

基于奇异谱分析的旅客运输量预测研究

方 成¹, 杨正儒², 任建宝³, 谭莹莹¹

(1. 安徽建筑大学数理学院, 合肥 230601; 2. 淮南市建发市政工程有限公司, 安徽 淮南 232007;

3. 安徽建筑大学电子与信息工程学院, 合肥 230601)

摘要: 对旅客运输量进行科学准确地预测, 可以为交通领域相关部门提供有效的借鉴。将旅客运输量作为研究对象, 基于 SSA(奇异谱分析), 结合 LSTM(长短时记忆神经网络)和 ARMA(自回归移动平均模型), 通过 SSA 降噪处理, 将旅客运输量时间序列分解为信号序列和噪声序列, 分别对其进行 LSTM 和 ARMA(2,3)建模, 预测其变化趋势。通过对比单一的 ARIMA(3,1,2)模型和 LSTM 模型的实验结果表明, SSA-LSTM-ARMA 预测旅客运输量效果更好, 预测精度更高。

关键词: 旅客运输量; 奇异谱分析; LSTM(长短时记忆神经网络); ARMA(自回归移动平均模型)

中图分类号: C813 **文献标志码:** A **文章编号:** 1671-1807(2024)03-0140-06

近年来,我国公路、铁路、水运以及民航运输体系日益完善,让人们出行更加方便快捷,为国家经济发展提供了有力支撑,形成了一种良性循环。这得益于我国经济社会发展模式逐步实现了由高速发展向高质量发展的转变,同时也提高了人们的生活水平。但是,这也给交通运输系统带来了更多的挑战,因此提高旅客运输量的预测准确性,就必须加强对交通运输系统的改进和优化,以保障人们出行的安全、舒适和便利。既有助于为我国交通运输系统的完善提供借鉴,也可以从侧面反映人民生活水平及国家经济的发展状况,具有一定的实际意义。

根据国家统计局发布的关于旅客运输量实际数据,由于新冠肺炎疫情的影响,2019 年,我国旅客运输量的人数为 1 760 429 万人,而 2020—2022 年我国的旅客运输量人数分别为 966 500 万人、830 249 万人及 558 451 万人,同比分别下降 45.1%、52.8%及 68.3%^①。2023 年上半年我国的旅客运输量有了显著的提高,截至 2023 年 5 月,我国旅客运输量的累计人数达 352 500 万人,因此对该问题的预测研究具有重要的意义。

对旅客运输量的预测研究相对较少,主要利用时间序列模型和神经网络模型进行分析和预测。有以平稳时间序列自回归移动平均模型(auto-regressive moving average, ARMA)模型为基本形式的 Box-Jenkins 模型以及各种指数平滑法为主的传统时间序列预测模型,它们在解决复杂问题时都具有重要的意义。例如,李黎莎^[1]根据中国民用航空局月度旅客运输量数据建立 ARIMA(3,1,3)模型进行预测,结果表明该模型较为准确地反映了民航旅客运输量的变化对航空公司、机场及相关民航运行保障单位的影响。周元炜和兰翔^[2]采用 LSTM-ARIMA(长短期记忆网络-自回归求和移动平均模式)加权模型预测旅客运输量,预测误差较小,预测精度较高。王晨晨^[3]利用主成分分析选取影响成都航空运输量的主要指标,建立线性回归模型,利用趋势外推法对主要指标预测并代入回归模型,得到航空运输量的预测结果。王飞和韩翔宇^[4]以分解-集成为基础提出了 VMD-LSTM(变分模态分解-长短期记忆网络)模型,并将其用于预测民航运输量,预测结果更佳。田郝青^[5]以铁路旅客运输

收稿日期: 2023-10-29

基金项目: 2022 年度省级研究生创新创业实践项目(2022cxcsj150); 2023 年度省级研究生质量工程项目(2022xxsfkc031)

作者简介: 方成(1998—),男,安徽六安人,硕士研究生,研究方向为大数据分析及应用;杨正儒(2000—),女,安徽六安人,质检员,研究方向为交通工程;任建宝(1998—),男,安徽六安人,硕士研究生,研究方向为推荐系统和数据挖掘;谭莹莹(1979—),女,安徽阜阳人,教授,研究方向为代数组合、谱图理论、复杂网络。

^①数据来自中国国家统计局官网 <http://www.data.stats.gov.cn/>。

量为研究对象,提出了基于季节性的预测模型。孔建国和李晓^[6]根据小波分析理论,利用 MATLAB 软件对民航旅客运输量进行预测。

奇异谱分析由 Colebrook^[7]提出并最早应用于海洋学的研究,然而许多时间序列数据分析的领域实际上都可以使用奇异谱分析(singular spectrum analysis, SSA)方法,但迄今为止将其用于旅客运输量的预测研究相对较少^[8-9]。梁小珍等^[10]提出基于 SSA 的集成模型预测中国航空客运量,并将其与经过 EMD(经验模态分解)降噪处理的预测结果对比,表明 SSA 的预测效果更好。胡茂政等^[11]提出了一种奇异谱分析(SSA)和改进萤火虫算法(IFA)优化 BP 神经网络结合的 SSA-IFA-BP 组合模型预测短时交通流情况。

1 理论基础

1.1 奇异谱分析

奇异谱分析(SSA)可以被广泛应用于处理复杂的非线性时间序列数据,它可以将原始序列进行分解与重构,从而得到一个消除噪声的时间序列,该序列不仅包含了原始序列基本的变化特征,而且比原始序列更加平滑,提高了预测模型的精度。奇异谱分析的步骤如下。

(1)嵌入。假设有一个长度为 T 的时间序列 $X(t) = (x_1, x_2, \dots, x_T)$,通过窗口长度 L 确定 SSA 中嵌入的维数,其中 L 为整数, $1 < L < T$,且 $X_i = (x_i, x_{i+1}, \dots, x_{i+L-1})$,根据选取的 L ,构造轨迹矩阵

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & & \vdots \\ x_L & x_{L+1} & \cdots & x_T \end{bmatrix}, \text{其中 } K = T - L + 1.$$

(2)奇异值分解(SVD)。对 $L \times L$ 阶矩阵 $\mathbf{X}\mathbf{X}^T$ 利用奇异值分解得到 L 个特征值,对其进行降序排列为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L \geq 0$,每个特征值对应的特征向量分别记为 $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_L$ 。令 $\mathbf{V}_i = \frac{\mathbf{X}^T \mathbf{U}_i}{\sqrt{\lambda_i}} (i = 1, 2, \dots, L)$, $d = \text{rank}(\mathbf{X}) = \max\{i; \lambda_i > 0\}$,则原始序列可分解为 $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_d$,其中 $\mathbf{X}_i = \sqrt{\lambda_i} \mathbf{U}_i \mathbf{V}_i^T (i = 1, 2, \dots, d)$ 。

(3)分组。在分组步骤中,可以选择分析周期图、右特征向量散点图或特征值函数图来区分信号和噪声。在信号重构的过程中,对于分组的方式没有特定的规则,可根据待重构时间序列的变化特征将下标 $\{1, 2, \dots, d\}$ 划分成 M 个不相交的子集,即 $I_1, I_2, \dots, I_M$ 。记 $I_i = \{i_1, i_2, \dots, i_p\}$ 为第 I_i 组对

应的下标,则有

$$\begin{cases} \mathbf{X}_{I_i} = \mathbf{X}_{i_1} + \mathbf{X}_{i_2} + \dots + \mathbf{X}_{i_p} \\ \mathbf{X} = \mathbf{X}_{I_1} + \mathbf{X}_{I_2} + \dots + \mathbf{X}_{I_M} \end{cases} \quad (1)$$

对于分量 $\mathbf{X}_{I_i}$ 的贡献率,可由其特征值所占的比例 $\sum_{i \in I_i} \lambda_i / \sum_{i=1}^d \lambda_i$ 来计算。

(4)对角平均化。取矩阵 $\mathbf{X}_{I_i} = (x_{ij})_{L \times K}$ 对角元素平均化,将其转化成长度为 T 的时间序列 $\text{RC}_{I_i} = (\text{rc}_1, \text{rc}_2, \dots, \text{rc}_T)$,其中 RC_{I_i} 第 k 个元素为满足 $i + j = k + 1$ 矩阵 $\mathbf{X}_{I_i}$ 中所有元素的均值。对角平均化的公式为

$$\text{rc}_k = \begin{cases} \frac{1}{k} \sum_{m=1}^k x_{m, k-m+1}, & 1 \leq k < L \\ \frac{1}{L} \sum_{m=1}^L x_{m, k-m+1}, & L \leq k < K \\ \frac{1}{T-k+1} \sum_{m=k-K+1}^L x_{m, k-m+1}, & K \leq k \leq T \end{cases} \quad (2)$$

通过对角平均化,原始时间序列 $X(t) = (x_1, x_2, \dots, x_T)$ 可分解为 M 个时间序列之和,即

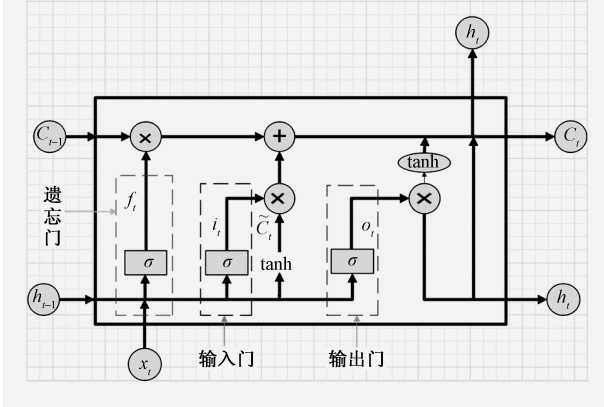
$$X(t) = \sum_{i=1}^M \text{RC}_{I_i} \quad (3)$$

原始时间序列通过奇异谱分析,可以有效地将其中的信号和噪声区分开,经过奇异值分解的时间序列,前 r 个特征值对应的分量中包含了时间序列的绝大多数的特征。将前 r 个分量归为一组,记为 $I_1 = \{1, 2, \dots, r\}$,将其称为信号序列,该组序列不仅包含了原始序列的重要特征,也是重构成预测模型的输入序列,将其余分量归为另一组,记为 $I_2 = \{r+1, r+2, \dots, L\}$,将其称为噪声序列,该组序列包含了原始序列的噪声成分。

1.2 长短时记忆神经网络

长短时记忆神经网络(long-short term memory, LSTM)是一个具备深度学习的递归神经网络,相比于传统的循环神经网络(recurrent neural network, RNN),LSTM 采用一个细胞状态和 3 个门的组合,大大提升了隐含层的性能,从而使得它能够更好地实现对时序数据的长期记忆,且梯度消失和梯度爆炸的问题都可以得到有效避免。LSTM 的结构原理如图 1 所示。具体计算公式为

$$\begin{cases} f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \\ C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \\ \sigma_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t * \tanh(C_t) \end{cases} \quad (4)$$



x_t 为输入序列; i_t 为 t 时刻输入门状态; f_t 为 t 时刻遗忘门状态; 数值 1 为完全保留, 数值 0 表示完全丢弃; o_t 为 t 时刻输出状态; h_t 、 C_t 分别为 t 时刻隐藏状态和单元状态; h_{t-1} 、 C_{t-1} 分别为 $t-1$ 时刻隐藏状态和单元状态; σ 为激活函数 tanh 和 sigmoid

图 1 LSTM 结构原理

式中: w_f 、 w_i 、 w_c 、 w_o 分别代表遗忘门、输入门、记忆单元和输出门相应的权值向量; b_f 、 b_i 、 b_c 、 b_o 分别代表遗忘门、输入门、记忆单元和输出门相应的偏差变量; $\tilde{C}_t$ 为当前输入的单元状态; σ 为激活函数 sigmoid, 其作用是将 f_t 、 i_t 、 o_t 的值限定在区间 $[0, 1]$ 的实数; $\tanh$ 表示 tanh 函数, 其作用是将输出值限定在区间 $[-1, 1]$ 上; $*$ 为矩阵的 Hadamard 积。

1.3 自回归移动平均模型

自回归移动平均模型 (auto-regressive moving average model, ARMA) 是基于线性的自回归模型 (auto-regressive model, AR) 和移动平均模型 (moving average model, MA) 建立起来的平稳时间序列模型, ARMA(p, q) 模型的结构如下:

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \\ \phi_p \neq 0, \theta_q \neq 0 \\ E(\epsilon_t) = 0, \text{Var}(\epsilon_t) = \sigma_\epsilon^2, E(\epsilon_t \epsilon_s) = 0, s \neq t \\ E(x_s \epsilon_t) = 0, \forall s < t. \end{cases} \quad (5)$$

式中: $E(\epsilon_t)$ 为 ϵ_t 的数学期望; $\text{Var}(\epsilon_t) = \sigma_\epsilon^2$ 为 ϵ_t 的方差为 σ_ϵ^2 。若 $\phi_0 = 0$, 该模型称为中心化 ARMA(p, q), 缺省默认条件, 中心化 ARMA(p, q) 模型可以简写为

$$x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \quad (6)$$

式中: x_t 为数据的预测值; $\{x_{t-1}, x_{t-2}, \dots, x_{t-p}\}$ 为前 p 个时刻的数据; $\{\epsilon_t, t = 0, 1, 2, \dots\}$ 为白噪声序列; $\{\phi_1, \phi_2, \dots, \phi_p\}$ 为 AR 模型的系数; $\{\theta_1, \theta_2, \dots, \theta_q\}$ 为 MA 模型的系数; p 和 q 的取值可根据样本自相关系数和偏自相关系数确定。

2 数据预处理及模型评价指标

2.1 数据来源

本文的数据来源于中国国家统计局官网发布的 2005 年 1 月 1 日至 2022 年 12 月 1 日旅客运输量的实际数据, 总计有 216 条数据, 涉及公路、铁路、水运以及航空 4 种运输方式, 分别记录了各种运输方式的客运量当期值 (其中 2007—2016 年每年的 12 月份以及 2018 年 12 月份, 共计有 11 个月缺少公路和水运的旅客运输量的实际数据, 采用均值法进行填充), 数据的单位为万人, 将 4 种运输方式的旅客运输量相加即可得到旅客运输量的总值。

2.2 数据归一化

旅客运输量数据是模型训练的核心, 由于旅客运输量数据具有很大的数量级, 因此需要对数据进行归一化处理。对收集的旅客运输量数据进行归一化处理所使用的是极值法, 即

$$Y' = \frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}} \quad (7)$$

式中: Y' 为归一化后的旅客运输量的数据; Y 为原始旅客运输量数据; $Y_{\min}$ 为旅客运输量数据中的最小值; $Y_{\max}$ 为旅客运输量数据中的最大值。

经过归一化处理后的数据将其输入到模型中进行训练与预测, 最后还需要对得到的预测值进行反归一化, 即

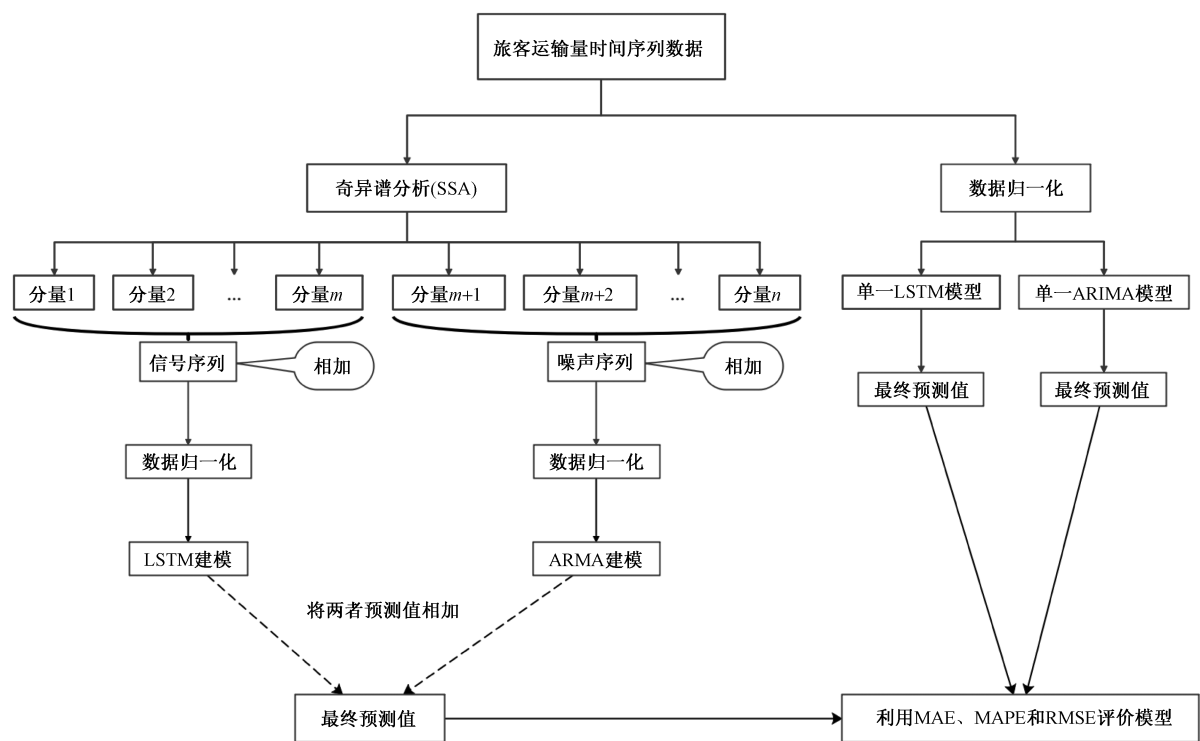
$$\hat{Y} = (Y_{\max} - Y_{\min}) Y_{\text{pred}} + Y_{\min} \quad (8)$$

式中: $\hat{Y}$ 为反归一化后的预测值; Y_{pred} 为归一化后的旅客运输量预测值。

2.3 实验内容

基于奇异谱分析建立的 SSA-LSTM-ARMA 模型的流程如图 2 所示。由图 2 可知, 首先利用 SSA 将旅客运输量的原始时间序列分解为 n 个分量。其次, 根据各个分量的方差贡献率进行排序, 将前 m 个分量重构为信号序列, 将剩余的 $n-m$ 个分量重构为噪声序列。最终的预测结果可以通过对信号序列进行 LSTM 建模和对噪声序列进行 ARMA 建模加总两者的预测值。

模型的主要参数设置包括学习率、神经元结点数、回溯时间步长、数据批处理、激活函数等。在网络结构设计中, 考虑到模型的计算复杂度和计算效率, 相关参数的设定范围为: 学习率 0.1~0.000 01; 神经元结点数 32~128; 回溯时间步长 5~30; 数据批处理 [64, 128, 256]; LSTM 激活函数 [relu, sigmoid, tanh, elu]。对旅客运输量数据进行预测的具体参数设置见表 1。



MAE 为平均绝对误差;MAPE 为平均绝对百分比误差;RMSE 为均方根误差

图 2 SSA-LSTM-ARMA 建模流程

表 1 SSA-LSTM-ARMA 模型参数设置

参数	参数描述	设定值
r	学习率	0.001
Batch_size	数据批处理	128
Activation	激活函数	tanh
Train:Test	训练集:测试集	206:10
LSTM_filters	神经元结点数	100
Look_back	回溯时间步长	12
Epoch	迭代次数	150

2.4 评价指标

预测模型的可靠程度和精确程度可以根据统计预测模型的绝对误差和相对误差进行综合评估。综合考虑选取常见的 3 种模型评价指标,分别为平均绝对误差(MAE)、平均绝对百分比误差(MAPE)和均方根误差(RMSE),其 MAE、MAPE 和 RMSE 的定义公式如式(9)~式(11)所示。

$$MAE = \frac{\sum_{i=1}^T |y_i - \hat{y}_i|}{T} \tag{9}$$

$$MAPE = \frac{\sum_{i=1}^T \frac{|y_i - \hat{y}_i|}{y_i}}{T} \tag{10}$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^T (y_i - \hat{y}_i)^2} \tag{11}$$

式中: T 为时间序列的长度; y_i 、 $\hat{y}_i$ 分别为第 i 期对应的真实值和预测值。由式(9)~式(11)可知,当 3 个模型评价指标对应的数值愈小,则表明预测值与真实值的误差就愈小,模型的拟合效果就愈佳。

3 实证分析

3.1 基于 SSA 的旅客运输量数据分解

通过设置选取窗口长度 $L=12$,即将序列分解为 12 个不同的分量,每个重构分量对应的特征值和方差贡献率见表 2。由表 2 可知,原始序列前两个特征值对应的方差贡献率累计达到了 99%,反映了原始序列的主要信息,且从第 3 个重构分量开始,其方差贡献率均小于 0.005。考虑到不应选取过多的重构分量,否则会削弱奇异谱分析进行降噪的效果,因此通过综合考虑,决定将前两个分量重构得到一个信号序列,将其余的 10 个分量重构得到一个噪声序列。

在图 3 中,由于统计口径的变化,城市公交和出租车不再计入公路运输量,其次还有高铁和民航的发展使得长距离出行减少了换乘次数,导致了 2013 年至 2014 年旅客运输量出现了骤降的趋势。2019 年、2020 年由于受到新冠肺炎疫情的影响,国家制定了严格的交通出行政策,因此旅客运输量亦呈现出大幅度下降的趋势。

表 2 SSA 重构分量对应的特征值和方差贡献率

序号	特征值	特征值累计求和	方差贡献率/%	累计方差贡献率/%
1	$9.872\ 283\ 25\times10^{13}$	$9.872\ 283\ 25\times10^{13}$	$9.845\ 204\ 11\times10^{-9}$	$9.845\ 204\ 11\times10^{-9}$
2	$5.503\ 540\ 01\times10^{11}$	$9.927\ 318\ 650\ 1\times10^{13}$	$5.488\ 444\times10^{-9}$	$9.900\ 088\ 55\times10^{-9}$
3	$2.064\ 673\ 33\times10^{11}$	$9.947\ 965\ 383\ 4\times10^{13}$	$2.059\ 01\times10^{-9}$	$9.920\ 678\ 65\times10^{-9}$
4	$1.425\ 376\ 53\times10^{11}$	$9.962\ 219\ 148\ 7\times10^{13}$	$1.421\ 467\times10^{-9}$	$9.934\ 893\ 32\times10^{-9}$
5	$1.184\ 723\ 03\times10^{11}$	$9.974\ 066\ 379\times10^{13}$	$1.181\ 473\times10^{-9}$	$9.946\ 708\ 05\times10^{-9}$
6	$1.073\ 979\ 14\times10^{11}$	$9.984\ 806\ 170\ 4\times10^{13}$	$1.071\ 033\times10^{-9}$	$9.957\ 418\ 39\times10^{-9}$
7	$9.634\ 632\ 77\times10^{10}$	$9.994\ 440\ 803\ 17\times10^{13}$	$9.608\ 21\times10^{-9}$	$9.967\ 026\ 59\times10^{-9}$
8	$8.859\ 017\ 92\times10^{10}$	$1.000\ 329\ 982\ 109\times10^{14}$	$8.834\ 72\times10^{-9}$	$9.975\ 861\ 31\times10^{-9}$
9	$8.627\ 148\ 77\times10^{10}$	$1.001\ 192\ 696\ 986\times10^{14}$	$8.603\ 48\times10^{-9}$	$9.984\ 464\ 79\times10^{-9}$
10	$5.755\ 533\ 89\times10^{10}$	$1.001\ 768\ 250\ 375\times10^{14}$	$5.739\ 75\times10^{-9}$	$9.990\ 204\ 54\times10^{-9}$
11	$5.228\ 742\ 78\times10^{10}$	$1.002\ 291\ 124\ 653\times10^{14}$	$5.214\ 4\times10^{-9}$	$9.995\ 418\ 94\times10^{-9}$
12	$4.593\ 658\ 73\times10^{10}$	$1.002\ 750\ 490\ 526\times10^{15}$	$4.581\ 06\times10^{-9}$	1

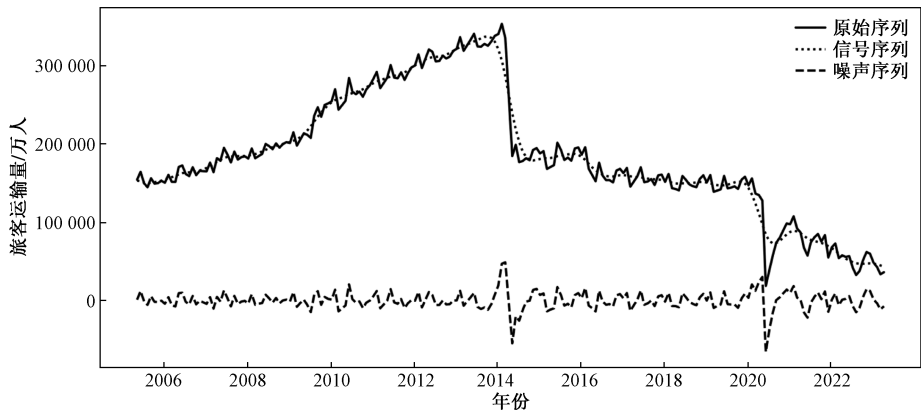


图 3 旅客运输量原始序列、信号序列和噪声序列时序图

通过观察时序图发现,信号序列比原始序列更加光滑,能够更好地捕捉到原始序列的变化特征。这表明,奇异谱分析在有效降低原始序列噪声的同时,还能够对原始序列的主要信息进行提取。

3.2 模型效果比较分析

为了体现基于奇异谱分析建立的 SSA-LSTM-ARMA 模型在旅客运输量预测研究的有效性和准确性,将单一的 LSTM 模型和 ARIMA 模型预测结果与其进行比较,对应的预测误差见表 3。

由表 3 可知,基于奇异谱分析建立的 SSA-LSTM-ARMA 模型对旅客运输量的预测效果明显优于单一的 ARIMA(3,1,2)模型和 LSTM 模型,其对应的 MAE、MAPE 和 RMSE 分别为 9 288.52、19.86 和 11 653.16,这 3 个评价指标的数值均优于其他两个模型。MAE 和 RMSE 的对比结果表明,SSA-LSTM-ARMA 模型的预测误差较小,模型的精度更高,MAPE 的对比结果表明,SSA-LSTM-ARMA 模型更加稳定。综合上述分析可知,SSA-LSTM-ARMA 模型的预测值更接近旅客运输量的

表 3 SSA-LSTM-ARMA、LSTM 和 ARIMA 对旅客运输量的预测效果

模型	MAE	MAPE	RMSE
ARIMA(3,1,2)	14 028.58	36.79	16 078.26
LSTM	10 875.12	27.05	13 063.31
SSA-LSTM-ARMA(2,3)	9 288.52	19.86	11 653.16

真实值,说明了该模型的有效性。进行 SSA 降噪处理后的旅客运输量数据能够更好地提取其变化特征,使模型的预测性能得到提升,预测效果更佳。

4 结论

(1)通过比较单一模型的预测效果,经过 SSA 降噪处理后的重构序列对原始序列的预测效果更好,说明 SSA 在获取序列的主要变化特征和降噪方面具有明显的优势,即 SSA 有助于模型在测试集中降低误差,提高预测精度。

(2)SSA-LSTM-ARMA 模型基于奇异谱分析理论,克服了利用单一模型预测的片面性,在更加准确预测旅客运输量变化的基础上,结合了各单一模型的优点。

参考文献

- [1] 李黎莎. 基于 ARIMA 模型的民航旅客运输量预测[J]. 科学技术创新, 2022(4): 9-12.
- [2] 周元伟, 兰翔. 2022 年民航客运市场旅客运输量预测报告[J]. 空运商务, 2021(11): 30-33.
- [3] 王晨晨. 回归分析在成都地区航空运输规模预测中的应用[J]. 中国科技信息, 2022(7): 46-48.
- [4] 王飞, 韩翔宇. 基于 VMD-LSTM 的民航运输量预测[C]//2022 世界交通运输大会(WTC2022)论文集(交通工程与航空运输篇). 北京: 人民交通出版社股份有限公司, 2022: 674-683.
- [5] 田郝青. 基于季节性预测模型的铁路旅客运输量预测[J]. 西部交通科技, 2015(8): 81-85.
- [6] 孔建国, 李骁. 基于 MATLAB 的小波分析在民航旅客运输量预测中的应用研究[J]. 中国科技信息, 2014(17): 40-41.
- [7] COLEBROOK J M. Continuous plankton records : zooplankton and environment, northeast atlantic and north sea, 1948-1975[J]. Oceanologica Acta, 1978, 5(1): 473-480.
- [8] HASSANI H, WEBSTER A, SILVA E, et al. Forecasting US tourist arrivals using optimal singular spectrum analysis[J]. Tourism Management, 2015, 46: 322-335.
- [9] BENEKI C, ECKELS B, LEON C. Signal extraction and forecasting of the UK tourism income time series: a singular spectrum analysis approach[J]. Journal of Forecasting, 2009, 31(5): 391-400.
- [10] 梁小珍, 乔晗, 汪寿阳, 等. 基于奇异谱分析的我国航空客运量集成预测模型[J]. 系统工程理论与实践, 2017, 37(6): 1479-1488.
- [11] 胡茂政, 张剑, 王文翠. 基于奇异谱分析和改进的 BP 神经网络交通流预测[J]. 智能计算机与应用, 2023, 13(9): 89-94.

Research on Passenger Traffic Forecast Based on Singular Spectrum Analysis

FANG Cheng¹, YANG Zhengru², REN Jianbao³, TAN Yingying¹

(1. College of Mathematics and Physics, Anhui Jianzhu University, Hefei 230601, China;

2. Huainan Jianfa Municipal Engineering Co., Ltd., Huainan 232007, Anhui, China;

3. College of Electronic and Information Engineering, Anhui Jianzhu University, Hefei 230601, China)

Abstract: Scientific and accurate prediction of passenger transport volume can provide effective reference for transportation related departments. Taking passenger transport volume as the research object, based on SSA (singular spectrum analysis), combined with LSTM (long-short term memory neural network) and ARMA (auto-regressive moving average model), the time series of passenger transport volume was decomposed into signal sequence and noise sequence through SSA noise reduction processing, and LSTM and ARMA(2,3) modeling were carried out on them respectively. Based on this, its changing trend is predicted. By comparing the experimental results of single ARIMA(3,1,2) model and LSTM model, it shows that SSA-LSTM-ARMA has better prediction effect and higher prediction accuracy in passenger traffic volume.

Keywords: passenger traffic; singular spectrum analysis; LSTM(long-short term memory neural network); ARMA(auto-regressive moving average model)