

基于深度神经网络的短时地铁客流预测

李安娜¹, 李永福²

¹贵州大学数学与统计学院, 贵州 贵阳

²中铁八局集团昆明铁路建设有限公司, 云南 昆明

收稿日期: 2021年12月21日; 录用日期: 2022年1月21日; 发布日期: 2022年1月29日

摘要

为更好推动城市轨道交通智慧化进程, 利用复杂的神经网络对城市轨道交通客流进行短时精准预测, 为客运交通运输规划提供科学的依据。针对传统的基于梯度下降法训练的长短期记忆网络交通流预测算法受网络初始值影响较大这一缺陷, 提出了一种基于遗传算法改进的长短期记忆网络(LSTM)对郑州地铁一号线客流量进行短时预测。首先, 搭建基于交通流预测的LSTM网络结构; 然后, 根据遗传算法通过选择、交叉、变异三种遗传操作不断迭代搜索得到LSTM最优的隐含层和全连接层层数以及对应的神经元个数; 最后, 利用训练好的LSTM神经网络模型对测试数据集进行预测。同时利用卷积神经网络从时间和空间两个维度进行预测, 提取更高维度的站间客流特征, 充分考虑了不同站之间客流量数据的关联性。实验结果表明, 神经网络模型在短时地铁客流预测中具有较高的预测精度。

关键词

短时地铁客流预测, 遗传算法, 长短期记忆网络, 卷积神经网络, 深度学习

Short-Term Subway Passenger Flow Prediction Based on Deep Neural Network

Anna Li¹, Yongfu Li²

¹School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

²Kunming Railway Construction Co., Ltd., China Railway Eighth Bureau Group, Kunming Yunnan

Received: Dec. 21st, 2021; accepted: Jan. 21st, 2022; published: Jan. 29th, 2022

Abstract

In order to better promote the intelligentization of urban rail, a complex deep neural network is used to make short-term accurate predictions of urban rail transit passenger flow, and provide a

scientific basis for passenger transportation planning. Aiming at the defect that the traditional Long Short-Term Memory network traffic flow prediction algorithm based on gradient descent training is greatly affected by the initial network value, a genetic algorithm-improved Long Short-Term Memory network (LSTM) is proposed for Zhengzhou Metro Line 1 Short-term forecast of passenger flow. First, building an LSTM network structure based on traffic flow prediction; then, according to the genetic algorithm, iteratively search through the three genetic operations of selection, crossover, and mutation to obtain the optimal number of hidden and fully connected layers of the LSTM and the corresponding number of neurons; finally, use the trained LSTM neural network model to make predictions on the test data set. At the same time, Convolutional Neural Network is used to predict from the two dimensions of time and space, and the higher-dimensional characteristics of passenger flow between stations are extracted, which fully considers the correlation of passenger flow data between different stations. The experimental results show that the deep neural network model has high prediction accuracy in short-term subway passenger flow prediction.

Keywords

Short-Term Subway Passenger Flow Prediction, Genetic Algorithm, Long Short-Term Memory, Convolutional Neural Network, Deep Learning

Copyright © 2022 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

短期客流预测是根据历史客流数据, 预测下一较短时刻的客流模型。短期客流预测可用于优化线路和客流组织, 为运营商提供预警和多方决策, 进一步提高城市的科学管理水平, 对于做好列车运营排班、提高乘客的出行效率, 减少轨道交通的压力都具有重要意义[1]。

地铁客流量的研究历程主要包括三个阶段: 统计学方法、传统机器学习方法和深度学习方法阶段。统计学方法主要包括 MA 移动平均模型, ARIMA 自回归差分移动平均模型、逻辑回归模型和卡尔曼滤波等。ARMA 模型[2]将移动平均和线性差分结合起来, 用来预测具有线性平稳性质的时间序列数据。1996年, 袁振洲[3]在预测货源货流发展趋势的基础上, 提出了自回归积分移动平均(ARIMA(p, d, q))模型。此外, 研究人员引入了传统机器学习方法来研究短时交通流预测。2008年, 姚[4]等根据主成分数据序列训练支持向量机从而预测道路网短时交通流量取得较好效果。此外, 也有不少学者在预测交通客流量时采用贝叶斯方法[5]、随机森林方法[6]等都取得了较好的效果。

随后深度学习也逐渐应用于地铁客流预测中。Zhao 等[7]利用长短期记忆网络对交通流做预测。李等[8]提出基于残差长短期记忆网络(LSTM)的交通流量预测方法, 显式建模特征维度之间的相互依赖关系。随着 LSTM 的发展, LSTM 演变出来了很多的变体, 其中一种为门控循环单元(GRU)神经网络, GRU 相比于 LSTM 少了一个门结构, 它的训练速度会更快, 但是它的训练能力有所下降。Fu [9]等构建了 LSTM 模型和 GRU 模型, 利用这两个模型单独预测了短时交通流量。Du [10]等提出了一种深度不规则卷积残差 LSTM 神经网络对城市交通客流进行预测, 具有很高的预测精度。

近年来, 基于标准分类器的方法将特征提取和预测两个过程分开进行, 然而特征提取阶段通常会采取人工设计方式。人工提取特征具有主观依赖性且存在耗时、费力的缺陷。与传统的机器学习方法相比, 深度学习方法具备自动提取特征和预测的优势, 可以无需显式的特征提取和选择过程而从海量数据中自

动挖掘深层次的高级特征, 具有强大的表征地铁客流数据特征的能力。但是深度学习的特点是含有多层隐藏层, 网络结构层次较深, 含有大量需要训练的参数, 想要将模型调整到一个最佳状态并非易事。在此基础上, 本文依托深度学习网络利用了遗传算法和 LSTM 神经网络对参数组合快速寻优的特性, 在快速寻参的同时, 可以得到更高的预测精度, 表现出更好的预测性能。并且本文从时间和空间两个维度对郑州地铁客流量进行预测, 使用卷积神经网络作为对照实验, 提取更高维度的站间客流特征, 利用不同站台间客流量数据的关联性, 分析下一时刻的客流情况。

2. 研究基础

2.1. LSTM 模型

传统的 RNN 模型在训练时会遇到梯度消失或者梯度爆炸的问题, LSTM 是循环神经网络的变体, 既能够在保持循环神经网络处理序列数据的能力, 也避免梯度消失和梯度爆炸现象。LSTM 通过门控机制, 达到学习时序数据的能力, 其结构见图 1 所示。

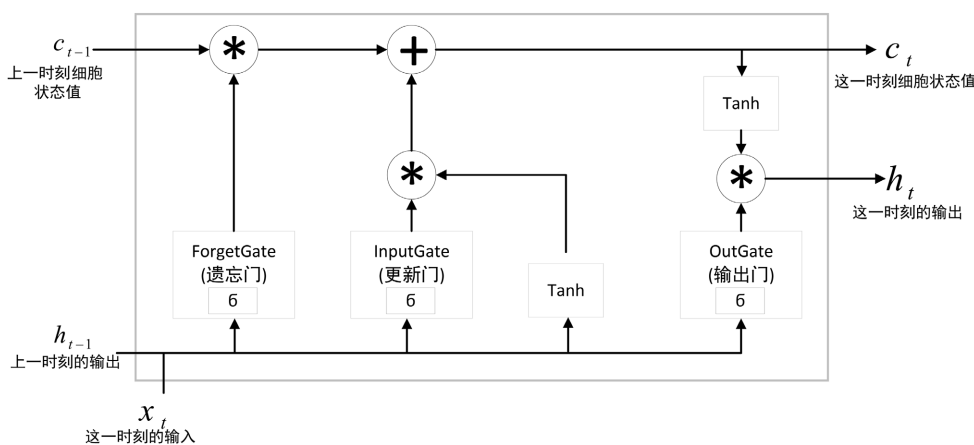


Figure 1. Model structure of LSTM
图 1. LSTM 模型结构图

LSTM 拥有三个门来保护和控制细胞状态, 包括更新门(Input Gate)、遗忘门(Forget Gate)和输出门(Out Gate)。其中遗忘门决定应丢弃或保留信息, 得到过滤后的历史信息。遗忘门的计算过程为:

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

其中 f_t 是遗忘门的输出, σ 是激活函数, 一般设置的激活函数为 sigmoid 函数。激活函数将激活值缩放到 0 和 1 之间, 0 表示“完全遗忘”, 1 表示“完全保留”; w_f 是权重矩阵, h_{t-1} 是上一时刻是输出, x_t 是当前时刻的输入, b_f 是遗忘门的偏置项。

数据经过遗忘门筛选历史信息后, 由更新门对以往信息和当前的信息进行更新。更新门包括确定更新的信息和计算备选细胞状态两个部分。新生信息的计算为:

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

备选细胞状态的计算公式为:

$$\tilde{c}_t = \tanh(W_c [h_{t-1}, x_t] + b_c) \quad (3)$$

i_t 为更新门计算的新生信息, w_i , w_c 为权重矩阵, b_i , b_c 是更新门的偏置项。此时通过计算新生信息 i_t 和 \tilde{c}_t 的乘积, 确定了添加到备选细胞里的新生信息, 同时计算遗忘门输出信息 f_t 与原细胞状态 c_{t-1} 的乘

积, 得到原细胞状态所保留的历史信息。将二者乘积相加, 得到新的更新细胞 c_t 。计算公式如下:

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \quad (4)$$

最后输出门选择信息输出。输出门的计算公式为:

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \tanh(c_t) \quad (6)$$

其中 o_t 是输出门的输出, w_o 是权重矩阵, b_o 是输出门的偏置项, h_t 是当前记忆单元的输出。

2.2. Dropout

深度神经网络的模型复杂且参数较多, 在训练过程中容易出现过拟合的现象。Srivastava 等人[11]提出 Dropout, 以防止神经网络过拟合。Dropout 方法是在训练过程中, 让某个神经元的激活值以一定的概率 p 停止工作, 从神经网络中随机丢弃神经元, 其中也包括神经元之间的链接, 这就防止了神经元之间过度的协同适应。Dropout 显著减少了深度学习中由于参数过多带来的过拟合问题, 并且相比于传统的正则化方法有了很大的改进。本文选择 $p = 0.2$ 的概率使 GA-LSTM 和 CNN 模型中部分神经元在工作中失活。

2.3. 遗传算法

遗传算法(Genetic Algorithm, GA)是一种根据适者生存、优胜劣汰等自然机制来进行搜索计算和问题求解的算法。遗传算法是一种基于全局搜索的方法, 一般用遗传算法中适应性函数(fitness function)来衡量这个解决方案的优劣。图 2 为遗传算法优化 LSTM 流程图。

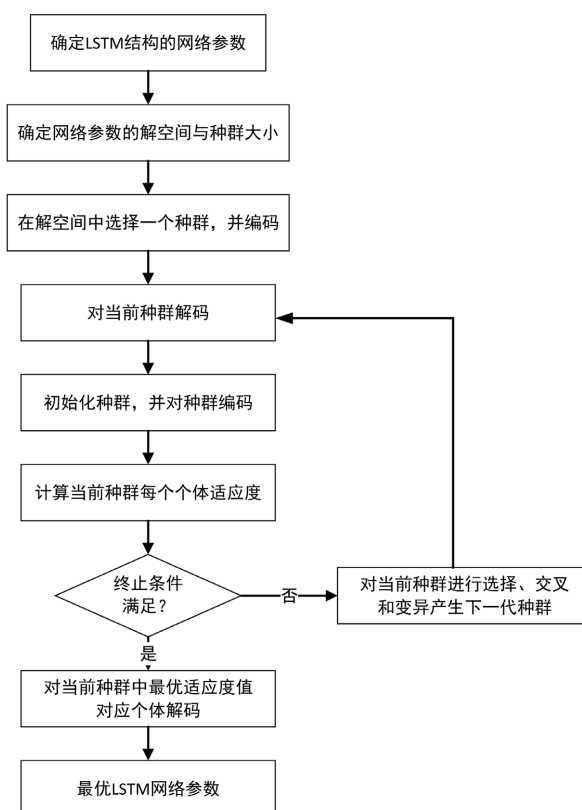


Figure 2. Genetic algorithm optimization LSTM flow chart

图 2. 遗传算法优化 LSTM 流程图

- 1) 确定优化问题的解空间, 即求解参数与不同参数的取值范围。同时, 还要确定种群的大小 N , 即每一代备选解的个数;
- 2) 在解空间中随机抽取 N 个备选解, 并对每个备选解进行编码。本文采用二进制进行编码;
- 3) 对种群中的每个个体求解适应度。适应度一般为优化问题中的目标函数或者对目标函数进行变换后的函数。
- 4) 根据当前种群的适应度判断是否终止产生下一代种群。一般判断条件包括: 算法最大迭代次数、适应度阈值、相邻两代适应度变化阈值。如果满足终止条件, 则对当前种群中最优适应度值对应的个体解码, 得到问题的最优解; 如果终止条件不满足, 根据当前种群的适应度通过选择、交叉和变异三种操作产生下一代种群。其中, “选择”为从父代选择优秀的个体直接进行遗传; “交叉”为对进行交配的两个个体的编码按照某种方式交换, 而产生两个新的个体; “交叉”指对编码后的个体的某些位的符号进行改变。

2.4. CNN 模型

卷积神经网络模型相比于 LSTM 模型, 能够提取高维的空间特征, 从而预测下一时刻的地铁客流。

卷积神经网络(CNN)是用于特征提取表达的深度学习模型, 可实现从输入到输出的功能映射。如图 3 所示, CNN 模型由输入层(Input)、卷积层(Convolution Layers)、池化层(Pooling)、全连接层(Fully Connected)和输出层(Output)组成。卷积层通过权值共享的卷积核与输入数据对应的感受野区域进行卷积计算, 从而提取输入特征。卷积的计算公式为:

$$V_j^l = \sigma \left(\sum_{s \in M_j} V_s^{l-1} * K_{sj}^l + b_j^l \right) \quad (7)$$

其中 V_s^{l-1} 用于表示输入层的第 s 个特征, 输出层为第 l 层, V_j^l 表示输出层的第 j 个特征, K_{sj}^l 表示卷积核的元素, b_j^l 为偏置项, σ 为激活函数。

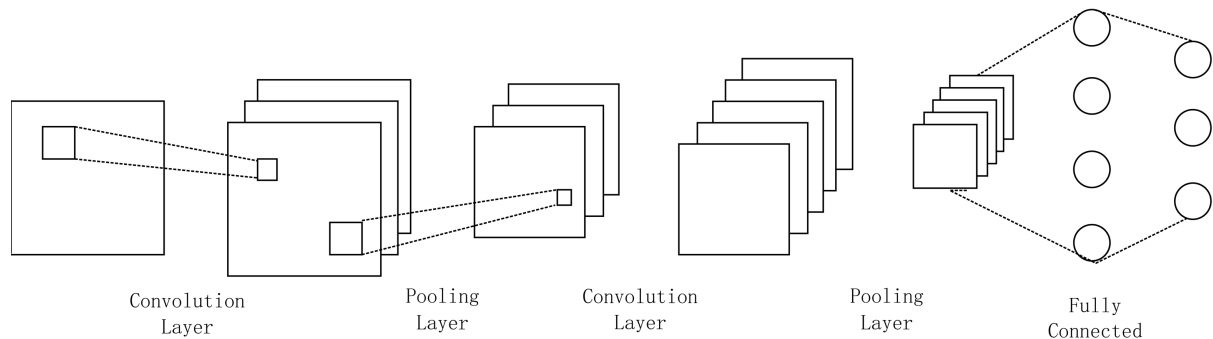


Figure 3. Convolutional neural network structure

图 3. 卷积神经网络结构图

将郑州地铁 1 号线客流量处理为以时间和不同站间客流的二维数据。此外, 郑州地铁 1 号线, 共有 20 个站, 设 CNN 模型的预测步长为 n , 则 CNN 模型的输入为一个 $20 \times n$ 的矩阵。

3. 实验及结果分析

3.1. 实验数据处理

原始的 AFC 数据是郑州市某年 9 月 1 日到 9 月 30 日地铁一号线 20 个站点全天客流量数据。根据

郑州地铁运行的实际时间, 筛选出了 05:30 至 23:30 之间的客流量数据作为模型数据集。另外在预测客流量时, 为提高预测精度, 以五分钟为间隔单位, 并对不同时段内的客流量数据进行标准化处理。此外, 同一天的不同时段, 客流量之间也存在明显差异。根据图 4、图 5 显示, 工作日客流量出现的峰值点明显多于双休日, 且双休日在同一时间点的客流也明显高于工作日。因此对于此次数据的分析, 只针对工作日进行讨论。针对工作日数据集划分: 将 9 月 29 (周二) 日和 9 月 30 (周三) 日划分为测试集, 其余工作日期间的客流量数据划分为训练集。

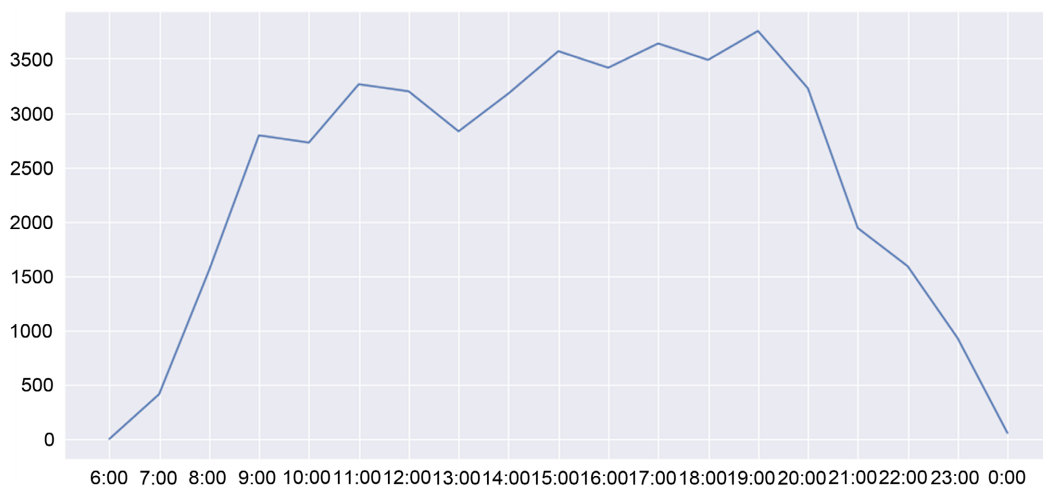


Figure 4. Weekday passenger flow of Xiliuhu Platform

图 4. 西流湖站台工作日客流量

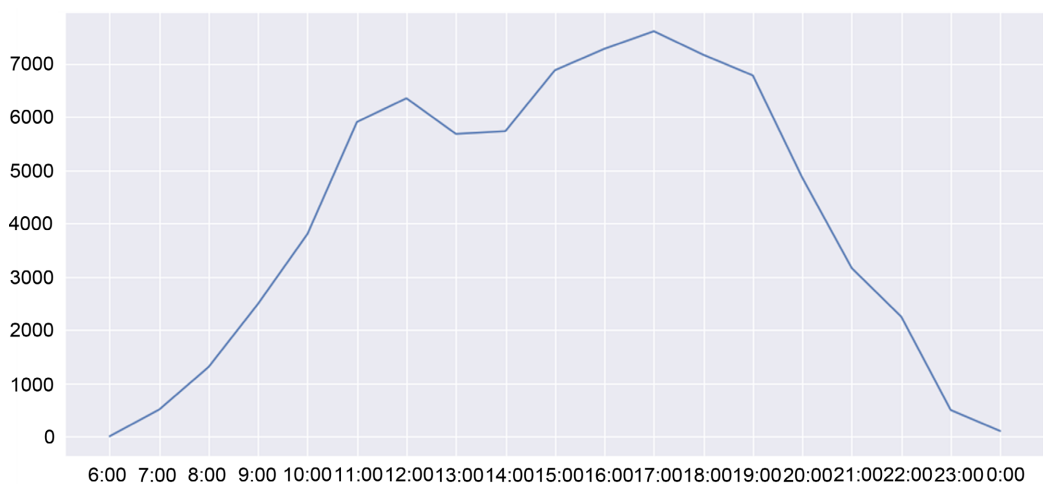


Figure 5. Weekend passenger flow of Xiliuhu Platform

图 5. 西流湖站台双休日客流量

3.2. 模型参数设置

时间序列预测包括单步预测、多步预测和滚动预测。从预测结果的效果来看, 单步预测的精度要高于多步预测和滚动预测。因此, 本文将使用单步预测的方法对未来 5 分钟内地铁客流量进行预测, 并将预测步长的初始值设置为 8。由于 3 层的网络足以拟合非线性数据, 因此本文利用遗传算法优化的 LSTM 隐藏层和 Dense (全连接层) 的层数都在 1~3 的范围内。其次, 地铁客流量预测问题属于回归问题, 因此在本次模型搭建中选择平方损失函数为评价指标。

$$\text{MSE} = \sum_N (y - \hat{y})^2 \quad (8)$$

式中, y 为真实值, \hat{y} 为预测值。

根据 2.3 中, 遗传算法首先通过选择、交叉、变异三种遗传操作不断迭代搜索得到 LSTM 最优的隐含层和全连接层层数以及对应的神经元个数; 最后, 利用训练好的 LSTM 神经网络模型对测试数据集进行预测。用遗传算法求 LSTM 模型参数的终止条件中, 设置的最大迭代次数为 20 次, 并找出适应度最大时对应的参数。由于在 LSTM 模型搭建中选择平方损失函数为评价指标, 因此在 GA-LSTM 模型中, 选择 $\frac{1}{\text{MSE}}$ 作为适应度函数, 选出适应度值最大下对应的参数值。其中, 种群数量、交叉率和变异率的初始值分别设置为 20、0.5 和 0.01。最终在 20 次迭代中, 适应度(即 $\frac{1}{\text{MSE}}$)最大值为 66.98, 此时的 LSTM 层和 dense 层的个数以及各层神经元数参数分别为[2, 1, 86, 83, 177]。表 1 详细展示了 GA-LSTM 网络训练中使用到的参数。

Table 1. Network training parameters of GA-LSTM

表 1. GA-LSTM 网络训练参数

固定参数	参数取值或对应方式
时间步长	8
预测步长	1
迭代次数	100
Dropout	0.2
损失函数	Mean-Squared-Error
激活函数	Relu
遗传算法优化参数	参数取值以及对应方式
迭代次数	20
最大适应度	66.98
模型结构	输入→LSTM 层→LSTM 层→Dropout 层→全连接层→输出
LSTM 隐藏层神经元数	86, 83
全连接层神经元数	177

与 LSTM 模型一致, 卷积神经网络模型在全连接层中加入 dropout 参数, 使全连接神经元以 20% 的概率随机失活, 利用随机梯度下降法寻找最后参数, 采用 MSE 衡量输出值与真实值之间的偏离程度, 模型迭代次数为 100 次。郑州地铁 1 号线共有 20 个站, 预测步长为 8, 因此输入的是一个 8×20 大小的矩阵。输出层的数据输出维度为 20×1 , 输出层中的 20 个神经元对应着预测时刻下 20 个不同站的地铁客流。图 6 为输入矩阵的示意图。

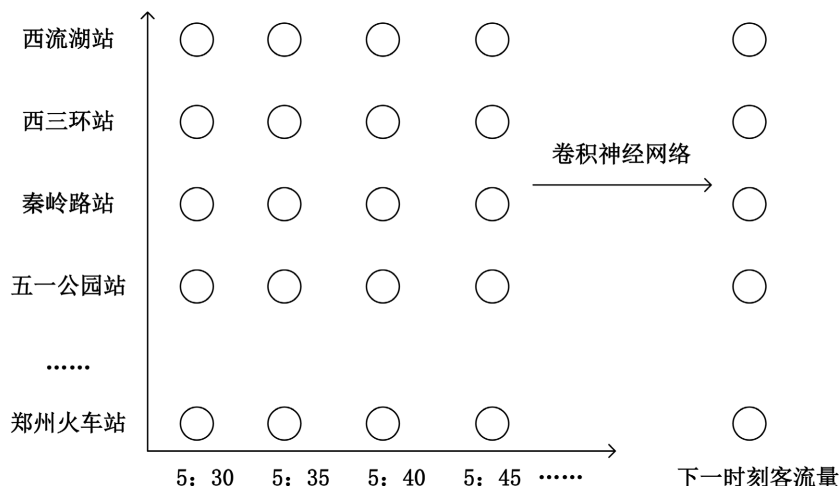


Figure 6. Two-dimensional data of subway passenger flow

图 6. 地铁客流量二维数据

3.3. 结果分析

根据 3.2 设定好参数以后, 对 GA-LSTM 网络模型和 CNN 模型进行迭代优化, 最后利用训练好的网络模型对测试数据集进行预测。图 7 表示 GA-LSTM 模型分别在工作日训练集和测试集上, 均方误差随迭代次数增加的变化趋势。由误差的变化趋势可知, 在迭代前 10 次时误差值快速下降, 当迭代到 100 次时, 误差达到最小。最终误差值趋向平稳, 起伏波动较小, 训练集和测试集误差都较小, 并且误差相差不大, 模型预测能力精度较高, 说明 GA-LSTM 在预测中具有良好的平稳性。

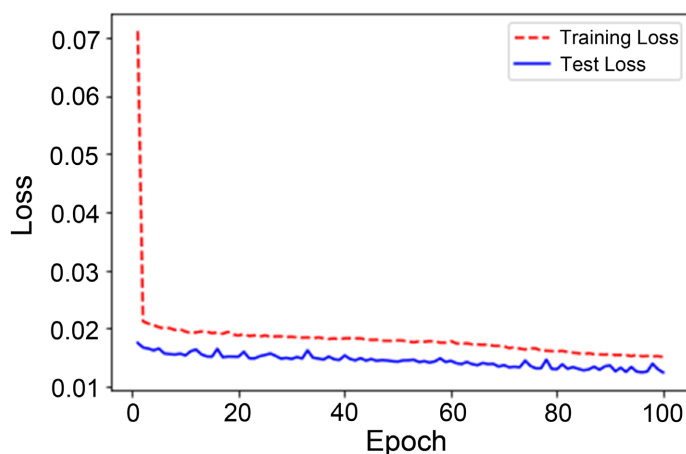


Figure 7. Mean square error of GA-LSTM iteration 100

图 7. GA-LSTM 迭代 100 次均方误差图

图 8 表示 CNN 模型在工作日训练集和测试集上, MSE 随迭代次数增加的变化趋势。根据图 8 中均方误差的变化趋势可知, 在模型迭代前 10 次时, 均方误差快速下降, 之后逐渐趋于平缓, 最终随着迭代次数的增加, 均方误差最终处于相对稳定的状态。由结果显示, CNN 模型在输入矩阵时, 充分考虑了地铁客流的时序特征和空间特征, 同时预测出郑州地铁 1 号线 20 个站台的客流量数据, 在工作日的客流量预测中具有较好效果。

为进一步评估深度学习模型的短时客流预测能力, 同时, 也通过十折交叉验证下的 SVR 回归作为对

比预测, 结果表示模型在训练集和测试集上的均方误差分别为 0.016 和 0.014。通过三个模型在测试集上的误差对比(见表 2)可以看出基于遗传算法改进的 LSTM 模型具有更高的精确度。与传统的机器学习方法(SVR)相比, 深度学习具备自动提取特征和预测的优势, 可以无需显式的特征提取和选择过程。如果针对海量数据, 传统的机器学习方法表现精度可能会大大下降, 但深度学习网络可以自动挖掘深层次的高级特征, 具有强大的表征地铁客流数据特征的能力。

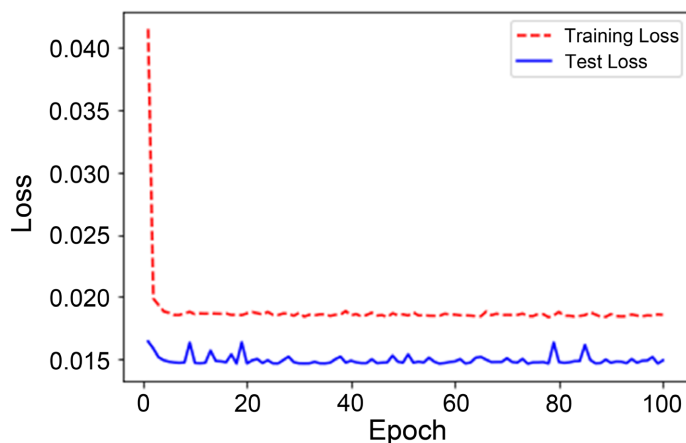


Figure 8. Mean square error of CNN iteration 100

图 8. CNN 迭代 100 次均方误差图

Table 2. Comparison of prediction errors between models

表 2. 模型间预测误差对比

模型	MSE
GA-LSTM	0.012
CNN	0.014
SVR	0.014

4. 结语

文章通过实例验证了通过遗传算法改进的 LSTM 深度神经网络模型在短时地铁客流中拥有更高的预测精度, 扩展了 LSTM 的模型应用范围, 同时也说明深度学习在时间序列预测中具有广泛应用价值。针对 LSTM 模型, 首先采用遗传算法优化 LSTM 网络模型的各个参数, 再对 LSTM 网络模型进行迭代优化, 最后利用训练好的 LSTM 网络模型对测试数据集进行预测, 预测出来的均方误差为 0.012, 相比 SVR 回归, 均方误差有所下降。以此说明利用遗传算法和 LSTM 神经网络对参数组合快速寻优的特性, 可以得到更高的预测精度, 表现出更好的预测性能。此外, CNN 网络模型在处理时间序列的同时, 也能提取高维特征, 说明不同特征间的关系, 通过实验数据, CNN 模型预测的均方误差值也仅为 0.014。但此次实验, 仅考虑地铁客流时间序列的本身的内部变化规律, 并未考虑天气、节假日、季节等影响客流的因素, 下一步将考虑更多因素的影响, 结合时间序列本身, 实现基于深度学习的多因素地铁客流预测。

参考文献

- [1] 赵明伟. 改进粒子群算法优化 LSTM 神经网络的地铁短时客流预测[D]: [硕士学位论文]. 石家庄: 石家庄铁道大

- 学, 2021.
- [2] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (2008) *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, New Jersey, 677-678.
 - [3] 袁振洲. 应用自回归积分移动平均法预测铁路货源货流发展趋势[J]. 铁道学报, 1996(S1): 52-56.
 - [4] 姚智胜, 邵春福, 熊志华, 岳昊. 基于主成分分析和支持向量机的道路网短时交通流量预测[J]. 吉林大学学报(工学版), 2008(1): 48-52.
 - [5] Roos, J., Bonnevey, S. and Gavin, G. (2016) Short-Term Urban Rail Passenger Flow Forecasting: A Dynamic Bayesian Network Approach. 2016 15th *IEEE International Conference on Machine Learning and Applications (ICMLA)*, Anaheim, 18-20 December 2016, 1034-1039. <https://doi.org/10.1109/ICMLA.2016.0187>
 - [6] Liu, L., Chen, R.C., Zhao, Q., *et al.* (2019) Applying a Multistage of Input Feature Combination to Random Forest for Improving MRT Passenger Flow Prediction. *Journal of Ambient Intelligence and Humanized Computing*, **10**, 4515-4532. <https://doi.org/10.1007/s12652-018-1135-2>
 - [7] Zhao, Z., Chen, W., Wu, X., *et al.* (2017) LSTM Network: A Deep Learning Approach for Short-Term Traffic Forecast. *IET Intelligent Transport Systems*, **11**, 68-75. <https://doi.org/10.1049/iet-its.2016.0208>
 - [8] 李月龙, 唐德华, 姜桂圆, 肖志涛, 耿磊, 张芳, 吴骏. 基于维度加权的残差 LSTM 短期交通流量预测[J]. 计算机工程, 2019, 45(6): 1-5.
 - [9] Fu, R., Zhang, Z. and Li, L. (2017) Using LSTM and GRU Neural Network Methods for Traffic Flow Prediction. 2016 31st *Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, Wuhan, 11-13 November 2016, 324-328. <https://doi.org/10.1109/YAC.2016.7804912>
 - [10] Du, B., Peng, H., Wang, S., *et al.* (2020) Deep Irregular Convolutional Residual LSTM for Urban Traffic Passenger Flows Prediction. *IEEE Transactions on Intelligent Transportation Systems*, **21**, 972-985. <https://doi.org/10.1109/TITS.2019.2900481>
 - [11] Srivastava, N., Hinton, G., Krizhevsky, A., *et al.* (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *The Journal of Machine Learning Research*, **15**, 1929-1958.