

文本复制检测报告单(全文标明引文)

№:ADBD2019R_20190603233136444135751116

检测时间:2019-06-03 23:31:36

检测文献: 基于深度学习的台风路径预测技术的研究与实现

作者: 朱强

检测范围: 中国学术期刊网络出版总库

中国博士学位论文全文数据库/中国优秀硕士学位论文全文数据库

中国重要会议论文全文数据库

中国重要报纸全文数据库

中国专利全文数据库

图书资源

优先出版文献库

大学生论文联合比对库

互联网资源(包含贴吧等论坛资源)

英文数据库(涵盖期刊、博硕、会议的英文数据以及德国Springer、英国Taylor&Francis 期刊数据库等)

港澳台学术文献库

互联网文档资源

CNKI大成编客-原创作品库

个人比对库

时间范围: 1900-01-01至2019-06-03

检测结果

去除本人已发表文献复制比: 5.9%

跨语言检测结果: 0%

去除引用文献复制比: 5.9%

总文字复制比: 5.9%

单篇最大文字复制比: 1.6% (预报准确率迈入国际先进行列)

重复字数: [1626]

总字数: [27658]

单篇最大重复字数: [433]

总段落数: [3]

前部重合字数: [1197]

疑似段落最大重合字数: [1276]

疑似段落数: [3]

后部重合字数: [429]

疑似段落最小重合字数: [43]

指标: ☐ 疑似剽窃观点 ☒ 疑似剽窃文字表述 ☐ 疑似自我剽窃 ☐ 疑似整体剽窃 ☐ 过度引用

表格: 0 公式: 0 疑似文字的图片: 0 脚注与尾注: 0

12.2% (1276) 基于深度学习的台风路径预测技术的研究与实现_第1部分 (总10494字)

3% (307) 基于深度学习的台风路径预测技术的研究与实现_第2部分 (总10332字)

0.6% (43) 基于深度学习的台风路径预测技术的研究与实现_第3部分 (总6832字)

(注释: ■ 无问题部分 ■ 文字复制部分 ■ 引用部分)

指导教师审查结果

指导教师: 乔百友

审阅结果: 通过

审阅意见: 通过, 总体重复率符合要求, 建议进行适当调整, 并引用相关文献。

1. 基于深度学习的台风路径预测技术的研究与实现_第1部分

总字数: 10494

相似文献列表

去除本人已发表文献复制比: 12.2%(1276) 文字复制比: 12.2%(1276) 疑似剽窃观点: (0)

1	预报准确率迈入国际先进行列 本报记者 黄彬 郝静 刘淑乔 - 《中国气象报》 - 2018-03-01	4.1% (433) 是否引证: 否
2	51241766279385751_徐博_我国东南沿海地区近20年台风灾情研究 徐博 - 《高职高专院校联合比对库》 - 2018-05-15	2.4% (249) 是否引证: 否
3	台风莲花、灿鸿数值模拟试验 程秋人 - 《大学生论文联合比对库》 - 2016-06-24	1.7% (178) 是否引证: 否
4	对现行防抗台风措施的应用研究和改良--以广州市为例	1.7% (178)

	韦奇 - 《大学生论文联合比对库》 - 2017-05-10	是否引证：否
5	北京气象局 - 《网络 (http://www.bjmb.gov) 》 - 2013	1.7% (178) 是否引证：否
6	台风钻鱼会造成什么危害_碧海如云 - 《网络 (http://blog.sina.com) 》 - 2013	1.4% (152) 是否引证：否
7	台风的危害_小荷才露尖尖角 - 《网络 (http://blog.sina.com) 》 - 2015	1.4% (151) 是否引证：否
8	台风的危害_小荷才露尖尖角 - 《网络 (http://blog.sina.com) 》 - 2018	1.4% (151) 是否引证：否
9	美国飓风破坏力为何如此巨大 李永平;陈晶茹; - 《生命与灾害》 - 2012-11-20	1.3% (139) 是否引证：否
10	台风与台风危害 - 《中国防汛抗旱》 - 2010-10-15	0.9% (94) 是否引证：否
11	高三地理自然地理专题学习指导-百度文库 - 《互联网文档资源 (http://wenku.baidu.c) 》 - 2012	0.9% (92) 是否引证：否
12	高中地理热点知识解析 - 《互联网文档资源 (http://wenku.baidu.c) 》 - 2017	0.9% (92) 是否引证：否
13	21-气象 - 《互联网文档资源 (http://wenku.baidu.c) 》 - 2017	0.9% (92) 是否引证：否
14	高考自然地理专题 - 《互联网文档资源 (http://wenku.baidu.c) 》 - 2017	0.9% (92) 是否引证：否
15	台风监测预报系统关键技术获国家科学技术进步奖 记者 崔国辉 - 《中国气象报》 - 2019-01-09	0.9% (91) 是否引证：否
16	台风期间的海洋观测及其新技术发展 朱伯康,刘增宏,许建平 - 《海洋技术》 - 2005-06-15	0.9% (91) 是否引证：否
17	台风林奔凡-百度文库 - 《互联网文档资源 (http://wenku.baidu.c) 》 - 2012	0.9% (91) 是否引证：否
18	梅雨季结束 酷暑模式开启_滚动读报_地方频道 - 《网络 (http://difang.gmw.cn) 》 - 2016	0.8% (81) 是否引证：否
19	1211号台风“海葵”降水特征分析 梁卉林;唐晓东;袁慧玲; - 《创新驱动发展 提高气象灾害防御能力——S18第四届研究生年会》 - 2013-10-22	0.7% (78) 是否引证：否
20	基于深度学习和MFCC特征的脑卒中预测方法 陈鸥宇;刘怡俊;叶武剑;牟志伟;李琪; - 《信息与电脑(理论版)》 - 2019-02-15	0.7% (76) 是否引证：否
21	影响我国的主要气象灾害及其发展态势 秦大河; - 《中国应急救援》 - 2008-11-20	0.6% (62) 是否引证：否
22	山东人民版小学四年级品德与社会上册《面对突发事件》教案 - 豆丁网 - 《互联网文档资源 (http://www.docin.com) 》 - 2015	0.5% (55) 是否引证：否
23	公共事业管理1301101123赖艳洁 赖艳洁 - 《大学生论文联合比对库》 - 2017-05-18	0.4% (47) 是否引证：否
24	上层海洋对台风的响应分析 ——以台风“浣熊”为例 赖颖 - 《大学生论文联合比对库》 - 2018-05-26	0.3% (33) 是否引证：否
25	主动学习与半监督技术相结合的海冰图像分类 韩彦岭;李鹏;张云;徐利军;王静; - 《遥感信息》 - 2019-04-20	0.3% (31) 是否引证：否

原文内容

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, 据美国海军的联合台风警报中心统计, 西北太平洋及中国的南海海域的台风发生的个数与月份, 平均每年有好多的台风生成, 出现最多台风的月份是8月, 其次是7月和9月6月5月。科学家曾预算, 一个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波的人数约为2000到3000人。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成为一个重要的课题。

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形式的温度, 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，实际的测试表明，该方法达到了平均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，数据训练，数据预处理，台风的路径演示等诸多功能。

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; MachineLearning; DeepLearning;

The neural network;Long and short term memory neural network

目录

摘要.....I

ABSTRACT.....II

第1章.....绪论.....1

1.1项目研究背景及意义.....1

.....1

1.2国内外台风预测方法现状.....2

1.3国内外台风预测方法研究.....4

1.4研究内容.....5

1.5论文组织结构.....6

1.6本章小结.....6

第2章.....相关技术.....7

2.1 机器学习.....7

2.2 循环神经网络RNN.....8

2.3神经网络LSTM.....11

2.4 foilum.....12

2.5.Python-tkinter.....12

2.6大气动力学.....13

2.7本章小结.....14

第3章.....深度学习预测台风路径.....15

3.1本文的研究内容.....15

3.2台风路径预测研究.....17

3.2.1数据预处理.....17

3.2.2采用LSTM对温度位势高湿度比进行预测.....21

3.3.3运用神经网络的结果来预测路径.....26

第4章台风等级播报.....31

4.1 SVM支持分类向量机.....31

4.2.svm用于台风等级播报.....32

第5章系统分析和设计.....33

5.1 系统开发目标.....34

5.2 录入数据的系统界面.....34

5.2.1录入数据的图形界面.....34

5.2.2matplotlib的结果对比.....	35
5.2.3folium台风路径预测系统演示.....	36
5.2.4预测系统测试部分.....	37
第6章系统使用手册.....	42
6.1 系统的使用.....	42
6.2 代码文件说明.....	44
参考文献.....	46
致谢.....	50

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暴雨对于交通而言，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。她引起房屋倒塌、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。

例如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，而且给守在人民带来了焦灼、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人意。

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 国内外台风预测方法现状

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、台风风场释用技术、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 (Weather rResearch And ForecastModel)，是一种较为有效的台风数值模型，他是非静太力平衡的数值模式，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，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。

1.3国内外台风预测方法研究:

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]和风险评价相结合,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],台风结构:

$$V(r) = V_m (R_{gale} / RMW)^{-x} \quad (1.1)$$

$$V(r) \text{ - 半径为 } r \text{ 的台风的风速} \quad (1.2)$$

$$V_m \text{ - 最大风速} \quad (1.3)$$

$$R_{gale} \text{ - 最大半径} \quad (1.4)$$

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,再根据就个场台风的气象站点结合国家气象数据裤,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’,‘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,这是一类包含将前面的结果进行综合考虑的,且具有深度结构的前馈神经网络[15](Feedforward Neural Networks),是深度学习(deep learning)的代表算法之一。

但在另外一方面,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]。基于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],本次设计的工作应用机器学习的知识,应用于南海台风路径模型的建立。

1.4本文研究内容

本文通过对台风的初始坐标(x0,y0),以及初始坐标点的海洋表面温度SST(Sea Surface Temperature),位势高度数据HGT(geopotential height),比湿数据(specific humidity)对接下来的台风路径坐标每隔6小时一次预测。

训练模型所用数据主要是来源于美国JTWC(Joint Typhoon Warning Center,美国联合台风预警中心)上下载的,将2001年~2005年的这5年的所有成形的台风路径数据,比湿数据,位势高matlab数据下载下来,并通过matlab初步处理,形成数据文件‘track.mat’。其中包括三个部分,海表面温度数据(Sea Surface Temperature, SST, (°C))位势高度数据(geopotential height, hgt.(m))。路径数据主要是从美国JTWC(Joint Typhoon Warning Center),美国联合台风预警中心)上下载的,其网址如下 <https://metoc.ndbc.noaa.gov/JTWC>。

我们这个毕业设计主要是应用三个数据集,141个台风的路径数据集,近几年的温度数据集,压强的数据集,温度的数据集对台风的这几个影响因素进行分析。数据预处理之后,我们将所有的数据进行压缩挑选。

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]。可以看出误差很大。

目前大多数用户例如如海捕捞的渔民,沿海的百姓等等,很难时刻关注官网的台风动向,目前的模型对台风的预测还难做到精准和即时,我们的这个系统容量小,渔民等用户可通过这个软件来简单的输入初始点坐标,得到台风路径的坐标,进而在短时间内将风险降到最小。效率比台风官网通报来的及时。我们可以采用2种模型进行对比明显发现RNN的效果不如LSTM,无论是路径预测还是对单个因素的预测均不是很理想。我们拿出‘Marge’号台风,(马琦)坐标起始位于菲律宾吕宋岛之后再凌晨一直向大陆移动,之后登陆了海南岛的琼海,造成了将近903人死亡,5000多人受伤。我们拿这个台风来测试2个神经网络的模型,我们并不难发现,两种模型的误差对比,LSTM明显准确率要比RNN这种传统的模型来的高,所以我们本次的基于深度学习的台风路径的预测打算使用LSTM这个神经网络。

我们本次毕业设计进行对台风路径的预测主要应用的就是深度学习(deeplearning)的知识,和机器学习的相关内容,打破传统的物理公式推导的弊端,适用于沿海的渔民等进行对台风动向的实时预测。

1.5论文组织结构

第一张绪论主要是对所用的技术进行大体阐述,我们就是对所用技术的进行具体的分析。第二章主要是相关技术简介,对于所用的神经网络LSTM,还有机器学习的SVM算法对于台风起始坐标的对台风等级的播报,来预测出是24小时警戒线还是48小时警戒线。第三章是本文的研究内容和技术路线对主要的技术具体讲述了LSTM的神经网络的进行讲解,具体讲解了LSTM如何训练,送入数据的具体讲解。第四章台风等级播报通过采用SVM对等级进行播报,由卫星传过来的中心台风风眼坐标。来区分该台风到底是24小时警戒线还是48小时警戒线。第五章的系统分析和设计是主要对本系统进行了大致的讲解以及如何使用,对我们预测的路径进行和实际值的比对,我们大致列出了和实际的几场成形的台风和我们预测的路径进行的误差计算,列出了一个大致表格,用于系统的分析。第6章是系统使用手册对本系统进行的详细的讲解。

1.6本章小结

本小章大致论述了本次设计的大致路线,首先对系统的功能性需求及非功能性需求分别进行了分析,针对系统需求给出了系统的整体设计讲述了传统模型的弊端,对所有的技术进行了大致的介绍,进而为后面的深度学习模型引出做出铺垫。

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(computer vision),自然语言处理(natural language processing),语音识别(Speech Recognazation)[18]。所用的Sklearn也是广受大家喜好的机器学习的框架,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(supervize learning),无监督学习(unsupervised learning),半监督学习(semi supervised

learning),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(machine learning)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’,非监督学习的无人为的标签(label),常见的应用场景应用比如聚类[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。

其中的编辑器选用pycharm,它是一种高效率的可靠编辑器,是JetBeans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。

其中机器学习的一个重要分支就是深度学习。深度学习在预测(predict),回归(regression),分类(classify)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。

DeepLearning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(linear algebra)[25],最优化理论(optimization theory),逼近论(Approximation theory),高等数学(advanced mathematics),微积分(calculus),高等代数(advanced algebra),离散数学(discrete mathematics),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。

图2.1 网络结构

2.2 循环神经网络RNN:

图2.2 工作原理

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的也有权值链接,RNN是一种传统的序列数据的构建模型:

$$H_t=f(W_{xh} \cdot X_t + W_{hh} \cdot H_{t-1}) \quad (2.1)$$

$$Y_t=g(W_{hy} \cdot H_t) \quad (2.2)$$

我们列出RNN的必要的推导公式,其中G()函数作为激活函数

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

$$S_t=f(U \cdot X_t + W \cdot S_{t-1}) \quad (2.4) \quad O_t=g(V \cdot S_t) = g(Vf(U \cdot X_t + W \cdot S_{t-1})) = g(Vf(U \cdot X_t + W \cdot f(U \cdot X_{t-1} + W \cdot S_{t-2}))) = g(Vf(U \cdot X_t + W \cdot f(U \cdot X_{t-1} + W \cdot f(U \cdot X_{t-2} + \dots)))) \quad (2.4)$$

图2.3 网络的展开

如上图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。

指 标
疑似剽窃文字表述
<div><div>1. 据美国海军的联合台风警报中心统计,西北太平洋及中国的南海海域的台风发生的个数与月份,平均每年</div><div>2. 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波的人数约为2000到3000人。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。</div><div>3. 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。</div><div>4. 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。</div><div>5. 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。</div><div>6. 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、台风风场释用技术、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。</div><div>7. 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。</div></div>

相似文献列表

去除本人已发表文献复制比：3%(307) 文字复制比：3%(307) 疑似剽窃观点：(0)

1	XSS漏洞分析与安全加固实现 方泽 - 《大学生论文联合比对库》 - 2018-05-31	2.5% (257) 是否引证：否
2	基于神经网络的中文事件时序关系的识别 赵雪峰 - 《大学生论文联合比对库》 - 2018-05-13	2.2% (226) 是否引证：否
3	基于神经网络的中文事件时序关系识别方法研究 赵雪峰 - 《大学生论文联合比对库》 - 2018-05-19	2.0% (205) 是否引证：否
4	基于深度学习的交通流预测方法研究 徐宁 - 《大学生论文联合比对库》 - 2018-06-16	1.9% (196) 是否引证：否
5	藏-汉神经网络机器翻译系统设计与实现 次仁巴姆 - 《大学生论文联合比对库》 - 2018-06-04	1.9% (196) 是否引证：否
6	基于序列标注的引语识别初探 贾泓昊;罗智勇; - 《中文信息学报》 - 2019-02-15	0.5% (51) 是否引证：否
7	基于机器学习的城乡统筹供水水质异常检测方法研究 付亚涛(导师：秦文虎) - 《东南大学博士论文》 - 2018-05-27	0.4% (45) 是否引证：否
8	基于长短期记忆神经网络的暖通空调系统能耗预测 廖文强;王江宇;陈焕新;丁新磊;尚鹏涛;魏文天;周镇新; - 《制冷技术》 - 2019-02-15	0.4% (43) 是否引证：否
9	pm2_高旭 高旭 - 《高职高专院校联合比对库》 - 2018-05-30	0.4% (40) 是否引证：否
10	数据驱动的航空发动机余寿预测方法 唐王(导师：皮德常) - 《南京航空航天大学博士论文》 - 2018-03-01	0.3% (31) 是否引证：否

原文内容

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 (sequence data) 的有效处理，语音识别(Speech Recognition)，还有音乐生成，自动写诗，DNA遗传物质分析(DNA sequence analysis) 视频活动识别[28] (Video active recognition)，对于句子的识别，自然语言处理的任务名称的识别 (Name entity recognition)

2.3神经网络LSTM

我们不妨思考一个问题-长期依赖的问题， (long-Term Dependence) 问题，如果我们仅仅需要知道先前的信息来预测，那么RNN的效果有些时候很好，但是更多的时候需要更多的信息依赖[29][30]，例如我们针对一个英语句子的最后 There is a bird flying in the sky我们很轻松就会预测出那个最后的一个词语是sky，这种场景的RNN在依赖不是很远信息的时候效果比较好。但是事实上的情况往往比这个要来的复杂的多，I am Chinese...I speak Chinese 这样中间的信息隔了诸多距离，那么RNN在记忆这种像个很远的信息的能力很差。

我们可以打一个比方，一个老爷爷记忆力并不是很好，但是对自己的记忆力盲目自信，实际上只能记住前几天发生的事情太过久远的事情就记不住了，每次别人问起他这几天发生的事情，他只能支支吾吾的回答出这前几天的发生的事情，但是他的儿子就吸取他父亲的教训，他损失那这个本子记下啦这几天发生的事情，当本子写满了以后，那么关键的技术来了，就是选择性的将这些记忆信息去掉，记住下当前的最重要的事情。那么得益于这个小本子，儿子可以根据这个本子回忆很久以前发生的事情，对这个序列化的效果明显优于老爷爷的记忆力。

其中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’(gate)来实现的。‘门’(gate)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’[34][35][36](forget gate)，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进行对因素的预测，训练模型，进而探究对台风路径的预测。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，我们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][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。

除此之外，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（supplement）。

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(loading)和训练(train),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最长于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][38][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):

图2.5 大气动力学方程

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][41][42]。

2.7 本章小结

本章我们主要对所应用的各种技术进行了分析对比，并且把RNN模型和LSTM进行对比，我们选择了更好的深度学习模型LSTM，将于时间相关的信息进行更好的预测。

3 基于深度学习的方法预测台风路径

3.1 研究思路

本文通过对台风的初始坐标(x0,y0),以及初始坐标点的海洋表面温度SST(Sea Surface Temperature)，位势高度数据HGT (geopotential height) ,比湿数据 (specific humidity) 对接下来的台风路径坐标每隔6小时一次预测。

首先将.mat的数据读入，并进行预处理，将matlab数据集中nan的数据剔除。然后其中采用keras框架，对台风路径的三个影响因素分别采用LSTM算法进行预测，Istm (Long Short-Term Memory) 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。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]。

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：

$$X_{t+5}=f(x_t,sst,shum,hgt) \quad (3.1)$$

$$Y_{t+5}=g(y_t,sst,shum,hgt) \quad (3.2)$$

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v0=(x0,y0,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。

下面给出思维导图：

图3.1 思维导图

3.2 台风路径预测研究

3.2.1 数据预处理：

下面对所用数据进行处理：

<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坐标图1：

图3.2 路径图

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：对应着上面台风的名字 141

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，最低气压

<2>海表面温度数据 (Sea Surface Temperature, SST, (°C)):

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121×161×7304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:

图3.3温度场

<3>比湿数据 (specific humidity)

“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121×161×7304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。如下所示

图3.4比湿数据场

<4>位势高: geopotential height, hgt,(m):

hgt数据文件“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”

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

hgt2: 表示200hPa位势高度数据, 是121×161×7304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, 接着进行预处理流程。Loadmat函数返回的是一个字典, 在140场台风中我们用构造出向量:

V01=(xt,sst, shum,hgt) (3.3)

V02=(yt,sst,shum,hgt) (3.4)

以及标签:

Xt+5

Yt+5

3.2.2采用LSTM对影响因素进行预测

LSTM(Long Short Term Memory)长短期记忆神经网络是当下最同行的RNN,

RNN:

$H_t = \text{sigmoid}(W_{xh} * X_t + W_{hh} * H_{t-1})$ (3.5)

其中 W_{xh} 和 W_{hh} 是可学习的参数

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。

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, 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)OutputLayer

图3.11 输出层

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)

3.3.3数据输入

我们将温度sst.mat,这个三维数组121X161X7304个数据进行送入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，就没有冗余，而且比较快，并且可以新增样本。但是SGD及有可能的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,当梯度大的时候，学习率变小，当梯度小的时候，学习率变大，以确保网络的稳定性。有效解决了Adagrad学习率极度下降的缺点，经过对比发现rmsprop效果更好一些。

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svae_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送入的数据进行预测，得到下一个时刻的温度湿度比和压强数据。

下面我们查看一下神经网络对这几个因素的预测效果：

(1)采用LSTM对SST进行预测的图像：

图3.12 温度处理

(2)采用LSTM对HGT进行预测的图像：

图3.13 位势高

3.3.4 运用神经网络的结果进行预测路径：

刚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一年的海洋数据就是好几万，由于科研经费有限，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我开始放弃对于大气物理推导加上lstm的模型。我转而把目光投向了机器学习的算法，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，我们仅仅讨论温度，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既是是大型计算机运算起来也很复杂，是通过自然条件的多变也导致了台风的走向的不确定性，更加能体现出模型的优越性。

运用上面得到的数据，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。

我们对这个递推方程进行训练：

(一直当前坐标对接下来的坐标进行预测)

$$X_{t+5}=C1+C2*X_t+C3*shum+C4*hgt+C5*sst \quad (3.6)$$

$$Y_{t+5}=C7+C8*Y_t+C9*shum+C10*hgt+C11*sst \quad (3.7)$$

调用sklearn中的LinearRegression进行对全体数据的拟合，其中抽象原理：

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

最这些数据进行拟合成：

$$(w,b)=()$$

对其进行求偏导数：

令上面2个公式等于0，可求出对应的w b。

既是求出对应的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的文件中在前面界面传入初始向量：

$$V0=(x_t,y_t,shum,sst,hgt) \quad (3.9)$$

将这个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，分别保存在save_hgt_model.pkl, save_sst_model.pkl, save_shum_model.pkl 我们调用这三个文件。将他们已经训练好的模型调用出来，将这个经纬度几个小时之前的温度压强湿度比带入这个模型中。将其预测出来，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，一次求出每隔6小时一次的坐标,然后传入foiliun，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。

我们对已经运行的模型进行预测并计算误差，我们对比一下已经有的官网数据的对我们训练出来的模型进行误差对比，下面我们给出误差分析公式：

$$/n \quad (3.10)$$

$$(3.11)$$

图3.14 台风路径对比图

误差表

台风6小时误差 12小时误差 18小时误差
chebi15.213 15.37888 16.712
Wukong19.213333333 19.0234 19.293
Muifa21.09812 20.1213 21.345
Langka12.223409 13.22213 12.442
Saomai40.21334 39.021 39.214
Usagi37.7212 38.214 38.213
Hato30.00021 30.213 30.213
Rammasun31.2131 31.234 30.6001
Mujigae32.11123 32.1021 32.0123
Haiyan20.1233 20.2134 20.409
Nepartak19.213 20.34 20.1
Mirinae20.213 20.123 20.78222
Maysak19.0213 20.13 20.1
Soudelor12.345 15.87 16.901
Fengshen13.4455 14.67 15.2
Maemi32.121214 33.112 38.21
Hagupit45.67091 40.213 49.21
Vongfong23.213 20.14 26.57
Phanfone23.2131 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’。我们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(Support Vector Machine)是一种对样本的进行二维分类的广义分类器 (generalized linear classifier)，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：

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

之后采用拉格朗日乘数法，求解最优的解

图4.1 支持向量机

如图也叫感知机，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，那么我们此时可以使用‘核变换’，我们可以将低纬度的数据映射到高纬度的数据。假设我们有2个数据集：

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

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

d对于这三个纬度数据，此时的3D空间，已经很难弄对其进行线性化分了，那么我们可以见这个空间中的所有数据，可以采用一个函数进行数据映射到高纬度的空间，其中那个的‘核函数’便是变量x变换：

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

支持向量机在人脸识别(face recognition)也有着重要的应用。

指 标		
疑似剽窃文字表述		
1. 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，		
3. 基于深度学习的台风路径预测技术的研究与实现_第3部分		总字数：6832
相似文献列表		
去除本人已发表文献复制比：0.6%(43) 文字复制比：0.6%(43) 疑似剽窃观点：(0)		
1	大红山铜矿空间数据库的设计 金怡杉(导师：方源敏) - 《昆明理工大学博士论文》 - 2009-03-10	0.6% (42) 是否引证：否
2	房产空间数据整合研究 张鑫(导师：闫国年) - 《南京师范大学硕士论文》 - 2004-05-01	0.6% (42) 是否引证：否
3	MAPGIS67操作技巧汇总 - 豆丁网	0.6% (42)

	- 《互联网文档资源 (http://www.docin.com) 》 - 2017	是否引证：否
4	ArcGIS的三种文件：shapefile,Coverage,Geodatabase - 豆丁网 - 《互联网文档资源 (http://www.docin.com) 》 - 2015	0.6% (42) 是否引证：否
5	基于ArcGIS Server的移动客户端路径导航系统 邓中亮;崔艳雯; - 《软件》 - 2011-02-15	0.6% (42) 是否引证：否

原文内容

那么我们初试向量是(xt, yt, 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其中的纬度 (latitude) 间接影响的是温度(sst), 越靠近赤道, 温度越高, 那么我们可以将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。

如图所示便是24小时:

图4.2 台风等级播报

4.2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进行对坐标带你们进行划分, 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。做到了不止可以运用这个系统进行台风路径的播报, 还可以进行台风的等级预测。

5系统分析和设计

5.1需求分析

5.1.1系统开发目标

力求实现在卫星传来的风眼的坐标(Xt, Yt), 传入后台数据库的根据这个坐标找出温度, 湿度比, 位势高。

可以实现如下功能:

数据录入系统

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

台风路径演示系统

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

5.1.2系统的开发

硬件 (hardware) :

- (1) CPU : corei5CPU : 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

功能性需求:

(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。

5.2录入数据的图形界面(GUI)

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。

5.2.1数据训练(train):

如图以上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)。

数据录入：

图5.1 数据录入

5.2.2通过matplotlib观看轨迹：

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散点：

5.2.3 将路径数组传入ArcGis绘制散点图：

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。

图5.5 散点图

其中最后的点由于已经登陆大陆，可以酌情舍去。

5.2.4预测系统测试部分：

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

将坐标打入系统来显示出台风的坐标，来显示出轨迹。显示出登陆坐标。

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。‘山竹’如果对这种台风均难以预测，那带来的损失将难以预测如图我们将35年来号称最强台风，山竹号台风与我们的Typhoon_Predict系统来一个对比蓝色既是我们预测的，绿色是真实值，对比我们发现相差不大：

对比一下我们从国家中央气象局获得的坐标：

```
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

我们再来看一下‘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，那么我们模拟一下‘chebi’台风的路径：

图5.8 台风飞燕路径对比

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。

```
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5.9 莫拉菲台风对比

除此之外，我们再看一下东南沿海的新主角，玛莉亚 201808，对日本造成了很大影响

```
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 ]]
```

图5.9 玛莉亚台风对比

6系统使用手册

6.1系统使用

先把我们的带训练的数据写入文件中，格式是matlab形式，运行pycharm，使得程序开始运行，接着点击loading，开始加载数据，接着点击train，练习出线性参数，输入经纬度坐标，输入温度，输入压强，输入湿度比，输入坐标点击start to predict。开始训练，并调用显示地图的文件，并显示出台风预警的播报。

6.2代码文件说明

如图是所有的文件代码：

图 6.1 文件说明

图6.2文件演示

表6.1 文件功能阐述

campare_map.py通过传入路径坐标对台风预测的进行界面化的对比。

Clock.pyGUI的显示当前标准时间，便于用户对比路径的时间，提前预知登录时间的。

compare_molika.py对比台风monika来显示出本系统的优越性。

Heatmap.html保存的显示路径的html图

kit.py工具包文件，用来剔除nan的数据，使得数据处理方便

LinearCoef.csv 线性存储文件，存放我们的坐标递推方程的系数。

LinearIteration.py 线性递推的文件，用来求出坐标，然后打到散点图上

predict_Hgt.py预测位势高，调用已经训练好的LSTM模型进行对位势高预测。

predict_hgt_lstm.py 采用LSTM对位势高进行预测，我们搭建好网络，选择合适的batch_size，合适的epoch，优化器采用sgd，loss采用mse 将预测好的模型保存到好，再次调用的时候，直接采用本文件存放的模型

predict_SHUM.py 预测比湿数据，调用已经训练好的神经网络，我们对依赖时间的比湿进行预测。

predict_shum_lstm.py 采用LSTM对比湿数据进行预测，我们搭建好网络，选择合适的batch_size，合适的epoch，优化器采用sgd，loss采用mse 将预测好的模型保存到好，再次调用的时候，直接采用本文件存放的模型

predict_SST.py 调用训练好的神经网络来预测温度

predict_sst_lstm.py 采用LSTM对温度进行预测，我们搭建好网络，选择合适的batch_size，合适的epoch，优化器采用sgd，loss采用mse 将预测好的模型保存到好，再次调用的时候，直接采用本文件存放的模型

(续表6.1)

ProcessBar.py 这个文件将GUI的进度条显示出来，便于用户了解自己的数据训练(training)和加载(loading)的进度

README.md 本文件用来存放本系统的演示数据和使用说明属于使用前必读

save_hgt_model.pkl 存放的已经训练好神经网络的位势高的文件.pkl

save_shum_model.pkl 保存训练好的神经网络的比湿数据.pkl

save_sst_model.pkl 保存好的训练好的温度的神经网络数据.pkl

TrainAndSaveLinearCoefficients.py 采用机器学习(MachineLearning)的多元拟合，对给出的模型进行训练，之后存入线性存储文件csv

typhoon.gif 存放台风预测系统的头文件的图片

Heatmap_compare_chebi.html 生成对比台风'chebi'飞燕的坐标图

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.

[6] Gers F A, Schmidhuber J, Cummins F. Learning to forget: continual prediction with LSTM [C]

[7] 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.

[8] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF Models for Sequence Tagging[J]. Computer Science, 2015.

[9] Liu J, Shahroudy A, Xu D, et al. Spatio-Temporal LSTM with Trust Gates for 3D Human Action Recognition[J]. 2016.

[10] Gers F A, Schmidhuber, Jürgen, Cummins F. Learning to Forget: Continual Prediction with LSTM[J]. Neural Computation, 2000, 12(10):2451-2471.

[11] Gers F A. Learning to forget: continual prediction with LSTM[C]// 9th International Conference on Artificial Neural Networks: ICANN '99. IET, 1999.

[12] Shi X, Chen Z, Wang H, et al. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting[J]. 2015.

[13] 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.

[14] Ma X, Hovy E. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF[J]. 2016.

[15] Ordóñez Francisco, Daniel R. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition[J]. Sensors, 2016, 16(1):115-.

[16] Graves A, Jürgen Schmidhuber. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. Neural Networks, 2005, 18(5-6):602-610.

[17] 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).

[18] 高文胜, 周瑞旭, 张博文, et al. 台风登陆预测方法:.

[19] 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.

[20] 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.

[21] 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.

[22] 段昊书. 台风是如何预报的[J]. 生命与灾害, 2017(10):12-13.

- [23] 李晴岚, 兰红平, 陈训来, et al. 一种根据台风预报降雨的系统及方法:.
- [24] 耿焕同, 周利发, 陈哲, et al. 一种台风路径预测方法:.
- [25] 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.
- [26] 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.
- [27] 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
- [28] 高文胜, 周瑞旭, 张博文, et al. 台风登陆预测方法:.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] 段昊书. 台风是如何预报的[J]. 生命与灾害, 2017(10):12-13.
- [33] 李晴岚, 兰红平, 陈训来, et al. 一种根据台风预报降雨的系统及方法:.
- [34] 耿焕同, 周利发, 陈哲, et al. 一种台风路径预测方法:.
- [35] 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.
- [36] 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.
- [37] 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. 0.
- [38] 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.
- [39] 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.
- [40] 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.
- [41] 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.
- [42] 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).
- [43] 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).
- [44] 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.
- [45] 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.
- [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, 毛节奏, 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4 5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细心照料。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。

说明：1.总文字复制比：被检测论文总重合字数在总字数中所占的比例

2.去除引用文献复制比：去除系统识别为引用的文献后，计算出来的重合字数在总字数中所占的比例

3.去除本人已发表文献复制比：去除作者本人已发表文献后，计算出来的重合字数在总字数中所占的比例

4.单篇最大文字复制比：被检测文献与所有相似文献对比后，重合字数占总字数的比例最大的那一篇文献的文字复制比

5.指标是由系统根据《学术论文不端行为的界定标准》自动生成的

6.红色文字表示文字复制部分;绿色文字表示引用部分;棕灰色文字表示作者本人已发表文献部分

7.本报告单仅对您所选择比对资源范围内检测结果负责



 amlc@cnki.net

 <http://check.cnki.net/>

 <http://e.weibo.com/u/3194559873/>