|  |  |
| --- | --- |
|  |  |
| 2029320381452802095400 |

**基于BP神经网络的气象预测模型**

|  |  |
| --- | --- |
| 题 目： | 基于BP神经网络的气象预测模型 |
| 小组名单： | 钟祺 |
| 学 号： | 19820201153583 |
| 专业班级： | 20级电子信息2班 |
| 任课教师： | 丁兴号 |
| 日 期： | 2021/1/13 |

**基于BP神经网络的气象预测模型**

摘要

天气状态时时刻刻影响着人类的生存和发展，是所有生命活动所要考虑的因素。天气的好坏，天气的差异直接影响到人类的生产活动，同时，良好的天气条件或者合适的天气条件能给人类实践活动给予参考。但是，随着人类文明的发展，气象状况受到人类活动的影响，能够检测和预测较为准确的气象数据显得尤为重要。

近年来，随着气象数据采集技术的进步，各种探测器和收集器的出现和应用，使得气象数据得到爆炸性的增长，这就使得如何处理这些海量的气象数据成为一个问题。机器学习和数据挖掘相关技术给气象预测的研究提供了新的思路和新的途径，

本文主要以BP人工神经网络理论为依据，在考虑了多种气象数据的影响下，对厦门市近十年的气象数据构建一个基于BP神经网络的气象数据回归预测模型。本文首先介绍了相关的研究背景和意义，研究现状和本文主要内容。接着，对人工神经网络和BP神经网络模型进行了介绍，包括了BP神经网络的算法步骤。最后， 在MATLAB上实现了基于BP神经网络的气象数据回归预测模型。

**关键词**：气象数据，机器学习，数据挖掘，BP神经网络，回归预测模型

**目 录**

[第一章 绪论 4](#_Toc61531784)

[1.1 研究背景及意义 4](#_Toc61531785)

[1.2 研究现状 5](#_Toc61531786)

[1.3 本文主要内容 5](#_Toc61531787)

[第二章 相关基础理论和技术 6](#_Toc61531788)

[2.1 人工神经网络 6](#_Toc61531789)

[2.1.1人工神经网络模型 6](#_Toc61531790)

[2.1.2人工神经网络学习 7](#_Toc61531791)

[2.2BP神经网络 8](#_Toc61531792)

[2.2.1BP神经网络模型 9](#_Toc61531793)

[2.2.2BP神经网络算法 9](#_Toc61531794)

[2.3本章小结 11](#_Toc61531795)

[第三章 基于BP神经网络气象数据回归预测模型 12](#_Toc61531796)

[3.1 气象数据处理 12](#_Toc61531797)

[3.2 实验过程 12](#_Toc61531798)

[3.3 实验结果 14](#_Toc61531799)

[3.4 本章小结 16](#_Toc61531800)

[第四章 总结 17](#_Toc61531801)

[参考文献 18](#_Toc61531802)

# 第一章 绪论

## 1.1 研究背景及意义

天气条件时时刻刻影响着人类的社会活动，会对人类社会造成经济，精神上的影响。在环境多变的今天，对天气进行预测，从而具备一定应对突发情况的准备已经变得十分重要。从十七世纪开始，就有科学家使用相关的科学仪器来测量天气的状况，得到相关的气象数据，并由此数据来推测预测天气。所谓的天气预报，就是根据气象观测资料，应用天气学，动力气象学、统计学等原理和方法，在参考某一区域的气候背景和天气演变规律的基础上，对该区域未来一定时段的天气状况做出定性或者定量的预测。作为一门预测学科，由于受到各种天气系统之间的制约和相互作用影响，影响天气变化的原因因素众多，且它们之间的关系极为复杂，相互有一定的相关性，这使得气象预报存在一定程度的不确定性。随着计算机技术的快速发展，以及相关硬件的更新迭代，使得新型气象观测设备投入使用并得到推广普及，气象观测的空间密度得到进一步的提高，越来越多的信息被记录，这就使得通过这些记录进而较高准确地预测天气成为可能，数以万计的气象观测站点的地面和高空实况资料，以及卫星雷达探测资料的快速增长。这些通过气象观测设备获得的数据变得越来越具有现实价值，是十分能够体现天气状况的数据，然而实际上这些大规模的气象资料并未能够得到有效的处理和利用，以致不能在天气预报中发挥更多的积极作用。随着近些年机器学习、大数据和数据挖掘相关技术的热门，它们作为兴起的计算机技术，为充分利用这些气象数据并从中得到预期的规律性结果提供了可能性。

把机器学习和数据挖掘相关的技术应用在气象预测中，近些年也已经有在兴起了。从研究的方法方面划分，通常分为以下两种方法，第一种方法是通过使用机器学习、数据挖掘技术，对海量的历史气象数据进行统计训练学习，最终得到一个成熟的预测模型，然后使用预测模型对未来一段时间天气进行预测，这个过程不需要气象相关的专家参与学习过程。这类算法的常用技术有贝叶斯网络、神经网络、SVM分类器、决策树等等。第二种方法是在气象专家总结出来的气象规则的基础上构建一个完整的知识推理系统，这个方法虽然在一定程度上能够预测气象状况，但是在处理大量的数据时的预测准确度还存在着各种各样的问题，毕竟大量数据中存在着许许多多相关关系，是不可能被气象专家全部发掘发现的。

反向传播（Backpropagation，缩写为BP）是“误差反向传播”的简称，是一种与最优化方法（如梯度下降法）结合使用的，用来训练人工神经网络的常见方法。该方法对网络中所有权重计算损失函数的梯度。这个梯度会反馈给最优化方法，用来更新权值以最小化损失函数。反向传播要求有对每个输入值想得到的已知输出，来计算损失函数梯度。因此，它通常被认为是一种监督式学习方法。反向传播要求人工神经元（或“节点”）的激励函数可微。本文将应用BP神经网络对大量气象数据进行训练，从而得到一种气象预测模型。

## 1.2 研究现状

近年来，在理论水平和科技水平迅速发展的同时，气象领域也取得了惊人的成果。随着新型气象观测手段的发展，气象信息化程度也越来越高，气象数据资料的存储量也在不断增加，如何有效地组织、使用这些海里数据是众多气象工作者目前面临的一个严峻考验。同时也是预警灾害天气和提高预测准确率的关键所在。现如今气象预测的方法主要有三种，分为数值预测方法、天气学预报法和统计预报法等。国内外研究学者将数据挖掘应用到气象预测中，并取得了丰硕的成果。目前应用于气象预报的数据挖掘方法主要有关联规则挖掘、决策树、人工神经网络、遗传算法、贝叶斯和支持向量机等等。

比如哈尔滨工业大学的杨函[1]在基于深度学习下对气象预测进行研究，Gordon Reikard[2]利用神经网络和频域算法构建了一个气象预测模型，其预测的准确性有着显著的提高。Siva Venkadesh[3]建立了以遗传算法确定作为神经网络模型输入的每个环境变量的最佳时段和分辨率，改进预测效果。

相关研究成果表面了将机器学习相关技术应用在气象预测领域是有可能的，并且有效性有所保证。

## 1.3 本文主要内容

本文把BP神经网络应用在分析气象数据上，形成一种基于BP神经网络的气象预测模型，数据来源为开放的气象数据网站。

第一章，为绪论。首先阐述了研究内容的背景及其意义，介绍了相关的研究进展，并分析其不同。最后对本文主要内容进行介绍。

第二章，主要是介绍了本文应用的相关基础理论和技术。从人工神经网络和BP神经网络两大方面进行阐述。

第三章，为实验部分，应用BP神经网络构成一种气象预测回归模型，将厦门近十年的数据分为训练集和测试集，得到最后的对比数据图，预测的气象数据在趋势上大致符合，有良好的预测性。

第四章，为总结部分，对全文的内容进行总结概述，并思考了几处需要改进的地方。

# 第二章 相关基础理论和技术

## 2.1 人工神经网络

神经网络（Neural Network, NN）,又称作人工神经网络（Artificial Neural Network， ANN）。相当多的单元相互连接构成了这样一种网络，进而，通过抽象、简化和模拟人脑的活动来反映人脑的一些基本特性，并不是真实描述人脑神经系统，而是经过模仿这样的生物过程从而实现能够表现一定人脑特性的目的。研究神经网络模型，其实就是从人的生理结构角度看问题，探讨人的一些智能行为，以及实现模拟这样处理信息的能力和功能[4]。神经网络具备许多优点，如知识的分布存储、固有的并行结构、强大的非线性、优良的容错能力和自组织、自学习、自适应等。综上所述，神经网络在非线性映射、联想记忆、优化计算以及分类识别等许许多多方面都取得了很好的应用效果。

### 人工神经网络模型

人工神经网络模型中的基础就是人工神经元，其是抽象和模仿生物神经元结构和的功能，是形式化下的概述描述处理信息过程。作为人工神经网络的基本处理单元，人工神经元的功能是:对每个输入的信号进行处理以确定其强度，即加权。确定所有的输入信号的组合效果，即求和。确定其输出，又称为转移特性。它一般是一个多输入或单输出的非线性器件，其结构模型如图2所示。如同生物神经元有许多树突输入一样，人工神经元也有很多输入信号，并同时作用到人工神经元上，生物神经元中大量的突触具有不同的性质和强度，使得不同的输入的激励作用各不相同，因此在人工神经元中，对每一个输入都有一个可变的加权，用于模拟生物神经元中突触的不同连接强度及突触的可变传递特性，生物神经元的时空整合功能，使得人工神经元必须对所有的输入进行累加求和来全部输入作用的总效果，该和类似与生物神经元的膜电位；在生物神经元中，只有在膜电位超过动作电位的阈值时，生物神经元才能产生神经冲动，反之则不能，因此在人工神经元中，也必须考虑该动作的电位阈值，与生物神经元一样，人工神经元只有一个轴突输出，同时，由于生物神经元的膜电位与神经脉冲冲动之间存在这一种数模转换关系，因此在人工神经元中要考虑输入与输出之间的非线性关系。

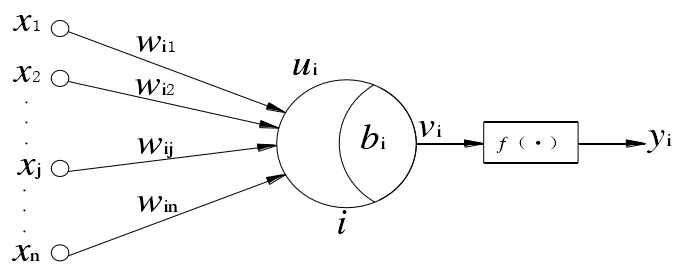


图2-1 人工神经元模型

设神经元的输入信号所对应的向量为

(2-1)

其中，为神经元在神经网络所对应的位置序号，n表示输入神经元的信号个数，为第个神经元的输入信号。连接输入神经元节点到神经元节点j的加权向量为

(2-2)

其中为输入神经元从第个节点到第个节点的加权值，也可以称作连接强度。

进而，可以得到神经元的输入加权和u为

(2-3)

其中为线性组合输入向量后得到的输出向量，作为净输入传递到神经元。在这样的线性组合后，一般还得调整偏差得到局部的感应区。可以认为是在这个基础上，加上了一个权值为1的偏差，如下式子

(2-4)

其中，为神经元的偏差，也叫阈值。为局部感应区。

神经元在输入向量经过转移函数的激励后，才能够得到最终的处理单元的最终输出值，神经元的输出信号如下

(2-5)

其中，为神经元的输出信号。F为转移函数，描述了神经元的输入向量和输出向量的某种非线性关系。神经网络模型的基本处理单元就是人工神经元，它的作用体现在要处理每个输入信号，来确定相应的强度，也就是加权值。从而计算全部输入向量组合在一起后得到的结果，也就是进行求和。最后，可以计算神经元的终端输出信号，也就是转移特性。

### 人工神经网络学习

仅仅具备拓扑结构这对一个人工神经网络模型来说是远远不够的，还必须具有一套完整的工作和学习规则相匹配，才能够具备智能的特性。神经网络的学习也叫训练，就像人类学习知识和事物一样，在获取信息后，要加以实践和训练，从而完善知识储备和认知。学习是人工神经网络的重要特征之一。神经网络必须先通过训练，从而学习到要做的工作，进而才能具备某些特性。在神经网络模型中，训练的本质就是对网络连接权值进行调整的过程，实际上的神经网络也可以认为是拟合一种函数映射关系的过程，训练的完成，也就是完成了对连接权值的调整，从而学习到的知识就会分布存储到网络上的各个连接权上。

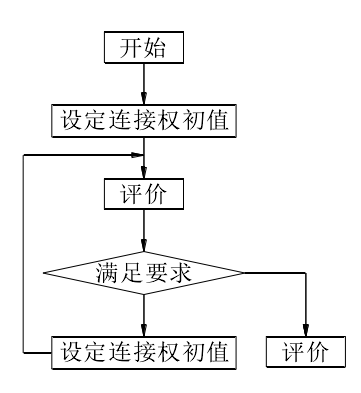


图2-2 人工神经网络学习过程

如上图所示，该流程图展示了人工神经网络的学习过程。在网络的训练过程中，无论采用哪种的学习算法和规则去调整网络连接权值，在调整之前，都需要评价人工神经网络的训练结果，也就是对训练的效果进行评价，从而可以根据这个分数进而调整训练方案。评价的标准可以由外部提供，也可以由人工神经网络自身来提供。由此可以把人工神经网络的学习方式分为两类：有监督学习和无监督学习。

有监督学习也可以叫有导师指导下的学习，它指的是在网络学习训练过程中，必须要有一个对网络的学习训练结果的正确性的评价标准，也就是训练的结果需要给它一个是否符合答案的判断，也就是这时正确答案是确定的。因此需要不断地在网络学习的过程中为神经网络提供输入和期望输出数据对，通过期望输出与实际输出的差异来确定其数值和方向，以此来调整连接权值，其目的是为了让网络的实际输出结果和期望输出结果想接近。在学习过程中，外界把期望输出作为评价标准提供给网络。这就相当于有一位明确知道正确答案的导师给予了网络指导，所以也叫做有导师指导下的学习。

同样的，无监督学习也叫无导师指导下的学习，它是指在人工神经网络的学习过程中，没有来自外来的对于网络的学习结果的正确性的评价标准，换句话说，这时没有所谓的正确答案，正确的答案需要通过网络的自我训练得到。因此在网络学习的过程中需要不停地向提供动态的输入信号，神经网络根据其特有的结构及学习规则，来寻找可能存在的模式和规律，自动分类，同时，根据网络的输入信号与功能自动调整连接权值。在无监督学习中，如果网络外部没有导师的指导示范教学，就可以认为网络的学习评价标准在网络的内容中隐含。所以这种学习方式也叫做无导师指导的学习。

## 2.2BP神经网络

反向传播（Backpropagation，缩写为BP）是误差反向传播的简称，是一种与最优化方法，比如梯度下降法，结合使用的，用来训练人工神经网络的常见方法[5]。该方法对网络中所有权重计算损失函数的梯度。这个梯度会反馈给最优化方法，用来更新权值以最小化损失函数。反向传播要求有对每个输入值想得到的已知输出，来计算损失函数梯度。因此，它通常被认为是一种监督式学习方法。反向传播要求人工神经元，或者说是节点的激励函数可微。

### 2.2.1BP神经网络模型

BP神经网络是一种多层前馈神经网络，主要特点是信号前向传递，误差反向传播。在前向传递中，输入信号从输入层经隐含层逐层处理，直至输出层。每一层的神经元状态只影响下一层神经元状态。如果输出层得不到期望输出，则转入反向传播，根据预测误差调整网络权值和阈值，从而使BP神经网络预测输出不断逼近期望输出。

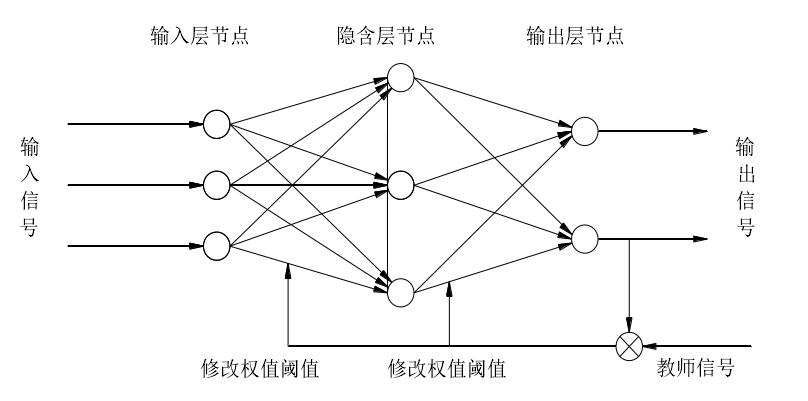


图2-3 三层BP神经网络的拓扑结构

如上图就为三层BP神经网络的拓扑结构，随着误差的反向传播会不停地调整修正BP神经网络，使之能够逐渐增大对输入信号的识别的准确率。所以，BP神经网络模型运用的这类学习算法就被人们成为误差方向传播算法，在这种学习方法中误差函数值根据梯度不断减低。实际上，这是一个回归的过程，并且，这是个前向网络，之所以的反向，指的是误差的反向，误差从结果反推给前面各个层的神经元，从而使它们根据误差权重的分配来调整它们各自的权值，进而实现减少最终误差的目的。

### 2.2.2BP神经网络算法

BP神经网络算法分为正向传播和反向传播两步进行，其步骤过程为：正向传播时，样本从输入层经过全部隐含层传向输出层，在输出层对当前输出和期望输出进行比较，如果满足误差要求，则计算结束，如果不满足，则进入反向传播过程：这里指的是误差的反向传播，把误差信号反向传回，并对每层的各个神经元的权值和阈值进行修改，以期望误差趋于最小。

举例，BP神经网络有N个输入层单元，M个输出层单元，H个隐含层单元；i表示第i个输入层神经元，i=1,2,…,N; h表示第h个隐含神经元，h=1,2,…H; j表示第j个输出层神经元，j=1,2,…M; K为数据样本的个数，k=1,2,…,K;

为BP神经网络的输入向量；

为BP神经网络的实际输出向量；

为BP神经网络的期望输出向量；

为输入层第i个神经元到隐含层第h个神经元的连接权值；

为隐含层第h个神经元的阈值；

为隐含层第h个神经元到输出层第j个神经元的连接权值；

为输出层第j个神经元的阈值；

为BP神经网络的学习速率；

为BP神经网络的误差要求；

1. 选取一组输入样本和目标样本提供给网络；
2. 正向计算

输入层第i个神经元的输出为

(2-6)

隐含层第h个神经元的输入为

(2-7)

隐含层第h个神经元的输出为

(2-8)

1. 计算输出层第j个神经元的输入为

(2-9)

输出层第j个神经元的真实输出为

(2-10)

1. 计算该样本的输出误差

(2-11)

1. 修正输出层、隐含层的所有权值和阈值

(2-12)

(2-13)

(2-14)

(2-15)

其中，为输出层神经元的局部梯度，隐含层神经元的局部梯度。

（6） 判断是否训练完所有的样本，若不满足则选取下一个学习样本提供给网络，返回到步骤（2），若满足则转至步骤（7）。

（7） 计算BP神经网络的总误差

(2-16)

判断误差E是否满足，若满足则结束训练，若不满足则转向步骤（8）。

（8） 判断是否达到预定的训练次数，若满足则结束训练，若不满足则返回到步骤（2）继续进行训练。

## 2.3本章小结

本章主要介绍了人工神经网络模型和BP神经网络模型，包括了人工神经网络模型的基本理论知识和学习方法。重点介绍了在之后建立气象数据回归模型所要用到的BP神经网络模型的结构和算法步骤。

# 第三章 基于BP神经网络的气象数据回归预测模型

本章将在前几章介绍的基础上，在MATLAB上实现基于BP神经网络的气象数据回归预测模型。

## 3.1 气象数据处理

气象的数据来源是rp5.ru气象数据库网站，从中获得了大约近十年厦门的气象数据情况，对于年份较早的数据网站大多数都是一天四次每隔六个小时收集补充到网站的数据库中，年份较近的数据是隔三个小时记录一次收集到网站的气象数据中。其中的气象数据包括了：地面以上2米处的大气温度T、气象站水平的大气压Po、平均海平面的大气压P、地面高度2米处的相对湿度U、总云量N、水平能见度VV、地面高度2米处的露点温度Td、降水量RRR等等。为了尽可能的准确预测这些天气情况，就需要综合考虑这些气象数据的相关指标。第一，在气象数据中，存在一些数据的缺失值，这可能是由于各种复杂的原因导致的，那么该如何去处理这些缺失值。在这里，处理的方式是取前后相邻值的平均值，在缺失值较少时，为人工手动在文件数据中处理，较多时通过编程方式解决。第二，对于不同维度的数据，这些数据存在量纲的差异，那么就应该考虑进行标准化来减小量纲对结果产生的影响。第三，在对气象数据的回归进行预测时，考虑到气象数据各维度之间通常具有很强的依赖性和关联性，这时就应该要进行维度的约减，也就是降维的处理，来减少对预测结果的影响。第四，气象数据处理时，可能不同类型数据跨度有差异，这时就需要对数据进行归一化处理。总结如下表所示。

|  |  |
| --- | --- |
| 数据问题 | 处理方法 |
| 数据的缺失值 | 取相邻数据的平均值填充 |
| 不同维度数据的量纲差异 | 对数据进行标准化 |
| 各维度数据的依赖性和关联性 | 降维处理 |
| 不同数据量的跨度差异 | 归一化处理 |

表3-1 数据的处理方法

## 3.2 实验过程

实验在Windows操作系统下，通过MATLAB仿真软件和Deep Learning Toolbox等相关的工具箱构建神经网络预测回归模型。

为了验证所提出的气象数据预测回归模型的有效性，本文借助相关的数据库中的数据进行训练和测试。实验中的数据来自rp5.ru网站中关于厦门2010至2020年间的气象数据。所收集到的气象数据有地面以上2米处的大气温度T、气象站水平的大气压Po、平均海平面的大气压P、地面高度2米处的相对湿度U、总云量N、水平能见度VV、地面高度2米处的露点温度Td、降水量RRR等等。这里只选取其中的T、Po、P、U、VV和Td这六个比较关键的气象数据进行神经网络的构建，将它们作为输入。程序流程图如下。

载入xls数据并进行数据集的划分。

对数据进行初始化。

如缺失值处理。

构造时间函数曲线和较为相关的两两属性关系图。

划分训练集和测试集。

对数据进行降维，剔除贡献度低的属性。

构造神经网络模型。并设置相关参数。

BP神经网络训练。

仿真测试。并绘图。

图3-1程序流程图

## 3.3 实验结果

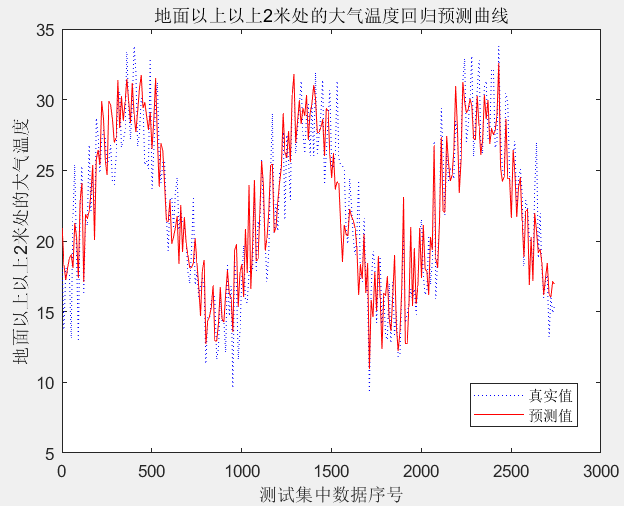


图3-2 地面以上以上2米处的大气温度回归预测曲线

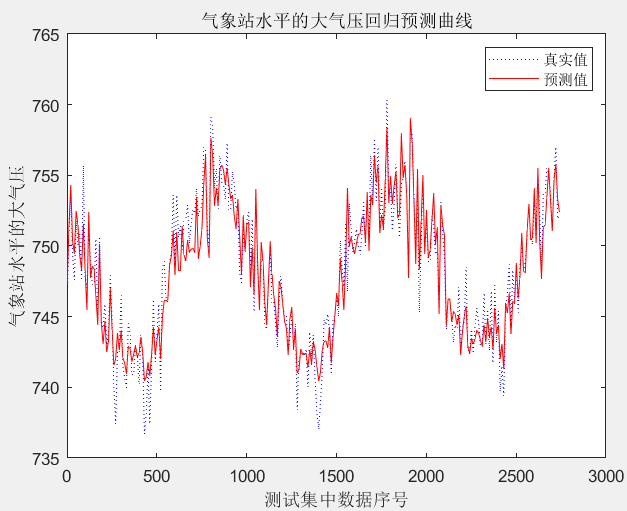


图3-3 气象站水平的大气压回归预测曲线

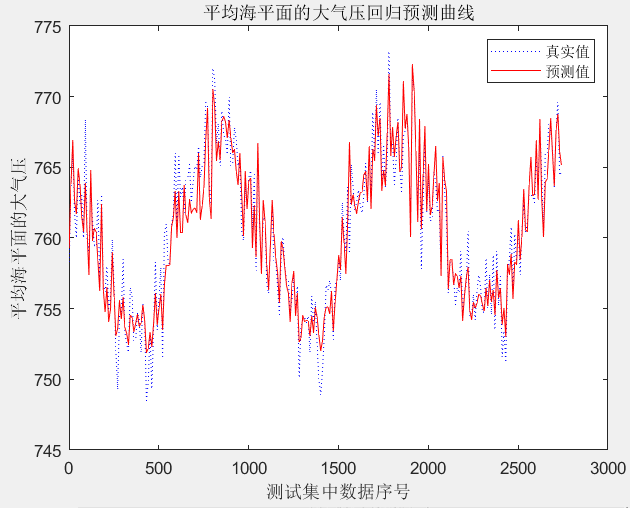


图3-4 平均海平面的大气压回归预测曲线

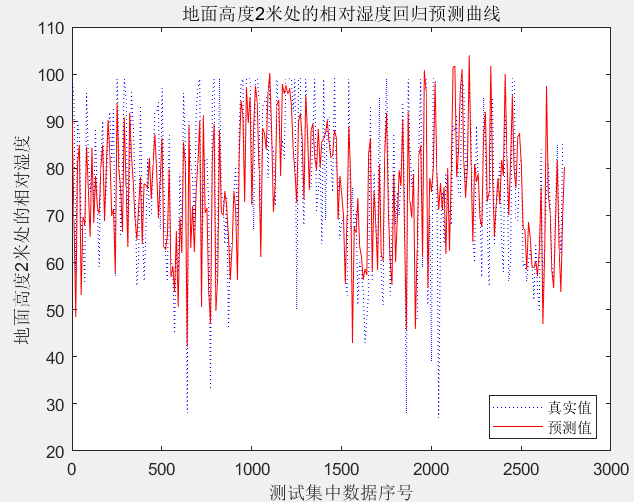


图3-5 地面高度2米处的相对湿度回归预测曲线

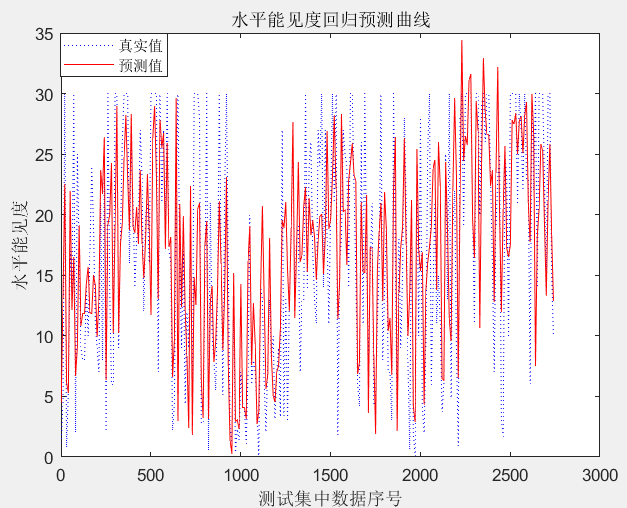


图3-6 水平能见度回归预测曲线

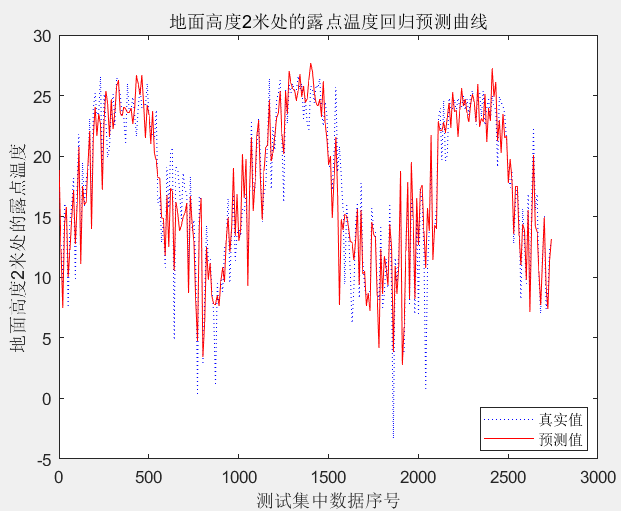


图3-7 地面高度2米处的露点温度回归预测曲线

本文实验环境为MATLAB R2020b，通过收集到的厦门气象数据，实验结果如上图所示，以上展示了通过基于BP神经网络建立的气象数据回归预测模型得到的回归预测曲线。根据结果可以知道，构建一个前馈性的BP网络结构，由助于搭建天气气象数据预测研究，能符合预期实现结果。对于六类气象属性，在实验中，实现了曲线变化的预测，并且在测试集的验证中，有良好的重合度。

## 3.4 本章小结

本章从气象数据库收集厦门近十年的气象数据，对其建立基于BP神经网络的气象数据预测回归模型。综合考虑了气象数据中的一些相关指标，解决了一些潜在问题，如数据缺失值的影响，不同维度数据量纲之间差异的影响，气象数据各维度之间极强依赖性和关联性的影响等等。在预测天气状况建模的过程中，运用各类数据进行分析处理，建立合适的模型，使得预测结果更加准确，最终得到了六类属性的回归预测曲线，与测试集相比，变化趋势近似，有可靠的预测性。

# 第四章 总结

本文提出了一种基于BP神经网络的气象数据预测回归模型。根据开放网站的数据库，结合数据库内有关厦门近十年的气象数据，利用BP神经网络算法完成对气象数据情况的预测。

在本文的内容中，首先简要叙述了气象预测研究的背景，相关研究进展和本文内容框架，简述了本文研究的意义和实用价值，证明了BP神经网络应用在气象数据预测回归上的可行性。然后分析了人工神经网络和BP神经网络相关知识，重点介绍了BP神经网络的算法步骤。接着对实验数据的来源和处理进行了介绍，给出了程序的主要流程图，通过MATLAB和其有关深度学习的工具箱构建了基于BP神经网络的气象数据预测回归模型，对厦门近十年的数据集进行处理预测，在结果证明了该模型能大致预测出相关气象数据的变化曲线，实现了最初的目标任务。不过，也有一些不足之处，如预测准确性还有待提高，最终评估准确性的标准应该用一些数据指标来体现，如均方根误差、平均绝对误差和平均绝对百分误差等等。

# 参考文献

[1]杨函. 基于深度学习的气象预测研究[D].哈尔滨工业大学,2017.

[2] Gordon Reikard. Combining frequency and time domain models to forecast space weather[J]. Advances in Space Research, 2013(52):622-632.

[3] Venkadesh S, Hoogenboom G, PotterW, et al. A genetic algorithm to refine input data selection for selection for air temperature using artificial neural networks[J]. Applied Soft Computing, 2013 13(5):2253-2260.

[4]艾洪福,石莹.基于BP人工神经网络的雾霾天气预测研究[J].计算机仿真,2015,32(01):402-405+415.

[5] 任谢楠. 基于遗传算法的BP神经网络的优化研究及MATLAB仿真[D].天津师范大学,2014.

[6] 刘鑫达. 基于深度学习的气象温度预测研究[D].宁夏大学,2016.