

# 基于多源数据与LSTM的大豆期货价格预测方法

缪鑫挺

杭州电子科技大学 经济学院 (浙江 杭州 310020)

**摘要:** 大连商品交易所的大豆期货交易, 为大豆产业发展增添了新的推动力, 融合多源数据大豆期货价格的预测则可以市场的稳定做出参考。本研究考量了影响我国大豆期货价格的核心因素, 利用大豆期货的历史交易价格数据及其他多源数据, 对于长短期记忆网络(LSTM)的价格预测模型。为提高价格预测效果, 模型针对不同时间窗口参数做了对比实验。结果表明, 时间窗口设定为20时, LSTM模型的预测效果最佳; 将此价格预测模型用于当年大豆期货价格预测, 也能较好地反映价格变动情况。

**关键词:** 多源数据; 大豆期货; LSTM; 价格预测

## 1 引言

国家统计局的数据显示, 2024年我国大豆总产量为2065.24万吨, 在全球产量排名中位列第四, 仅次于巴西、美国 and 阿根廷。回溯至1993年11月8日, 大豆期货在大连商品交易所正式挂牌交易, 经过多年发展, 对市场的稳定做出了杰出的贡献, 并且成为了农业经济中的支柱性农产品和重要“晴雨表”。大豆价格的波动直接关系到整个农业产业的发展走向, 因此深入研究其价格预测方法, 可为农业经济与产业的健康可持续发展提供有力的技术支持。

影响大豆期货价格的因素众多, 既包括市场内部因素, 如供需关系变化、投机活动活跃度、交易规模大小等, 也涉及外部环境因素, 像国际大豆期货行情、原油价格波动、宏观经济形势及政策调控措施等。在这些因素的交织作用下, 大豆价格呈现出显著的不稳定性和复杂波动特征。当前关于大豆期货价格的研究中, 普遍存在对产业上下游关联影响及国际期货价格中潜在预测信息考虑不足的问题。与现有研究相比, 本研究的特点在于综合考量影响我国大豆期货价格的多方面因素, 选取国内大豆期货历史价格、国际大豆期货价格、国内豆粕期货价格这三类核心指标作为多源输入数据, 进而构建基于LSTM的大豆期货价格预测模型。

## 2 文献综述

在金融市场预测领域, 相关研究多围绕模型与数据应用展开<sup>[1-2]</sup>。有研究将多源数据情感分析融入金融产品价格的预测研究, 挖掘了市场参与者情感因素对价格的影响, 开辟了新的研究路径<sup>[3]</sup>。而把多源数据与具有自注意力机制的LSTM结合用于股票价格预测的探索, 既发挥了多源数据的优势, 又优化了LSTM模型性能, 为模型改进提供了方向<sup>[4]</sup>。

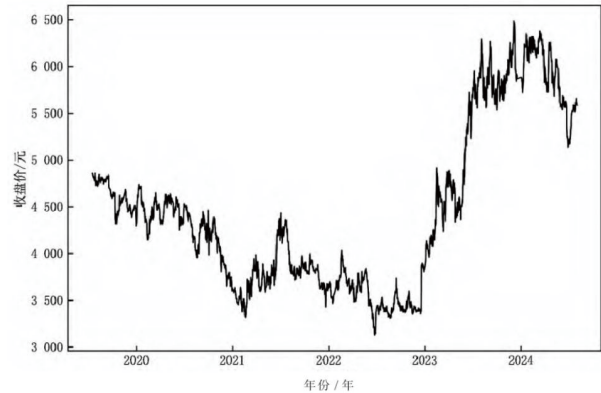


图1 国内大豆期货收盘价走势

基于深度学习对金融时间序列预测的研究, 展示了深度学习处理复杂金融数据的潜力, 为LSTM等深度学习模型在金融预测中的应用奠定了理论基础<sup>[5]</sup>。基于神经网络的股票价格预测研究, 探索了神经网络结构调整与优化对预测的影响, 为其他金融产品价格预测模型的构建提供了参考<sup>[6]</sup>。提出创新神经网络方法用于股票市场预测的研究, 其结构与应用方式为金融市场预测模型创新提供了思路, 激励研究者改进现有模型<sup>[7]</sup>。运用LSTM对河南省花生价格进行预测的实践, 证明了LSTM在农产品价格预测中的可行性, 为其应用于金融产品价格预测提供了实践参考<sup>[8]</sup>。

## 3 LSTM的大豆期货多源数据预处理方法

### 3.1 多源数据预处理

为确保LSTM模型在大豆期货价格预测中的有效性, 需对选取的多源输入数据(国内大豆期货历史价格、国际大豆期货价格、国内豆粕期货价格)进行系统性预处理, 以消除数据噪声、统一数据格式并提升数据质量<sup>[9]</sup>。由于不同来源的数据可能存在量纲差异, 需通过标准化处理将所有数据映射到[0,1]区间, 采用的公式为:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

其中X为原始数据， $\min(x)$ 和 $\max(x)$ 分别为该类数据中的最小值和最大值， $X'$ 为标准化后的数据。这一步骤可消除量纲对模型训练的影响，加快模型收敛速度。

此外，需对数据进行时间对齐处理，确保三类输入数据的时间维度一致，并剔除异常值，可通过箱线图法识别并替换超出合理范围的数值，以保证输入数据的稳定性与合理性。

### 3.2 模型构建

基于预处理后的多源数据构建的大豆期货价格预测LSTM模型，整体框架包含输入层、LSTM层和输出层<sup>[10]</sup>。输入层接收预处理后的多源时序数据，转换为三维张量格式（样本数、时间步长即时间窗口参数、特征数，其中特征数对应中国大豆期货价格、国际大豆期货价格、国内豆粕期货价格三类数据，故为3）；LSTM层作为核心特征提取层，设64个隐藏层神经元以学习时序数据特征、捕捉多源数据与大豆期货价格的动态关联，其输出经dropout率为0.2的Dropout层正则化处理，以减少过拟合、增强泛化能力；输出层采用单个神经元结合线性激活函数，直接输出未来某一时刻的价格预测值。

模型训练以预测值与实际价格的均方误差（MSE）为损失函数，用Adam优化器迭代优化参数，通过调整初始为0.001的学习率和训练轮次，使损失函数收敛至最小值，得到稳定的预测模型。

本研究参考了文献[11]中选择窗口期的方法，以及LSTM网络层数的实验思路，最终搭建了3层LSTM网络模型，并用该模型对10至50范围内不同的时间窗口长度w进行测试。选用均方根误差（RMSE）、平均绝对误差（MAE）和模型预测准确率作为评价模型预测性能的指标：

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

式中，N表示数据量； $\hat{y}_i$ 表示预测值； $y_i$ 表示真实值。

$$RMSE = \sqrt{MSE}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|$$

## 4 实验方案与效果

### 4.1 实验方案

研究所用数据取自大连商品交易所大豆期货主力合约的行情信息，具体划分方式为：2022到2024年的大豆期货数据被用作训练组，供LSTM模型开展训练；2025年上半年的数据则作为效用组，用于对模型结构进行论证。关于LSTM模型的预测流程，详见图3。

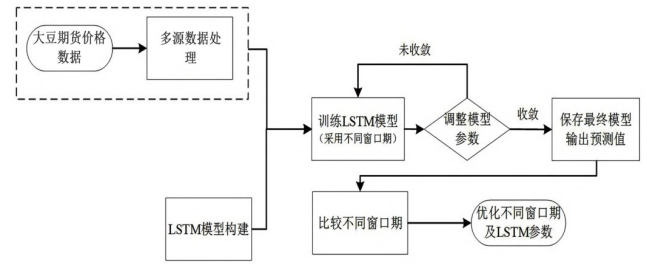


图3 LSTM模型预测流程图

### 4.2 实验结果

#### 4.2.1 2022—2024年大豆期货价格预测实验

构建不同时间窗口10—50长度的模型。取各项指标的平均值作为最终评判标准，通过对比确定最佳时间窗口长度，相关预测结果及指标详见表1：

表1 不同时间窗口LSTM模型预测性能指标均值对比

时间窗口	均方根误差 (RMSE) (元/吨)	平均绝对误差 (MAE) (元/吨)	准确率 (%)
10	68.25	52.36	86.42
15	55.78	41.59	89.76
20	38.62	29.45	94.38
25	42.87	33.12	92.65
30	49.53	38.74	90.21
35	56.89	44.27	88.57
40	63.15	49.83	87.34
45	69.72	55.68	85.89
50	75.36	61.24	84.56

由表1可知，在综合考量下，时间窗口为20时模型性能最为出色，因此最终选定该窗口长度。预测效果最优与最差的情况可参考图4：

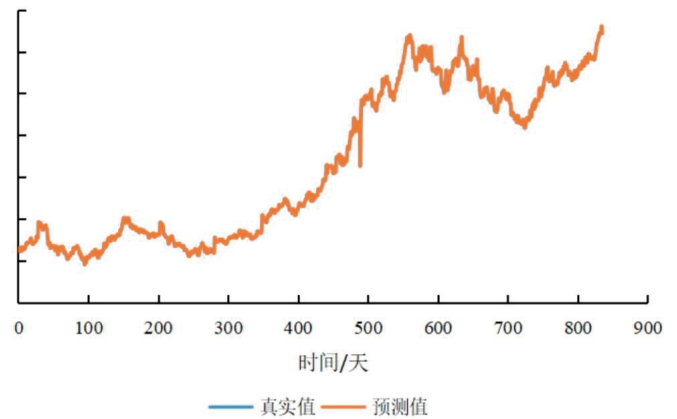


图4 时间窗口期为20的LSTM模型预测结果图

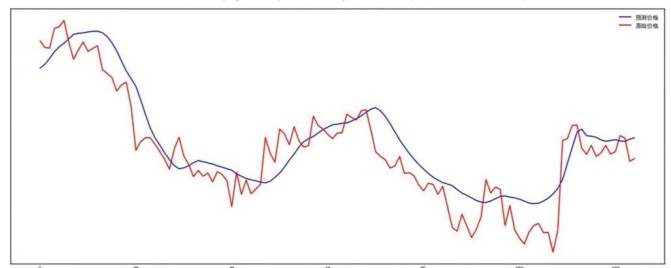


图5 2025年上半年大豆期货价格预测

4.2.2 参数化模型对2025年大豆期货价格预测实验

构建的参数化LSTM模型包含3层网络，时间窗口长度为20，将其应用于2025年大豆期货价格的预测，预测结果如图5所示：

为进一步验证本文所构建LSTM模型在大豆期货价格预测中的优越性，选取BP神经网络、ARIMA时间序列模型、多元线性回归三种传统预测模型作为参照，在相同数据集（2022—2024年大连商品交易所大豆期货主力合约日度收盘价，含国内大豆期货、国际大豆期货、国内豆粕期货三类多源特征）、相同硬件环境下开展对比实验，采用均方根误差（RMSE）、平均绝对误差（MAE）、预测准确率、模型训练耗时四项核心指标对各模型性能进行量化评估，对比结果如表2所示。

表2 LSTM模型与传统预测模型性能对比

预测模型	均方根误差 (RMSE) (元/吨)	平均绝对误差 (MAE) (元/吨)	预测准确率 (%)	模型训练耗时 (s)
LSTM (时间窗口=20)	38.62	29.45	94.38	18.56
BP神经网络	62.39	48.72	87.65	12.34
ARIMA(p=2,d=1,q=1)	70.58	56.39	85.21	8.79
多元线性回归	85.42	68.95	81.76	5.23

由表2可见，本文构建的时间窗口为20的LSTM模型在预

测精度上显著优于三类传统模型：其RMSE、MAE分别为38.62元/吨、29.45元/吨，较BP神经网络分别降低38.10%、39.51%，较ARIMA模型分别降低45.28%、47.87%，较多元线性回归分别降低54.79%、57.28%；同时预测准确率达94.38%，较BP神经网络、ARIMA模型、多元线性回归分别提升6.73、9.17、12.62个百分点。尽管LSTM模型因网络结构复杂度较高，训练耗时略高于传统模型，但其在价格预测精度上的优势足以抵消时间成本，充分验证了该模型在大豆期货价格时序预测任务中的有效性与适用性，能够为大豆期货价格的精准预测提供可靠的技术支撑<sup>[12]</sup>。

结语

综上所述，时间窗口长度为20的LSTM神经网络模型在大豆期货价格预测方面表现更优，且该模型对2025年的大豆期货价格也能起到一定的指示作用。2025年中央一号文件明确提出“多措并举巩固大豆扩种成果”，完善生产者补贴与保险政策，还通过推广饲用豆粕减量替代技术，推动产业优化升级。海关也通过严格检验检疫政策，加强对进口大豆的监管，保障国内市场有序运行。本研究的LSTM模型预测成果能为大豆产业相关企业提供更精准的价格趋势参考。

参考文献

[1]刘会滔.基于二次分解和多源数据融合的股票价格预测研究[D].南京:南京信息工程大学,2024.  
 [2]李辰宇.基于多源数据融合的股票价格预测模型及应用研究[D].济南:山东财经大学,2023.  
 [3]李惠迪.融合多源数据情感分析的金融产品价格预测研究[D].长春:吉林财经大学,2022.  
 [4]康瑞雪.融合多源数据和具有自注意力机制的LSTM股票价格预测[D].太原:太原理工大学,2022.  
 [5]Yan H,Ouyang H.Financial time series prediction based on deep learning [J].Wireless Personal Communications,2018,102(2):683-700.  
 [6]Zhou Dan.Research on prediction of stock price based on neural network[D].Chongqing:Chongqing University of Mathematics and Statistics,2018.  
 [7]Pang X,Zhou Y,Wang P,et al.An innovative neural network approach for stock market prediction [J].The Journal of Supercomputing,2020,76(3):2098-2118.  
 [8]张璇.基于LSTM的河南省花生价格预测研究[D].郑州:华北水利水电大学,2021.  
 [9]郭佳丽,邢双云,栾昊,等.基于改进的LSTM算法的时间序列流量预测[J].南京信息工程大学学报(自然科学版),2021,13(5):571-575.  
 [10]尹力博,杨清元,韩立岩.技术指标能够预测商品期货价格吗?来自中国的证据[J].管理科学学报,2018,21(6):99-109.  
 [11]韩金磊,熊萍萍,孙继红.基于LSTM和灰色模型的股价时间序列预测研究[J].南京信息工程大学学报(自然科学版),2023,15(6):631-642.  
 [12]赵雪伶.基于卷积神经网络的农产品大宗商品价格预测模型研究[D].成都:西南财经大学,2020.