文章编号：2095-2163(2022)04-0054-08 中图分类号：TP391 文献标志码：A

基于模糊 K 线的 FCLSTM-vSVR 模型的股票价格预测

刘茜阳[1]，宋 燕[2]，张亚萌[2]
（1 上海理工大学 理学院，上海 200093；2 上海理工大学 光电信息与计算机工程学院，上海 200093）

摘 要：股票市场具有不确定性和非线性等特点，因此准确地预测股票价格对投资者来说是一项重大挑战。现有的股价预测模型较为单一，预测精度不高。针对这一问题，提出一种基于模糊 K 线的长短期记忆（LSTM）网络和支持向量回归多阶段混合模型（FCLSTM-vSVR）。研究第一阶段，基于遗传算法对 LSTM 网络进行参数寻优，找到时间窗口和隐藏层神经元的最佳值，并利用训练好的 LSTM 进行股票价格初步预测，计算出股票价格的残差值。第二阶段，利用模糊 K 线将原始价格序列转换为模糊数据，并作为 vSVR 模型的输入，利用 vSVR 模型预测残差值。综合前文论述后，再将两阶段的预测值之和作为最终的股票价格预测值。通过对比实验得出，该模型具有更高的预测准确率，在股票价格的一步预测方面优于其他对比模型。

关键词：LSTM 神经网络；vSVR 模型；模糊 K 线；残差预测；股票价格预测

Stock price forecasting using FCLSTM-vSVR based on fuzzy Candlestick

LIU Xiyang[1], SONG Yan[2], ZHANG Yameng[2]
(1 College of Science, University of Shanghai for Science and Technology, Shanghai 200093, China; 2 School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

【Abstract】 Due to the uncertainty and nonlinearity of stock market, accurately predicting stock prices is always a major challenge for investors. The existing stock price prediction models are relatively single and the prediction accuracy is undesirable. In order to solve this problem, a multi-stage hybrid model of long short-term memory (LSTM) networks and support vector regression based on the fuzzy Candlestick, namely FCLSTM-vSVR, is presented. In the first stage, some parameters of LSTM network are optimized based on Genetic Algorithm. The stock price is preliminarily predicted by LSTM, and the residual values of stock price are calculated. In the second stage, original price series are transformed to ambiguous outputs by applying fuzzy Candlestick. Then the fuzzy outputs are used as the inputs of vSVR model to predict the residual values. Finally, the final forecast values of stock price are taken by summing the predicted values of the two stages. Experimental results show that the model presented has higher prediction accuracy and is superior to the baseline model in one-step prediction of stock price.

【Key words】 LSTM network; vSVR; fuzzy Candlestick; residual prediction; stock price forecasting

0 引 言

金融数据是现代经济学不可分割的一部分。在金融市场中，由于具有较高的回报率，股票市场成为最热门的投资领域。股票价格反映了公司的经营状况，为投资者提供了重要的投资参考。为了使利益最大化，降低投资风险，投资者有必要对股票价格进行预测。然而，由于股票市场受到国外市场行情、时事生态和投资者行为及心理等多方面因素的影响，其数据呈现非线性和非平稳特征，这使得预测更具挑战性[1]。因此，长期以来，股票价格预测一直是金融学者研究的重点。

目前，股票预测方法主要包括基本分析法、技术分析法、组合分析法、时间序列分析法、机器学习和神经网络等几大类[2]。传统的股票市场预测技术主要是基于历史股票数据的统计分析，如自回归综合移动平均模型（ARIMA）、自回归条件异方差模型（ARCH）和广义自回归条件异方差模型（GARCH）模型，已被广泛用于金融市场的预测中[3~6]。但由于股票自身的非平稳与非线性特征，这些统计方法并不能在预测时达到较好的效果。

近几年，随着人工智能领域的发展，机器学习方法在股票市场预测中被广泛应用，并取得了一定的研究成果。其中，人工神经网络（ANN）和支持向量回归（SVR）是预测金融时间序列流行的技术，因为不需要做任何的统计假设条件，可以直接提取数据间的非线性关系[7-8]。同时，在小样本预测方面，与 ANN 方法相比，SVR 不容易陷入局部最优，因此有

作者简介：刘茜阳（1996-），女，硕士研究生，主要研究方向：人工智能、大数据分析；宋 燕（1979-），女，博士，教授，主要研究方向：大数据算法、图像处理、预测控制；张亚萌（1992-），女，博士研究生，主要研究方向：人工智能、大数据分析。
通讯作者：宋 燕 Email：sonya_usst@163.com

收稿日期：2021-11-27

较大的优越性[9]。但这些方法在处理输入数据时并不能捕获序列数据的顺序信息，对时间序列问题没有优秀的泛化能力，所以预测效果仍然受到了一些限制。

循环神经网络(RNN)解决了这一问题，因其具备了时序概念，对股票的预测性能更好，但 RNN 在训练时往往会出现梯度消失或梯度爆炸的问题，这导致时间序列的长期依赖关系很难学习[10]。1997 年，Hochreiter 等人[11]在论文《Long Short-Term Memory》中，针对 RNN 不能解决数据的长序依赖的问题进行研究并提出了 LSTM 模型。但是此 LSTM 的记忆存储会随序列长度的延伸而增长，最终可能会导致网络崩溃，因此，在 2000 年针对该问题，Felix 等人[12]在 LSTM 神经元内部增加了遗忘门，使数据在传输时可以保持长时记忆。因此 LSTM 神经网络被越来越多地应用到金融时间序列的预测中。

此外，基于 LSTM 神经网络的混合方法也被广泛应用于金融时间序列分析中，并通过与单一模型方法相比获得更高精度的预测结果[13-14]。然而，大多数混合方法虽然在一定程度上提高了预测精度，但这些方法都是通过对 LSTM 预测模型的输入进行分析，而没有对于 LSTM 模型产生的残差进行分析预测。

基于上述问题，本文提出了一种将遗传算法、LSTM 网络、模糊 K 线和改进的支持向量回归算法(vSVR)相结合的混合股价预测模型。第一阶段，该模型首先利用遗传算法对 LSTM 神经网络的参数进行优化，然后用训练好的 LSTM 网络产生预测输出。第二阶段，基于模糊 K 线模型提取到的模糊信息，采用 vSVR 模型预测误差。最后，将两阶段的预测值之和作为最终的股票价格预测值。本文的贡献主要如下：

(1)对于 LSTM 网络的部分参数、如时间窗口大小和结构参数的估计，以往通常采用试错法，但效率较低[15]，本文利用遗传算法优化 LSTM 网络，以选择最佳的窗口大小、神经元数目。

(2)由于股票价格序列可以由 K 线表示，而且会受到多方面因素的影响，本文利用模糊理论将股票价格数据转换为模糊数据，采用 vSVR 模型建立预测误差与股票模糊信息之间的映射关系，减小了模型的固有误差。

(3)实验结果显示本文提出模型的预测效果更好，拟合程度更优。

1　背景知识

1.1　LSTM 神经网络

近年来，LSTM 神经网络已陆续应用于时间序列预测、文本分类等领域中，具有较高的精度[13]。LSTM 单元结构如图 1 所示。由图 1 可知，LSTM 单元控制门结构主要包含遗忘门、输入门与输出门。遗忘门 f_t 控制细胞在 $t-1$ 时刻的状态有多少信息会被遗忘，输入门 i_t 决定有多少新信息将被保存到 t 时刻的细胞状态，输出门 o_t 决定输出新细胞状态的信息。其计算过程如下：

$$f_t = \sigma(\boldsymbol{W}_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \sigma(\boldsymbol{W}_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = \tanh(\boldsymbol{W}_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

$$C_t = C_{t-1} \circ f_i + i_t \circ \tilde{C}_t \tag{4}$$

$$o_t = \sigma(\boldsymbol{W}_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t \circ \tanh(C_t) \tag{6}$$

其中，C_{t-1}，C_t 分别表示 $t-1$ 和 t 时刻的单元状态；$\tilde{C}_t$ 为 t 时刻的状态候选值；$\boldsymbol{W}_f$，$\boldsymbol{W}_i$，$\boldsymbol{W}_c$，$\boldsymbol{W}_o$ 为各项门的权重矩阵；h_{t-1} 和 h_t 分别为 $t-1$ 和 t 时刻的输出值；x_t 为 t 时刻的输入值；b_f，b_i，b_c，b_o 为偏置项；$\sigma(*)$ 和 $\tanh(*)$ 为激活函数；“$\circ$”表示向量元素乘积。

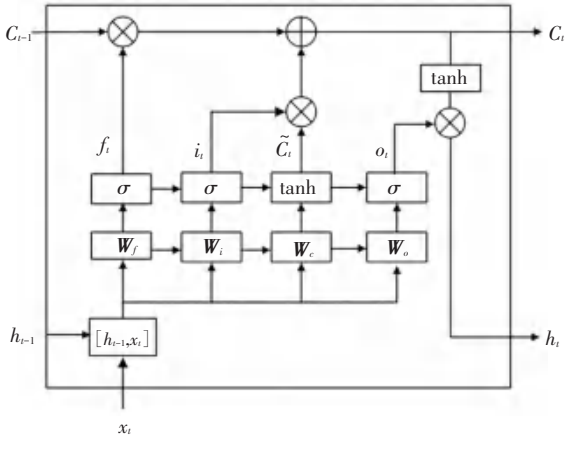


图 1　LSTM 单元结构

Fig. 1　The LSTM cell structure

当使用 LSTM 网络进行股票价格预测时，前期多个时刻的收盘价将作为输入，当前时刻的收盘价作为输出。对于 LSTM 模型预测而言，时间窗口起着较为重要的作用，每层神经元数也是 LSTM 模型优化的重要参数。本文利用遗传算法对 LSTM 模型进行优化，得到时间窗口和隐藏层神经元数的最佳值或接近最佳值。

1.2 vSVR 模型

SVR 是支持向量机(SVM)在回归问题领域的推广,其中 vSVR 模型是一种改进的 SVR,新参数 v 控制误差分数的上界和支持向量分数的下界,并自动最小化误差参数 ε,这使得更容易通过手动校准来调整参数。对于给定的一组数据点 $\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$,其中 $x_i \in R^n$ 是输入,$y_i \in R$ 是目标输出,vSVR 模型使用核函数将不可分割的输入数据 x_i 映射到高维空间,从而使得目标值与训练数据得到的回归函数 $f(x) = (w,\varphi(x)) + b$ 的距离最小,即极小化目标函数[16]:

$$\min \frac{1}{2}\|w\|^2 + C\cdot(v\varepsilon + \frac{1}{N}\sum_{i=1}^{N}(\xi_i + \xi_i^*)) \quad (7)$$

$$\text{s.t.}\begin{cases} y_i - ((w_i,\varphi(x_i)) + b) \leqslant \varepsilon + \xi_i \\ ((w_i,\varphi(x_i)) + b) - y_i \leqslant \varepsilon + \xi_i^* \\ \varepsilon \geqslant 0,\ \xi_i \geqslant 0,\ \xi_i^* \geqslant 0,\ i = 1,2,\cdots,N \end{cases} \quad (8)$$

其中,常数 v 表示错误样本个数占总样本个数份额的上界或支持向量与总样本数比值的下界;$\varphi(x)$ 为核函数;w 和 b 分别表示权重和偏置;ε 表示误差;ξ_i 和 ξ_i^* 为松弛变量;C 为惩罚参数。

其对偶问题为:

$$\min_{a,a^*} \frac{1}{2}\sum_{i,j=1}^{N}(a_i - a_i^*)(a_j - a_j^*)k(x_i,x_j) - \sum_{i=1}^{N}(a_i - a_i^*)y_i \quad (9)$$

$$\text{s.t.}\begin{cases} \sum_{i=1}^{N}(a_i - a_i^*) = 0 \\ a_i, a_i^* \in [0,\frac{C}{N}],\ i = 1,2,\cdots,N \\ \sum_{i=1}^{N}(a_i + a_i^*) \leqslant Cv \end{cases} \quad (10)$$

其中,a 和 a^* 为拉格朗日乘子,且其值都不为 0,$k(x_i,x_j)$ 为核函数。决策函数变为:

$$f(x) = \sum_{i=1}^{N}(a_i - a_i^*)k(x_i,x) + b \quad (11)$$

相对于传统的经济学模型和基本的机器学习方法,单一的 LSTM 神经网络的预测精度虽然有所提高,但这并不能满足人们的需求。针对这个问题,本文利用 vSVR 模型进行残差分析来提高 LSTM 神经网络的预测精度。

2 基于模糊 K 线和遗传算法的 FCLSTM-vSVR 股票价格预测方法

2.1 模糊 K 线

研究可知,经过不断演变,K 线图现已形成了拥有完整形式和分析理论的技术分析方法。在 K 线图中,影线长度和实体长度在识别 K 线模式上发挥了重要作用。但是,在日常生活中,人们对于 K 线的描述往往难以做到精确,甚至是模糊的,例如长的、中等的或者是短的[17~19]。因此,本文将引入 Naranjo 等人[20]在股票预测中所提到的模糊变量法,即运用模糊集合理论将本文的股票时间序列数据转化为模糊化的数据,并作为 vSVR 模型的输入,LSTM 神经网络的残差作为输出构建 vSVR 模型。

首先,将 K 线的 3 个变量包括上影线、下影线和实体的长度(L_u、L_l 和 L_b)作为输入数据,R_s 和 R_p 作为模糊输出数据,分别表示实体与烛台整体之间的相对大小和相对位置。交易时间 t 中的 3 个模糊输入可以定义如下:

$$L_u(t) = \frac{Hight(t) - \max(Open(t),\ Close(t))}{Open(t)} \quad (12)$$

$$L_l(t) = \frac{\min(Open(t),\ Close(t)) - Low(t)}{Open(t)} \quad (13)$$

$$L_b(t) = \frac{|\ Open(t) - Close(t)\ |}{Open(t)} \quad (14)$$

其中,$Open(t)$,$Close(t)$,$Hight(t)$,$Low(t)$ 分别表示 t 时刻的开盘价、收盘价、最高价和最低价。然后利用式(15)将这 3 个变量缩放到[0,100]之间,数学公式具体如下:

$$L'(t) = \frac{L(t) - L_{\min}}{L_{\max} - L_{\min}} \times 100\% \quad (15)$$

其次,用定义的 4 个模糊语言变量来描述 3 个输入变量:NULL、SHORT、MIDDLE、LONG,如图 2 所示。用 5 个模糊子集来描述 R_p:DOWN、CENTER_DOWN、CENTER、CENTER_UP、UP,如图 3 所示。用 5 个模糊子集来描述输出变量 R_s:LOW、MEDIUM_LOW、MEDIUM、MEDIUM_EQUAL、EQUAL,如图 4 所示。

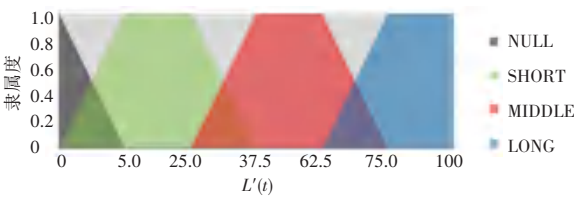


图 2 L_u、L_l 和 L_b 的隶属度函数

Fig. 2 Membership function for L_u、L_l and L_b

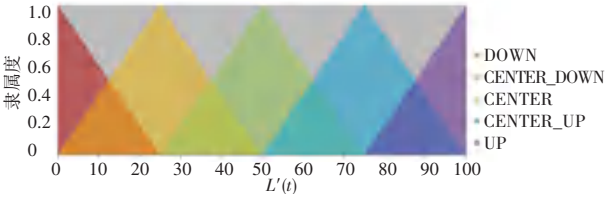


图 3　R_p 的隶属度函数

Fig. 3　Membership function for R_p

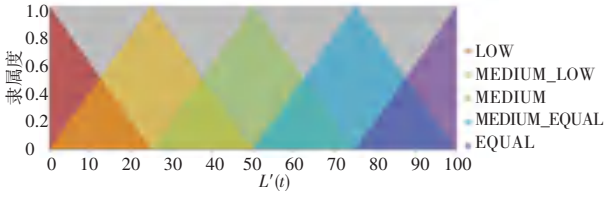


图 4　R_s 的隶属度函数

Fig. 4　Membership function for R_s

最后，基于 128 条 IF-THEN 模糊规则，详见表 1 中的部分模糊规则[20]，并用质心法去模糊化得到输出数据。

表 1　模糊规则

Tab. 1　Fuzzy rules

L_u	L_l	L_b	R_s	R_p
LONG	LONG	LONG	MEDIUM_EQUAL	CENTER
LONG	LONG	MIDDLE	MEDIUM	CENTER
LONG	LONG	SHORT	MEDIUM_LOW	CENTER
LONG	LONG	NULL	LOW	CENTER
…	…	…	…	…

2.2　GLSTM 模型

由于 LSTM 网络在学习过程中使用过去的信息，不同的时间窗口会对模型学习性能的提高起不同的作用。窗口过小，模型会忽略重要信息；窗口过大，模型会对训练数据过度拟合。所以将遗传算法用于优化 LSTM 模型，以选择最佳的窗口大小、神经元数目，即 GLSTM 模型。图 5 给出了 GLSTM 模型的流程图。

在上述的 GLSTM 模型的流程图中，LSTM 网络的结构如图 6 所示，主要包括一个输入层，一个 LSTM 层和一个输出层(完全连接层)。其中，l 表示时间窗口的大小，P_t 表示 t 时刻的收盘价，$\hat{y}_t$ 表示 LSTM 网络模型的预测值。

2.3　FCLSTM-vSVR 模型

由于 GLSTM 网络模型较为单一，本文提出了一种基于模糊 K 线的 FCLSTM-vSVR 模型的股票价格预测方法。其中，主模型是基于遗传算法的 LSTM 网络模型，次模型是 vSVR 模型。

对于主模型，基于不同的参数设置，LSTM 网络模型的预测结果和性能会有所不同。所以本文首先通过遗传算法进行参数寻优，找到模型的最佳时间窗口大小和隐藏层单元数，如 2.2 节所示。

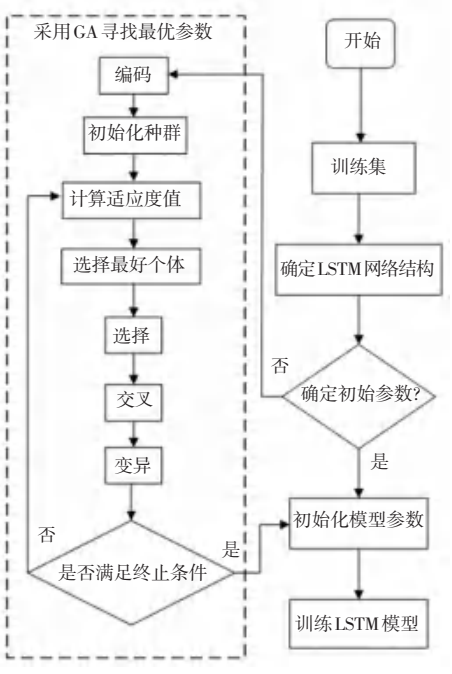


图 5　GLSTM 模型流程图

Fig. 5　The flowchart of the GLSTM model

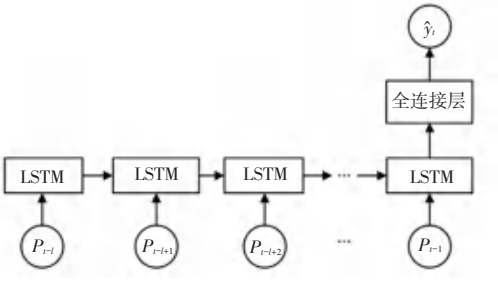


图 6　LSTM 模型结构

Fig. 6　The structure of the LSTM model

对于次模型，vSVR 模型是一种不依赖于任何先验知识的机器学习非线性回归方法，该模型参数具有明确意义，更利于得到精确的回归解，因此在小样本预测方面具有显著优势。图 7 描述了FCLSTM-vSVR 模型的流程图。

图 7 中，FCLSTM-vSVR 残差修正模型的处理流程如下：

(1)利用数据集原始价格数据建立模糊烛台模型，得到 2 个模糊输出数据 R_s 和 R_p。然后将数据集拆分为训练集和测试集，并利用最大-最小标准化公式进行归一化处理。

(2)将训练集部分划分为验证集，利用训练集建立 GLSTM 预测模型：

① 使用二进制位编码表示时间窗的大小和

LSTM 神经单元数。

② 随机生成初始种群,根据适应度函数和选择进行评估,然后进行交叉和变异,使用赌轮盘选择。在本文中使用均方误差(*MSE*)来计算每个染色体的适应度,输入 LSTM 模型,在验证集上计算 *MSE*,并返回该值将其作为当前遗传算法解决方案的适应度值,得出最优时间窗口大小及最优神经网络隐藏层单元数。

③ 重复该过程直至满足终止条件。

(3)将经过良好训练的 GLSTM 预测模型应用于整个训练集的股票价格预测中,得到不同时刻的预测残差值 e_t, 形成历史残差,称为残差集。

(4)利用历史残差作为真实值,2 个模糊输出变量 R_s 和 R_p 作为输入,训练 vSVR 模型。将经过良好训练的 vSVR 残差预测模型应用于预测残差 $\hat{e}_t$。

(5)应用 vSVR 残差模型的残差值 $\hat{e}_t$ 和 LSTM 模型的预测值 $\hat{y}_t$ 得到最终的预测值 $\hat{p}_t$。

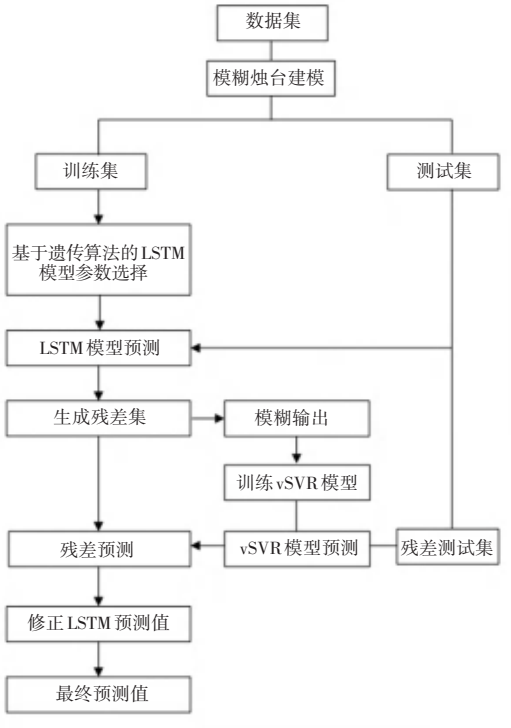


图 7 FCLSTM-vSVR 模型流程图

Fig. 7 The flowchart of the FCLSTM-vSVR model

3 实验

为了验证实验结果,本文将设置 GLSTM 模型、vSVR 模型、BPNN 模型、ARIMA 模型、GLSTM-BPNN 模型作为对照,其中 GLSTM 模型的输入为收盘价历史数据,并对 ARIMA 模型使用网格搜索法确定最优参数。本节通过股票历史数据验证了 FCLSTM-vSVR 模型的优越性和稳健性,也对 FCLSTM-vSVR 模型和其他模型进行了性能比较。所有模型都由 Python3.6 和 TensorFlow 实现,电脑的操作系统 Windows 10,处理器是英特尔 i7-7820HQ (2.90 千兆赫)。模型的主要最佳参数是通过在验证集上不断试验和参考相关文献确定的[15],见表 2。

表 2 模型的部分参数设置

Tab. 2 Parameters setting of the model

模型	参数设置
LSTM	训练次数:100;批量大小:32; *dropout*:0.2
BPNN	层数:2; 第一层神经元:128; 第二层神经元:64; 训练次数:50; 批量大小:50; *dropout*: 0.2
vSVR	核函数:径向基函数; 惩罚参数 C:1.0; 参数 v:0.1; 参数 *gamma*:0.1
遗传算法	种群数量:20; 迭代次数:10; 交叉概率:0.7; 变异概率:0.15

3.1 数据来源

本文使用从雅虎财经网站获得的 5 只股票数据集:AAPL、ADI、WTI、GSPC、IXIC。每只股票数据集的时间跨度为 2015 年 1 月 4 日至 2020 年 10 月 22 日。数据集划分为训练集和测试集,其中训练集为 80%,测试集为 20%。这里,训练集中选择 20%数据作为验证集来确定超参数。

3.2 评价指标

为了量化模型的性能, 引入了 5 个指标: 平均绝对误差(*MAE*)、平均绝对百分比误差(*MAPE*)、均方误差(*MSE*)、均方根误差(*RMSE*)、决定系数(R^2),这些指标可以定义如下:

$$MAE = \frac{1}{n}\sum_{t=1}^{n} | P_t - \hat{P}_t | \tag{16}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left| \frac{P_t - \hat{P}_t}{P_t} \right| \times 100\% \tag{17}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (P_t - \hat{P}_t)^2} \tag{18}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n} (P_t - \hat{P}_t)^2}{\sum_{t=1}^{n} (P_t - \bar{P})^2} \tag{19}$$

其中, n 是时间序列的长度;P_t 是实际值;$\hat{P}_t$ 是预测值;$\bar{P}$ 是平均值 ($t = 1,2,\cdots,n$)。

3.3 结果分析

首先通过遗传算法,本文得到 LSTM 网络的最优结构因素,包括 LSTM 网络的时间窗口大小和隐

藏层的单元数。其中，时间窗口大小取值范围为[5,30]，隐藏层单元数取值范围为[10,100]。表 3 显示了 GLSTM 模型在各个数据集上的训练得到的参数结果。例如股票 AAPL 预测的最佳时间窗口大小为 6，最佳 LSTM 隐藏层单元数为 80。也就是说，对于股票 AAPL，利用过去 6 个交易日的信息来分析预测是最有效的。

表 3　GLSTM 模型的时间窗口大小和隐藏层单元数

Tab. 3　Time window size and hidden layer units of GLSTM model

数据集	时间窗口	隐藏层单元数
AAPL	6	80
ADI	6	40
WTI	8	43
GSPC	5	75
IXIC	11	38

在得到模型最佳参数基础上，利用提出 FCLSTM-vSVR 预测模型对各个股票数据集进行股价预测，并与其他模型进行对比。表 4~表 8 给出了不同模型在各个数据集上的平均预测误差。对于单一模型，传统的时间序列模型 ARIMA 的预测效果最差。机器学习 vSVR 模型的预测效果与 BPNN 模型的效果相近，而 GLSTM 模型的预测效果最好。这表明 GLSTM 模型在具有相同的输入变量下可以获得相比于普通机器学习的模型 vSVR 和 BPNN 更好的性能。所以在股票预测中，GLSTM 网络模型是一种较好的预测方法。

表 4　AAPL 上各模型比较结果

Tab. 4　The evaluation results of different models on AAPL

模型	R^2	*RMSE*	*MAPE*	*MAE*
ARIMA	0.319	16.346	19.783	13.446
vSVR	0.964	3.808	3.046	2.728
BPNN	0.955	4.262	3.555	3.232
GLSTM	0.974	3.247	2.652	2.334
GLSTM-BPNN	0.958	4.107	3.443	2.888
FCLSTM-vSVR	**0.974**	**3.224**	**2.640**	**2.329**

表 5　ADI 上各模型比较结果

Tab. 5　The evaluation results of different models on ADI

模型	R^2	*RMSE*	*MAPE*	*MAE*
ARIMA	-1.769	16.403	13.521	14.021
vSVR	0.835	4.048	2.854	3.093
BPNN	0.851	3.895	2.680	2.886
GLSTM	0.870	3.593	2.441	2.619
GLSTM-BPNN	0.845	3.916	2.586	2.733
FCLSTM-vSVR	**0.876**	**3.512**	**2.382**	**2.558**

表 6　WTI 上各模型比较结果

Tab. 6　The evaluation results of different models on WTI

模型	R^2	*RMSE*	*MAPE*	*MAE*
ARIMA	-15.892	3.114	140.180	2.993
vSVR	0.851	0.244	8.842	0.186
BPNN	0.898	0.202	6.585	0.147
GLSTM	0.917	0.183	6.080	0.134
GLSTM-BPNN	0.878	0.221	7.938	0.171
FCLSTM-vSVR	**0.928**	**0.170**	**5.560**	**0.121**

表 7　GSPC 上各模型比较结果

Tab. 7　The evaluation results of different models on GSPC

模型	R^2	*RMSE*	*MAPE*	*MAE*
ARIMA	-1.104	422.567	12.184	338.167
vSVR	0.912	87.155	2.263	67.621
BPNN	0.921	82.688	2.019	59.435
GLSTM	0.915	85.235	2.172	64.878
GLSTM-BPNN	0.923	81.553	**1.925**	**56.305**
FCLSTM-vSVR	**0.925**	**80.271**	2.016	59.848

表 8　IXIC 上各模型比较结果

Tab. 8　The evaluation results of different models on IXIC

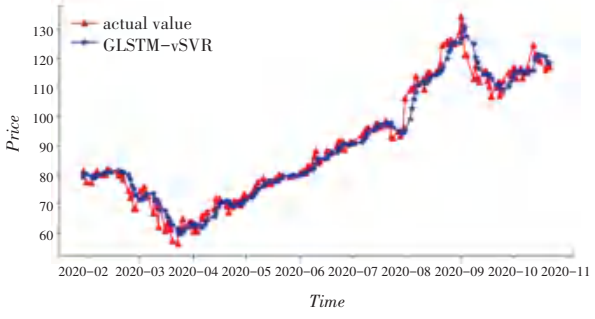
模型	R^2	*RMSE*	*MAPE*	*MAE*
ARIMA	0.366	1 014.204	12.093	785.554
vSVR	0.931	342.198	2.896	270.081
BPNN	0.931	342.186	2.902	269.034
GLSTM	0.946	304.324	2.599	244.293
GLSTM-BPNN	0.934	334.584	2.532	228.303
FCLSTM-vSVR	**0.952**	**286.123**	**2.422**	**225.780**

对于混合模型 GLSTM-BPNN，在股票 GSPC 上的结果比其他单一模型的预测效果好，R^2、*RMSE*、*MAPE*、*MAE* 分别为 0.923 23、81.552 46、1.924 91、56.304 93，与 GLSTM 模型相比，精度分别提高了 0.86%、4.32%、11.39%、13.21%。但在其他数据集上，模型 GLSTM-BPNN 结果并不理想，这可能与神经网络参数设置不合理有关。

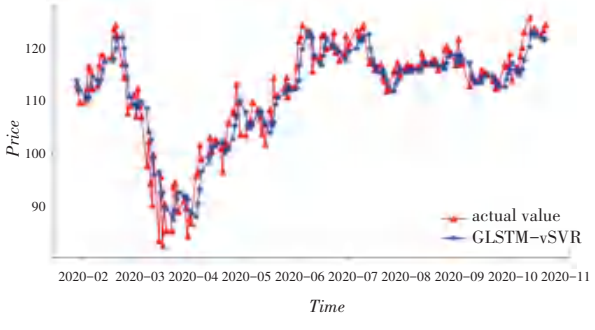
所有数据集中，与 GLSTM 模型相比，FCLSTM-vSVR 的误差指标均保持较低值，R^2 也更接近于 1，拟合效果更好。这一结果的主要原因是增加了残差分析。结果表明，残差分析是一种能显著提高股价预测精度的方法，残差中具有重要的信息价值，值得进行深入研究。基于模糊烛台建模，2 个模糊输出变量（R_s 和 R_p）在残差模型中成功应用，这一结果表明，在数据中的模糊信息对于残差序列也具有影响。与 GLSTM-BPNN 模型相比，FCLSTM-vSVR 的所有指标结果都较好，这表明 vSVR 模型与 BPNN

相比,对于小样本具有更精确的回归解。所以,对于股票价格预测,FCLSTM-vSVR 模型与其他模型相比具有更好的预测效果。

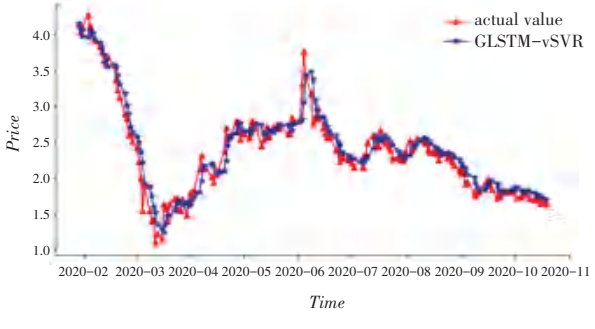
为了更好地观察模型 GLATM-vSVR 的性能,在测试集上实际值对于模型预测值的拟合效果如图 8 所示。



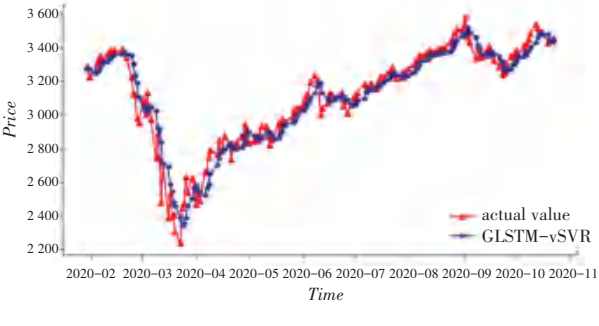
(a) AAPL



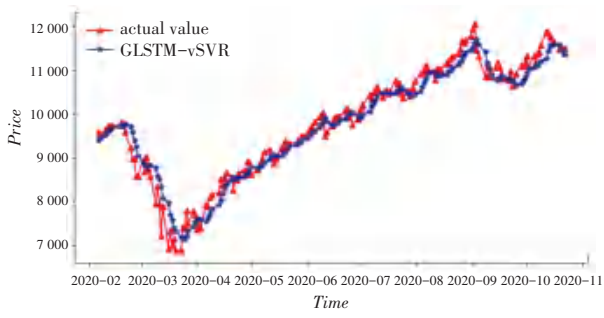
(b) ADI



(c) WTI



(d) GSPC



(e) IXIC

图 8 收盘价真实值与 FCLSTM-vSVR 模型的预测值对比

Fig. 8 The comparison of the actual closing price values with the predicted value of FCLSTM-vSVR model

由图 8 可知,提出 FCLSTM-vSVR 模型得到的预测值与真实值较为接近。但由于市场环境和投资人行为等因素的存在,所以零误差的预测无法实现。但本文提出模型与其他对照模型相比,取得了较为理想的预测效果。

4 结束语

对股票市场的预测可以产生实际的盈利或亏损,因此提高模型的可预测性对投资者来说是非常重要的。在本文中,提出了一种新的基于模糊 K 线的混合股票价格预测模型,即 FCLSTM-vSVR 模型。具体来说,首先本文将遗传算法和 LSTM 网络结合起来考虑股票市场的时间特性和模型的自定义结构因素。利用遗传算法搜索时间窗口大小和神经网络隐藏层单元数的最优或接近最优值。然后,提出了一种降低预测误差的方法来提高 GLSTM 模型的预测精度。该方法采用模糊 K 线模型来表示原始价格序列中的模糊信息,vSVR 模型建立预测误差与模糊输出因素之间的映射关系。本文选取 5 个股票数据集,通过对比试验,验证了该模型的可行性。实验结果表明,与基线模型相比,所提出的模型具有更高的预测精度。但使用遗传算法寻找 LSTM 神经网络模型参数所需时间与计算资源较大。因此,寻找一种更简单的方法来优化 LSTM 网络参数将是下一步研究的方向。

参考文献

[1] HUANG Yusheng, GAO Yelin, GAN Yan, et al. A new financial data forecasting model using genetic algorithm and long short-term memory network[J]. Neurocomputing, 2021,425:207-218.

[2] 武大硕,张传雷,陈佳,等. 基于遗传算法改进 LSTM 神经网络股指预测分析[J]. 计算机应用研究, 2020, 37(S1): 86-87, 107.

(下转第 69 页)