

基于CNN-LSTM-Attention模型的股票预测及系统开发

周晓娅^{1*}, 何松芝²

¹重庆对外经贸学院数学与计算机科学学院, 重庆

²重庆对外经贸学院大数据与智能工程学院, 重庆

收稿日期: 2025年7月21日; 录用日期: 2025年8月18日; 发布日期: 2025年8月27日

摘要

近年来, 中国A股市场呈现出“高噪声、厚尾分布、政策扰动”等复杂特征, 传统计量模型难以刻画其特征。深度学习技术凭借其强大的非线性建模能力, 为金融时间序列预测提供了新范式。而现有研究多聚焦于美股或指数层面, 对A股个股的研究较少, 且缺乏系统化的模型部署与交互式展示。本文以京粮控股(000505)为研究对象, 构建CNN-LSTM-Attention混合模型, 融合CNN的局部特征提取能力、LSTM的长期依赖建模能力与Attention的动态权重分配机制, 提出一种适用于A股市场的次日收盘价预测框架。实验结果表明, 该模型在测试集上R²高达0.92, 显著优于传统的LSTM模型, 并通过FastAPI与ECharts实现了可部署、可交互的预测系统。

关键词

CNN, LSTM, 深度学习, 股票预测

Stock Prediction and System Development Based on the CNN-LSTM-Attention Model

Xiaoya Zhou^{1*}, Songzhi He²

¹School of Mathematics and Computer Science, Chongqing College of International Business and Economics, Chongqing

²School of Big Data and Intelligent Engineering, Chongqing College of International Business and Economics, Chongqing

Received: Jul. 21st, 2025; accepted: Aug. 18th, 2025; published: Aug. 27th, 2025

*通讯作者。

文章引用: 周晓娅, 何松芝. 基于 CNN-LSTM-Attention 模型的股票预测及系统开发[J]. 软件工程与应用, 2025, 14(4): 919-927. DOI: 10.12677/sea.2025.144081

Abstract

In recent years, China's A-share market presents complex features such as "high noise, thick-tailed distribution, and policy perturbations", which are difficult to be characterized by traditional econometric models. Deep learning technology provides a new paradigm for financial time series forecasting by virtue of its powerful nonlinear modeling capability. Existing research focuses on the U.S. stock or index level, with less research on A-share stocks and a lack of systematic model deployment and interactive display. In this paper, we take Beijing Food Holdings (000505) as the research object, construct a hybrid CNN-LSTM-Attention model, integrate the local feature extraction capability of CNN, the long-term dependence modeling capability of LSTM and the dynamic weight allocation mechanism of Attention, and propose a next-day closing price prediction framework for A-share market. The experimental results show that the model has a high R^2 of 0.92 on the test set, which significantly outperforms the traditional LSTM models, and realizes a deployable and interactive prediction system through FastAPI and ECharts.

Keywords

CNN, LSTM, Deep Learning, Stock Prediction

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

金融时间序列预测研究历经了从传统计量模型、机器学习到深度学习的演进。中国 A 股市场受 T + 1 交易机制、涨跌停限制及散户主导等因素影响,呈现出高噪声、厚尾分布和政策扰动显著的特征。这些特性使得基于线性假设的传统 ARIMA-GARCH 类模型难以有效捕捉市场中的跳空缺口与突发性政策冲击。在此背景下,机器学习凭借其强大的非线性建模能力、海量数据处理潜力以及在金融预测、算法交易、风险管理和投资组合优化等领域的广泛应用前景,吸引了研究者的广泛关注。作为金融科技革新的重要驱动力,深度学习为金融领域提供了众多创新工具与方法。股票市场作为“宏观经济的晴雨表”,其预测不仅关乎金融行业,也间接影响社会经济生活。

在众多深度学习方法中,循环神经网络(RNN)因其对时序依赖性的建模能力而被广泛用于金融序列分析。然而,标准 RNN 面对长序列时易受梯度消失或爆炸问题困扰[1]。Hochreiter S [2]在 RNN 的基础上提出了 LSTM 神经网络, LSTM 神经网络通过设置遗忘门机制,有效地解决了 RNN 出现的梯度消失和梯度爆炸问题。耿晶晶[3]等人将 CNN-LSTM 模型与 BPNN 模型、SVR 模型、LSTM 模型分别进行股票价格预测对比,结果表明 CNN-LSTM 模型在股票价格预测领域更加具有参考价值。注意力机制通过动态生成注意力权重分布,使模型能够自适应地评估输入信息中不同元素对于当前任务的相关性权重,从而实现对关键特征的聚焦。这种基于注意力权重的特征选择机制,能够有效提升模型对重要信息的表征能力,进而优化神经网络的整体性能表现。林杰等[4]针对时间序列动态不稳定和长期依赖特性,构建的基于注意力机制的 LSTM 股价预测模型,对于股价的预测效果比无注意力机制的 LSTM 股价预测效果更好。

为更精准地捕捉 A 股市场的复杂动态,本文提出一种融合卷积神经网络(CNN)、LSTM 与注意力机制(Attention)的新型预测模型(CNN-LSTM-Attention)。该模型旨在结合 CNN 的局部特征提取能力、LSTM

的长时记忆能力以及注意力机制的关键信息聚焦优势。本文选取 2016 年 7 月 1 日至 2025 年 7 月 1 日中国 A 股市场的股票收盘价数据进行实证分析, 并将所提模型与传统机器学习基准模型进行对比。实证结果表明, 该融合模型在预测精度上取得了显著提升。不仅如此本次研究还将模型落地为网页端, 可直接访问网页进行简单的操作便可使用该模型。

2. 模型介绍

2.1. LSTM 模型

LSTM 网络是在递归神经网络(RNN)的拓扑结构上进行改进得来的[5]。作为 RNN 的变体, LSTM 模型通过引入可控自循环, 有效解决了 RNN 因网络层数增多和时间流逝而带来的梯度消失问题, 其巧妙的设计结构, 特别适合处理延迟和时序间隔较长的任务。LSTM 神经网络结构不同于其他深度学习算法之处在于其特殊的神经元细胞状态, 可在长期状态下学习需要记录和遗忘的序列数据信息。细胞单元内部由遗忘门、输入门和输出门三大门限单元构成[6]。整体结构如图 1 所示。

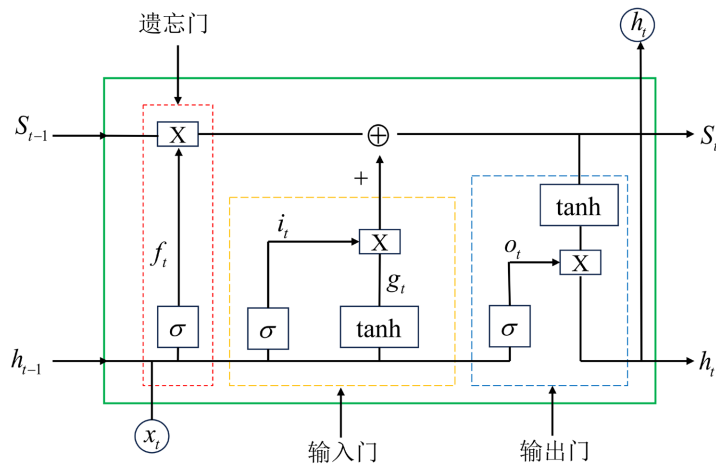


Figure 1. Overall structure of the LSTM network

图 1. LSTM 网络的整体结构

在遗忘门中, 当前输入 x_t 与上一时刻的状态记忆单元 S_{t-1} 以及中间输出 h_{t-1} 共同决定了状态记忆单元中被遗忘的部分。输入门中的 x_t 分别通过 sigmoid 和 tanh 函数变化后共同决定状态记忆单元中保留向量。中间输出 h_t 由更新后的状态记忆单元 S_t 和输出门 o_t 共同确定。具体公式如下:

$$\text{遗忘门: } f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f)$$

$$\text{输入门: } i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i)$$

$$\text{输入节点: } g_t = \phi(W_{gx}x_t + W_{gh}h_{t-1} + b_g)$$

$$\text{输出门: } o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o)$$

$$\text{状态单元的状态: } S_t = g_t \odot i_t + S_{t-1} \odot f_t$$

$$\text{中间输出: } h_t = \phi(S_t) \odot o_t$$

其中, W 和 b 分别为三个门权重矩阵与偏置项, 通过训练得到; $\odot$ 表示向量中元素按位相乘; σ 表示 sigmoid 函数变化; ϕ 表示 tanh 函数变化[7]。

2.2. CNN 模型

卷积神经网络(CNN 模型)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。

CNN 由输入层(Input)、卷积层(Convolution Layers)、池化层(Pooling)、扁平化全连接层(Flattening)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:

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

其中,规定输入层为第 $l-1$ 层, V_s^{l-1} 用于表示输入层第 s 个特征,输出层为第 l 层, V_j^l 用于表示输出层第 j 个特征, K_{sj}^l 用于表示卷积核的元素, b_j^l 为偏置项, σ 为激活函数[8]。

卷积层接受输入的数据并将数据提取特征值,采取池化操作将抽象卷积层输出的特征信息全连接层得到的输出与标签作对比,不断迭代更新权值,从而实现反向传播。本文选取 CNN 模型的卷积层和池化层,保留了模型特征提取和信息抽象的功能,提取数据中含有因子间强相关的特征,为后续预测做准备。

2.3. 注意力机制

注意力机制通过动态分配权重聚焦关键时序特征,有效解决了长序列中信息稀释的问题。其核心思想是让模型在每一时间步自动学习输入序列不同位置的关注强度,从而增强重要特征的贡献度。在 LSTM 层输出的时序特征基础上, Attention 机制的计算分为三步分别为评分函数,权重分配,上下文向量。

2.4. CNN-LSTM-Attention 模型

本文将 CNN,注意力机制与 LSTM 进行融合构建 CNN-LSTM-Attention 深度学习预测模型先利用 CNN 模型的卷积层(Convolution Layer)、池化层(Pooling Layer)对输入数据提取特征、抽象信息得到特征数据,再经由 Flattening 层后输入到 LSTM 模型进行关联预测,最后接入全连接层(Full Connection Layer)得到模型的预测值。卷积和池化操作降低了输入数据的复杂性,可有效处理异常值数据,防止过拟合。接入 LSTM 层,利用 LSTM 的遗忘门与输入门可对关联数据信息进行有效的筛选与更新,更好地预测关联数据,最后引入多头注意力机制对 LSTM 输出的隐藏状态进行自适应加权,通过计算自注意力权重矩阵聚焦关键时间步的特征表示,生成具有判别性的上下文向量更好的提高本模型的性能。其流程图如图 2 所示。

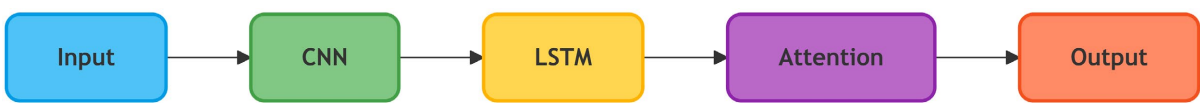


Figure 2. CNN-LSTM-Attention model flowchart

图 2. CNN-LSTM-Attention 模型流程图

3. 实证分析

首先利用 akshare 接口,通过 python 脚本自动获取数据并存储为 csv 格式。其次进行数据预处理包括数据清洗,缺失值填充,指标体系构建并进行可视化展示,然后构建 lstm 模型以及 cnn-lstm-attention 模型并对其性能进行对比以及前段的构建。

3.1. 数据收集及可视化

数据收集是实证分析的第一步。通过编写 Python 脚本,利用 akshare 库从指定的接口自动获取指京粮控股(000505)的价格数据。数据包括该股票从 2016 年 7 月 1 日到 2025 年 7 月 1 日间的开盘日的开盘价、收盘价、最高价、最低价、成交量、成交额等数据。数据收集后,使用 pandas 库进行数据清洗和预

处理, 包括去除无效数据、填充缺失值、转换数据类型以及将时间戳转换为日期格式等, 并对收盘价进行可视化如图 3 所示。由图可知该股票价格在 2016 年 7 月 1 日至 2025 年 7 月 1 日总体不平稳经历过多次涨跌。

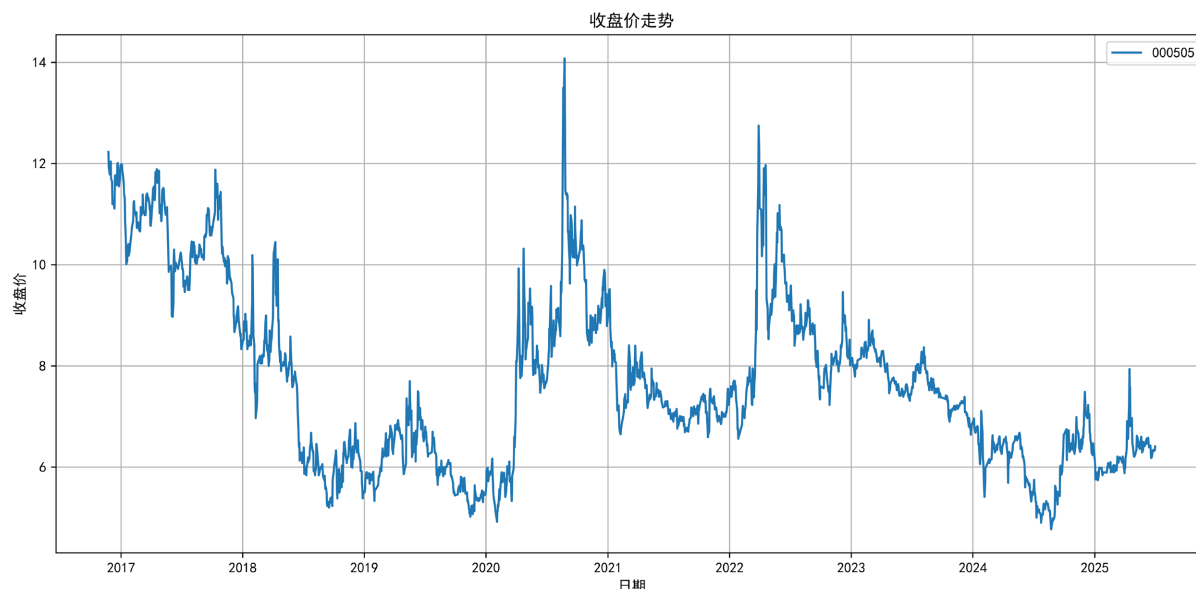


Figure 3. Stock closing price chart

图 3. 股票收盘价格图

3.2. LSTM 模型构建

本研究将 LSTM 模型作为基准模型在原有的数据集上划分训练集和测试集, 测试集占原有数据的 20%, 且对收盘价进行单独的归一化防止信息泄露在后续实验中造成过拟合。本文初始学习率设置为 0.001 且采用网格搜索, 早停机制以及 Adam 优化器, 防止过拟合以及寻找到最佳参数。最终该模型 MSE 为 0.000408, R^2 为 0.86。

3.3. CNN-LSTM-Attention 模型

为进一步提升预测性能, 本研究提出融合 CNN、LSTM 和注意力机制的混合模型。该模型采用层次化架构设计, 首先通过一维卷积层提取局部时序特征, 设置卷积核数量在 32~128 之间动态调整, 配合 ReLU 激活函数捕捉收盘价序列中的短期波动模式。随后使用步长为的最大池化层进行特征降维, 在保留关键信息的同时增强模型抗噪能力。经过卷积处理的特征输入到 LSTM 层进行时序建模, 其中 LSTM 单元数利用网格搜索调优, 并设置 `return_sequences = True` 以保留完整的时间步输出。

模型创新性地引入注意力机制层, 通过 Dense 层配合 softmax 激活函数生成各时间步的注意力权重, 再与 LSTM 输出进行逐元素相乘实现特征加权。这种设计使得模型能够自适应地聚焦于重要时间节点, 如财报发布、政策调整等关键事件窗口。采用 Tuner 工具进行超参数优化, 通过多轮随机搜索确定最佳参数组合, 最终选定 64 个卷积核和 96 个 LSTM 单元的最优配置。实验结果显示, CNN-LSTM-Attention 模型在测试集上的 MSE 降至 0.000331, R^2 提升至 0.92, 显著优于基础 LSTM 模型。最后通过逐步剔除 CNN 或 Attention 的消融实验得出结果如表 1 所示。

根据表 1 可知, CNN 模型的加入为仅有 LSTM 模型的 R^2 提高了 0.04, attention 的加入也为 LSTM 模型提高了 0.04。综合下来 CNN-LSTM-Attention 模型性能优于 LSTM 模型。 R^2 高达 0.92 且 MSE 低于

LSTM 模型。

Table 1. Melting experiment results
表 1. 消融实验结果

模型名称	MSE	R^2
CNN-LSTM-Attention	0.000331	0.9200
LSTM-Attention	0.000385	0.9000
CNN-LSTM	0.000412	0.8984

3.4. 模型预测

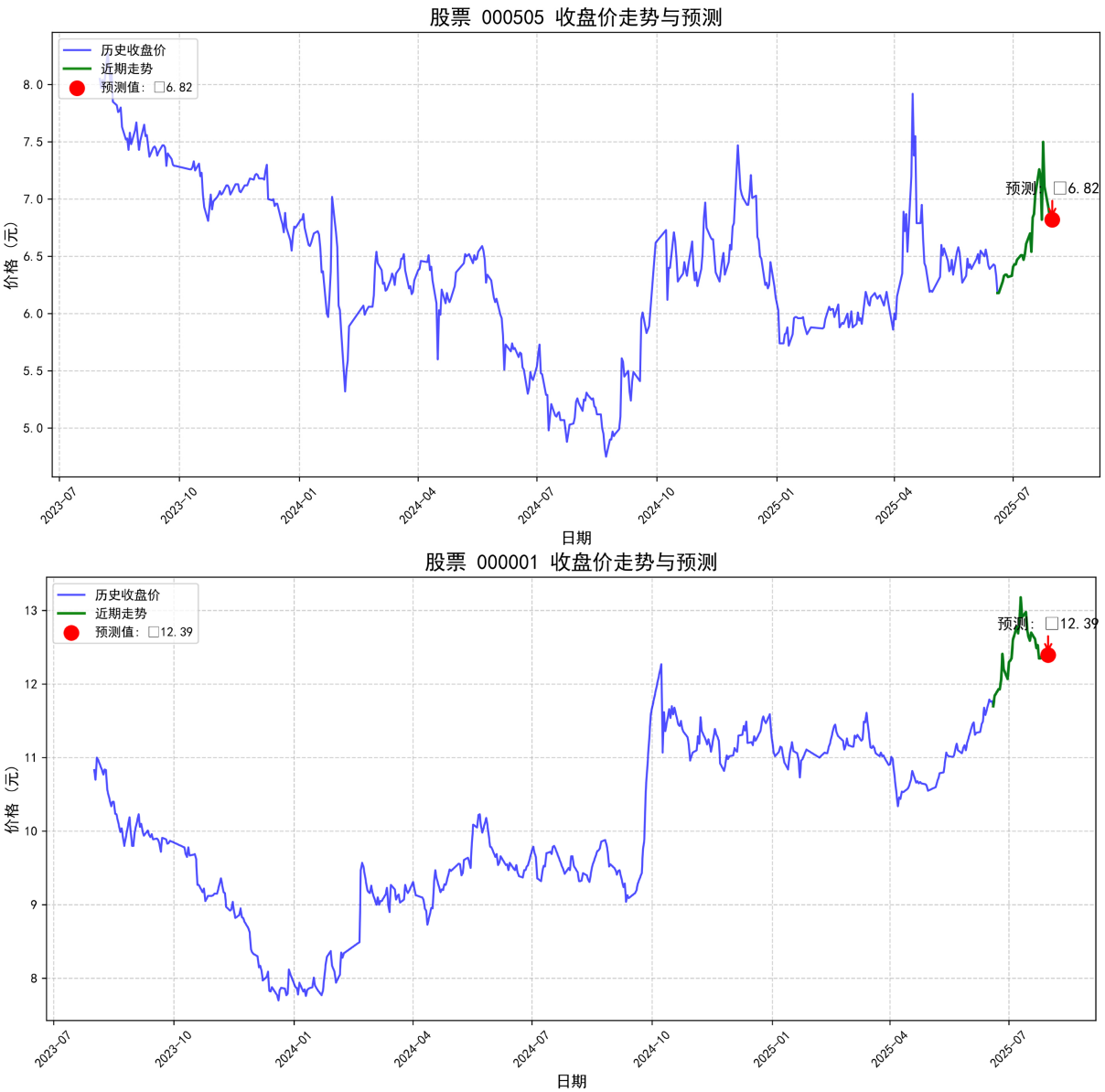


Figure 4. Stock closing price forecast chart
图 4. 股票的收盘价预测图

本研究将该模型保存为 `best-model.pk` 以及权重为 `scaler-close.pkl` 后利用该模型对京粮控股(000505)和平安银行(000001)股票进行预测其结果如图 4 所示。

3.5. 系统设计与部署

本研究基于 Gradio 框架开发了一套面向金融领域的智能化股票价格预测系统, 该框架的创新性应用彻底改变了传统机器学习模型部署的复杂流程。如图 5 所示的系统架构图清晰展现了从数据输入到结果输出的完整技术路线, 其中 Gradio 的核心价值体现在其革命性的快速开发能力和极简部署方案上。相较于需要专业前端开发团队参与的 Flask 或 Django 等传统框架, Gradio 的自动化组件仅需数十行代码就能构建出功能完备的交互界面。在系统设计方面, 本文充分利用了 Gradio 的多模态输入输出特性, 构建了高度智能化的用户交互体验。

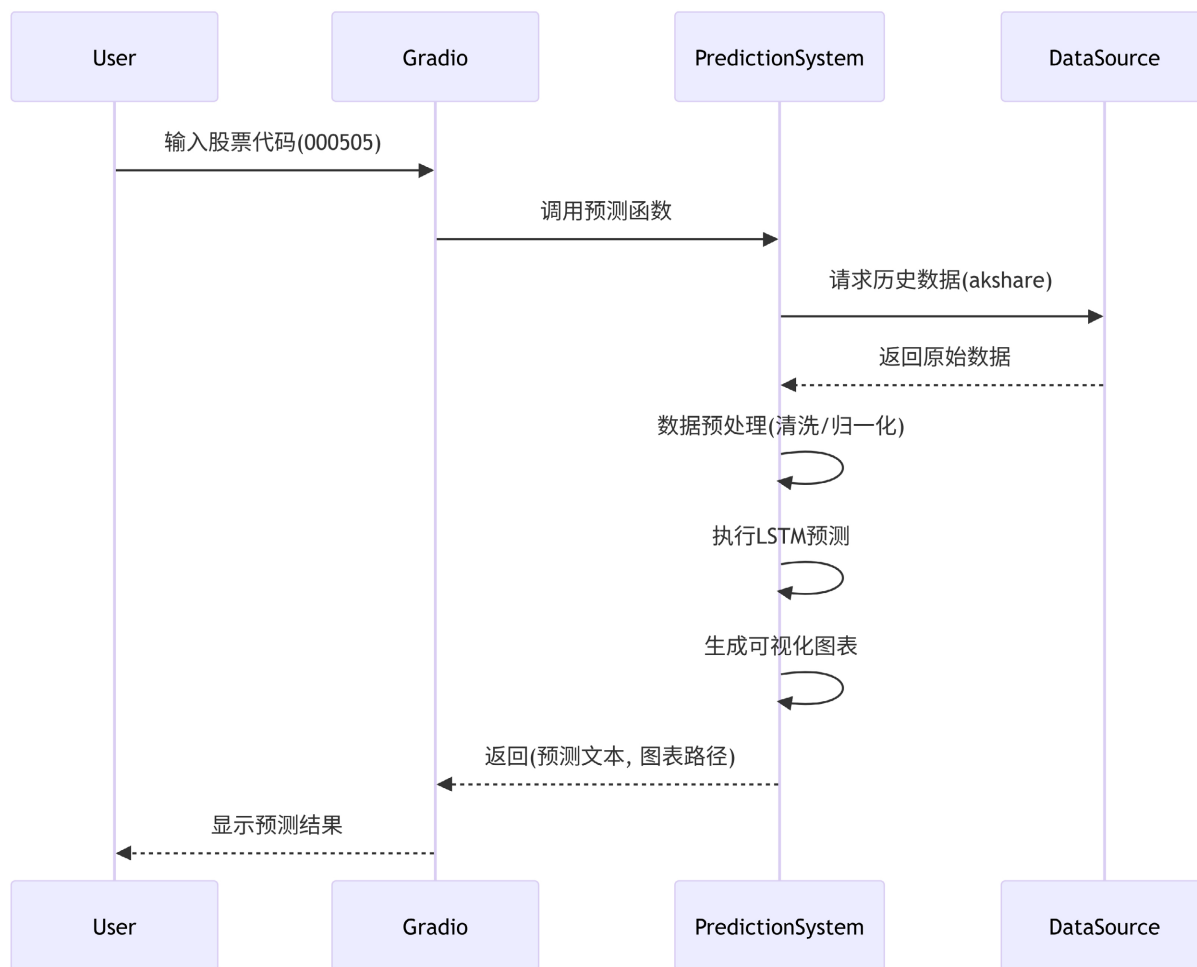


Figure 5. System architecture diagram

图 5. 系统总架构图

系统前端采用响应式布局设计, 适配各类终端设备, 用户界面如图 6 所示, 其核心交互流程经过精心优化: 投资者只需在醒目的输入框内输入目标股票代码(支持沪深 A 股), 点击智能预测按钮后, 系统将在毫秒级时间内完成从数据获取到结果展示的全流程处理。

系统的核心技术优势体现在其强大的实时计算能力上。基于 Gradio 的高效异步处理机制, 系统能够

并行处理多个预测请求，平均响应时间控制在 500 毫秒以内。

A股次日收盘价预测系统

基于深度学习融合模型开发，输入股票代码获取次日收盘价预测和可视化分析

注：首次加载可能需要30秒初始化模型和数据



Examples

000505 000001 000002 000003

模型说明: 本系统采用融合CNN特征提取、LSTM时序建模和Attention关键信息聚焦的三重神经网络架构，能够有效捕捉A股市场的非线性特征和政策扰动影响。

Figure 6. System interface diagram

图 6. 系统界面图

在部署运维方面，该系统充分展现了 Gradio 的独特优势。通过简单的 API 调用即可启动高性能的嵌入式 Web 服务器，自动生成可全球访问的 HTTPS 安全链接。系统支持多种访问控制策略，包括密码保护、IP 白名单等功能，确保商业应用的隐私性和安全性。此外，Gradio 的版本管理功能使得模型迭代更新过程完全无缝，研究人员可以随时部署新模型而不会中断现有服务。

4. 总结

本文将构建的 CNN-LSTM-Attention 模型预测方法用于京粮控股(000505)进行股价预测, 并与 LSTM、LSTM-Attention、CNN-LSTM 三个模型进行对比, 结果表示本文构建的 CNN-LSTM-Attention 模型性能优于其他模型, 其 R^2 能够达到 0.92, 高于其他模型。在此基础上, 本文保存该模型且利用 Gradio 框架构建了股票预测系统并将其部署在网络上。最后将多支股票(如平安银行, 万科 A 等)的股票价格预测中该模型均表现出良好的性能。由于网络的迅速发展, 舆论对股票市场的影响越来越大, 本研究暂未考虑该方面的外部影响因素, 如投资者情绪数据、国家战略政策调整等, 后续将继续研究试验更好的模型、算法等使得可以考虑到更多的影响因素并对其进行深入、详细的剖析, 进一步增强模型的预测准确、抗噪、泛华能力, 使得模型性能得到提升。此外, 本研究主设备仅在一台普通主机上, 在多人同时访问等情况下可能会有数据泄露程序奔溃等问题, 未来将进一步探索 Gradio 在分布式计算和边缘设备部署等方面的创新应用, 持续提升系统的性能和适用范围满足多人同时使用的性能且加入模型部署功能使得私人或公司可以离线部署本模型防止数据泄露等。

基金项目

2024~2025 年重庆对外经贸学院科研项目: 基于机器学习的股票价格预测系统研究(KYZK2024018)。

参考文献

- [1] 管健. 基于 RNN-CNN 模型股票价格预测方法研究[D]: [硕士学位论文]. 南京: 南京信息工程大学, 2023.
- [2] 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>
- [3] 耿晶晶, 刘玉敏, 李洋, 等. 基于 CNN-LSTM 的股票指数预测模型[J]. 统计与决策, 2021, 37(5): 134-138.
- [4] 林杰, 康慧琳. 基于注意力机制的 LSTM 股价趋势预测研究[J]. 上海管理科学, 2020, 42(1): 109-115.
- [5] Bukhari, A.H., Raja, M.A.Z., Sulaiman, M., Islam, S., Shoaib, M. and Kumam, P. (2020) Fractional Neuro-Sequential ARFIMA-LSTM for Financial Market Forecasting. *IEEE Access*, **8**, 71326-71338.
<https://doi.org/10.1109/access.2020.2985763>
- [6] 汪定, 邹云开, 陶义, 等. 基于循环神经网络和生成式对抗网络的口令猜测模型研究[J]. 计算机学报, 2021, 44(8): 1519-1534.
- [7] 周章元, 何小灵. 基于优化 LSTM 模型的股价预测方法[J]. 统计与决策, 2023, 39(6): 143-148.
- [8] 曹超凡, 罗泽南, 谢佳鑫, 等. MDT-CNN-LSTM 模型的股价预测研究[J]. 计算机工程与应用, 2022, 58(5): 280-286.