

基于机器学习的农产品期货价格预测模型研究

丁雅芳, 孙娜, 李靓靓

河北金融学院河北省科技金融重点实验室, 河北 保定

收稿日期: 2025年11月1日; 录用日期: 2025年12月1日; 发布日期: 2025年12月5日

摘要

面向农产品期货价格波动特征, 基于机器学习构建预测模型, 实现趋势识别与波动响应的高精度拟合。研究内容涵盖价格数据预编码、时序因子滑动窗口编码、LSTM深度网络配置与偏差残差修正等技术环节, 形成完整预测流程。实验以真实期货交易数据为样本, 在多时间步区间内检验模型预测效果。结果表明, 真实价格与预测价格在趋势与波动阶段高度一致, 残差分布集中稳定, 误差幅度较低, 验证方法在捕捉价格动态特征方面具备较强适应性与时序稳定性。

关键词

机器学习, 农产品期货价格, 预测模型, 残差修正

Research on Agricultural Product Futures Price Prediction Model Based on Machine Learning

Yafang Ding, Na Sun, Liangliang Li

Science and Technology Finance Key Laboratory of Hebei Province, Hebei Finance University, Baoding Hebei

Received: November 1, 2025; accepted: December 1, 2025; published: December 5, 2025

Abstract

Focusing on the fluctuation characteristics of agricultural product futures prices, a prediction model is constructed based on machine learning to achieve high-precision fitting of trend identification and volatility response. The research content covers technical links such as price data pre-coding, time series factor sliding window encoding, LSTM deep network configuration and deviation residual correction, forming a complete prediction process. Experiments use real futures trading data as samples and test the model's prediction effect within multiple time steps. The results show that the real price

文章引用: 丁雅芳, 孙娜, 李靓靓. 基于机器学习的农产品期货价格预测模型研究[J]. 计算机科学与应用, 2025, 15(12): 133-137. DOI: 10.12677/csa.2025.1512329

and the predicted price are highly consistent in terms of trend and volatility stages, the residual distribution is concentrated and stable, and the error range is low. This verification method has strong adaptability and temporal stability in capturing price dynamic characteristics.

Keywords

Machine Learning, Agricultural Product Futures Price, Prediction Model, Residual Correction

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

农产品期货市场在价格形成与风险管理中具有关键作用，其价格受气候波动、政策调控与供需变化等多重因素影响，序列波动性强、非线性特征显著。近年来，机器学习在非线建模与特征提取方面展现出较强优势，为农产品期货价格预测提供新思路。向永辉等[1]总结了机器学习在国际贸易与大宗商品预测中的应用规律，提出基于 LSTM 与 GRU 的多层预测框架以捕捉价格波动趋势。冯景等[2]构建了融合 KNN、决策树与随机森林算法的农产品进销预测模型，实现新疆棉花市场的动态供需匹配。王润周等[3]提出基于 CEEMD-TCN-BiSeq2seq 的多阶段深度学习模型，在猪肉、菠菜、苹果等农产品价格预测中验证了信号分解与时序建模结合的优越性能。现有研究虽在算法集成与特征提取方面取得进展，但多集中于单品种短期预测，残差修正与多时间尺度趋势捕捉仍待完善。基于此，本文创新在于引入滑动窗口自适应与残差动态加权策略，构建多维时序编码与误差闭环修正模型，旨在提升农产品期货价格预测的稳定性与泛化能力。

2. 机器学习概述

机器学习是一类依托算法实现数据自动识别、特征提取、规律挖掘的计算方法，能够在缺乏明确函数关系时完成高维数据映射与趋势预测[4]。其核心思想在于利用训练数据集提取潜在统计关系，再利用分类、回归、聚类、降维等多种算法实现目标预测。监督学习(SL)依赖标签信息完成映射关系学习，无监督学习(UL)侧重潜在分布模式识别，强化学习(RL)基于动态反馈优化预测策略[5]。在农产品期货价格预测模型研究中，支持向量机(SVM)、随机森林(RF)、梯度提升(GB)、循环神经网络(RNN)、长短期记忆网络(LSTM)等方法具备识别复杂时序关联关系能力。结合滑窗编码、特征工程、误差修正等技术手段，可增强价格趋势刻画精度，为动态波动环境下的预测任务提供高效数值基础与可靠量化支撑。

3. 基于机器学习的农产品期货价格预测模型构建

3.1. 期货价格数据预编码

机器学习输入阶段完成原始价格序列规整化与数值尺度统一处理，农产品期货价格预测模型需对结算价、开盘价、收盘价、最高价、最低价等多维时序数据进行统一编码。原始价格序列记为 R_t ，价格数值区间位于 2800~3000 CNY/吨，为提取连续时间步的价格变化特征，对数收益率定义为(1)：

$$R_t = \ln \frac{P_t}{P_{t-1}} \quad (1)$$

其中 P_t 表示当前时刻价格, P_{t-1} 表示前一时刻价格。时序长度设定为 180, 采样频率为 1 天, 时间序列在原始维度中保留结算价主线。为了减少数值波动影响并统一尺度, 对 R_t 实施标准化运算, 公式为(2):

$$\bar{R}_t = \frac{R_t - \mu}{\sigma} \quad (2)$$

其中 μ 为滑动窗口均值, σ 为滑动窗口标准差, 滑动窗口长度设置为 20, 步长设置为 1。该预编码步骤对价格序列完成连续化与数值归一, 使农产品期货价格预测模型能够接收满足时序特征要求的稳定输入矩阵[6]。

3.2. 时序因子滑窗编码

农产品期货价格标准化收益序列在机器学习预测输入阶段执行时间维度特征展开与多尺度信息聚合[7]。编码过程设定滑动窗口长度在 20 至 60 区间, 时间步长在 1 至 5 区间, 通过固定时间间隔对价格变化率序列递进抽取。实验数据以某主力农产品期货合约 180 个连续交易日价格为基础, 每个时间窗口覆盖 20 至 60 个交易日价格波动信息, 并在相邻窗口间设置步长为 1 至 5 的交叠区间, 实现时间维度连续滑移。每个窗口在时间维度上同步提取均值、方差、自相关系数(ACF)、偏度(Skew)、峰度(Kurt)等统计因子, 窗口样本容量与滑动频率共同决定特征提取密度。统计量提取阶段在每个窗口中生成不超过 30 维的特征向量, 因子数量随窗口配置调整[8]。编码过程在全时序范围形成长度为 180 的滑移序列, 窗口与步长组合形成多重时间尺度映射。

3.3. LSTM 模型结构配置

针对农产品期货价格预测中时间依赖关系与趋势变化特征的复杂性, LSTM 在机器学习序列建模过程中具备稳定捕捉非线性时序规律的能力, 因此在模型配置阶段需针对价格波动特性进行精细参数设计与信息流动控制[9]。输入信号来源于经过滑窗编码处理的价格序列, 时间维度信息在输入层完成顺序展开与特征对齐后进入记忆单元, 实现跨时间步的动态关联捕捉, 如图 1 所示。

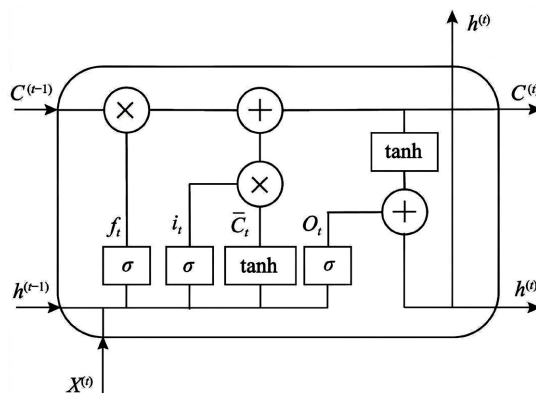


Figure 1. Schematic diagram of LSTM network structure
图 1. LSTM 网络结构示意图

输入序列长度可取 100 至 300, 特征维度 10 至 40, 批次大小 32 至 128。隐藏层采用 2 至 4 层 LSTM 单元, 单层隐藏单元数 64 至 256, 门控机制在输入门、遗忘门、输出门间实现筛选与更新, 使季节性波动与供需扰动信号保持动态响应。激活函数使用 Tanh, 梯度传播维持稳定收敛[10]。Dropout 比例 0.2 至 0.5, 权重采用分布式初始化。优化策略基于 Adam, 学习率 0.0005 至 0.005, β_1 取 0.9, β_2 取 0.999。损失函数选用 MSE, 输出层神经元数量 1 至 3, 用于价格趋势映射与信号响应控制。

3.4. 回测偏差残差修正

在农产品期货价格预测任务中,机器学习输出结果与真实价格间常存在系统性偏差与波动残差[11]。为提高预测结果的时序一致性与数值稳定性,回测偏差残差修正环节对预测序列与实盘价格序列执行差异校准与统计补偿。价格采样长度设定 100~300,时间步 1~5,取样间隔 1 d、5 d 或 10 d。预测输出与真实价格逐点对齐,经标准化与一阶差分处理形成残差序列,残差窗口长度 20~60,滚动间隔 1~3。偏差识别环节采用线性漂移修正策略,偏移幅度控制在 ± 5 至 ± 20 ,步长 0.5~2。残差部分经平滑滤波与自回归残差拟合提取短期扰动信号,滤波跨度 3~10,拟合阶数 1~5。修正计算采用动态加权与漂移补偿结合形式,表达为(3):

$$\bar{P}_t = P_t^{pred} + \alpha_t (P_t^{real} - P_t^{pred}) + \beta_t (\varepsilon_t - \bar{\varepsilon}) \quad (3)$$

其中, \hat{P}_t 为修正预测价格, P_t^{pred} 为原始预测值, P_t^{real} 为真实价格, $\varepsilon_t = P_t^{real} - P_t^{pred}$ 为残差项, $\bar{\varepsilon}$ 为滑动窗口残差均值, α_t 为漂移修正系数(0.2~0.6), β_t 为残差动态权重(0.1~0.5)。公式反映模型在误差分解与动态加权下的双层补偿机制,使预测值在时间维度与数值维度上保持与真实走势的同步关系,实现农产品价格波动特征的稳健捕捉与差异修正。

4. 实验验证

选取成交量高、价格波动显著的主力农产品期货合约作为样本,数据来源于国家级期货交易所公开历史数据库,时间区间为 2024 年 1 月至 2024 年 12 月,采样频率为日度,共 300 个时间步,变量涵盖收盘价、开盘价、最高价、最低价与成交量。异常值经箱线检验剔除,缺失段采用线性插值修复,时间戳对齐保证序列连续性。样本按 7:2:1 划分训练、验证与测试集,验证区间价格波动幅度较高,具备代表性。特征编码阶段采用对数收益率变换与标准化处理,滑窗长度设为 30,窗口内提取均值、方差、自相关、偏度、峰度五类统计指标构建输入矩阵。预测模型采用三层 LSTM 网络,单层神经元数量 64,激活函数为 Tanh,优化算法为 Adam,学习率 1×10^{-3} ,Batch 大小 128,训练轮次 200,Dropout 比例 0.2,损失函数为 MSE。偏差残差修正环节依据公式(3)对预测输出执行线性漂移补偿与残差拟合,漂移系数 $\alpha_t = 0.3$ 、残差权重 $\beta_t = 0.2$ 、残差窗口长度 30。实验在 Python3.9 与 TensorFlow2.15 环境下完成,模型训练与推理在 GPU 加速环境下运行。真实价格序列与预测结果逐点对齐后绘制对比折线图,如图 2 所示。

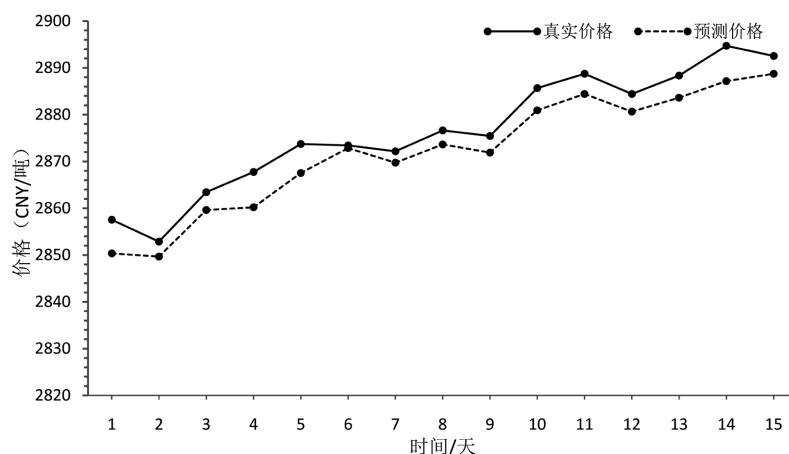


Figure 2. Comparison curve of actual prices and predicted prices of agricultural commodities futures

图 2. 农产品期货真实价格与预测价格对比曲线

真实价格与预测价格在 15 个时间步中走势接近, 整体数值区间从 2850 逐步攀升至 2890, 趋势稳定。1~3 时间步预测值低于真实值, 偏差保持在 4~7 范围内, 体现模型在低价起始阶段的响应稳定性。4 至 8 时间步进入波动增长区间, 预测价格与真实价格同步上扬, 峰值差距最大不超过 8, 时序依赖映射在中段区间得到体现。9~13 时间步上行趋势增强, 预测与真实几乎重叠, 偏差收敛至 3 以内, 残差接近零均值, 偏差残差修正有效抑制系统误差并维持一致性。14~15 时间步峰值区间预测与真实差距继续缩小, 波峰波谷位置重合, 价格走势高度吻合。整体趋势中预测价格始终紧贴真实价格, 未出现相位漂移与偏移累积, 验证模型在趋势识别、波动响应与非线性拟合环节具备稳定性与适应性。

5. 结论

基于机器学习的农产品期货价格预测模型在时序趋势捕捉、短期波动响应与残差偏差抑制环节展现稳定预测性能。数据预编码与时序因子滑窗编码增强了价格信号时间依赖特征表达, LSTM 对价格序列中趋势段与震荡段的非线性映射较为精准, 偏差残差修正策略有效减弱了预测偏离。真实价格与预测价格曲线在峰谷位置保持高度一致, 短期扰动期间预测轨迹未出现系统性漂移, 残差分布稳定集中, 误差幅度维持低位区间。未来, 研究可在扩展数据样本规模、引入多因子特征和优化残差修正策略等方向进一步提升预测能力与应用适应性。

参考文献

- [1] 向永辉, 王喆. 机器学习在国际贸易高质量预测中的应用[J]. 浙江科技大学学报, 2025, 37(4): 365-376.
- [2] 冯景, 王长军, 杨东, 等. 机器学习驱动的农产品进销决策——以新疆棉为例[J]. 工业工程与管理, 2024, 29(5): 94-103.
- [3] 王润周, 张新生, 王明虎. 基于信号分解和深度学习的农产品价格预测[J]. 农业工程学报, 2022, 38(24): 256-267.
- [4] 许珠路, 王兴芬, 刘亚辉. 融合 CNN-BiLSTM-Attention 的集成学习价格预测[J]. 计算机系统应用, 2023, 32(6): 32-41.
- [5] 于晓华, 刘爽. “没有免费午餐定理”与农业政策研究中的算法选择: 以机器学习预测生猪价格变动为例[J]. 农业经济问题, 2024(5): 20-32.
- [6] 杨聪. 人工智能预测农产品价格的研究[J]. 南方农机, 2023, 54(12): 167-169+187.
- [7] 胡世洋, 刘威. 机器学习在农产品价格预测方面的研究与应用[J]. 电脑知识与技术, 2023, 19(33): 14-17.
- [8] 李慧东. 基于集成模型的农产品价格预测[J]. 信息技术与信息化, 2022(3): 44-47.
- [9] 胡越, 王桑原, 覃浩恒, 等. 基于双重 XGBoost 模型的农产品期货波动率预测——以玉米期货为例[J]. 系统管理学报, 2023, 32(2): 332-342.
- [10] 李忻怡. 乡村振兴背景下蔬菜价格波动因素分析及预测[J]. 科技和产业, 2023, 23(8): 126-130.
- [11] 张大斌, 黄均杰, 凌立文, 等. 基于注意力时间卷积网络的农产品期货分解集成预测[J]. 南京信息工程大学学报, 2024, 16(3): 311-320.