

基于神经网络回归模型的降雨量问题研究

李沐阳 易领程 钟绍恒

2024 年 5 月 2 日

目录

1 摘要	3
2 问题重述	4
2.1 问题背景	4
3 模型假设和符号说明	4
4 模型建立与求解	4
5 模型的优缺点与改进方法	5
6 参考文献	5
附录	5
.1 源代码	5
.2 数据来源	5

1 摘要

本文针对局部地区的降雨量与其它气象相关的因素进行了数学模型的建立和求解的通用算法的设计,运用**线性回归**,**障碍计数**,**时间序列**和**神经网络**等模型分别进行了尝试,并且使用**遗传算法**优化了模型的求解。

我们首先针对与降水量关联度最大的**气压**和**气温**等条件,将数据按照 30 天为一组进行平均处理整理后,根据气象学上的基本原理,运用线性回归模型拟合真实数据,完成了该简单模型的求解,取得了**正确率 76%** 的良好的预测效果,在这里同时使用了**最小二乘法**和**遗传算法**进行拟合,均能够取得很好的效果。在此时还尝试了**障碍计数模型**来应对未降雨的情况,但尝试后发现准确率极低,只有 10% 左右,整理数据后分析得到数据的趋势具有连续变量的特征,趋势上大致更加符合回归模型这一类。

随后扩大研究的范围,添加了**云层覆盖**,**湿度**,**风速**为新的自变量,由于此时线性回归在多个自变量的相互影响下适应性较差,转变模型,根据深度学习的基本原理和神经网络中防止过拟合的技术,运用了神经网络回归模型进行拟合,在仅使用 4 个神经层,共 200 多个神经元的情况下取得了**绝对平均误差仅为 9mm** 的优异结果。同时,我们也再次展开了遗传算法的尝试,在仅有 4 层共 10 个神经元的网络中取得了 50% 的正确率,由于遗传算法在一定程度上的局限无法将神经元数量扩充到太大,但依旧提供了一种解决问题的新思路。

在数据具有明显季节性质,且相邻的日期可能相互影响的情况下,我们也对**时间序列模型**进行了一定的尝试和研究,最后得到的效果也不如神经网络回归模型,原因可能是相互影响因素已经体现在了自变量中。经过进一步的分析,由于数据间并非存在明显的非线性关系,排除了**决策树模型**的可能性。尝试和分析过后,最后认为该神经网络模型更加准确。

该模型具有预测率较高,和可以轻易扩展到新的自变量的优点。生成该模型的方法可以轻易地被用于生成**任何局部降雨模型**,非常简单易行。在考虑到气象的周期性而忽略人为影响的情况下,本模型具有较为准确的预测能力。

关键词:神经网络回归模型 线性回归模型 遗传算法

2 问题重述

2.1 问题背景

在过去的几十年中，随着各行各业的迅速发展，气象对于人类生产生活的影响日益显著，对于气象预测的精确程度需要愈来愈高，降雨作为人们出行和农业生产的客观需要是非常关键的一部分。在当今时代，由于人类活动的影响日益增大和气象全球性变化的影响下，气象的预测越来越有必要与时俱进，因地制宜的适应局部气候的变化。

3 模型假设和符号说明

该模型假设气象活动完全是周期性的，不考虑气候的逐渐变化，也不考虑人为因素的气候的干扰。

表 1: 符号说明

符号	说明
r	降雨量
y	模型的因变量

4 模型建立与求解

首先，从附录 2 中获取关于 11 号气象站 (一个奥地利气象站) 中的相关数据，考虑到尽可能减少人类活动的影响，本文主要研究该地区从 1877 年到 1917 的局部降水量。首先研究降水量与气压和温度的关系，获取完数据之后，按照 30 天为一组进行平均，画出各个自变量的图像如下

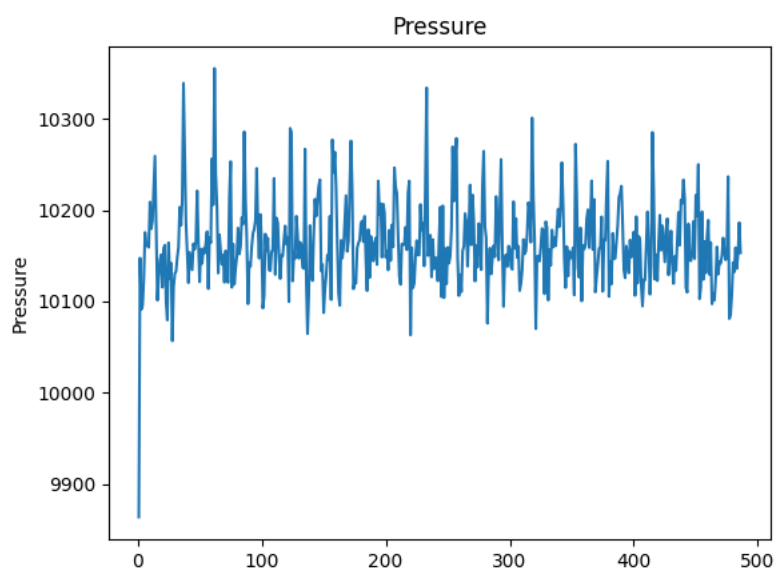


图 1: 气压覆盖关于时间的折线图

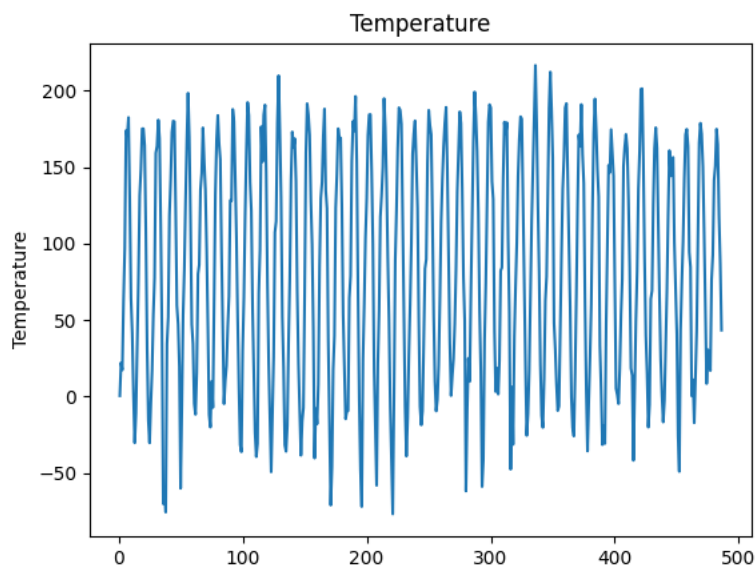


图 2: 气温关于时间的折线图

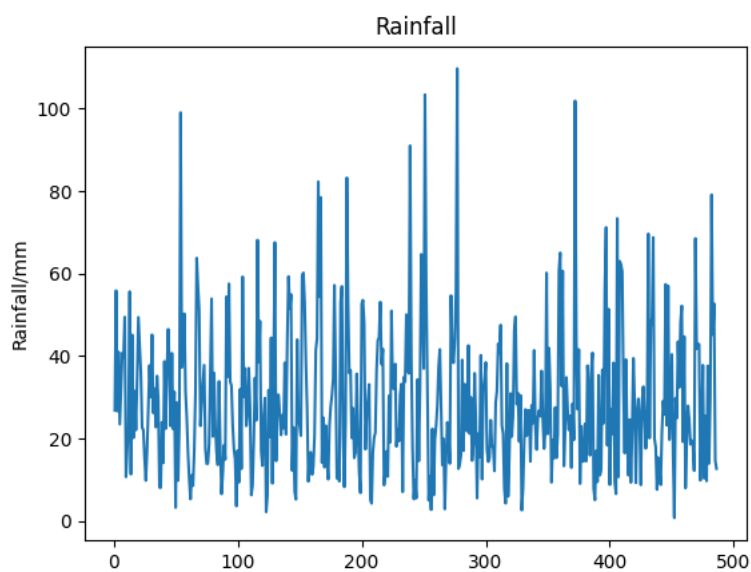


图 3: 降水量关于时间的折线图

简单研究,发现这 3 个量之间很可能呈线性关系,设降雨量为 y , 气压为 a , 温度为 b , 建立线性回归模型 $y = pa + qb$, 使用最小二乘法让程序生成对应的模型 (该程序见附录 3), 从而拟合最多的数据, 得到结果为 $y =$, 画出拟合图像与真实图像:

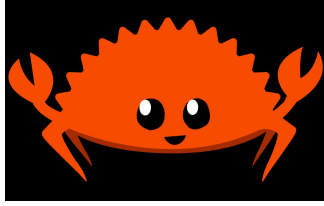


图 4: 拟合降水量和真是降水量的折线统计图

5 模型的优缺点与改进方法

6 参考文献

附录

.1 源代码

.2 数据来源

European Climate Assessment Dataset