

基于神经网络的雾霾预警系统研究与实现

黄 威,王星捷,阳青青

(成都理工大学 工程技术学院,四川 乐山 614007)

摘 要:传统的雾霾预警系统缺乏预知性,无法实现提前预警以规划预防,其次雾霾物理结构异常复杂且具有强烈非线性的结构、预测难度较高。为了更好地反映雾霾在时间及空间的分布状况,为预防工作提供充足的时间准备,实现了一种预测与报警模型相结合的预警系统。基于BP神经网络算法,对乐山市未来5天各站点PM_{2.5}含量进行了高精度的预测,并设计了各县及各城区预警的功能模块,对各县预警专题图、空间插值分布图及相关数据信息进行展示。采用插值算法对市中区各城区监测点进行数据离散化处理,生成预警专题图及相关预测数据可视化。系统以乐山市的PM_{2.5}的监测站实时监测的数据为源数据,通过对乐山监测点历史PM_{2.5}数据进行实验预测,并与实际值进行对比分析,达到了较好的预测效果。建立的预警模型具有一定的实用和研究价值,可为相关部门提供可靠的数据支撑。

关键词:神经网络;雾霾预警系统;雾霾预测;预警专题图;插值算法

中图分类号:TP302.1

文献标识码:A

文章编号:1673-629X(2019)10-0026-05

doi:10.3969/j.issn.1673-629X.2019.10.006

Research and Implementation of Haze Early Warning System Based on Neural Network

HUANG Wei, WANG Xing-jie, YANG Qing-qing

(Engineering Technical College, Chengdu University of Technology, Leshan 614007, China)

Abstract: The traditional haze warning system lacks predictability and cannot realize early warning for planning and prevention. Secondly, the physical structure of haze is extremely complex and has a strong nonlinear structure, which is difficult to predict. In order to better reflect the distribution of haze in time and space and provide sufficient time preparation for prevention work, an early warning system combining prediction and alarm models is implemented. Based on the BP neural network algorithm, we have realized the high-precision prediction of PM_{2.5} content in each site in the next 5 days of Leshan City, and designed the functional modules of the early warning of each county and each city, and displayed the early warning thematic map and spatial interpolation distribution of each county, and related data information. The interpolation algorithm is used to discretize the data of the monitoring points in the urban areas of the city, and the early warning thematic map and related prediction data are generated. The system takes the real-time monitoring data of the monitoring station of Leshan City PM_{2.5} as the source data. Through the experimental prediction of the historical PM_{2.5} data of the Leshan monitoring point, and the comparison and analysis with the actual value, a better prediction effect is achieved. The early warning model established has certain practical and research value, and can provide reliable data support for relevant departments.

Key words: neural networks; haze warning system; haze prediction; early warning thematic map; interpolation algorithm

0 引 言

PM_{2.5}^[1-2]是指大气中直径小于等于2.5微米的可入肺颗粒物,其对空气的质量和能见度有着重要的影响。大量研究表明,大气中的PM_{2.5}浓度与心血管疾病和肺癌的发病率存在一定的相关性。近年来,随

着工厂的发展和机动车的增加以及城镇化节奏,大气污染问题的加剧,研究PM_{2.5}的预测模型^[3]有着重要的意义。

在预测方面,主要有时间序列、遗传算法、小波神经网络、线性回归、深度学习、人工鱼群、贝叶斯、灰色

收稿日期:2018-12-09

修回日期:2019-04-10

网络出版时间:2019-06-26

基金项目:2015年度四川省教育科研项目(15ZB0366);2018年度乐山市重点研究项目(18JZD051);2018年度学院苗子工程项目(2018MK09)

作者简介:黄 威(1997-),男,研究方向为地理信息系统、机器学习;王星捷,硕士,副教授,通信作者,研究方向为数字城市、智慧城市、三维空间数据处理、大数据应用。

网络出版地址:<http://kns.cnki.net/kcms/detail/61.1450.TP.20190626.0833.054.html>

理论、基于数据挖掘、模糊网络、支持向量机、随机森林等^[4-6],以及后期学者对这些算法的合成与改进。总体来说可归为传统方法与近期火热的人工智能机器学习。传统方法虽然简单、解释能力强,但是对环境参数进行相关性分析需要大量的数据支持,往往最少的也需要风速、风向、温度等环境参数,数据繁杂,收集困难,且很难反映 PM2.5 这种复杂的物理变化过程,预测能力有限。神经网络机器学习算法具有处理复杂非线性问题和超强学习的能力,其在预测结果上的准确率也很高。BP 神经网络^[7-8]是信息向前传播和误差向后传播相结合变换的计算过程,是每层权值不断调试更新的过程,也是神经网络学习训练的过程,此过程一直进行到网络输出的误差减小到可以拟定的数值,或者进行至预先设置好的学习次数结束^[9-10]。通过反复实验测试结果表明,控制程序在误差小于 0.5 或者学习训练次数大于 5 000 作为结束条件时具有优良的预测效果。

文中应用 BP 神经网络对雾霾预测进行了研究,对报警系统进行了设计和实现。值得注意的是,在做参数优化时误差不能设定的太小。在参考其他学者的研究时发现,有学者将其设置的很小,虽然在单站点小范围数据内不会出现问题,也取得了不错的预测值,但是在实际情况下的多站点大量数据应用时会出现收敛速度极慢的问题,导致程序假死。

1 软件系统设计

系统框架结构如图 1 所示。该系统分为表现层、业务层、逻辑层、数据层。表现层是用户与系统进行友好交互的界面窗口;业务层是按系统的不同需求定制的不同的功能业务;逻辑层为数据及功能之间的逻辑结构;数据层的数据分为空间数据和非空间数据。

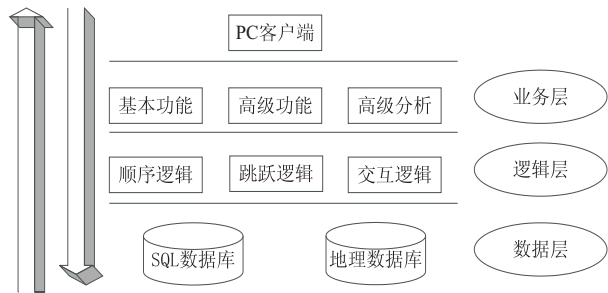


图 1 系统架构

数据层主要包括空间数据和非空间数据(属性数据)的存取,以及各数据间的关联。非空间数据使用 Microsoft 提供的 SQL 进行数据库长时间存取和 Microsoft Excel 数据临时存取及程序临时作业。空间数据使用 MapGIS 进行地理数据库存取。

逻辑层包括设计该系统的各个功能间的相互信息

传递,以及各个功能之间的协调工作,各功能板块之间的顺序逻辑、跳跃逻辑、交互逻辑。通过逻辑层的设计,使得程序科学化、简洁化、人性化,实现系统的完整性。

业务层包括各种业务功能,主要实现了基本功能、高级功能、预测分析、治理分析等功能,实现此系统的功能需求。

表现层是系统与用户的交互关注点,通过对用户使用规则来设定。实现用户最理想的需求表现层,是用户与系统交互的入口。该系统的表现层采用 MapGIS 接口开发,主要包括地图组件、业务交互组件和缓存管理组件以及 .NET 提供的各种类库、功能组件及交互控件。界面整体简洁美观,符合大众审美需求。

2 预测模型设计

雾霾天气是一种大气污染状态,雾霾是对大气中各种悬浮颗粒物含量超标的笼统表述,雾霾的主要成分是 PM2.5。文中以 PM2.5 预测模型为雾霾预测模型。该系统用到的非地理数据通过 PM2.5 云监测平台、国家环保部的数据中心收集到之后按照程序需求有序地入 SQL 数据库和临时 Excel 表。

2.1 BP 神经网络原理

BP 神经网络模型拓扑结构包括输入层(input layer)、输出层(output layer)和多个隐层(hidden layer)。BP 算法的学习过程由信息的正向传播和误差的反向传播两个过程组成。为了简单起见,这里只做一个隐层的 BP 网络分析,多个隐层的原理类似。其结构如图 2 所示。

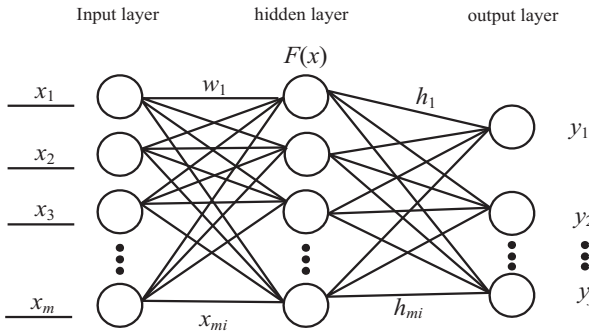


图 2 三层神经网络结构

首先,输入一组 x_1, x_2, \dots, x_m 到 input layer,然后通过 hidden layer 的连接权重产生一组数据 s_1, s_2, \dots, s_n 作为 hidden layer 的输入,然后通过 hidden layer 节点的 $f(x)$ 激活函数后变为 $\theta(s_j)$ 。其中 s_j 表示 hidden layer 的第 j 个节点产生的输出,这些输出将通过 hidden layer 与 output layer 的连接权重产生 output layer 的输入。这里 output layer 的处理过程和 hidden

layer 是一样的,最后会在 output layer 产生输出 y_j , 这里 j 是指 output layer 第 j 个节点的输出。然后计算误差,如果误差大于拟定的再反向误差传递,其原理和前面类似,依此调节各层之间的权重和阈值。直到误差或学习次数达到所拟定的参数,最后将 output layer 层的结果作为最终的预测结果。理论上只要有足够多的隐层神经元,就可以逼近任意复杂的非线性函数。

2.2 数据预处理

将收集到的历史 PM2.5 数据(样本数据)做预处理(数据归一化),统一样本的统计分布性。即将样本数据映射到更小的区间内,归一化一般可将样本数据映射至(0,1)或者(-1,1)。本次实验是为了统计坐标的分布,因此将数据归一化到(-1,1)区间,然后用非线性标准方程形式进行预测。但是传统的标准归一化公式并不能满足文中的需求,经过反复的实验对比分析,最终使用符合本次实验的归一化公式:

$$y = 2 * \frac{x - \min}{\max - \min} - 1 \quad (1)$$

其中, x 为转换前的值; y 为转换后的值; \max 、 \min 分别为样本的最大值和最小值。

2.3 学习训练数据

BP 神经网络的建立包括数据预处理、函数的选择、参数优化。

目前的神经网络激活函数有阶跃函数、准线性函数、双曲正切函数、Sigmoid 函数等等。由于雾霾的物理结构复杂,选取的激活函数为:

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (-1 < f(x) < 1) \quad (2)$$

参数优化在目前为止并没有系统的理论来解释说明参数的选择规范。通常都是借助经验和实验来确定最优参数。文中建立训练数据和历史数据预测对比来确定最优参数。训练数据如式 3 所示,测试对比数据如公式 4 所示。

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ x_2 & x_3 & x_4 & x_5 & x_6 \\ \dots & \dots & \dots & \dots & \dots \\ x_{n-4} & x_{n-3} & x_{n-2} & x_{n-1} & x_n \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ x_2 & x_3 & x_4 & x_5 \\ \dots & \dots & \dots & \dots \\ x_{n-3} & x_{n-2} & x_{n-1} & x_n \end{bmatrix} \quad (4)$$

其中, $x_1 \sim x_n$ 为单组样本数据,实验证明该训练方法较为有效。

2.4 预测模型的建立

BP 神经网络可以说是迄今为止最为成功的神经网络算法之一。基于 BP 神经网络优越的非线性处理

能力,建立 PM2.5 预测模型,模型流程如下:

- (1) 数据归一化;
- (2) 在(0,1)区间内随机初始化网络中的权值和阈值;
- (3) 根据当前随机的权值和阈值参数和激活函数公式 2 计算当前样本的输出值 y ;
- (4) 计算代价误差函数值;
- (5) 更新网络中的连接权重和阈值(以最小化累计误差为目的的更新);
- (6) 直至满足结束条件,以输出层的结果作为临时预测结果;
- (7) 将预测数据反归一化作为实际预测结果。

传统的 BP 神经网络是缺乏记忆性的,即每一组输入的学习数据之间是没有联系的,但如果经过数据的循环叠加的程序逻辑处理可以让其网络具有一定的往前记忆效果。

2.5 参数调优过程

BP 神经网络虽然应用广泛,但是在其不同的领域有着不同的使用方法,其主核心主要在于函数的选择和参数的优化。使用传统的调优方式不一定能真正地找到最优参数。文中使用程序逻辑的方法基本可以最快速获取最优越的参数。以误差的确定为例,其具体流程如下:

- (1) 准备样本数据(少量多组);
- (2) 以 2.4 节方法处理每组数据;
- (3) 通过循环记录从 0 开始到理论误差最大值的所有预测值;
- (4) 将预测值与实际值进行对比,取其最接近实际值的误差作为拟定的误差;
- (5) 对每组数据,重复步骤 1 ~ 步骤 3(为防止偶然情况);
- (6) 将所有的拟定误差均值作为该系统的累计误差判断。

此方法应用计算机的优越计算能力进行调参,适用于大多数参数的调优,既快速又可以基本达到最优。

3 功能结构

该系统包括基础操作、基本功能、数据管理、预测分析、栅格出图、专题图与治理六个部分^[11]。预测分析包括了可行性分析、预测分析,主要是测试网络和测试训练,训练网络数据、测试网络数据是使用文中预测模型处理好的数据,采用 BP 神经网络进行计算分析,最后通过程序输出数据和实测与预测对比曲线图来分析其预测精确度^[12-13],以此来判断此次运行的可行性,并提供用户交互式输入需要预测的天数。栅格出图^[14]包括了乐山出图、周边出图、输出工作区,通过多

层 B 样条法插值法对预测^[15]的未来 5 天数据进行空间插值以反映 PM2.5 的空间分布。专题与治理包括治理路线和预测专题以及预警,该功能块对乐山各县和各城城区进行预警专题图的可视化,便于使用者观察,做出需求判断。功能结构如图 3 所示。

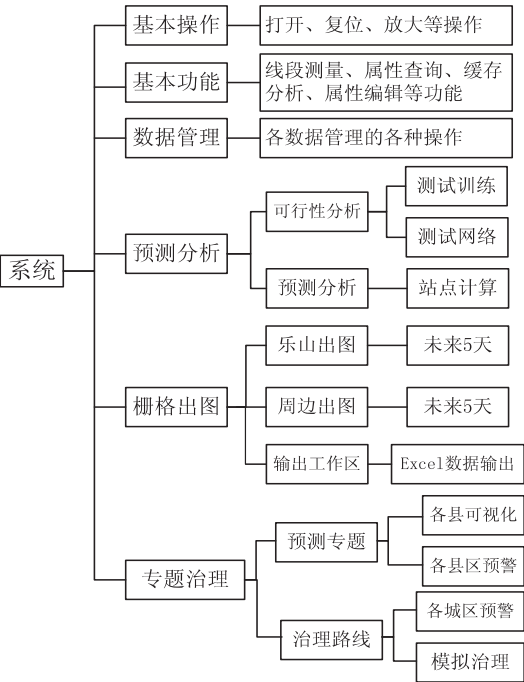


图 3 功能结构

基础操作、基本功能、数据管理是关于 GIS 软件地图操作的基本操作和基本功能等的实现,是系统的辅助功能。其他核心功能分别实现预测、预警、治理等功能。该系统在具有基本功能的情况下具有一套完整有序的处理流程。

4 系统实验与分析

文中采用乐山市整个行政区的雾霾数据进行系统测试,通过神经网络算法对数据进行数据的预测^[16-18],实现了县区预警分析和区县雾霾专题图^[19-20]。重点采集了乐山市中区的数据进程进行系统测试,对市中区进行了预测分析和预警。

4.1 县区预警分析

系统设计一键式实现乐山各县的预警模块。点击相应的功能键可按照程序的标准色卡表进行等级专题图可视化显示。也可对当前超出黄色预警线或者红色预警线的县区在程序地图中标记闪烁,并且把各县的 PM2.5 浓度值显示出,如图 4 所示。

4.2 城区预警

利用 MapGIS 二次开发等技术设计了乐山市中区各城区的黄色预警和红色预警模块。并使用插值算法计算出各城区的 PM2.5 浓度,使得各城区 PM2.5 数据可查。点击相应的功能键可标记出市中区各城区超

出红色预警线或者黄色预警线的城区,并可根据用户需求查出各城区的 PM2.5 浓度。并可通过模拟洒水车治理对严重城区进行洒水路线动态作业演示,给出相应的治理信息。黄色预警如图 5 所示。

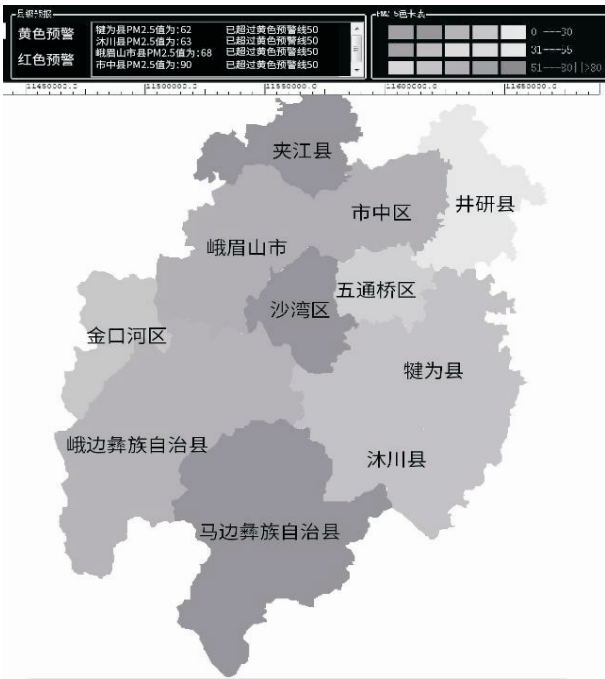


图 4 各县预警



图 5 黄色预警

该功能模块提供了多地区 PM2.5 数据可查,使用者可随意查出各个城区的 PM2.5 浓度,可为人们出行提供一定的决策。

4.3 插值出图

采用多层 B 样条法插值法对监测点未来 5 天的预测数据进行空间插值分析,以此来反映 PM2.5 的浓度空间分布情况。以乐山市 2018 年 8 月 1 日的一天

为例,进行空间插值,如图 6 所示。

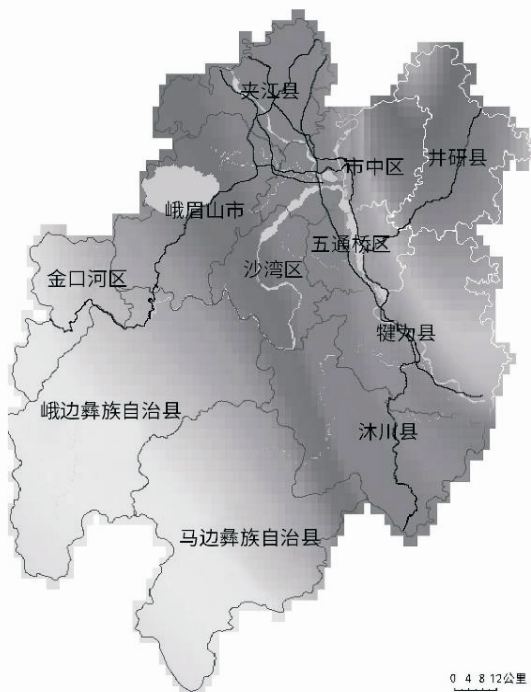


图 6 2018 年 8 月 1 日的预测空间分布

此次插值的数据是在 7 月份数据基础上经过预测模型预测出来的数据。从实验结果可以看出,乐山东部地区 PM_{2.5} 污染较为严重,笔者对乐山地区的地形环境及社会环境进行了多方资料查阅,其预测结果的分布基本与实际吻合。表明了预测精度足够的情况下,该功能模块具有一定的参考价值。

5 结束语

基于 .NET 平台采用 C# 程序设计语言、MapGIS 二次开发技术相结合,利用 BP 神经网络预测模型与报警模型设计了雾霾预报警系统。预报警能力与所选的预测模型的预测准确度有着紧密的联系,通过反复实验训练调整网络结构、改进相关参数、优化预测模型,提高了 BP 神经网络算法对 PM_{2.5} 的预测能力,并设计了科学合理的报警模型。通过实际数据进行了实验分析,证明了该预测模型的准确性和实用性,为雾霾预测系统提供了技术支持。在特殊的时段(比如换季时节、某些突变情况)预报警分析表明,该系统还存在一些不足,有待提高。后续研究中,应对季节数据、风力数据等一些自然因素进行进一步的研究与分析。

参考文献:

[1] 胡彬,陈瑞,徐建勋,等.雾霾超细颗粒物的健康效应[J].科学通报,2015,60(30):2808-2823.
[2] 谢志祥,秦耀辰,李亚男,等.基于 PM_{2.5} 的中国雾霾灾害风险评价[J].环境科学学报,2017,37(12):4503-4510.

[3] 童英华,田立勤,李靖.基于贝叶斯网络的雾霾重点污染源排放预测[J].计算机工程与设计,2018,39(9):2894-2901.
[4] 苗启广,李宇楠.图像去雾霾算法的研究现状与展望[J].计算机科学,2017,44(11):1-8.
[5] 丁凡,马敏劲,丁峰,等.基于 Spark+YARN 的空气质量预报云平台设计[J].计算机工程与应用,2017,53(19):98-101.
[6] 张志强,何文春,朱江,等.基于 B/S 架构的雾霾专题数据可视化服务系统设计与实现[J].计算机应用,2014,34(s2):140-142.
[7] 周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(6):1229-1251.
[8] 刘亚彬,刘大有,王金鹤.利用人工神经网络推断曲线间的空间关系[J].中国图象图形学报,2012,7(1):63-68.
[9] 王伦文,冯彦卿,张铃.动态数据挖掘的构造性学习方法综述[J].小型微型计算机系统,2016,37(9):1953-1958.
[10] 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(1):402-405.
[11] 张梦瑶,廉东本,赵奎,等.海量大气颗粒物成分分析系统[J].计算机系统应用,2015,24(10):53-57.
[12] 李飞,丁若修,张志佳.基于曲波变换的图像去雾算法研究[J].计算机技术与发展,2017,27(7):65-67.
[13] 麦嘉铭,王美华,梁云,等.特征学习的单幅图像去雾算法[J].中国图象图形学报,2016,21(4):464-474.
[14] 陈莹,刘冬冬.结合物理模型和 Retinex 的图像去雾算法[J].小型微型计算机系统,2016,37(10):2355-2360.
[15] 李力,金伟其,徐超,等.雾霾天气彩色图像的局域非线性变换增强算法[J].北京理工大学学报,2013,33(5):516-522.
[16] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks[J]. Journal of Machine Learning Research, 2017, 17(1):2096-2030.
[17] TAJBAKSHI N, SHIN J Y, GURUDU S R, et al. Convolutional neural networks for medical image analysis: fine tuning or full training[J]. IEEE Transactions on Medical Imaging, 2016, 35(5):1299-1312.
[18] CHEN Y H, KRISHNA T, EMER J S, et al. Eyeriss: an energy-efficient reconfigurable accelerator for deep convolutional neural networks[J]. IEEE Journal of Solid-State Circuits, 2017, 52(1):127-138.
[19] TSENG C W, LIN Y H, CHENG C H, et al. Depolarized haze of nano-porous AAO film via porosity and aspect control[J]. Optical Materials, 2018, 75:850-857.
[20] YANG T, GBAGUIDI A, YAN P, et al. Model elucidating the sources and formation mechanisms of severe haze pollution over Northeast mega-city cluster in China[J]. Environmental Pollution, 2017, 230:692-700.