

基于 LSTM 深度学习的大豆期货价格预测

范俊明 刘洪久 胡彦蓉

(浙江农林大学, 浙江杭州 311300)

【摘要】面对越来越复杂的金融市场环境,以传统统计学和计量学为主的时间序列预测模型在发现序列中的长期依赖关系方面存在一定局限性,而深度学习中的长短期记忆(LSTM)网络有望克服这一问题。通过构造一个多层 LSTM 网络价格预测模型,使用中国 2007—2019 年大豆期货价格数据进行了实证研究。结果显示,参数调优对 LSTM 网络模型预测效果有着较大影响,其中影响较大的主要参数包括迭代次数、学习率、窗口大小和网络层数等;与 ARIMA 模型、MLP 模型、SVR 模型相比,LSTM 网络模型的预测结果准确性更高,在拟合优度(R^2)上分别提高了 1.064%、2.147%、1.674%。LSTM 网络模型在价格预测方面的良好表现,为预测大豆期货价格提供了新思路。

【关键词】深度学习 LSTM 神经网络 价格预测 大豆期货

【中图分类号】F737.7 **【文献标识码】**A **【文章编号】**1006-2025(2021)02-0007-09

一、引言及文献综述

中国作为消费大国,对大豆的需求量巨大。统计数据显示,2011—2017 年,中国大豆进口量从 5263.41 万吨增至 9552.6 万吨,增长了约 81.4%,2018 年、2019 年大豆进口量虽然有所回落,但依然保持在 8800 万吨左右,中国对大豆消费的刚需性不言而喻。目前,受国内大豆产量限制,中国大豆市

场依然处于供不应求、高度依赖进口的状况。大豆进口价格受到诸多因素影响,如大豆生产国的产量变动、大豆消费国的需求变动、大豆替代品的供求变化、国内外贸易政策^[1]及国际局势等。为了应对这些因素对国内大豆市场价格的影响,中国积极发展并完善大豆期货市场,以期在大豆价格战中取得一定话语权。

【收稿日期】2020-09-14

【基金项目】教育部人文社会科学研究规划基金项目“社会阻抑对情绪劳动、顾客导向跨界行为的影响机制研究——以一线服务员工为例”(编号:18YJA630030);浙江省公益研究计划项目“基于系统动力学的并购 DCF 价值评估方法优化研究”(编号:LY18G010005);浙江省哲学社会科学规划课题“基于卷积神经网络的企业财务风险特征识别和评价预警研究”(编号:19NDJC240YB)。

【作者简介】范俊明,浙江农林大学硕士研究生,主要研究方向为人工智能预测与决策研究;通信作者:刘洪久,博士,浙江农林大学信息工程学院教授,硕士生导师,主要研究方向为预测管理;胡彦蓉,博士,浙江农林大学信息工程学院副教授,硕士生导师,主要研究方向为预测与评价。

梳理国内外研究文献可以发现,现有研究方法大致可以划分为线性与非线性两大类预测模型。其中线性模型主要包括一些传统统计和计量经济学模型,如向量自回归模型(VAR)、差分自回归移动平均模型(ARIMA)、广义自回归条件异方差模型(GARCH)等。王锐和陈倬(2011)运用 VAR 模型研究了各种因素对中国农产品价格波动的影响。^[2]吴玉霞和温欣(2016)于志军和杨善林(2013)分别运用 ARIMA 模型和误差修正 GARCH 模型对股价进行了预测研究。^[3,4]其研究所使用的原始数据均需满足平稳性假设,且模型的性能非常依赖参数选择,如有时必须服从正态分布才能进行建模。期货市场价格受标的物价格、供求状况、信息不对称、周期变化、社会政策、不可抗力等因素影响,波动极其不规律,导致试验数据在测试过程中存在显著噪声并表现出不平稳特征。这些问题使得线性研究方法在对该类型数据进行预测时显得有些力不从心,尤其在面对更高精度预测需求时,线性模型更是会面临瓶颈,较大程度限制时间序列模型的运用与拓展。

相较于线性模型,非线性模型具有良好的预测准确性和变量非线性关系挖掘能力,近些年来,其在金融预测任务中优势明显,受到了各界的广泛青睐。^[5]其中,包括决策树、支持向量回归模型(SVR)、随机森林(RF)、人工神经网络(ANN)等在内的以机器学习为代表的各类模型在非线性预测模型中扮演着越来越重要的角色。机器学习是一个以经验数据和学习算法为基础、进而实现模型性能自我提升的过程,该方法已经被广泛运用于各大领域的研究中,而金融数据预测正是运用的重要方向之一。杨建辉和李龙(2011)提出了非参数方法 SVR 与改进的期货价格预测模型,其在预测精度方面明显优于传统的参数方法。^[6]Dhar S. et al.使用决策树和随机森林集成算法来预测股票的回报率,以期降低投资风险。^[7]杨海民和潘志松(2019)通过多次试验发现,传统的机器学习预测模型虽然能在一定程度上弥补线性模型的不足,但在处理过程中对样本分布要求较高,后续结果显示的模型预测精度较低,且泛化能力不足。^[8]Basak S. (2019), Moghaddam A. H. (2016), 魏文轩(2013)分别使用贝叶斯学习(BL)、BP 神经网络和改进的径向基函数(RBF)对印度、美国、中国等 3 个国家的股票市场价格变化趋势进行

了预测分析。^[9-11]贝叶斯学习、BP 神经网络和改进的径向基函数属于早期的传统浅层神经网络(SNN)预测模型,存在收敛速度不理想、易陷入局部极值以及难以找到全局最优解等缺陷。

近年来, Lecun Y. et al. (2015) 从理论层面阐述了深度学习是如何突破局部极值障碍的。^[12]有别于传统的浅层机器学习模型,深度学习模型隐含的层数更多且对数据的依赖度更高,具有更好的数据表征能力和预测效果,已经成为目前最受欢迎的研究方法。深度学习模型主要包括卷积神经网络(CNN)、深度信念网络(DBN)和循环神经网络(RNN)及变体模型。这些研究模型在图像识别、语音识别、自然语言处理等方面均取得了重大成功,在金融领域研究方面也获得了一定进展,但将其应用于期货市场研究的并不多见。笔者选取 2007—2019 年中国大豆期货价格数据,运用 LSTM 神经网络进行相关研究。本文的主要贡献在于:(1)将对历史信息具有良好记忆能力的 LSTM 深度学习模型引入到大豆期货价格预测分析中,并取得了比较符合预期的实验结果,为金融市场价格预测研究提供了一种新思路;(2)构建了一个多层 LSTM 神经网络模型,并采用目前较为先进的 Dropout 正则技术,有效降低了实验过程中过拟合发生的概率,切实保障了实验预测的精准度;(3)选取 ARIMA 模型作为线性模型的代表,选取 SVR 模型作为非线性模型中机器学习的代表,选取 MLP 模型作为非线性模型中浅层神经网络的代表,将通过 3 种模型得出的实验结果与本文所构建的多层 LSTM 神经网络模型预测结果进行对比,论证了多层 LSTM 神经网络模型在预测精度和预测性能上具有更好的效果。

二、研究方法

(一) LSTM 网络

Hochreiter S. & Schmidhuber J. 于 1997 年提出了长短期记忆(LSTM)网络,^[13]后经 Graves A. (2012) 的改进,使得 LSTM 得到了进一步推广。^[14]作为 RNN 的变体,LSTM 模型具有独特的设计结构:巧妙地引入可控自循环,以产生让梯度能够长时间可持续流动的路径,有效克服了 RNN 在机器学习中因时间流逝、网络层数增多而带来的梯度消失问题。因此,该模型特别适合处理时序间隔和延迟非常长的任务,且性能优异。LSTM 记忆单元结构如图 1 所示。^[15]

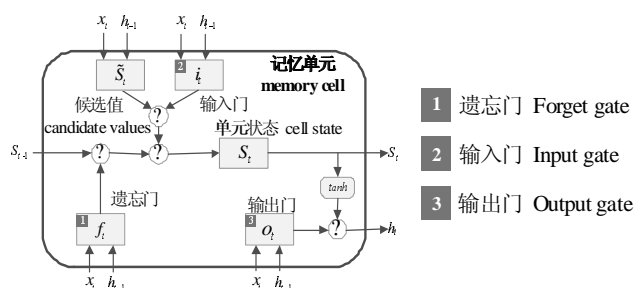


图 1 LSTM 记忆单元结构图

LSTM 网络由一个输入层、一个或多个隐藏层及一个输出层组成。输入层中,神经元的数量等于解释变量(特征空间)的数量,输出层中神经元的数量反映了输出空间的广度。相较于传统的 RNN 模型,LSTM 网络中的每个记忆神经元均新增了 1 个记忆单元状态 s_t 和 3 个控制门(gate),即遗忘门、输入门和输出门。其中单元状态 s_t 用以取代传统的隐含层神经元节点,主要负责将记忆信息从序列的初始位置传递到序列的末端。单元状态 s_t 随着时间的变化而变化,其存储的信息由门控制决定与更新。

(二)门控机制

在每个时间步长 t 时,3 个门中的每 1 个门都具有输入 x_t (输入序列的 1 个元素)及在前 1 个时间步长 $t-1$ 时的记忆单元输出 h_{t-1} 。因此,这些门充当了过滤器,每个过滤器都实现各自不同的功能。

1.遗忘门:遗忘门指定从单元状态中丢弃多少信息。

2.输入门:输入门确定哪些信息添加到单元状态中。

3.输出门:输出门决定哪些单元状态信息作为输出。

在时间步长为 t 的 LSTM 层单元状态 s_t 和输出 h_t 的前向计算步骤如下。

第一步:确定应从先前的单元状态 s_{t-1} 中删除哪些信息。遗忘门在时间步长为 t 时的激活值 f_t 可通过当前输入 x_t 、单元在前 1 个时间步长的输出 h_{t-1} 和遗忘门的偏置项 b_f 计算得到,计算如公式(1)所示。

$$f_t = \sigma(W_{fx} \times x_t + W_{fh} \times h_{t-1} + b_f) \quad (1)$$

其中, σ 为 sigmoid 激活函数,形式为 $\sigma(x) = (1 + e^{-x})^{-1}$,能将所有激活值 f_t 缩放到 0(完全忘记)和 1(完全记住)范围。 W_{fx} 表示遗忘门输入层与隐藏层之间的权重矩阵。

第二步:确定将哪些信息添加到网络单元状态 s_t 中。该过程又包括以下两个步骤:首先计算有可能被添加到单元状态的候选值 \tilde{s}_t ;其次计算输入门的激活值 i_t 。具体步骤如公式(2)和公式(3)所示。

$$\tilde{s}_t = \tanh(W_{sx} \times x_t + W_{sh} \times h_{t-1} + b_s) \quad (2)$$

$$i_t = \sigma(W_{ix} \times x_t + W_{ih} \times h_{t-1} + b_i) \quad (3)$$

第三步:根据前两个步骤得到的结果计算新的单元状态 s_t ,如公式(4)所示。

$$s_t = f_t \circ s_{t-1} + i_t \circ \tilde{s}_t \quad (4)$$

其中, \tanh 为双曲正切激励函数,形式为 $\tanh(x) = (e^x - e^{-x}) / (e^x + e^{-x})$;符号 \circ 为矩阵哈达玛积。

第四步:记忆单元的输出 h_t 如公式(5)和公式(6)所示。

$$o_t = \sigma(W_{ox} \times x_t + W_{oh} \times h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t \times \tanh(s_t) \quad (6)$$

处理输入序列时,其门控机制会一步一步呈现给 LSTM 网络。因此,网络将根据上述步骤来处理每个时间步长 t 的输入,处理完序列最后 1 个元素后,将返回整个序列的最终输出。

(三)标准 Dropout 正则化

为应对大规模复杂神经网络的过拟合问题,Srivastava N.et al.(2014)提出了标准 Dropout 正则化技术,^[16] 该技术随后被广泛地应用于神经网络正则化及模型压缩等任务中。^[17]模型训练期间,标准的 Dropout 数学表达式如公式(7)所示。

$$y = f(Wx) \circ m, m_i \sim \text{Bernoulli}(p) \quad (7)$$

其中, $f(\cdot)$ 是激活函数, x 是输入, W 是权值矩阵, y 为输出, m 为 Dropout 掩膜(mask),mask 中每个元素为 1 的概率为 p 。在测试阶段,该层的输出如公式(8)所示。

$$y = pf(Wx) \quad (8)$$

标准 Dropout 在正向传递和权值更新过程中,按照一定的概率将一部分神经元输入和递归连接进行暂时性丢弃。

这样,Dropout 就像用相同的数据训练了多个子网络,产生了多种拟合状态。但这些网络共享一个损失函数(loss),相当于对网络自身进行了优化,求取了所有状态的平均值;同时,弱化了神经元节点间的耦合性,增强了网络的泛化性,改善了模型性能。

(四)性能指标

笔者在评估模型预测性能时选取 MAE、RMSE、

MAPE、 R^2 等 4 项评价指标, 计算方法分别如公式(9)至公式(13)所示。

1. 平均绝对误差(MAE)。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - p_i| \quad (9)$$

2. 均方根误差(RMSE)。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2} \quad (10)$$

3. 百分比误差绝对值的平均值(MAPE)

$$MAPE = \frac{1}{n} \times \sum_{i=1}^n \left| \frac{y_i - p_i}{p_i} \right| \times 100\% \quad (11)$$

4. 决定系数 R^2 (拟合优度)

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (12)$$

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^n (y_i - p_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (13)$$

其中, y 是真实值, p 是预测值, i 为样本个数, \bar{y} 是 n 个样本的平均值。

三、实证研究

(一) 实验平台

基于 Python3.6 语言环境, 以 Keras 作为深度学习的框架进行训练及预测。Keras 封装了目前最优秀的两个开源深度学习框架 TensorFlow 和 Theano 的诸多小组件, 用户只需通过 API 将构建的模块排在一起, 就可以设计出各种类型的神经网络, 进而快速更改网络层数和层类型, 在优化网络时非常方便。

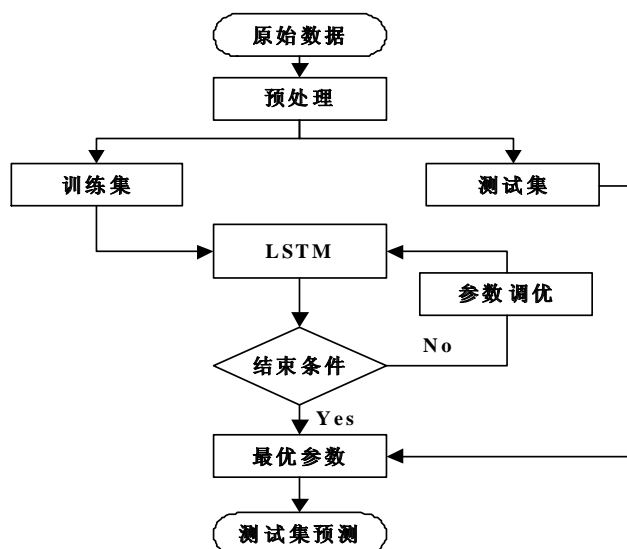


图2 整体流程图

(二) 实验总体流程

实验的工作任务主要包括数据预处理、模型构建、参数调优及模型对比, 整体流程结构如图 2 所示。

(三) 数据预处理

1. 样本区间及训练集、测试集划分

全文数据来源于新浪财经-期货-大连商品交易所大豆连续价格指数。样本区间: 选取现存日成交量数据即 2007 年 8 月 21 日—2019 年 12 月 31 日的所有交易数据, 剔除日成交量为 0 的所有收盘价格, 最后收集的样本共计 2731 条。将原始数据集的前 85% 作为训练集用于训练模型, 后 15% 作为测试集用于评估模型性能, 另外从训练集中再划分出 15% 数据作为验证集用于确定网络结构。

2. 数据归一化

为了使模型能够得到更好的收敛, 需要对每个训练和测试数据窗口 n 进行归一化处理, 以反映该窗口开始时的百分比变化。使用公式(14)进行归一化, 并在预测过程结束时使用公式(15)进行反归一化, 得以从预测中获得真实的数值。

$$n_i = \left(\frac{p_i}{p_0} \right) \quad (14)$$

$$p_i = p_0(n_i + 1) \quad (15)$$

其中, i 代表窗口数据中第 i 个数据, n 代表价格转化后的窗口归一化列表 ($i=0$ 处的归一化数据始终为 0), p 代表调整后返回的每日价格窗口原始列表, p_0 代表窗口原始列表的首个数据。

(四) 模型构建与仿真

实验目标是根据大豆期货历史收盘价构建一个回归预测模型, 故选取均方差(MSE)作为模型训练时的损失函数(loss)。在优化器选择方面, 采用 Adam 优化器进行优化训练。与其他随机优化算法相比, Adam 算法计算效率更高, 内存需求更低, 适合求解大规模数据问题。^[18]

1. 模型数据加载

在训练过程中, 假设移动窗口大小 w 设为 3 (即数据序列长度为 4), 模型加载训练数据如图 3 所示。由图 3 可知, 时间窗口每次截取的数据将作为模型输入特征 $X_{feature}$, 而 $w+1$ 项作为需要预测的、未来 $t+1$ 天的价格标签 Y_{lable} , 时间窗口每次移动的步数为 1。这样将在训练数据集上依次取得训练所需的模型输入样本数据共 $n-w$ 个, 将这些窗口数

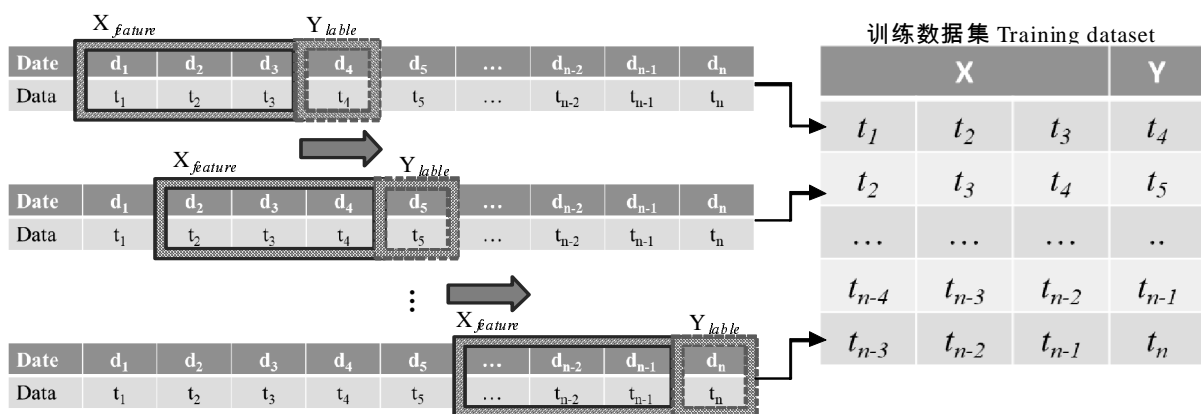


图3 模型加载训练数据示意图

据随机打乱输入模型训练中,随机打乱数据能够使训练更切合数据的真实分布,从而提高预测模型的泛化能力。

2. 模型参数的确定

实验中,超参数设定对预测结果影响较大,需要对影响预测精度的超参数进行实验对比进而确定 LSTM 网络模型的最优网络结构,主要有 3 方面内容:(1)学习率与迭代次数;(2)时间窗口长度与网络层数;(3)隐藏层节点数与 Dropout 率。此外,训

练期间表现最好的模型会被自动保存下来。

(1) 学习率与迭代次数

迭代是指将所有训练集数据样本进行一次完整训练测试的过程,训练测试的次数被称为迭代次数。

学习率(lr)决定网络梯度更新的幅度大小。若 lr 过大,容易导致最终在最优值处震荡但不收敛;反之,会使网络训练速度过慢,从而增加时间成本。

实验中,以 1 层 LSTM 结构的网络模型为基准,预设训练迭代周期相对足够大,同时在模型中设置

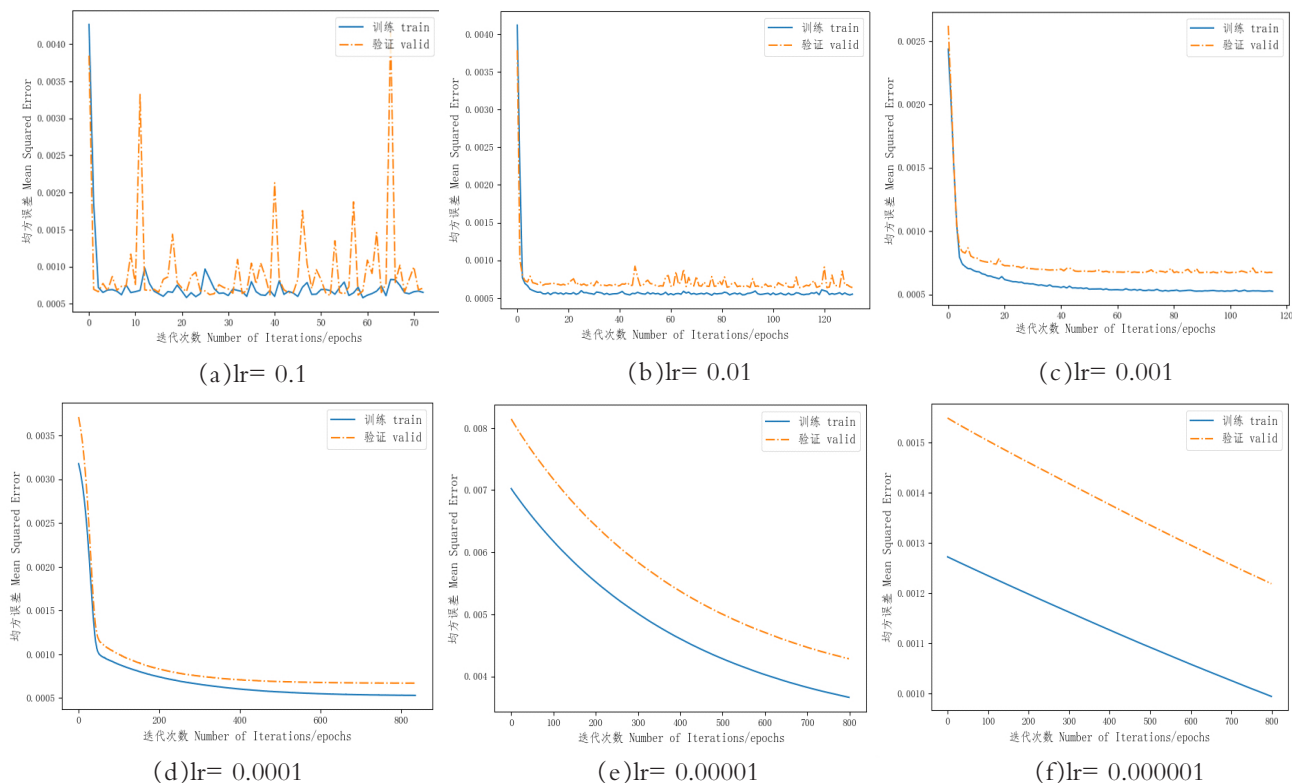


图4 不同学习率下的损失函数曲线

早停(Early-Stopping)机制防止模型因训练过长而导致过拟合,设置忍耐参数 $\text{patience}=20$,即当模型在 20 个 epochs 验证集上的损失函数没有得到改善时,则模型训练自动结束,从而有效防止模型训练过度而导致过拟合。 lr 从典型值 0.1 到 0.000001,以 10 倍为间隔降低,通过观察学习曲线对应的 loss 值变化来确定 lr 量级,再进行微调确定最佳的 lr 。不同学习率对应的损失函数曲线如图 4 所示,其中训练和验证数据均为归一化后的数据。

从图 4 可以发现, lr 为 0.1 和 0.01 时,学习率过大,同时模型训练迭代次数较少,在最优处不断振荡无法正常收敛,数据在验证集上表现得更为明显。 lr 为 0.00001 和 0.000001 时,学习率过小,模型欠拟合,在 800 个 epochs 时 loss 曲线仍在下降且未收敛。 lr 为 0.01 和 0.0001 时,loss 曲线逐渐平滑下降收敛。在此规律下,微调最终确定的模型最佳 lr 为 0.0006。

在对 LSTM 网络模型进行训练时,过少的迭代次数无法使模型获得准确预测结果,为此,需要将训练的数据集在同样的神经网络中传递多次。为寻找合适的迭代次数,LSTM 网络模型被进行了多次训练,不同迭代次数相对应的误差统计如图 5 所示。

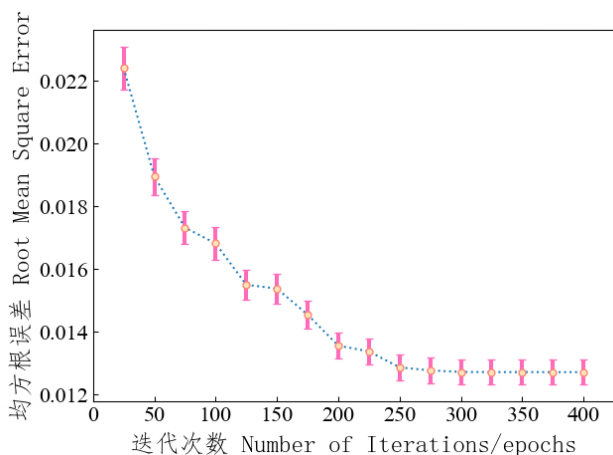


图 5 不同迭代训练次数对应的均方根误差图

由图 5 可知,随着模型迭代次数的增加,LSTM 网络模型的 RMSE 值在前 100 个 epochs 中快速下降,在第 200 个 epoch 中开始逐渐收敛并趋于稳定,结合本文实验的复杂程度,300 个 epochs 即可满足训练精度要求。

(2)时间窗口长度与网络层数

Chen K,et al.(2015)指出,增加 LSTM 网络隐藏层数能够增强对序列特征的提取能力、提高模型预测精度。^[19]彭燕等(2019)的实验结果表明,过度增加网络深度对预测准确率提升效果并不明显,反而会增加计算冗余度、延长模型训练时间。^[20]由杨青和王晨蔚研究成果可知,对于未来 1 天交易日的大豆期货收盘价,前 60 天的交易数据已经包含了足够的信息,因此对应的时间窗口长度不必过长。笔者分别构建 1~3 层 LSTM 网络模型对不同时间窗口长度 w (10~50)进行测试实验,并将 5 次实验结果得到的各项指标平均值作为最终指标标准,通过比较来确定网络层数结构及最佳窗口大小。不同时间窗口长度与 LSTM 网络隐藏层数对应的各项指标均值如表 1 所示,其中各项误差指标数据均由反归一化数据计算得到。

表 1 不同时间窗口长度与 LSTM 网络隐藏层数各项指标均值

窗口大小	MAE	RMSE	MAPE	R ² /%
Layer1				
10	11.2313200	16.1651400	0.3516020	99.4040
20	13.9206597	19.3102133	0.4375827	99.1489
30	13.1513440	19.6971660	0.4106760	99.1140
40	12.9672503	17.8589530	0.4079830	99.2657
50	20.5012360	25.2014600	0.6546253	98.4576
Layer2				
10	10.5299283	15.0509530	0.3297477	99.4827
20	16.9381363	21.8294867	0.5351983	98.9138
30	13.8221810	19.2374463	0.4358157	99.1572
40	13.6943103	18.5449670	0.4303787	99.2068
50	14.4745310	19.6099533	0.4584800	99.0777
Layer3				
10	10.0591407	14.2504733	0.3151757	99.5365
20	20.7673297	25.3927273	0.6567210	98.5388
30	12.7451137	17.7448160	0.4005433	99.2811
40	12.4615607	17.4958547	0.3913803	99.2953
50	13.9653010	19.2123483	0.4402020	99.1152

注:加粗部分为结果最佳的项。

由表 1 可知,除了窗口长度为 20 的模型,随着网络层数 l 的增加和时间窗口 w 的增长,模型的性能都具有一定程度提升,其中窗口长度为 10 的模型性能最优,最终选定模型的网络层数为 3,窗口大小为 10。

(3)隐藏层节点与 Dropout 率

若隐藏层节点数过少,网络模型将无法充分学

习到数据之间隐含的特征联系;若隐藏层节点数过多,将会增加网络的复杂度,进而增加训练网络的计算难度和时间成本,且容易出现过拟合现象。对于隐藏层节点,一般都是根据经验得出一个大致范围,通过多次试验选取一个合适的数值。夏克文等(2005)在研究中提供了一些确定隐藏层节点数的经验公式,^[21]如公式(16)、公式(17)所示。

$$n_h = \sqrt{i + o + \alpha} \quad (16)$$

$$n_h = 2m + 1 \quad (17)$$

其中, n_h 、 i 和 o 分别为神经网络的输入层、隐含层和输出层节点数, α 为1~10之间的常数。本实验参照经验公式,通过多次测试依次找到对应RMSE最小的 n_h ,分别为100、100、10。

对于Dropout在模型中的添加位置,笔者提出的最优方案是在第1层LSTM层和最后1层LSTM层后分别添加Dropout,且Dropout率均设置为0.2。

综上所述,结合以上实验确定的各个参数,模型最终结构如图6所示,各层对应的具体参数如表2所示。

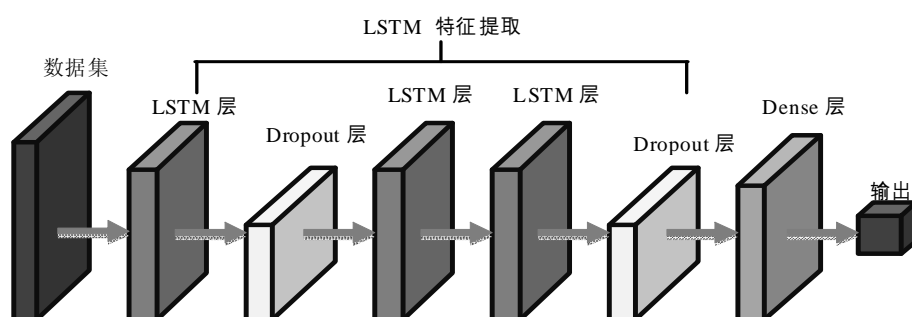


图6 多层LSTM网络模型结构

表2 模型各层具体参数

LSTM-layer1 neurons	LSTM-layer2 neurons	LSTM-layer3 neurons	Dense neurons	Activation	Dropout rate
100	100	10	1	linear	0.2

表3 不同模型对应评估指标汇总

	MAE	RMSE	MAPE	R ² /%
ARIMA (5, 1, 5)	23.91108	37.71460	0.733937	97.9694
MLP (BP)	35.18553	49.62300	1.080520	96.9299
SVR	24.41697	38.34014	0.749243	97.3816
LSTM	17.90572	23.57394	0.551580	99.0001

3.模型预测

采用训练好的LSTM网络模型和ARIMA线性模型、SVR模型、MLP模型同时对大豆期货价格进

行预测,并计算出各模型的评估指标,结果如表3所示(表3中显示的数据均由反归一化数据计算得到)。

由表3可知,LSTM网络模型的各项误差指标数值均小于其他模型,拟合优度R²在所有模型中是最优的。此外,为了更加直观地体现LSTM网络模型的适用性和优越性,图7给出了LSTM网络模型最优拟合图以及上述4种模型预测结果对比图。

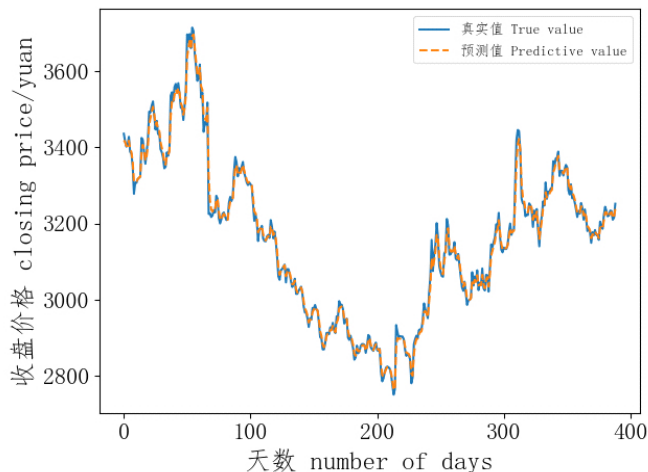
从图7(1)中可观察到,LSTM网络模型的预测值与真实值吻合度较为满足预期,且大豆期货价格走势能够较好地还原;从图7(2)中不难发现,与另外3种基准模型相比,LSTM网络模型的预测结果最接近真实值,且该模型拥有较好的预测极值的能力。

四、结论与展望

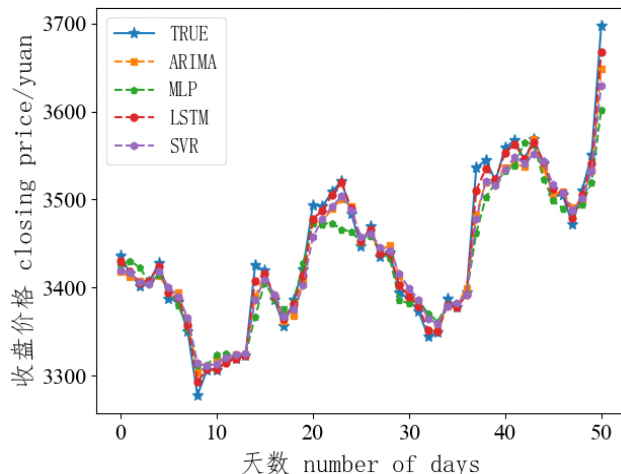
通过实验选取了合适的学习率、迭代次数、窗口大小及隐藏层数等参数,结合Dropout正则技术,构建了一个较优化的多层LSTM网络模型,将其应用于中国大豆期货价格预测分析,并与引言中提到的

3种基准模型(ARIMA模型、SVR模型、MLP神经网络模型)的预测结果进行了对比,得到以下研究结论:(1)多层LSTM网络模型在大豆期货价格预测的应用方面具有可行性。(2)多层LSTM网络模型具有优秀的预测精度。相较于ARIMA、SVR、MLP模型,由多层LSTM网络模型最优拟合结果预测出来的大豆期货价格,从平均MAE角度看,预测性能分别提高了25.12%、26.67%和49.96%;从平均RMSE角度看,误差分

别降低了37.49%、38.51%和52.49%;从平均MAPE角度看,百分比误差分别降低了24.85%、26.38%和48.95%;从决定系数(R²)数据看,拟合优度分别提高了1.064%、1.674%和2.147%。表明多层LSTM网络模型更适用于期货价格预测,在大豆期货价格预测精度方面显现出全方位优越性。(3)实验证明,参数调节对多层LSTM模型的预测性能影响较大,故合



(1)LSTM 网络最优拟合图



(2)四种模型预测结果对比图

图 7 LSTM 网络最优拟合图及四种模型预测结果对比图

注:图 8(2)由于数值较为接近,为了便于观察各模型曲线,图中只放大展示测试集中前 50 个数据

理选择参数显得尤为重要。此外,鉴于 LSTM 网络强大的自我学习能力、良好的泛化能力及模型的高度可调节性,其模型在金融时间序列预测的发展中具有广阔前景。

关于 LSTM 模型的研究方法,可做进一步拓展:(1)结合自然语言处理(如时政新闻、论坛评论)、图像识别技术(如价格曲线走势图像),挖掘并构建更多维度的期货价格特征;(2)采用组合模型方式,结合其他处理时间序列模型优势,建立一个预测效果更好的模型。

【参考文献】

- [1]李开鹏.中美贸易摩擦背景下大豆价格波动的影响及策略[J].对外经贸实务,2018(11):41-44.
- [2]王锐,陈倬.“十一五”期间我国农产品价格波动的影响因素分析——基于协整和向量自回归模型的实证研究[J].财经论丛,2011(3):8-13.
- [3]吴玉霞,温欣.基于 ARIMA 模型的短期股票价格预测[J].统计与决策,2016(23):83-86.
- [4]于志军,杨善林.基于误差校正的 GARCH 股票价格预测模型[J].中国管理科学,2013(S1):341-345.
- [5]杨青,王晨蔚.基于深度学习 LSTM 神经网络的全球股票指数预测研究[J].统计研究,2019(3):65-77.
- [6]杨建辉,李龙.基于 SVR 的期权价格预测模型[J].系统工程理论与实践,2011(5):848-854.
- [7]Dhar S,Mukherjee T. and Ghoshal A K. Perfor-

mance evaluation of Neural Network approach in financial prediction: Evidence from Indian Market [C]// 2010 International Conference on Communication and Computational Intelligence (INCOCCI 2010).Erode,2010:597-602.

[8]杨海民,潘志松,白玮.时间序列预测方法综述[J].计算机科学,2019(1):21-28.

[9]Basak S,Kar S, Saha S, et al. Predicting the direction of stock market prices using tree-based classifiers[J]. The North American Journal of Economics and Finance, 2019(47):552-567.

[10]Moghaddam A H,Moghaddam M H, Esfandiyari M. Stock market index prediction using artificial neural network [J].Journal of Economics Finance & Administrative Science,2016(41):89-93.

[11]魏文轩.改进型 RBF 神经网络在股票市场预测中的应用[J].统计与决策,2013(15):70-72.

[12]Lecun Y, Bengio Y,Hinton G.Deep learning[J]. Nature,2015(521):436-444.

[13]Hochreiter S, Schmidhuber J.Long Short-Term Memory[J].Neural Computation, 1997(8):1735-1780.

[14]Graves A. Supervised Sequence Labelling with Recurrent Neural Networks[M].Springer Berlin Heidelberg,2012.

[15]OlahC. Understanding LSTM Networks[EB/OL].<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>,2015-08.

[16] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting [J]. Journal of Machine Learning Research, 2014(56):1929–1958.

[17] Park S, Kwak N. Analysis on the Dropout Effect in Convolutional Neural Networks [C]// Asian Conference on Computer Vision. Springer, Cham, 2016:189–204.

[18] Zhang Y Q, Wang X Y. A new image encryption algorithm based on non-adjacent coupled map lattices[J]. Applied Soft Computing, 2015(26):10–20.

[19] Chen K, Zhou Y, Dai F. A LSTM-based method for stock returns prediction: A case study of China stock market[C]//2015 IEEE International Conference on Big Data (Big Data). Santa Clara, CA, 2015:2823–2824.

[20] 彭燕, 刘宇红, 张荣芬. 基于 LSTM 的股票价格预测建模与分析[J]. 计算机工程与应用, 2019(11):209–212.

[21] 夏克文, 李昌彪, 沈钧毅. 前向神经网络隐含层节点数的一种优化算法[J]. 计算机科学, 2005(10):143–145.

Soybean future prices forecasting based on LSTM deep learning

FAN Jun-ming, LIU Hong-jiu, HU Yan-rong

(Zhejiang A&F University, Hangzhou, Zhejiang 311300)

Abstract: Faced with an increasingly complex financial market environment, time series forecasting models based on traditional statistics and metrology have certain limitations in discovering long-term dependencies in the series, but Long Short-Term Memory (LSTM) in deep learning is expected to overcome this problem. By constructing a multi-layer LSTM prices forecasting model, this paper made an empirical study on the China's 2007–2019 soybean future prices data. The results show that the parameter optimization has a great impact on the forecasting effect of the LSTM model, and the main parameters that have a greater impact include the number of iterations, learning rate, window size and the number of network layers. Compared with ARIMA model, MLP model and SVR model, the forecasting results of the LSTM model are more accurate, and the goodness of fit (R^2) is improved by 1.064%, 2.147% and 1.674%, respectively.

Key words: deep learning; LSTM; prices forecasting; soybean future