邢雪, 路康, 罗发辉,等. 融合多头注意力机制及 GWO-LSTM 的地铁流量预测[J]. 智能计算机与应用,2026,16(8):26-32.
DOI:10. 20169/j. issn. 2095-2163. 24121804

融合多头注意力机制及 GWO-LSTM 的地铁流量预测

邢 雪[1], 路 康[1], 罗发辉[1], 路 义[2]
(1 吉林化工大学 人工智能学院, 吉林 吉林 132022; 2 河南科技大学 土木建筑学院, 河南 洛阳 471023)

摘 要: 本文针对杭州地铁流量预测进行了研究,报告了在运营优化、路径规划中预测精度不足的现状,对地铁客流量的动态变化特征进行了深入调查,指出了传统方法在处理非线性和长时序依赖关系时的局限性,为模型设计提供依据。本研究提出了一种融合灰狼优化(Grey Wolf Optimization,GWO)算法、长短时记忆网络(Long Short-Term Memory,LSTM)与多头注意力机制(Multi-Head Attention ,MHA)的预测模型。该模型结合了 LSTM 对时间序列建模的长依赖优势、MHA 对全局特征的捕获能力以及 GWO 的全局搜索能力,对超参数进行优化。通过对杭州某地铁站点客流量数据的实例分析,选取天气、时间等多维度因素,建立了预测模型并开展模型对比实验。结果表明,GWO-LSTM-MHA 模型在平均绝对误差、均方根误差和模型拟合优度上均优于其他四种模型,验证了模型在预测中的稳定性和适用性。

关键词: 地铁站点客流量预测; 灰狼优化算法; 长短时记忆网络; 多头注意力机制; 时间序列分析; 模型优化
中图分类号: U495 **文献标志码:** A **文章编号:** 2095-2163(2026)08-0026-07

GWO-LSTM subway passenger flow prediction based on multi-head attention mechanism

XING Xue[1], LU Kang[1], LUO Fahui[1], LU Yi[2]
(1 College of Artificial Intelligence, Jilin University of Chemical Technology, Jilin 132022, Jilin, China;
2 School of Civil Engineering, Henan University of Science and Technology, Luoyang 471023, Henan, China)

Abstract: This paper conducts research on passenger flow prediction for Hangzhou Metro, analyzing the current problem of insufficient prediction accuracy in operation optimization and path planning. It investigates the dynamic characteristics of subway passenger flow and highlights the limitations of traditional methods in handling nonlinearity and long-term dependencies, providing a basis for model design. A prediction model integrating the Grey Wolf Optimization (GWO) Algorithm, Long Short-Term Memory (LSTM) Network, and Multi-Head Attention Mechanism (MHA) is proposed. The model leverages the long-term dependence modeling capability of LSTM, the global feature extraction ability of MHA, and the global search and hyperparameter optimization capability of GWO. A case study using passenger flow data from a Hangzhou subway station incorporates multidimensional factors such as weather and time to construct and compare prediction models. Results demonstrate that the GWO-LSTM-MHA model outperforms four other models in root mean square error, mean absolute error, and goodness of fit, confirming its stability and applicability in subway passenger flow prediction.

Key words: subway station passenger flow prediction; Grey Wolf Optimization Algorithm; Long Short-Term Memory Networks; Multi-Head Attention Mechanism; time series analysis; model optimization

0 引 言

精准预测地铁站点客流量不仅能提升地铁运营管控的精准性和服务质量,提高运输资源的利用率和调度效率,还能更好地应对客流高峰期、突发事件和拥堵问题。杭州作为中国重要的经济中心和创新城市之一,随着国家政策的推动,大批国有企业和集团公司入驻杭州,进一步推动了杭州地铁产业的发展。然而,当前杭州地铁相关行业的发展迫切需要优化与地铁其他交通运输工具的接驳。因此,对未

基金项目: 吉林省教育厅科学技术研究项目(JJKH20230306CY);2025 年吉林省高等教育教学改革研究课题(JLJY202579322416)。
作者简介: 邢 雪(1983—),女,博士,教授。主要研究方向:基于人工智能的交通数据挖掘,智能交通关键技术; 罗发辉(2002—),男,硕士。主要研究方向:人工智能; 路 义(2002—),男,硕士。主要研究方向:交通建筑设计。
通信作者: 路 康(2002—),男,硕士。主要研究方向:人工智能; Email:3196304696@ qq. com。
收稿日期: 2024-12-18

来杭州客流量的预测显得尤为重要，并且对提升城市交通管理水平和优化路径推荐具有重要意义。

随着信息技术与人工智能的不断发展，越来越多的研究致力于探索新的预测方法，当前对客流量的预测主要是基于深度学习的方法，最常用的有时间序列预测模型、时空建模方法、混合模型等。为解决传统循环神经网络（RNN）难以捕捉长期依赖的问题，Hochreiter 等[1] 于 1997 年创建了 LSTM（Long Short-Term Memory）模型。2000 年，Gers 等[2] 对原始 LSTM 进行改进，提出了遗忘门（Forget Gate）结构，增加了门控机制与记忆单元状态之间的直接连接，提高了 LSTM 对短期和长期依赖关系的建模能力。2013 年，Graves 等[3] 将 LSTM 应用于语音识别任务，并展示了其在处理长时间序列数据时的优势。2014 年，Cho 等[4] 提出门控循环单元（Gated Recurrent Unit，GRU）作为 LSTM 的一种简化变体，通过合并输入门和遗忘门减少参数数量。同年，Bahdanau 等[5] 将注意力机制引入 LSTM 框架，提升了模型对长序列中重要特征的捕捉能力。2017 年，Vaswani 等[6] 提出 Transformer 模型，其核心是多头注意力机制，完全摒弃了循环结构，实现了更高效的并行计算。

随着对预测精度要求的不断提高，多数学者通过构建组合模型来提高模型的预测精度，最常见的是引入智能优化算法来优化模型的参数，例如 Dorigo 等[7] 于 1992 年提出了蚁群算法（Ant Colony Optimization，ACO），该算法受蚂蚁觅食行为启发，通过模拟蚂蚁群体中基于信息素（pheromone）浓度的路径选择机制，来解决复杂的组合优化问题。1995 年，Kennedy 等[8] 引入粒子群优化算法（Particle Swarm Optimization，PSO），用以解决遗传算法在某些问题上参数调整复杂、计算开销大的缺陷。2020 年，Xue 等[9] 引入麻雀搜索算法（Sparrow Search Algorithm，SSA）解决传统优化算法中常见的早熟收敛和局部最优问题。基于此，本研究将灰狼优化算法（Grey Wolf Optimization，GWO）[10] 引入建模流程，对 LSTM 模型的超参数（如学习率、神经元个数和迭代次数）进行优化，同时，考虑到天气、气温等多重变量因素，采用多头注意力机制对这些特征进行综合预测，从而提升模型的预测精度与泛化能力。

1　GWO-LSTM-MHA 地铁客流预测模型

1.1　预测模型

LSTM 是一种面向时序序列数据设计的深度学习模型，通过引入输入门、遗忘门和输出门来解决传统循环神经网络（RNN）的梯度消失问题，从而更好地保留长距离依赖信息[11]。在预测客流量时，LSTM 能有效捕捉长短期信息，适应多维度和非线性关系，提高预测准确性[12]。

LSTM 结构由输入门、遗忘门、输出门和记忆细胞组成，其内部结构如图 1 所示。

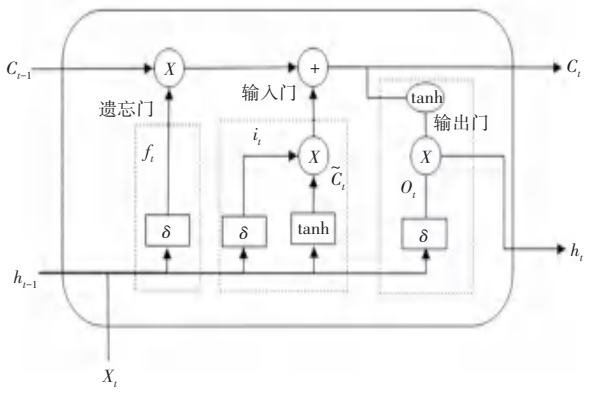


图 1　LSTM 单元内部结构图

Fig. 1　Internal structure of LSTM unit

遗忘门决定细胞状态中需要遗忘哪些信息，即

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_f \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + b_f) \tag{1}$$

式中：$\boldsymbol{f}_t$ 表示遗忘门的输出，值在[0,1]之间，控制遗忘比例；σ 是 Sigmoid 激活函数；$\boldsymbol{h}_t$ 表示当前时刻的隐藏状态；$\boldsymbol{h}_{t-1}$ 表示上一时刻的隐藏状态；$\boldsymbol{x}_t$ 表示当前时刻的输入；$\boldsymbol{W}_f$ 表示遗忘门的权重矩阵；b_f 表示遗忘门的偏置项。

输入门决定细胞状态中需要写入哪些信息，更新候选记忆内容和细胞状态，即

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_i \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + b_i) \tag{2}$$

$$\widetilde{\boldsymbol{C}}_t = \tanh(\boldsymbol{W}_C \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + b_c) \tag{3}$$

$$\boldsymbol{C}_t = \boldsymbol{f}_t \odot \boldsymbol{C}_t - 1 + i_t \odot \widetilde{\boldsymbol{C}}_t \tag{4}$$

式中：$\boldsymbol{i}_t$ 表示输入门的输出，值在[0,1]之间，控制写入比例；$\boldsymbol{W}_t$、b_t 表示输入门的权重矩阵和偏置项；$\widetilde{\boldsymbol{C}}_t$ 表示候选记忆内容；$\boldsymbol{W}_c$、b_c 表示候选记忆的权重矩阵和偏置项；$\boldsymbol{C}_t$ 表示更新后的细胞状态。

输出门决定当前时刻的隐藏状态输出值，即

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_o \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + b_o) \tag{5}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \odot \tanh(\boldsymbol{C}_t) \tag{6}$$

式中：$\boldsymbol{o}_t$ 是输出门的输出，值在[0,1]之间，控制输出比例；$\boldsymbol{W}_o$ 和 b_o 代表输出门的权重矩阵和偏置项；$\boldsymbol{h}_t$ 表示当前时刻的隐藏状态[13]。

1.2　灰狼优化算法

传统手动调整超参数的方法效率低下，通常仅能获得局部较优参数组合，难以搜寻到最优配置，会

限制 LSTM 模型的预测性能上线。灰狼优化算法(GWO)是一种基于元启发式的优化算法,凭借其收敛精度高、全局搜索能力强、收敛速度快的优势,可对 LSTM 模型的超参数进行自动寻优,充分挖掘模型的预测潜力[14-15]。该算法的设计灵感来源于对灰狼群体协作捕猎行为的模拟[16]。

1)社会中的层级构成。在灰狼群体中,等级被细分为 α,β,δ,ω 四个等级,α 个体对应寻优问题的全局最优解,β,δ 个体分别对应次优解、第三优解,其余所有候选解的个体均为 ω 个体。在 α 的指导下,捕食活动得以顺利完成。在 α 为正整数且满足一定条件下,α 与 β 成线性关系。在狼群的 α 值出现缺失的情况下,β 会取代 α 的位置。当 α 和 β 的适应性降低时,其数值也将减少至 δ。

2) 围攻猎物。通过动态调整灰狼的位置,逐步缩小搜索范围。将灰狼在狩猎过程中围捕猎物的行为定义为:

$$\boldsymbol{D} = |\boldsymbol{C} \cdot \boldsymbol{X}_{\text{prey}} - \boldsymbol{X}| \tag{7}$$

$$\boldsymbol{X}(t+1) = \boldsymbol{X}_{\text{prey}} - \boldsymbol{A} \cdot \boldsymbol{D} \tag{8}$$

$$\boldsymbol{A} = 2\boldsymbol{a} \cdot \boldsymbol{r}_1 - \boldsymbol{a} \tag{9}$$

$$\boldsymbol{C} = 2 \cdot \boldsymbol{r}_2 \tag{10}$$

式中:$\boldsymbol{X}_{\text{prey}}$ 为猎物位置(目前最优解);$\boldsymbol{X}$ 为灰狼个体位置;$\boldsymbol{A}$、$\boldsymbol{C}$ 为动态调整系数,分别控制搜索和步长的方向;$\boldsymbol{a}$ 从 2 逐步减小到 0,控制算法的探索与开发;$\boldsymbol{r}_1$、$\boldsymbol{r}_2$ 为随机数,值在[0,1]之间[17]。

3) 狩猎行为。灰狼通过评估距离选择目标,并在 α 的引导下,由 β 和 δ 带领狼群包围猎物。该算法将灰狼群体视为粒子群模型,提出基于种群动态演化的求解方法。假设 α、β 和 δ 能更好地揭示猎物位置,保存 3 个最优解,使其他灰狼根据这些解更新位置,逐步接近猎物。数学模型如下:

$$\boldsymbol{X}_1 = |\ C_1 \cdot \boldsymbol{X}_\alpha - \boldsymbol{X}\ | \tag{11}$$

$$\boldsymbol{X}_2 = |\ C_2 \cdot \boldsymbol{X}_\beta - \boldsymbol{X}\ | \tag{12}$$

$$\boldsymbol{X}_3 = |\ C_3 \cdot \boldsymbol{X}_\delta - \boldsymbol{X}\ | \tag{13}$$

式中:$\boldsymbol{X}_\alpha$、$\boldsymbol{X}_\beta$、$\boldsymbol{X}_\delta$ 分别表示 α、β 和 δ 的个体位置,$\boldsymbol{X}_1$,$\boldsymbol{X}_2$,$\boldsymbol{X}_3$ 为灰狼个体与 α、β 和 δ 的距离。

4) 灰狼个体根据 α、β 和 δ 的引导更新其位置:

$$\boldsymbol{X}_1 = \boldsymbol{X}_\alpha - \boldsymbol{A}_1\boldsymbol{D}_\alpha \tag{14}$$

$$\boldsymbol{X}_2 = \boldsymbol{X}_\beta - \boldsymbol{A}_2 \cdot \boldsymbol{D}_\beta \tag{15}$$

$$\boldsymbol{X}_3 = \boldsymbol{X}_\delta - \boldsymbol{A}_3 \cdot \boldsymbol{D}_\delta \tag{16}$$

最终更新群体中的每个灰狼位置为

$$\boldsymbol{X}(t+1) = \frac{\boldsymbol{X}_1 + \boldsymbol{X}_2 + \boldsymbol{X}_3}{3} \tag{17}$$

1.3 多头注意力机制

多头注意力机制(Multi-Head Attention MHA)[18]是一种用于捕获序列数据中全局依赖关系的核心方法,最早在 Transformer 模型中提出,广泛应用于自然语言处理、时间序列分析等领域。其主要思想是通过多个独立的注意力头,捕获输入序列中不同部分之间的关系,从而增强模型对全局上下文的感知能力。

自注意力机制的基础是通过查询(Query)、键(Key)和值(Value)三个矩阵计算输入序列的权重分布。每个注意力头分别计算输入的加权和,以突出重要信息的贡献。通过多头机制(Multi-Head),模型可以从不同的子空间视角并行学习序列特征,获得更丰富的表达能力[19]。

自注意力机制公式:

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{18}$$

式中:$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别是查询、键和值矩阵,d_k 是键向量的维度,缩放因子 $\sqrt{d_k}$ 用于防止 $\boldsymbol{Q}\boldsymbol{K}^{\text{T}}$ 内积值过大,导致 Softmax 梯度消失。

多头机制公式为:

$$\text{MultiHead}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{Concat}(\text{Attention}_1,\cdots,\text{Attention}_H)\boldsymbol{W}_o \tag{19}$$

式中:$\boldsymbol{W}_o$ 是线性变换矩阵,Concat 表示拼接操作。

MHA 会对同一个输入序列并行执行多个注意力计算,再将各头的输出拼接后做线性变换,生成最终注意力特征。相比单一注意力机制,MHA 能从不同特征子空间捕获更多的信息模式,大幅提升对长时间依赖关系的建模能力,尤其适配时间序列、多变量数据等场景,多头注意力机制如图 2 所示。

1.4 GWO-LSTM-MHA 预测模型的构建

为提升地铁站点客流量预测的准确性,本研究构建了 GWO-LSTM-多头注意力机制(MHA)多变量预测模型。由于地铁站点客流量受多种因素影响,且包含大量非线性和动态变化数据,因此需要一种高效的预测方法。LSTM 模型因其在处理多变量输入和非线性数据方面的显著优势成为首选[20],而多头注意力机制(MHA)通过捕获输入序列中各时间步之间的全局依赖关系,进一步提升了对复杂动态关系的建模能力[21]。同时,模型性能在很大程度上依赖于超参数的合理设置。为此,引入灰狼优化算法(GWO),其具备全局收敛与强大的局部搜索能力,适合解决非线性函数的超参数寻优问题[22]。

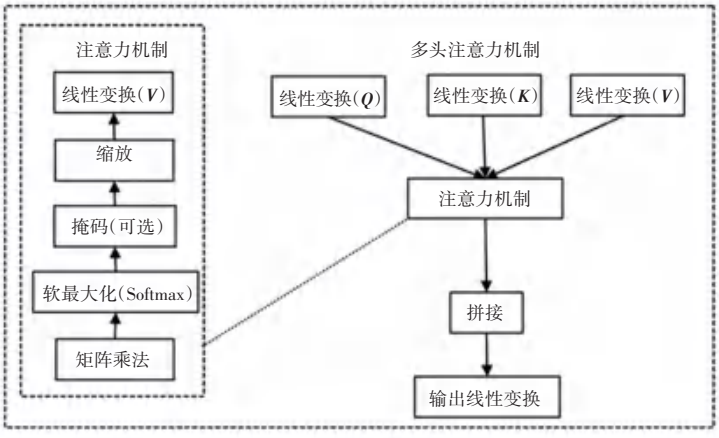


图 2　多头注意力机制

Fig. 2　Multi-Head Attention Mechanism

模型结构由输入层、隐含层和输出层构成。其中，输入层包含影响地铁站点客流量的多源数据与历史客流量，隐含层包括多头注意力模块与多个 LSTM 单元，输出层用于预测未来的地铁站点客流量。在 GWO-LSTM-MHA 模型中，MHA 通过自注意力机制捕获输入特征的全局相关性，增强了对长时间序列依赖关系的建模能力。与此同时，GWO 用于优化 LSTM 的关键超参数（如学习率、迭代次数、隐含层节点数），并以预测值与实际值的均方误差（MSE）作为适应度函数。通过动态调整学习率与优化超参数，模型避免陷入局部最优解，提高了训练过程的稳定性与预测准确性。

在模型建立过程中，通过对输入数据进行特征工程和相关性分析，初始化 LSTM 网络的结构与参数，并结合多头注意力机制进行全局信息建模。在训练过程中，神经元通过权重调整学习输入数据的内在属性，MHA 层对特征进行加权分配，提取重要的时间序列关系。隐含层节点数等超参数决定模型的容量及可学习函数的复杂性[23]。为避免过度拟合，本研究设计了紧凑的网络结构，并采用适当的学习率策略提升训练效率和预测性能。模型具体步骤与流程如图 3 所示。

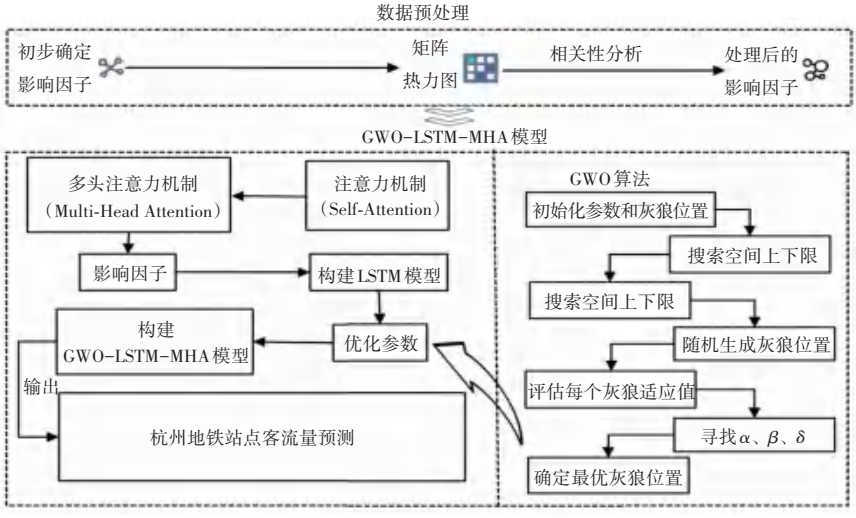


图 3　模型预测流程

Fig. 3　Model prediction process

2　案例分析

2.1　数据准备

本研究采用杭州某地铁站点 2019 年 1 月 1 日至 25 日的客流数据，包含时间、风向、气温等影响因素，共获取闸机刷卡有效信息 452 832 条。数据以小时为单位进行训练、预测，选取 80% 的地铁客流数据作为训练集样本，20% 为测试集，在训练集中按比例抽取 10% 作为验证集，用于验证提出的预测模型的预测精度。

2.2 影响因素选取和相关性分析

选取客流量(flow)、工作日(workday)、非工作日(nonworkdays)、最高气温(Maximum_temperature)、最低气温(Minimum_temperature)、风向(wind_direction)、风力(wind_power)、天气(weather)作为分析变量,绘制相关性矩阵热力图(见图 4),分析结果如下:

1)工作日与客流量呈正相关(相关系数 0.8),工作日人们因通勤、活动频繁导致流量增加;非工作日和客流量的相关性为负(相关系数-0.8),人们出行减少,流量随之降低,非工作日与流量呈反向关系。

2)最高温度对客流量的直接影响较小,最低温度对客流量有微弱影响,但某些极端温度(例如极端低温)可能会间接影响客流量。

3)天气与客流量相关性为 0.11,说明天气状况会影响流量,较好的天气条件通常会伴随较高的流量。例如晴天时,出行活动增多,而恶劣天气(如风、雨)可能会降低流量。

4)风力与客流量存在一定相关性,可能影响出行模式,风向与客流量的相关系数接近 0,表示风向对客流量无直接影响,并未直接影响到人们的出行或者活动安排。

本研究将上述影响因素融入时间序列分析中,探索历史客流变化特征。在 LSTM 模型中加入多头注意力机制,以便更有效地学习特征,并重点分析下一时间段的客流特征,从而实现更精确的客流量预测。

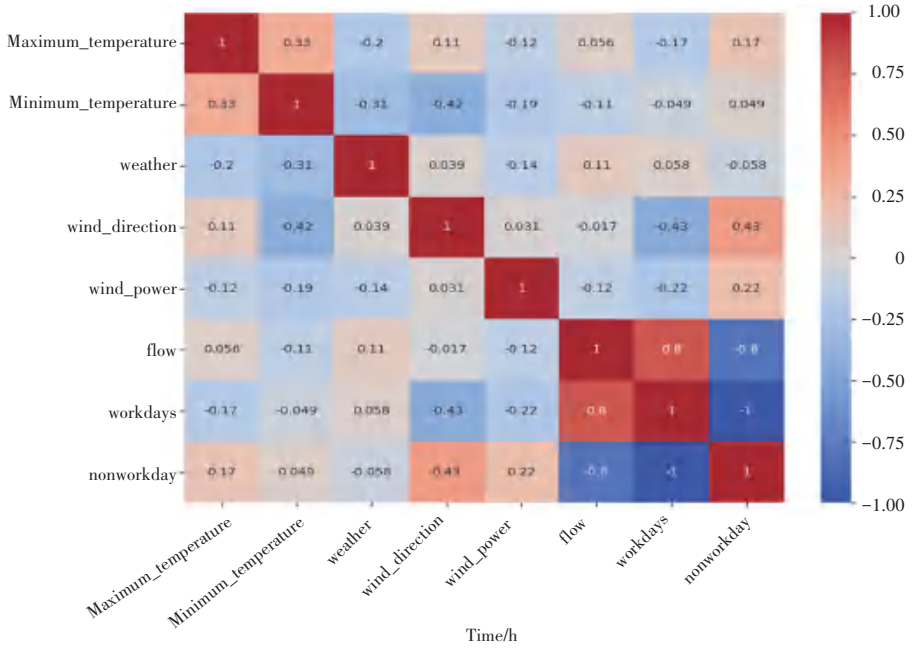


图 4 影响因素矩阵热力图

Fig. 4 Influencing factor matrix heat map

3 实验结果与分析

3.1 模型参数寻优

本实验结合 Adam 优化算法与灰狼优化算法(GWO)开展神经网络的训练与优化。Adam 优化方法用于隐藏层参数的梯度更新,通过结合动量法和 RMSProp 算法,动态调整每个参数的学习率,以适应不同维度的梯度变化,从而提升参数更新的效率[24]。同时,为进一步提升模型性能,采用灰狼优化算法对神经网络的超参数(包括学习率、批次大小、隐藏层神经元个数)进行优化,寻找最优的参数组合。超参数优化算法的设置如下:

灰狼算法(GWO):设置狼的种群数量为 5、迭代次数 300 次。为提高搜索效率和准确性,在对参数空间进行初步估计后,设隐藏层神经元个数的搜索范围为[16, 128],学习率的搜索范围为[0.001, 0.09],以及批量大小的搜索范围为[10, 56]。

3.2 最优参数组合确定

本研究以个体最优适应度的变化反映算法的收敛状态与训练过程[25]。使用灰狼算法优化参数,通过计算灰狼个体适应度,选出 3 匹适应度最优的狼。在每次迭代中,更新最优个体适应度和位置并重新计算适应度,直至达到最大迭代次数,最终获得最优参数[26]。

GWO-LSTM-MHA 预测模型在迭代寻优过程中表现稳定,个体适应度值较低,最优灰狼适应度值的

变化趋势如图 5 所示，Iteration 表示迭代次数，Fitness Value 表示个体适应度值。最优超参数组合见表 1。

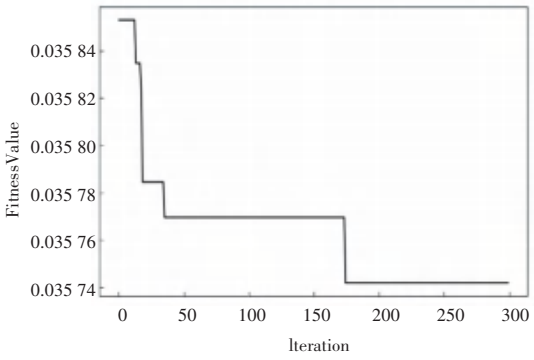


图 5　最优灰狼适应度值变化

Fig. 5　Optimal Grey Wolf fitness value variation

表 1　最优超参数组合

Table 1　Optimal combination hyperparameters

参数	GWO
第一层神经元	119.000
第二层神经元	32.000
学习率(Droput)	0.061
批次大小	19.000

3.3　预测结果对比验证

为更好地表征 GWO-LSTM-MHA 模型在杭州地铁站点客流预测中的预测效果，除了对比客流数据实际值与 GWO-LSTM-MHA 模型预测值的差异之外，本实验引入 BP、LSTM、SSA-LSTM、PSO-LSTM 和 GWO-LSTM 五种模型与 GWO-LSTM-MHA 模型进行横向对比，在使用相同测试集的基础上比较不同类型算法下的客流预测结果与性能。

为评估模型预测结果的准确程度，本实验采用实际值相对均方根误差、平均绝对误差与模型拟合优度对预测结果进行评估。

1）预测结果对比。如图 6 所示，BP 模型在高峰时段预测偏差较大，而 LSTM 和 PSO-LSTM 模型对流量趋势变化时适应性较差，表现出不稳定性，SSA-LSTM、GWO-LSTM 模型的预测精度在低谷和高峰时段仍存在预测误差，GWO-LSTM-MHA 模型预测误差更小、精度更高，在客流预测任务中综合性能最优。

2）模型性能评价。本研究通过比较 6 种模型的 RMSE、MAE 与 R^2 指标，来评估预测的准确性，各模型评价指标结果见表 2。数据显示：GWO-LSTM 模型与 SSA-LSTM 和 PSO-LSTM 模型相比 MAE 和 RMSE 差值较小，但拟合度较低，因此，本实验在 GWO-LSTM 模型基础上加入多头注意力机制，以提升模型拟合度。

由表 2 数据对比可得：GWO-LSTM-MHA 模型在各项指标上相较于其他模型有显著的优化。在 R^2 值上，GWO-LSTM-MHA 达到了 0.802，优于 GWO-LSTM 的 0.739，拟合优度更强；同时，RMSE 和 MAE 值分别为 206.592 和 137.884，明显低于 BP、LSTM、SSA-LSTM、PSO-LSTM 和 GWO-LSTM 模型，这些结果充分证明了 GWO-LSTM-MHA 模型在提升预测精度和模型稳定性方面的优势。

表 2　模型预测误差计算

Table 2　Calculation of model prediction error

优化模型	MAE	RMSE	R^2
BP	245.191	355.527	0.714
LSTM	287.252	427.537	0.721
SSA-LSTM	196.661	298.734	0.798
PSO-LSTM	207.143	314.014	0.777
GWO-LSTM	211.254	297.665	0.739
GWO-LSTM-MHA	137.884	206.592	0.802

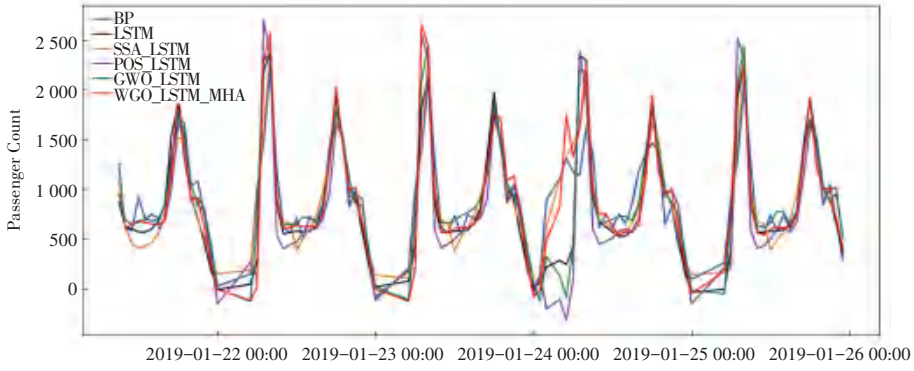


图 6　多种模型预测结果对比

Fig. 6　Comparison of prediction results across various models

4 结束语

本研究提出一种融合灰狼优化算法(GWO)、长短时记忆网络(LSTM)和多头注意力机制(MHA)的预测方法,用于精准预测地铁站点客流量,主要结论如下:

(1)通过引入灰狼优化算法对 LSTM 模型的超参数(如学习率、隐藏层节点数、迭代次数)进行优化,有效提升了模型的预测精度。与其余模型相比,GWO-LSTM-MHA 模型表现出更小的预测误差,其 MAE 和 RMSE 指标均显著降低。

(2)GWO-LSTM-MHA 模型为多变量输入,预测时充分考虑了时间、天气、气温等多种因素的影响,同时通过多头注意力机制捕获输入序列中不同时间步之间的全局依赖关系,进一步提升了预测的准确性。研究结果表明,该模型对历史客流数据和外部环境因素的拟合效果优于传统方法,综合多元因素建模展现出更强的泛化能力。

(3)在实例分析中,模型在完成训练和验证后进一步对客流量进行了预测,研究结果显示,GWO-LSTM-MHA 模型能够对历史数据进行有效拟合。为未来客流量变化提供了科学的预测依据,为地铁站点运营管理、资源调配以及路径优化提供了更为精确的理论支撑和实践指导。

(4)未来研究中,可进一步引入空气湿度、地理位置、特殊事件等因素进行综合建模,以进一步提升模型的预测精准度和实用性,同时探索模型在更复杂场景中的应用潜力。

本研究创新性地将 GWO 优化、LSTM 模型与多头注意力机制结合,提出了一种面向地铁站点客流量预测的高效方法,为复杂多变量时间序列预测提供了新的解决方案。

参考文献

[1] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[2] Gers F A, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM[J]. Neural Computation, 2000, 12(10): 2451-2471.

[3] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks[C]//Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE, 2013: 6645-6649.

[4] Cho K. On the properties of neural machine translation: Encoder-decoder approaches[J]. arXiv preprint arXiv, 1409.1259, 2014.

[5] Bahdanau D. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv, 1409.0473, 2014.

[6] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. arXiv preprint arXiv, 1706.03762, 2017. DOI: 10: S0140525X16001837.

[7] Dorigo M, Manizeeoa V, Colorni A. Ant system: Optimization by a colony of cooperating agents[J]. IEEE Transactions on Systems, Man, and Cybernetics: Part B (cybernetics), 1996, 26(1): 29-41.

[8] Kennedy J, Eberhart R. Particle swarm optimization[C]// Proceedings of ICNN′95-International Conference on Neural Networks. Piscataway, NJ: IEEE, 1995: 1942-1948.

[9] Xue J P, Shen B. A novel swarm intelligence optimization approach: Sparrow search algorithm[J]. Systems Science & Control Engineering, 2020, 8(1): 22-34.

[10] Mirjalili S, Mirjalili S M, Lewis A. Grey wolf optimizer[J]. Advances in Engineering Software, 2014, 69: 46-61.

[11] 姜垚先, 蒋硕, 李键, 等. 基于 MSW-LSTM 的卫星功率负荷预测方法[J]. 航天器工程, 2024, 33(5): 43-49.

[12] 侯慧, 吴文杰, 魏瑞增, 等. 基于注意力机制的 CNN-LSTM-XGBoost 台风暴雨电力气象混合预测模型[J]. 智慧电力, 2024, 52(10): 96-102. DOI: 10.20204/j.sp.2024.10013.

[13] 张舒. 基于 NWD 集成算法的多粒度微博用户兴趣画像研究[D]. 广州: 广东工业大学, 2020. DOI: 10.27029/d.cnki.ggdgu.2020.001483.

[14] 张晓凤, 王秀英. 灰狼优化算法研究综述[J]. 计算机科学, 2019, 46(3): 30-38.

[15] 蒋恣源, 罗敏, 刘翰林, 等. 基于 PCC-GWO-SVM 算法的刀具磨损预测[J]. 工具技术, 2024, 58(11): 131-138.

[16] Ghalambaz M, Yengejeh R J, Davami A H. Building energy optimization using grey wolf optimizer[J]. Case Studies in Thermal Engineering, 2021, 27: 101250.

[17] 谭方青, 张天骐, 黄烈超, 等. 基于 QPSO 的两通道正交镜像滤波器组的优化设计[J]. 计算机应用研究, 2011, 28(9): 3432-3435.

[18] 孙溪晨, 李伟兵, 黄昌伟, 等. 基于自注意力机制增强的 CNN-LSTM 的榴弹轨迹多步超前预测[J]. 兵工学报, 2024, 45(S1): 51-59.

[19] 高凯, 李勋豪, 胡林, 等. 基于多头注意力的 CNN-LSTM 的换道意图预测[J]. 机械工程学报, 2022, 58(22): 369-378.

[20] 刘聪. 改进的灰狼优化算法及其应用研究[D]. 哈尔滨: 哈尔滨师范大学, 2023. DOI: 10.27064/d.cnki.ghasu.2023.001018.

[21] 仵晓聪, 冯鑫, 蒋豪. 基于多头注意力 CNN-LSTM 碳排放量预测研究[J]. 重庆工商大学学报(自然科学版), 2026, 43(3): 116-124. DOI: 10.16055/j.issn.1672-058X.2026.0003.014.

[22] 李帅辰, 武建锋. 基于智能优化算法及其优化 BP 神经网络的室内定位[J]. 科学技术与工程, 2024, 24(20): 8568-8576.

[23] 杨玉好, 杨斌, 胡军, 等. 基于 MHA-BiLSTM 的尾矿坝位移预测[J]. 有色金属工程, 2024, 14(10): 147-157.

[24] 吴青云, 高景辉, 李昭, 等. 基于 Adam 优化卷积神经网络的火电厂大数据平台预警模型测试与应用[J]. 科学技术与工程, 2023, 23(35): 15075-15083.

[25] 翁剑成, 陈旭蕊, 潘晓芳, 等. 基于超参数优化 WOA-Bi-LSTM 模型的客运枢纽抵站客流预测方法[J]. 交通信息与安全, 2023, 41(5): 148-157.

[26] 崔彤. 基于物联网的图像异常检测及能耗优化研究[D]. 天津: 河北工业大学, 2020. DOI: 10.27105/d.cnki.ghbgu.2020.000754.