

基于ARIMA-Adam-LSTM模型的广州市空气质量预测研究

王思坤¹ 王慧² 叶超³

成都理工大学数理学院

[摘要] 空气质量指数(AQI)是评估空气质量对健康影响程度的重要指标之一,为掌握AQI的波动情况,本文收集了广州市空气质量相关数据,并对AQI建立空气质量预测模型,具体包括ARIMA模型、Adam-LSTM模型以及ARIMA-Adam-LSTM混合模型进行预测,根据预测结果可知,ARIMA-Adam-LSTM组合模型相对ARIMA模型和Adam-LSTM模型预测效果更佳,评价指标也都显著优于ARIMA模型和Adam-LSTM模型,并且ARIMA-Adam-LSTM组合模型可以有效地捕获AQI数据的极值变化情况,使得结果更加接近实际情况,可为空气质量指数预测提供新的科学方法。

[关键词] 空气质量指数; ARIMA模型; Adam-LSTM模型; ARIMA-Adam-LSTM组合模型

[DOI] 10.12252/j.issn.2096-627X.2021.12.1597

一、引言

我国空气质量现状在不同地区状况不尽相同,本文基于当前空气质量的研究现状和社会现状,根据《环境空气质量标准》选取空气质量指数(AQI)进行研究,涉及的大气污染物包括PM₁₀、NO₂、SO₂、PM_{2.5}、CO和O₃,然后基于上述指标建立AQI的预测模型,模型包括ARIMA模型和Adam-LSTM模型,并结合两者的优势创新提出ARIMA-Adam-LSTM组合模型对空气质量数据进行预测,为相关环保或政府部门做出相应的决策提供了有力的参考。

二、研究现状

黄进等人利用ARIMA模型对苏州市AQI进行预测^[1]。范竣翔等人利用PM_{2.5}建立了有BP与LSTM的组合预测模型并应用于空气质量预报^[2]。石晓文等人分别建立了AQI的SVM、BP、LSTM、多元线性回归预测模型,结果表明LSTM模型预测效果更好^[3]。郑洋洋等人利用不同的优化方法优化LSTM预测模型,结果表明利用Adam优化后的预测模型效果更佳^[4]。杨其等人针对ARMA和LSTM进行了精度和实时性的对比分析,结果表明实时LSTM模型预测效果更佳^[5]。赵明珠等人分别建立了ARIMA模型、LSTM模型、Bi-LSTM预测模型并进行模型对比^[6]。本文将LSTM模型与ARIMA模型结合,采用Adam算法对模型进行优化,最后将预测结果与实际值进行比较,充分发挥单一模型的优势,从而达到预期的效果和精度。

三、模型介绍

(一) ARIMA模型

移动平均自回归模型(ARIMA)是结合差分运算和自回归移动平均模型的时间序列预测分析方法之一,即ARIMA(p, d, q)模型, p为自回归项数, q为滑动平均项数, d为差分阶数, ARIMA是一类重要的时间序列预测模型,在很多领域得到了广泛的运用。

(二) LSTM模型

LSTM长短期记忆网络模型以循环神经网络为基础的序列预测模型,在LSTM模型中每个数据段与前一段有一定的相关性,前一时刻的信息神经元将与当前的信息神经元相连,因此可以利用LSTM模型来解决对数据的长期依赖问题。同时自适应矩估计(Adam)是一种基于低阶自适应矩估计的随机目标函数一阶梯度下降优化算法,Adam对于LSTM模型的优化表现得十分出色,能够通过找到一系列参数最小化LSTM的损失函数,从而有效地提高LSTM模型的预测精度。

(三) ARIMA-Adam-LSTM模型

本文提出了ARIMA-Adam-LSTM组合模型预测广州市空气质量指数AQI, ARIMA模型主要提取数据中的线性部分, LSTM模型主要提取残差序列中的非线性部分,对ARIMA模型的预测结果进行补偿。ARIMA-Adam-LSTM组合模型的算法流程如下,

- 1、将数据集划分成训练集和测试集,然后对训练集拟合ARIMA模型;
- 2、计算ARIMA模型的残差,然后对测试集数据进行预测,并计算残差测试集残差序列;
- 3、利用Adam优化的LSTM模型对训练集残差序列进行建模,然后通过建立的模型对测试集残差序列进行预测,然后与ARIMA模型的残差序列进行比较,进行模型的分析;
- 4、将3中得到的残差序列和2中得到的测试集预测序列对应求和,得到测试集最终的预测序列,然后与原始测试集进行对比分析判定模型的性能。

(四) 空气质量预测效果评价

本文将采用RMSE(均方根误差)、MSE(均方误差)、MAE(平均绝对误差)和R方(判定系数)来评价预测模型的优劣。

四、实证分析

(一) 数据来源与说明

通过网络爬虫技术在中国空气质量在线监测分析平台选取广州市2015年1月到2021年6月空气质量相关数据,其中包括日

期、AQI、PM_{2.5}、PM₁₀、SO₂、CO、NO₂、O₃共计2373条数据。然后按照时间顺序将数据按照4:1划分为训练集和测试集。

(二) ARIMA模型

对AQI进行平稳性检验和模型定阶，通过计算ADF=-21.976, p=0.000<0.05, 表明AQI为一个平稳时间序列，不需要进行差分运算。然后根据自相关图和偏自相关图确定模型为ARIMA(1, 0, 1)。对模型参数进行检验，结果如下表所示，

表 参数显著性检验

	系数	标准误差	检验统计量	显著性水平
常量	70.1414	1.273	55.085	0.000
AR. 1	0.5830	0.025	23.392	0.000
MA. 1	0.1308	0.030	4.329	0.000

从上表可知，模型的系数均通过了显著性检验(p<0.05)，表明模型有效。检验残差是否为白噪声序列，LB检验统计量为6.323, p=0.898>0.1, 无法拒绝原假设，说明残差序列为白噪声序列，建立的ARIMA(1, 0, 1)模型显著有效。

(三) Adam-LSTM模型

使用AQI建立Adam-LSTM模型，通过多次迭代分析，最终确定模型参数：时间步长为1、迭代次数100、批量大小为32、采用Relu激活函数、2层隐藏层、每层为300个神经元，丢弃率为0.2。

(四) ARIMA-Adam-LSTM模型

首先使用ARIMA(1, 0, 1)来拟合时间序列得到线性成分，再利用LSTM模型对残差序列进行拟合来得到非线性成分，此时得到组合模型。模型的参数和Adam-LSTM模型的参数保持一致。

(五) 模型对比

本文采用多个评价指标对预测结果进行评价，对比结果如下表所示。

表 三种模型拟合效果比较

模型	RMSE	MSE	MAE	R方
ARIMA	23.635	558.601	17.047	0.247
Adam-LSTM	12.766	162.961	9.614	0.814
ARIMA-Adam-LSTM	2.199	4.836	2.199	0.995

从上表中可知，模型的预测性能从弱到强依次为ARIMA模型、LSTM模型、ARIMA-Adam-LSTM模型。利用ARIMA-Adam-LSTM组合模型来预测AQI，所有的评价指标都要显著优于ARIMA模型和LSTM模型。由此也说明了对于AQI这样复杂的序列而言，组

合模型可以在精度和效率上得到显著提升，使得结果更加接近实际情况。如下图所示，在测试集上预测值和原值对比的折线图，蓝线为原始值，红线为预测值，

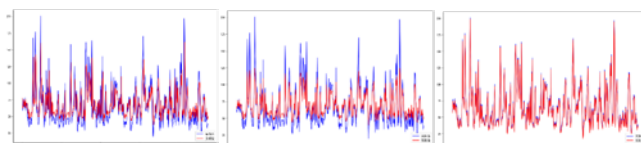


图 ARIMA(1, 0, 1) (左) LSTM (中) ARIMA-Adam-LSTM (右)

从图中可以看出，使用ARIMA-Adam-LSTM模型对广州市的AQI进行预测，组合模型不仅可以预测到数据的波动情况，还有效地克服了对AQI的峰值和谷值的变化未能有效把握的问题，完美拟合了空气质量指数的变化情况，预测效果明显优于ARIMA模型和Adam-LSTM模型。

五、结语

针对广州市的空气质量指数AQI，通过建立了ARIMA和LSTM的混合模型，并且使用Adam作为优化器进行了模型的训练，最终得到了一个较准确的空气质量指数预测模型，其中ARIMA模型主要提取数据中的线性部分，LSTM模型主要提取数据中的非线性部分，非线性部分以残差序列为数据基础，通过对残差序列的分析，对ARIMA模型的预测结果进行补偿，进而达到较好的预测效果。

通过对比分析发现，相比于ARIMA模型和Adam-LSTM模型，ARIMA-Adam-LSTM组合模型的预测效果更好，更加能够获取到空气质量指数AQI中的绝大部分信息，模型的迁移性较强，针对不同的地区可以采用相同的混合模型的进行训练和预测，该组合模型具有较强的泛化能力。

参考文献：

[1]黄进,张金池.苏州市空气质量的时间序列变化过程研究[J].环境科学与技术,2009,32(06):49-52.
 [2]范峻翔,李琦,朱亚杰,侯俊雄,冯道.基于RNN的空气污染时空预报模型研究[J].测绘科学,2017,42(07):76-83+120.

作者简介：王思坤(1996-)，男，汉族，四川广元人，硕士在读，职称：无，研究方向：大数据技术与运用。

王慧(1997-)，女，汉族，四川眉山人，硕士在读，职称：无，研究方向：数理金融统计。

叶超(1996-)，女，汉族，四川达州人，硕士在读，职称：无，研究方向：数理金融统计。