

基于增强式集成学习模型的铁矿石价格预测研究^{*}

刘仕强, 潘威旭

(福州大学 经济与管理学院, 福州 350108)

摘要:随着经济建设进程的加速以及基础建设对原材料需求的持续增长,原材料生产企业面临着日益严峻的挑战。为维持企业的稳定运营并提升竞争力,探寻一种有效的原材料商品价格预测方法成为企业的迫切需求。铁矿石作为基础建设的重要原材料之一,期货价格波动呈现出复杂多变的特性,这给价格预测工作带来了巨大挑战。尽管当前已存在多种预测方法,但预测精度和稳定性仍有待提高。因此,为探索一种更为准确且有效的铁矿石价格预测方法,基于多元时间序列的铁矿石价格序列,设计了一个由多头卷积堆叠长短期记忆编码器解码器网络和模态信号分解技术与注意力机制组合的增强式集成学习模型。该方法首先利用模态信号分解技术降低价格序列的波动性,然后通过多头卷积网络自动从多元序列数据中提取复杂空间特征,接着利用堆叠长短期记忆网络从历史序列中提取时间特征,最后通过注意力机制提取关键特征,并得出最终的预测结果。实验结果表明,提出的增强式集成模型在预测精度和稳定性方面均优于未考虑分解技术的其他模型。研究结果具有实际应用价值,可助力钢铁企业更为精准地预测铁矿石价格,进而维持高质量的发展与运营。

关键词:铁矿石价格预测;增强式集成学习模型;模态分解技术;编码器解码器网络;多步时间序列

中图分类号:TP181;F724.5;F764.2

文献标志码:A

文章编号:1672-6693(2025)03-0001-16

铁矿石作为钢铁生产企业的重要原材料,价格波动将直接影响钢铁行业及下游产业链的稳定性和盈利能力^[1]。党的二十大报告强调“坚持把发展经济的着力点放在实体经济上,推进新型工业化”,钢铁工业作为国民经济的重要基础产业,它的稳健发展对于建设现代化强国至关重要^[2]。中国是铁矿石消费大国,尽管国内资源储量丰富但受限于品位不一等因素,国内产能难以满足日益增长的市场需求,导致所需铁矿石高度依赖于从国际市场进口^[3]。这种对进口铁矿石的高度依赖使中国经济不仅易受国际市场波动影响,还面临着国际矿业巨头通过垄断来操控铁矿石价格的风险,进而对中国矿产资源安全和经济稳定发展构成潜在威胁^[4]。从全球范围来看,世界铁矿石产量的 56% 以及出口量的 74% 严重依赖澳大利亚和巴西,而中国是最大的单一进口国,占世界铁矿石进口量的 66%;且中国铁矿石进口量的 62.3% 来自澳大利亚,对澳大利亚铁矿石的依存度较高,这表明铁矿石贸易中买卖双方的多样性存在局限^[5-6]。因此,铁矿石价格的波动不仅直接影响钢铁企业的生产成本和利润,还对中国的经济利益产生深远影响。

自 2006 年起,全球铁矿石价格呈现出 3 轮波动周期态势^[7]:前 2 轮的价格变动主要受到供需基本面的直接影响,价格变动相对稳定;然而,自 2016 年开始的第 3 轮价格波动显得更为复杂,表现为价格先经历缓慢上涨随后急剧下跌,并且在这一过程中金融因素对价格的影响明显增强^[8]。随着铁矿石金融属性的持续加强,价格波动特征变得愈加复杂,其中一方面受供需基本面驱动,另一方面又受诸多非基本面因素的干扰,总体呈现出非线性、波动集聚以及高度不确定性的特点。

传统的铁矿石价格预测方法主要依赖于计量经济学模型或简单的统计方法,这些方法在捕捉价格趋势方面虽有一定的效果,但在处理非线性、高维特性和复杂时间序列关系时存在局限性。价格趋势的演变具有典型的时间序列特征,数据特征主要分为 2 类:性质特征和模式特征。性质特征从整体上揭示数据的内在规律,而模式特征则侧重于挖掘数据中隐藏的成分或影响因素^[9]。为了有效探索数据的内在因素,时频分析方法在剖析联合时频域中信号的频率组成及时变模式方面得到广泛应用^[10]。该方法主要借助时频分析中的信号分解技术,将时间序列分解为频率成分相对简单的子序列,随后针对每个子序列分别建模。这种基于多时间尺度的建模方法能

^{*} 收稿日期:2023-08-10 修回日期:2025-03-29 网络出版时间:2025-06-13T16:57

资助项目:国家自然科学基金面上项目(No. 71871064)

第一作者简介:刘仕强,男,教授,博士生导师,研究方向为矿业管理,E-mail:samsqliu@fzu.edu.cn

网络出版地址:https://link.cnki.net/urlid/50.1165.N.20250613.1607.004

够充分挖掘不同频率级别的信息,从而更有效地刻画数据的特征^[11]。许多学者采取“分而治之”的策略,通过降低建模的复杂性并结合不同预测模型的优势,有效提升了预测精度^[12]。由于变量模式分解方法通过自适应确定分解层数,有效改善了经验模态分解方法中存在的模态混叠问题,从而降低了原始时间序列的复杂性,并使得对分解后各分量的建模更为便利^[13]。

近些年来,随着人工智能和大数据技术的快速发展,基于数据驱动的人工神经网络与深度学习模型在价格预测领域取得了较大进展。特别是在铁矿石价格预测方面,深度学习模型凭借强大的数据处理能力和非线性关系建模能力,展现出卓越的性能和鲁棒性,受到了广泛关注^[14]。成金华等人^[15]通过构建向量自回归模型,分析了铁矿石价格波动的动态关系,指出铁矿石价格波动对钢铁行业产出具有明显的时滞效应。张松波等人^[1]的研究表明,当前中国钢铁生产与需求已步入峰值平台期,致使铁矿石需求将随钢铁消费的下降以及废钢供应量的上升而减少。姜赛平等^[16]针对矿产市场,提出了调动市场主体积极性的激励机制以及优化矿产资源管理的政策建议。现有研究主要集中在铁矿石市场的理论分析、定价机制、国际市场关系以及影响价格波动的各类变量等方面。在价格预测方面,张登耀^[17]设计了基于反向传播神经网络的期货价格预测模型,并运用 3 种评价指标对预测效果进行对比分析。Li 等人^[18]采用了多种神经网络模型对铁矿石价格进行预测,实验结果表明提出的预测模型在预测精度方面表现最优。Ewees 等人^[19]构建了一种由混沌蚱蜢优化算法训练的多层感知器神经网络模型,并通过与其他预测方法的对比实验验证了该模型的优越性。然而,传统神经网络存在过拟合、隐藏节点难以确定、最优参数搜索复杂等缺点。此外,每种预测技术都有特定应用场景和局限性。为了克服这些不足,研究人员提出技术集成和特征选择的思路,通过结合多个机器学习模型来减少预测误差并提升模型性能^[20]。孙景云等人^[21]采用基于多种机器学习的集成模型对期货价格进行预测,并利用多种指标对预测结果进行了分析。集成学习方法的广泛应用,为铁矿石价格预测提供了新的思路和方法。集成模型通过综合不同模型的预测结果,实现了预测精度的提升,降低了预测误差的方差分量^[22]。然而,现有集成模型多侧重于设计序列模型以捕捉铁矿石历史价格的时间依赖性,对与铁矿石相关的其他变量信息的挖掘尚不充分。此外,当输入数据量较大且存在多重共线性时,网络的训练速度和效率会明显下降^[23]。因此,本文提出了一种增强式集成学习的铁矿石价格预测模型。该模型结合了多头卷积堆叠长短期记忆(multi-head convolutional neural network-stacked long short-term memory, MCNN-SLSTM)编码器解码器网络、模态信号分解技术与注意力机制,能够全面捕捉铁矿石价格波动的复杂性和非线性特征,从而更准确地预测铁矿石价格的波动,并有效控制学习模型的复杂性。相较于传统方法和常规集成模型,本文的增强式集成模型具有以下优势:

1) 通过整合多种技术和方法,增强式集成模型能够更全面地捕捉铁矿石价格波动的相关信息,包括历史价格数据、市场供需状况、国际政治经济因素等,从而提高了预测的准确性和可靠性。

2) 模型采用多头卷积堆叠长短期记忆编码器-解码器网络,能够处理大规模输入数据,并有效应对多重共线性问题。同时,模态信号分解技术与注意力机制的集成进一步提升了模型对复杂数据结构的理解和处理能力。

3) 采用纠偏自适应矩估计优化技术对模型进行训练和优化,能够在更少的训练过程中获得更高的精度,显著提高了模型的训练速度和效率。

这一模型的提出和应用为铁矿石市场的价格预测和风险管理给出了新的思路和方法,期望能够为钢铁企业及相关决策者提供一种更加科学、准确的铁矿石价格预测工具,从而更好地应对市场变化和风险挑战。

1 模型介绍与建模流程

增强式集成模型由 5 个核心模块构成:模态分解层、基于多头卷积神经网络(multi-head convolutional neural network, MCNN)的编码器网络、基于长短期记忆神经网络(long short-term memory, LSTM)的解码器网络、注意力机制层以及输出层。在训练阶段,增强式集成模型通过纠偏自适应矩估计优化算法学习输入变量和输出变量之间的潜在关系。下面将详细阐述增强式集成模型的结构设计与各模块的功能,并重点介绍纠偏自适应矩估计优化算法的原理及其在模型训练中的应用。

1.1 增强式集成模型

本文提出的增强式集成模型架构见图 1。该架构首先从原始复杂的铁矿石时间序列中提取模态特征,利用时频分析技术中的信号分解技术将时间序列数据进行模态分解,分解成频率成分较为简单的子序列;其次,构建基于多头卷积神经网络的编码器网络,通过逐步处理输入时间步长,将时序数据编码为固定维度的特征向量,实

现模态特征的自动提取;随后,采用堆叠式长短期记忆网络(stacked long short-term memory, SLSTM)构建解码器网络,对编码特征进行时序建模;最后引入注意力机制层,动态加权关键特征以生成预测输出。此外,采用纠偏自适应矩估计优化技术对模型进行训练和优化,能够在更少的训练过程中获得更高的精度,有效提高了模型的训练速度和效率。这种增强式集成架构通过融合信号处理技术与深度学习方法的优点,实现了对铁矿石价格时序数据多层次特征的协同挖掘与建模。

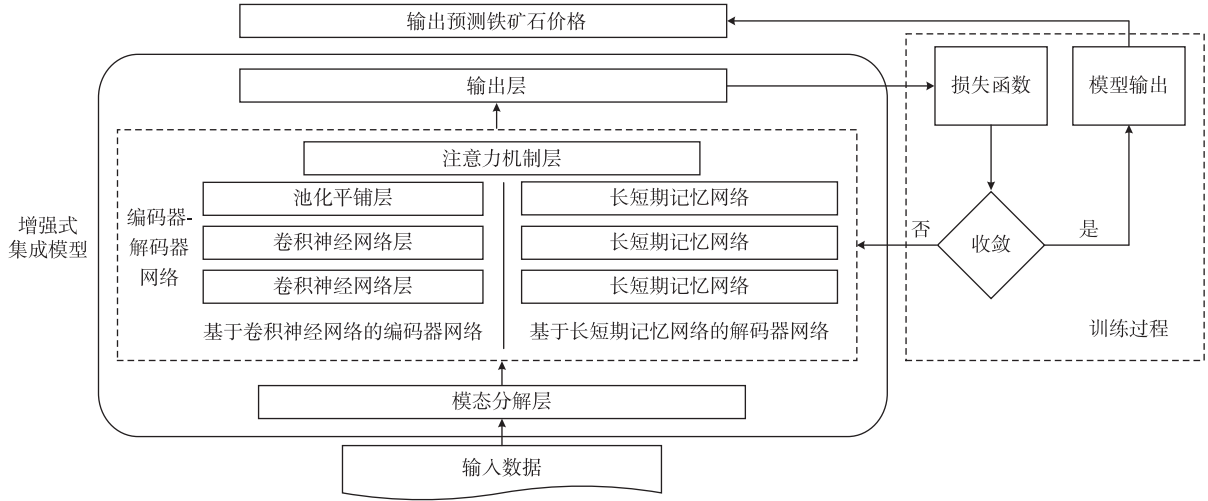


图1 增强式集成模型结构

Fig. 1 Structure of the enhanced ensemble model

1.1.1 变分模态分解

模态分解层采用变分模态分解(variational mode decomposition, VMD)方法来构建,主要用于对铁矿石价格数据的模态分解。VMD方法是一种自适应信号处理方法,对原始复杂时间序列进行分解,得到数据各分量序列对应频率特征信息:本征模态分量(intrinsic mode function, IMF)和趋势项,再对各分量数据进一步挖掘并有效地刻画对应特征。VMD克服了经验模态分解的端点混乱效应和模态大小问题,可以有效分解低频和高频信号,降低时间序列的高复杂性、非线性和非平滑性^[24]。VMD的目标是将铁矿石价格数据作为输入信号 $f(x)$,将它分解为 k 个稀疏特征的离散子信号(模态) $u_k(x)$,假设每个模态 $u_k(x)$ 都是在频率中心 ω_k 附近均为紧凑的分布,每个模态 $u_k(x)$ 在分解过程中被自适应地确定。所有的离散子信号(模态) $u_k(x)$ 的带宽之和最小为目标函数,具体步骤如下。

步骤1:信号分解。对于每个模态 $u_k(x)$,通过傅里叶变换计算相关的解析信号,并构造出单边频谱,每个模态 $u_k(x)$ 的解析信号为:

$$\left(\delta(x) + \frac{j}{\pi x}\right) * u_k(x), \quad (1)$$

式中: $\delta(x)$ 是单位脉冲函数,*表示函数的点积操作, j 代表虚数单位, x 表示时间戳。

步骤2:信号调制。通过每个模态 $u_k(x)$ 解析信号的中心频率 ω_k ,将模态 $u_k(x)$ 的频谱移于指数项混合调到对应的基带,表达式为:

$$\left[\left(\delta(x) + \frac{j}{\pi x}\right) * u_k(x)\right] e^{-j\omega_k x}, \quad (2)$$

式中: ω_k 为第 k 个模态对应的频率中心。

步骤3:变分优化模型构建。通过解调信号的高斯平滑度,即梯度的 L_2 范数规则化来估计模态 $u_k(x)$ 的带宽。并探索输入信号的 k 个离散子信号(模态) $u_k(x)$,使得估计的带宽总和最小。因此,变分优化模型的估计带宽模式表达式为:

$$\begin{cases} \min \left\{ \sum_k \left\| \partial_x \left[\left(\delta(x) + \frac{j}{\pi x}\right) * u_k(x) \right] e^{-j\omega_k x} \right\|_2^2 \right\}, \\ \sum_{k=1}^K u_k(x) = f(x). \end{cases} \quad (3)$$

式中: ∂_x 表示偏导数算子。

步骤 4: 约束优化问题转化。为了求解该优化问题, 考虑通过构造增广拉格朗日函数来寻找上述问题的最优解, 将带约束条件的变分问题转换为如下无约束条件的变分问题:

$$L(\{u_k(x)\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_x \left[\left(\delta(x) + \frac{j}{\pi x} \right) * u_k(x) \right] e^{-j\omega_k x} \right\|_2^2 + \left\| f(x) - \sum_k u_k(x) \right\|_2^2 + \langle \lambda(x), f(x) - \sum_k u_k(x) \rangle, \quad (4)$$

其中: α 是二次惩罚约束的平衡因子, λ 是拉格朗日乘数。 $\left\| f(x) - \sum_k u_k(x) \right\|_2^2$ 是为了迅速收敛的约束项。对于上述无约束变分问题, 可采用交替方向乘子法来求解增广拉格朗日函数的鞍点, 并交替地更新变分优化模型参数 $u_k(x)$, ω_k , λ , 如下所示:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(x) - \sum_{i \neq k} \hat{u}_i^{n+1}(\omega) + \hat{\lambda}^n(\omega)/2}{1 + 2\alpha(\omega - \omega_k^n)^2}, \quad (5)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega}, \quad (6)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \gamma(\hat{f}(x) - \sum_k \hat{u}_i^{n+1}(\omega)). \quad (7)$$

其中: γ 是噪声容量, $\hat{u}_i^{n+1}(\omega)$, $\hat{u}_i(\omega)$, $\hat{f}(\omega)$, $\hat{\lambda}(\omega)$ 分别是 $u_i^{n+1}(x)$, $u_i(x)$, $f(x)$, $\lambda(x)$ 的傅里叶转换值。迭代的收敛条件是 $\sum_k \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2 / \|\hat{u}_k^n\|_2^2 < \epsilon$, 其中 n 表示迭代的次数, ϵ 表示迭代的判别度。

1.1.2 编码器-解码器网络

通过模态分解层处理后的铁矿石价格数据被分解为若干频率成分简化的本征模态函数 (intrinsic mode functions, IMFs) 子序列。然而, 在处理这些具有不同尺度和维度的子序列时, 传统的序列到序列 (Seq2Seq) 预测方法面临挑战, 主要体现在输入序列与输出序列间可能存在非对称长度映射问题。为解决这一技术难题, 本研究设计了一种端到端的编码器-解码器神经网络架构。该架构包含 2 个核心组件: 1) 编码器网络采用多头卷积神经网络结构, 负责将变长输入子序列编码为固定维度的潜在空间表征; 2) 解码器网络基于堆叠长短期记忆单元构建, 通过自回归方式将潜在表征解码为预测序列。这种双网络协同机制有效解决了模态分解后子序列的异质性处理问题, 同时保持了时序预测的连续性。

在编码器网络架构设计中, 本文采用 MCNN 作为核心特征提取模块 (图 2)。该设计基于以下关键考量: 每个独立的卷积神经网络 (convolutional neural network, CNN) 头部由并行的一维卷积层构成, 可针对不同模态子序列进行专属的特征提取。具体而言, 各 CNN 头部通过独立的卷积核组对输入时序数据进行空间特征映射, 经 ReLU 激活函数非线性变换后, 输出的特征图通过展平 (flatten) 操作进行维度压缩, 最终在特征融合层进行拼接与重塑。这种多头架构相较单路 CNN 具有明显优势: 不仅通过并行特征提取保留各子序列的独有模式特征; 而且增强了模型对跨模态关联特征的捕获能力。其中卷积层是基于 MCNN 的编码器结构的关键组成部分, 其中包含多个卷积核, 这些卷积层常被当作是卷积过程中的小窗口, 将数据分解层分解得到的模态特征进行特征映射操作 (图 2a), 然后再使用卷积核进行卷积操作, 并经过激活函数得到输出特征 (图 2b)。通过基于 MCNN 的编码器结构生成的新特征通常比数据分解层分解得到的模态特征更有效, 可以包含更多的关键特征。卷积层的提取过程是一个线性操作, 涉及到一组权重的乘法。具体的卷积操作结果的表达式如下:

$$\mathcal{C} = \sigma_{\text{sig}} \left(\sum_i \mathcal{N}_i * \mathbf{x}_i \right), \quad (8)$$

其中: $\mathbf{x}_i$ 表示子序列数据的第 i 个特征, $\mathcal{N}_i$ 表示卷积核的第 i 个特征, σ_{sig} 表示激活函数。随后, 将提取的卷积特征传输到池化层进行压缩。接下来, 经展平层去除多余的不必要信息后传输到全连接层, 进行特征信息的聚合。全连接层 $\mathbf{v}$ 的计算如下所示:

$$\mathbf{v} = \sigma_{\text{sig}} (\mathbf{W}_c * \mathcal{C} + \mathbf{b}_c), \quad (9)$$

式中: $\mathbf{W}_c$ 表示过滤器的权重, $\mathbf{b}_c$ 代表偏置。卷积层的操作是将模态特征的长序列编码为具有向量特征的较短序

列,特征提取过程如图2所示。每个卷积核扫描输入图像的不同区域,生成特征映射。

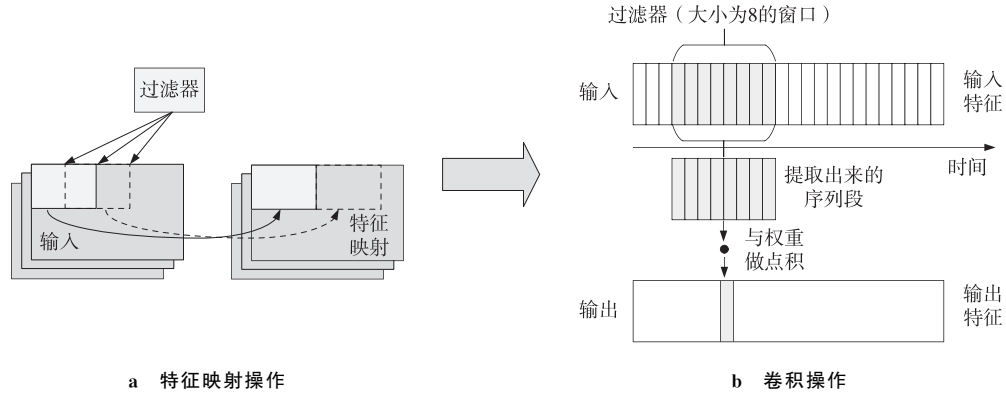


图2 卷积层的特征提取过程

Fig. 2 Process of feature extraction in convolutional layer

传统一维卷积神经网络(1D-CNN)在时序数据处理中存在固有局限性:局部感受的特性导致对时间步序的敏感性不足。为有效捕捉铁矿石价格序列中的长期依赖关系,本研究提出卷积-递归混合的编码器-解码器网络架构。该设计融合了2种神经网络的互补优势:1D-CNN通过分层卷积核执行局部特征提取,将原始序列转换为高阶特征表示;随后SLSTM对这些特征序列进行时序建模。由于传统基于循环神经网络(recurrent neural network, RNN)模型存在梯度消失和爆炸的局限性,所以Hochreiter等人^[25]引入了LSTM神经网络。LSTM网络是共享参数,每一时刻的输出结果与之前所有的输入有关,这种更深层的共享参数方式更有利于处理长时间序列数据。为了捕捉铁矿石价格和相关变量的长期依赖关系,本文模型的解码器网络中使用了多层LSTM网络的堆叠结构,其中单个LSTM网络结构单元如图3所示。解码器网络结构中采用LSTM来读取上下文向量,并逐步执行输出操作。LSTM网络是一种特殊的循环神经网络结构。它通过使用3个“门”结构(遗忘门、输入门、输出门)来控制信息进出网络单元,有效地避免了传统循环神经网络产生的梯度消失与梯度爆炸问题。其中,遗忘门决定LSTM需要从细胞状态中保留或忘记什么信息,通过引入遗忘门来学习长期依赖关系。

输入门决定LSTM需要在单元状态中记录哪些新信息。根据记忆单元的状态,输出门决定LSTM单元状态中哪些信息将被输出。LSTM中的这些门是通过输入、记忆和输出的信息进行加权,通过加权值来处理信息。记忆单元是LSTM的核心部分,它负责存储和传递序列中的信息。记忆单元代表一个长期的存储单元,它在序列数据的不同时刻进行接收和处理信息的操作,并决定哪些信息应该保留或丢弃。通过LSTM的门控结构,记忆单元可以长期地保持一些信息,而快速地遗忘一些无用信息。

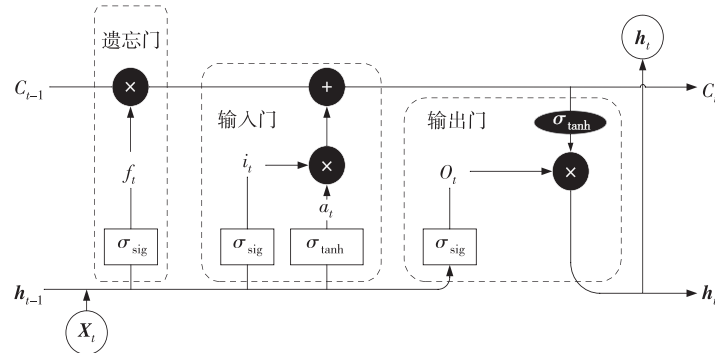


图3 单个LSTM单元的结构

Fig. 3 Structure of a single LSTM unit

首先,遗忘门主要是对上一个节点传进来的输入进行忘记不重要信息并记住重要信息的操作。通过接受输入向量 X_t 和隐藏状态向量 h_{t-1} ,并输出一个值 f_t ,该值决定从单元状态 C_{t-1} 中保留哪些信息到当前状态,遗忘门的表达式为:

$$f_t = \sigma_{\text{sig}}(\mathbf{W}_f * \mathbf{X}_t + \mathbf{W}_f * \mathbf{h}_{t-1} + \mathbf{W}_f * \mathbf{v} + \mathbf{b}_f), \quad (9)$$

式中: $\mathbf{v}$ 表示编码器输入信息的向量, σ_{sig} 表达式为:

$$\sigma_{\text{sig}} = \frac{1}{1 + e^{-x}}. \quad (10)$$

式中: x 表示输入的变量, σ_{sig} 将输入的变量映射到 $[0, 1]$ 之间。输入门决定什么新信息将被存储在 LSTM 记忆单元状态中, 表达式为:

$$i_t = \sigma_{\text{sig}}(\mathbf{W}_i * \mathbf{X}_t + \mathbf{W}_i * \mathbf{h}_{t-1} + \mathbf{W}_i * \mathbf{v} + \mathbf{b}_i). \quad (11)$$

通过输入门的计算, 会产生一个 LSTM 记忆单元状态中的信息候选值 a_t , 表达式为:

$$a_t = \sigma_{\text{tanh}}(\mathbf{W}_a * \mathbf{X}_t + \mathbf{W}_a * \mathbf{h}_{t-1} + \mathbf{W}_a * \mathbf{v} + \mathbf{b}_a), \quad (12)$$

式中: σ_{tanh} 表示双曲正切函数, 表达式为:

$$\sigma_{\text{tanh}} = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (13)$$

通过剔除过去单元状态的部分信息以及融合信息候选值, 过去单元状态 C_{t-1} 将被融入到当前的单元状态 C_t 中(即哪些当前数据将保存在单元状态中), 表达式为:

$$C_t = f_t * C_{t-1} + i_t * a_t. \quad (14)$$

最后, 输出门对单元状态 C_t 进行筛选并决定哪些单元状态信息被输出, 表达式为:

$$o_t = \sigma_{\text{sig}}(\mathbf{W}_o * \mathbf{X}_t + \mathbf{W}_o * \mathbf{h}_{t-1} + \mathbf{W}_o * \mathbf{v} + \mathbf{b}_o), \quad (15)$$

$$\mathbf{h}_t = o_t * \sigma_{\text{tanh}}(C_t). \quad (16)$$

式(9)、(11)、(12)及(15)中: 参数 $\mathbf{W}_f, \mathbf{W}_i, \mathbf{W}_o, \mathbf{W}_a$ 表示合适的权重矩阵, 参数 $\mathbf{b}_f, \mathbf{b}_i, \mathbf{b}_o, \mathbf{b}_a$ 表示对应的偏置向量, 这些参数训练过程中可由优化算法确定。* 是向量中对应元素的乘法运算。记忆单元是 LSTM 的核心部分, 它负责存储和传递序列中的信息。LSTM 的记忆单元本质上是具有状态持久性的时序存储单元, 通过微分门控实现信息的选择性保持与遗忘。通过 LSTM 的门控结构, 记忆单元可以长期地保持有用信息, 并且快速地遗忘无用信息。然而, 传统 LSTM 对所有时序特征采用均匀处理策略, 未能充分考虑输入数据的局部重要性差异。为此, 本研究引入注意力机制构建层次化特征选择架构。

注意力层是由注意力机制构成, 每个注意力层的神经元都有一个权重数组, 大小与前一层的神经元数量相同。其中注意力层的引入, 是让预测模型从产生的输出变量中进行矿产价格预测, 对产生的变量中不同局部信息进行筛选, 将足够的注意力集中在关键的特征信息上。在每个时间步, 注意力机制都会通过计算的权重对解码器的先前隐藏状态和编码向量列表进行缩放, 实现更高质量的特征提取。然后对缩放后的向量生成非归一化的得分值, 该得分值指示输入序列的元素与当前输出序列元素之间的对齐程度, 以此来表明信息的重要性。

将注意力机制层与编码器-解码器网络相结合, 以提高编码器-解码器模型的信息利用率, 进一步提高模型的预测性能。注意力机制层将一个来自编码器和解码器的输入序列 $\mathbf{y}$, 通过以下迭代的计算步骤, 映射得到一个预测输出序列 $\mathbf{y} = (y_0, y_1, y_2, \dots, y_t)$, 其中 t 是预测周期。对于时间序列 $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_t)$, 使用注意力层中的注意力机制计算每个时刻 $\mathbf{h}_i$ 的注意力权重, 通过计算查询向量 $\mathbf{q}$, 来计算它们之间的相关性, 并为时间序列赋予注意力权重 $\mathbf{a}_i$, 其中查询向量与模型要达到的目标相关。为了使注意力机制的权重向量变成一个概率分布, 通常使用激活函数 Softmax 对该权重向量进行处理, 计算公式如下:

$$\mathbf{a}_i = \text{Softmax}(s(\mathbf{q}, \mathbf{h}_i)), \quad (17)$$

$$\text{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_{i=1}^t \exp(x_i)}. \quad (18)$$

式中: $s(\cdot)$ 表示 $\mathbf{q}$ 和 $\mathbf{h}_i$ 之间的相关性计算公式。表达式为:

$$s(\mathbf{q}, \mathbf{h}_i) = \mathbf{v}^T \tanh(\mathbf{W}\mathbf{h}_i + \mathbf{U}\mathbf{q}) \quad (19)$$

式中: $\mathbf{W}, \mathbf{U}$ 代表可训练的参数矩阵。在计算出相关性 $s(\mathbf{q}, \mathbf{h}_i)$ 以及注意力权重 $\mathbf{a}_i$ 后, 将编码器网络最后一个时刻输出 $\mathbf{h}_t$ 作为查询向量 $\mathbf{q}$, 采用注意力机制加权融合的方式, 将注意力权重 $\mathbf{a}_i$ 与 Encoder 在每个时刻的输出 $\mathbf{h}_i$, 作为 Encoder 的编码结果, 融合加权的计算公式如下:

$$\mathbf{c} = \sum_{i=1}^t \mathbf{a}_i h_i \quad (20)$$

式中: $\mathbf{a}_i$ 是每个时刻输出的权重, $\mathbf{c}$ 为编码器网络的输出结果。然后将编码器网络的输出结果与解码器网络每个时刻的隐含层状态进行拼接,得到新的每个时刻的隐含层状态的输出 $\mathbf{h}'_i$ 。计算注意力的隐含层状态过程如下:

$$\mathbf{h}'_i = \tanh(\mathbf{W} \cdot \text{concat}(\mathbf{c}, \mathbf{h}_i)) \quad (21)$$

由于在使用注意力机制拼接后,状态的维度进行改变,所以通过 concat 方式,将拼接后的结果与全连接矩阵 $\mathbf{W}$ 进行相乘,来对状态的维度进行恢复。最后,通过使用新的注意力机制得到解码器每个时刻的输出 $\mathbf{h}'_i$,计算出每个时刻的输出 Y_i 。输出过程的计算如下:

$$Y_i = \mathbf{W}_{\text{De}} \mathbf{h}'_i + \mathbf{b}_{\text{De}} \quad (22)$$

式中: $\mathbf{W}_{\text{De}}$ 和 $\mathbf{b}_{\text{De}}$ 分别表示解码器的参数矩阵和偏置向量,可以在训练过程进行优化中,因此选择合适的优化算法对于提高模型的收敛速度和预测精度至关重要。

1.2 纠偏自适应矩估计优化算法

自适应矩估计(adaptive moment estimation, Adam)算法是常见的神经网络训练优化算法,通过将梯度下降与动量相结合,对一阶和二阶矩估计进行自适应调整,提升了预测模型的泛化能力^[26]。然而,由于 Adam 算法与许多自适应学习率优化器一样,在训练初期,如果不使用预热方法,学习率可能存在方差较大的情况,从而导致初期训练不太稳定,可能会收敛到局部最优解。而预热方法是在模型训练的初始阶段使用一个极小的学习率,来抵消方差过大的问题。因此,本文采用纠偏自适应矩估计(rectified Adam, RAdam)算法来修正自适应学习率的方差,RAdam 通过插入一个估计自由度的选项解决了 Adam 算法在训练初期遇到的稳定性问题^[26]。这一改进使得 RAdam 在保持 Adam 算法快速收敛优点的同时,进一步提高了训练的稳定性和准确性。本文在搭建预测模型时,为了防止数据过拟合现象,引入了 Dropout 层。Dropout 层通过随机丢弃网络中的部分神经元,减少神经元之间的复杂共适应关系,从而提高模型的泛化能力,具体的优化步骤如下。

算法 1 RAdam 算法

输入:初始参数 θ ,定义学习率 α 为数值稳定量,一阶矩移动平均衰减系数 β_1 ,二阶矩移动平均衰减系数 β_2 ,优化目标函数 $f_t(\theta)$

输出:最终参数 θ_t

1. 初始化 $\mathbf{m}_0 \leftarrow 0, \mathbf{v}_0 \leftarrow 0$

2. 初始化 $\theta_0 \leftarrow \text{Gauss}()$

3. 初始化指数加权平均长度 $\rho_\infty = 2/(1-\beta_2) - 1$

For $t \leftarrow 1$ to T do

4. 计算第 t 步迭代过程中的梯度 $\mathbf{g}_t = \nabla_{\theta} f_t(\theta_{t-1})$

5. 计算一阶矩 $\mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + (1-\beta_1) \mathbf{g}_t$

6. 计算二阶矩 $\mathbf{v}_t = \beta_2 \mathbf{v}_{t-1} + (1-\beta_2) \mathbf{g}_t^2$

7. 计算学习率 $\alpha_t = (\alpha \sqrt{1-\beta_2^t}) / (1-\beta_1^t)$

8. 计算指数移动加权平均值的修正 $\hat{\mathbf{m}}_t = \mathbf{m}_t / (1-\beta_1^t)$

9. 计算近似指数加权平均长度 $\rho_t = \rho_\infty - 2t\beta_2^t / (1-\beta_2^t)$

10. If $\rho_t > 4$ then

11. 计算二阶矩的修正值 $\hat{\mathbf{v}}_t = \mathbf{v}_t / (1-\beta_2^t)$

12. 计算方差校正项 $r_t = \sqrt{\frac{(\rho_t-4)(\rho_t-2)\rho_\infty}{(\rho_\infty-4)(\rho_\infty-4)\rho_t}}$

13. 使用自适应动量进行参数更新 $\theta_t = \theta_{t-1} - \alpha_t r_t \hat{\mathbf{m}}_t / \sqrt{\hat{\mathbf{v}}_t}$

14. Else

使用非自适应动量进行参数更新 $\theta_t = \theta_{t-1} - \alpha_t \hat{\mathbf{m}}_t$

End for

算法的输入是自定义学习率 α_t 、用于生成一阶矩移动平均衰减系数 β_1 、移动二阶矩平均衰减系数 β_2 和优化目标函数。该算法返回输出参数 θ_t 。该算法首先将移动一阶矩 $\mathbf{m}_0$ 和移动二阶矩 $\mathbf{v}_0$ 初始化为 0。然后初始化指数加权平均长度 ρ_∞ ，指数加权移动平均可以近似地表示简单移动平均长度。然后，计算自适应学习率 α_t 。算法更新指数移动二阶矩 $\mathbf{v}_t$ 和指数移动一阶矩 $\mathbf{m}_t$ ，其中 $\mathbf{g}_t$ 是对应于时间步长 t 的随机梯度。当平均值初始化为 0 时，Adam 中的估计量偏向于 0，特别是衰减率很低。因此，RAdam 通过计算经过偏差校正的移动平均 $\hat{\mathbf{m}}_t$ 来应用偏差校正技术，该技术稍后用于更新参数。该算法还计算了近似指数加权移动平均线的长度 ρ_t 。使用上述计算的值，该算法解决方差判断与参数更新。如果方差很大(即 $\rho_t > 4$)，则算法处于预热阶段，此时，计算移动量二阶矩的修正值 $\hat{\mathbf{v}}_t$ 和方差校正项 r_t ，使用这些值和自适应动量来更新参数；如果方差很小(即预热阶段已经完成)，则参数会通过非自适应动量进行更新(不使用移动动量二阶矩和方差校正项)。

2 实证分析

本节首先介绍了实验设置的细节，包括数据集的细节、用于评估的指标和对比模型的参数。然后，通过实验对模型的性能进行评估分析。本文模型在基于 Tensorflow1.15.0 的深度学习框架上进行实验，实验运行环境为 Intel(R) Core(TM) i7-7700 CPU @ 3.60 GHz 的 Windows 系统，内存为 32 GB。

2.1 实验设置

2.1.1 实验数据集

本文选取的 2 组铁矿石数据集，分别来自大连商品期货交易所和新加坡交易所，代表了国内外铁矿石市场的 2 个重要视角，数据均来源于万得数据库(<https://www.wind.com.cn/portal/zh/EDB/index.html>)。其中，数据集 A 的数据是大连商品期货交易所铁矿石期货合约每日结算价，因为大连商品期货交易所是全球交易量最大的铁矿石期货交易市场，反映了国内市场对铁矿石价格的预期和交易情况；时间范围是 2014 年 1 月 2 日至 2024 年 1 月 3 日，共 2 492 组数据；这些价格数据被分为训练集和测试集 2 部分，前 1 994 d 的价格数据作为训练集，剩下 498 d 的价格数据作为测试集。数据集 B 的数据是新加坡交易所铁矿石价格指数，因为新加坡交易所是国际铁矿石衍生产品的发源地，铁矿石指数在国际贸易中占有重要地位，特别是普式铁矿石指数的 62% 品位 CFR 中国青岛港价格是国际铁矿石贸易中的主要定价基准；时间范围同样为 2014 年 1 月 2 日至 2024 年 1 月 3 日，共 2 439 组数据；这些价格数据同样被分为训练集和测试集 2 部分，前 1 951 d 的价格数据作为训练集，剩下 488 d 的价格数据作为测试集。在所有的实验中，首先用最小/最大归一化方法对每个时间序列的输入数据进行归一化处理。

2.1.2 评价指标

本文使用 4 种最常用的评估指标来评估不同预测模型的性能：均方根误差(e_{RMSE})、平均绝对误差(e_{MAE})、平均绝对百分比误差(e_{MAPE})和决定系数(R^2)，表达式分别为：

$$e_{\text{RMSE}} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}, \quad (23)$$

$$e_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (24)$$

$$e_{\text{MAPE}} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%, \quad (25)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}. \quad (26)$$

式中： N 为测试集中值的个数， $\hat{y}_i$ 为第 i 个测试用例的预测铁矿石价格， y_i 是第 i 个测试用例的实际铁矿石价格， $\hat{y}_i$ 和 y_i 为归一化值， $\bar{y}$ 表示观察到的样本的平均值。

2.1.3 基线方法和参数设置

为了验证所提模型的优越性,需要与多种基线模型进行比较,并在模型之间进行详细分析。具体来说,本文模型与对比模型的实验参数设置如表1所示。参与对比模型有:长短期记忆网络(LSTM)、双向长短期记忆网络(BiLSTM)、堆叠长短期记忆网络(SLSTM)、卷积神经网络(CNN)、多头卷积神经网络(MCNN)、卷积堆叠神经网络(CNN-SLSTM)、多头卷积堆叠神经网络(MCNN-SLSTM)、堆叠注意力网络(SLSTM-AT)、多头卷积注意力网络(MCNN-AT)、基于注意力的多头卷积堆叠长短期记忆网络(MCNN-SLSTM-AT)和增强式集成模型(Ensemble)。其中,将 BiLSTM 和 SLSTM 与单个 LSTM 进行比较,以反映不同长短期记忆层模型对提高预测性能的影响。这些实验为本研究选择最佳的长短期记忆层模型提供了依据。同样,将 MCNN-SLSTM、CNN-SLSTM 和 SLSTM 的比较,证明加入 CNN 结构的 SLSTM 可以进一步提高精度,特别是 MCNN 可以自动提取多个时间序列中的特征信息。然后,将 AT 与 SLSTM 和 MCNN 分别结合进行预测分析和比较,进一步证明加入注意机制的 SLSTM 和 MCNN 模型可以对预测精度产生提高的作用。之后,将 MCNN-SLSTM-AT 与 SLSTM-AT 和 MCNN-AT 进行比较,证明前者的预测精度更高。最后,将 MCNN-SLSTM-AT 与本文模型的预测结果进行比较,证明选择增强式集成方法的精度最高,它将各神经网络模型的模式结合起来进行预测。

表1 基线模型的实验参数

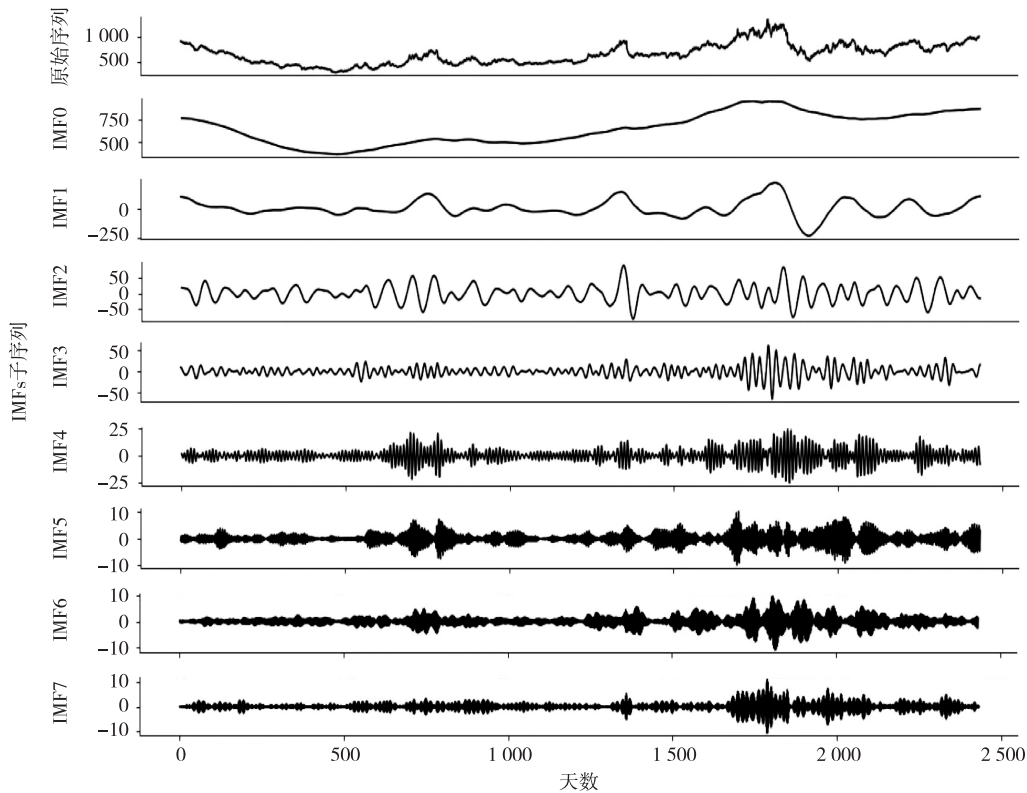
Tab.1 Experimental parameters of the baseline model

模型	层数	单元数	丢弃率	迭代次数	批量数	学习率
LSTM	1	50	0.08	200	32	0.1
CNN	1	50	0.09	200	32	0.1
BiLSTM	2	50	0.085	200	32	0.1
MCNN	2	50	0.085	200	32	0.1
SLSTM	3	50	0.085	200	32	0.1
MCNN-AT	3	50	0.085	200	32	0.1
CNN-SLSTM	4	64	0.09	200	32	0.1
SLSTM-AT	4	64	0.09	200	32	0.1
MCNN-SLSTM	5	64	0.09	200	32	0.1
MCNN-SLSTM-AT	5	64	0.09	200	32	0.1
Ensemble	6	64	0.09	200	32	0.1

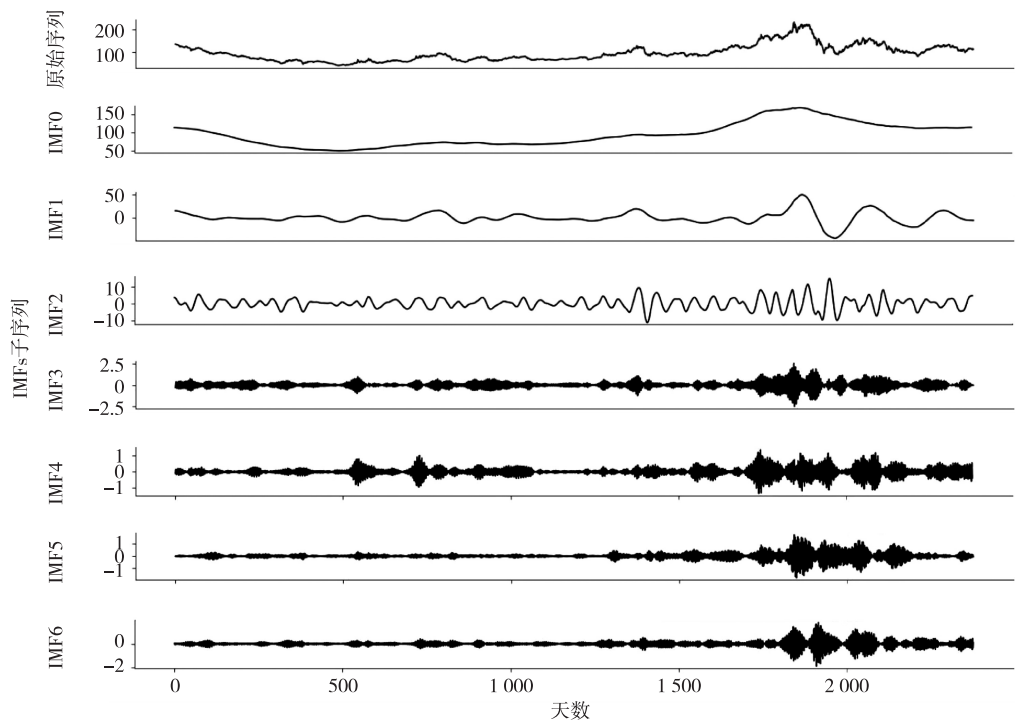
超参数设置对模型性能有重要影响。对于本文提出的增强式集成模型,信号分解层的参数设置主要取决于人工设定的模数 K 值^[27]。如果 K 值太小,可能会忽略原始数据中的一些重要信息,导致预测精度不足;如果 K 值太大,则相邻模态分量的中心频率可能很接近,从而导致模态重复或产生额外的噪声。这都不利于实现高准确性。因此,正确确定模态数 K 值至关重要。目前,已经提出了许多方法来确定 K 值,其中常用的方法是参考 EMD 算法自适应分解的模式数来确定。本文为了提高参数选择的效率,首先将原始数据通过 EMD 自动分解为多个模态,可以为 VMD 提供 K 的初始值^[28]。然后,通过实验进一步优化 K 值。这样,通过迭代实验验证,将数据集 A 和数据集 B 的 K 值分别设置为 8 和 7。VMD 算法对 2 个数据集的第 1 次分解结果如图 4 所示。除 K 值外,其他 VMD 参数如惩罚因子 α 、收敛精度 ϵ 均初始化为默认值;例如, $\alpha=2\ 000$, $\epsilon=10^{-7}$ 。为了每个模型的预测性能,对各实验模型的参数进行了优化,保证所有模型都是单独训练来实现最佳的预测性能,每个模型运行 100 次,并使用预测模型的平均结果。RAdam 模型的参数设定:最大迭代次数 $T=5\ 000$, $\alpha=0.01$, $\beta_1=0.01$, $\beta_2=0.02$ 。

2.2 结果与讨论

本节详细介绍了所提出的增强式集成模型与其他基准模型之间的性能比较。时间步长的选择将直接影响模型的效果。时间步长太短会导致模型无法完全选择输入特征,而时间步长太长可能会导致过拟合。首先对时间窗口设置为 1 d,进行预测分析。然后对不同时间窗口的预测结果进行广泛讨论。



a 数据集 A 的分解结果



b 数据集 B 的分解结果

图 4 数据集的分解结果

Fig. 4 Decomposition results of the dataset

2. 2. 1 提前一步预测

表 2 显示了在不同数据集下所有模型的预测结果。通过观察发现本文提出的模型(Ensemble)在所有的评估指标(e_{RMSE} , e_{MAE} , e_{MAPE} 和 R^2)方面始终优于对比模型。

表2 不同模型在4个指标下的比较结果

Tab. 2 Comparative analysis of different models under four metrics

模型	数据集 A				数据集 B			
	e_{RMSE}	e_{MAE}	e_{MAPE}	R^2	e_{RMSE}	e_{MAE}	e_{MAPE}	R^2
LSTM	13.813	8.825	6.281	0.902	12.934	8.726	6.137	0.907
CNN	11.951	7.973	5.899	0.911	11.737	7.821	5.691	0.910
BiLSTM	10.034	6.912	5.385	0.918	10.549	6.688	5.646	0.920
MCNN	9.816	6.731	4.797	0.921	9.941	6.917	4.851	0.925
SLSTM	9.432	6.511	4.541	0.928	9.533	6.781	4.687	0.926
MCNN-AT	9.038	6.321	4.237	0.934	8.979	6.297	4.188	0.932
CNN-SLSTM	8.221	5.471	3.816	0.942	8.243	5.297	3.848	0.942
SLSTM-AT	7.394	4.968	3.161	0.957	7.521	5.071	3.316	0.955
MCNN-SLSTM	6.525	4.577	2.794	0.961	6.941	4.869	2.963	0.957
MCNN-SLSTM-AT	5.373	3.509	2.525	0.966	5.571	3.671	2.639	0.963
Ensemble	4.847	3.331	2.485	0.969	4.513	3.227	2.382	0.971

根据表2的实验结果分析,可以得出以下重要结论。

首先,Ensemble模型的整体性能具有优势。在数据集A上的 e_{RMSE} 、 e_{MAE} 和 e_{MAPE} 分别为4.847、3.331和2.485;在数据集B上的 e_{RMSE} 、 e_{MAE} 和 e_{MAPE} 分别为4.513、3.227和2.382。误差指标明显低于其他基准模型。同时,Ensemble模型的 R^2 在数据集A上为0.969,在数据集B上为0.971,是所有基准模型中最高的。这表明本文提出的Ensemble模型具有最优的综合预测性能。

其次,对单个LSTM、BiLSTM和SLSTM的结果比较表明,后者在2个数据集上都具有较低的 e_{RMSE} 、 e_{MAE} 、 e_{MAPE} 和较高的 R^2 。误差指标降低大约10%, R^2 提高大约0.021。多层LSTM模型的效果比单一模型要好。将CNN和MCNN的结果进行比较后发现,基于多层的CNN模型优于单个CNN模型。MCNN相较基础CNN实现在数据集A的 e_{RMSE} 上降低了14.2%,在数据集B的 e_{MAE} 上降低了9.8%。这是因为多层CNN可以对原始铁矿石序列数据的特征进行多次提取。

第三,将SLSTM-AT和SLTM相比较,前者在2个数据集上都具有较低的 e_{RMSE} 、 e_{MAE} 、 e_{MAPE} 和较高的 R^2 。在数据集A上的误差降低了9.6%,在数据集B上的 R^2 提升了0.017。这是因为注意机制的加入使SLSTM模型能够从输入数据中提取更重要的时间信息。而MCNN-AT与SLSTM-AT的实验结果显示,多层CNN在数据集A上的预测误差降低了10%左右, R^2 提升了0.015。多层模型(SLSTM-AT)可以提高铁矿石价格预测的精度。

第四,将CNN-SLSTM和MCNN-SLSTM相比较,后者在2个数据集上都具有较低的 e_{RMSE} 、 e_{MAE} 、 e_{MAPE} 和较高的 R^2 。这是因为多层CNN的加入使SLSTM模型能够从输入数据中提取更重要的时间信息。这表明多层CNN可以有效地提高SLSTM的预测精度。MCNN-SLSTM-AT和Ensemble相比较,后者在2个数据集上都具有较低的 e_{RMSE} 、 e_{MAE} 、 e_{MAPE} 和较高的 R^2 。模态分解技术使误差降低了8.2%。这是因为加入模态分解后,分解得到的相对高频的模态包含了原始铁矿石价格数据的随机性,高频模态捕获价格波动细节。而相对低频的模态主要对应原始铁矿石价格数据的长期趋势。然后将分解的模态与其他多变量数据结合,减少了不同尺度特征信息之间的干扰。因此,可以更有效地处理原始铁矿石价格中的复杂信息,使得Ensemble比其他基准模型具有更好的预测性能。

为了明确基准模型与本文模型之间的差距,通过计算本文模型与其他基线模型之间的精度指标的改进率。 e_{RMSE} 、 e_{MAE} 、 e_{MAPE} 和 R^2 的改进率,可以统一表示为:

$$R_M = \left| \frac{M_p - M_b}{M_b} \right| \times 100\% \quad (27)$$

其中: M_p 表示 Ensemble 模型的各种指标(e_{RMSE} , e_{MAE} , e_{MAPE} 和 R^2), M_b 表示其他基线模型的指标(e_{RMSE} , e_{MAE} , e_{MAPE} 和 R^2)。图 5 描述了本文提出的增强式集成模型在一步超前预测中相对于其他模型的改进率。

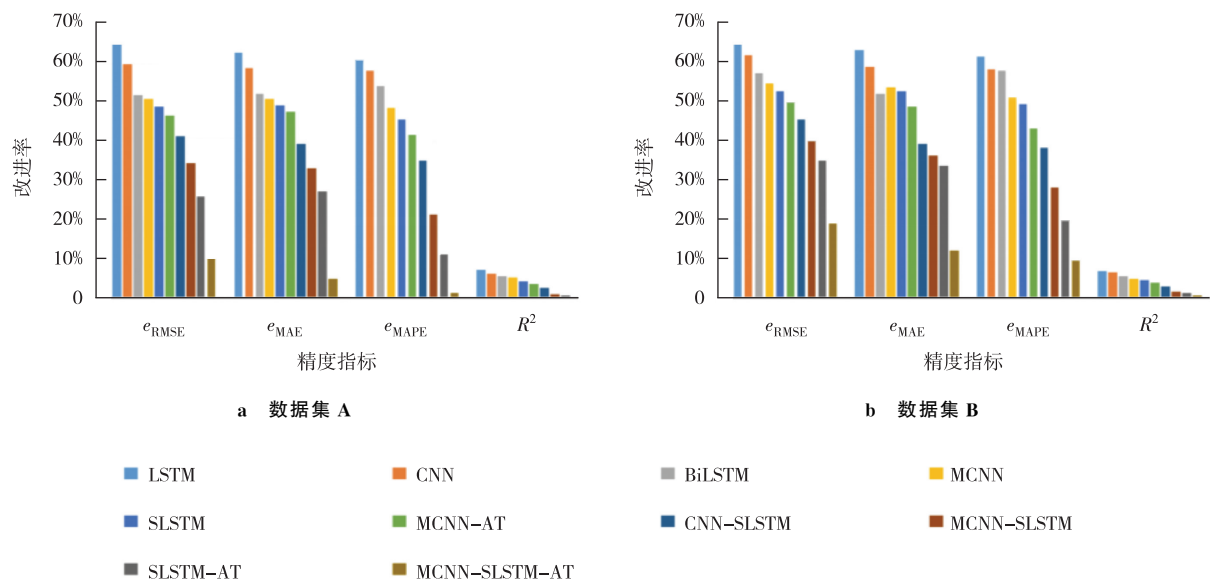


图 5 提出的模型对时间窗为一步预测的不同指标的改进率

Fig. 5 The improved rate of the proposed model for different indicators within one-step prediction

2.2.2 多步预测

精确的多步超前预测在铁矿石市场决策支持系统中具有关键作用,为决策者提供了更多时间来有效应对铁矿石价格波动,从而为决策者提供了理论支持。例如,矿业企业特别是钢铁企业可以根据铁矿石市场预测调整产品结构,生产利润更高、适销对路的铁矿石产品。生产调整需要响应时间,因此提前多步时间窗口预测对铁矿石企业来说是有价值的。本文实施多步时间窗口下的预测,表 3 总结了时间窗口为 4 d 的预测结果。

表 3 不同模型在时间窗口为 4 d 的预测结果比较

Tab. 3 Comparative analysis of different models under a four-day time window

模型	数据集 A				数据集 B			
	e_{RMSE}	e_{MAE}	e_{MAPE}	R^2	e_{RMSE}	e_{MAE}	e_{MAPE}	R^2
LSTM	14.283	10.255	6.791	0.891	13.913	10.426	6.751	0.892
CNN	12.591	9.733	6.179	0.918	12.337	9.6962	6.387	0.912
BiLSTM	11.683	9.291	5.996	0.926	11.481	9.168	5.972	0.929
MCNN	11.432	9.161	5.641	0.933	11.313	9.015	5.459	0.936
SLSTM	10.592	8.851	5.512	0.938	10.816	8.741	5.212	0.939
MCNN-AT	10.318	8.181	5.421	0.941	10.541	8.347	5.169	0.942
CNN-SLSTM	10.128	7.821	5.143	0.945	10.218	7.812	5.014	0.946
SLSTM-AT	9.221	6.571	4.891	0.949	9.432	6.697	4.919	0.952
MCNN-SLSTM	8.525	5.877	4.648	0.955	8.913	5.971	4.736	0.956
MCNN-SLSTM-AT	7.373	5.116	4.216	0.963	8.397	5.771	4.595	0.959
Ensemble	5.547	3.251	3.493	0.966	6.897	3.319	4.512	0.961

如表 3 所示,本文提出的增强式集成模型仍然可以完成准确的预测,并在指标对比中结果最佳。表 4 总结了时间窗口为 7 d 的预测结果。随着预测时间窗口长度的增加,每个模型的性能逐渐恶化。考虑到多步时间窗口预测误差的积累,这是可以理解的。

如表 4 所示,本文提出的增强式集成模型仍然可以产生较为理想的预测结果,并在与其他基准模型的指标对比中结果仍是最佳。表 5 总结了时间窗口为 30 d 的预测结果,此时,本文的增强式集成模型的性能退化率明显低于其他模型。在 4 个指标方面,它始终优于其他基准模型。各模型的性能排名与之前的一步超前预测的性能排名基本一致。综上所述,本文提出的增强式集成模型精度较高,在相对长期的铁矿石价格预测中具有明显优势。

表 4 不同模型在时间窗口为 7 d 的预测结果比较

Tab. 4 Comparative analysis of different models under a seven-day time window

模型	数据集 A				数据集 B			
	e_{RMSE}	e_{MAE}	e_{MAPE}	R^2	e_{RMSE}	e_{MAE}	e_{MAPE}	R^2
LSTM	16.834	12.855	7.151	0.875	16.373	12.524	7.191	0.873
CNN	16.591	12.173	6.987	0.891	16.163	12.461	6.989	0.889
BiLSTM	14.513	10.273	6.876	0.903	14.913	11.186	6.723	0.907
MCNN	14.304	10.608	6.726	0.912	14.025	10.383	6.654	0.917
SLSTM	12.125	9.159	6.174	0.917	11.935	9.548	5.891	0.918
MCNN-AT	11.381	9.018	5.913	0.921	11.235	9.012	5.772	0.924
CNN-SLSTM	10.812	8.317	5.717	0.926	10.807	8.317	5.701	0.928
SLSTM-AT	10.252	8.157	5.625	0.929	10.112	8.023	5.593	0.931
MCNN-SLSTM	9.159	6.936	5.041	0.936	8.907	6.664	4.928	0.941
MCNN-SLSTM-AT	8.816	6.616	4.855	0.941	7.613	6.313	4.316	0.955
Ensemble	7.218	5.621	4.162	0.954	7.105	6.015	4.088	0.959

表 5 不同模型在时间窗口为 30 d 的预测结果比较

Tab. 5 Comparative analysis of different models under a thirty-day time window

模型	数据集 A				数据集 B			
	e_{RMSE}	e_{MAE}	e_{MAPE}	R^2	e_{RMSE}	e_{MAE}	e_{MAPE}	R^2
LSTM	18.343	13.352	7.371	0.853	18.163	13.116	7.351	0.855
CNN	17.991	12.978	7.113	0.858	17.835	12.728	7.204	0.851
BiLSTM	16.483	12.091	7.025	0.877	15.525	11.403	6.827	0.883
MCNN	15.873	11.971	6.974	0.894	16.161	12.016	7.025	0.879
SLSTM	14.598	10.584	6.713	0.901	14.759	10.719	6.819	0.899
MCNN-AT	13.214	10.394	6.501	0.892	13.597	10.518	6.694	0.903
CNN-SLSTM	12.981	9.987	6.321	0.917	12.426	10.159	6.444	0.914
SLSTM-AT	12.261	9.216	6.115	0.911	12.361	10.011	6.294	0.913
MCNN-SLSTM	11.572	9.357	5.991	0.915	11.971	9.781	6.001	0.916
MCNN-SLSTM-AT	9.112	6.597	5.104	0.931	9.647	7.315	5.369	0.925
Ensemble	8.421	6.456	4.856	0.945	8.679	6.624	4.846	0.937

表 5 的实验结果表明,本文所提出的增强式集成模型在预测性能上明显优于其他单一神经网络模型(LSTM、CNN)和多层神经网络模型(BiLSTM、MCNN、SLSTM、MCNN-AT、CNN-SLST、SLSTM-AT、MCNN-SLSTM、MCNN-SLSTM-AT)。通过数据集 A 的 e_{RMSE} 结果表明,堆叠式 SLSTM 比单层 LSTM 提升了

12.1%;通过数据集 B 的 e_{MAE} 结果表明,多层模型(CNN-SLSTM)比单层模型(CNN)提升了 14.9%。根据多步时间窗口的实验结果表明,本文提出的增强式集成模型可以有效提取铁矿石数据的关键特征,提升预测模型的鲁棒性。

3 结论

针对铁矿石价格非线性强、动态波动明显以及单一预测模型性能受限的问题,本文提出了一种结合变分模态分解技术和编码器-解码器网络的铁矿石价格预测模型。该模型充分利用了变分模态分解技术在信号处理领域的优势,使用变分模态分解技术将铁矿石时间序列中的复杂波动性进行有效分解,提取出包含不同频率特征的子序列。随后,在特征提取层中设计了基于多头卷积神经网络结构来学习并提取这些子序列中的关键模态特征。然后采用基于堆叠长短期记忆网络结构来捕捉铁矿石价格中的中长期依赖关系和顺序信息特征,并对敏感信息进行分析和提取。通过注意力机制来将各分量的预测模型进行叠加和集成,并采用纠偏自适应矩估计算法作为优化算法对预测模型进行训练。纠偏自适应矩估计算法通过修正自适应学习率的方差,提高了训练的稳定性 and 收敛速度。通过在多种数据集的实验结果表明,本文提出的模型在不同长度的时间窗口下均表现出优异的预测性能,所有评估指标均明显优于基准模型。这一成果不仅验证了模型设计的合理性和有效性,更为钢铁企业准确地把握未来市场走势和制定适销对路的产品和市场交易提供了有力的支持。

参考文献:

- [1] 张松波,张磊,蔡启端,等.中国钢铁工业演化和铁矿石供需展望[J].中国矿业,2025,34(2):204-211.
ZHANG S B,ZHANG L,CAI Q D,et al. China's iron and steel industry evolution and iron ore supply and demand outlook[J]. China Mining Magazine,2025,34(2):204-211.
- [2] 刘清梅,张福明.中国钢铁制造流程发展与关键问题[J].钢铁,2025,60(1):15-28.
LIU Q M,ZHANG F M. Iron and steel manufacturing process development and critical issues in China[J]. Iron & Steel,2025,60(1):15-28.
- [3] 王安建.全球矿产资源战略研究:“战略性关键矿产研究”专辑特邀主编寄语[J].地球学报,2023,44(2):257-260.
WANG A J. Strategies in global mineral resources:guest editor's preface to “studies on strategic and critical mineral resources” [J]. Acta Geoscientica Sinica,2023,44(2):257-260.
- [4] 干勇,彭苏萍,毛景文,等.我国关键矿产及其材料产业供应链高质量发展战略研究[J].中国工程科学,2022,24(3):1-9.
GAN Y,PENG S P,MAO J W,et al. High-quality development strategy for the supply chain of critical minerals and its material industry in China[J]. Strategic Study of CAE,2022,24(3):1-9.
- [5] 卢延纯,赵公正,左更,等.破解铁矿石价格困境 构建钢铁产业新发展格局[J].价格理论与实践,2021(1):4-9.
LU Y C,ZHAO G Z,ZUO G,et al. Solving the iron ore price dilemma and building a new development pattern for the iron and steel industry[J]. Price; Theory & Practice,2021(1):4-9.
- [6] 周静言.新世纪以来国际铁矿石价格波动及对中国的启示[J].价格月刊,2024(5):77-83.
ZHOU J Y. The fluctuation of international iron ore price since the new century and its enlightenment to China[J]. Prices Monthly,2024(5):77-83.
- [7] 惠长虹,张华琨.铁矿石价格形成机制的演化机理及相关建议[J].宏观经济管理,2023(4):72-79.
HUI C H,ZHANG H J. The evolution of the pricing mechanism for iron ore[J]. Macroeconomic Management,2023(4):72-79.
- [8] 王晓峰,林立超.国际铁矿石期货价格传导比较研究:兼析加快我国铁矿石期货市场国际化建设[J].价格理论与实践,2020(3):95-98.
WANG X F,LIN L C. Comparative study on price transmission of international iron ore futures: concurrent analysis of accelerating the internationalization of my country's iron ore futures market[J]. Price (Theory & Practice),2020(3):95-98.
- [9] GUIDOTTI R. Counterfactual explanations and how to find them: literature review and benchmarking[J]. Data Mining and Knowledge Discovery,2024,38(5):2770-2824.
- [10] HOU L L,ZHANG Q,DU Y L. Width estimation of hidden cracks in tunnel lining based on time-frequency analysis of GPR data and back propagation neural network optimized by genetic algorithm[J]. Automation in Construction,2024,162:105394.
- [11] 袁晔,汤春阳,张博轩,等.融合时频特征的通信辐射源个体识别方法[J].电子测量技术,2025,48(1):129-136.

- YUAN Y, TANG C Y, ZHANG B X, et al. Individual identification method for communication radiation sources by integrating time-frequency characteristics[J]. *Electronic Measurement Technology*, 2025, 48(1): 129-136.
- [12] ABU DOUSH I, AHMED B, AWADALLAH M A, et al. Improving multilayer perceptron neural network using two enhanced moth-flame optimizers to forecast iron ore prices[J]. *Journal of Intelligent Systems*, 2024, 33(1): 20230068.
- [13] 刘帅帅. 基于模式分解的结构时变模态参数和非线性参数辨识方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2022.
- LIU S S. Research on mode decomposition based methods for structure parameter identification of time-varying mode and nonlinearity [D]. Harbin: Harbin Institute of Technology, 2022.
- [14] 石敏, 许莉钧, 彭华东. 可穿戴生理传感器驱动的深度情绪识别模型在心理健康评估中的应用[J]. *西南大学学报(自然科学版)*, 2024, 46(12): 189-201.
- SHI M, XU L J, PENG H D. Application of wearable physiological sensor-driven deep learning emotion recognition model in mental health assessment[J]. *Journal of Southwest University (Natural Science Edition)*, 2024, 46(12): 189-201.
- [15] 成金华, 刘凯雷, 邵诗峰, 等. 铁矿石价格波动对钢铁行业产量的影响研究[J]. *中国矿业*, 2025, 34(2): 212-221.
- CHENG J H, LIU K L, SHAO S F, et al. Research on the impact of iron ore price fluctuations on the steel industry output[J]. *China Mining Magazine*, 2025, 34(2): 212-221.
- [16] 姜赛平, 刘天科. 基于矿产资源管理视角的政府与市场关系研究[J]. *中国矿业*, 2024, 33(1): 43-51.
- JIANG S P, LIU T K. Research on the relationship between government and market based on the perspective of mineral resources management[J]. *China Mining Magazine*, 2024, 33(1): 43-51.
- [17] 张登耀. 基于 BPNN 及优化方法的期货价格预测效果的比较[J]. *统计与决策*, 2024, 40(23): 161-166.
- ZHANG D Y. Comparison of futures price forecasting effects based on BPNN and optimization methods[J]. *Statistics & Decision*, 2024, 40(23): 161-166.
- [18] LI D Y, MOGHADDAM M R, MONJEZI M, et al. Development of a group method of data handling technique to forecast iron ore price[J]. *Applied Sciences*, 2020, 10(7): 2364.
- [19] EWEES A A, ELAZIZ M A, ALAMEER Z, et al. Improving multilayer perceptron neural network using chaotic grasshopper optimization algorithm to forecast iron ore price volatility[J]. *Resources Policy*, 2020, 65: 101555.
- [20] 刘超, 吕三和, 周圣川, 等. 多源三维地质模型集成方法研究与应用[J]. *地理空间信息*, 2025, 23(3): 77-80.
- LIU C, LYU S H, ZHOU S C, et al. Research and application of multi-source 3D geological model integration method[J]. *Geospatial Information*, 2025, 23(3): 77-80.
- [21] 孙景云, 邴贵英. 一种融合多源数据信息的沪铜期货价格预测新方法[J/OL]. *运筹与管理*, 2024: 1-8. (2024-06-29). <https://kns.cnki.net/KCMS/detail/detail.aspx?filename=YCGL20240627002&dbname=CJFD&dbcode=CJFQ>.
- SUN J Y, BING G Y. A new prediction method for Shanghai copper futures price integrating multi-source data information[J/OL]. *Operations Research and Management Science*, 2024: 1-8. (2024-06-29). <https://kns.cnki.net/KCMS/detail/detail.aspx?filename=YCGL20240627002&dbname=CJFD&dbcode=CJFQ>.
- [22] 王绍琛, 王超, 李士栋, 等. 基于自适应集成学习的煤矿微震时序预测模型[J/OL]. *采矿与岩层控制工程学报*, 2025: 1-13. (2025-03-19). <https://link.cnki.net/doi/10.13532/j.jmsce.cn10-1638/td.2024-1364>.
- WANG S C, WANG C, LI S D, et al. Time series prediction model of microseism in coal mine based on adaptive ensemble learning[J/OL]. *Journal of Mining and Strata Control Engineering*, 2025: 1-13. (2025-03-19). <https://link.cnki.net/doi/10.13532/j.jmsce.cn10-1638/td.2024-1364>.
- [23] 贺安阳. 基于神经网络集成的离心压缩机迁移建模方法[J]. *设备管理与维修*, 2025(5): 56-59.
- HE A Y. Modeling method of centrifugal compressor migration based on neural network integration[J]. *Plant Maintenance Engineering*, 2025(5): 56-59.
- [24] 严利鑫, 高雅婷, 郭军华, 等. 基于改进深度极限学习机的电动汽车行驶里程预测[J/OL]. *吉林大学学报(工学版)*, 2024: 1-12. (2024-11-28). <https://link.cnki.net/doi/10.13229/j.cnki.jdxbgxb.20240985>.
- YAN L X, GAO Y T, GUO J H, et al. Electric vehicle driving range prediction based on improved deep extreme learning machine [J/OL]. *Journal of Jilin University (Engineering and Technology Edition)*, 2024: 1-12. (2024-11-28). <https://link.cnki.net/doi/10.13229/j.cnki.jdxbgxb.20240985>.
- [25] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [26] LIU L Y, JIANG H M, HE P C, et al. On the variance of the adaptive learning rate and beyond[EB/OL]. (2020-02-14)[2023-08-10]. <https://openreview.net/pdf?id=rkgz2aEKDr>.

- [27] 张海霞,李瑞,王霞,等. 基于 VMD 和 GA-BiLSTM 组合模型的河流水质预测[J/OL]. 环境工程技术学报, 2025:1-14. (2025-03-24). <https://kns.cnki.net/KCMS/detail/detail.aspx?filename=HKWZ20250321003&dbname=CJFD&dbcode=CJFQ>.
ZHANG H X, LI R, WANG X, et al. River water quality prediction based on the combined model of VMD and GA-BiLSTM[J/OL]. Journal of Environmental Engineering Technology, 2025:1-14. (2025-03-24). <https://kns.cnki.net/KCMS/detail/detail.aspx?filename=HKWZ20250321003&dbname=CJFD&dbcode=CJFQ>.
- [28] 张礼艳,刘增力,彭艺. 基于改进小波阈值和优化 VMD 算法的语音增强方法[J]. 吉林大学学报(理学版), 2025, 63(2):608-621.
ZHANG L Y, LIU Z L, PENG Y. Speech enhancement method based on improved wavelet threshold and optimized VMD algorithm [J]. Journal of Jilin University (Science Edition), 2025, 63(2):608-621.

Operations Research and Cybernetics

An Enhanced Ensemble Learning Model for Iron Ore Price Prediction

LIU Shiqiang, PAN Weixu

(School of Economics and Management, Fuzhou University, Fuzhou 350108, China)

Abstract: With the acceleration of economic construction and the sustained growth in demand for raw materials driven by infrastructure development, raw material production enterprises are facing increasingly severe challenges. To maintain stable operations and enhance competitiveness, there is an urgent need for these enterprises to explore effective methods for predicting commodity prices of raw materials. As one of the essential raw materials for infrastructure construction, iron ore exhibits complex and volatile characteristics in its futures price fluctuations, posing significant challenges to price forecasting. Although various prediction methods currently exist, their accuracy and stability still require improvement. Therefore, it aims to explore a more accurate and effective method for forecasting iron ore prices. It proposes an enhanced ensemble learning model that integrates a multi-head convolutional stacked long short-term memory (LSTM) encoder-decoder network with modal signal decomposition and attention mechanisms, based on multivariate time series of iron ore prices. The proposed method first employs modal signal decomposition to reduce the volatility of the price series. Subsequently, a multi-head convolutional network automatically extracts complex spatial features from the multivariate sequence data, while a stacked LSTM (SLSTM) network captures temporal dependencies from historical sequences. Finally, an attention mechanism is applied to identify key features and generate the final prediction results. The experimental results demonstrate that the proposed enhanced ensemble model outperforms other benchmark models without decomposition techniques in terms of both prediction accuracy and stability. The findings possess significant practical implications, enabling mining enterprises to achieve more accurate iron ore price predictions, thereby facilitating sustained high-quality development and operational stability.

Keywords: iron ore price prediction; enhanced ensemble learning model; modal decomposition technique; encoder-decoder network; multi-step time series

(责任编辑 黄 颖)