

基于PSO-LSTM模型对上海某院糖尿病患者住院量的预测研究

周 琪¹, 费哲为^{2*}, 樊重俊¹, 刘 红^{1,3}, 熊红林⁴

¹上海理工大学管理学院, 上海

²上海健康医学院附属崇明医院普外科, 上海

³上海健康医学院附属崇明医院信息科, 上海

⁴上海健康医学院协同科研中心, 上海

收稿日期: 2024年11月24日; 录用日期: 2025年2月4日; 发布日期: 2025年2月12日

摘 要

目的: 当前医疗管理在资源分配和效率提升方面缺乏精确的住院量预测方法, 制约了卫生系统的优化。为解决这一问题, 本研究提出一种基于粒子群优化的长短期记忆网络(PSO-LSTM)模型, 为医疗资源配置提供科学支持。方法: 采用2013年至2023上半年上海崇明地区某医院的住院量数据, 构建并优化PSO-LSTM模型。PSO算法通过100次迭代对LSTM模型的关键超参数进行全局优化, 并与ARIMA、SARIMA和SVR等模型进行对比分析。结果: PSO-LSTM模型在MSE、RMSE、MAE和R²等指标上表现最佳, 相较于SVR模型, MSE降低81.7%, R²提高26.2%。结论: 研究结果表明, PSO-LSTM模型有效弥补了传统方法的局限性, 为精准预测住院量提供了一种科学的解决方案, 有效提高医疗资源配置效率, 为卫生系统管理决策提供支持。

关键词

PSO-LSTM模型, 统计预测, 时间序列分析, 住院人次

Prediction Study of Hospitalization Volume in a Hospital in Shanghai Based on PSO-LSTM Model

Qi Zhou¹, Zhewei Fei^{2*}, Chongjun Fan¹, Hong Liu^{1,3}, Honglin Xiong⁴

¹Business School, University of Shanghai for Science and Technology, Shanghai

²General Surgery Department, Shanghai University of Medicine & Health Science Affiliated Chongming Hospital, Shanghai

*通讯作者。

文章引用: 周琪, 费哲为, 樊重俊, 刘红, 熊红林. 基于 PSO-LSTM 模型对上海某院糖尿病患者住院量的预测研究[J]. 运筹与模糊学, 2025, 15(1): 128-137. DOI: 10.12677/orf.2025.151013

³Information Department, Shanghai University of Medicine & Health Science Affiliated Chongming Hospital, Shanghai

⁴Collaborative Research Center of Shanghai University of Medicine & Health Science, Shanghai

Received: Nov. 24th, 2024; accepted: Feb. 4th, 2025; published: Feb. 12th, 2025

Abstract

Objective: The lack of accurate inpatient volume prediction methods in current healthcare management hinders effective resource allocation and system optimization. To address this issue, this study proposes a Particle Swarm Optimization-based Long Short-Term Memory (PSO-LSTM) model to provide scientific support for the allocation of medical resources. **Methods:** Using inpatient data from a hospital in Chongming, Shanghai, spanning 2013 to the first half of 2023, the PSO-LSTM model was developed and optimized. The PSO algorithm conducted 100 iterations to globally optimize key hyperparameters of the LSTM model. The performance of the proposed model was compared against ARIMA, SARIMA, and SVR models. **Results:** The PSO-LSTM model demonstrated superior performance across all metrics, including MSE, RMSE, MAE, and R^2 . Compared to the SVR model, the PSO-LSTM model achieved an 81.7% reduction in MSE and a 26.2% improvement in R^2 . **Conclusion:** The results indicate that the PSO-LSTM model effectively addresses the limitations of traditional methods, providing a robust solution for accurate inpatient volume prediction. This model enhances the efficiency of medical resource allocation and supports healthcare system management decisions.

Keywords

PSO-LSTM Model, Statistical Forecasting, Time Series Analysis, Hospitalization Volume

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,糖尿病在我国的患病率持续攀升,已成为影响公共健康的重大慢性疾病之一。根据《中华人民共和国卫生健康事业发展统计公报(2023 年)》,我国糖尿病患者人数已居世界前列,且患者呈现年轻化趋势。现代社会中不健康的饮食结构、久坐生活方式及肥胖率的增加,进一步推动了糖尿病的高发态势。糖尿病不仅显著降低患者的生活质量,还可能导致心血管疾病、肾功能衰竭、视网膜病变等严重并发症,是导致死亡和致残的主要原因之一。这些特性决定了糖尿病的防治具有长期性和复杂性,亟需医疗系统提供精准且高效的资源支持[1]。在此背景下,糖尿病患者的住院需求也在快速增长,对床位、医护人员及医疗设备的需求不断增加,使医疗系统面临巨大的供需压力。此外,糖尿病相关住院费用占医疗总支出的比例逐年升高,对卫生系统的经济负担日益加重。为应对这一挑战,《“健康中国 2030”规划纲要》和《“十四五”国家卫生健康规划》明确提出,加强慢性病防控及优化医疗资源配置[2] [3]。然而,现有的医疗管理方法缺乏科学的预测模型,无法有效应对糖尿病住院需求的激增。这一问题严重制约了医疗资源的合理分配,影响了卫生系统的运行效率[4]。

科学预测糖尿病患者的住院趋势,对优化医疗资源配置至关重要。合理的预测能够帮助医疗管理者

在床位、医护人员、药品等资源的分配上做出更科学的决策,从而更好地满足患者需求,避免资源浪费,并提升整体服务质量。刘焰等人[5]利用移动平均和季节指数法结合最小二乘回归模型,对某医院门诊量数据进行分析,揭示了门诊量的逐年增长趋势和季节波动规律,并对2019年的门诊量进行了预测。传统的数理统计方法虽然在揭示数据的季节性变化上有一定成效,但在处理复杂、非线性数据和突发事件的动态变化方面存在明显不足。尤其是,这些方法在长期趋势预测、异常值处理等方面效果有限,难以满足现代医疗管理对高精度和灵活性的需求[6]。

鉴于传统方法的局限性,现代时间序列预测模型已逐渐被引入到住院趋势研究中,其中包括自回归移动平均模型(ARIMA)、支持向量回归(SVR)和长短期记忆神经网络(LSTM)等。廉恒丽等人[7]利用ARIMA和SARIMA模型对某医院门诊人次和出院人次数据进行预测,证明了这两种模型在短期趋势预测中的实用价值。陈文娟等人[8]利用季节ARIMA模型对浙江省某三级综合性医院的门诊量进行建模和预测,揭示了门诊量的逐年上升趋势及周期性波动特征。ARIMA模型、SARIMA模型在揭示短期波动方面表现良好,但在应对高非线性复杂数据时效果不佳[9]。相比之下,LSTM模型作为一种深度学习方法,能够捕捉长期依赖特征,在非线性时间序列预测方面表现更为出色[10]。研究表明,LSTM在处理住院趋势数据的长期依赖和非线性方面具有显著优势,但其超参数调优复杂、训练时间长且在动态数据变化时效果有限[11]。为提高LSTM模型的预测性能,近年来有研究者将优化算法与LSTM结合,提出了诸如PSO-LSTM(基于粒子群优化的LSTM)模型。Kumar等人[12]研究表明,PSO-LSTM在超参数优化和应对复杂非线性数据方面具有优势,使得该模型在住院预测中展现了更高的精确度与稳定性。

基于此,本研究旨在构建一个能够精确预测糖尿病患者住院趋势的模型,以支持医疗资源的合理配置和管理效率的提升。本文首先分析了2013年至2023年上半年上海崇明地区糖尿病患者的住院数据,创新性地采用以周为单位的住院人次数据,以避免月度数据可能掩盖波动性的风险,同时揭示住院人次的季节性波动与长期趋势及其成因。在此基础上,研究采用PSO-LSTM模型对糖尿病住院数据进行建模,并与ARIMA、SARIMA、SVR及未优化的LSTM模型进行精度对比。通过结合LSTM对时间序列非线性和周期性规律的捕捉能力以及PSO算法在优化超参数方面的优势,进一步提升了模型的预测性能。实验结果验证,PSO-LSTM模型在MSE、MAE等指标上的表现显著优于对比模型,展现出卓越的预测精度与鲁棒性。研究结果为医疗管理者提供了可靠的依据,能够帮助其提前应对住院高峰期,优化资源配置,并进一步提升医疗服务质量。

2. 资料与方法

2.1. 数据来源

研究数据来源于上海市崇明区某医院信息统计科,其中包含2013年至2023年上半年每周糖尿病患者住院人次,总的记录共计547条记录。纳入标准:入院主要诊断包含符合WHO制定的糖尿病患者,年龄 >18 岁,且具有完整的住院记录。排除标准:妊娠糖尿病、言语意识障碍、心力衰竭、恶性肿瘤患者。

2.2. 研究方法

2.2.1. LSTM模型

在处理季节性强、趋势明显且具有非线性关系的时间序列数据时,传统的循环神经网络(RNN)由于其简单的结构和梯度传播机制,往往表现出一定的局限性,尤其是像住院人次这类长时间跨度时序数据容易出现梯度消失或梯度爆炸的问题[13]。为了解决这些挑战,Hochreiter和Schmidhuber(1997)提出了长短期记忆(LSTM)网络。LSTM通过引入输入门、遗忘门和输出门这三个核心门控机制,能够精细控制细胞状态(Cell State, C_t)和隐藏状态(Hidden State, h_t)的更新,从而在较长的时间跨度内有效保留和提取关键

的时间序列特征[14]。在 LSTM 结构中,记忆单元包括输入门、遗忘门和输出门,这些门控机制决定了输入信息如何影响当前的记忆状态、前一时刻的记忆如何在当前时刻保留,以及最终的输出状态如何生成。在 LSTM 单元内部,首先,遗忘门(Forget Gate, f_t)决定前一时刻的细胞状态有多少应被保留。遗忘门的计算如下:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

其中, W_f 是遗忘门的权重矩阵, b_f 为偏置项, σ 是 sigmoid 激活函数。遗忘门根据当前输入和前一时刻的隐藏状态决定哪些信息需要在当前细胞状态中保留或舍弃。接下来,输入门(Input Gate, i_t)控制当前时刻的新信息如何影响细胞状态,其计算公式为:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

与此同时,候选细胞状态($\tilde{C}_t$)由 tanh 函数计算得出:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

而输入门的作用是决定这些候选信息有多少会被引入到细胞状态中。最后,输出门(Output Gate, o_t)控制当前细胞状态中的信息输出到下一个时间步的隐藏状态 o_t :

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (4)$$

随后,通过计算得到当前的隐藏状态:

$$h_t = o_t \times \tanh(C_t) \quad (5)$$

这些门的组合使得 LSTM 能够在复杂的时间序列预测任务中有效捕捉长期依赖关系,并且在处理季节性和趋势性显著的非线性数据时表现出色。

2.2.2. PSO 模型

粒子群优化(Particle Swarm Optimization, PSO)算法是一种基于群体智能的全局优化算法,由 Kennedy 和 Eberhart 于 1995 年提出[15]。PSO 的灵感来源于自然界中群体协作行为,如鸟群觅食或鱼群迁徙,通过模拟群体中个体(即“粒子”)之间的合作和竞争,以集体智慧搜索最优解。

在 PSO 中,假设有一个由 N 个粒子组成的群体,每个粒子在 D 维搜索空间中表示为 $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$, 其中 $i = 1, 2, \dots, N$ 。粒子的速度也表示为一个向量 $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})$, 该速度决定了粒子在搜索空间中的移动方向和距离。每个粒子都会记录自己在搜索过程中找到的历史最佳位置,称为个体最优(pBest),同时,粒子群中也会保留一个全局最优位置(gBest),即群体中所有粒子所经历的最优位置。在 PSO 中,粒子的速度和位置更新遵循:

$$v_{ij}(t+1) = w \cdot v_{ij}(t) + c_1 \cdot r_1 \cdot (p_{ij}(t) - x_{ij}(t)) + c_2 \cdot r_2 \cdot (p_{gj}(t) - x_{ij}(t)) \quad (6)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (7)$$

其中, w 为惯性权重,用于平衡全局与局部搜索能力; c_1 和 c_2 为学习因子,常用值在 0.1 至 2 之间; r_1 和 r_2 为随机数,取值范围为[0, 1]; $p_{ij}(t)$ 和 $p_{gj}(t)$ 分别代表个体最优和全局最优位置。通过迭代更新,粒子群逐渐趋于全局最优解。

2.2.3. 数据预处理

本研究首先导入了 2013~2023 上半年时间序列数据,包含每日的住院人数。本研究使用周为单位的数据进行分析,而非月度数据,主要原因在于住院人次在月内可能存在剧烈波动,按月统计会掩盖这些

波动,而周数据则可以更准确地反映这些动态变化。因此,为确保分析的准确性和决策的有效性,本研究选择以周为单位进行时间序列分析。在数据导入后,首先进行了数据清洗,处理了缺失值和异常值。针对缺失值,采用了向前填充(forward fill)方法,以保持时间序列的连续性,避免破坏时序模式[16]。为减少数据噪声及疫情封控期间(2022年3月至6月)数据异常对模型的影响,使用了线性插值法和移动平均平滑技术。线性插值法用于填补因疫情封控缺失的数据,结合窗口大小为12的移动平均法进行平滑处理,从而减小异常值对数据的干扰。此外,采用三点平滑法进一步处理住院人次数据中的极端值,生成平滑后的数据列(Smoothed_Patient_Count)。为了提高模型的收敛性和预测稳定性,使用 Scikit-learn 库中的 MinMaxScaler 将 Smoothed_Patient_Count 归一化至[0,1]区间,避免不同特征量级差异的影响[17]。最后,为进一步提高数据处理的全面性,应用了基于树的无监督异常检测算法——IsolationForest,自动识别并处理数据中的异常点,以减少其对模型训练和预测的负面影响。

2.2.4. 模型构建

LSTM 模型通过记忆单元和门控机制,能够有效捕捉时间序列中的长期依赖关系,广泛用于处理包含季节性和趋势性的数据[18]。本研究采用了两层 LSTM 结构,每层 50 个神经元,旨在捕捉数据中的长期依赖关系。输入数据为三维形式:样本数、时间步长(设为 3)、每个时间步的特征维度。这一时间步长是通过交叉验证实验得出的,旨在平衡数据的周期性变化与模型的学习效率。输出层使用线性激活函数,预测每周住院人次。LSTM 层使用 tanh 激活函数捕捉非线性特征,tanh 函数有助于避免梯度消失问题并提升长短期记忆的学习能力;细胞状态更新采用 sigmoid 函数控制门控开关[18]。在模型构建过程中,选择了适当的时间步长(time_step)、隐藏层神经元数(units)和批量大小(batch_size)作为关键超参数。为平衡模型的复杂度与计算成本,初始设置为 50 个神经元,后续进行超参数调优。为防止过拟合,在 LSTM 层之间引入 Dropout 层,Dropout 率在 0.2 到 0.5 之间,这一范围是通过实验调优得出的,旨在平衡模型的学习能力和泛化能力。与此同时,还应用了 L2 正则化,通过权重衰减项控制过拟合的风险,保证模型的稳定性和泛化能力[18]。

在模型训练之前,时间序列数据按照时间顺序被划分为训练集和测试集,划分比例为 8:2。测试集用于监控模型性能,防止过拟合。为保持时间序列特性,数据划分时严格按照时间顺序进行,确保测试集中的数据始终比训练集数据晚。此外,本研究采用了滚动窗口交叉验证,通过测试集来监督训练过程以更好地评估模型的泛化能力。

LSTM 超参数的选择通过滚动验证和 PSO 优化算法进行调优,多次实验后得出了最优的超参数组合。本研究中,PSO 的种群规模设为 20,最大迭代次数为 100,适应度函数基于验证集的均方误差(MSE)进行优化[19]。通过动态调整关键超参数(time_step、units、learning_rate、dropout_rate 和 batch_size),PSO 显著提升了 LSTM 模型的性能。适应度函数的收敛条件为:当连续 5 次迭代的 MSE 下降不足 0.01 时停止迭代。

在训练过程中,实验使用 Adam 优化器最小化 MSE 损失函数,初始学习率结合学习率调度器逐步降低,以确保模型平稳收敛。为防止过拟合,实验中使用了早停法(Early Stopping)。在测试集损失不再下降时,提前停止训练,避免过拟合。同时,动态调整批量大小,训练初期使用较小批量加速收敛,后期逐步增大批量以稳定训练过程。训练完成后,对预测结果进行反归一化处理,恢复数据的原始尺度。

2.2.5. 模型评价指标

在完成模型训练后,模型的性能通过测试集进行全面评估。本研究采用了均方误差(MSE)、均方根误差(RMSE)、平均绝对误差(MAE)及决定系数(R^2)用于全面衡量模型预测精度,其中 RMSE 更关注大误差,MAE 反映平均误差, R^2 则反映模型对趋势的拟合能力。这些指标从不同角度衡量了模型的预测效果,能

够为模型性能的全面评估提供依据。

3. 结果

3.1. 住院人次趋势与时间序列分解分析

基于平滑处理后的数据，糖尿病患者住院人次自 2013 年以来整体呈上升趋势。2013 年至 2020 年期间，住院人次波动较大，并且表现出明显的季节性波动。2020 年初，受到新冠疫情影响，住院人次经历了短暂的剧烈波动；2022 年中期，上海市疫情封控措施导致住院人次显著下降。然而，自 2020 年起，住院人次增长加速，并于 2023 年一月初达到了历史新高周住院人次为 46 人，这一趋势与疫情后患者需求增加和医院收治能力提升有关。见图 1。

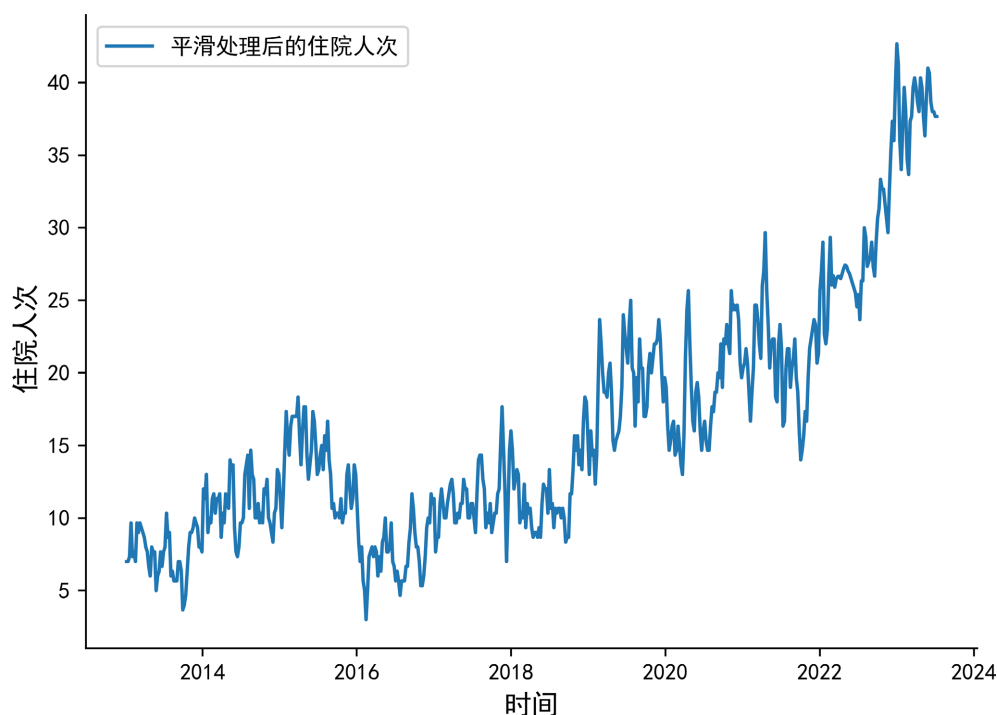


Figure 1. 2013~2023 year trend chart of hospital admissions for diabetic patients in a certain Hospital in Shanghai
图 1. 2013~2023 年上海市某医院糖尿病患者住院人次趋势图

图 1 展示了 2013~2023 年上半年间上海市崇明区某医院糖尿病患者住院人次的总体趋势。为了更全面地揭示数据中的季节性模式和随机波动，进一步对时间序列进行了分解分析。图 2 展示了糖尿病患者住院人次的时间序列分解结果，包括趋势成分、季节性成分和残差成分，以便更深入地理解数据波动特征。

从趋势成分来看，住院人次自 2013 年至 2023 年呈现波动性上升趋势。2013 年至 2016 年间，住院人次总体保持平稳或略微下降；自 2017 年起，住院人次逐步上升，且在 2020 年后显著增加。季节性成分反映了住院人次中的周期性波动。图 2 中显示，住院人次存在明显的季节性模式，每年均出现固定的波动周期，这个周期大致为一年。住院人次的峰值通常出现在冬季(12 月至 2 月)。12 月至次年 2 月，崇明区的气候寒冷湿润。低温和湿度增加了糖尿病患者的血糖波动，血液循环减缓，可能加重并发症，导致住院人次在此期间显著上升。低谷通常出现在每年的夏季(6 月至 8 月期间)。这时期崇明区的气温较高且湿度高，人体新陈代谢加快，部分患者的血糖控制相对稳定，住院人次减少。此外，崇明区的生态环

境较好，夏季患者的户外活动增加，这种生活方式的改善也有助于稳定血糖，从而降低住院人次。从残差成分来看，该时间序列的残差值主要分布在-10 到 10 之间，并集中在零附近，表明时间序列与均值之间的偏差较小，整体波动性较低。残差在时间轴上的分布表现出随机性，没有明显的周期性或趋势，说明时间序列中的噪声是随机的，未表现出系统性偏差。见图 2。

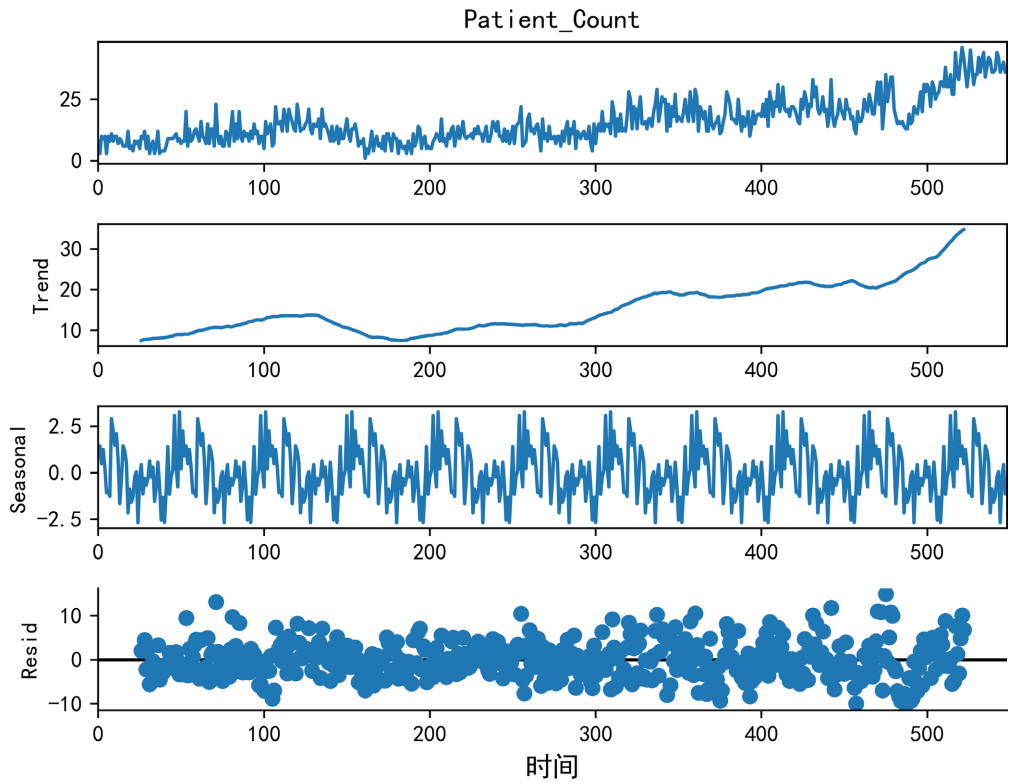


Figure 2. Decomposition of time series of hospitalizations of diabetic patients
图 2. 糖尿病患者住院人次时间序列分解图

3.2. PSO-LSTM 模型预测结果

为了验证所提出模型的性能并为结果提供解释性支持，本研究采用了将预测结果与实际观测数据进行可视化对比的方法。图 3 展示了模型在各个时间段的拟合效能。分析结果表明，在模型训练的初期阶段，预测值与实际观测值之间具有较高的一致性，这反映了模型成功捕捉了数据的内在趋势。然而，2020 年初和 2022 年中期的预测值与实际观测值之间出现了偏差，这可能与新冠疫情期间住院人次的剧烈波动有关。为了降低此类异常波动对模型预测精度的影响，研究中在数据预处理阶段实施了特定策略，包括在 2022 年 3 月至 6 月疫情期间采用线性插值方法来处理缺失数据，并应用移动平均技术以减轻异常值的影响，从而增强了模型的稳定性。此外，为了更准确地捕捉数据中的季节性波动，本研究通过滚动验证实验确定了最优的 `time_step` 参数，使得模型能够更精确地识别并预测年度内的季节性变化。图 3 中的季节性趋势揭示了住院人次在冬季达到高峰而在夏季降低至低谷的规律。利用 LSTM 模型的记忆单元特性以及通过 PSO 算法优化的超参数配置，模型在预测这些季节性模式方面展现了卓越的性能。这些优化措施确保了 PSO-LSTM 模型在大多数时间段内能够精确地拟合实际数据，为医院资源的合理配置和决策提供了坚实的数据基础。最终，图 3 直观地呈现了 PSO-LSTM 模型预测结果与实际观测值之间的对比。在大多数时间段内，预测曲线与实际曲线紧密贴合，特别是在非疫情影响的时期，模型的预测精度尤为显著。

在性能评估方面，模型在多个关键指标上均表现优异，具体包括：均方误差(MSE)为 3.61，均方根误差(RMSE)为 1.90，平均绝对误差(MAE)为 1.44，以及决定系数(R^2)达到 0.93。这些指标共同证实了 PSO-LSTM 模型在捕捉数据的时间依赖性和非线性特征方面的能力，尤其是在处理长时间序列预测任务时，其性能表现尤为卓越。见图 3。

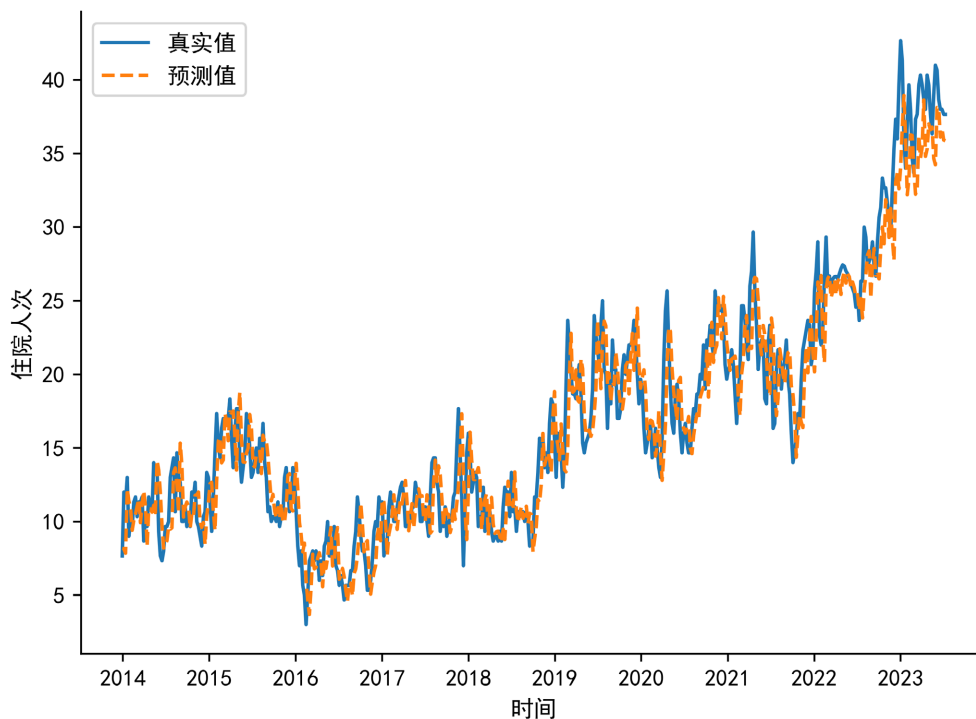


Figure 3. Comparison of PSO-LSTM model prediction results

图 3. PSO-LSTM 模型预测结果对比图

3.3. 模型对比与性能分析

为了验证 PSO-LSTM 模型在时间序列预测任务中的有效性，本研究还对其他几种常见的时间序列预测模型包括 ARIMA、SARIMAX、SVR 和 LSTM 模型。各模型的性能通过相同的数据集和评估指标进行测量，以确保比较的公平性。为了全面评估各模型的性能，本研究采用了均方误差(MSE)、均方根误差(RMSE)、平均绝对误差(MAE)、以及决定系数(R^2)四个指标。表 1 列出了不同模型在测试集上的性能表现：

Table 1. Comparison of the performance of each model on the test set

表 1. 各模型在测试集上的性能对比

模型	MSE	RMSE	MAE	R^2
ARIMA	34.25	5.85	2.81	0.48
SARIMA	29.93	5.47	4.26	0.62
SVR	19.74	4.44	3.53	0.74
LSTM	29.09	5.39	4.22	0.63
PSO-LSTM	3.61	1.90	1.44	0.93

从表 1 的结果可以得知 PSO-LSTM 模型在所有指标上均表现最优, MSE, RMSE, MAE 和 R^2 指标分别为 3.61、1.90、1.44 和 0.93。这表明 PSO-LSTM 模型能够非常准确地捕捉数据中的时间依赖性和非线性关系, 在长时间序列预测中的表现尤为突出。而 SVR 模型在非线性数据处理上表现出色, 具有较低的 RMSE 和 MAE 值, 分别为 4.44 和 3.53, R^2 值也相对较高, 值为 0.74, 但仍不及最优模型 PSO-LSTM 模型。SARIMA 模型以往在捕捉季节性波动方面有优势。然而, 与 PSO-LSTM 模型相比, 模型在该复杂依赖关系的捕捉上有所欠缺, 且 SARIMA 模型的 MAE 较高, 表明其预测的误差较大。LSTM 模型的表现与 SARIMA 接近, 但在某些指标上稍有优势。ARIMA 模型在各项指标中总体表现最差, 特别是 RMSE 值 5.86 较高且 R^2 最低为 0.48, 表明该模型难以有效处理复杂的时间序列数据, 尤其是在数据的长期预测上效果不佳, 该模型可能更适用于简单线性时间序列数据的分析。因此, 基于 PSO-LSTM 模型的住院人次预测模型总体预测表现更佳。该模型方法能够为医院提供更加准确的住院需求预测, 帮助医院管理者更好地规划和分配医疗资源, 如病房安排、医护人员调度等。

4. 结论

本研究利用 2013 年至 2023 上半年上海市崇明地区某医院的糖尿病患者住院数据, 构建了基于 PSO-LSTM 模型的住院人次时间序列预测模型。通过粒子群优化(PSO)算法调优 LSTM 模型的超参数, 显著提升了其在捕捉非线性特征和长时依赖性方面的精度与稳定性。与 ARIMA 和 SVR 等传统模型相比, PSO-LSTM 模型在预测精度方面表现优异, 在测试集上 MSE 降低至 3.61, R^2 提高至 0.93, 相较于 SVR 模型, MSE 降低了 81.7%, R^2 提升了 26.2%。研究结果为预测住院需求趋势和优化医疗资源配置提供了科学依据, 有助于管理者提前应对住院高峰, 合理分配资源并提升医疗服务质量。

尽管 PSO-LSTM 模型在住院量预测方面展示出较好的效果, 但研究存在一定局限性。首先, 数据仅来自上海崇明地区的单一医院, 区域性限制可能导致模型泛化性较差, 难以在其他地区直接应用。未来应扩展数据来源, 涵盖更多地区和多样化人群, 以提高模型的适用性。此外, 数据主要为结构化数值, 缺乏患者个体特征和临床病历等非结构化信息的支持, 限制了模型对个体化因素的预测能力。后续研究可以结合临床文本数据和个体健康记录, 利用自然语言处理技术进一步提升预测深度和精度。同时, 建议在不同医院环境中进行跨机构验证, 以确保模型在多种环境下的稳定性和适用性。

研究结果显示冬季糖尿病患者住院人次显著增加, 提示医院在该季节应提前准备资源, 增加病房床位和医护人员配置, 特别是针对老年和高危患者群体, 实施更为严格的健康监控, 确保医疗服务的连续性。相较之下, 夏季住院人次较少, 医院可以在这一时期加强糖尿病患者的健康教育, 进行血糖控制管理的宣教活动, 帮助患者提升自我管理能力。此外, 夏季也可作为医护人员轮休或技能培训的时机, 优化资源利用率, 为即将到来的冬季高峰做好准备。

基金项目

本工作得到上海健康医学院地方高水平大学建设项目(22MC2022001), 上海市公共卫生重点学科项目(GWVI-11.1-49), 上海市公共卫生体系建设三年行动计划项目(GWVI-6)和上海市卫生健康委员会 2024 年卫生健康政策研究课题(2024HP72)的资助。

参考文献

- [1] 中国政府. 2023 年我国卫生健康事业发展统计公报[EB/OL]. https://www.gov.cn/lianbo/bumen/202408/content_6971241.htm, 2024-10-20.
- [2] 中国政府. “健康中国 2030”规划纲要[EB/OL]. https://www.gov.cn/zhengce/2016-10/25/content_5124174.htm, 2024-10-20.

- [3] 中国政府. “十四五”国民健康规划[EB/OL].
https://www.gov.cn/zhengce/content/2022-05/20/content_5691424.htm, 2024-10-20.
- [4] King, Z., Farrington, J., Utley, M., Kung, E., Elkhodair, S., Harris, S., *et al.* (2022) Machine Learning for Real-Time Aggregated Prediction of Hospital Admission for Emergency Patients. *npj Digital Medicine*, 5, Article 104.
<https://doi.org/10.1038/s41746-022-00649-y>
- [5] 刘焰, 卢萍萍. 基于移动平均季节指数法的门诊量分析及预测[J]. 医学信息, 2021, 34(23): 156-158.
- [6] 陈龙. 南方某省 2010~2019 年住院人次季节变动趋势分析[J]. 中国医院统计, 2020, 27(4): 375-378.
- [7] 廉恒丽, 俞剑琴, 傅映晖. ARIMA 与 SARIMA 模型在医院门诊人次与出院人次预测中的应用[J]. 中国医院统计, 2017, 24(2): 81-84.
- [8] 陈文娟, 林建潮. 基于季节 ARIMA 模型对某三级综合性医院门诊量的预测研究[J]. 中国医院统计, 2024, 31(3): 185-188.
- [9] 张驰, 郭媛, 黎明. 人工神经网络模型发展及应用综述[J]. 计算机工程与应用, 2021, 57(11): 57-69.
- [10] 王淑平, 李敏, 杜敏, 等. ARIMA 与 LSTM 模型在医院出院人次预测中的比较研究[J]. 公共卫生与预防医学, 2021, 32(1): 18-21.
- [11] Hamiane, S., Ghanou, Y., Khalifi, H. and Telmem, M. (2024) Comparative Analysis of LSTM, ARIMA, and Hybrid Models for Forecasting Future GDP. *Ingénierie des systèmes d'information*, 29, 853-861.
<https://doi.org/10.18280/isi.290306>
- [12] Kumar, G., Singh, U.P. and Jain, S. (2022) An Adaptive Particle Swarm Optimization-Based Hybrid Long Short-Term Memory Model for Stock Price Time Series Forecasting. *Soft Computing*, 26, 12115-12135.
<https://doi.org/10.1007/s00500-022-07451-8>
- [13] Duong, N., Nguyen, M., Nguyen, T. and Tran, T. (2023) Application of Seasonal Trend Decomposition Using Loess and Long Short-Term Memory in Peak Load Forecasting Model in Tien Giang. *Engineering, Technology & Applied Science Research*, 13, 11628-11634. <https://doi.org/10.48084/etasr.6181>
- [14] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, 9, 1735-1780.
<https://doi.org/10.1162/neco.1997.9.8.1735>
- [15] Kennedy, J. and Eberhart, R. (1995) Particle Swarm Optimization. *Proceedings of ICNN'95—International Conference on Neural Networks*, Perth, 27 November-1 December 1995, 1942-1948. <https://doi.org/10.1109/icnn.1995.488968>
- [16] Jose, R., Syed, F., Thomas, A. and Toma, M. (2024) Cardiovascular Health Management in Diabetic Patients with Machine-Learning-Driven Predictions and Interventions. *Applied Sciences*, 14, Article 2132.
<https://doi.org/10.3390/app14052132>
- [17] Pedregosa, F., Varoquaux, G., Gramfort, A., *et al.* (2011) Scikit-Learn: Machine Learning in Python. *The Journal of Machine Learning Research*, 12, 2825-2830.
- [18] 李琳, 王哲, 张学良, 等. 基于 LSTM 深度神经网络的月门诊量预测精度研究[J]. 中国数字医学, 2019, 14(1): 14-17.
- [19] Cansu, T., Kolemen, E., Karahasan, Ö., Bas, E. and Egrioglu, E. (2023) A New Training Algorithm for Long Short-Term Memory Artificial Neural Network Based on Particle Swarm Optimization. *Granular Computing*, 8, 1645-1658.
<https://doi.org/10.1007/s41066-023-00389-8>