

项目编号 S2017212461

# 武汉大学大学生科研项目结题报告

## 基于智能气象传感器系统监测数据的短时小 区域气象预测

院（系）名称：电子信息学院

专业名称：通信工程

学生姓名：邱子濛 张凌霄 郑睦炜  
姜 蕾

指导教师：卜方玲 副教授

二〇一八年四月

**FINAL REPORT OF UNDERGRADUATE  
SCIENCE RESEARCH PROJECT OF  
WUHAN UNIVERSITY**

**Short-term and Small-area Weather  
Forecast Based on Monitoring Data of  
Intelligent Meteorological Sensor System**

College : Electronic Information School

Subject : Communication Engineering

Name : Zimeng Qiu, Lingxiao Zhang,  
Muwei Zheng, Lei Jiang

Director : Professor Fangling Pu

April 2018

# 郑 重 申 明

本人呈交的结题报告，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，本报告的研究成果不包含他人享有著作权的内容。对本报告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。本报告的知识产权归属于培养单位。

本人签名：\_\_\_\_\_

日期：\_\_\_\_\_

# 摘 要

当前,随着中国经济社会迅速发展,政府和广大人民群众对天气预报的准确率和精细化提出了更高的要求,也对天气预报的针对性、通俗性和指导性提出了新的需求;同时,气象防灾减灾也需要对灾害性天气和气象灾害的监测、预报预警能力的提升。而现在的天气预报都集中在报道一些省、市等大范围的区域,忽略了一些局部的小区域。事实证明,并不是该省市所有地区的天气都和天气预报上的吻合,天气预报报道的是一个大范围的趋势,而很多地区的天气还是可能会与其有所差别。

此时对于小地区做出区域性的检测就尤为重要,因为大范围的天气预报在这个小区域并不适用,对人们的生产生活造成了极大的影响。因此,日晒强度和降雨概率等更具有小范围的区域特性,更适合在小范围内设网络进行特定区域的监测与预报。

本项目致力于打造更加精准的短时小区域气象信息服务。针对现阶段气象服务在小区域不准确的问题,本项目基于自主研发智能化的小区域气象监测传感器系统,分析高时间分辨率气象监测数据、大尺度天气预报数据和卫星数据,建立短时小区域的降雨和太阳辐射预测模型,实现小区域气象信息精准服务。

**关键词:** 天气; 短时; 小区域; 预测

## ABSTRACT

At present, with the rapid development of China's economy and society, the government and the general public have put forward higher requirements for the accuracy and refinement of weather forecasts, and have also put forward new demands for the pertinence, popularity, and guidance of weather forecasts; At the same time, meteorological disaster prevention and mitigation also require the monitoring of severe weather and meteorological disasters and the improvement of forecasting and early warning capabilities. Today's weather forecasts have focused on covering a large number of provinces, cities and other areas, ignoring some of the local small areas. Facts have proved that the weather in all areas of the province and city is not consistent with the weather forecast. The weather report covers a wide range of trends, and the weather in many areas may differ from it.

At this time, it is particularly important to make regional inspections for small areas, because a wide range of weather forecasts do not apply in this small area, which has a great impact on people's production and life. Therefore, the solar intensity and the rainfall probability have more regional characteristics and are more suitable for network monitoring and forecasting in a small area.

The project is dedicated to creating more accurate short-term and small regional weather information services. In view of the inaccuracy of meteorological services at this stage in small areas, this project is based on a self-developed intelligent small-area meteorological monitoring sensor system to analyze high-resolution meteorological monitoring data, large-scale weather forecast data, and satellite data to establish short-term small The regional rainfall and solar radiation prediction model enables small regional meteorological information precision services.

**Key words:** weather; short-time; small-area; forecast

# 目 录

<b>第 1 章 绪论</b> .....	1
1.1 研究背景与研究现状 .....	1
1.2 本项目研究内容 .....	4
<b>第 2 章 短时小区域太阳辐射预测模型</b> .....	6
2.1 整体概述 .....	6
2.2 基于 MLP 的短时小区域太阳辐射预测算法 .....	6
2.2.1 网络构造 .....	6
2.2.2 MLP 的前向传导 .....	7
2.2.3 MLP 的后向传导 .....	7
2.3 实验验证 .....	8
2.3.1 数据集 .....	8
2.3.2 实验结果 .....	8
<b>第 3 章 短时小区域降雨预测模型</b> .....	10
3.1 整体概述 .....	10
3.2 基于 MLP 的短时小区域降雨预测模型 .....	10
3.3 基于 SVR 的短时小区域降雨预测模型 .....	11
3.4 实验验证 .....	12
3.4.1 数据集 .....	12
3.4.2 实验结果 .....	12
<b>第 4 章 总结与展望</b> .....	15
<b>参考文献</b> .....	16
<b>致谢</b> .....	17

# 第 1 章 绪论

## 1.1 研究背景与研究现状

### 1. 研究背景

当前,随着中国经济社会迅速发展,政府和广大人民群众对天气预报的准确率和精细化提出了更高的要求,也对天气预报的针对性、通俗性和指导性提出了新的需求;同时,气象防灾减灾也需要对灾害性天气和气象灾害的监测、预报预警能力的提升。而现在的天气预报都集中在报道一些省、市等大范围的区域,忽略了一些局部的小区域。事实证明,并不是该省市所有地区的天气都和天气预报上的吻合,天气预报报道的是一个大范围的趋势,而很多地区的天气还是可能会与其有所差别。

此时对于小地区做出区域性的检测就尤为重要,因为大范围的天气预报在这个小区域并不适用,对人们的生产生活造成了极大的影响。因此,日晒强度和降雨概率等更具有小范围的区域特性,更适合在小范围内设网络进行特定区域的监测与预报。

近年来,由于平流层臭氧遭到日趋严重的破坏,地面接受的日晒辐射量增多,引起人们广泛的关注。为此,世界各国的环境科学家都提醒人们应该十分注意日晒——紫外线辐射对人体的危害并采取必要的预防措施。

当皮肤受到紫外线的照射时,人体表皮层中的黑色素细胞开始产生黑色素来吸收紫外线,以防止皮肤受到伤害。长时间的紫外线照射会引起大量黑色素沉积在表皮层中,成为永久性的“晒黑”痕迹。人们现在都已经普遍地认识到,过多地遭受紫外线辐射后容易引起皮肤癌和白内障。

现有的紫外线指数标准是指当太阳在天空中的位置最高时(一般是在中午前后,即从上午十时至下午三时的时间段里),到达地球表面的太阳光线中的紫外线辐射对人体皮肤的可能损伤程度。但这个标准无法做到实时提供当前时刻、小区域的日晒强度或者是预警 1 或 2 小时之内的紫外线辐射强度,这为人们的出行造成一定程度的困扰,早上中午和晚上日晒强度都有明显的区别,而在小区域的天气状况(如乌云、降雨等)也会影响当前的日晒强度,因此我们研究的则是当前小区域未来半小时内分钟级的日晒强度,一般市民需要根据这些状况来决定自己是否出

行或者更换自己的着装，必要的时候需要擦涂防晒品。

而大区域、大范围的降雨概率预测也会和一部分小区域的天气状况产生矛盾。武汉大学的校运会就是一个例子。天气预报称武汉市在校运会要举办的周六有雨，于是校运会延期了，但是出乎大家意料的是那个周六是大晴天。之后连续一周的阴雨天气让秋季校运会延成了冬运会，为了避免一拖再拖，武汉大学只得在一个阴雨的周六举办校运会。这样的降雨概率预报很大程度的影响了公众的出行安排问题。

总而言之，公众接触到的气象信息多为当地有关部门基于本地区某几个观测站的数据对外统一发布，此类信息对于大部分小区域并不十分准确，有时偏差甚至会很大。

## 2. 研究现状

### 1) 国内外目前气象监测技术手段时空分辨率低

我国已初步建立了地基、空基、天基观测相结合的综合观测系统，目前地基的研究和应用技术已基本成熟，并从技术研究走向业务应用，在空间探测方面，国际上很多国家和地区都实施并发射了 GPS 应用气象小卫星，美国国家极轨环境卫星（NPOES）和欧盟极轨气象卫星系列（METOP）也都将搭载星载的 GPS 接收机，用于对地的掩星探测[1]；对灾害性天气的预报预测水平和质量都有了较大提高如有研究引进能够描述气象灾害的非线性和突变性的门限回归模型，并采用基于加速遗传算法的简便通用方案，使预测性能稳定，精度提高[2]；初步建立了较完整的数值预报体系，但是也还存在着一些问题[3]，主要有：

一是局部性区域气象变化的监测能力不足。地基观测多采用气象站观测，在小区域假设成本较高，空间密集度不足，对于大气中的水汽监测误差源较多，但是基本上其应用技术已成熟；空基观测多采用有人机遥感平台与低空无人机系统的多平台协同或雷达观测且其观测范围覆盖广高垂直分辨率，天基观测多采用光学监视系统和卫星观测，但是二者对于小区域地区的气象辨识程度都不足，不利于局部性小区域（如学校，公园，一个单独的村子）的精确观测。因此，小区域的气象变化与大型观测站网络得出的大范围的区域（如整个城市）的气象变化经常有所出入。

二是软件系统相对滞后，探测产品的集成度不高。气象站等观测手段的互联网紧密度较低，一般最高也只能达到半小时更新一次，达不到需求的 15~30 分钟更新一次，使得观测信息的实时性低，观测要素需求远远不能满足。

三是突发灾害性天气事件发布渠道单一，发布窗口不畅，特别是对农村和弱势



群体的信息发布尚存在盲点，使其不能及时知晓突发性灾害天气的到来，易造成不必要的人员伤亡。气象灾害的预报时间范围较为宽泛，无法提供相关行业更为精准的灾害天气预警。

因此，通过传感器监测网络获取数据并建立短时小区域气象预测模型，在气象领域将会有重大影响。

## 2) 中国气象局提出在 2020 年将我国天气预报精确到“小区”

中国气象局于 2014 年发布了《国家气象科技创新工程(2014—2020 年)实施方案》。随着一系列气象科技攻关任务的实施，到 2020 年，中国天气预报精细化程度可能从现在的“区县级”提高到“小区级”。

《方案》提出，通过技术攻关，建立覆盖整个中国区域 1—3 千米分辨率的精细化区域数值预报模式系统，带动典型区域 1 千米分辨率区域数值模式发展。这也意味着，未来中国的气象预报不仅将精细到 1 至 3 公里范围内的某个城区，在重点城市更可能精细到方圆 1 公里范围内的某个街道甚至某个具体小区。

提高气象预报准确率和精细化水平，核心是先进的数值预报模式。记者了解到，中国自主研发的数值天气预报模式与国际先进水平相比，在精细化程度上存在差距。目前，国内预报模式的分辨率约为 50 公里，国际先进水平则在 16 至 25 公里。另一方面，国外的数值预报模式不是针对中国研发，本地适用性较差。

## 3) 短时小区域降雨预测模型发展缓慢

目前降雨量的预测方法主要有小波神经网络法、序位集对分析法、均生函数法、时间序列法、马尔科夫预测法、贝叶斯概率预报、支持向量机预测法等，而实际的降雨量数据并非纯粹呈指数的增长规律，是一种具有剧烈波动的非线性变化规律，利用单一的预测方法虽然在预测过程中取得了较好的效果，但自身难免都存在一些不足，单独使用预测精度不是很理想，误差较大。比如时间序列法在数据波动较大时预测效果较差，人工神经网络存在局部收敛、过拟合、网络层数和各层神经元数没有统一的确定标准等问题[6]，最小二乘支持向量机(Least Square Support Vector Machines, LS-SVM)继承了标准 SVM 优异的泛化性能和全局收敛能力，用等式约束代替了不等式约束，并用训练误差的平方代替了松弛变量，从而优化了 SVM 复杂的运算过程，极大地提高了训练速率[7]。

国外有学者研究使用前馈神经网络 (FFNN) 和高斯过程 (GP) 建立短时降雨预测模型，通过对比高斯过程 (GP Kernel) 和两种前馈神经网络 (Bayesian Approach

FFNN, Heuristic Approach FFNN), 发现使用 Bayesian Approach FFNN 的预测精度最高。实验证实前馈神经网络 FFNN 对于短时降雨预测也是有效的。

#### 4) 短时小区域太阳辐射预测模型不够精准

由于太阳辐射时间序列周期性的存在, 因此辐射序列本身具有高度自相关性。在同一地理纬度, 类似的太阳赤纬角、时角和天气状况下, 地面的太阳辐射量和日变化将十分类似。因此, 利用过去的辐射资料序列和天气状况要素进行辐射预测的方法是可行的。

影响太阳辐射的自然因子主要有三类, 即天文因子、大气因子和地表因子。另外, 人类活动所引起的大气成分和地表状况的改变也对辐射气候产生附加影响。

太阳辐射预测模型主要分为两个方面: 一类是瞬时太阳辐照度的模型, 即短时太阳辐射预测模型, 另一类是计算日、时太阳辐射总量的模型。

太阳辐射的预测方法主要: 传统统计、神经网络、卫星遥感和数值模拟等。

基于传统统计的方法: 利用太阳能发电量的实时和历史数据的统计预报, 它是一种统计外推方法, 相对简单但仅适用于 0-1 小时的临近预报。传统统计方法主要是以时间序列为基础, 对于较长时间的预测效果较差。

基于人工神经网络的预测方法, 利用神经网络的非线性函数逼近及自学习、自适应的特性, 将天气类型预报参数化来进行太阳能预报, 其预测时效与输入的因子和数据有关, 月预报结果具有随机性。如传统的 BP 算法或是较为新颖的遗传算法, 通过带入前一日的逐时辐射还有往年的历史数据, 来不断地训练算法, 得到较为精确的预测值。

基于卫星云图资料的外推方法, 它的优点是能处理尺度较小的对流云系统, 通过具体的云量来判断地面的辐射值, 但由于天气系统和相关云系发展移动过程的非线性, 这种方法的预报时效为 0-5 小时。

利用数值天气预报结果进行统计订正的方法, 预报时效可达数天。其优点是可分析每个大气过程, 并根据分析结果优化预测模型。国际上常用的数值模型主要有: MM5、WRF 等。目前, 晴天太阳辐射的预测多采用 HOTTEL 模型、ASHRAE 模型、REST 模型, 而有云天气多采用 Nielsen 模型、云遮修正系数模型等。

## 1.2 本项目研究内容

本项目致力于打造更加精准的短时小区域气象信息服务。针对现阶段气象服

务在小区域不准确的问题，本项目通过分析高时间分辨率气象监测数据、大尺度天气预报数据和卫星数据，基于机器学习的方法，建立短时小区域的降雨和太阳辐射预测模型。

## 第 2 章 短时小区域太阳辐射预测模型

### 2.1 整体概述

由于太阳辐射时间序列周期性的存在，因此辐射序列本身具有高度自相关性。在同一地理纬度，类似的太阳赤纬角、时角和天气状况下，地面的太阳辐射量和日变化将十分类似。因此，利用过去的辐射资料序列和天气状况要素进行辐射预测的方法是可行的。

影响太阳辐射的自然因子主要有三类，即天文因子、大气因子和地表因子。另外，人类活动所引起的大气成分和地表状况的改变也对辐射气候产生附加影响。

太阳辐射预测模型主要分为两个方面：一类是瞬时太阳辐照度的模型，即短时太阳辐射预测模型，另一类是计算日、时太阳辐射总量的模型。

我们针对目前太阳辐射预测模型存在缺陷以及太阳辐射数据时间序列周期性的特点，基于多层感知器 (Multi-layer Perception, MLP)，利用机器学习在处理原始未加工且单独不可解释的数据特征时的优势，比较得出适合我们数据的最优算法。

### 2.2 基于 MLP 的短时小区域太阳辐射预测算法

MLP (Multi-layer Perceptron)，即多层感知器，是一种前向结构的人工神经网络，映射一组输入向量到一组输出向量。MLP 可以被看做是一个有向图，由多个节点层组成，每一层全连接到下一层。除了输入节点，每个节点都是一个带有非线性激活函数的神经元（或称处理单元）。一种被称为反向传播算法的监督学习方法常被用来训练 MLP。MLP 是感知器的推广，克服了感知器不能对线性不可分数据进行识别的弱点。

#### 2.2.1 网络构造

网络分为三层，输入层 (Input Layer)，隐藏层 (Hidden Layer) 和输出层 (Output Layer)，每层之间采用全连接 (Full-connection)，除了输入层外，其余的每层激活函数均采用 sigmoid，网络结构如下图所示：

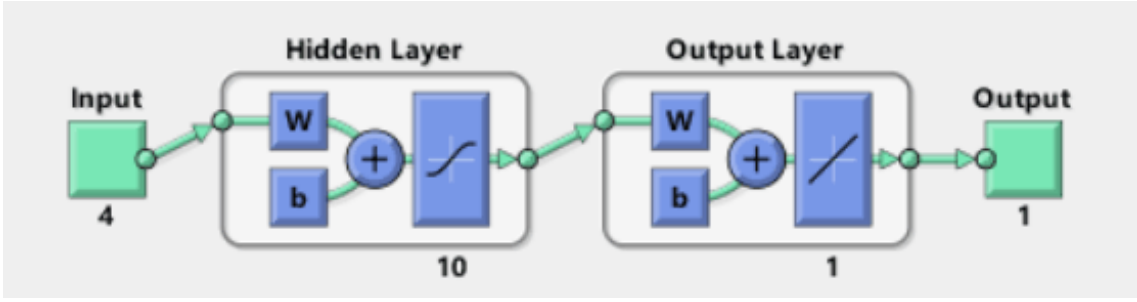


图 2.1 三层 MLP 网络架构

输入层包含四个参数，即时刻  $t$  前 4 个时刻的降雨数据，隐藏层有 10 个神经元，输入经过隐藏层和输出层以后，输出数值作为  $t$  时刻降雨的预测结果。

### 2.2.2 MLP 的前向传导

MLP 的 BP 算法基于经典的链式求导法则，首先看前向传导(Forward Pass)，对于输入层有  $I$  个单元，对于输入样本  $(x, z)$ ，隐层的输入有：

$$a_h = \sum_{i=1}^I w_{ih} x_i \quad (2.1)$$

$$b_h = f(a_h) \quad (2.2)$$

这里函数  $f$  为非线性激活函数，常见的有 *sigmoid* 或者 *tanh*，这里我们选用 *sigmoid* 作为激活函数。若有多个隐层，则计算完输入层向第一个隐层的传导以后，剩下的隐层以此类推。为了减少计算量，我们的算法只设计了一个隐层，效果已经达到要求。

隐层得到的结果传到输出层时，由于降雨量预测是个回归问题，因此我们直接输出隐层的结果，得到预测降雨量。

$$y = \sum_{h'} w_{hh'} b_{h'} \quad (2.3)$$

采用极大似然估计 (Maximum Likelihood Estimation)，构造如下函数：

$$o = \sum_{(x,z)} \frac{1}{2} (y - z)^2 \quad (2.4)$$

$$O = -\log(o) = -[\sum_{(x,z)} \log(y - z)] \quad (2.5)$$

这就是前向传导的损失函数，要寻找最优参数即对该函数求最小值。

### 2.2.3 MLP 的后向传导

有了以上前向传导的过程，接下来看误差的反向传递，对于 *sigmoid*，隐藏层输出和输入的损失函数可以表示为：

$$O = -\log(\sigma(b) - z) \quad (2.6)$$

链式求导过程如下：

$$\frac{\partial O}{\partial w_h} = \frac{\partial O}{\partial b} \cdot \frac{\partial b}{\partial w_h} = \frac{\partial O}{\partial b} \cdot a_h \quad (2.7)$$

其中，

$$\frac{\partial O}{\partial b} = -\frac{\partial \log(\sigma(b)-z)}{\partial b} = \frac{\sigma'(b)}{\sigma(b)-z} = \frac{y'}{y-z} \quad (2.8)$$

由此可以得到残差函数，接下来残差反向传递即可，传递形式形同 softmax。

## 2.3 实验验证

### 2.3.1 数据集

我们选用知名数据挖掘竞赛网站 Kaggle 提供的莫斯科四个月时间内的太阳辐射数据，每一条数据包含以下属性：“wind direction”，“wind speed”，“humidity” 和 “temperature”。

训练集和测试集选取按照 10-fold 交叉选择的原则，每次实验重复十次，每次随机按训练集 32000 条数据，测试集 682 条数据的比例划分训练集和数据集。

训练集中每一条标记数据包括 t 时刻前 4 个时刻的太阳辐射数据，t 时刻的真实太阳辐射数据作为标记，带入 MLP 中进行训练，直到误差收敛为止。然后将训练好的 MLP 网络在测试集上验证，得到实验结果。

### 2.3.2 实验结果

我们把十次实验的数据汇总，取平均值后得到如下曲线：

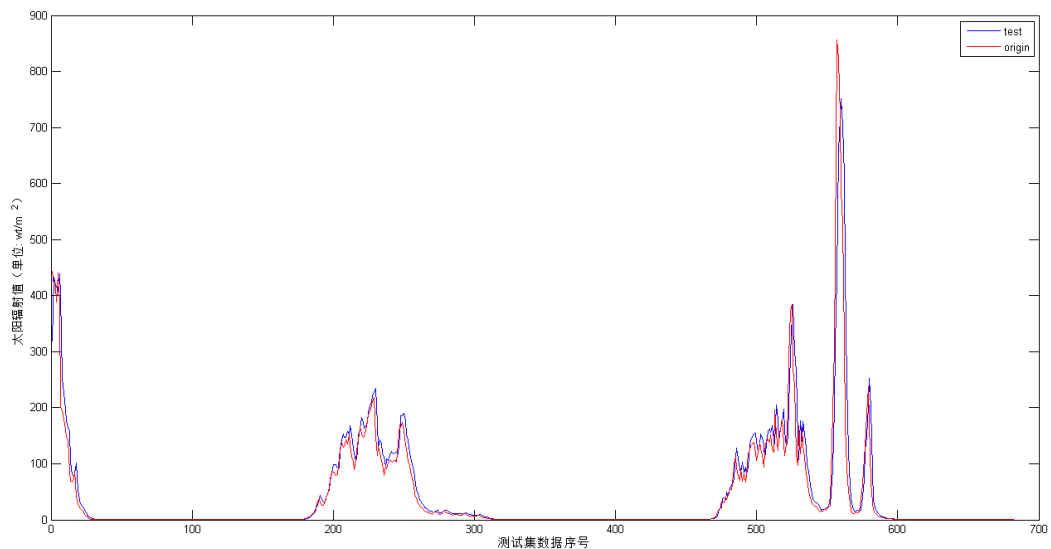


图 2.2 基于 MLP 的太阳辐射预测模型实验结果

其中红色曲线代表真实数据，蓝色曲线代表测试数据，由图可以明显看出我们

的模型具有很强的预测能力，经计算，预测平均相对误差为 0.20。这一误差水平完全可以满足需求。

## 第 3 章 短时小区域降雨预测模型

### 3.1 整体概述

目前，国内外有关降水量预测方法，主要有概率统计法、灰色预测方法、马尔科夫链、时间序列法、人工神经网络等，但预测精度差，特别是在水文中长期预测方面。提高降雨预测精度关键在于有限样本包含信息的充分挖掘，提取反映变化规律成分，再建立模型，组合预测。现实降水量并非呈指数增长规律，而是剧烈波动。现有预测方法未充分考虑降水量变化时间序列属性，无法反映时序动态特征，预测精度不高。我们针对目前降水量预测模型存在缺陷以及降水量数据非平稳性和多尺度特点，基于多层感知器 (Multi-layer Perception, MLP) 和支持向量机回归 (Support Vector Regression), 利用机器学习在处理原始未加工且单独不可解释的数据特征时的优势，比较得出适合我们数据的最优算法。

### 3.2 基于 MLP 的短时小区域降雨预测模型

类似我们在预测太阳辐射时采用的基于 MLP 的模型，这里我们同样使用 3 层全连接的神经网络，包括输入层，隐藏层和输出层。但与太阳辐射预测不同，预测降雨量的输入层有 6 个神经元，分别对应 6 个输入：前 4 单位个时间降雨量，当前时间一小时内平均风速，当前时间相对湿度；我们设计隐藏层有 100 个节点，隐藏层的激活函数为 ReLU，输出层与隐藏层的输出相同，将输出值作为降雨量的预测值。

ReLU 激活函数的灵感来源于电子电路中的二极管门限导通效应，只有达到门限以后，ReLU 函数才会有大于 0 的相应，否则都是 0 值。标准的 sigmoid 函数输出不具备稀疏性，需要用一些惩罚因子来训练出一大堆接近 0 的冗余数据来，从而产生稀疏数据，例如 L1、L1/L2 或 Student-t 作惩罚因子。因此需要进行无监督的预训练。

而 ReLU 是线性修正，是 purelin 的折线版。它的作用是如果计算出的值小于 0，就让它等于 0，否则保持原来的值不变。这是一种简单粗暴地强制某些数据为 0 的方法，然而经实践证明，训练后的网络完全具备适度的稀疏性。而且训练后的可视化效果和传统方式预训练出的效果很相似，这也说明了 ReLU 具备引导适度稀疏的能力。



因此，ReLU 的使用，使得网络可以自行引入稀疏性，同时大大地提高了训练速度。

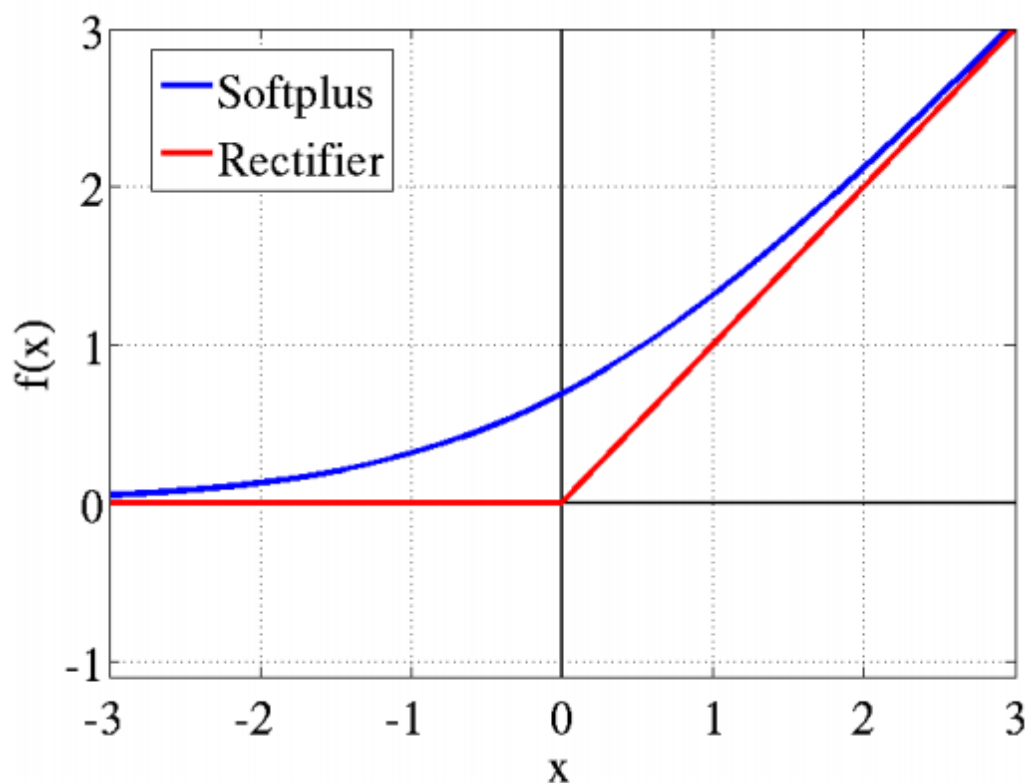


图 3.1 ReLU 函数示意图

### 3.3 基于 SVR 的短时小区域降雨预测模型

在分类问题中，支持向量机 (Support Vector Machine, SVM) 应用非常普遍，但对于回归问题，SVM 就不再适用，我们选择了 SVM 的回归算法 SVR 来预测降雨量。

回归和分类从某种意义上讲，本质上是一回事。SVM 分类，就是找到一个平面，让两个分类集合的支持向量或者所有的数据 (LSSVM) 离分类平面最远；SVR 回归，就是找到一个回归平面，让一个集合的所有数据到该平面的距离最近。

根据支持向量机二分类，数据集合归一化后，某个元素到回归平面的距离为  $r=d(x)-g(x)$ 。另外，由于数据不可能都在回归平面上，距离之和还是挺大，因此所有数据到回归平面的距离可以给定一个容忍值  $\epsilon$  防止过拟合。该参数是经验参数，需要人工给定。如果数据元素到回归平面的距离小于  $\epsilon$ ，则代价为 0。SVR 的代价函数可以表示为：

$$cost(x)=max(0, |d(x)-g(x)|-\epsilon) \quad (3.1)$$

其中  $d$  是标准答案。考虑松弛变量  $\xi_i, \xi_i^*$ , 分别代表上下边界的松弛因子。有约束条件:

$$\begin{cases} d(xi) - g(xi) < \varepsilon + \xi_i, \xi_i \geq 0 \\ g(xi) - d(xi) < \varepsilon + \xi_i^*, \xi_i^* \geq 0 \end{cases} \quad (3.2)$$

我们为了获得  $w$  的稀疏解, 且假设  $w$  的计算结果满足正态分布, 根据贝叶斯线性回归模型, 对  $w$  有 L2 范数约束。

SVR 可以转变为最优化问题, 考虑约束条件, 引入拉格朗日算子, 将最优化问题转为对偶问题, 再分别求导, 可得到最终结果。

SVR 的代价函数和 SVM 的很相似, 但是最优化的对象却不同, 对偶式有很大不同, 解法同样都是基于拉格朗日的最优化问题解法。求解这类问题的早期解法非常复杂, 后来出来很多新的较为简单的解法, 对数学和编程水平要求高, 对大部分工程学人士来说还是颇为复杂和难以实现, 因此大牛们推出了一些 SVM 库。比较出名的有 libSVM, 该库同时实现了 SVM 和 SVR。

## 3.4 实验验证

### 3.4.1 数据集

我们选用 Kaggle 中的降雨数据集作为我们的数据集, 训练集和测试集选取按照 10-fold 交叉选择的原则, 每次实验重复十次, 每次随机按 4: 1 的比例划分训练集和测试集。

训练集中每一条标记数据包括  $t$  时刻前 4 个时刻的降雨数据、当前时间一小时内平均风速和当前时间相对湿度,  $t$  时刻的真实降雨数据作为标记, 带入模型中进行训练, 直到误差收敛为止。然后将训练好的模型在测试集上验证, 得到实验结果。

### 3.4.2 实验结果

MLP 模型的实验结果如下图所示:

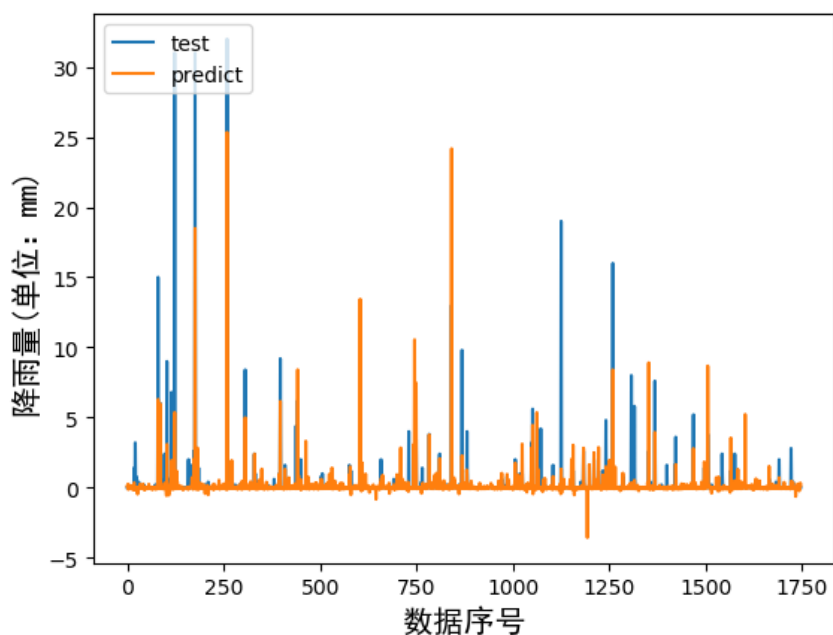


图 3.2 基于 MLP 的降雨预测模型实验结果

平均相对误差 0.13，相对降水的基数而言已经足以忽略不计。

而对于基于 SVR 的降雨预测模型，我们同时比较了 SVR 的两种核函数，Linear 和 RBF，结果如下：

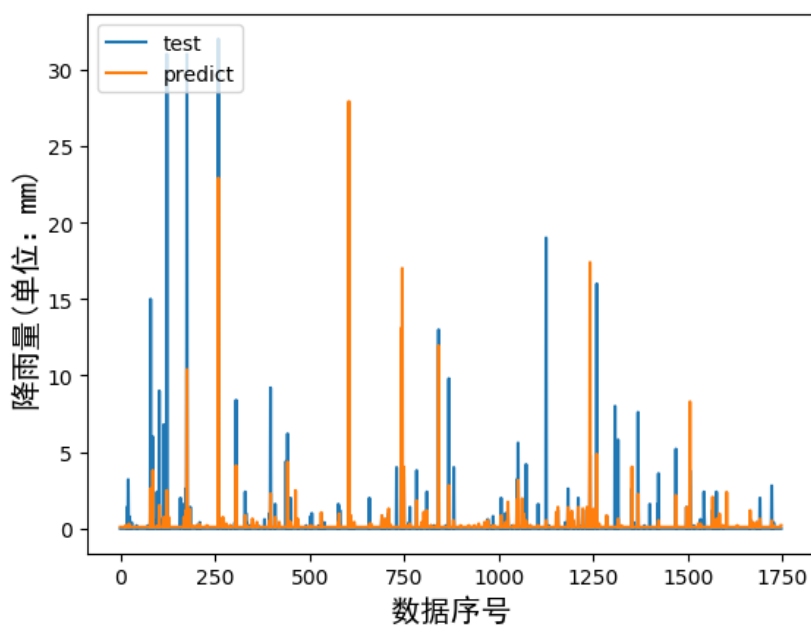


图 3.3 基于 Linear SVR 的降雨预测模型实验结果

Linear SVR 预测模型平均相对误差为 1.27。

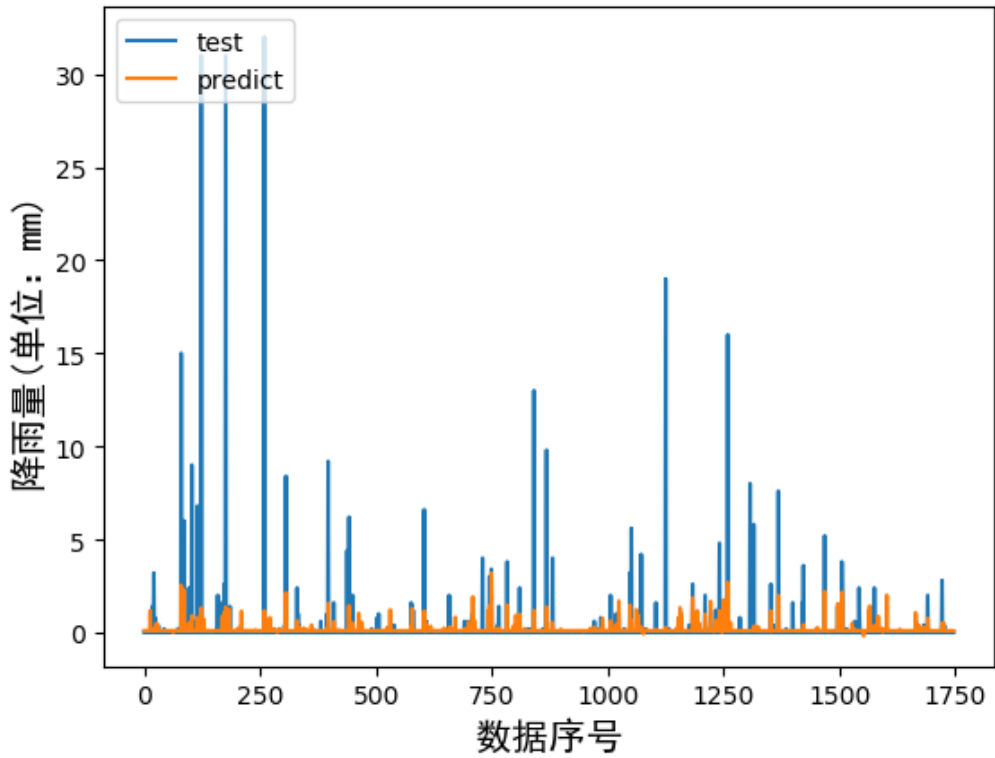


图 3.4 基于 RBF SVR 的降雨预测模型实验结果

RBF SVR 预测模型平均相对误差为 1.36。

显然，两种 SVR 的核函数中，Linear 核更适合我们的问题。但相比较而言，最优的算法还是 MLP。

## 第 4 章 总结与展望

### 4.1 项目获得的成果

表 4.1 项目成果

成果类型	名称	状态
竞赛获奖	武汉大学暑期实践大赛	校二等奖
竞赛获奖	武汉市“互联网+”大学生创新创业大赛	全国 30 强

### 4.2 研究展望

本项目与武汉大学 WSN 实验室承接的国家 863 计划、973 计划子课题成果相结合，承接小组的前一个大学生科研项目，针对现阶段气象和环境服务在小区域不准确的问题，通过自主开发的 APP 和网站，接入传感器网络监测数据并利用自主建立的算法模型对数据进行处理，实现小区域气象环境信息精准服务。

过去的一年里，我们团队提出并基本实现了短时小区域降雨和太阳辐射的预测算法，并基于 Kaggle 上的公开数据集进行了一系列实验，经检验，我们的模型具有鲁棒性和良好的准确率。

结题阶段，本项目完全达到预期目标，在结题以后的时间里，我们将继续专注于与项目相关课题，在创新上力求突破，进一步优化算法，完善实验，将论文投稿到相关会议。

## 参考文献

- [1]曹云昌, 方宗义, 夏青, 地空基 GPS 探测应用研究进展. 北京. 2004.
- [2]杨晓华, 金菊良, 魏一鸣, 预测低温冷害的门限回归模型. 灾害学, 2002, 17 (1): 10~14.
- [3]白玉洁. 改进时间序列模型在降雨量预测中的应用研究 [J]. 计算机真, 2011, (10): 141-145.
- [4]章颢. 基于 ZigBee 无线传感器网络的室内检测环境系统设计. 2013.
- [5]钱志鸿. 面向物联网的无线传感器网络综述. 2013.
- [6]李丽娟. 最小二乘支持向量机建模及预测控制算法研究 [D]. 杭州: 浙江大学, 2008.
- [7]陈冰廉, 林健玲, 林开平. 神经网络在小区域强降水预报应用研究 [J]. 计算机应用, 2006. 6, (26): 135-147.
- [8]梁山, 胡颖, 王可之, 鲜晓东. 基于无线传感器网络的山体滑坡预警系统设计 [J]. 传感技术学报, 2010 (08).
- [9]曹诗咏. 基于无线传感器网络的滑坡监测研究 [J]. 西南石油大学, 2012.
- [10]瞿雷, 刘盛德, 胡咸斌. ZigBee 技术及运用 [M]. 北京: 北京航空航天大学出版社, 2007.
- [11]刘玉华. 摄影测量系统中三维电子罗盘的设计与实现 [J]. 华中师范大学. 2013.
- [12]高效曾. 用于地层倾角仪的加速度计和磁力计 [J]. 地球物理测井. 1989.

## 致谢

在本项目将要结题之际，我们向本项目的指导老师致以诚挚的感谢，本论文是在导师卜方玲副教授的悉心指导下完成的。导师渊博的专业知识，严谨的治学态度，精益求精的工作作风，诲人不倦的高尚师德，严以律己、宽以待人的崇高风范，朴实无华、平易近人的人格魅力对我影响深远。不仅使我树立了远大的学术目标、掌握了基本的研究方法，还使我明白了许多待人接物与为人处世的道理。

同时卜方玲老师会定期聚集项目组成员进行科研进展及问题讨论例会。并安排从事相近研究方向的研究生对本项目组进行科研前的基本课程讲授和基本技能指导，使得本组成员能够较快的进入科学研究状态并动手参与工程实践。另一方面，科研导师也在项目组成员的学业培养中倾注了极大的精力，尤其是在如何把握大学生活，如何学好电子信息类专业上给予了极大的帮助。

通过项目的完成，使我能够更系统、全面地学习有关机器学习和神经网络等新型的、先进的前沿理论知识，并得以借鉴众多专家学者的宝贵经验，这对于我今后的工作和我的学业，无疑是不可多得的宝贵财富。感谢在整个项目期间和我们密切合作的同学，和曾经在各个方面给予过我帮助的伙伴们，在此，我们再一次真诚地向帮助过我们的老师和同学表示感谢！