用他的系统时间，帮助渔民等用户来实现的对有效的台风路径进行规避。如图所示我们有数据录入功能针对经纬度 (longitude and latitude) 。

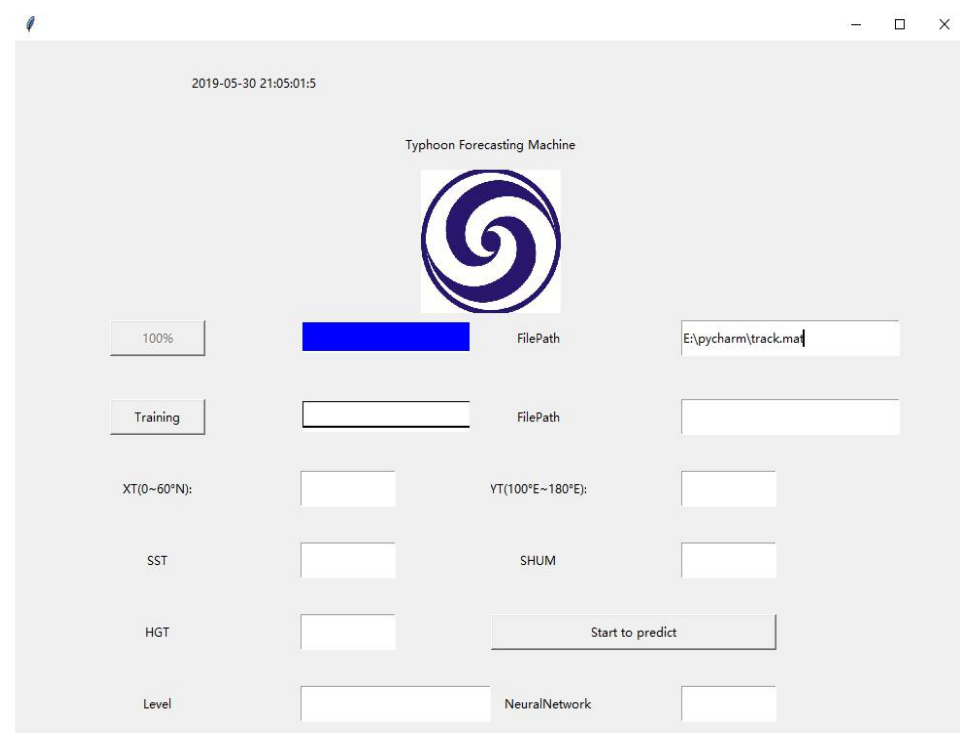


图 5.4 数据录入

5.4.3 台风轨迹预测展示

对上个初始向量的坐标位于鹿儿岛的成型台风进行预测，得到预测轨迹如图 5.5 所示。

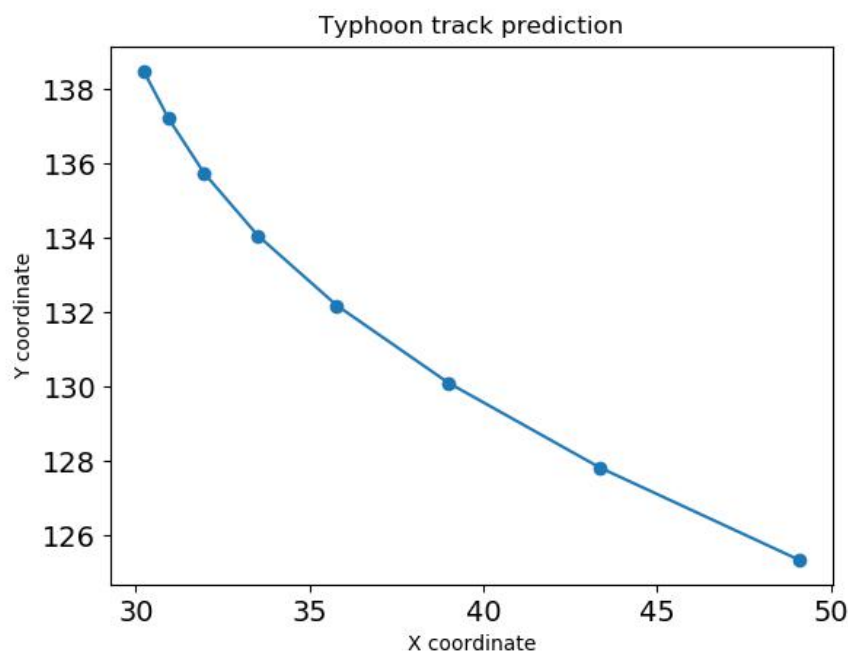


图 5.5 基于坐标的展示

5.4.4 散点图展示

ArcGis 软件是地理信息专业绘图软件，可以显示地形，可以显示海洋的情况，可以支持第三方切片的地图的服务等更多的服务类型，还能提供组织的用户托管的 GIS 服务，具备大数据实时分析，和处理能力。其中的定位器(locator)，是定位参考和定位方法的组合，不同的定位信息有不同的表达方法，在其后台数据库中，其中的 TIN 表面数据模型主要有节点、边、三角形、包面、拓扑。他通过交互图，将数据，人，位置联系起来，我们将上面的经纬度数组传入这个软件绘制大致的散点坐标图，进而达到预测台风路径的效果，如图所示的散点图便是台风路径图，每隔 6 小时便绘制一个散点，如图 5.5 所示。

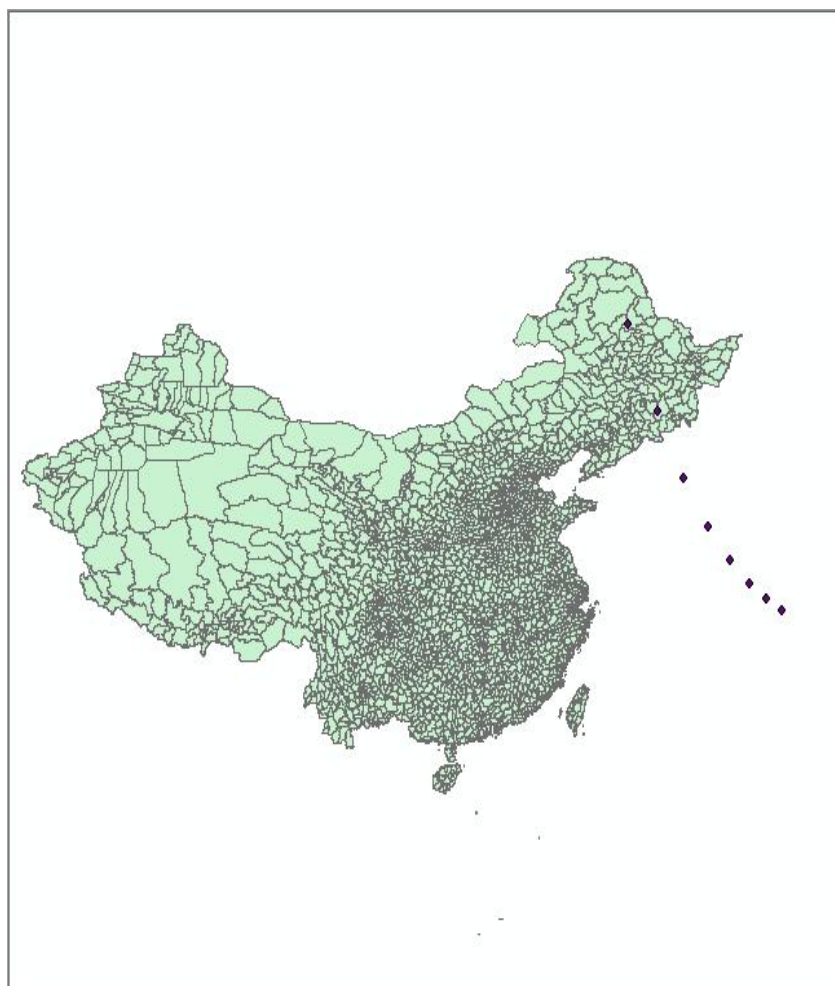


图 5.5 散点图展示

5.5 系统功能测试

调用 folium 并且联网加载西北太平洋的实况图，将等到的经纬度散点打到

上面，显示出台风轨迹，将海岸线矢量化即可得到交点的登录坐标，如图 5.6 所示，就是我做的台风预报系统，由于需要联网所以加载比较慢，因为大量的气象学物理数据本系统从录入数据到显示结果需要 10 分钟，演示起始坐标为冲绳纬度-25.9，经度 126.6，初试温度 20，比湿数据 0.0000213，位势高 10234。本系统还支持在这个系统地球上标记坐标，标记路径等工能，方便用户进行观察路径。将坐标打入系统，可以显示出台风的坐标、轨迹、登陆坐标等。

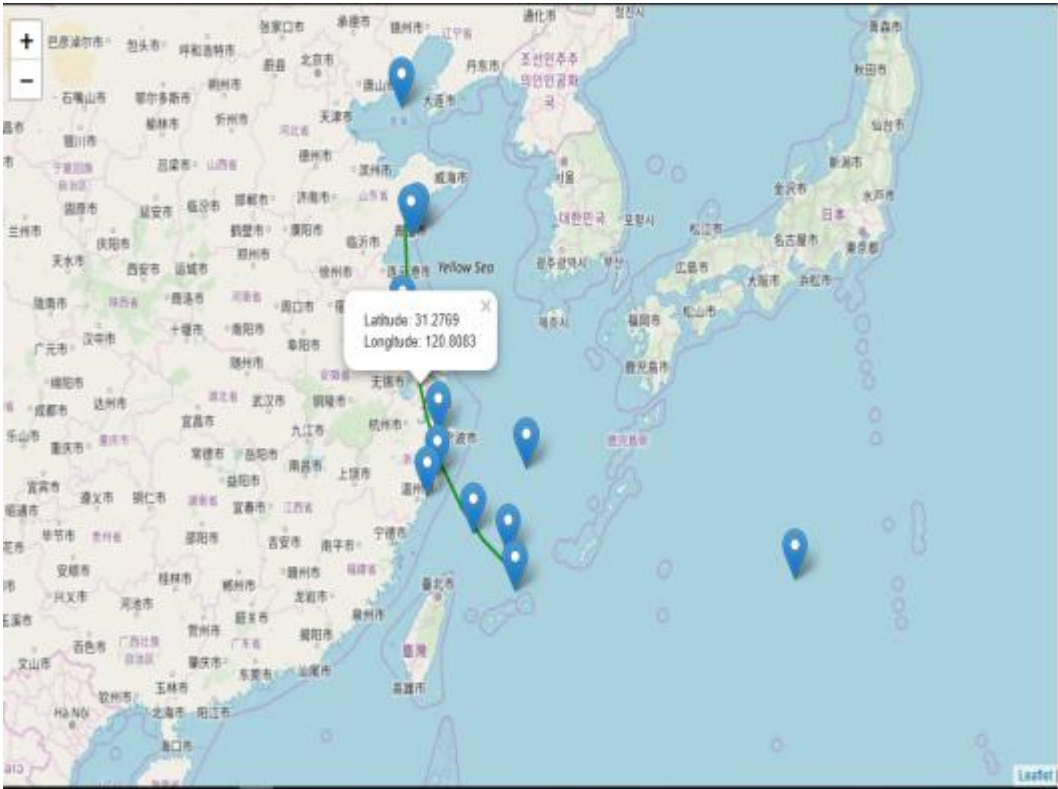


图 5.6 演示系统

众所周知‘山竹’是有史以来破坏力数一数二的台风，对我国东南沿海造成了难以磨灭的伤害，早在台风到来之前，多数擦后市早已清空货架，人民曲终将房屋禁闭，大多数的学校，工厂等均停工、停课、停产，澳门赌场更是停业避风头。深圳更是相爱纳入了停摆状态给大所数的深圳人民造成了巨大生活的不便。如果对这种台风均不进行准确预测，那带来的损失将难以估计。如图 5.7 是对 35 年来号称最强台风山竹号台风的预测，蓝色既是预测台风移动路径，绿色是真实值，对比发现相差不大。对比一下我们从国家中央气象局获得的坐标为：

Location1=[[13,113], [14,112], [17,109], [17.5,109.2], [17.9,108], [18.8,107], [18.6,107], [19.4,106.4]]

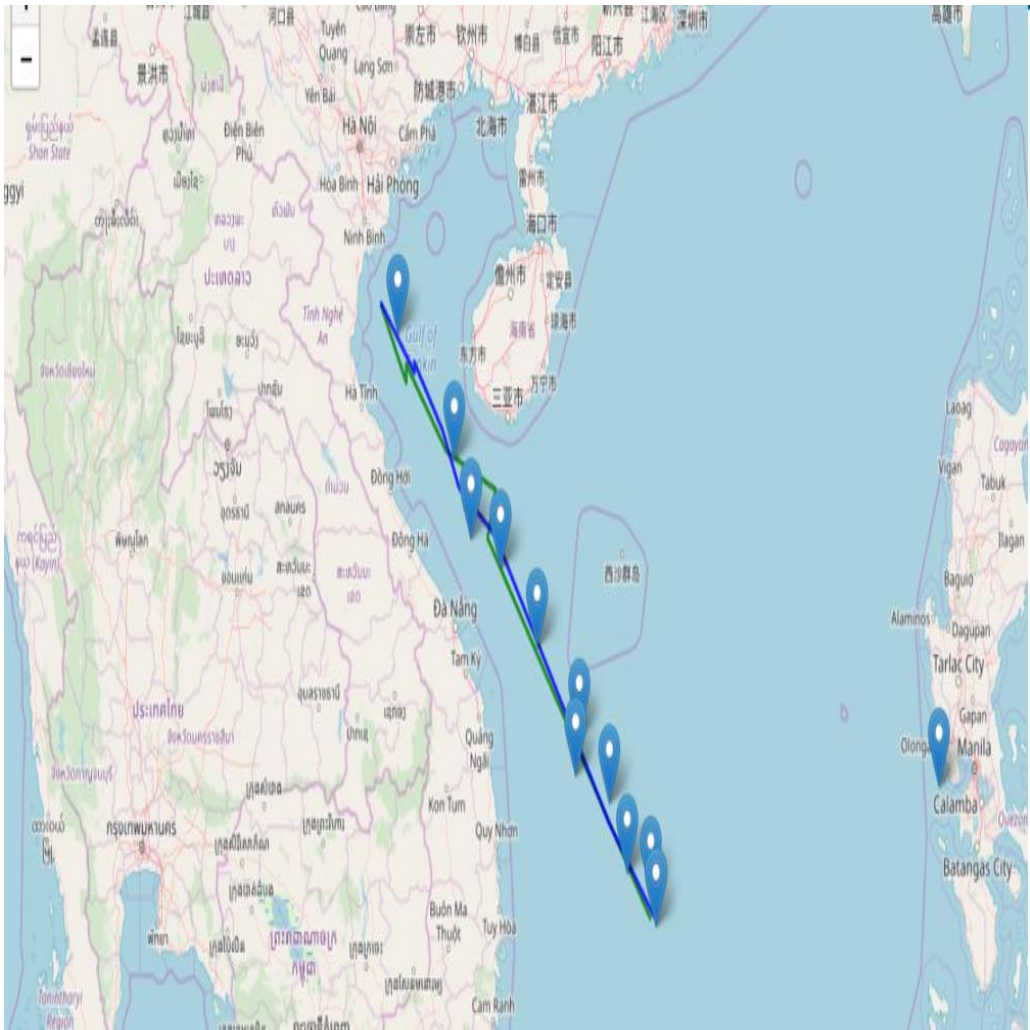


图 5.7 台风路径对比

再来看一下‘chebi’飞燕号台风，也是一场破坏性极强的台风，“飞燕”打破了 53 个日本气象站 10 分钟的历史记录，飞燕的增强让大 boss 的副热带气压也受不了，登录日本造成了巨大的灾害，给人民带来巨大的损失，影响了多个国际航班，韩国的金枪鱼养殖场也遭受了巨大的损失。‘飞燕’也成为了日本在二战后台风历史上破坏力数一数二的台风，那么模拟一下‘chebi’台风的路径如图 5.8 所示。

接下来我们在看一下肆虐广东的，莫拉菲 200906，(Typhoon Molave)，国际的编号：0906，JTWC：07W），成型的坐标是一个菲律宾的海域的形成的热带的低气压，在 2009 年 7 月 15 日被联合过的气象相关的部门首次发出台风预警。对于这个台风，台湾的交通气象局，莫拉菲路过台湾省的时候，带来了大量的降雨，经过巴士海峡。对于这次台风香港(hongkong)气象局也发出警戒信息。次日，香港的天文台取消了预警信号，澳门等地也接着发出了预警，台风在路过品

菲律宾吕宋岛的时候，导致多人死亡，汕头等地均有人员伤亡，如图 5.9 所示。
location=[[16.6 , 125] , [17.9 , 124] , [18.4 , 123.9] , [18.9 , 123.3] , [19.5 , 122.7] , [19.5 , 122.7] , [21.2 , 119.1] , [21.3 , 118.7] , [21.6 , 117.7] , [21.6 , 117.7] , [21.8 , 117] , [22.5 , 114.5]]:

如图对比图：可以看出我们预测有些偏差，在登陆点上有一些和实际不符合的散点。误差稍微大了些。

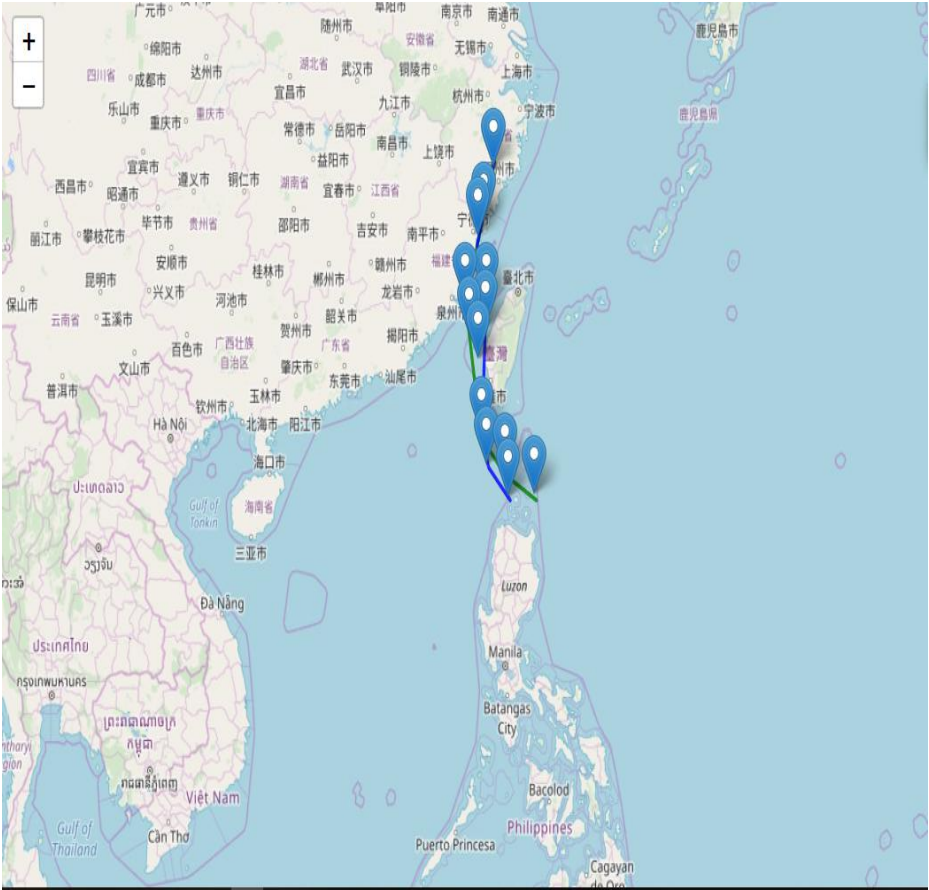


图 5.8 台风飞燕路径对比

除此之外，我们再看一下东南沿海的新主角，玛莉亚 201808，如图 5.10 所示，对日本造成了很大影响，其路径如下：

location= [[14.6 , 143.3] , [15.4 , 142.30] , [17.1 , 140.80] , [18.7 , 139] , [20.1 , 137.2] , [22.4 , 132] , [24 , 127]]

5.6 本章小节

本章主要阐述了系统分析需求、系统功能设计、系统的实现，和系统测试部分，本节给出了设计流程图、功能模块图等，形象的展示了本系统所用的算法。通过对台风山竹、飞燕、玛利亚、莫拉菲等真实路径值和预测的路径值进行对比，并计算误差距离，来凸显出本系统的准确率高的特点。对系统各个功能模块实现

的供能进行具体的阐述，将坐标点代入专业的地理信息软件，方便观察台风的预测登陆点，便于用户及时对台风的登录做出合理的对策。

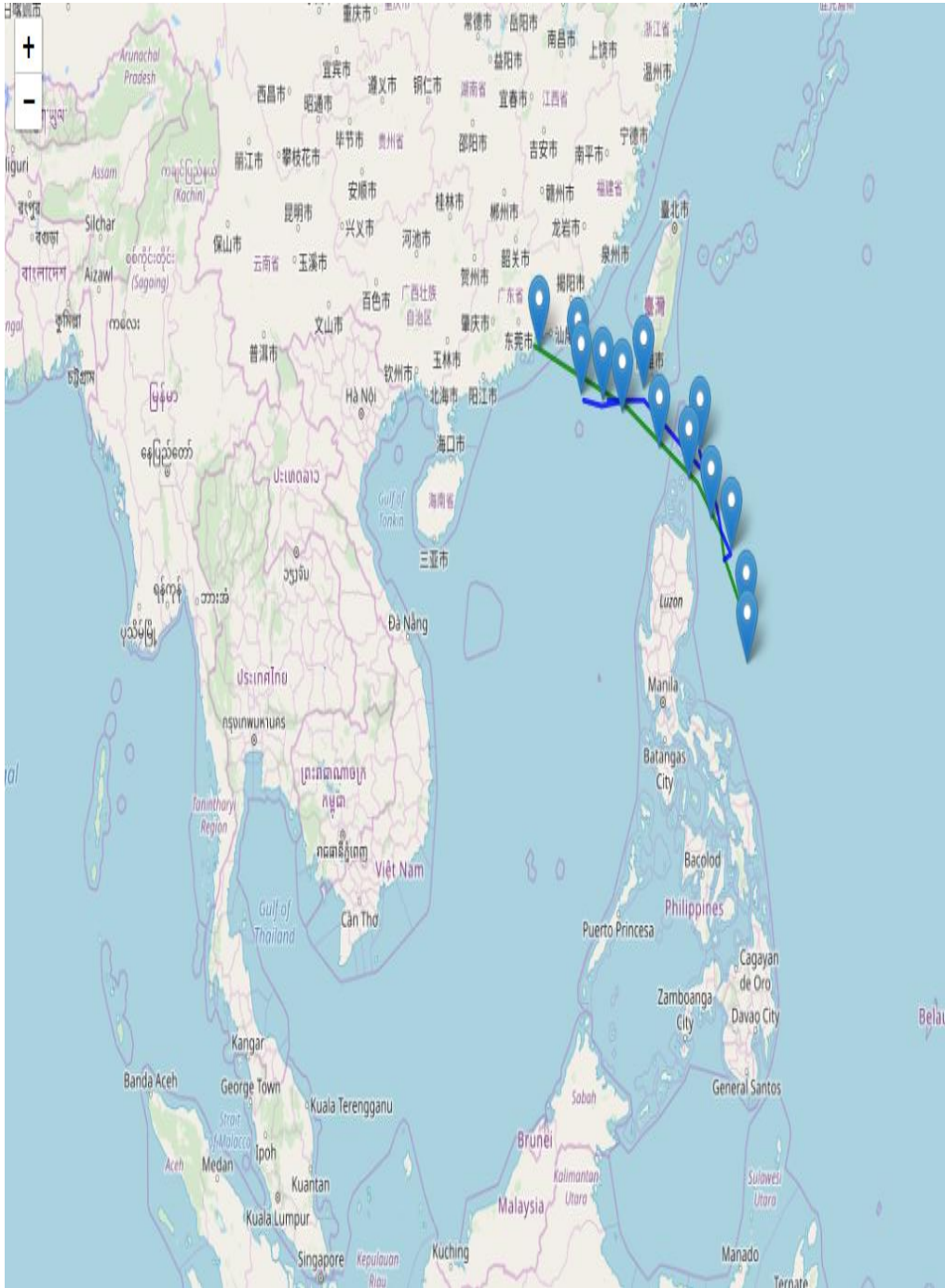


图 5.9 莫拉菲台风对比



图 5.10 玛利亚台风对比

6 结论

6.1 本文工作总结

本文运用机器学习的多元拟合回归,深度学习的长短期记忆神经网络对台风路径进行了预测。并采用了支持分类向量机对台风等级进行播报。

当卫星检测到台风成形的坐标,立马在中央气象数据库调这个经纬度的海域 sst, hgt, shum, 得到初始矢量。再根据这个坐标到后台的数据库中找到对应的前 6 小时的温度,湿度比,位势高带入.pkl 文件已经保存好的神经网络模型中进而得到这个时间戳的温度、湿度比、位势高,组成新的矢量,进而接着循环得到坐标点。之后再把得到的经纬度数组传给前端的台风演示系统,将经纬度的散点坐标连接起来,得到一个台风的路径图,便于用户较为直观的观看到登陆点的坐标。

台风的分类,分为 12 小时警戒线,24 小时警戒线 2 种情况。通过经验公式观察发现,位于南海部分的台风大多会进入 12 小时警戒线,而西北太平洋的基本会只停留在 24 小时警戒线,这个场景很类似于 SVM 支持分类向量机对坐标系的坐标点进行分类的原理,我们将按照台风的坐标点进行分类。本文对比一下朴素贝叶斯的分类效果可以发现 SVM 的分类效果比朴素贝叶斯的好很多。

通过路径对比图可以发现本系统实现的效果和官网的效果很接近,对台风等级的预测准确率高,本系统对路径的预测平均误差为 15.52 公里。

6.2 前景展望

1.网络结构: LSTM 在时间序列的预测效果已经明显优于传统的 RNN。但是在网络结构上还可以进一步的改进。可以在以后的系统版本中应用改进版的 LSTM。改进版的 LSTM 可以在选择遗忘的权重计算中更加智能,效果会比现有的模型好很多。

2.台风路径演示系统的改进: 优于在给用户演示台风路径的时候需要加载地球的地图,当网络较差的时候,加载的速度会很慢。在未来的系统版本中,可以添加用户自定义选择网络节点,当前网络节点网速较差的时候,支持用户对网络节点进行更改和切换 IP 通道的操作

参考文献

- [1]丁燕, 史培军. 台风灾害的模糊风险评估模型[J]. 自然灾害学报, 2002, 11 (1):34-43.
- [2] 葛耀君, 赵林, 项海帆. 基于极值风速预测的台风数值模型评述[J]. 自然灾害学报, 2003, 12 (3) :31-40.
- [3]余锦华, 唐家翔, 戴雨涵,等. 我国台风路径业务预报误差及成因分析[J]. 气象, 2012, 38(6) : 695-700.
- [3]邹亮, 任爱珠, 徐峰,等. 基于 GIS 空间分析的台风路径预测 [J]. 清华大学学报: 自然科学版, 2008 (12): 2036 - 2040.
- [4] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF Models for Sequence Tagging [J]. Computer Science, 2015.
- [5] Liu J, Shahroudy A, Xu D, et al. Spatio-Temporal LSTM with Trust Gates for 3D Human Action Recognition [J] . 2016.
- Gers F A, Schmidhuber J, Cummins F. Learning to forget: continual prediction with LSTM [C].
- [5] 佚名. 应用卫星云导风进行台风路径预报试验[J]. 热带气象学报, 1999 (4) : 347-355.
- [6] Huang Z , Xu W , Yu K . Bidirectional LSTM-CRF Models for Sequence Tagging[J]. Computer Science, 2015.
- [7] Liu J , Shahroudy A , Xu D , et al. Spatio-Temporal LSTM with Trust Gates for 3D Human Action Recognition[J]. 2016.
- [8] Gers F A , Schmidhuber, Jürgen, Cummins F . Learning to Forget: Continual Prediction with LSTM[J]. Neural Computation, 2000, 12(10):2451-2471.
- [9] Gers F A . Learning to forget: continual prediction with LSTM[C]// 9th International Conference on Artificial Neural Networks: ICANN '99. IET, 1999.
- [10] Shi X , Chen Z , Wang H , et al. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting[J]. 2015.

- [11] Graves A, Jaitly N, Mohamed A R. Hybrid speech recognition with Deep Bidirectional LSTM[C]// Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on. IEEE, 2013.
- [12] Ma X, Hovy E. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF[J]. 2016.
- [13] Ordóñez Francisco, Daniel R. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition[J]. Sensors, 2016, 16(1):115-.
- [14] Graves A, Jürgen Schmidhuber. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. Neural Networks, 2005, 18(5-6):602-610.
- [15] Ordóñez, Javier F, Roggen, et al. Sensors, Vol. 16, Pages 115: Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition[J]. Sensors, 2016, 16(1).
- [17] 高文胜, 周瑞旭, 张博文, et al. 台风登陆预测方法:.
- [18] 谭燕, 梁旭东, 黄伟. 集合预报技术在台风路径预报中的应用[C]// 海峡两岸气象科学技术研讨会. 2008.
- [19] 黄世成. 台风转向预测技术和强度变化机理的初步研究[D]. 南京气象学院, 2003.
- [20] 许富祥. 一种台风浪长期预测方法的探讨[J]. 海洋预报, 1995(2):67-73.
- [21] 段昊书. 台风是如何预报的[J]. 生命与灾害, 2017(10):12-13.
- [22] 李晴岚, 兰红平, 陈训来, 一种根据台风预报降雨的系统及方法:.
- [23] 叶小岭, 刘程波, 张颖超, . 基于 BP 神经网络的浙江台风损失预测[J]. 信息技术, 2011(10):59-61.
- [24] 黄杰华, 何龙, 张明棣, 应用数据挖掘算法预测台风条件下 PM_{2.5}质量浓度[J]. 环境监测管理与技术, 2016, 28(3):54-57.
- [25] 王璟琰, 李风华. 台风生成水下噪声强度估计及频率关系研究[C]// 2014 年中国声学学会全国声学学术会议.
- [26] 高文胜, 周瑞旭, 张博文, 台风登陆预测方法.

- [27] 谭燕, 梁旭东, 黄伟. 集合预报技术在台风路径预报中的应用[C]// 海峡两岸气象科学技术研讨会. 2008.
- [28] 黄世成. 台风转向预测技术和强度变化机理的初步研究[D]. 南京气象学院, 2003.
- [29] 许富祥. 一种台风浪长期预测方法的探讨[J]. 海洋预报, 1995(2):67-73.
- [30] 段昊书. 台风是如何预报的[J]. 生命与灾害, 2017(10):12-13.
- [31] 李晴岚, 兰红平, 陈训来. 一种根据台风预报降雨的系统及方法:.
- [32] 耿焕同, 周利发, 陈哲, et al. 一种台风路径预测方法:.
- [33] 叶小岭, 刘程波, 张颖超, et al. 基于 BP 神经网络的浙江台风损失预测[J]. 信息技术, 2011(10):59-61.
- [34] 黄杰华, 何龙, 张明棣, et al. 应用数据挖掘算法预测台风条件下 PM_{2.5} 质量浓度[J]. 环境监测管理与技术, 2016, 28(3):54-57.
- [35] 王璟琰, 李风华. 台风生成水下噪声强度估计及频率关系研究[C]// 2014 年中国声学学会全国声学学术会议.
- [36] 李明晓, 张恒才, 仇培元, et al. 一种基于模糊长短期神经网络的移动对象轨迹预测算法[J]. 测绘学报, 2018, 47(12):102-111.
- [37] 刘桐彤. 基于长短期记忆神经网络的短期负荷预测方法[J]. 科学技术创新, 2016(31):81-81.
- [38] 盛永, 黄阁, 廖国进. 0509 号台风波雨过程分析与暴雨灾害评估[C]// 中国气象学会年会. 2006.
- [39] 盛永, 陈艳秋, 廖国进, et al. 0509 号台风波雨过程分析与暴雨灾害评估[J]. 气象与环境学报, 2006, 22(6):29-33.
- [40] 孙兴池, 王文毅, 王业宏, et al. 0509 号台风麦莎影响山东分析[J]. 热带气象学报, 2007(3).
- [41] 刘学刚, 李庆宝, 张金艳, et al. 近年来引发青岛暴雨的台风特征分析[J]. 气象, 2011(9).
- [42] 张建海, 于忠凯, 何勇. 两个路径相似台风波雨过程的模拟分析[J]. 热带气象学报, 2010, 26(4):392-402.
- [43] 陈联寿, 孟智勇. 我国热带气旋研究十年进展[J]. 大气科学, 2001, 25(3):420-432.

- [44] Xilinx Inc. IP Catalog Documentation and Tutorials[S]. Xilinx Vivado Documentation. 2018.1.
- [45] Chen Zhang, PengLi, Guangyu Sun.Optimizing FPGA.based Accelerator Design for Deep Convolutional Neural Networks[C].Center for Energy.Efficient Computing and Applications, Peking University, China; Computer Science Department, University of California, Los Angeles, USA; PKU/UCLA Joint Research Institute in Science and Engineering.
- [46] 韦志刚, 黄荣辉, 董文杰. 青藏高原气温和降水的年际和年代际变化[J]. 大气科学, 2003, 27(2):157-170.
- [47] 李成才, 毛节泰, 刘启汉, et al. 利用 MODIS 光学厚度遥感产品研究北京及周边地区的大气污染[J]. 大气科学, 2003, 27(5):869-880.
- [48] 孙建华, 赵思雄. 华南"94 • 6"特大暴雨的中尺度对流系统及其环境场研究 I. 引发暴雨的 β 中尺度对流系统的数值模拟研究[J]. 大气科学, 2002, 26(4).
钱维宏, 张玮玮. 我国近 46 年来的寒潮时空变化与冬季增暖[J]. 大气科学, 2007, 31(6):1266-1278.
- [49] 孙建华, 赵思雄. 华南“94 • 6”特大暴雨的中尺度对流系统及其环境场研究 II.物理过程、环境场以及地形对中尺度对流系统的作用[J]. 大气科学, 2002, 26(5):633-646.
- [50] 张小玲, 陶诗言, 张顺利. 梅雨锋上的三类暴雨[J]. 大气科学, 2004, 28(2):187-205.
- [51] 武炳义, 黄荣辉, 高登义. 冬季北极喀拉海、巴伦支海海冰面积变化对东亚冬季风的影响[J]. 大气科学, 1999, 23(3):267-275.

致谢

四年的生活即将画上一个句号，在东北大学的4年年华短暂而又快乐，但是对于我的人生来说，只是一个逗号，对于计算机专业毕业的学生应该做出自己的系统，而不是只会理论的研读，我首先学习了机器学习经典算法，首先学习了C4.5算法的经典的分类决策树，决策树的构造方法就是将一个选择好的特征点进行作为当前的分类节点的分类的特征，原来的决策树基本采用熵的衡量标准，在学习这里的时候我多次询问我以前的人工智能的老师。在他的指导下我又进一步学习了K-means算法(The K means algorithm)这个经典的分类算法。把n个数据分为K个类。在学习PageRank的算法的时候，也要感谢同是计算机毕业的表格的指导，他教会了我这个(Larry Page)算法来衡量网站的火热程度，采用对一个网站的链接的数目来衡量是否是比较火。进而我的表哥还教会了我朴素贝叶斯的分类算法(Naive Bayes)。贝叶斯网络表示的是一组联合概率的分布，其中步骤就是贝叶斯网络结构学习之后就是贝叶斯网络参数学习。之后有学习了无监督学习，聚类就是分类的样本标签是未知的。组成的簇内部的样本相似程度较高。通过对目标函数最小化来将对象划分为K各组。但是本算法对初试的质心很敏感。

在5月份的时候学习了强化学习。众所周知，机器学习分为三大类，监督学习，非监督学习，强化学习。而强化学习是在没有标签的情形下，先自己尝试一个结果。通过结果的反馈，来得到强化学习的最终的模型。‘知错’便是强化学习的核心点之一。试错搜索，延迟回报是关键的两个特征。当模型的未知的情况下，我们需要对<状态，动作对>进行累计奖赏。

这个阳光明媚的季节，我即将结束了自己本科的四年生活，四年的本科经历，走的艰辛也成绩颇丰，在毕业论文即将结束之际，我感慨万分，在本次毕业设计的研究中，我首先要感谢乔百友老师对我的细心指导，他也是我的计算机体系结构的老师。刚开始对深度学习一无所知，拿到这个深度学习预测台风路径的课题时，毫无思路，对气象学和深度学习都是一无所知在乔百友导师的指导下逐步学会LSTM的使用，进而对深度学习有了很深的理解，向乔老师多次询问课题的思路，在乔老师的指导下，思路拓展了许多，也纠正了之前的许多做法，乔老师经常询问我的进度，关心我的研究情况，在我的系统的最终演示时候提出了宝

贵建议，将本系统的演示结果进一步的详细的显示出来，做出了很好看的界面，演示效果大大增强，每次指导都能得到新的观点，使得当前的难题迎刃而解。也要感谢姚来刚学长，他也是台风气象的有关研究，和学长多次探讨，解决了数据的瓶颈问题，并确定了 Keras 等学习框架，将大量的气象数据进行高效率的分析，达到对台风路径预测的准确率，并认真探讨了当前的几个流行框架的弊端和优势。还要感谢和我一起科研的各个同学，在遇到 bug 时正是有他们的陪伴，才是得我正确修复 bug，才让我的程序得以运行。在图书管查阅资料的时候，图书馆的老师也给了我诸多指导。同事也要感谢引用这的这些论文的作者，如果没有这些前辈在学术上的上的铺垫，我将走很多弯路，从论文到选题几次修改，最终定稿。

从做出毕业设计到最终写出论文，前前后后大致有 45 个月的时间，这期间有了同学的陪伴一起学习深度学习，一起学习深度学习的框架，一起给我思路，一起来寻找准确率底下的原因，一起来寻找合适的优化器，损失函数，合适的 batch_size。合适的 epoch，在寻找效果理想的路径图，相比于自己独自科研，效率会更高。对于每个机器学习的模型，

当然也要感谢我的父母，在做毕业设计的期间不慎感染了支原体肺炎，一遍忙于治病，一边忙于毕业设计的构思这期间，医院的护士，医院的大夫对我无微不至的关怀也是我做毕业设计动力，当然还要感谢我的主治大夫。要感谢自己的父母在这段煎熬的时间对自己的心细照料。如果没有他们，我很难度过这样的年华。在医院的无数个日日夜夜对数据进行调参数，多亏了医院护士的细心照料。

当然还要感谢我的母校，给了我这个精良的科研环境，让我能一心一意的专心于毕业设计的项目研究，安静的实验室，资料齐全的图书馆。本次论文的顺利完成离不开温冬，王治郅，孙亮的悉心指导，还有王新东提供的思路。才能将本次的毕业设计论文顺利写出来。希望读研以后也会在科研路上越走越远。

