

## 基于深度学习的中国黄金期货价格区间预测

李雨涵<sup>1</sup>, 孙景云<sup>1,2</sup>

1. 兰州财经大学 统计与数据科学学院, 兰州 730020

2. 甘肃省经济发展数量分析研究中心, 兰州 730020

**摘要:**目的 为解决中国黄金期货价格由于固有的非线性、不稳定和高波动性特征而无法精确预测的问题,提出了一个基于深度学习的区间预测模型,以更精确地描绘和预测黄金期货价格的变化趋势。方法 首先,采用门控循环单元(GRU)、长短期记忆(LSTM)和双向长短期记忆(BiLSTM)3种深度学习模型,分别通过核密度估计(KDE)和分位数回归(QR)方法构建独立的区间预测模型;然后,基于这两种区间预测模型,进一步整合成一个组合预测模型,以预测区间覆盖率作为约束条件,并利用网格搜索算法(GS)来优化模型权重配置,以确保预测区间的平均宽度达到最小。结果 实证分析显示:采用BiLSTM方法构建的组合区间预测模型在预测性能上显著优于其他基准模型,在各种置信水平下,都显示出更广泛的预测覆盖率和更窄的区间平均宽度。结论 所提出的区间预测组合模型能够有效预测未来黄金期货价格的波动范围,为投资者提供更为可靠的风险评估和决策支持。

**关键词:**黄金期货;深度学习;核密度估计;分位数回归;区间预测

**中图分类号:**TP18;F832.54;F832.5 **文献标识码:**A **doi:**10.16055/j.issn.1672-058X.2026.0003.016

### Interval Prediction of Chinese Gold Futures Prices Based on Deep Learning

LI Yuhan<sup>1</sup>, SUN Jingyun<sup>1,2</sup>

1. School of Statistics and Data Science, Lanzhou University of Finance and Economics, Lanzhou 730020, China

2. Gansu Research Center for Quantitative Analysis of Economic Development, Lanzhou 730020, China

**Abstract: Objective** To tackle the problem that the Chinese gold futures prices cannot be accurately predicted due to their inherent characteristics of non-linearity, instability, and high volatility, an interval prediction model based on deep learning is proposed. This model aims to more precisely depict and forecast the changing trends of gold futures prices. **Methods** First, three deep-learning models, namely the gated recurrent unit (GRU), long short-term memory (LSTM), and bidirectional long short-term memory (BiLSTM), were employed. Independent interval prediction models were constructed using the kernel density estimation (KDE) and quantile regression (QR) methods respectively. Then, based on these two types of interval prediction models, a combined prediction model was further integrated. Taking the prediction interval coverage rate as a constraint condition, the grid search(GS) algorithm was used to optimize the model weight configuration to minimize the average width of the prediction interval. **Results** Empirical analysis showed that the combined interval prediction model constructed using the BiLSTM method significantly outperformed other benchmark models in terms of prediction performance. At various confidence levels, it exhibited a wider prediction coverage rate and a narrower average interval width. **Conclusion** The proposed combined interval prediction model can more effectively predict the future fluctuation range of gold futures prices, providing more reliable risk assessment and decision-making support for investors.

**Keywords:** gold futures; deep learning; kernel density estimation; quantile regression; interval prediction

收稿日期:2024-05-22 修回日期:2024-09-20 文章编号:1672-058X(2026)03-0135-09

基金项目:国家自然科学基金项目资助(72061020).

作者简介:李雨涵(2001—),女,安徽阜阳人,硕士研究生,从事机器学习及区间预测研究. Email:17775305790@163.com.

通信作者:孙景云(1982—),男,甘肃临潭人,博士,教授,从事组合预测及文本挖掘研究. Email:sunjy@lzufe.edu.cn.

引用格式:李雨涵,孙景云.基于深度学习的中国黄金期货价格区间预测[J].重庆工商大学学报(自然科学版),2026,43(3):135-143.

Li Yuhan, Sun Jingyun. Interval prediction of Chinese gold futures prices based on deep learning[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(3): 135-143.

黄金是一种特别的贵金属,它同时拥有货币和金融的双重属性,具有高流动性、高安全性和低收益性的特点。近年来,世界经济市场动荡多变,由于黄金特殊的双重属性,具有一定的规避风险作用。因此,黄金期货已成为一种有效的投资手段,研究和预测黄金期货价格具有重要的现实意义。

目前,黄金期货价格预测的研究主要围绕点预测开展,传统的统计模型可用于数据预测和推测,它们依赖完整且丰富的历史统计数据,通过数学方法系统地处理这些数据,以探索不同变量间的规律性联系。统计模型主要涵盖自回归移动平均模型(ARMA)<sup>[1]</sup>、差分整合移动平均自回归模型(ARIMA)<sup>[2]</sup>、广义自回归条件异方差模型(GARCH)<sup>[3]</sup>等。然而,当面临非线性、非平稳的时序数据时,这些传统模型的有效性便受到了挑战,它们往往难以精确分析并揭示其中的复杂动态结构。

传统时序模型在进行预测时往往需要对数据的分布进行假设,这在一定程度上限制了其预测的准确性。相对而言,机器学习模型则无需对总体进行分布假设,且所需估计的参数较少,计算效率更高。因此,越来越多的学者开始采用机器学习方法构建预测模型,并成功获得了良好的预测效果。比如,Guo等<sup>[4]</sup>采用支持向量回归机(SVR)来进行股票价格数据的预测,并通过粒子群优化算法实时调整学习参数,以获得更准确的预测结果;Yang等<sup>[5]</sup>结合人工神经网络(ANN)、长短期记忆(LSTM)和支持向量回归3种机器学习算法,并依据传统指数、新兴指标、大宗商品以及黄金的历史价格时间序列,对黄金价格进行预测。

需要注意的是,黄金期货价格呈现出非常复杂的非线性趋势,传统的线性模型预测方法在使用时通常存在各种困难,普通的神经网络模型在时间序列预测方面也存在许多局限性和缺点。近年来,深度学习理论发展迅速,其非线性计算和函数拟合能力显著超越浅层机器学习模型<sup>[6]</sup>,并已成功应用于许多领域<sup>[7-8]</sup>。尤其在金融时间序列预测的处理方面,深度学习展现出了卓越的能力。例如,李梦等<sup>[9]</sup>在预测股价时,采用了小波-长短记忆网络(LSTM-Wavelet)模型,充分发挥了LSTM模型处理长期依赖关系的优势,并深入挖掘小波系数中的潜在信息,从而显著提升了股价预测的精度。

然而,以上研究考虑的都是点预测,缺乏对预测结果波动范围的评估,无法充分捕捉黄金期货数据的随机性。区间预测却能够从本质上解决这个问题,Meira等<sup>[10]</sup>提出,可靠的区间预测比单点预测更重要,因为区间预测量化了预测中的不确定性,旨在多个置信水平下,将数据的不确定性限制在可控区间内。

在现有文献中,区间预测的主流思路通常遵循的

流程是首先进行点预测,随后基于这些点预测结果,进一步构建出预测区间。在这一过程中,研究者们主要运用模糊信息粒化(Fuzzy Information Granulation, FIG)、核密度估计(Kernel Density Estimation, KDE)等方法。譬如,熊景华等<sup>[11]</sup>在汇率预测的研究中,采用随机森林模型整合3种模型,构建组合预测模型,并通过模糊信息粒化将点预测拓展为区间预测;郎伟明等<sup>[12]</sup>采用渐进积分均方误差准则的窗宽优化策略,以确定在多个置信水平下,风电功率波动范围的非参数核密度估计;Li等<sup>[13]</sup>采用非参数核密度估计方法来扩大预测区间,并使用速度受限的多目标粒子群优化(MOPSO)对区间的上下界进行调整。但这种区间预测方法是基于点预测的结果,结合概率计算方法来产生轨迹点的分布概率,从而使其过度依赖于点预测模型的性能。

为了克服这些限制,分位数回归(Quantile Regression, QR)作为一种更为灵活和高效的预测方法,逐渐受到学者们的青睐。分位数回归可以在不对分布进行任何假设的情况下提供一组不同分位数的值,这一特性使得它能够更全面地反映数据的分布情况。同时,分位数回归模型还可以与简单的线性模型或复杂的机器学习模型结合,进一步提升其预测性能。譬如,Taylor<sup>[14]</sup>提出了分位数回归神经网络(QRNN),它以每个分位数的弹球损失为目标函数,指导神经网络的训练,然后通过训练每个分位数的神经网络来获得未来数据的分位数估计。之后,学者们开发了许多基于弹球损失的非线性机器学习方法。例如,Zheng<sup>[15]</sup>在函数梯度下降的框架下,提出了分位数提升算法(QBoost),该算法可预测感兴趣的响应量化值,用于回归和二元分类;He等<sup>[16]</sup>构建了支持向量分位数回归模型(SVQR),通过将分位数回归(QR)纳入支持向量回归机(SVR)中以构建非线性分位数回归模型。与传统分位数回归相比,这些方法在区间预测方面显示出优势。

由于深度学习网络能够出色地刻画时间序列数据中复杂且多变的非线性关系,因此在预测领域取得了卓越成就。目前,将深度学习神经网络与分位数回归相融合,已成为主流的区间预测方法。例如,Chen等<sup>[17]</sup>构建了一种融合最小门控单元(MGRU)和分位数回归的混合模型,该模型巧妙地将深度学习领域中的最小门控单元网络与统计领域的分位数回归相结合,实现了更为精准的预测;Hu等<sup>[18]</sup>提出基于时序卷积网络(TCN)架构和共形分位数回归(Conformal Quantile Regression, CQR)算法的风电预测区间构建方法,这种方法不仅有效结合了两者的优势,而且无须对数据进行任何分布假设,显著提升了预测精度;Xiong等<sup>[19]</sup>进一步探索了最小门控内存网络(MGM)与分位数回归的混

合模型,通过引入开关损耗函数机制,使模型能够在不同分位数下进行灵活的训练和预测。尽管分位数回归结合深度学习模型能够带来较为准确的区间预测结果,但由于每个分位数的神经网络都需要单独训练,这也导致了计算复杂度的增加<sup>[20]</sup>。

在组合模型权重参数的确定上,目前广泛采用的方法包括粒子群优化(PSO)、网格搜索(GS)以及遗传算法(GA)等。其中,Xiao等<sup>[21]</sup>分别探讨了基于粒子群优化和网格搜索法的参数选择策略,发现网格搜索方法具有全面搜索参数空间的能力,但此过程常常较为耗时,而PSO在求解过程中可能存在陷入局部最优解的风险;Sun等<sup>[22]</sup>聚焦于遗传算法与网格搜索在参数优化中的应用,并指出,在数据集规模较小且待优化参数不多的情况下,网格搜索在参数优化方面的表现优于遗传算法,显示出更高的效率和准确性。

综上,本文提出一种基于深度学习的区间预测组合模型。首先,利用门控循环单元(GRU)、长短期记忆(LSTM)和双向长短期记忆(BiLSTM)模型,分别结合核密度估计和分位数回归方法构建区间预测模型。评估后发现,核密度估计模型区间窄但覆盖率不足,而分位数回归模型覆盖率高但区间宽。因此,进一步利用两者优势进行加权组合,利用网格搜索法(Grid Search, GS)优化权重,确保高覆盖率的同时最小化区间宽度。最后,通过覆盖宽度准则(Coverage Width Criterion, CWC)指标<sup>[23]</sup>评估模型的预测性能。

与以往研究相比,本文的创新点和主要贡献如下:

(1) 将BiLSTM和分位数回归相融合,充分利用BiLSTM模型强大的拟合能力和分位数回归的非对称形式,以最小化绝对值残差。这一创新不仅使得模型能够直接生成黄金期货价格的预测区间,而且有效避免了传统参数法中因误差分布假设不准确而导致的预测偏差。

(2) 提出了一种新颖的区间预测组合模型,突破了传统点预测无法捕捉黄金期货价格数据中随机性信息的局限。该模型融合了核密度估计和分位数回归的优点,使得构建的预测区间能够精准地反映黄金期货收盘价数据中的不确定性,从而更准确地描绘出未来黄金期货价格的波动趋势。

(3) 本文提出的组合模型在保证预测区间的区间覆盖率高于设定置信度的同时,还实现了较低的区间平均宽度。经过实证研究,最终证明该模型在CWC指标上表现优异,明显优于其他对比模型,为黄金期货价格预测提供了新的有效工具。

## 1 研究方法

### 1.1 双向长短期记忆

传统循环神经网络(RNN)在处理长序列时易遭遇

梯度消失与爆炸,影响对长时间序列特征的捕捉,降低预测性能。为解决这一问题,LSTM通过增设输入门、遗忘门和输出门,巧妙控制了信息的遗忘、更新与输出,有效缓解了这些问题,提高了预测的准确性。

基于LSTM的黄金期货价格预测方法在一定程度上提高了黄金期货价格的准确性,但该方法在对黄金期货价格进行学习时存在单一方向性等问题,忽略了黄金期货价格每日之间双向的联系,难以充分利用整个黄金期货收盘价序列的特征。

综上,为精确预测黄金期货价格,本文使用基于双向长短期记忆(BiLSTM)的预测方法。此方法通过对黄金期货收盘价序列进行双向学习,有效解决了单向学习在早期特征提取和记忆方面的局限。双向学习的策略能够深入挖掘黄金期货价格日与日之间的双向关联,从而更全面、准确地捕捉到数据的整体特征,提升预测的准确性。

BiLSTM结合两个LSTM,其中一个LSTM按正向顺序处理输入序列,而另一个则按反向顺序处理,随后组合两个LSTM的输出。具体地,正向LSTM按时间步从左到右处理输入序列,每个时间步的隐藏状态 $h_t$ 和单元状态 $c_t$ 计算步骤如下:

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

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \quad (3)$$

$$\tilde{c}_t = \tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

其中, $i_t$ 、 $f_t$ 和 $o_t$ 分别代表了 $t$ 时刻下的输入门、遗忘门和输出门的值, $\tilde{c}_t$ 表示输入细胞的备选状态, $\sigma$ 和 $\tanh$ 是Sigmoid和双曲正切函数, $\odot$ 表示逐元素乘积。

类似地,反向LSTM按时间步从右到左处理输入序列,每个时间步的隐藏状态 $h'_t$ 和单元状态 $c'_t$ 可以由类似的公式计算。最终,BiLSTM的输出由两个方向的隐藏状态拼接而成 $y_t = [h_t, h'_t]$ 。

### 1.2 核密度估计

核密度估计是一种非参数方法,它直接利用样本数据本身进行预测和估计,无须对数据的分布形式做出任何先验假设。若随机变量 $X$ 的密度函数 $p(x) = F'(x)$ ,根据核密度估计的基本思想,可以得到 $f$ 的简单估计,即

$$p_n(x) = \frac{F(x+h) - F(x-h)}{2h} \quad (7)$$

其中, $h = h(n)$ 为非负常数, $n$ 为总的待估计样本量; $F(x)$ 为随机变量 $X$ 的经验分布函数。在选择 $h$ 使 $n \rightarrow \infty$ 、 $h \rightarrow 0$ 且 $nh \rightarrow +\infty$ 时,可以让估计的密度函数 $p_n(x)$ 表现

出良好的特性。核密度函数的估计表达如下:

$$\hat{p}(x) = \frac{1}{nh} \sum_{i=1}^N K\left(\frac{x - X_i}{h}\right) \quad (8)$$

其中,  $n$  代表样本的总数,  $h$  代表窗口宽度,  $X_i$  为指定的样本,  $K(\cdot)$  为选择的高斯核函数, 其公式为

$$K(\cdot) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (9)$$

### 1.3 基于双向长短期记忆的分位数回归模型

#### 1.3.1 分位数回归

在探究自变量对因变量分布特征的影响时, 分位数回归展现出了独特优势。与传统的最小二乘回归方法相比, 后者主要聚焦于最小化残差平方和, 从而得出因变量的条件分布均值。然而, 分位数回归以其非对称绝对值残差加权和最小化方法, 不仅涵盖了均值估计, 还能全面展现因变量条件分布的全貌。它的重要之处在于, 它是基于误差数据驱动的, 无须假设特定的误差概率分布, 因此在处理不符合某一固定概率分布的误差时, 分位数回归的估计结果更可靠。这使得分位数回归在统计分析能力上相较于传统最小二乘法更为全面和优越。

分位数回归具有如下形式:

$$Q_y(\tau | X) = X'\beta(\tau) \quad (10)$$

式(10)中,  $\tau$  是分位点,  $\tau \in (0, 1)$ ;  $Q_y(\tau | X)$  是被解释变量  $X' = [x_1, x_2, \dots, x_t]$  在给定条件下的  $\tau$  分位数;  $\beta(\tau) = [\beta_0(\tau), \beta_1(\tau), \dots, \beta_t(\tau)]'$  为回归系数向量, 随着分位点  $\tau$  改变而改变;  $t = 1, 2, 3, \dots, N$ ,  $N$  为训练样本数。

利用式(10)计算多个分位点处的预测值, 为了最小化真实值和预测值的误差, 采用 pinball 损失函数来构建目标函数, 其表达式如式(11)所示:

$$\min_{\beta} \left\{ \sum_{y \geq X'\beta} \tau |y - X'\beta| + \sum_{y < X'\beta} (1 - \tau) |y - X'\beta| \right\} \quad (11)$$

#### 1.3.2 双向长短期记忆分位数回归模型

在应用过程中, 反应和描述因素之间存在着多样的、非线性的联系特征, 但是, 由于式(10)中的分位数回归模型是基于线性模式构建的, 它限制了解释因素对反应因素的影响方式, 使得它们更偏向于线性框架。人工神经网络能模拟从输入到输出之间的非线性结构, Taylor<sup>[13]</sup> 提出非线性神经网络分位数模型, 实现了由解释变量至响应变量的非线性关联。

分位数回归不需要事先定义数据的分布情况, 其利用所选定的分位点来实现对输入变量的回归分析, 这有助于完整地展示出目标  $Y$  波动的可能性分布范围。因此, 这个特点使分位数回归能精准把握黄金期货预测过程中不稳定因素的变化, 并产生可靠的预测

区间, 有效解决了传统区间预测方式模型过于繁复、预设数据分布和参数估计偏差等难题。

本文整合 QR 与 BiLSTM, 形成 BiLSTM-QR 模型, 以产生多分位点预测值。该模型可以在不同置信度下提供多分位数预测, 增强预测的全面性。设 BiLSTM 神经网络隐藏层有  $K$  个神经, 得到 BiLSTM-QR 模型的输出为

$$Q_{y_t}(\tau | x_t, W, b) = f\left(\sum_{k=1}^K W_k(\tau) h_k(\tau) + b_k(\tau)\right) \quad (12)$$

式(12)中,  $Q_{y_t}(\tau | x_t, W, b)$  代表 BiLSTM-QR 模型  $t$  时刻的输出,  $W_k(\tau)$  和  $b_k(\tau)$  分别为  $\tau$  分位点下神经网络隐含层输出与全连接层输入之间的权重和偏置,  $h_k(\tau)$  为  $\tau$  分位点下隐含层的输出值,  $f(\cdot)$  为激活函数。

假定的置信概率为  $1 - \alpha$ , 选取分位点  $\tau = \alpha/2$ , 标记为下分位点, 并根据式(12)确定黄金期货价格的  $\alpha/2$  条件分位数, 即下分位数, 用以设定预测范围的下界; 接着, 再挑选另一个分位点  $\tau = 1 - \alpha/2$ , 标记为上分位点, 同样利用式(12)计算出黄金期货价格的  $1 - \alpha/2$  条件分位数, 即上分位数, 以此设立预测范围的上界。

## 2 基于双向长短期记忆的区间预测组合模型

### 2.1 区间预测组合模型

提出一种基于深度学习的区间预测组合模型。首先构建门控循环单元、长短期记忆以及双向长短期记忆这 3 种主流的深度学习模型, 采用核密度估计和分位数回归技术, 构建多个区间预测模型。经过对模型的全面评估, 精选出 GRU-KDE、LSTM-KDE、BiLSTM-KDE 这 3 个核密度估计模型中的最佳区间预测模型, 以及 GRU-QR、LSTM-QR、BiLSTM-QR 这 3 个分位数回归模型中的最优模型。接下来, 将这些最优模型进行加权组合, 以提升预测精度。在确定组合权重时, 采用网格搜索法, 通过遍历不同的权重组合, 在确保组合模型预测区间覆盖率超过预设置信度水平的基础上, 寻找能使预测区间平均宽度达到最小的组合模型权重。这一方法旨在实现预测准确性和稳定性的最佳平衡, 为实际应用提供更为可靠和精准的区间预测解决方案。最终的区间上下界如式(13)所示:

$$\begin{cases} L = wL_1 + (1-w)L_2 \\ U = wU_1 + (1-w)U_2 \end{cases} \quad (13)$$

其中,  $U_1, L_1$  是由最优核密度估计模型得到的预测区间上下界,  $U_2, L_2$  是由最优分位数回归模型得到的预测区间上下界,  $U, L$  是组合模型的预测区间上下界。

### 2.2 预测步骤

所构建的预测流程图如图 1 所示。

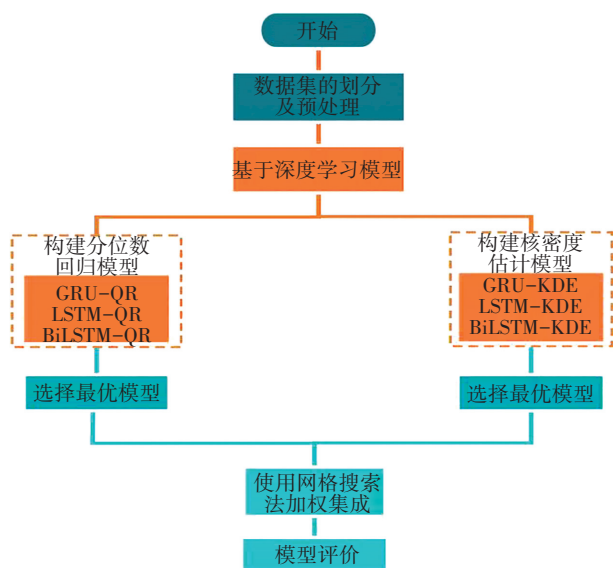


图 1 组合模型区间预测流程图

Fig. 1 Flowchart of the interval prediction of the combined model

预测模型的主要步骤:

(1) 数据处理。黄金收盘价数据划分及预处理。

(2) 构建深度学习模型。构建 GRU、LSTM、BiLSTM 模型,进行训练和预测。

(3) 构建核密度估计模型。对 GRU、LSTM、BiLSTM 预测结果进行核密度估计,得到各模型区间预测结果。

(4) 构建分位数回归模型。同时构建 GRU-QR、LSTM-QR、BiLSTM-QR 预测模型,通过训练与预测,得出相应的区间预测结果。

(5) 评估与选择模型。根据 PICP、PINAW 指标,分别选出核密度估计和分位数回归的最优区间预测模型。

(6) 加权集成处理。运用网格搜索法,对选择出的最优核密度估计和分位数回归模型的区间预测结果加权集成。

(7) 评估组合模型性能。通过 PICP、PINAW、CWC 指标,来全面评判所提出的组合模型和其他模型的优劣。

### 2.3 评价指标

#### 2.3.1 点预测评价指标

选取平均绝对百分比误差 (MAPE)、平均绝对误差 (MAE) 和均方根误差 (RMSE) 3 个统计指标对深度学习点预测模型的预测性能进行客观评价,其定义如下:

$$R_{MAPE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \times 100\% \quad (14)$$

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

$$R_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (16)$$

其中,  $N$  表示测试样本数量,  $\hat{y}_i$  表示预测值,  $y_i$  表示实际

值。这 3 个指标取值越小说明所构建模型的预测性能越好。

#### 2.3.2 区间预测评价指标

为对预测区间的覆盖率和宽度作出评判,引入 PICP 和 PINAW 两个评价指标。其中, PICP 代表区间覆盖率,它用于评估预测区间覆盖未来黄金期货价格的可能性。当预测区间能更精确地覆盖实际的黄金期货收盘价时,其可信度便随之提升。简而言之, PICP 越高,表明模型对未来黄金期货价格的预测能力越强。

$$\zeta_{PICP} = \frac{1}{N} \sum_{i=1}^N \delta_i \quad (17)$$

$$\delta_i = \begin{cases} 1, & y_i \in [L_i^\alpha, U_i^\alpha] \\ 0, & y_i \notin [L_i^\alpha, U_i^\alpha] \end{cases} \quad (18)$$

其中,  $U_i^\alpha$  和  $L_i^\alpha$  为预测区间的上下限。尽管 PICP 是衡量预测区间准确性的重要指标,但仅高覆盖率而无紧凑性的预测区间对决策参考价值有限,因此,引入 PINAW 来全面评估预测区间的质量。

$$\zeta_{PINAW} = \frac{1}{N \times \psi} \sum_{i=1}^N (U_i^\alpha - L_i^\alpha) \quad (19)$$

$$\psi = y_{\max} - y_{\min} \quad (20)$$

其中,  $\psi$  为归一化因子,  $y_{\max}$  和  $y_{\min}$  分别为当日黄金期货收盘价实际的最高价和最低价。PINAW 值越小,区间宽度越窄,区间预测性能越好。

PICP 和 PINAW 都是预测区间的关键特征,但两者此消彼长。具体来说,若要提高覆盖率,不可避免地会导致预测区间的宽度增大;相反,若要减小区间宽度,则覆盖率会相应降低。为了有效平衡这一矛盾,此处引入覆盖宽度准则 (CWC)<sup>[20]</sup>,该准则能够综合考虑区间覆盖率和宽度的双重维度,从而全面评估预测区间的质量。当 CWC 值越趋近于 0,意味着模型的预测区间质量越高,具体计算公式如下:

$$\zeta_{CWC} = \zeta_{PINAW} (1 + \gamma e^{-\eta(\zeta_{PICP} - \mu)}) \quad (21)$$

$$\gamma = \begin{cases} 0 & \zeta_{PICP} \geq \mu \\ 1 & \zeta_{PICP} < \mu \end{cases} \quad (22)$$

其中,  $\eta, \mu$  是用来决定惩罚程度的参数;  $\mu$  是由置信区间决定的,比如当置信水平为 90% 时,  $\mu = 0.9$ 。根据 CWC 的定义, CWC 在 PICP 未能达到预先设定的区间置信水平  $\mu$  时施加惩罚,从而提高 PICP 的权重。这个惩罚由  $\eta$  控制,  $\eta$  越小,说明 PICP 没有达到置信水平的惩罚程度越小,即其重要程度越低。若 PICP 达到了置信水平,则 PICP 等于 PINAW。

### 3 实证分析

选取上海黄金交易所黄金期货收盘价日度数据,如图 2 所示。选取的数据范围为 2008 年 1 月 9 日到 2022 年 5 月 13 日,共 3 485 个数据,所有数据均来自

Wind 数据库。即实验是在 MATLAB 和 Python 计算环境下实现的。即 GRU、LSTM 以及 BiLSTM 这 3 种深度学习模型,以及它们各自对应的核密度估计和分位数回归区间预测模型,均通过 MATLAB 程序实现。在组合模型权重分配环节,采用网格搜索算法,并借助 Python 程序实现。

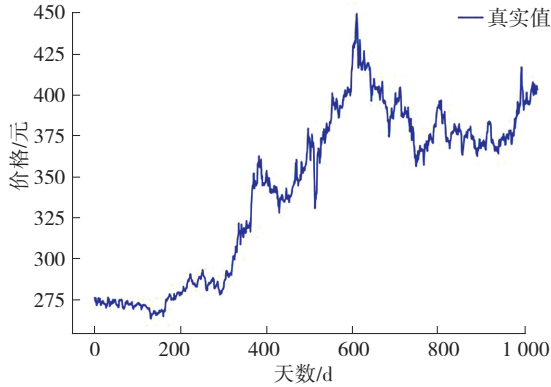


图 2 黄金期货收盘价格趋势图

Fig. 2 Trend chart of the closing prices of gold futures

### 3.1 数据处理

在对序列进行量化分析前,首先将数据进行归一化处理。根据选取数据的特点和模型构建的要求,采用线性函数归一化方法对数据进行预处理。公式如下:

$$x'_i = \frac{x_i - \min x_i}{\max x_i - \min x_i} \quad (23)$$

其中,  $x_i$  代表第  $i$  个变量,  $\min x_i$ 、 $\max x_i$  分别代表  $x_i$  的最小值和最大值。

### 3.2 模型参数设置

训练集与测试集划分为 7 : 3, 即 2 440 个训练样本, 1 045 个测试样本。依次验证模型的最佳输入步长, 最终设置滑动窗口的大小为 8, 即用过去 8 d 的黄金期货收盘价数据预测未来 1 d 的黄金期货价格。GRU、LSTM、BiLSTM 网络均采用 Adam 为优化算法, 并应用 dropout 机制来规避模型训练的过拟合。GRU、LSTM、BiLSTM 具体超参数设置如表 1 所示。分位数回归分位点设置为 [0.02, 0.97]、[0.05, 0.95]、[0.07, 0.92]。网格搜索法权重的搜索范围为 [0, 1], 步长为 0.01。在计算覆盖宽度准则(CWC)时, 惩罚参数  $\eta$  设置为 2, 而  $\mu$  则分别对应置信水平 0.95、0.9 和 0.85, 以便在不同置信水平下评估预测区间的质量。

表 1 模型参数设置表

Table 1 Model parameter settings

模 型	参 数 设 置		
	隐含层节点数	初始学习率	学习率下降因子
GRU	50	0.01	0.5
LSTM	10	0.01	0.2
BiLSTM	10	0.005	0.2

### 3.3 模型性能评价与结果比较

#### 3.3.1 深度学习预测模型

基于 RNN 模型选择 3 种常见的深度学习模型 GRU、LSTM、BiLSTM。由表 2 的点预测结果可知, BiLSTM 模型的 MAPE、MAE、RMSE 均低于 GRU 模型和 LSTM 模型, 其模型预测性能最佳。如图 3 可见, BiLSTM 模型的预测曲线更接近实际值, 表明其拟合能力较强。

表 2 GRU、LSTM、BiLSTM 模型的评价指标对比

Table 2 Comparison of evaluation metrics for GRU, LSTM and BiLSTM models

预测模型	$R_{MAPE}$	$R_{MAE}$	$R_{RMSE}$
GRU	1.479	5.987 9	8.357 3
LSTM	0.711	2.578 3	3.640 7
BiLSTM	0.569	2.522 6	3.628 4

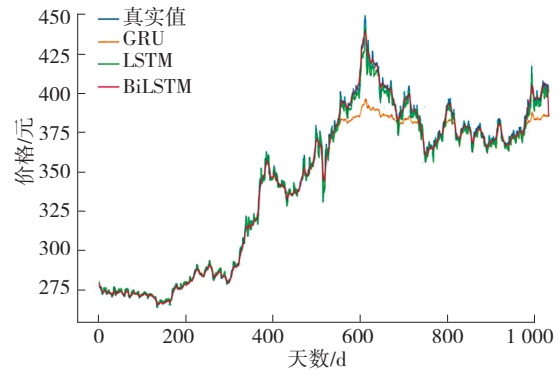


图 3 GRU、LSTM、BiLSTM 预测结果拟合图

Fig. 3 Prediction results of GRU, LSTM, and BiLSTM models

#### 3.3.2 核密度估计区间预测模型

鉴于点预测方法无法充分捕捉黄金期货数据的随机性, 可进一步探索区间预测方法, 以期提供更全面、准确的预测结果。为此, 本文对 3 个模型的预测结果在不同置信水平下进行核密度估计, 结果如表 3 所示。

表 3 不同核密度估计模型的区间预测评价指标对比

Table 3 Comparison of evaluation metrics for interval prediction of different kernel density estimation models /%

置信水平	GRU-KDE		LSTM-KDE		BiLSTM-KDE	
	$\zeta_{PICP}$	$\zeta_{PINAW}$	$\zeta_{PICP}$	$\zeta_{PINAW}$	$\zeta_{PICP}$	$\zeta_{PINAW}$
95	73.87	4.11	91.51	3.75	94.34	3.62
90	68.37	3.02	82.73	2.83	89.87	2.71
85	63.07	2.46	76.76	2.33	85.82	2.25

由表 3 可知, 在 95% 置信水平下, BiLSTM-KDE 模型的 PICP 指标分别高出 GRU-KDE、LSTM-KDE 模型 20.47%、2.83%; PINAW 指标分别低于 GRU-KDE、LSTM-KDE 模型 0.49%、0.13%。在 90% 置信水平下, BiLSTM-KDE 模型的 PICP 指标分别高出 GRU-KDE、

LSTM-KDE 模型 21.5%、7.14%;PINAW 指标分别低于 GRU-KDE、LSTM-KDE 模型 0.31%、0.12%。在 85% 置信水平下,BiLSTM-KDE 模型的 PICP 指标分别高出 GRU-KDE、LSTM-KDE 模型 22.75%、9.06%;PINAW 指标分别低于 GRU-KDE、LSTM-KDE 模型 0.21%、0.08%。如图 4 所示,相比于其他两个模型,BiLSTM-KDE 模型的预测区间更接近真实值曲线。

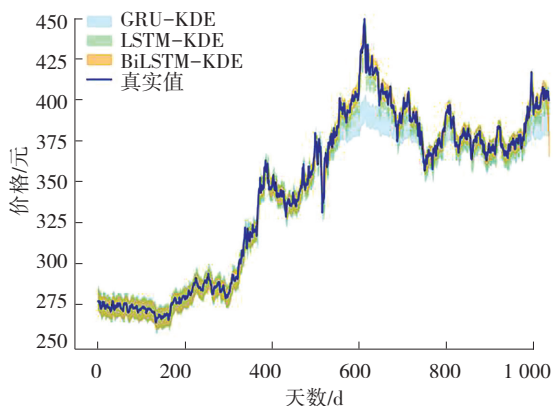


图 4 95%置信水平下不同核密度估计模型的区间预测结果拟合图

Fig. 4 Fitting chart of interval prediction results of different kernel density estimation models at the 95% confidence level

可见 BiLSTM-KDE 模型区间预测效果最佳。这是由于核密度估计是基于点预测的结果,结合概率计算方法,产生轨迹点的分布概率,从而使其依赖于点预测模型的性能。BiLSTM 的点预测结果优于 LSTM、GRU 模型,从而其核密度估计的区间预测结果也优于 LSTM、GRU 模型。

综上所述,虽然 BiLSTM-KDE 模型在预测性能上表现最佳,但仍存在预测区间宽度较窄,导致其区间覆盖率不够高的问题。

### 3.3.3 分位数回归区间预测模型

由于核密度估计模型存在预测区间宽度较窄的局限性,导致其区间覆盖率相对较低。为了弥补这一不足,本文进一步采用深度学习模型与分位数回归相结合的方法来进行区间预测。表 4 为不同分位数回归模型的区间预测评价指标对比。

表 4 不同分位数回归模型的区间预测评价指标对比

Table 4 Comparison of interval prediction evaluation metrics for different quantile regression models /%

置信水平	GRU-QR		LSTM-QR		BiLSTM-QR	
	$\zeta_{PICP}$	$\zeta_{PINAW}$	$\zeta_{PICP}$	$\zeta_{PINAW}$	$\zeta_{PICP}$	$\zeta_{PINAW}$
95	90.16	14.4	97.48	8.48	97.01	7.88
90	74.06	7.21	86.79	5.68	86.59	2.31
85	70.01	9.65	68.45	3.32	88.81	2.65

由表 4 可知,在 95%置信水平下,BiLSTM-QR 模型的 PICP 指标高于 GRU-QR 模型 6.85%,低于 LSTM-QR 模型 0.47%;PINAW 指标分别低于 GRU-QR、LSTM-QR 模型 6.52%、0.6%。在 90%置信水平下,BiLSTM-QR 模型的 PICP 指标高于 GRU-QR 模型 12.53%,低于 LSTM-QR 模型 0.2%;PINAW 指标分别低于 GRU-QR、LSTM-QR 模型 4.9%、3.37%。在 85%置信水平下,BiLSTM-QR 模型的 PICP 指标分别高出 GRU-QR、LSTM-QR 模型 18.8%、20.36%;PINAW 指标分别低于 GRU-QR、LSTM-QR 模型 7%、0.67%。如图 5 所示,相比于其他两个模型,BiLSTM-QR 模型的预测效果更好,其预测区间能更好地覆盖真实值曲线,且区间宽度较窄。

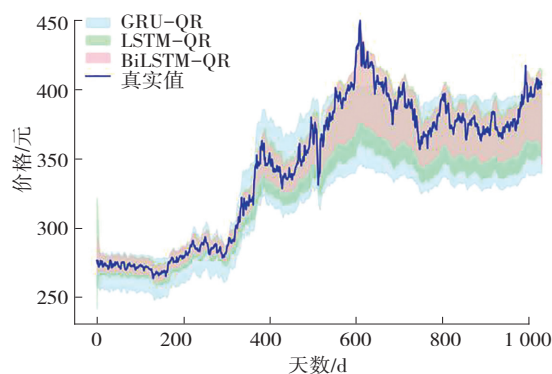


图 5 95%置信水平下不同分位数回归模型的区间预测结果拟合图

Fig. 5 Fitting chart of interval prediction results of different quantile regression models at the 95% confidence level

综上所述,虽然 BiLSTM-QR 模型在预测性能上表现最佳,但与采用核密度估计的区间预测模型相比,其预测区间宽度相对较宽,这在一定程度上影响了预测精度和实用性。因此,需进一步探索和研究如何优化 BiLSTM-QR 模型,以减小预测区间的宽度,提高其预测精度和稳定性。

### 3.3.4 组合模型性能评价

鉴于双向长短期记忆(BiLSTM)的核密度估计模型(KDE)和分位数回归模型(QR)在预测区间覆盖率和区间平均宽度方面各有优缺点,本文创新性地将两者的区间预测结果进行加权组合。为了确保组合模型的效果最优,进一步利用网格搜索法来确定组合权重。通过遍历不同的权重组合,在确保组合模型区间覆盖率高于所设定置信度水平的同时,力求找到使预测区间平均宽度达到最小的最佳权重配置。经过实证研究发现,在 95%、90%以及 85%的置信水平下,BiLSTM-KDE、BiLSTM-QR 及其组合模型的最佳权重比分别确定为 0.941、0.963 和 0.927。表 5 为 BiLSTM-KDE、BiLSTM-QR 及其组合模型的区间预测评价指标对比。

表 5 BiLSTM-KDE、BiLSTM-QR 和组合模型的区间预测评价指标对比

Table 5 Comparison of interval prediction evaluation metrics for BiLSTM-KDE, BiLSTM-QR and the combined model /%

预测模型	95			90			85		
	$\zeta_{PICP}$	$\zeta_{PINAW}$	$\zeta_{CWC}$	$\zeta_{PICP}$	$\zeta_{PINAW}$	$\zeta_{CWC}$	$\zeta_{PICP}$	$\zeta_{PINAW}$	$\zeta_{CWC}$
BiLSTM-KDE	94.34	3.62	7.28	89.87	2.71	5.43	85.82	2.25	2.25
BiLSTM-QR	97.01	7.88	7.88	86.59	2.31	4.79	88.81	2.65	2.65
BiLSTM-QR-KDE	96.62	4.75	4.75	90.26	2.52	2.52	85.92	2.20	2.22

由表 5 可知,在 95%置信水平下,本文提出组合模型的 PICP 指标值高于 BiLSTM-KDE 模型 2.28%,低于 BiLSTM-QR 模型 0.39%;PINAW 指标高值于 BiLSTM-KDE 模型 1.13%,低于 BiLSTM-QR 模型 3.13%。如图 6 所示,BiLSTM-QR 模型虽然具备较高的覆盖率,但区间宽度较大,从而增加了预测结果的不确定性。相比之下,BiLSTM-KDE 模型虽然区间宽度较窄,但其区间覆盖率较低,这可能导致预测精度下降,有时预测结果甚至无法涵盖真实值的范围。而本文提出的组合模型在预测区间上表现出色,不仅基本覆盖了真实值,还有效降低了区间宽度,从而在提升预测精度的同时,也增强了模型的稳定性。

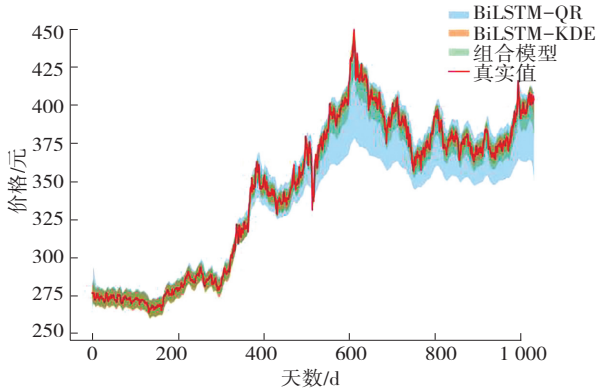


图 6 95%置信水平下 BiLSTM-QR、BiLSTM-KDE 和组合模型的预测区间

Fig. 6 Prediction intervals for BiLSTM-QR, BiLSTM-KDE and the combined model at the 95% confidence level

在 90%置信水平下,本文提出组合模型的 PICP 指标值分别高于 BiLSTM-KDE、BiLSTM-QR 模型 0.39%、3.67%;组合模型的 PINAW 指标低于 BiLSTM-KDE 模型 0.19%,高于 BiLSTM-QR 模型 0.21%。在 85%置信水平下,本文提出组合模型的 PICP 指标值高于 BiLSTM-KDE 模型 0.1%,低于 BiLSTM-QR 模型 2.89%;PINAW 指标分别低于 BiLSTM-KDE、BiLSTM-QR 模型 0.05%、0.45%;

在 95%、90%、85%置信水平下,组合模型均展现出最优的 CWC 指标性能。如图 7 所示,在不同置信水平下,该组合模型不仅表现出高区间覆盖率,同时拥有相对较窄的区间宽度。其兼顾了区间覆盖率和区间宽度两个指标,使得预测模型的性能得以显著提升。这一结果充分证明了本文所提组合模型在预测精度和稳定性方面的卓越表现。

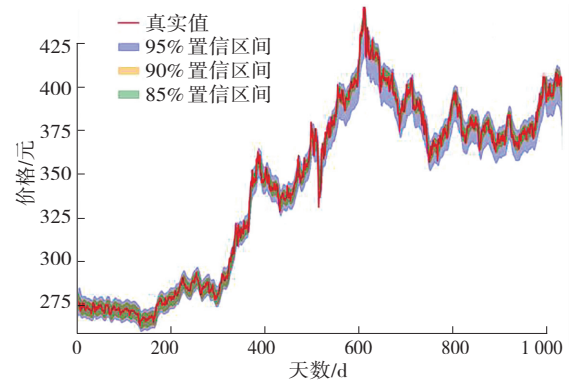


图 7 组合模型在 95%、90%和 85%置信水平的预测区间

Fig. 7 Prediction intervals for the combined model at the 95%, 90% and 85% confidence levels

### 4 结论

传统点预测在捕捉黄金期货价格差异性和不确定性上存在局限,而区间预测则能直观展示未来价格的波动范围及潜在趋势。为更精准把握数据中的不确定性,本文提出一种基于双向长短期记忆的组合区间预测模型。通过一系列实验验证,得出以下结论:

(1) 在对比 GRU-KDE、LSTM-KDE 和 BiLSTM-KDE 3 种模型时,发现 BiLSTM-KDE 的预测效果最为出色。然而,它仍然面临区间覆盖率不足的问题,这意味着其预测范围的准确性有待提升。

(2) 在对比 GRU-QR、LSTM-QR 和 BiLSTM-QR 3 种模型时,发现 BiLSTM-QR 在预测表现上同样领先。但与 BiLSTM-KDE 模型相比,其预测区间的宽度偏大,这可能导致预测精度受到一定影响。

(3) 进一步对比 BiLSTM-KDE 与 BiLSTM-QR 模型,发现前者在预测区间平均宽度上更胜一筹,而后者则在预测区间覆盖率上表现更佳。这表明两者各有优势,但也存在明显的互补性。

(4) 提出的组合区间预测模型巧妙地将 BiLSTM-KDE 和 BiLSTM-QR 两者的优点结合起来。实验结果显示:该组合模型的 CWC 指标在所有对比模型中表现最佳,成功地在区间覆盖率和区间平均宽度之间找到了一个理想的平衡点,为黄金期货价格的预测提供了更为准确和可靠的解决方案。

本文提出的组合模型在区间预测领域展现出了卓越的性能,不仅覆盖率高且宽度适中,其高精度预测和强大泛化能力为黄金期货价格预测的稳定性提供了有力保

障,为决策者提供了精准可靠的数据支持。此外,该区间预测方法具备广泛的应用潜力,不仅限于黄金期货,还可拓展至原油价格、交通流量等多领域预测问题。然而,目前模型仅基于单变量时间序列分析,未来研究可进一步融入多元复杂因素,以提升预测精度和广泛适用性。

#### 参考文献(References):

- [1] ASKARI M, ASKARI H. Time series grey system prediction-based models: gold price forecasting[J]. Trends in Applied Sciences Research, 2011, 6(11): 1287-1292.
- [2] ARIYO A A, ADEWUMI A O, AYO C K. Stock price prediction using the ARIMA model[C]//Proceedings of the UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. Piscataway: IEEE Press, 2015: 106-112.
- [3] 韦艳华, 张世英. 金融市场的相关性分析: Copula-GARCH模型及其应用[J]. 系统工程, 2004, 22(4): 7-12.  
WEI Yan-hua, ZHANG Shi-ying. Dependence analysis of finance markets: copula-GARCH model and its application[J]. Systems Engineering, 2004, 22(4): 7-12.
- [4] GUO Y, HAN S, SHEN C, et al. An adaptive SVR for high-frequency stock price forecasting [J]. IEEE Access, 2018, 6: 11397-11404.
- [5] YANG J, DE MONTIGNY D, TRELEAVEN P. ANN, LSTM and SVR for gold price forecasting[C]//Proceedings of the IEEE Symposium on Computational Intelligence for Financial Engineering and Economics. Piscataway: IEEE Press, 2022: 1-7.
- [6] LYU M, WANG J, NIU X, et al. A newly combination model based on data denoising strategy and advanced optimization algorithm for short-term wind speed prediction[J]. Journal of Ambient Intelligence and Humanized Computing, 2023, 14(7): 8271-8290.
- [7] CHEN S, GE L. Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction [J]. Quantitative Finance, 2019, 19(9): 1507-1515.
- [8] JIANG J, WU L, ZHAO H, et al. Forecasting movements of stock time series based on hidden state guided deep learning approach[J]. Information Processing & Management, 2023, 60(3): 103328.
- [9] 李梦, 黄章杰, 徐健晖. 基于深度学习和小波分析的 LSTM-Wavelet 模型股价预测[J]. 重庆工商大学学报(自然科学版), 2023, 40(2): 99-105.  
LI Meng, HUANG Zhang-jie, XU Jian-hui. Stock price prediction with LSTM-wavelet model based on deep learning and wavelet analysis [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2023, 40(2): 99-105.
- [10] MEIRA E, CYRINO OLIVEIRA F L, DE MENEZES L M. Point and interval forecasting of electricity supply via pruned ensembles[J]. Energy, 2021, 232: 121009.
- [11] 熊景华, 茹璟. 基于随机森林算法和模糊信息粒化的汇率预测组合模型研究[J]. 数量经济技术经济研究, 2021, 38(1): 135-156.
- [12] 郎伟明, 麻向津, 周博文, 等. 基于 LSTM 和非参数核密度估计的风电功率概率率区间预测[J]. 智慧电力, 2020, 48(2): 31-37, 103.  
LANG Wei-ming, MA Xiang-jin, ZHOU Bo-wen, et al. Wind power probabilistic intervals prediction based on LSTM and nonparametric kernel density estimation [J]. Smart Power, 2020, 48(2): 31-37, 103.
- [13] LI H, YU Y, HUANG Z, et al. A multi-step ahead point-interval forecasting system for hourly PM<sub>2.5</sub> concentrations based on multivariate decomposition and kernel density estimation[J]. Expert Systems with Applications, 2023, 226: 120140.
- [14] TAYLOR J W. A quantile regression neural network approach to estimating the conditional density of multiperiod returns[J]. Journal of Forecasting, 2000, 19(4): 299-311.
- [15] ZHENG S. QBoost: predicting quantiles with boosting for regression and binary classification [J]. Expert Systems with Applications, 2012, 39(2): 1687-1697.
- [16] HE Y, LIU R, LI H, et al. Short-term power load probability density forecasting method using kernel-based support vector quantile regression and Copula theory[J]. Applied Energy, 2017, 185: 254-266.
- [17] CHEN T, QIAN Z, JING B, et al. Probabilistic wind speed forecasting based on minimal gated unit and quantile regression[J]. Journal of Physics: Conference Series, 2020, 1659(1): 012039.
- [18] HU J, LUO Q, TANG J, et al. Conformalized temporal convolutional quantile regression networks for wind power interval forecasting[J]. Energy, 2022, 248: 123497.
- [19] XIONG Z, YAO J, HUANG Y, et al. A wind speed forecasting method based on EMD-MGM with switching QR loss function and novel subsequence superposition [J]. Applied Energy, 2024, 353: 122248.
- [20] ZHU J, HE Y, GAO Z. Wind power interval and point prediction model using neural network based multi-objective optimization[J]. Energy, 2023, 283: 129079.
- [21] XIAO T, REN D, LEI S, et al. Based on grid-search and PSO parameter optimization for Support Vector Machine [C]//Proceedings of the 11th World Congress on Intelligent Control and Automation. Piscataway: IEEE Press, 2015: 1529-1533.
- [22] SUN Y, WANG Y, GUO L, et al. The comparison of optimizing SVM by GA and grid search[C]//Proceedings of the 13th IEEE International Conference on Electronic Measurement & Instruments. Piscataway: IEEE Press, 2018: 354-360.
- [23] KHOSRAVI A, NAHAVANDI S, CREIGHTON D, et al. Lower upper bound estimation method for construction of neural network-based prediction intervals [J]. IEEE Transactions on Neural Networks, 2011, 22(3): 337-346.

责任编辑:李翠薇