

融合 CNN-BiLSTM-Attention 的集成学习价格预测^①



许珠路^{1,2}, 王兴芬^{2,3}, 刘亚辉³

¹(北京信息科技大学 计算机学院, 北京 100192)

²(北京信息科技大学 商务智能研究所, 北京 100192)

³(北京信息科技大学 信息管理学院, 北京 100192)

通信作者: 王兴芬, E-mail: xfwang@bistu.edu.cn

摘要: 价格预测对于大宗农产品市场的稳定具有重要意义, 但是大宗农产品价格与多种因素有着复杂的相关关系. 针对当前价格预测中对数据完整性依赖性强与单一模型难以全面利用多种数据特征等问题, 提出了一种将基于注意力机制的卷积双向长短期记忆神经网络 (CNN-BiLSTM-Attention)、支持向量机回归 (SVR) 与 LightGBM 组合的增强式集成学习方法, 并分别在包含历史交易、天气、汇率、油价等多种特征数据的数据集上进行了实验. 实验以小麦和棉花价格预测为目标任务, 使用互信息法进行特征选择, 选择误差较低的 CNN-BiLSTM-Attention 模型作为基模型, 与机器学习模型通过线性回归进行增强式集成学习. 实验结果表明该集成学习方法在小麦及棉花数据集上预测结果的均方根误差 (RMSE) 值分别为 12.812, 74.365, 较之 3 个基模型分别降低 11.00%, 0.94%, 4.44%, 1.99% 与 13.03%, 4.39%, 能够有效降低价格预测的误差.

关键词: 集成学习; 双向长短期记忆神经网络 (BiLSTM); 卷积神经网络; 注意力机制; 价格预测; 支持向量机回归 (SVR); LightGBM

引用格式: 许珠路, 王兴芬, 刘亚辉. 融合 CNN-BiLSTM-Attention 的集成学习价格预测. 计算机系统应用, 2023, 32(6): 32-41. <http://www.c-s-a.org.cn/1003-3254/9137.html>

Emsemble Learning Fused with CNN-BiLSTM-Attention for Price Forecasting

XU Zhu-Lu^{1,2}, WANG Xing-Fen^{2,3}, LIU Ya-Hui³

¹(Computer School, Beijing Information Science & Technology University, Beijing 100192, China)

²(Institute of Business Intelligence, Beijing Information Science & Technology University, Beijing 100192, China)

³(School of Information Management, Beijing Information Science & Technology University, Beijing 100192, China)

Abstract: Price forecasting is important for the stability of bulk agricultural commodity markets, but bulk agricultural commodity prices have complex correlations with multiple factors. In order to address the current problems of strong dependence on data integrity and the difficulty of single models to fully utilize multiple data features in price forecasting, a boosting ensemble learning method that combines the attention mechanism-based convolutional bi-directional long short-term memory neural network (CNN-BiLSTM-Attention), support vector regression (SVR), and LightGBM is proposed, and experiments are conducted on the datasets containing historical trades, weather, exchange rate, oil price, and other features data, respectively. The experiment takes the price forecasting of wheat and cotton as the target task, uses the mutual information method for feature selection, selects the CNN-BiLSTM-Attention model with low error as the base model, and performs boosting ensemble learning with the machine learning model through linear regression. The experimental results show that the root mean square error (RMSE) of the ensemble learning method is 12.812 and 74.356

① 基金项目: 国家重点研发计划 (2019YFB1405003)

收稿时间: 2022-12-07; 修改时间: 2023-01-06; 采用时间: 2023-01-19; csa 在线出版时间: 2023-04-07

CNKI 网络首发时间: 2023-04-10

for wheat and cotton datasets, which are 11.00%, 0.94%, 4.44%, 1.99%, 13.03%, and 4.39% lower than the three base models, respectively. The method can effectively improve the accuracy of price forecasting.

Key words: ensemble learning; BiLSTM; convolutional neural networks (CNN); attention mechanism; price forecasting; support vector regression (SVR); LightGBM

价格预测对于大宗商品市场的稳定具有重要意义,商品价格的剧烈波动可能会导致商品市场发生多种风险,尤其是在农产品等领域,对于价格的波动更为敏感,一旦受到影响引发价格波动风险,会对社会造成巨大的损失.因此,准确的预测商品价格可以提前发现大宗农产品市场的变化趋势,为商品生产者以及消费者提供参考,降低损失,稳定商品市场,对其进行研究是十分必要的.大宗农产品市场受到供求关系、金融、政策及其他如品种联动等多种因素的影响,具体包括供求、其他大宗商品、汇率、期货市场、国际形势及天气等因素^[1].由于大宗农产品市场具有受上下游产业链、天气以及金融因素影响较大且相关关系复杂的特点,因此准确预测大宗农产品价格仍然是十分困难的.目前,使用广泛的大宗农产品价格预测方法多使用深度学习模型,虽然可以在一定程度上克服机器学习模型难以利用复杂相关关系的局限性,但是存在依赖数据完整性及过拟合等问题.针对以上问题,本文提出通过集成深度学习模型与机器学习模型以减少模型对于数据的依赖性,从而提升对价格进行预测的鲁棒性,进而准确地预测大宗农产品价格.

1 价格预测相关研究

深度学习因其可更充分地利用数据集中的大量特征,在数据量充足且相关特征较多的情况下预测效果通常较传统机器学习好,目前被很多学者用于农产品价格预测^[2].深度学习使用方法使用多层神经网络如卷积神经网络(convolutional neural networks, CNN),长短期记忆神经网络(long short-term memory neural network, LSTM)等进行预测,如Jaiswal等人^[3]使用长短期记忆神经网络,时延神经网络(time delay neural network, TDNN)和自回归移动平均模型(autoregressive integrated moving average, ARIMA)模型对玉米和棕榈油的国际月价进行预测及模型对比分析,发现LSTM有着比ARIMA和TDNN更小的误差;江知航等人^[4]使用LSTM与BiLSTM模型对2016–2019年的棉花价格数

据进行预测,取得了更低的预测误差;Chuluunsaikhan等人^[5]使用多层感知机(multi-layer perceptron, MLP)、卷积神经网络、LSTM与机器学习模型岭回归(ridge)、随机森林(random forest, RF)、梯度提升(gradient boosting, GB)及统计学模型带额外输入的自回归移动平均模型(autoregressive integrated moving average with extra input, ARIMAX)分别对韩国猪肉价格进行预测,发现深度学习模型LSTM的误差低于其他模型.

同时,以CNN-LSTM模型为代表的基于深度学习的混合模型在价格预测领域逐渐受到关注^[6–10],如Nassar等人^[11]使用CNN-LSTM等模型预测新鲜农产品价格,发现在其上添加注意力机制的CNN-LSTM-Attention模型表现较好,在价格预测的准确性方面优于机器学习模型和简单的深度模型.Lu等人^[12]搭建CNN-BiLSTM-Attention模型预测股票价格并搭建LSTM, CNN-BiLSTM, BiLSTM-Attention等深度学习模型进行比较,发现CNN-BiLSTM-Attention模型较其他模型决定系数(R^2)最大且均方根误差(RMSE)和绝对误差(MAE)最小.如曹超凡等人^[13]构建了MDT-CNN-LSTM模型,对4类股票指数数据使用多向延迟嵌入(MDT)进行数学变换并使用CNN-LSTM进行预测,取得了比传统方法更低的误差.Yu等人^[14]提出了一种相空间重建长短期记忆神经网络(PSR-LSTM)方法,将金融产品价格数据利用时间序列相空间重建(PSR)方法进行重构并使用LSTM进行预测,实验结果表明他们所提出的预测模型方法具有较低的误差.

由于混合深度学习模型仍然具有表现依赖完整时间序列及容易产生过拟合等问题^[15],而集成学习因其可以克服单一模型的这些局限性,有可能比单一模型有着更好的效果^[16].集成学习方法常见的有分解集成学习方法、堆叠式集成学习方法、增强式集成学习方法等.其中分解集成学习方法考虑数据非平稳及非线性对数据进行分解与集成,如Rezaei等人^[17]使用经验模态分解(empirical mode decomposition, EMD)和完全集成经验模态分解(complete ensemble empirical mode

decomposition, CEEMD) 对股票数据进行分解后分别使用 CNN-LSTM 模型进行预测并集成, 实验结果表明混合模型误差较 CNN-LSTM 模型低且 CEEMD-CNN-LSTM 最低. 还有许多学者使用集合经验模态分解 (ensemble empirical mode decomposition, EEMD)、变分模态分解 (variational mode decomposition, VMD) 等其他分解方法构建分解集成模型进行研究^[18,19], 但是, 由于分解集成方法需要分解出多类数据并分别使用多种模型训练及选择, 所以导致运算速度较慢. 因此, 目前主流的集成学习方法仍为易于搭建及运算速度较快的堆叠式集成学习方法, 如 Silva 等人^[20] 对巴西玉米和糖的价格使用 ARIMA、季节性差分自回归滑动平均模型 (seasonal autoregressive integrated moving average, SARIMA)、支持向量回归 (support vector regression, SVR)、AdaBoost 和 LSTM 模型及它们的部分集成进行预测, 发现 LSTM 分别和 SVR 及 ARIMA 的集成模型有着更小的平均绝对误差 (MAE), 均方误差 (MSE) 及更高的 R^2 分数. Parida 等人^[21] 通过基于 CNN 和堆叠自动编码器 (stacked autoencoder, SAE) 的深度学习集成模型对 29 mm 棉和一种农用商品瓜尔胶的收盘价进行预测, 在价格波动较大的价格预测中误差较传统方法小. 如 Xiao 等人^[22] 提出 LSTM-EL 方法, 该方法将数据先通过 LSTM 神经网络, 生成特征嵌入, 然后将它作为 GBRT 和 RF 模型的输入分别进行预测

并集成. 最后将它们与普通的 LSTM, GBRT, RF 模型进行了比较, 发现最终的混合集成模型拥有比单一模型更低的误差. 增强式集成学习方法将误差较大的数据权重调高并再次搭建模型进行迭代, 如 Ribeiro 等人^[23] 使用基于随机森林的套袋式集成方法、梯度增压机 (gradient boosting machine, GBM)、极端梯度增压机 (extreme gradient boosting machine, XGB), 堆叠 (stacking) 以支持向量机回归 (SVR)、多层感知器神经网络 (MLP) 和 K-最近邻 (KNN) 作为参考模型, 发现使用增强式集成学习方法 XGB、GBM 综合 MAPE 及绝对百分比误差 (APE) 的性能最好.

综上所述, 基于混合深度学习模型的增强式集成学习方法由于可以在一定程度上减少对数据完整性的依赖性, 并且更加充分地利用数据特征, 在农产品价格领域有着比传统方法更低的误差. 为了更加准确地预测价格, 为价格波动风险的预警提供支持, 本文结合在深度学习模型中误差较低的 CNN-BiLSTM-Attention 模型与机器学习模型 SVR 和 LightGBM 进行增强式集成, 对不同品类的农产品进行预测与分析.

2 本文方法

本文提出了一种 CNN-BiLSTM-Attention 与两种机器学习方法的增强式集成学习方法, 方法流程图如图 1 所示.

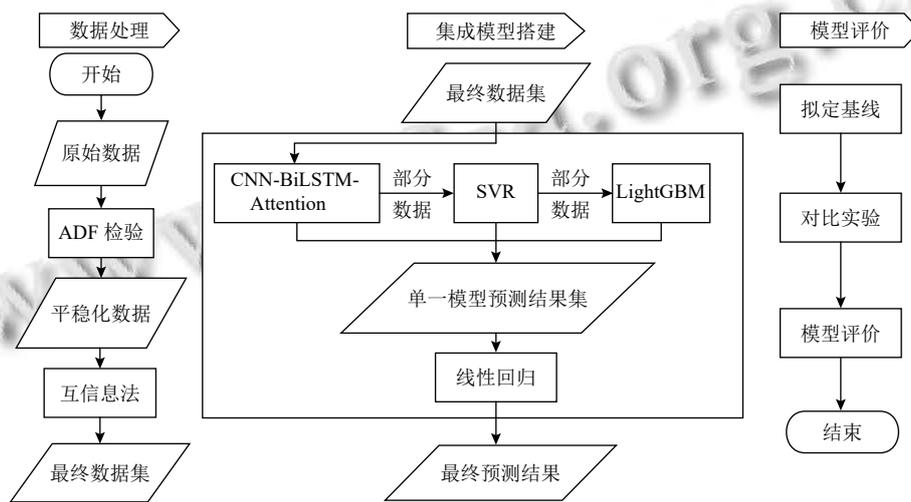


图 1 方法流程图

首先进行数据处理, 包括对数据进行缺失值填充, 时间对齐和线性相关分析去重的数据预处理; 基于互信息法的特征选择. 然后分别训练多种深度学习模型

进行模型选择, 依据模型的 $RMSE$, MAE , R^2 等评价结果选择可以生成时空特征并关注关键特征的 CNN-BiLSTM-Attention 与 SVR, LightGBM 使用增强式集

成学习方法以线性回归为元学习模型进行集成. 最后使用 $RMSE$, MAE , R^2 与单一模型进行比较与评价.

2.1 数据处理方法

大宗农产品受到包括天气, 需求等众多因素的影响, 且不同大宗农产品的影响因素与程度不同. 由于无关变量在数据集中产生了额外的噪声, 会引发神经网络的预测精度下降的问题^[24]. 因此对于收集到包含多种影响因素的特征数据, 需要通过互信息法进行特征选择, 找出影响程度较高的特征, 以提升模型的计算效率与预测精度.

对棉花与小麦的数据使用互信息法计算目标数据与各个特征数据的互信息值以筛选出影响较明显的特征. 一般而言, 两个时间序列的相关性越大, 其互信息值就会越大^[25]. 在信息论中, 互信息的描述如下: 在信道的一端有着信息源 X , 它向信道中发射信息 x , 在另一端的接收端则会收到 x 经过信道影响后转变的 y 或 $f(x)$. 接收端通过 y 推断信息源 X 发射 x 的概率即后验概率 $p(x|y)$, 同时信息源 X 发射 x 的概率称为先验概率 $p(x)$. 定义 $p(x)$ 与 $p(x|y)$ 比值的对数为 y 对 x 的互信息值.

互信息值的计算方法如下.

1) 计算元素长度为 n 的随机变量 X 的熵 $H(X)$:

$$H(X) = - \sum_{i=0}^n p(x = a_i) \ln p(x = a_i) \quad (1)$$

其中, $p(x=a_i)$ 为变量 X 内第 i 个离散点 a_i 的概率密度函数.

2) 计算随机变量 X 和 Y 的联合熵 $H(X, Y)$ 的方法如下:

$$H(X, Y) = - \sum_{i=0}^n p(x = a_i, y = b_i) \ln p(x = a_i, y = b_i) \quad (2)$$

则有 $H(X|Y)=H(X)+H(Y|X)=H(Y)+H(X|Y)$, 可以推导出 $H(X)-H(X|Y)=H(Y)-H(Y|X)$, 而这个差就叫作 X 和 Y 的互信息, 记作 $I(X, Y)$.

3) 可得 $I(X, Y)=H(X)-H(X|Y)=H(X)+H(Y)-H(X, Y)$, 计算互信息 $I(X, Y)$ 的方法如下:

$$I(X, Y) = - \sum_{i=0}^n p(x = a_i, y = b_i) \cdot \ln \frac{p(x = a_i)p(y = b_i)}{p(x = a_i, y = b_i)} \quad (3)$$

最后使用不同阈值筛选去除互信息值过小的与价格变化关系较小特征, 留下互信息值较大的几个与价

格变化相关度较高的特征.

2.2 集成模型搭建方法

本文提出的预测模型方法首先训练多种深度学习模型进行模型选择. 将选择的基模型在训练集上误差大于 MAE 的数据作为训练集训练机器学习模型, 使用相同方式将机器学习模型预测效果较差的数据取出再训练下一个机器学习模型, 然后使用线性回归作为元学习模型对 3 个模型的结果进行集成, 最后在测试集上与单一模型进行评价与比较.

2.2.1 深度学习模型

本文在卷积神经网络、双向长短期记忆神经网络及它们的混合模型如带注意力机制的卷积双向长短期记忆神经网络模型等进行模型选择.

添加注意力机制的卷积双向长短期记忆神经网络模型 (CNN-BiLSTM-Attention) 由卷积神经网络层, 长短期记忆神经网络层以及注意力层组成, 其结构如图 2 所示. 输入的大宗农产品价格相关特征数据先经过卷积层进行卷积操作以提取数据空间特征. 由于卷积后提取的数据特征维度过高, 因此在其后加上池化层以降低特征维度. 然后再通过 BiLSTM 层, 将前面的输出分别输入给一个正向及一个逆向的 LSTM 层, 以获取正向及逆向的时间序列时间特征. 最后将提取的时间与空间特征输入到 Attention 层, 以关注影响输出目标及后一天农产品价格的特征, 提高模型预测的精度.

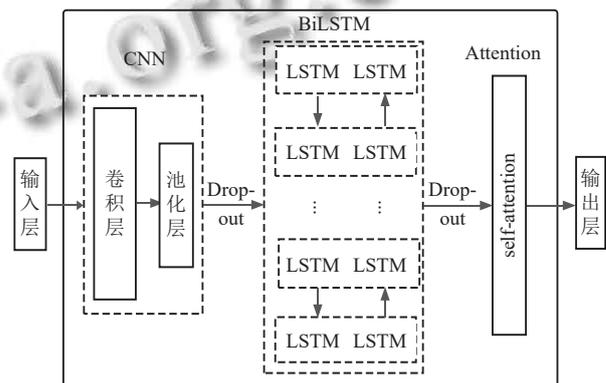


图 2 添加注意力机制的卷积双向长短期记忆神经网络

本文使用 1D-CNN 处理价格时间序列数据, 以有效地提取数据的局部特征. 卷积神经网络 (CNN) 是深度学习中很常用的一种, 多用于图片识别, 自然语言处理等领域. 其根据处理的数据不同又可分为一维卷积 (1D-CNN), 二维卷积 (2D-CNN) 等, 其中 1D-CNN

主要用于序列建模与自然语言处理领域. CNN 一般包括输入层, 卷积层, 池化层, 全连接层和输出层. 它的结构如图 3 所示, 其核心为卷积层, 池化层和全连接层. 由于农产品价格相关特征数据多为时序序列数据, 通过卷积层的可以提取数据在某一方向的平移特征, 在时间序列的表现为提取序列的空间特征, 因此使用一维卷积神经网络可以提取输入的时间序列数据的空间特征.

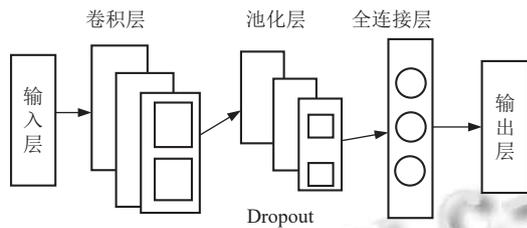


图 3 卷积神经网络

本文使用双向长短期记忆神经网络提取序列时间特征. 双向长短期记忆神经网络 (BiLSTM) 是长短期记忆神经网络 (LSTM) 的一种改进, 堆叠了一个前向传播及一个后向传播的 LSTM 层, 分别重复训练输入的大宗农产品价格相关特征时间序列数据及其倒序后的时间序列数据以提取时间特征. LSTM 单元结构如图 4 所示.

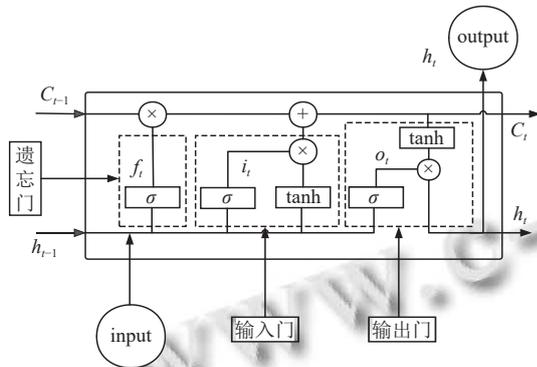


图 4 长短期记忆神经网络单元

其计算过程如下:

$$\text{遗忘门: } f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

$$\text{输入门: } i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (5)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (6)$$

$$\text{输出门: } o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (7)$$

$$\text{长期记忆: } C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (8)$$

$$\text{短期记忆: } h_t = o_t \cdot \tanh(C_t) \quad (9)$$

其中, σ 表示激活函数, X_t 表示输入数据; h_{t-1} 表示前一时刻输出; b_f, b_i, b_C, b_o 分别为遗忘门, 输入门, 细胞状态, 输出门的偏置; W_f, W_i, W_C, W_o 分别表示遗忘门, 输入门, 细胞状态, 输出门的权重参数矩阵. 价格相关特征序列数据输入后, 组成输入组合 $[h_{t-1}, X_t]$. $[h_{t-1}, X_t]$ 输入组合经过遗忘门得到决定保留 C_{t-1} 哪部分特征的概率值 f_t ; $[h_{t-1}, X_t]$ 经过输入门得到的表示现在学到的特征中哪些可以保留的概率值 i_t 与本次学习到的所有特征 \tilde{C}_t ; 将 C_{t-1} 与 f_t 的乘积即上一时刻保留的特征和 \tilde{C}_t 与 f_t 的乘积即这一时刻选择的特征求和得到长期记忆 C_t . $[h_{t-1}, X_t]$ 经过输出门得到表示对长期记忆 C_t 即学习到的特征进行筛选后找到的解决当前问题的信息概率值 o_t , o_t 与 $\tanh(C_t)$ 的结果相乘为短期记忆 h_t 即需要的输出. 输入的时间序列数据经过 BiLSTM 即可提取出大宗农产品数据的时间特征用于预测价格.

由于提取的时间与空间特征数量较多, 本文通过注意力机制对模型进行优化. 注意力机制 (Attention) 通过计算注意力的概率分布, 从大量特征中选择关键的特征, 增加模型计算效率. 其计算分为 3 步.

1) 首先输入特征 h_t 与 Attention 的权重 W_h 相乘再加上 Attention 的偏置 b_h 通过 \tanh 函数计算输出与输入特征的相关性量 s_t . 如式 (10) 所示:

$$s_t = \tanh(h_t \cdot W_h + b_h) \quad (10)$$

2) 再将其归一化并经过 Softmax 函数得到注意力分数量 a_t . 如式 (11) 所示:

$$a_t = \text{Softmax} \left(\exp(s_t) \middle/ \sum \exp(s_t) \right) \quad (11)$$

3) 最后 a_t 与 h_t 相乘以加权相加得到最终输出注意力值向量 s . 如式 (12) 所示:

$$s = \sum a_t h_t \quad (12)$$

训练过程中深度学习模型使用的损失函数为 MSE , 其计算方法如式 (13) 所示:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (13)$$

2.2.2 集成学习方法

由于深度学习模型对数据数量及完整性等依赖性较强, 因此选择机器学习模型 SVR 和 LightGBM 与深度学习模型进行增强式集成, 使用线性回归作为元学习

模型,通过增强式的集成学习方法改善深度学习模型的预测结果。

其中支持向量机回归 (SVR) 是支持向量机 (SVM) 在回归任务上的一个分支,是一种经典的机器学习方法,有着良好的预测性能。它通过将输入的商品价格数据映射进高维特征空间并找到一个与样本点总偏差最小且线性可分的超平面,再将超平面映射回原空间,以实现回归预测。与传统的回归模型不同,它设置一个阈值,只有误差程度超过阈值才计算损失,并通过最大化间隔带的宽度与最小化总损失来优化模型。可以有效地拟合特征及输出的非线性关系,以准确得预测农产品价格。

LightGBM 是微软提出的一个实现梯度提升决策树 (gradient boosting decision tree, GBDT) 算法的增强式集成学习框架,它在 XGBoost 方法的基础上进行优化,通过迭代训练弱分类器以得到最优的模型,且限定只能使用回归树模型作为弱分类器。但是通过使用基于直方图的决策树算法避免过拟合且具有更好的推广性,以及使用带有深度限制的节点展开方法 (leaf-wise) 来提升模型预测精度。拥有更快的计算速度及更好的预测精度,可以准确地预测农产品价格。

将深度学习模型预测结果较差的数据输入到 SVR 模型中,再将 SVR 模型预测结果较差的数据输入到 LightGBM 模型中,通过增强的集成学习方法弥补单一模型难以全面利用数据间复杂相关关系的缺陷,减少预测结果对于数据完整性的依赖性,提升预测价格的精度。

2.2.3 模型评价方法

模型使用 $RMSE$ 、 MAE 及 R^2 进行评价,它们的计算方式如式 (14)–式 (16) 所示:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (14)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (15)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (16)$$

其中, y_i 表示真实值, \hat{y}_i 表示预测值, \bar{y} 表示真实值的平均值, n 为数据条数。 $RMSE$ 常用于表示回归任务模型

预测结果与真实值间的误差,对于较大误差的权重较高,值为 0 到无穷大,越小越好; MAE 表示预测结果与真实值绝对值的误差,误差值的权重一致,值为 0 到无穷大,越小越好; R^2 则是对预测结果与真实值间差距的平方进行归一化,方便比较模型间的差距,值在 0 到 1 之间,越接近 1 表示模型越好。

3 实验

本节首先介绍实验过程中的数据分析与处理过程,包括数据预处理,特征选择等;然后介绍集成模型的构建,分别为基模型搭建以及模型集成,最后对实验的结果进行分析。

3.1 数据分析与处理

由于取得的棉花与小麦的数据来源不同,所以需要先对其分别进行数据预处理,经过时间对齐,缺失值填充等操作后,将目标数据差分后,再对数据使用互信息法进行特征处理,最后划分得到模型训练及验证使用的数据。

3.1.1 数据来源及数据预处理

根据文献 [1] 中对大宗农产品市场影响因素的研究,小麦及棉花市场受到相关商品如上下游大宗商品,金融因素如能源、期货市场、汇率等以及主要产地天气因素的影响,因此从 Kaggle、万德数据、RP5 地方天气数据、英伟财经等来源收集需要的数据。其中小麦价格数据及其相关商品数据来源于 Kaggle,按日期对齐后共 2 339 条,数据特征及说明如表 1 所示。棉花价格数据来源于万德数据,按日期对齐后共 2 883 条,天气数据来源于 RP5 地方天气数据,油价数据以及汇率数据来源于英伟财经数据,数据特征及说明如表 2 所示。

表 1 小麦数据特征

序号	特征类别	说明
1	相关农产品	美国玉米和大豆的开盘价,闭盘价,最高价,最低价,交易量
2	能源	美国布伦特石油、WTI原油和天然气的开盘价,闭盘价,最高价,最低价,交易量
3	金属	美国黄金、钯和镍的开盘价,闭盘价,最高价,最低价,交易量
4	经济	美元-人民币、美国历史通货膨胀率、海洋尼诺指数
5	小麦	美国小麦及其期货的开盘价,闭盘价,最高价,最低价,交易量;全球小麦产量、收获面积

表2 棉花数据特征

序号	特征类别	说明
1	时间与天气	交易年份、交易月份、山东温度、大气压、湿度、阵风风速、能见度、降雨量
2	相关商品	山东B籽棉、绞纱、毛籽、棉粕、全棉纱价格, 新疆棉花价格, 全国棉花均价, 中国纱线价格指数
3	能源与经济	XBR、WTI原油与美元-人民币的开盘价, 收盘价, 最高价, 最低价, 交易量
4	山东棉花	当天棉花现货价, 当月棉花进口数量, 当月棉花进口美元值及其增量

对于取得的小麦价格数据、相关商品数据以及油价数据等, 先经过如下步骤。

1) 时间对齐, 将时间跨度不同的数据修改时间跨度为天, 并取有交集的数据。

2) 缺失值填充, 对存在缺失值的小麦与棉花价格数据使用 KNN 方法进行填充, 对其他数据进行前向填充。

3) 对处理过后的商品价格数据进行滞后, 使其后一天的价格作为目标数据, 再加入前两天的价格作为新特征。

4) 之后对特征数据进行相关性分析, 去除线性相关较高的特征, 或通过差分以及作差等运算将其转换为新特征, 保留线性相关性较小的非重复特征。

5) 对特征数据与目标数据使用互信息法计算互信息值进行特征选择, 对不同模型调整阈值选出合适的特征, 去除无关特征。

6) 最后将得到的数据按 7:3 划分为训练集及测试集。

经过预处理的数据即可分别供多种深度学习及机器学习的基模型进行训练, 以选出性能较好且可互补的模型使用线性回归进行集成, 分配权重, 生成最终的集成模型。

3.1.2 基于互信息法的特征选择

通过对数据采用互信息法通过调整不同阈值, 以找出与目标相关度较低的几项特征进行去除, 使得更重要的特征可以被保留下来, 达到加快训练速度, 提高预测精度的目的。将拆分为特征及目标的数据, 输入到互信息函数中, 生成各特征与目标数据的互信息值, 然后设置不同阈值, 通过保留互信息值超出阈值的特征以进行特征选择。表3和表4分别为小麦与棉花互信息值阈值为0.01的特征选择结果说明。

通过表3和表4可以发现, 小麦价格波动受到相关商品、布伦特原油及通货膨胀率的影响较大, 而棉

花价格波动受到天气, 商品价格波动, 纱线价格, 交易年份的影响较大。

表3 小麦保留特征

序号	互信息值	说明
1	0.01-0.02	黄金交易量, 布伦特石油开盘价, 布伦特石油交易量, 天然气开盘价, 天然气最低价, 天然气交易量, 交易日, 玉米价格, 大豆价格, 布伦特原油开盘价
2	0.02-0.03	小麦交易量, 黄金开盘价, 布伦特石油收盘价, 交易年份, 小麦交易价格, 小麦期货交易量, WTI原油收盘价, WTI原油价格, 小麦期货价格, 美国历史通货膨胀率, 小麦价格
3	0.03-0.04	小麦期货收盘价, 美国历史通货膨胀率, 小麦最高最低价差, 今日镍开盘价增量
4	>0.04	小麦今日开盘价增量

表4 棉花保留特征

序号	互信息值	说明
1	0.01-0.02	3128B籽棉、2227棉花、21S绞纱、毛籽、棉粕、21s/2全棉纱价格, 新疆棉花现货价格, 全国棉花平均价格, 大气压, 交易月份
2	0.02-0.03	山东棉花现货价格, 降雨量, 交易日, 山东32S全棉纱市场价, 3128B棉花价格, 美元-人民币收盘价, 棉花进口量
	0.03-0.04	WTI原油收盘价
3	>0.04	山东棉花现货价增量, 交易年份, 中国纱线价格指数, 阵风风速

3.2 模型构建及调参

3.2.1 基模型搭建及调参

使用贝叶斯调参方法重复训练以调整模型超参数, 调参过程中, 基础层包括输入层、中间层及输出层各一层, 同时混合模型每种神经网络层各一层, 调整的深度学习模型参数为神经网络隐藏层数 0-3, 神经网络细胞数 32-256, dropout 值 0.0-0.5, 迭代次数 epochs 5-50, 块大小 batch_size 30-200, 学习率 learn_rates 0.001-0.01; 机器学习模型中调整的 SVR 参数为 3 种核函数, 惩罚系数 C 0.0001-1.0; LightGBM 参数为学习率 learning_rate 0.001-0.01, 树模型的最大深度 3-7, subsample 0.5-1.0. 使用相同的训练集、测试集、时间步长、特征与相同的参数范围通过贝叶斯调参方法进行训练及调参, 最终获得的深度学习模型在性能上是有可比性的。

最后调整出的深度学习模型参数如表5所示。

对于训练出的深度学习模型, 使用测试集进行评价, 综合考虑 R^2 等评价指标以选择合适的基模型, 分

别训练 CNN、LSTM、BiLSTM、CNN-BiLSTM、BiLSTM-Attention、CNN-BiLSTM-Attention 等深度学习模型. 在时间序列模型中棉花采用 15 天的历史数据预测后一天的价格, 小麦采用 10 天的历史数据预测后一天的价格. 训练后使用测试集进行评价, 各深度学习模型评价结果如表 6 所示.

从表 6 可以发现 3 种混合深度学习模型 BiLSTM-

ATT、CNN-BiLSTM 和 CNN-BiLSTM-Attention 在小麦及棉花价格预测上的预测结果准确性方面均优于传统的深度学习模型 LSTM、BiLSTM 及 CNN, R^2 较大; 而混合深度学习模型中 CNN-BiLSTM-Attention 的预测结果 R^2 最小, 因此在混合深度学习模型中选择 CNN-BiLSTM-Attention 作为基模型, 与 SVR 和 LightGBM 进行集成.

表 5 深度学习模型参数

数据	模型	隐藏层数	细胞数	Dropout	迭代次数	块大小	学习率
小麦	LSTM	3	69	0.4	30	113	0.0326
小麦	BiLSTM	3	124	0.5	30	50	0.001
小麦	CNN	3	154	0.06	36	146	0.01
小麦	BiLSTM-ATT	2	114	0.28	43	131	0.0054
小麦	CNN-BiLSTM	2	100	0.2	49	44	0.0079
小麦	CNN-BiLSTM-Attention	1	162	0.41	49	135	0.0021
棉花	LSTM	2	90	0.15	23	129	0.0018
棉花	BiLSTM	2	144	0.01	15	45	0.0070
棉花	CNN	3	112	0.2	30	50	0.01
棉花	BiLSTM-ATT	1	200	0.35	40	42	0.0890
棉花	CNN-BiLSTM	2	131	0.41	22	135	0.0097
棉花	CNN-BiLSTM-Attention	1	230	0.11	30	38	0.0027

表 6 深度学习模型结果

数据	模型	RMSE	MAE	R^2
小麦	LSTM	14.488	8.230	0.928558
小麦	BiLSTM	14.523	8.287	0.927230
小麦	CNN	14.476	8.255	0.928435
小麦	BiLSTM-ATT	14.386	8.122	0.929073
小麦	CNN-BiLSTM	14.394	8.136	0.929840
小麦	CNN-BiLSTM-Attention	14.396	8.130	0.938913
棉花	LSTM	76.518	47.543	0.997530
棉花	BiLSTM	76.619	47.719	0.997269
棉花	CNN	76.697	47.064	0.997353
棉花	BiLSTM-ATT	75.667	47.524	0.997584
棉花	CNN-BiLSTM	76.266	48.232	0.997477
棉花	CNN-BiLSTM-Attention	75.068	48.137	0.997578

3.2.2 模型的集成

将 CNN-BiLSTM-Attention 模型在训练集上误差大于 MAE 的数据作为训练集训练 SVR 模型, 再将 SVR 模型在其训练集是预测结果误差大于 MAE 的数据取出训练 LightGBM 模型, 然后使用线性回归作为元学习模型对 3 个模型的结果进行集成, 最后在测试集上进行评价, 与单一模型进行比较.

4 结果及分析

本实验各单一模型及集成模型评价结果如表 7 所示.

通过表 5 和表 6 的预测结果可以发现机器学习模型一般较深度学习模型误差更大, 而混合模型的集成模型预测结果误差比传统深度学习模型小. 同时也可以分析 CNN-BiLSTM-Attention 与 SVR、LightGBM 进行集成后可以有效地改善预测结果, 能够有效集合各自的优点, 克服单一模型的局限性, 有着比单一的机器学习及深度学习模型更低的 RMSE 及更高的 R^2 . 同时因为是与机器学习模型进行集成, 计算速度较之单一的混合深度学习模型并没有明显下降.

表 7 集成模型结果

数据	模型	RMSE	MAE	R^2
小麦	LightGBM	14.732	8.742	0.936025
小麦	SVR	13.407	8.326	0.947014
小麦	CNN-BiLSTM-Attention	14.396	8.130	0.938913
小麦	Ensemble	12.812	7.807	0.949747
棉花	LightGBM	77.759	49.396	0.993115
棉花	SVR	75.873	46.723	0.981320
棉花	CNN-BiLSTM-Attention	75.068	48.137	0.997578
棉花	Ensemble	74.365	47.341	0.998025

最终集成模型对小麦价格部分预测结果如图 5 所示; 最终集成模型对棉花价格部分预测结果如图 6 所示.

图 5 和图 6 分别显示了最终模型的预测结果, 可

以发现模型的预测值与真实值相差较小, 可认为该集成模型在农产品价格预测上有着较低的预测误差。

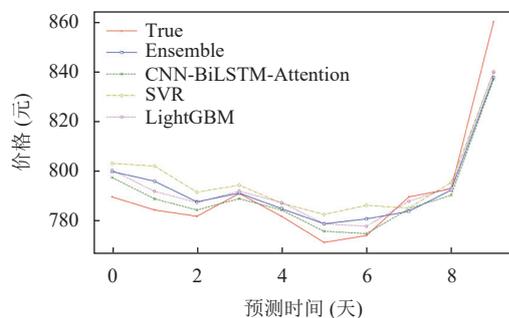


图5 小麦价格预测结果

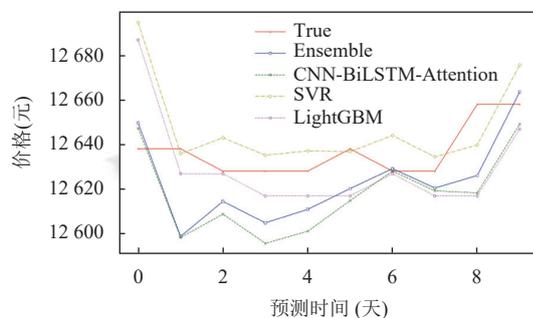


图6 棉花价格预测结果

通过表6、表7、图5、图6的结果可以发现, 本文提出的模型在性能上优于其他几种模型。相对于单一的深度学习模型, 基于CNN-BiLSTM-Attention的集成学习模型对于数据完整性的依赖较小, 即使有部分数据存在大量缺失, 导致深度学习模型在这部分数据中效果较差, 也可以通过与其他机器学习模型集成减少误差; 同时, 通过对误差较大的数据进行增强式集成, 可以对深度学习模型预测结果中过拟合的部分进行改善, 因此, 提出的集成学习模型可以在原有深度学习混合模型基础上, 提升价格预测的准确性。

5 结束语

本文针对价格预测问题提出了一种基于CNN-BiLSTM-Attention的集成学习方法, 旨在通过深度学习模型与机器学习模型进行集成, 改善深度学习模型的预测结果。首先对棉花及小麦的价格数据使用各种深度学习模型进行建模, 最后选出了误差最低的CNN-BiLSTM-Attention作为基模型, 将其与SVR、LightGBM模型使用线性回归进行增强式集成。最终的实验结果

表明通过与机器学习模型进行增强式集成可以有效改善所选混合深度学习模型的预测结果, 所生成的集成学习模型有着比单一模型更好的预测精度。

参考文献

- 1 孙玉奎, 高苗苗. 大宗商品价格影响因素及分类型比较研究. 价格理论与实践, 2018, (9): 86–89, 102. [doi: 10.19851/j.cnki.cn11-1010/f.2018.09.022]
- 2 Sun Y, Guo J, Shan S, *et al.* Wheat futures prices prediction in China: A hybrid approach. *Discrete Dynamics in Nature and Society*, 2021, 2021: 5545802. [doi: 10.1155/2021/5545802]
- 3 Jaiswal R, Jha GK, Kumar RR, *et al.* Deep long short-term memory based model for agricultural price forecasting. *Neural Computing and Applications*, 2022, 34(6): 4661–4676. [doi: 10.1007/s00521-021-06621-3]
- 4 江知航, 王艳霞, 颜家均, 等. 基于BiLSTM的棉花价格预测建模与分析. *中国农机化学报*, 2021, 42(8): 151–160. [doi: 10.13733/j.jcam.issn.2095-5553.2021.08.21]
- 5 Chuluunsai Khan T, Ryu GA, Yoo KH, *et al.* Incorporating deep learning and news topic modeling for forecasting pork prices: The case of South Korea. *Agriculture*, 2020, 10(11): 513. [doi: 10.3390/agriculture10110513]
- 6 Cai RC, Zhu BJ, Ji L, *et al.* An CNN-LSTM attention approach to understanding user query intent from online health communities. *Proceedings of the 2017 IEEE International Conference on Data Mining Workshops (ICDMW)*. New Orleans: IEEE, 2017. 430–437. [doi: 10.1109/ICDMW.2017.62]
- 7 Kim TY, Cho SB. Particle swarm optimization-based CNN-LSTM networks for forecasting energy consumption. *Proceedings of the 2019 IEEE Congress on Evolutionary Computation (CEC)*. Wellington: IEEE, 2019. 1510–1516. [doi: 10.1109/CEC.2019.8789968]
- 8 Kim TY, Cho SB. Predicting the household power consumption using CNN-LSTM hybrid networks. *Proceedings of the 19th International Conference on Intelligent Data Engineering and Automated Learning*. Madrid: Springer, 2018. 481–490. [doi: 10.1007/978-3-030-03493-1_50]
- 9 Hasan MN, Toma RN, Nahid AA, *et al.* Electricity theft detection in smart grid systems: A CNN-LSTM based approach. *Energies*, 2019, 12(17): 3310. [doi: 10.3390/en12173310]
- 10 Zheng L, Xue WH, Chen F, *et al.* A fault prediction of equipment based on CNN-LSTM network. *Proceedings of*

- the 2019 IEEE International Conference on Energy Internet (ICEI). Nanjing: IEEE, 2019. 537–541. [doi: [10.1109/ICEI.2019.00101](https://doi.org/10.1109/ICEI.2019.00101)]
- 11 Nassar L, Okwuchi IE, Saad M, *et al.* Deep learning based approach for fresh produce market price prediction. Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN). Glasgow: IEEE, 2020. 1–7. [doi: [10.1109/IJCNN48605.2020.9207537](https://doi.org/10.1109/IJCNN48605.2020.9207537)]
- 12 Lu WJ, Li JZ, Wang JY, *et al.* A CNN-BiLSTM-AM method for stock price prediction. Neural Computing and Applications, 2021, 33(10): 4741–4753. [doi: [10.1007/s00521-020-05532-z](https://doi.org/10.1007/s00521-020-05532-z)]
- 13 曹超凡, 罗泽南, 谢佳鑫, 等. MDT-CNN-LSTM 模型的股价预测研究. 计算机工程与应用, 2022, 58(5): 280–286. [doi: [10.3778/j.issn.1002-8331.2108-0104](https://doi.org/10.3778/j.issn.1002-8331.2108-0104)]
- 14 Yu PF, Yan XS. Stock price prediction based on deep neural networks. Neural Computing and Applications, 2020, 32(6): 1609–1628. [doi: [10.1007/s00521-019-04212-x](https://doi.org/10.1007/s00521-019-04212-x)]
- 15 Semwal VB, Gupta A, Lalwani P. An optimized hybrid deep learning model using ensemble learning approach for human walking activities recognition. The Journal of Supercomputing, 2021, 77(11): 12256–12279. [doi: [10.1007/s11227-021-03768-7](https://doi.org/10.1007/s11227-021-03768-7)]
- 16 Li Y, Pan Y. A novel ensemble deep learning model for stock prediction based on stock prices and news. International Journal of Data Science and Analytics, 2022, 13(2): 139–149. [doi: [10.1007/s41060-021-00279-9](https://doi.org/10.1007/s41060-021-00279-9)]
- 17 Rezaei H, Faaljou H, Mansourfar G. Stock price prediction using deep learning and frequency decomposition. Expert Systems with Applications, 2021, 169: 114332. [doi: [10.1016/j.eswa.2020.114332](https://doi.org/10.1016/j.eswa.2020.114332)]
- 18 方雪清, 吴春胤, 俞守华, 等. 基于 EEMD-LSTM 的农产品价格短期预测模型研究. 中国管理科学, 2021, 29(11): 68–77. [doi: [10.16381/j.cnki.issn1003-207x.2019.0765](https://doi.org/10.16381/j.cnki.issn1003-207x.2019.0765)]
- 19 Wang JJ, Chen Y, Zhu SZ, *et al.* Depth feature extraction-based deep ensemble learning framework for high frequency futures price forecasting. Digital Signal Processing, 2022, 127: 103567. [doi: [10.1016/j.dsp.2022.103567](https://doi.org/10.1016/j.dsp.2022.103567)]
- 20 Silva RF, Barreira BL, Cugnasca CE. Prediction of corn and sugar prices using machine learning, econometrics, and ensemble models. Engineering Proceedings, 2021, 9(1): 31. [doi: [10.3390/engproc2021009031](https://doi.org/10.3390/engproc2021009031)]
- 21 Parida N, Mishra D, Das K, *et al.* On deep ensemble CNN-SAE based novel agro-market price forecasting. Evolutionary Intelligence, 2021, 14(2): 851–862. [doi: [10.1007/s12065-020-00466-w](https://doi.org/10.1007/s12065-020-00466-w)]
- 22 Xiao W, Xu C, Liu HL, *et al.* A hybrid LSTM-based ensemble learning approach for China coastal bulk coal freight index prediction. Journal of Advanced Transportation, 2021, 2021: 5573650. [doi: [10.1155/2021/5573650](https://doi.org/10.1155/2021/5573650)]
- 23 Ribeiro MHD, dos Santos Coelho L. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. Applied Soft Computing, 2020, 86: 105837. [doi: [10.1016/j.asoc.2019.105837](https://doi.org/10.1016/j.asoc.2019.105837)]
- 24 Li DF, Li ZR, Sun K. Development of a novel soft sensor with long short-term memory network and normalized mutual information feature selection. Mathematical Problems in Engineering, 2020, 2020: 7617010. [doi: [10.1155/2020/7617010](https://doi.org/10.1155/2020/7617010)]
- 25 Li CW, Zhan LW, Shen LQ. Friction signal denoising using complete ensemble EMD with adaptive noise and mutual information. Entropy, 2015, 17(9): 5965–5979. [doi: [10.3390/e17095965](https://doi.org/10.3390/e17095965)]

(校对责编: 牛欣悦)