267号 投稿日期2019-08-29

基于时间序列的网格化城市管理案件预测模型研究

陈栾杰1，吴 同2，郑建春3，彭 玲2※，杨艳英3※

（1. 北京工业大学信息学部，北京，100124；

2. 中国科学院空天信息研究院，北京，100094；

3. 北京城市系统工程研究中心，北京，100089）

**摘要：**针对传统或流行的基于时间序列的预测模型，探索出一种适用于网格化城市管理的成体系的案件预测方法。分别采用博克斯-詹金斯法、Auto-ARIMA以及LSTM模型，对近几年北京市六个城区各站点网格化管理问题的案件数量进行预测，通过对比不同模型方法间的准确度和实用性，以MAPE为精度评价指标，分析各个模型应用在城市网格化问题预测方面的优势与劣势。研究发现，Auto-ARIMA适合进行对网格化管理问题的数量趋势预测，博克斯-詹金斯法在解决滞后性问题后预测准确率很高，但由于预测流程繁琐，因此实用性较差，LSTM预测效果相对准确且平稳，可以在样本输入量、参数以及自身架构上进行进一步优化。

**关键词：**网格化城市管理；案件数量预测；LSTM；博克斯-詹金斯法；Auto-ARIMA

**中图分类号：**C931; TP391 **文献标识码：**A

# 引言

随着中国城市化进程的加快，如暴露垃圾、无照经营游商等一系列的城市管理问题也开始逐渐增多。而要解决城市发展过程中的问题，就离不开科学、完善的城市管理体制与城市管理手段。网格化的城市管理模式由于结合了先进的信息技术与科技手段，已经在城市化进程中逐渐兴起[1]。由于网格化城市管理以高效、敏捷、精确为目的，因此对于网格化城市管理问题的预测的研究就显得更加有必要。

另一方面，由于城市管理问题的案件数量需要定期进行总结分析，所以将案件数量按时间顺序排列的数据集合符合时间序列方法的数据特征。因此，可以将时间序列方法作为一种必要的新思路与网格化城市管理问题相结合，从而更高效的进行城市管理问题的解决。

目前，国内少有学者将基于时间序列的预测应用到网格化城市管理当中。佘冰[2]等将网格化城市管理问题以空间的角度进行了可视化分析，为进一步在空间上的统计建模分析奠定基础；常燕军等[3]对ARIMA算法在网格化管理预测上的应用进行了初探，但尚未形成体系化的方法。本文旨在探究适用于网格化城市管理问题预测的模型，根据不同模型的优势与劣势，从而研究出一种实用性强、准确率高、成体系的预测方法。

# 与网格化管理相关的时间序列模型概述

## 由于网格化城市管理问题的案件发生存在一定的随机性，且不存在明显的变化规律，不同站点的案件数量也并不统一，因此在选择模型时，应当选择适用于非季节性的时间序列模型。针对这一数据特点，本实验采用博克斯-詹金斯法中的ARMA模型、ARIMA模型，以及Auto-ARIMA模型和LSTM模型，它们对于时间序列的变化特点没有特别的要求。 博克斯-詹金斯法

博克斯-詹金斯法（Box-Jenkins Method）是由乔治·博克斯（George E.P. Box）和格威利姆·詹金斯（Gwilym Jenkins）于20世纪70年代提出的著名时间序列预测方法[4]。该预测方法经历了模式识别、模型参数估计、应用模型预测三个阶段，是一种精度相当高的短期时间序列预测方法[5]。其中在模式识别阶段，根据数据预处理的方式不同，可将预测模型分为自回归移动平均模型（Auto-Regressive Moving Average Model, ARMA）与差分自回归移动平均模型（Autoregressive Integrated Moving Average Model, ARIMA）。在参数估计阶段，ARMA模型所需要确定的参数为p、q，其中p代表自回归项数，它表示样本数据的滞后数，q代表移动平均项数，它表示预测误差的滞后数。ARIMA模型在保留ARMA的参数p、q的基础上，多出了参数d，参数d代表时间序列趋于平稳时所需要的差分次数，而ARMA模型与ARIMA模型的唯一区别在于是否对样本时间序列进行差分处理。

## Auto-ARIMA

博克斯-詹金斯法是一种被证明过的、有一定准确度的时间序列预测方法，但在实际应用中，对于时间序列的平稳化处理和参数调整过程是十分耗时的。在实现博克斯-詹金斯法时，需要让数据保持平稳，同时p和q值的确定过程十分繁琐且不易达到精确值，这会使博克斯-詹金斯法的ARMA或ARIMA模型的使用条件变得十分苛刻。因此，针对博克斯-詹金斯法在实际应用中造成的困扰，而Auto ARIMA[6]能够省去p和q值确定的过程，使模型的数据接受能力更强。Auto-ARIMA 采用了Hyndman-Khandakar算法，它最早被封装为R语言的包[6]，之后被应用于python语言的包中。

## LSTM

人工神经网络（Artificial Neural Network, ANN）是人工智能领域的研究热点[7]，它对人脑神经元进行抽象化表示，从而形成各类人工神经网络模型。其中，循环神经网络（Recurrent Neural Networks, RNN）作为其中的重要分支，被广泛应用于语音识别领域[8，9]。长短期记忆神经网络（Long Short-Term Memory, LSTM）作为对RNN的改进，解决了RNN存在的梯度爆炸或梯度消失的问题[10]。LSTM 由塞普·霍克赖特（Sepp Hochreiter）[10]团队于20世纪90年代提出，并在近期被克亚历山大·格雷夫斯（Alexander Graves）[11]进行了推广和改良。与传统的RNN相比，LSTM使用记忆块（Memory blocks）替代RNN的隐层单元（Hidden units），这些记忆块中的记忆单元（Memory cell）用于存储之前序列的信息。同时为了去除或增加信息到记忆块，LSTM增加了三种“门”（gate）的结构：输入门（Input gate）、输出门（Output gate）、遗忘门（Forget gate）。它们用于记住重要信息，忘记不重要的信息。LSTM的记忆块结构如图1所示。

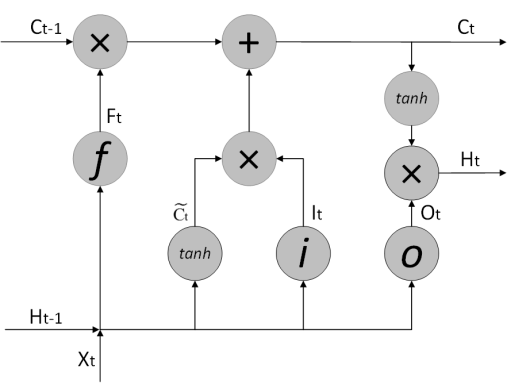


图1 LSTM记忆单元结构

Fig.1 LSTM memory cell structure

# 网格化城市管理统计数据实证分析

## 数据来源

研究中所用到的实验数据为2015—2018年北京市六个城区各个站点的无照经营游商、乱堆物堆料、暴露垃圾、积存垃圾渣土、非法小广告等五种网格化管理问题的案件数量，其中案件数量的统计以月为单位。数据来源于北京城市系统工程研究中心。本实验将以“无照经营游商”这一种网格化管理问题为例，测试与比较各个时间序列模型的适用性。此外，由于各站点的案件数据有一定的敏感性，故在本文中隐藏真实的站点名。

## 评价指标

本文对于各个模型间与模型内部的比较均使用平均绝对百分误差值（Mean Absolute Percentage Error, MAPE）作为评价指标。计算方法如下：

其中，指的是样本实际数据值，指的是样本预测数据值，*n*指的是预测的样本数量。由公式5可知，MAPE的值越小，表明预测值与实际值之间偏差越小，预测性能越好，反之则预测性能越差。

## 基于时间序列的模型预测

### 博克斯-詹金斯法

#### 数据平稳性分析与处理

由于博克斯-詹金斯法要求数据序列有较强的平稳性，因此，需要对样本数据序列进行平稳性检验。表1展现了样本数据序列的ADF（Augmented Dickey-Fuller test）单位根检验[12]结果。由于部分站点的ADF检验结果均大于99%，95%，90%置信区间下临界的ADF检验值，因此样本数据序列不平稳，故需要对样本数据序列进行平稳化处理。本实验将样本数据序列进行对数化，从而提高样本数据序列的平稳性。由表2可知，在对数化后，各站点的ADF检验结果均小于99%，95%，90%置信区间下临界的ADF检验值，且P-value值接近于零，因此数据经过平稳化处理后平稳性达到了ADF根检验的标准。

表1 博克斯-詹金斯法样本数据序列各站点ADF单位根检验表

Tab.1 The ADF test result of each site with Box-Jenkins Method

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 站点编号 | ADF Test Result | %1统计值 | ％5统计值 | ％10统计值 | P-value |
| 1 | -2.44 | -3.58 | -2.92 | -2.60 | 0.13 |
| 2 | -3.26 | -3.58 | -2.92 | -2.60 | 0.16 |
| 3 | -3.25 | -3.58 | -2.92 | -2.60 | 0.01 |
| 4 | -3.58 | -3.58 | -2.92 | -2.60 | 0.01 |
| 5 | -3.65 | -3.58 | -2.92 | -2.60 | 0.01 |
| 6 | -2.43 | -3.58 | -2.92 | -2.60 | 0.13 |
| 7 | -3.34 | -3.58 | -2.92 | -2.60 | 0.01 |
| 8 | -3.01 | -3.58 | -2.92 | -2.60 | 0.03 |
| 9 | -2.68 | -3.58 | -2.92 | -2.60 | 0.07 |
| 10 | -2.54 | -3.58 | -2.92 | -2.60 | 0.10 |
| 11 | -1.59 | -3.58 | -2.92 | -2.60 | 0.48 |
| 12 | -2.99 | -3.58 | -2.92 | -2.60 | 0.03 |
| 13 | -2.38 | -3.58 | -2.92 | -2.60 | 0.14 |
| 14 | -3.88 | -3.58 | -2.9 | -2.60 | 0.01 |
| 15 | -3.02 | -3.58 | -2.92 | -2.60 | 0.03 |

表2 博克斯-詹金斯法样本数据序列平稳化处理后各站点ADF单位根检验表

Tab.2 The ADF test result of each site with Box-Jenkins Method after stationary processing

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 站点编号 | ADF Test Result | %1统计值 | ％5统计值 | ％10统计值 | P-value |
| 1 | -5.04 | -3.58 | -2.92 | -2.60 | 1.84e-05 |
| 2 | -6.22 | -3.58 | -2.92 | -2.60 | 5.10e-08 |
| 3 | -5.88 | -3.58 | -2.92 | -2.60 | 3.02e-07 |
| 4 | -5.76 | -3.58 | -2.92 | -2.60 | 5.54e-07 |
| 5 | -6.92 | -3.58 | -2.92 | -2.60 | 1.11e-09 |
| 6 | -5.96 | -3.58 | -2.92 | -2.60 | 2.02e-07 |
| 7 | -5.39 | -3.58 | -2.92 | -2.60 | 3.54e-06 |
| 8 | -7.06 | -3.58 | -2.92 | -2.60 | 5.04e-10 |
| 9 | -6.58 | -3.58 | -2.92 | -2.60 | 7.45e-09 |
| 10 | -5.07 | -3.58 | -2.92 | -2.60 | 1.57e-05 |
| 11 | -6.69 | -3.58 | -2.92 | -2.60 | 4.09e-09 |
| 12 | -6.69 | -3.58 | -2.92 | -2.60 | 3.93e-09 |
| 13 | -5.78 | -3.58 | -2.92 | -2.60 | 5.10e-07 |
| 14 | -5.80 | -3.58 | -2.92 | -2.60 | 4.64e-07 |
| 15 | -5.92 | -3.58 | -2.92 | -2.60 | 2.43e-07 |

#### 构建ARMA模型

在对样本数据序列进行平稳化处理后，需要进行p，q的参数估计。由于没有进行差分操作数据便已经通过了ADF根检验，因此将采用ARMA模型进行建模。对于p，q的参数估计，采用传统的AIC、BIC统计量自动确定[13]，通过借助Python语言的statsmodels包实现对于各个站点阶数的确定。

#### 预测结果分析

采用博克斯-詹金斯法时，“无照经营游商”网格化管理问题各个站点的MAPE值均小于1。与此同时，以图2的6个站点为例，将预测值与实际值对比分析后可以发现，使用博克斯-詹金斯法预测站点案件数量存在明显的滞后性，且预测值相对实际值滞后一个单位。于是在手动将预测值提前一个单位后，各站点MAPE值下降到0.4以下，预测效果有明显的改观。

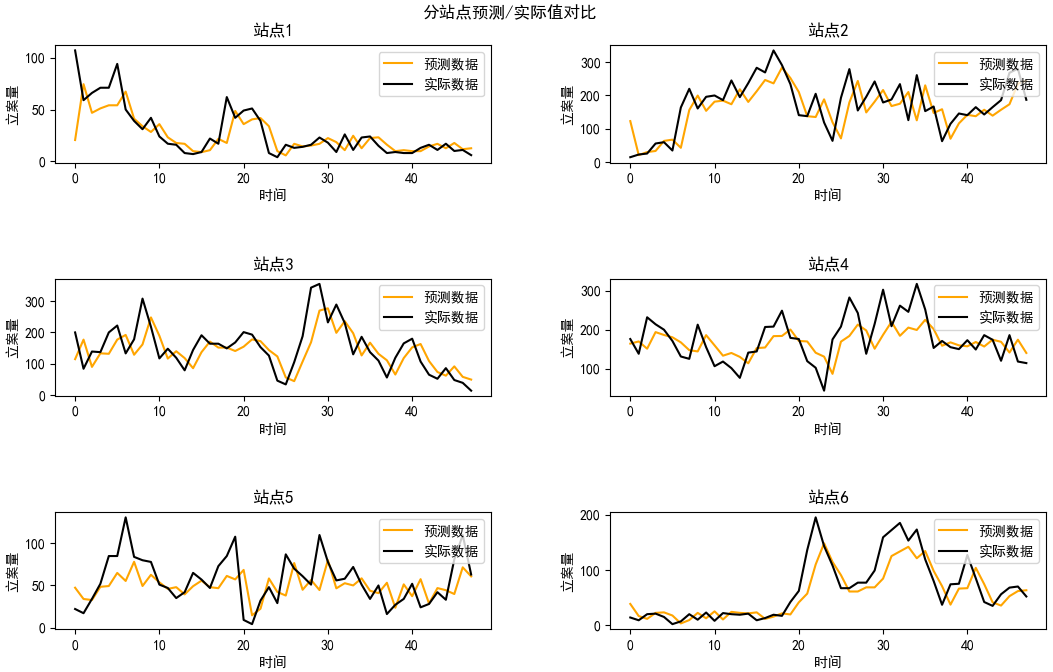


图2 博克斯-詹金斯法部分站点预测值与实际值对比图

Fig.2 The comparison of predicted value and actual value in some of sites with Box-Jenkins Method

### Auto-ARIMA

采用Auto-ARIMA模型解决数据预处理繁琐、滞后性等问题，得到的部分站点预测值与实际值对比数据如图3所示。从图中可以看出，采用Auto-ARIMA模型无准确性可言，仅仅能够预测样本时间序列的趋势。

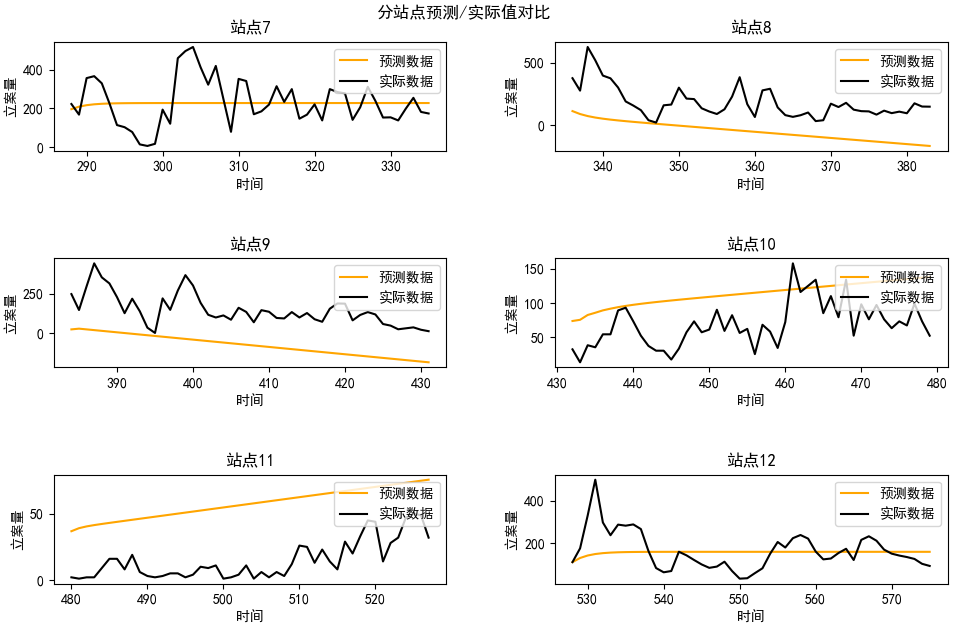


图3 Auto-ARIMA部分站点预测值与实际值对比图

Fig.3 The comparison of predicted value and actual value in some of sites with Auto-ARIMA model

### LSTM

#### ① 数据相关性分析

LSTM支持相关性强的数据间的统一预测，即在各站点数据相关性强的条件下，可以将所有站点的所有数据作为统一的输入进行预测。因此首先采用皮尔逊相关系数法度量各站点案件量的相关性。皮尔逊相关系数法是一种精确度量两个变量间关系密切程度的统计学方法。

将两个站点定义为*x*和*y*，将*i*定义为第*i*个月的案件数量，记为(，)(*i*=1，2，···，n)，则应用皮尔逊相关系数法比较各站点的相关系数的数学表达式为：

|  |  |
| --- | --- |
|  | 公式2 |

相关系数r的取值范围在-1和+1之间，即|*r*|≤1。|*r*|越接近于1，则表明*x*站点与*y*站点线性相关程度越高。当r=-1时，*x*站点与*y*站点之间为完全负线性相关关系；当*r*=+1，*x*站点与*y*站点之间为完全正线性相关关系；当r=0时，*x*站点与*y*站点之间不存在线性相关关系。

任意两个站点间的相关程度可分为以下几种情况：当|*r*|≥0.8时，可视为两站点高度相关；当0.5≤|*r*|＜0.8时，可视为两战点中度相关；当0.3≤|*r*|＜0.5时，可视为两站点低度相关；当|*r*|＜0.3时，说明两站点之间的相关程度极弱，可视为非线性相关[14]。

在无照经营游商网格化管理问题中，每两个站点间的相关系数值如图4所示。它的纵坐标代表相关系数r的值，横坐标代表站点*x*，任意站点*x*对应的散点为站点*y*。从图中可知，多数站点间呈中度相关或者低度相关，数据间的相关性强较弱，因此不适用于LSTM的统一预测，应该采用LTSM的分站点预测。

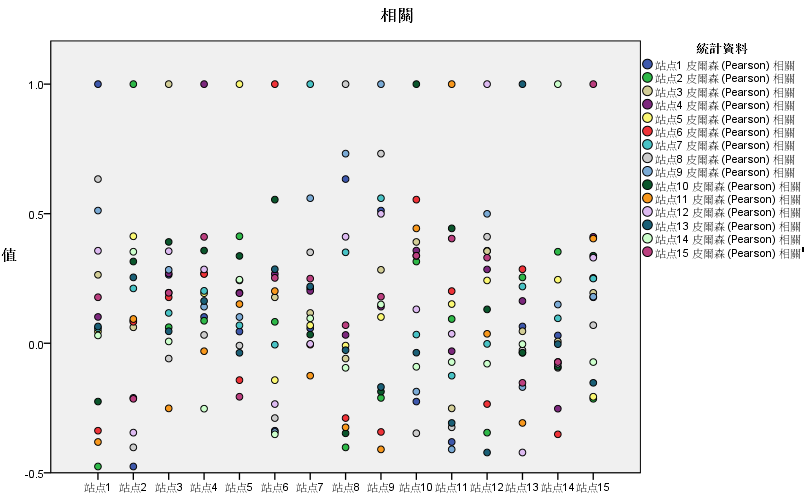


图4 “无照经营游商”网格化管理问题各站点案件量相关性散点图

#### Fig.4 The correlation scatter plot about grid city management problem of unlicensed business activities② 模型构建与参数选择

在建模前，由于神经网络对输入数据十分敏感，同时也由于单个站点内的不同月份中，案件数量可能有很大差别，因此需要对样本数据进行归一化处理。本实验采用的归一化函数如下：

其中，中为归一化后的数据，为样本数据，、分别为样本数据集的最大值和最小值。

在进行输入数据的归一化后，为了使模型具有较强的泛化能力，还需要利用随机函数将样本数据序列完全打乱，从而增强预测的合理性和准确性。实验表明，相比于不打乱样本数据序列，打乱样本数据序列的预测结果的精度要更高。在进行数据打乱后，将整个数据集分为训练集与测试集两部分。其中训练集占所有数据的80%，测试集占所有数据的20%。在模型构建时，本实验的后端采用Google的第二代分布式机器学习系统Tensorflow[15]来实现，前端采用Keras框架来实现。网络的逻辑结构如图5所示。

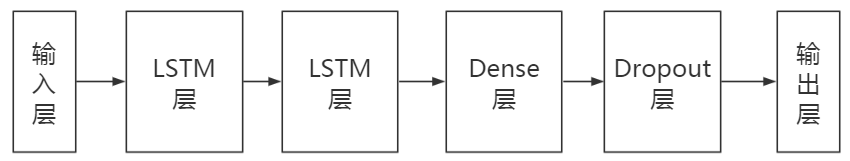


图5 本实验采用的LSTM网络逻辑结构图

Fig.5 The neural network structure used in the experiment

本模型采用随机梯度下降（Stochastic Gradient Descent, SGD）的方法进行优化，LSTM层包含的节点数为50，时间步长为10，学习率为0.001，批处理大小为16，迭代次数为800，激活函数采用ReLU激活。

#### ③ 预测结果分析

采用LSTM进行预测时，“无照经营游商”网格化管理问题中站点12的MAPE值相对最小，预测结果相比其他站点更加准确；而站点5的MAPE值相对最大，预测结果相比其他站点更加不准确。

在采用“无照经营游商”网格化管理问题站点12的案件量数据进行预测时，其MAPE值如图6所示；在采用“无照经营游商”网格化管理问题站点5的案件量数据进行预测时，其MAPE值如图7所示。

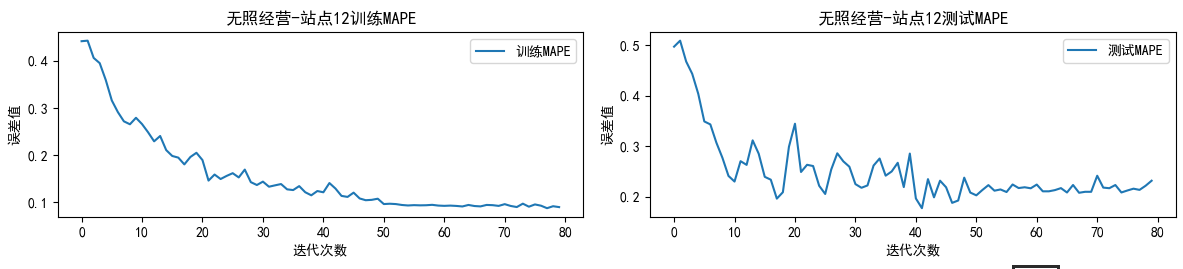


图6 采用LSTM时站点12的MAPE值图

Fig.6 MAPE of Site 12 with LSTM

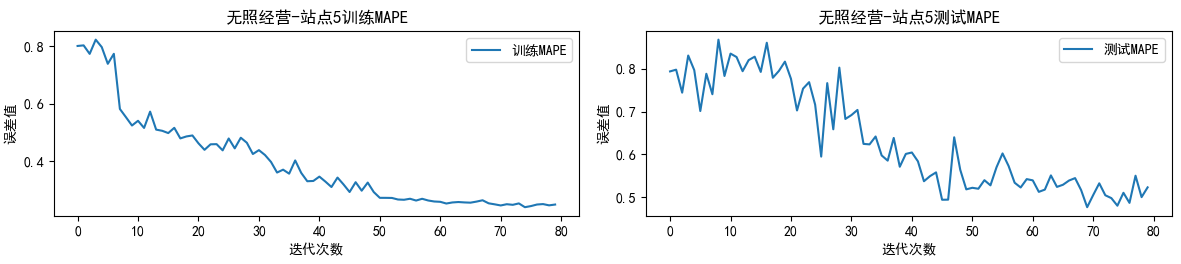


图7 采用LSTM时站点5的MAPE值图

Fig.7 MAPE of Site 5 with LSTM

可以看出，随着迭代次数的增加，站点12与站点5的MAPE逐渐趋于平稳。其中，站点12的训练数据的MAPE最终趋近于0.10，而测试数据的MAPE最终趋近于0.24。站点5的训练数据的MAPE最终趋近于0.31，而测试数据的MAPE最终趋近于0.51。由实验得出，在将“无照经营游商”网格化管理问题的所有站点进行预测后，各站点训练数据的预测结果MAPE均小于0.4，而测试数据的预测结果MAPE均小于0.6。

## 各时间序列模型的对比结果

根据各个模型的运行结果，本文将其MAPE汇总进行对比。“无照经营游商”网格化管理问题中各模型各站点的预测MAPE值如图8所示。从图中可知，若仅考虑预测效果，则在解决滞后性问题后的博克斯-詹金斯法相比于其他方法更加准确；若不解决博克斯-詹金斯法的滞后性，则LSTM相对更加准确。

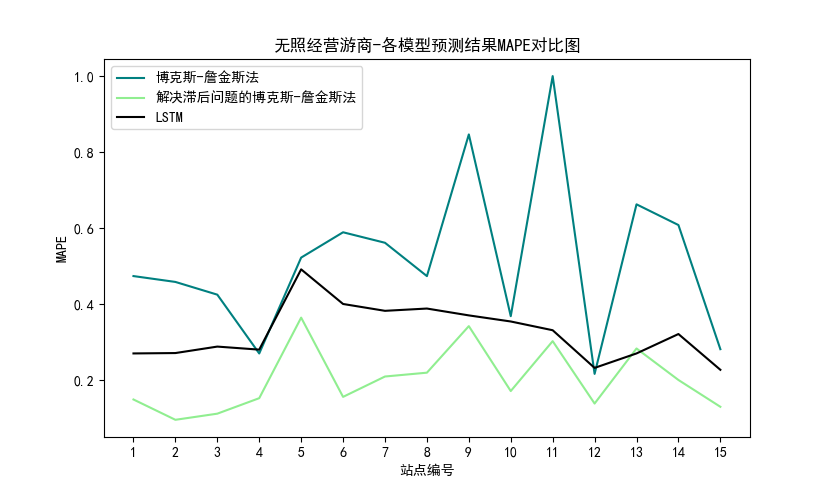


图8 不同模型之间预测结果MAPE对比图

Fig.8 MAPE comparison of predicted results between different models

# 结论

本文实现了四种基于时间序列的网格化管理问题的预测方法，并对比了这四种模型各自的优势与劣势。对于Auto-ARIMA来说，该模型适合对网格化管理问题的数量趋势进行预测，但不适合进行进一步的高精度预测。博克斯-詹金斯法虽然最终预测效果相比于其他三种模型更加准确，但在实际应用中，该模型相对繁琐的预测流程限制了其进一步的推广。博克斯-詹金斯法繁琐的预测流程主要体现在四个方面。第一，在进行预测前，需要对样本数据序列进行如对数化、差分等一系列的数据平稳化操作；第二，在对样本数据平稳化处理后还需要对数据进行ADF单位根检验，若未通过检验则会使预测结果有很大的偏差；第三，该模型的参数很难定论，本例中参数p和q会被限制在statsmodels包所给定的范围之内；第四，该模型产生的结果可能存在滞后性，人工观察滞后步数减缓了其预测速度。对于LSTM来说，其预测效果相对准确，预测精度较为平稳，且相比博克斯-詹金斯法，没有过多繁琐的预测流程。但根据上文提到的LSTM模型理论可知，若要进一步提高该模型的精度，则应该增加更多的样本数据。此外，LSTM可调整的参数过多，包括其架构本身也可以调整。因此在未来的研究中，可尝试进一步优化LSTM的参数和其本身的架构，从而找到最优参数的LSTM模型。同时，针对网格化城市管理，应当将其案件数量从以月为单位更改为以周为单位，增加LSTM输入的样本数量，从而进一步减少预测误差。

Urban Grid Management Cases Prediction Model Based on Time Series

【Abstract】 Specific to the traditional or popular predicting model based on time series, this paper explores the predicting method for systematic case suitable for grid city management. Respectively introduce Box-Jenkins method, Auto-ARIMA and LSTM models to predict the number of cases of grid management problems in six urban areas of Beijing in recent years, by comparing the accuracy and practicability between different model methods, taking MAPE as the accuracy evaluation indicator, analyzing in application of urban grid management. The research indicates, Auto-ARIMA is suitable for predicting the quantitative trend of grid management problems, Box-Jenkins method has a high prediction accuracy after solving the lag problem, however, in a poor practicability, LSTM prediction effect is relatively accurate and stable, it can be further improved on sample input, parameters, and its framework.

【Key words】 urban grid management；case number prediction ；LSTM；Box-Jenkins method；Auto-ARIMA

**参考文献**

[1] 田毅鹏．城市社会管理：网格化模式与基层秩序构建（专题讨论）——城市社会管理网格化模式的定位及其未来[J]．学习与探索，2012, (2): 28-32

[2] 佘冰，朱欣焰，呙维，等．基于空间点模式分析的城市管理事件空间分布及演化——以武汉市江汉区为例[J]．地理科学进展，2013, 32(6): 924-931

[3] 常燕军，周向红．基于ARIMA模型的城市网格化管理预测与仿真[J]．统计与决策，2016, (1): 54-57

[4] George E. P. Box，Gwilym M．Time series analysis : forecasting and control[J]．Journal of Time Jenkins，1971, 31(4): 303-303

[5] 王晓鹏，曹广超，丁生喜．基于Box-Jenkins方法的青南高原降水量时间序列分析建模与预测[J]．数理统计与管理，2008, (4): 565-570

[6] Rob J. Hyndman．Yeasmin. Automatic Time Series Forecasting: The forecast Package for R[J]．Journal of Statistical Software Khandakar，2008, 27(3): 1-22

[7] 朱大奇．人工神经网络研究现状及其展望[J]．江南大学学报（自然科学版），2004,3(1):103-110

[8] Tebelskis J．Speech recognition using neural networks[D]．Siemens AG，1995

[9] 朱小燕, 王昱, 徐伟．基于循环神经网络的语音识别模型[J]．计算机学报，2001, 24(2):213-218

[10] Hochreiter S，Schmidhuber, Jürgen．Long Short-Term Memory[J]．Neural Computation，1997, 9(8):1735-1780

[11] GRAVES A．Supervised Sequence Labelling with Recurrent Neural Networks[M]．Springer Berlin Heidelberg，2012

[12] 左秀霞．单位根检验的理论及应用研究[D]．华中科技大学，2012

[13] Ding Jie，Vahid Tarokh，Yuhong．Bridging AIC and BIC: a new criterion for autoregression[J]．IEEE Transactions on Information Theory Yang，2015, PP(99): 1-1

[14] 杨帆，冯翔，阮羚，等．基于皮尔逊相关系数法的水树枝与超低频介损的相关性研究 [J]．高压电器，2014, 50(6): 21-25+31

[15] Martín Abadi，Ashish Agarwal，Paul Barham. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems[J]. arXiv preprint arXiv:1603.04467，2016

作者简介：

陈栾杰（1997-），男，汉族，福建泉州人，北京工业大学本科生，主要研究方向为计算机科学与技术

吴同(1995-)，女，河北唐山人，中国科学院空天信息研究院硕士研究生，主要研究方向为遥感影像智能学习、智慧城市时空大数据脉动分析。

郑建春（1977-），男，河北省滦南县人，北京城市系统工程研究中心城市运行研究部主任，副研究员，博士，主要研究方向为城市公共安全、风险管理、应急管理。

通讯作者简介：

彭玲(1965-)，女，湖北武汉人，中国科学院空天信息研究院研究员，博士，主要研究方向为遥感影像智能学习、智慧城市时空大数据脉动分析、室内外定位导航

杨艳英（1976-），女，山西省忻州县人，北京城市系统工程研究中心城市运行研究部助理研究员，双硕士，主要研究方向为城市管理、城市领域数据分析、城市公共安全。

基金项目：

城市管理智能挖掘(Y9B0130H22)

通联信息：

地址：北京市朝阳区大屯路中国科学院遥感与数字地球研究所

邮政编码：100101

移动电话：13801072592

固定电话：01064843901

邮件：pengling@aircas.ac.cn