



中国管理科学
Chinese Journal of Management Science
ISSN 1003-207X, CN 11-2835/G3

《中国管理科学》网络首发论文

题目：基于注意力机制和 CNN-LSTM 模型的沪铜期货高频价格预测
作者：景楠，史紫荆，舒毓民
DOI：10.16381/j.cnki.issn1003-207x.2020.0342
网络首发日期：2020-08-13
引用格式：景楠，史紫荆，舒毓民. 基于注意力机制和 CNN-LSTM 模型的沪铜期货高频价格预测. 中国管理科学.
<https://doi.org/10.16381/j.cnki.issn1003-207x.2020.0342>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

DOI: 10.16381/j.cnki.issn1003-207x.2020.0342

基于注意力机制和 CNN-LSTM 模型的沪铜期货高频价格预测

景楠¹, 史紫荆¹, 舒毓民^{2*}

(1. 上海大学悉尼工商学院, 上海, 201800;

2. 哈尔滨工业大学(深圳)经济管理学院, 深圳, 518055)

摘要: 针对铜期货价格序列的复杂性和长期依赖性, 提出了一种结合卷积神经网络(CNN, Convolutional Neural Network)和长短期记忆(LSTM, Long-Short Term Memory)神经网络的期货价格预测模型(CNN-LSTM), 并在 LSTM 端引入注意力机制进行优化。模型结合了 CNN 时间扩展和 LSTM 长期记忆的优势。首先利用 CNN 模块从价格时间序列中提取不同权重的特征; 随后将特征作为 LSTM 模块的输入对未来价格进行预测, 通过 LSTM 神经网络的长期依赖特性来提高期货价格的预测精度; 最后在 LSTM 端引入注意力机制, 使模型集中学习更重要的数据特征, 进一步提高预测性能。实验选取上海期货交易所近三年的沪铜期货 5 分钟高频价格数据, 采用集成学习领域中广泛使用的 XGBoost 算法(eXtreme Gradient Boosting)评估和筛选预测指标, 并构建了引入注意力机制的 CNN-LSTM 预测模型, 对后一时点的收盘价进行预测。结果表明, 相比于 CNN、LSTM、CNN-LSTM 等基准模型, 引入注意力机制的 CNN-LSTM 混合模型提高了预测的准确性, 为后续开发相应的量化投资策略奠定了基础。

关键词: 注意力机制; LSTM; CNN; CNN-LSTM; 沪铜期货

1 引言

铜是一种工业生产的重要原材料, 关乎制造业乃至整个国民经济的发展。随着我国铜消费量的高速增长, 铜期货市场发展深受关注。作为上海期货交易所中最为活跃的品种之一, 沪铜期货合约以活跃的日内波动受到了诸多日内交易者的青睐。考虑到我国铜期货市场投机过度、高换手率的现状, 对高频沪铜期货数据进行量化建模, 对投资者有重要的实践意义。

自有效市场假说提出以来, 许多学者基于实证研究, 发现了与之违背的结论^{[1][2]}。季俊伟等学者^[3]使用时变状态空间模型度量沪铜期货市场, 证明了我国沪铜期货市场的无效性, 即沪铜期货的运动趋势和价格变化在一定程度上是可以预测的。目前, 对我国期货市场的研究主要基于低频数据, 所使用的指标多以日度或月度数据为主, 对于高频交易的参考价值极为有限。随着技术的

发展, 高频数据的获取变得可行和容易, 而技术指标能够高频、实时预测期货价格变化^[4]。因此, 基于技术指标对期货价格预测是有效而可行的。

由于我国商品期货市场起步较晚, 现有的铜等有色金属期货价格预测的文献相对有限, 但已经有许多学者研究出了大量金融时间序列分析模型, 这些模型概括起来可以分为两类: 数理统计方法和数据挖掘方法。数理统计方法最早应用于金融时间序列分析, 包括自回归模型 (AR)^[5]、移动平均模型 (MA)^[6]、自回归移动平均模型 (ARMA)^[7]、自回归条件异方差模型 (GARCH)^{[8][9]}等。然而, 铜期货价格的影响因素复杂多样, 数据呈现非线性、时变性等特征, 传统的统计方法很难揭示其内在的规律。随着数据挖掘技术和深度学习理论不断发展, 决策树^{[10][11]}、遗传算法^[12]、支持向量机^{[13][14]}、逻辑回归^[15]等机器学习算法都被应用于金融市场价格预测研究中。罗洪奔^[16]结合灰色预测和整合移动平均自回归模型 (ARIMA) 模型预测金融时间序列, 并采用遗传算法对 ARIMA 的系数进行优化。林杰和龚正^[17]采用 BP 神经网络和卷积神经网络 (CNN) 对沪铜期货价格涨跌进行预测, 两种方法均取得了较高的

作者简介: 舒毓民(1974-), 男, 湖北人, 哈尔滨工业大学博士研究生, 主要研究方向为管理科学。

准确率。近年来,长短期记忆 (LSTM)神经网络在时序预测任务中表现突出。Jin 等学者^[18]采用经验模态分解(EMD)和 LSTM 对股票价格进行预测,并引入了投资者情绪因子,进一步提高了预测精度。Kim 和 Chang^[19]将 LSTM 模型与各种 GARCH 模型相结合,提出了一种新的混合 LSTM 模型预测股价波动性。

随着人们对深度学习模型应用于金融市场上的研究不断深入,深度学习模型也被证实有一定的局限性,如 LSTM 在新的时间状态下不断叠加输入序列会导致前面的输入信息变得越来越模糊,即存在长程记忆能力有限问题。CNN 虽然在一定程度上可以缓解该问题,但由于滤波器一般不会选择太大,所以并不能很好地解决。使用注意力机制优化神经网络成为神经网络研究的一个热点。注意力机制最初应用在计算机视觉领域,通过对模型中不同关注部分赋予不同的权重,抽取出来更加重要和关键的信息,从而优化模型。Google Deep Mind 团队在 2014 年提出将注意力机制用在图像分类任务中^[20],验证了注意力机制在图像处理领域的有效性,同时也使结合注意力机制的神经网络成为研究的热点。随后,Banhdanau 等学者^[21]将注意力机制和循环神经网络 (RNN) 结合解决机器翻译任务,使注意力机制成功融入自然语言处理领域。

注意力机制根据每一时间步所关注的区域是一个区域还是所有的区域,可以分为硬注意力与软注意力。硬注意力机制采用直接限制输入内容的处理方法,在某一时刻只关注一个位置信息,这在期货预测领域并不适用。即使输入序列的重要程度存在差异,但由于各输入子序列均含有一定信息,这些信息存在于序列的不同位置,无法明确定位以去除。相较而言,软注意力机制以神经网络训练出的权值,在空间或通道上加权全局输入特征,从而达到关注特定空间区域或通道的目的。

2015 年, Luong 等学者^[22]在此基础上进一步提出针对上述两种注意力机制的改进版本,即全局注意力和局部注意力。

在金融时间序列预测中,由于序列中特征重要程度存在差异,重要的显著特征往往会包含更多信息量,对未来趋势影响程度更大。如果赋予 LSTM 更关注高重要度特征的能力,可以更好实现短期模式的有效提取,避免长期信息丢失问题。因此,本文提出一种将卷积神经网络 (CNN)和长短期记忆网络 (LSTM)相结合的期货价格预测模型(CNN-LSTM),并将注意力机制引入 LSTM 端,为 LSTM 所有时间步的隐藏状态训练权重,并进行加权求和。在实证研究中,建立完善的预测指标体系,利用近三年的上海期货交易所沪铜期货交易数据进行预测,并与传统的 CNN、LSTM 模型预测结果进行了对比,以验证模型的可行性和有效性。本文的创新点和主要贡献主要包括以下三点:

(1). CNN 适合提取数据的局部特征,并组合抽象成高层特征。LSTM 更适合进行时间扩展,具有长期记忆功能。在期货预测任务中,不仅需要考虑不同特征之间的空间联系,还需要考虑数据在时间维度上的变化,因此本文将两种模型相结合,使模型更适用于期货价格分析和预测。

(2). 本文将注意力机制应用于金融时序预测领域,可以全面获取时序信息,学习局部价格序列特征的重要程度。最终,通过提升模型中重要时序特征的影响力,降低最终模型中非重要特征的影响力,可以有效提升模型预测精度。

(3). 期货预测指标体系需建立在科学的基础上。本文基于 XGBoost(eXtreme Gradient Boosting)算法^{[23][24]}对影响价格变化的预测指标进行评估和筛选。XGBoost 算法能够大规模并行运行梯度增强决策树,在每个决策树构建的过程中,每一层贪心地选取一个指标分割点作为叶子节点,使在分割之后整棵树增益值最大。这意味着指标每次被分割时的平均增益越大,该指标越重要。因此,每个指标的平均增益可以量化指标重要度,为后续选择适用于沪铜期货的量化策略奠定基础^[25]。

2 模型构建

CNN 适合作用于空间上的抽象及泛化，LSTM 适用于扩展时间特征，处理具有先后顺序特征的数据。而构建沪铜期货价格预测模型，既需要考虑其空间上的特征联系，也要考虑时间维度上关联信息。因此本文结合两种算法方法的优点，使用串联的方式提出了 CNN-LSTM 算法，并将在计算机视觉、机器翻译领域中广泛采用的注意力机制引入金融时序预测任务中，以提升 CNN-LSTM 算法的预测精度。

本文构建的模型由 CNN 网络、LSTM 网络和注意力机制三部分构成。首先，在 CNN 网络中，

由一维卷积核对金融时间序列进行卷积操作，以提取期货数据在空间结构上的特征分量。随后，LSTM 网络基于提取出来的特征分量做序列预测。由于经过 CNN 网络提取出的特征仍具有时序特性，所以使用 LSTM 进行可以直接有效地对其建模。然后，注意力模块通过神经网络来自动化拟合权重分配，并对 LSTM 不同时间点的隐含层输出向量与对应的权重相乘并求和，为重要的特征分量赋予更大的权重，以此作为模型最终的特征表达。最终，注意力模块支路覆盖了每个神经元的输出，使模型获取更全面、更细致的特征信息。模型整体结构如图 1 所示。CNN 网络、LSTM 网络和注意力机制的原理将在下述小节中具体介绍。

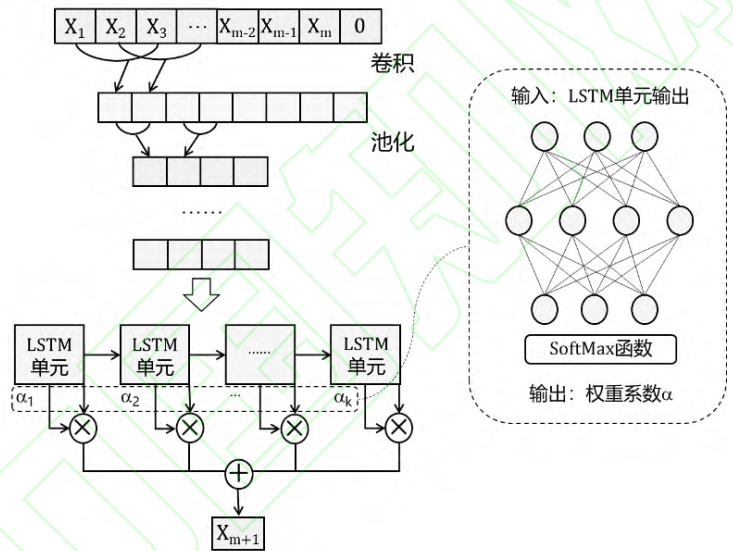


图 1 引入注意力机制的 CNN-LSTM 模型原理示意图

2.1 CNN 网络

CNN 最先在计算机视觉、自然语言处理领域取得了出色的表现，近年来，有学者尝试对其网络结构进行改进，用于时间序列预测^{[26][26]}。CNN 的本质在于构建多个能够提取数据特征的滤波器，通过对输入数据进行逐层卷积和池化操作来提取原

始数据的深层特征表示。考虑到铜期货数据为多元变量，且不同特征之间具有耦合性，本文采用的 CNN 网络中卷积核为一维结构，如图 2 所示。相比于直接使用 K 线图作为模型输入，本文将多元时间序列通过不同的通道输入，可以最大程度保留信息。CNN 主要由 3 个模块构成：卷积层、池化层和全连接层。

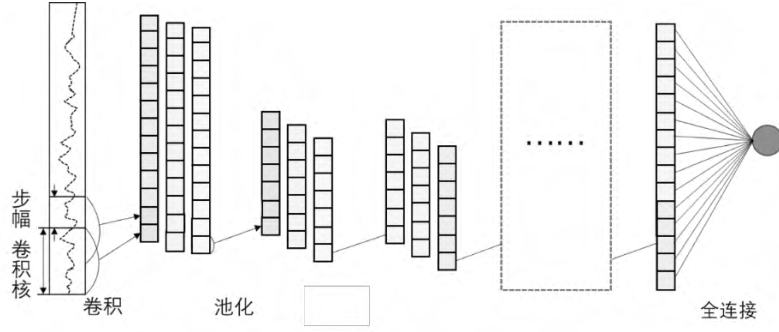


图 2 一维卷积神经网络结构

CNN 的输入是由连续的 M 个时间步上的特征向量组成的。设 x_t 表示第 t 个时间步的特征向量，则本节构造的序列形式为 $X_i = [x_{(t-M+1)}, x_{(t-M+2)}, \dots, x_t]^T$ ，即 $X_1 = [x_1, x_2, \dots, x_M]^T$ ， $X_2 = [x_2, x_3, \dots, x_{(M+1)}]^T$ ，依次可推出其它序列。

在卷积层中，卷积核可看成时间序列上的滑动窗口，提取的是序列间的短时特征。假设 c_j^l 是第 l 层卷积层的第 j 个特征图，其运算公式为：

$$c_j^l = f(w_{ij}^{(l)} \otimes \sum_{i=1}^k c_i^{l-1} + b^l) \quad (1)$$

其中， $w_{ij}^{(l)}$ 表示连接第 $l-1$ 层和第 l 层卷积层的权重矩阵， b^l 为第 l 层对应的偏置值，在同一层的所有区域内共享权值。 $f(\cdot)$ 是激活函数，运算符 \otimes 表示卷积。 c_i^{l-1} 表示第 $l-1$ 层的第 i 个特征图。

多个窗口的特征组合将由池化层进一步聚合，保留时间序列上的局部特征，得到降维后的特征矩阵。常见的池化函数有平均池化，最大池化等。池化层的输出作为全连接层或者下一个卷积层的输入。最后，全连接层通过非线性函数得到最终

预测结果。CNN 的权值更新使用误差的反向传播算法，不断调节网络连接权重以实现误差的最小化，从而使卷积层和池化层学习到最优参数矩阵。

2.2LSTM 网络

LSTM 是一种特殊的循环神经网络 (RNN)，它改进了传统 RNN 模型中的记忆模块，即单一的 \tanh 层或 sigmoid 层，避免了因为数据持续输入的影响而无法长期保存有效的历史信息的问题^[27]。通过门结构和记忆单元状态的设计，使得 LSTM 可以让时间序列中的关键信息进行有效的更新和传递。

LSTM 模型的基本单元是记忆模块，其中包含记忆单元和三个控制记忆单元状态的门结构，分别是忘记门、输入门和输出门。忘记门决定从记忆单元状态中忘记无用的历史信息，输入门决定当前输入数据对记忆单元状态的影响，输出门决定输出信息。LSTM 记忆模块的结构如图 3 所示。假设 x_t 表示 t 时刻的输入向量， h_{t-1} 表示 $t-1$ 时刻的输出， $W_f, W_i, W_c, W_o, U_f, U_i, U_c, U_o$ 和 b_f, b_i, b_c, b_o 分别表示权重矩阵和偏置向量，则记忆模块进行状态更新和信息输出的过程如下。

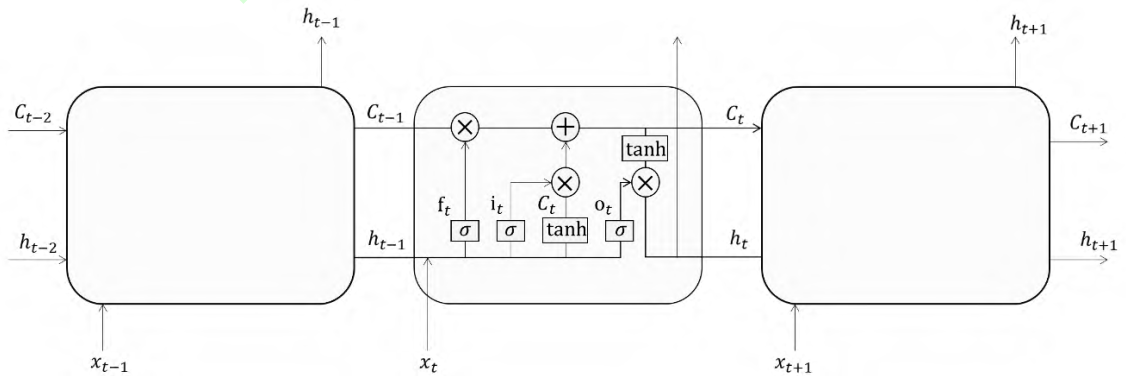


图 3 长短期记忆神经网络记忆模块结构图

首先，忘记门忘记无用的历史信息：

$$f_t = \sigma(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f) \quad (2)$$

然后，输入门根据输入数据和历史信息进行状态更新：

$$i_t = \sigma(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c) \quad (4)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (5)$$

最后，输出门输出当前时刻的信息：

$$o_t = \sigma(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o) \quad (6)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (7)$$

其中， σ 是 logistic sigmoid 函数， f_t , i_t 和 o_t 分别表示 t 时刻忘记门，输入门和输出门的输出状态， C_t 表示 t 时刻的记忆单元状态。

2.3 注意力机制

注意力机制把 LSTM 输出的隐层向量表达进行加权求和计算，其中权重的大小表示每个时间点上的特征重要程度。假设输入为 k 个特征向量

h_i , $i = 1, 2, \dots, k$ ，模型能基于 h_i 获得环境向量 c_i 。

这些环境向量可以和当前的隐藏状态一起实现预测。环境向量 c_i 可以由前面状态的加权平均数得出，如公式(8)所示。

$$c_i = \sum_{i=1}^k a_i h_i \quad (8)$$

其中状态所加的权重就是注意力权重 a_i ，为了得到 a_i ，我们训练一个全连接网络，它的输入是 LSTM 输出的隐层向量，通过计算每个隐层向量的得分 s_i ，评估其对输出的影响程度，即：

$$s_i = \tanh(w^T h_i + b_i) \quad (9)$$

式中， s_i 代表了 h_i 和 c_i 的相关程度，然后采用 SoftMax 函数对得分 s_i 进行归一化，得到最终的权重系数 a_i ，即：

$$a_i = \text{softmax}(s_i) = \frac{e^{s_i}}{\sum_j e^{s_j}} \quad (10)$$

3 实验及分析

3.1 实验环境

实验环境在 Windows10x64 操作系统下，使用基于 Python3.7 解释器的 PyCharm 进行，模型的搭建在 TensorFlow 框架下实现，支持 GPU 的 TensorFlow 版本为 1.13.0。硬件配置为 Intel(R)Core(TM)i5-8250UCPU@1.60 GHz, 1.80 GHz, 内存 16 G, 显卡 NVIDIA® GeForce MX250。

3.2 实验流程

引入注意力机制的 CNN-LSTM 模型的实验流程主要包括数据获取、数据预处理、特征重要性评估、模型训练、模型测试和模型评估等步骤，如图 4 实验流程图所示。其中数据预处理又包括数据归一化、划分训练集测试集和时间序列构建。对数据预处理完毕后，训练数据用于生成模型，即通过优化函数来调整网络权重，使模型损失函数达到最小，直至迭代次数达到设定值。随后，使用训练得到的预测模型对测试集数据进行测试，并通过 RMSE、MAE、MSE 等评估指标衡量模型性能。

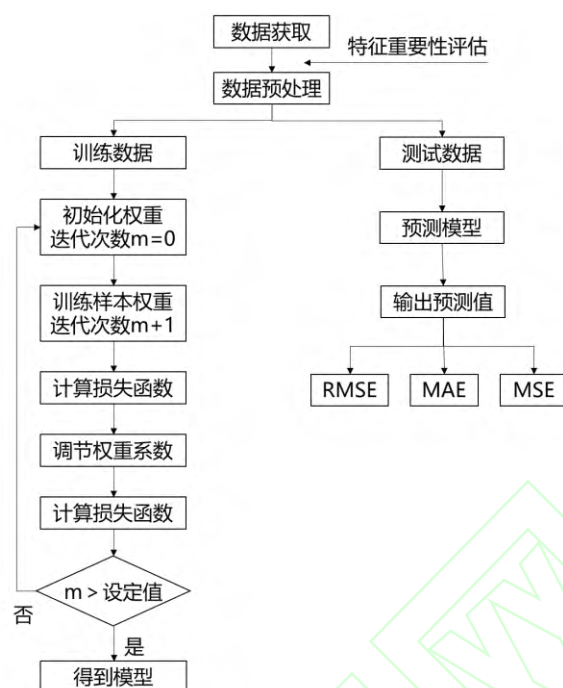


图 4 实验流程图

实验数据选取近三年沪铜近月合约的 5 分钟高频数据,时间范围从 2016-07-25 到 2019-07-25,共有 81917 个样本。输入特征选取了基本行情和经计算得出的技术指标两大类,其中,基本行情包括开盘价、最高价、最低价、收盘价、成交量、成交额、涨跌、涨跌幅和持仓量,技术指标包括乖离率、布林带、趋向标准、指数平滑移动平均、随机指标、简单移动平均、指数平滑异同平均、相对强弱指标等。

为了检验上述输入指标的有效性,基于 XGBoost 算法对所选取指标的有效性进行评估和筛选。在本次实验中,XGBoost 算法的基学习器采用 CART 回归树,通过 gini 不纯度计算特征重要性,迭代次数为 800,学习率为 0.05。由于 XGBoost 本质是一种树结构,数值缩放不影响分裂点位置,因此不同单位、量纲的指标对 XGBoost 的输出没有影响。经过一定迭代后,XGBoost 模型达到拟合。此时特征重要度数值如表 1 所示,特征重要度计算结果可视化如图 5 所示。

由 XGBoost 算法计算结果可知,所选取的指标均对收益率均有一定影响。其中,持仓量对下

一时刻的沪铜近月合约收益率的影响最大。持仓量是指交易者所持有的未平仓合约的双边数量,反映了市场对该合约兴趣的大小。相比于持仓量,交易量对收益率的影响较小,交易量反映的是该合约的活跃程度。在实际交易中,投资者可以通过分析持仓量的变化来了解市场情绪变化,并结合其他技术指标进一步分析。

表 1 基于 XGBoost 算法的特征重要度评估

类别	指标名称	指标解释	重要度
	open	开盘价	0.101
	high	最高价	0.041
	low	最低价	0.078

基本行情	close	收盘价	0.028
	volume	成交量	0.043
	amt	成交额	0.053
	chg	涨跌	0.092
	pctchg	涨跌幅	0.03
	oi	持仓量	0.158
技术指标	BIAS	乖离率	0.052
	BOLL	布林带	0.038
	DMI	趋向标准	0.060
	EXPMA	指数平滑移动平均	0.023
	KDJ	随机指标	0.072
	MA	简单移动平均	0.045
	MACD	指数平滑异同平均	0.038
	RSI	相对强弱指标	0.044

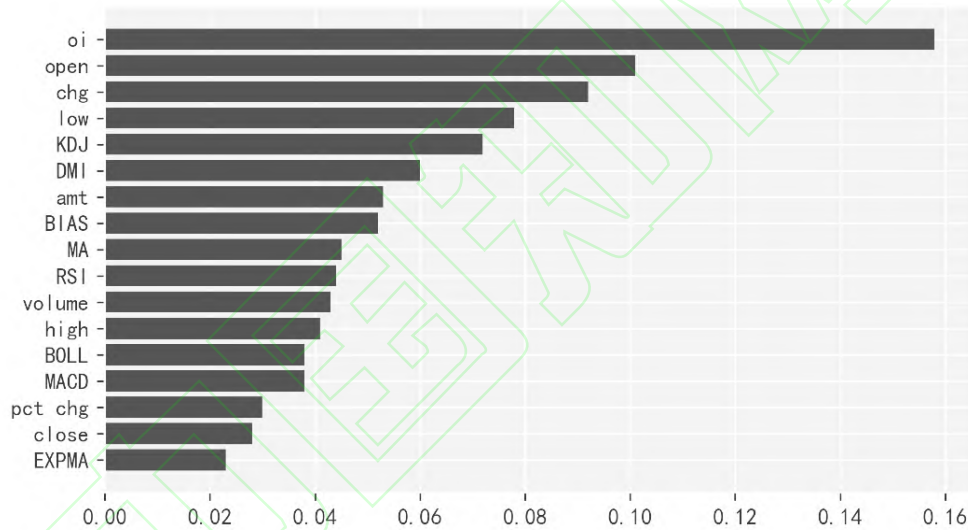


图 5 输入特征重要度评估结果排序

在训练之前，需要对原始数据进行预处理。数据预处理包括数据归一化、划分训练集测试集和时间序列构建。本节选取最大值最小值归一化方法对数据进行归一化处理，计算方法如公式(11)所示。随后，按照 7:3 的比例划分训练集和测试集，并按照 3.1 节所述构造神经网络的输入时间序列。

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (11)$$

其中 x 为某一具体输入特征值， x_{max} , x_{min} 为该特征的最大值，最小值。

在模型训练时，本文损失函数使用均方误差，采用Adam算法进行优化训练。为使均方误差收敛

到最小值，给定网络初始化的随机种子数、学习率以及训练步数，应用Adam优化算法不断更新网络权重，进而得到最终的隐藏层网络。Adam算法由Kingma和Ba提出^[28]，是目前最常用的算法。与其他自适应学习率算法相比，Adam算法收敛速度更快、学习效果更为有效^[29]。

3.3 参数设置

为验证引入注意力机制的 CNN-LSTM 的有效性，基于 TensorFlow 深度学习框架实现该模型，并与 LSTM、CNN、CNN-LSTM 模型作对比实验。

基于注意力机制的 CNN-LSTM 模型结构由输入层、一层卷积层、一层池化层、一层融合注意力机制的 LSTM 层、输出层组成。该模型将时

间窗口大小、卷积核数量、批处理大小、训练次数和隐藏层单元个数设置为超参数。为了减少人为因素对模型的影响，通过一系列实验得到超参数的取值如下：指定时间窗口大小取值为 6，卷积核数量为 64、批处理大小取值为 3000，LSTM 层隐藏层单元个数取值为 128。训练次数由模型误差损失情况直接确定，模型迭代到 100 次时误差损失函数达到收敛状态。因此模型的训练次数为 100。

金融价格时间序列先通过不同通道输入到一维卷积层，该层含有 64 个 1 维卷积核，激活函数为 ReLu 函数，此时张量形状变为(None, 6, 64)，再经过一个窗口为 5 的最大池化层，张量维度变为(None, 1, 64)，然后经一个含 128 个隐藏单元的 LSTM 层，并取其隐含层输出，将多个输出张量由注意力层融合，张量形状变为(None, 128)，得到数据融合特征表达。再最后由输出维度为 1 的全连接层激活输出。

3.4 评估指标

考虑到评价标准在实际应用领域中的特点和局限性，单个的评价指标很难全面、综合衡量模型训练结果。本文采用了三种评价指标来检验模型的训练效果，即 MSE (Mean Square Error)、RMSE (Root Mean Square Error) 和 MAE (Mean

Absolut error)，相比于 RMSE 来说，MAE 在数据里不利于预测结果异常值的情况下鲁棒性更好。但是 MAE 梯度基本是一样的，这会导致训练效果不好。作为损失函数来说，MSE 在预测误差大时学习速率就相应变大，预测误差小时学习速率就相应变小，所以训练效果比较精确。因此本文基于 MSE 进行训练，采用 MSE、MRE、RMSE 作为衡量指标，三者计算公式如下。

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - y_i^p)^2}{n}} \quad (12)$$

$$MSE = \frac{\sum_{i=1}^N (y_i - y_i^p)^2}{n} \quad (13)$$

$$MAE = \frac{\sum_{i=1}^N |y_i - y_i^p|}{n} \quad (14)$$

3.5 结果分析

本节首先对沪铜期货价格数据进行了描述性分析，图 6 展示了沪铜期货的日度对数收益率序列图、收益率分布图及收益率 QQ 图。由序列图和分布图可知，沪铜期货收益率存在波动集聚和杠杆效应，具有尖峰厚尾特征，偏态特征不明显。QQ 图表明沪铜期货收益率分布非正态分布。

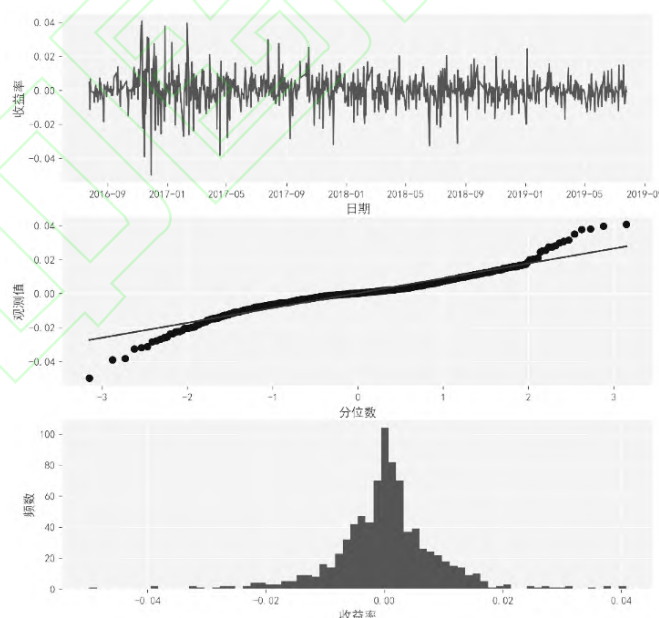


图 6 沪铜期货的收益率序列、收益率分布及 QQ 图

为了分析铜期货价格数据日内特征，借鉴 Amaya 等学者提出的指标构建方法^[30]，基于沪铜期货日内 5 分钟级高频数据计算其高频已实现波

动率 RVar、已实现偏度 RSkew 和已实现峰度 RKurtosis，对其进行描述性分析。表 2 对沪铜期货已实现高阶矩进行了描述性分析，图 7 展示了

指标序列图，各指标计算公式如下。

$$RVar_i = \sum_{j=1}^N r_{ij}^2 \quad (15)$$

$$RKurtosis_i = \frac{N \sum_{j=1}^N r_{ij}^4}{RVar_i^2} \quad (17)$$

$$RSkew_i = \frac{\sqrt{N} \sum_{j=1}^N r_{ij}^3}{RVar_i^{\frac{3}{2}}} \quad (16)$$

其中 r_{ij} 为铜期货合约日内 5 分钟对数收益率序列。

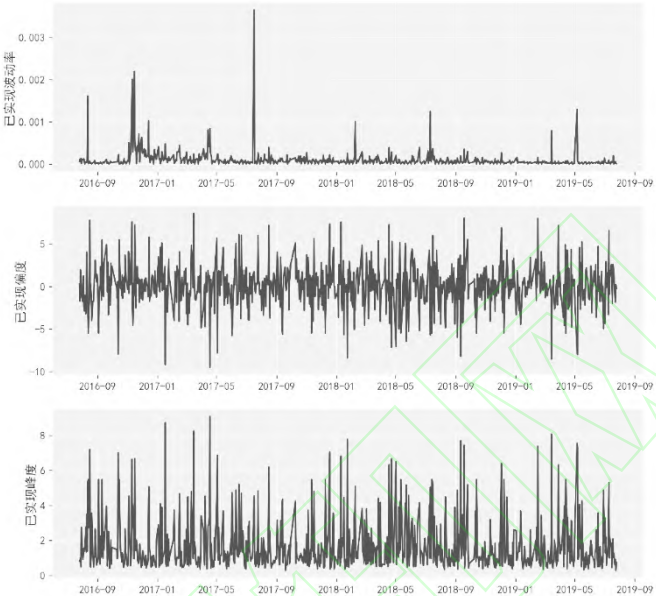


图 7 沪铜期货的高频已实现波动率、已实现偏度和和已实现峰度序列图

表 2 沪铜期货已实现高阶矩描述性分析

	RVar	RSkew	RKurtosis
均值	1.024e-04	0.061	1.657
中位数	5.300e-05	0.111	1.134
最大值	3.649e-03	8.616	9.084
最小值	7.190e-08	-9.478	0.29
标准差	2.164e-04	2.536	1.488
偏度	8.849	-0.106	2.119
峰度	109.959	1.604	4.664

通过一系列实验调整模型的结构和关键参数，以取得较优的预测精度，最终构造的引入注意力机制的 CNN-LSTM 模型结构和参数设置参见 3.3 节。图7为模型的训练过程中损失函数变化曲线，图 8 为测试集中模型输出的预测值及实际值对比图。

由图可知，引入注意力机制的 CNN-LSTM 模型在训练过程中较快达到收敛，最终损失函数值降为 0.00026，在测试过程中，模型中可以较好地拟合价格曲线，预测结果较为准确。

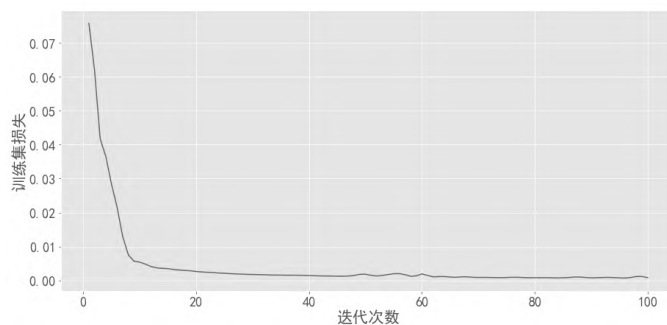


图 8 引入注意力机制的 CNN-LSTM 模型损失函数变化曲线

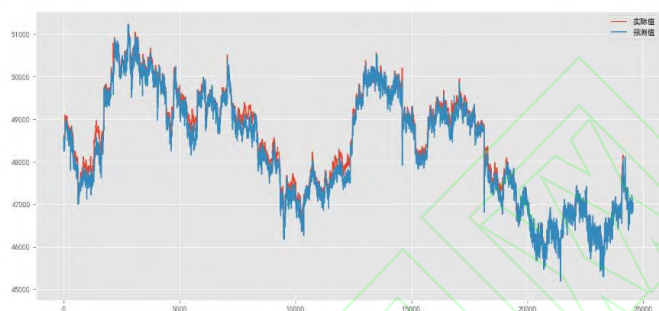


图 9 测试集中预测值及实际值对比

为了验证本文提出的模型的优越性，将该模型与 CNN、LSTM 和 CNN-LSTM 基准模型进行对比，采用 3.4 节介绍的指标来衡量不同模型的效果，结果如表 3 所示。由表可知，CNN-LSTM 模型的预测精度优于单一模型，降低了拟合误差，提高了预测的可靠性。此外，引入注意力机制的

混合模型在各项指标上结果均为最优，在测试集中 MAE 为 0.012893，RMSE 为 0.0162123，MSE 为 0.0002628。这表明本文所构建的铜期货价格预测模型是有效的，同时也看到了注意力机制在时序预测任务中的巨大潜力。

表 3 模型结果对比

数据集	模型	MAE	RMSE	MSE
沪铜期货	Attention-based CNN-LSTM	0.01289	0.01621	0.00026
	CNN-LSTM	0.01903	0.0195	0.00038
	CNN	0.04464	0.03555	0.00126
	LSTM	0.0231	0.02155	0.00046

实验结果表明，与 CNN、LSTM 模型对比，本文所提出的模型结合了两者的优势。从理论上来看，CNN 模型更适合进行空间扩展，提取数据的局部特征，并组合抽象成高层特征。而铜期货的价格信息在一段时间内能够反映出市场的状态，是对市场进行预测的有效特征。此外，其价格特征也符合局部视野和权值共享的思想，即每个神经元只需分析周围的数据。而 LSTM 更适合进

行时间扩展，具有长期记忆功能，更适合处理时间序列。在铜期货价格的特征提取中，不仅需要考虑不同特征之间的空间联系，还需要考虑数据在时间维度上的变化，结合后的 CNN-LSTM 模型具有时空特征表达能力，预测效果将更为准确。实验结果也验证了这一点；与 CNN-LSTM 模型对比，引入注意力机制的 CNN-LSTM 模型在沪铜期货价格预测任务中具有更好的精度，进而证明了

注意力机制可以调节网络结构,解决 CNN-LSTM 模型可能存在的重要特征丢失问题。

4 结束语

本文提出了一种基于注意力机制和 CNN-LSTM 神经网络的沪铜期货价格预测模型。CNN 结构用来提取特征, LSTM 模型用已经提取结果的特征来对时间序列做预测,能解决传统神经网络不能记忆和利用历史信息的问题。此外,引入注意力机制调节 LSTM 时间维度的特征表达,使模型获取更全面的特征信息,从而学习到当前局部序列特征的重要程度。为了验证模型的有效性,实验利用上海期货交易所近三年的高频 5 分钟铜期货数据,选取了 17 个指标,基于 XGBoost 方法进行了特征有效性和重要度评估,构建了沪铜期货高频价格预测模型。

实验主要有以下发现:

(1).根据对沪铜价格数据的统计检验,沪铜期货的收益率和已实现波动率均具有明显的波动聚

集效应,收益率具有明显的“尖峰厚尾”特性,已实现波动率具有偏态性,偏度为正。

(2).通过 XGBoost 算法计算特征重要性发现,本文选取的 17 个指标对预测收益率均有贡献,其中,持仓量对收益率的重要度最大。

(3).与基准模型相比,引入注意力机制的 CNN-LSTM 模型在铜期货价格预测任务上性能更优。但在预测过程中,也存在一定的滞后现象。

本研究的重点在于优化价格预测模型结构,提高价格预测精度。但由于沪铜期货高频市场存在较高噪声,对其价格进行准确预测仍是一项复杂的任务。在进一步的研究中,可考虑从数据去噪、特征工程构建阶段对模型进一步优化。此外,本文对铜期货日内高阶矩特征,即已实现波动率、已实现偏度和已实现峰度进行了初步描述性分析,在未来,可考虑继续探究其对金属期货价格、期货收益率的预测作用。

参考文献

- [1].Pan H. A basic theory of intelligent finance[J]. New Mathematics and Natural Computation, 2011, 7(02): 197-227.
- [2].Nguyen T H, Shirai K, Velcin J. Sentiment analysis on social media for stock movement prediction[J]. Expert Systems with Applications, 2015, 42(24): 9603-9611.
- [3].季俊伟,傅强,张兴敏.沪铜期货噪声交易与渐进有效性[J].管理评论,2019,31(02):17-35.
- [4].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(06):99-109.
- [5].李飞,黄凤.中证 500 股指期货价格发现功能实证研究[J].价格理论与实践,2019(02):88-91.
- [6].何迪,周勇.基于状态空间模型的宏观经济因素对股市流动性的建模分析[J].中国管理科学,2019(05):42-49.
- [7].Adebiyi A A, Adewumi A O, Ayo C K. Comparison of ARIMA and artificial neural networks models for stock price prediction[J]. Journal of Applied Mathematics, 2014, 1-7.
- [8].Basher S A, Sadorsky P. Hedging emerging market stock prices with oil, gold, VIX, and bonds: A comparison between DCC, ADCC and GO-GARCH[J]. Energy Economics, 2016, 54: 235-247.
- [9].侯利强,杨善林,王晓佳,陈志强.上证综指的股指波动:基于模糊FEGARCH模型及不同分布假设的预测研究[J].中国管理科学,2015,23(06):32-40.
- [10].Cavalcante R C, Brasileiro R C, Souza V L, Nobrega J P, Oliveira A L. Computational intelligence and financial markets: A survey and future directions[J]. Expert Systems with Applications, 2016, 55: 194-211. 7046-7056.
- [11].Chang P C, Fan C Y, Lin J L. Trend discovery in financial time series data using a case based fuzzy decision tree[J]. Expert Systems with Applications, 2011, 38(5):6070-6080.
- [12].Srinivasan N, Lakshmi C. Stock Price Prediction Using Fuzzy Time-Series Population Based Gravity Search Algorithm[J]. International Journal of Software Innovation, 2019, 7(2): 50-64.
- [13].喻胜华,肖雨峰.基于信息粒化和支持向量机的股票价格预测[J].财经理论与实践,2011,32(06):44-47.
- [14].胡波.商品期货高频交易价格趋势实证分析——以黄金期货市场为例[J].价格理论与实践,2018(01):114-117.
- [15].Ballings M, Van den Poel D, Hespeels N, Gryp R. Evaluating multiple classifiers for stock price direction prediction[J]. Expert Systems with Applications, 2015, 42(20): 7046-7056.
- [16].罗洪奔.基于灰色-ARIMA 的金融时间序列智能混合预测研究[J].财经理论与实践,2014,35(02):27-34.
- [17].林杰,龚正.基于人工神经网络的沪锌期货价格预测研究[J].财经理论与实践,2017,38(02):54-57.
- [18].Jin Z, Yang Y, Liu Y. Stock closing price prediction based on sentiment analysis and LSTM[J]. Neural Computing and Applications, 2019(3).
- [19].Kim H Y, Chang H W. Forecasting the Volatility of Stock Price Index: A Hybrid Model Integrating LSTM with Multiple GARCH-Type Models[J]. Expert Systems with Applications, 2018, 103:S0957417418301416.
- [20].Mnih V, Heess N, Graves A. Recurrent models of visual attention[C]//Advances in neural information processing systems. 2014: 2204-2212.
- [21].Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.
- [22].Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation[J]. arXiv preprint arXiv:1508.04025, 2015.
- [23].杨贵军,徐雪,赵富强.基于 XGBoost 算法的用户评分预测模型及应用[J].数据分析与知识发现,2019,3(01):118-126.
- [24].Zhang D, Qian L, Mao B, Huang C, Huang B, Si Y. A data-driven design for fault detection of wind turbines using random forests and XGboost[J]. IEEE Access, 2018, 6: 21020-21031.
- [25].Zheng H, Yuan J, Chen L. Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance

evaluation[J]. *Energies*, 2017, 10(8): 1168.

[26].Zhao B, Lu H, Chen S, Liu J, Wu D. Convolutional neural networks for time series classification[J]. *Journal of Systems Engineering and Electronics*, 2017, 28(1): 162-169.

[27].Cornia M, Baraldi L, Serra G, Cucchiara R. Predicting human eye fixations via an LSTM-based saliency attentive model[J]. *IEEE Transactions on Image Processing*, 2018, 27(10): 5142-5154.

[28].Kingma D, Ba J. Adam: A Method for

Stochastic Optimization[J]. *Computer Science*, 2014.

[29].Sabour S, Frosst N, Hinton G E. Dynamic Routing Between Capsules[J]. *Neural Information Processing Systems*, 2017: 3856—3866.

[30].Amaya D, Christoffersen P, Jacobs K, Vasquez A. Does realized skewness predict the cross-section of equity returns?[J]. *Journal of Financial Economics*, 2015, 118(1): 135-167

Forecasting High Frequency Price of Shanghai Copper Futures based on Attention Mechanism and CNN-LSTM

JING Nan¹, SHI Zijing¹, SHU Yumin^{2*}

(1. SHU-UTS SILC Business School, Shanghai University, Shanghai, 201800;

2. School of Economics and Management, Harbin Institute of Technology, Shenzhen, 518055)

Abstract:Allowing for the complexity and long-term dependence of Shanghai Copper Futures price, an Attention-based CNN-LSTM model is proposed in this paper for the forecast of futures price. In our study, a futures price prediction model that integrates Convolutional Neural Network (CNN) and long-term Memory (LSTM) Network is presented, while the attention mechanism is introduced at the LSTM end to optimize the model. The model proposed in this paper combines the advantages of CNN in time expansion and the long-term memory of LSTM. Firstly, the CNN module is applied to extract the features of different weights from the price time series. Then, the features are taken as the input of LSTM to predict the future price, and the long-term dependence of LSTM is used to improve the accuracy of prediction for futures prices. Finally, the attention mechanism is introduced at the LSTM end to keep the model focused on the learning of more significant data features, thus further enhancing the performance in prediction. In the process of experiment, the five-minute high-frequency price data on Shanghai Copper Futures over the past three years was collected, the XGBoost algorithm (eXtreme Gradient Boosting) was applied to evaluate and screen the indexes, and the Attention-based CNN-LSTM was constructed to predict the closing price at the next moment. According to the experimental results, the Attention-based CNN-LSTM model is advantageous over such benchmark models as CNN, LSTM and CNN-LSTM in the accuracy of price prediction, which lays a foundation for the further development of corresponding quantitative investment strategies.

Keywords: Attention; LSTM; CNN; CNN-LSTM; Shanghai Copper Futures