多模态数据驱动的煤炭价格智能化预测方法与应用

汪　莹[1],张若晗[1],朱红达[2],杨　洋[1],祖子帅[1]
(1. 中国矿业大学(北京)管理学院,北京　100083;2. 华电煤业集团数智技术有限公司,北京　100088)

摘要: 煤炭作为我国能源结构中的核心组成部分,其价格波动不仅影响上下游产业链发展,也对实现“双碳”目标和维护国家经济稳定具有重要意义. 为提高煤炭价格预测的准确性,本文构建了基于Informer的多模态交叉注意力预测模型,选取环渤海5 500 K动力煤价格为预测对象,综合考虑原煤产量、煤炭进口量与关税、地缘政治不确定性因素、煤炭企业库存水平、火力发电总量及居民消费者价格指数等影响因素,构建多模态数据集. 数值模态通过线性映射进行处理;文本模态采用RoBERTa编码器提取语义表示;跨模态部分引入交叉注意力机制,实现数值序列与文本信息的特征融合. 结果表明:Informer在长时间序列预测中整体优于多种基准模型;相比简单融合,引入跨模态交叉注意力显著提升了预测准确性;模型在噪声文本条件下通过了稳健性检验. 本文所构建的基于Informer的多模态交叉注意力预测模型能够为煤炭价格的科学预测提供支撑,对能源市场研究及相关政策制定具有一定的理论意义和实践价值.

关键词: 多模态数据;煤炭价格;Informer;交叉注意力
中图分类号: TP181;F426.21　　**文献标志码:** A　　**文章编号:** 1000-1964(2026)02-0546-13

Intelligent coal price prediction method and application driven by multimodal data

WANG Ying[1], ZHANG Ruohan[1], ZHU Hongda[2], YANG Yang[1], ZU Zishuai[1]
(1. School of Management, China University of Mining and Technology-Beijing, Beijing 100083, China;
2. Huadian Coal Industry Group Digital Intelligence Technology Co., Ltd., Beijing 100088, China)

Abstract: Coal is a cornerstone of China's energy structure, and fluctuations in its price not only affect the upstream and downstream of the industrial chain but also bear on achieving the “dual-carbon” goals and safeguarding macroeconomic stability. To improve forecasting accuracy, this study develops an Informer-based multimodal cross-attention forecasting model, using the Bohai-Rim 5 500 K thermal coal price as the prediction target. We build a multimodal dataset that integrates key drivers including raw coal output, coal imports and tariffs, geopolitical uncertainty, coal enterprise inventories, total thermal power generation, and the consumer price index. The numerical modality is processed via linear mapping; the textual modality employs a RoBERTa encoder to extract semantic representations; and a cross-attention mechanism is introduced for cross-modal fusion between numerical sequences and textual information. Empirical results show that Informer

收稿日期: 2025-07-29　　**修回日期:** 2025-12-16　　**DOI:** 10.13247/j.cnki.jcumt.20250295
基金项目: 国家社会科学基金项目(22BGL110)
第一作者: 汪　莹(1973—),女,江苏省徐州市人,教授,博士生导师.E-mail:2039819734@qq.com
通信作者: 祖子帅(1997—),男,山东省菏泽市人,硕士研究生. E-mail:zuzishuai@163.com
引用格式: 汪　莹,张若晗,朱红达,等.多模态数据驱动的煤炭价格智能化预测方法与应用[J].中国矿业大学学报,2026,55(2):546-558.
WANG Ying, ZHANG Ruohan, ZHU Hongda, et al. Intelligent coal price prediction method and application driven by multimodal data[J]. Journal of China University of Mining & Technology, 2026, 55(2): 546-558.

performs better than multiple benchmark models on long-horizon forecasting; relative to simple concatenation, introducing cross-modal cross-attention significantly improves predictive accuracy; the model also passes robustness tests under noisy text conditions. The proposed Informer-based multimodal cross-attention model provides methodological support for scientific coal price forecasting and offers theoretical and practical value for energy-market research and related policy formulation.

Key words: multimodal data; coal price; Informer; cross-attention

煤炭是我国能源结构中重要组成部分，其价格的变化对上下游产业的发展与“双碳”目标下国民经济的稳定增长有重大意义[1]. 煤炭消费在能源消费中占主导地位，2024年我国能源消费总量59.6亿t标准煤，煤炭消费量占能源消费总量比重为53.2%，煤炭消费量较2023年增长1.7%. 可以预见煤炭在我国能源结构中在短期内仍将保持主导地位[2]. 作为能源领域重要经济指标，煤炭价格不但实时反映煤炭行业发展状况，也和我国经济发展与社会稳定密切相关[3-6]. 因此，精确预测煤炭价格对保证我国能源安全和经济健康发展具有重要意义.

本文拟基于多模态数据特征融合以及深度学习组合模型，采用环渤海5 500 K动力煤指数作为煤炭价格参考值，对煤炭价格进行预测. 国内外学者在该研究领域已经取得了一定的研究成果，主要包括煤价影响因素识别和预测方法2个方面. 煤炭价格影响因素方面，文献[7]选取港口存量、销量、期货价格、生产者价格指数以及原油价格等作为煤炭价格影响，构建了向量自回归模型对煤炭价格进行预测. 文献[8]基于原煤产量、动力煤消费量、印尼动力煤指导价、火力发电、钢铁产量、煤炭企业库存、电力行业煤炭消费量、钢铁行业煤炭消费量、化工行业煤炭消费量、建材煤炭消费量工业和其他行业的煤炭消费等影响因素建立了煤炭价格预测模型. 文献[9]根据秦皇岛港煤炭库存、内蒙古东胜5 200 K原煤坑口价、美国煤炭期货价格、煤化工指数、中国大宗商品综合指数等煤炭价格影响因素对我国煤炭价格进行了短期预测.

在预测方法方面早期研究的方法多是聚焦于利用价格时间序列等单一特征维度的结构化数据，考虑多种影响因素的计量经济模型进行煤炭价格的预测[10-13]. 文献[14]通过差分整合移动平均自回归(ARIMA)模型对中国煤炭消费、价格和投资做了长期的预测，揭示了煤炭市场在未来一段时间内的趋势走向. 文献[15]基于偏最小二乘回归(PLSR)方法对煤炭价格指数进行建模预测，提出了一种新的PLSR预测模型，并验证了其在煤炭价格走势预测中的有效性. 文献[16]构建了模糊集成向量自回归与误差修正模型(Fuzzy-IVECM)，分析国际能源市场冲击对中国煤炭价格的影响机制，揭示了外部能源市场波动对中国煤价传导效应的复杂性与动态性. 尽管传统计量方法在描述和构建煤炭价格模型方面具备一定优势，但其难以捕捉宏观与微观经济背景下多种影响因素之间的耦合机制，通常只能基于已有历史数据进行拟合分析.

随着预测算法的发展，深度学习与人工智能方法能够自动提取复杂的非线性关系，相较于传统统计模型，在处理高维多因素驱动的数据时更具灵活性，逐渐成为预测研究中的重要工具[17-24]. 循环神经网络(RNN)由于其能够将信息进行“记忆”，在价格趋势预测领域被广泛应用[25]. 然而在RNN的反向传播过程中，梯度会随着时间步的推移而逐渐消失或爆炸. 为解决这类问题，文献[26]在RNN的基础之上创新地提出了LSTM神经网络的概念. 通过引入门控机制，有效地应对了RNN中的梯度消失与梯度爆炸的问题. 文献[27]提出了一种混合LSTM-DNN模型，首次将深度学习方法应用于煤炭价格多步预测中. 该模型融合LSTM对非线性关系的建模能力与DNN对高维特征表达的能力，实现对未来3,6,9,12个月煤价的精确预测，并验证了多种宏观经济变量和金属价格对煤价的影响. 文献[28]构建了WT-SAE-LSTM模型用于美国电价预测，通过交叉试验优化小波分解参数，验证了该模型在非线性时间序列预测中的有效性. Transformer是一种基于自注意力机制的深度学习模型，相较于RNN和LSTM，具有长距离依赖捕捉和并行计算的能力，表现出更高的效率和性能. 文献[29]将Transformer引入上海原油期货的开、高、低、收(OHLC)价格预测中，并结合结构化损失函数设计，实现了对多粒度价格序列的高精度预测. 针对多模态数

据,文献[30]构建了基于多层Transformer的交叉注意力网络,通过财经新闻、历史市场数据和技术指标之间的成对处理,预测股票市场价格的涨跌方向.由Transformer改进的Informer模型进一步凭借稀疏注意力机制降低了计算复杂度,通过层间蒸馏大幅减少计算量,从而提升了预测的效率和准确性[31].该模型在风电功率、铁路客运量、航班延误等研究领域已有较为广泛的应用[32-35].

综上所述,现有的深度学习方法主要集中在LSTM的组合算法上,缺乏Informer相关模型在煤炭价格预测问题中的应用.在指标选取方面,现有研究多以统计数据为主,忽略了文本数据对煤炭价格的影响.基于此,本文提出了RoBERTa+Informer模型,综合考虑相关统计数据和新闻文本数据,融合自注意力机制和跨模态交叉注意力机制对煤炭价格进行预测,旨在通过多模态数据融合和模型优化提升预测准确性.本研究的创新主要体现在3点:一是构建煤炭价格的多模态预测框架,将时序数据与新闻文本联合建模,全面反映煤炭价格变化规律;二是引入跨模态交叉注意力机制,实现文本与数值的特征融合,显著提升中长步长预测效果;三是采用适用于长序列的Informer架构,有效捕捉煤炭价格的长期依赖关系与周期性,提高了模型的预测准确性.

1 煤炭价格预测影响因素识别

1.1 供给方面影响因素识别

我国煤炭价格作为能源市场重要的经济指标,很大程度上取决于供需关系.2024年,全国动力煤有效供应量为429 915万t,同比增长2.49%.而动力煤总需求量为419 075万t,同比增长3.29%,因此,煤炭的供需格局近年来呈现出供略大于需的趋势.供给侧主要受原煤产量、煤炭进口量与关税、政治不确定性因素、煤炭企业库存水平、煤炭产能规划与供给政策等因素影响.

原煤产量方面,2024年,我国原煤产量延续增长态势,全年规模以上工业产量达47.6亿t,同比增长1.3%,连续第4年创历史新高.

进口量与关税水平方面,2024年,我国煤炭进口量累计达5.4亿t,较2023年同期增加0.68亿t,同比增长14.4%.从进口煤来源国占比来看,我国煤炭进口主要集中于印尼、俄罗斯、蒙古国和澳大利亚四国,我国从这4个国家进口的煤炭数量占我国煤炭进口总量的92.3%.关税也对煤炭价格有一定影响,但是2022—2023年能源危机期间我国对所有国家煤炭实施阶段性零关税,政策稳定性较高.2024年1月1日起,恢复煤炭进口关税.普通税率为20%;特惠税率为0;协定税率基本为0;最惠国税率中,无烟煤、炼焦煤、褐煤为3%,其他煤为6%.按照自贸协定,来自澳大利亚、印尼的进口煤适用的是协定税率,均为0;来自包括俄罗斯、美国、南非、蒙古国等在内的其他国家的进口煤实行最惠国税率,因此关税水平对我国煤炭价格波动影响有限.

地缘政治引起的不确定因素方面,以俄乌冲突为例,这场冲突短期内对我国煤价的影响呈现“先扬后抑”的双重机理.短期煤价因能源替代效应先呈现阶段性上涨,但受俄罗斯增供、国内保供政策及进口结构调整影响,涨幅有限且回落较快,整体呈现“脉冲式上涨后趋稳”的特征.冲突引发全球能源市场动荡,推动全球煤炭作为替代能源的需求激增.带动我国进口煤价同步上涨,进而传导至国内市场煤价短期走高.

煤炭企业库存水平方面,2024年,全年煤炭供需格局持续宽松,煤炭企业库存水平创历史新高,截至年底重点煤炭企业存煤6 800万t、火电厂存煤2.3亿t,可用天数达28 d.高库存抑制市场煤采购需求,尤其在传统旺季(夏季与冬季)因电厂长协资源充足叠加进口补充,煤价“旺季不旺”特征显著.

产能规划与供给政策方面,国家发改委于2024年4月出台的产能储备制度虽未在2024年完全落地,但常规产能释放与进口关税政策调整(协定税率为0)直接扩大了市场供给,导致全年煤价跌幅普遍超过10%.

1.2 需求方面影响因素识别

从需求侧看,动力煤的消耗主要涵盖火力发电、煤化工以及产业链下游(冶金与建材)3个方面.依据中国能源网2025年1月26日发布的数据如图1所示,2024年电力行业消耗动力煤265 106万t,约占63.26%,与2023年相比增长2.47%.化工行业消耗动力煤28 852万t,约占6.88%,与2023年相比增长15.43%.建材行业消耗动力煤27 008万t,约占6.44%,同比下降7.18%.冶金行业方面消耗动力煤17 234万t,约占4.11%,同比下降2.13%.

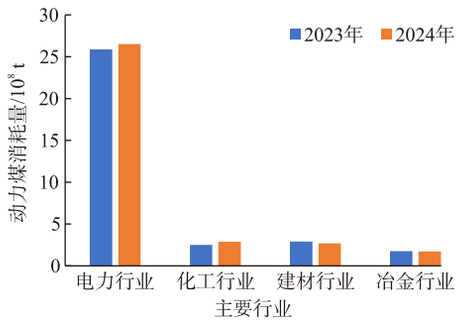


图 1　主要行业动力煤需求量

Fig. 1　Power coal demand in major industries

上述需求侧因素的需求量以及权重结构的变化都会引起需求关系的动态变化，从而影响煤炭价格未来的走势.

1.3　宏观经济方面影响因素识别

居民消费者价格指数(CPI)是衡量通货膨胀水平和居民生活成本变化的关键指标，本文引入CPI旨在反映我国宏观经济的周期性与增长性特征. 从周期性来看，CPI上涨通常表示宏观经济处于扩张阶段，此时企业生产规模扩大，居民收入增加，消费能力提升，这会使得对商品和服务的需求增加，从而导致煤价波动. 从增长性来看，如果CPI开始上涨，通常预示着经济增长动力增强，煤炭需求上涨，从而导致煤价波动.

2　多模态数据来源与预处理

根据前文影响因素的分析与识别，这里选取的数据主要包含统计数据与新闻文本数据，这2种数据具有显著的结构化和非结构化特征，本文将基于这2种数据的预处理结果，进行多模态数据特征融合，为后续的煤炭价格预测奠定数据基础.

2.1　基于关键影响因素的统计数据获取及预处理

本文选取了环渤海5 500 K动力煤指数、原煤产量、煤炭进口量、全国煤炭企业库存、火力发电量和居民消费者价格指数这5个核心数据指标对煤炭价格预测. 其中煤炭价格、原煤产量、煤炭进口量、全国煤炭企业库存来源于中国煤炭市场网，火力发电量和居民消费者价格指数来源于国家统计局. 另外，本文以2017年1月至2025年4月期间的环渤海5 500 K动力煤指数为研究样本，对其走势进行分析预测. 对收集到的煤炭价格与供给、需求、市场经济三大类变量整理与分析，其最大值、最小值、中位数、平均数和标准差相关统计指标详见表1.

表 1　影响因素数据统计信息

Table 1　Statistical information on impact factor data

影响因素	最大值	最小值	中位数	平均数	标准差
环渤海5 500K动力煤指数/(元·t^{-1})	848	526	599	643.49	78.87
原煤产量/万t	43 995.00	28 452.00	33 570.21	34 987.52	4 372.36
煤炭进口量/万t	5 498.20	277.20	2 739.85	2 879.83	1 068.17
全国煤炭企业库存/万t	8 900.00	4 765.00	6 180.00	6 235.40	875.48
火力发电量/(亿kW·h)	6 082.20	3 504.60	4 573.40	4 600.35	595.29
居民消费者价格指数	105.36	99.30	101.62	101.50	1.27

因为煤炭价格的各影响因素时间频度不一致，存在周度和月度2种频度数据，本文使用Python脚本对数据的频度进行转换，统一转换为周度数据，将2017年1月至2025年4月的原煤产量(月)、煤炭进口量(月)、全国煤炭企业库存(月)、火力发电量(月)每月的总量进行等比例均摊到各个周，其计算公式如下：

$$y_u=\sum_{i=1}^{n}\left(\frac{d_i}{D_i}\right)\times Y_i, \tag{1}$$

式中：y_u为第u周的值；n为该周跨越的月份数；d_i为该周在第i月的天数；D_i为第i月的总天数；Y_i为第i月的月度值.

对居民消费者价格指数进行线性插值，转换为周度数据，计算公式如下：

$$y_w=(1-\alpha)\cdot Y_m+\alpha\cdot Y_{m+1}, \tag{2}$$

式中：y_w为第w周的值；α为该周在该月度区间内的位置；Y_m为第m月的值；Y_{m+1}为第$m+1$月的值.

由于各类数据的量纲不一致，为消除各因素的量纲对权重的影响，训练之前对数据进行归一化处理，具体公式如下：

$$x_{ij'}=\frac{x_{ij}-x_{j,\min}}{x_{j,\max}-x_{j,\min}}, \tag{3}$$

式中：$x_{ij'}$为归一化后的第i个样本在第j个指标上的值；x_{ij}为归一化前的原始值；$x_{j,\min}$和$x_{j,\max}$分别为第j个指标的最小值和最大值.

2.2 基于关键影响因素的新闻文本数据获取及预处理

在分析煤炭价格走势时，新闻标题文本中的语义信息与煤炭价格的影响因素密切相关，能够有效辅助预测．根据前文对煤炭价格影响因素的识别，并参考《“十四五”现代能源体系规划》、《2024—2025年中国煤炭市场年度报告》、IEA《Coal 2024/2025》等文件，确定“产能规划、供给政策、行业发展战略、能源结构调整、煤炭进出口价格、动力煤需求、火电领域、火电消耗量占比、供热领域、建材行业、钢铁行业、化工行业、化工煤炭消费量、宏观经济形势、汇率、国际能源价格、俄乌冲突、黑海航运受阻”18个关键词，作为新闻标题文本数据爬取的依据．本文从中国煤炭市场网、中国煤炭经济网、中国能源网、电力网和今日头条等网站获取了2017年1月1日至2025年4月30日期间的行情信息、市场点评等新闻数据共47 193条（见表2）．

为提高文本数据的有效性与可利用性，需对原始数据进行预处理．首先，筛选核心字段，仅保留具有实际分析价值的“新闻标题”和“发布时间”2个属性，剔除与预测关联度较低的其他信息．其次，清洗文本内容，去除无实际意义的字符与图标，并删除常见停用词，以提升语料质量和模型输入的准确性．并通过Python中的Jieba库对语句分词，使用Set函数对分词后词序列进行去重处理．例如对“7月全国发电量增9.6%，火电增12.7%，水电降4.3%”清洗分词后得到的词序列为[“全国”、“发电量”、“增”、“火电”、“水电”、“降”]．分词后共获得26 437个不重复词．最后，按照日期对数据进行聚合处理，将同一周的多条新闻标题合并，便于后续统一输入模型进行分析．

表2 在线新闻数据集示例

Table 2 Example of online news dataset

日期	新闻标题
2017-01-16	1月上旬无烟煤价格上涨，部分动力煤价格持续下跌
2018-11-20	1～10月韩国煤炭进口同比下降0.5%
2019-03-11	2月中国进口煤炭1 764.1万t，同比下降15.6%
2020-01-08	河北今年将压减钢铁产能1 400万t
2021-08-17	7月全国发电量增9.6%，火电增12.7%，水电降4.3%
2022-03-16	俄乌冲突推动国际煤炭价格飙升，各方反应不一
2022-04-80	一季度国家能源集团生产煤炭1.52亿t，同比增长6.7%
2023-05-15	甘肃省印发碳达峰实施方案：推进煤炭消费替代和转型升级
2024-02-27	湖北能源全力应对新一轮冻雨寒潮
2025-04-09	未来中国能源结构呈现“煤降、油气稳、新能源升”的显著特征

3 多模态数据驱动下的组合预测模型

3.1 预测模型与算法逻辑框架

本文的模型架构如图2所示．首先将收集的煤炭价格影响因素统计数据和新闻文本分别进行数值投影和RoBERTa编码处理．随后，将处理后

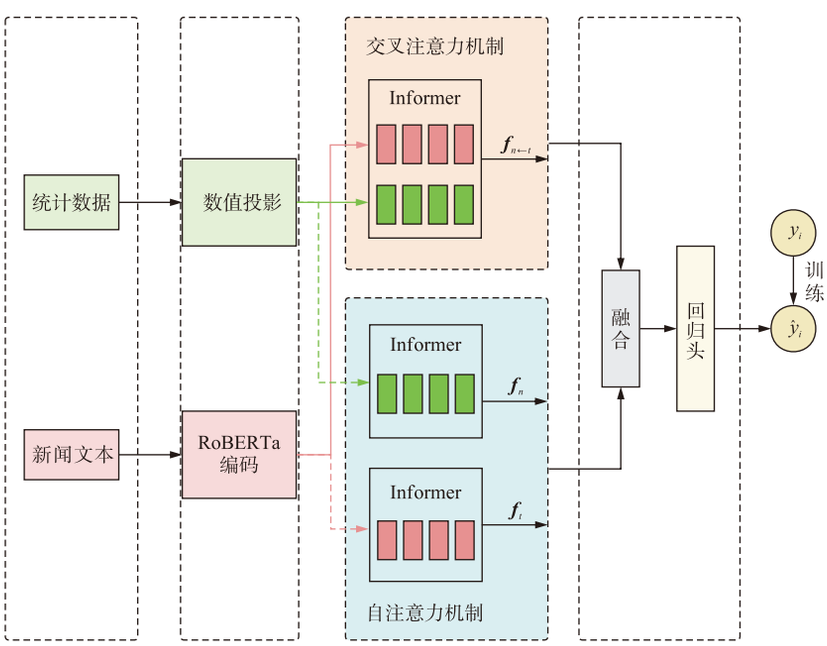


图2 多模态交叉注意力网络架构

Fig. 2 Multi-modal cross attention network architecture

的数据分别输入各自的Informer编码器，利用自注意力机制刻画模态内部规律，并通过交叉注意力机制实现数值与文本之间的信息融合，获得跨模态特征．接着，将数值特征、文本特征与跨模态特征进行融合．最后，融合后的特征输入回归预测头，输出对煤炭价格的预测结果．

3.2　模型输入

本文以煤炭价格及相关因素统计数据与新闻标题为数据源构建训练数据集．用于价格预测的2种模态分别为数值模态和文本模态，针对2种模态，分别构建2条输入分支．通过数值投影处理数值模态数据，对新闻文本采用RoBERTa编码，完成文本嵌入．将处理后的数值数据与句向量输入自注意力与交叉注意力模块，以同时捕捉模态内与模态间关系．

3.2.1　数值模态数据处理

将原始数值进行线性映射，便于与文本模态对齐和后续注意力模块输入，其表达式为

$$\boldsymbol{r}_t=\boldsymbol{W}_z\boldsymbol{z}_t+\boldsymbol{b}_z, \tag{4}$$

式中：$\boldsymbol{z}_t\in\boldsymbol{R}^{d_z}$为第$t$周的数值特征，即煤炭价格及其影响因素；$\boldsymbol{R}$为实数集合；$d_z$为数值模态输入向量的维度；$\boldsymbol{W}_z\in\boldsymbol{R}^{d\times d_z}$，$\boldsymbol{b}_z\in\boldsymbol{R}^d$为可学习参数；$d$为线性映射后的维度；$\boldsymbol{r}_t\in\boldsymbol{R}^{d_z}$为投影后的周向量．

在各个时间窗口下加入位置信息，本文采用正余弦位置编码对时间序列中的每个位置进行编码，其表达式如式(5)～(8)所示．

$$pe_t(t,i)=\begin{cases}\sin\left(\dfrac{t}{10\,000^{\frac{i}{d}}}\right), & i \bmod 2=0,\\ \cos\left(\dfrac{t}{10\,000^{\frac{i-1}{d}}}\right), & i \bmod 2=1,\end{cases} \tag{5}$$

式中：$pe_t(t,i)$表示第t周数据的位置编码，为时间序列中的位置索引值；$i\in[0,d)$为位置编码向量中的索引值；$i \bmod 2$表示i对2的取余结果，用于区分位置编码向量中的偶数维与奇数维．

$$\boldsymbol{R}_t=[\boldsymbol{r}_{t-L+1},\cdots,\boldsymbol{r}_t]\in\boldsymbol{R}^{L\times d}, \tag{6}$$

$$\boldsymbol{P}_t=[pe_{t-l+1},\ldots,pe_t]\in\boldsymbol{R}^{L\times d}, \tag{7}$$

$$\boldsymbol{X}^{(0)}=\boldsymbol{R}_t+\boldsymbol{P}_t, \tag{8}$$

式中：$\boldsymbol{R}_t$为窗口内的数值内容表示序列；L为时间窗口；$\boldsymbol{P}_t$为窗口内的时间位置表示序列；$\boldsymbol{X}^{(0)}$为融合内容与位置的数值序列，后续作为编码器的输入．

3.2.2　文本模态数据处理

使用RoBERTa编码器对文本数据编码，得到文本数据的表示．

$$\boldsymbol{e}_{t,j}=\mathrm{Pool}\left\{f_\theta\left[\mathrm{tokens}(s_{t,j})\right]\right\},\quad j=1,\cdots,108, \tag{9}$$

$$\tilde{\boldsymbol{e}}_{t,j}=\boldsymbol{W}_{\mathrm{text}}\boldsymbol{e}_{t,j}+\boldsymbol{b}_{\mathrm{text}}, \tag{10}$$

式中：$\boldsymbol{e}_{t,j}$为第t周第j条文本经过RoBERTa编码并经池化得到的原始文本句向量；$\tilde{\boldsymbol{e}}_{t,j}$为对$\boldsymbol{e}_{t,j}$进行线性投影后的文本向量；$s_{t,j}$为第$t$周第$j$条标题；$f_\theta$为RoBERTa编码器；Pool为句向量汇聚表示；$\boldsymbol{W}_{\mathrm{text}}\in\boldsymbol{R}^{d\times d_{\mathrm{text}}}$，$\boldsymbol{b}_{\mathrm{text}}\in\boldsymbol{R}^d$将RoBERTa输出与数值分支对齐到同一维度$d$．

对同一周内的多条标题做加权，得到周级文本表示，即

$$\alpha_{t,j}=\frac{\exp(\boldsymbol{q}^{\top}\boldsymbol{W}_a\tilde{\boldsymbol{e}}_{t,j})}{\sum_{k=1}^{5}\exp(\boldsymbol{q}^{\top}\boldsymbol{W}_a\tilde{\boldsymbol{e}}_{t,k})},\qquad j=1,\cdots,5, \tag{11}$$

$$\bar{\boldsymbol{e}}_t=\sum_{j=1}^{5}\alpha_{t,j}\tilde{\boldsymbol{e}}_{t,j}, \tag{12}$$

式中：$\alpha_{t,j}$为该周第j条新闻的重要性权重；$\boldsymbol{q}\in\boldsymbol{R}^d$，$\boldsymbol{W}_a\in\boldsymbol{R}^{d\times d}$为周内打分参数；$\bar{\boldsymbol{e}}_t\in\boldsymbol{R}^d$为$t$周汇聚后的文本向量．

对周文本向量在各个时间窗口下加入位置信息：

$$\boldsymbol{E}_t=[\bar{\boldsymbol{e}}_{t-L+1},\cdots,\bar{\boldsymbol{e}}_t]+\boldsymbol{P}_t\in\boldsymbol{R}^{L\times d}, \tag{13}$$

式中：$\boldsymbol{E}_t$即为文本时间序列；$\boldsymbol{P}_t$的位置编码方式与数值模态数据的编码方式相同，见式(5)．这里得到的$\boldsymbol{E}_t$与数值时间序列在长度和维度上对齐，便于后续做跨模态注意力．

3.2.3　各模态编码

2条分支分别送入各自的编码器，得到模态内时序表示，如下式所示．

$$\boldsymbol{X}=\mathrm{Encoder}_{\mathrm{num}}(\boldsymbol{X}^{(0)})\in\boldsymbol{R}^{L\times d}, \tag{14}$$

$$\boldsymbol{U}=\mathrm{Encoder}_{\mathrm{text}}(\boldsymbol{E}_t)\in\boldsymbol{R}^{L\times d}, \tag{15}$$

式中：$\boldsymbol{X}$与$\boldsymbol{U}$为后续自注意力和“数值与文本交叉注意力”的输入；$\mathrm{Encoder}_{\mathrm{num}}$，$\mathrm{Encoder}_{\mathrm{text}}$分别为数值模态编码器和文本模态编码器．

3.3　Informer模型

Informer是一种基于Transformer的长序列预测改进模型，主要对Transformer进行了3点优化[36]．标准自注意力在长度为N的序列上计算复杂度为$O(N^2)$，Informer通过概率稀疏自注意力把复杂度降到近似$O(N\log N)$．针对深层堆叠会造成显存冗余，Informer采用注意力蒸馏在时间维做下采样以压缩特征．此外，Informer还使用生成式的一步解码器来提升多步预测的推理效率．模型总体结构见图3．

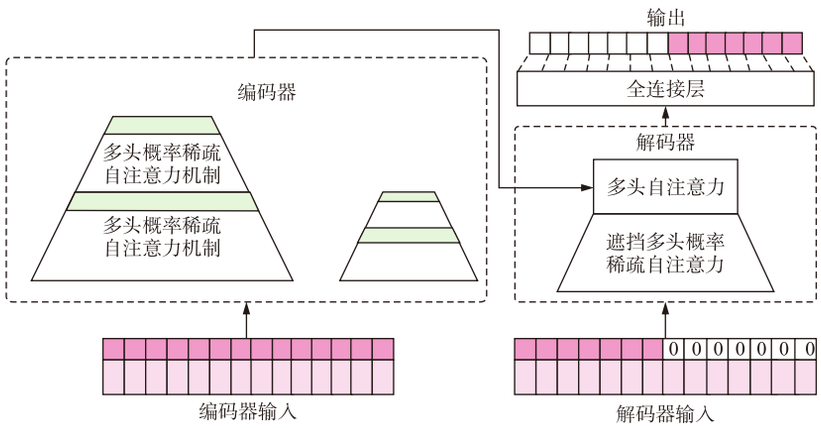


图3 Informer模型结构

Fig. 3 Informer model architecture

Transformer中提出的自注意力机制是通过使用查询$\boldsymbol{Q}$(Query)、键$\boldsymbol{K}$(Key)和值V(Value)来计算序列中各个元素之间的相关性,公式为

$$A(\boldsymbol{Q},\boldsymbol{K},V)=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\top}}{\sqrt{d}}\right)V, \tag{16}$$

式中:$\boldsymbol{Q}\in\boldsymbol{R}^{L_Q\times d}$,$\boldsymbol{K}\in\boldsymbol{R}^{L_K\times d}$,$V\in\boldsymbol{R}^{L_K\times d}$;$L_Q$,$L_K$,$L_V$分别为查询、键、值的矩阵维度;Softmax为激活函数.

概率稀疏自注意力机制是以概率稀疏化自注意力矩阵,将时间复杂度降低至对数的线性复杂度,其第i个$\boldsymbol{Q}$的注意力系数通过概率p的形式定义为

$$A(\boldsymbol{q}_i,\boldsymbol{K},V)=\sum_j\frac{\kappa(\boldsymbol{q}_i,\boldsymbol{k}_j)}{\sum_l\kappa(\boldsymbol{q}_i,\boldsymbol{k}_l)}\boldsymbol{v}_j=E_{p(\boldsymbol{k}_j|\boldsymbol{q}_i)}\left[\boldsymbol{v}_j\right], \tag{17}$$

式中:令$p(\boldsymbol{q}_i,\boldsymbol{k}_j)=\dfrac{\kappa(\boldsymbol{q}_i,\boldsymbol{k}_j)}{\sum_l\kappa(\boldsymbol{q}_i,\boldsymbol{k}_l)}$;$\kappa$为相似度函数;$E$为期望;$p$为条件概率;$\boldsymbol{q}_i$为矩阵$\boldsymbol{Q}$第$i$行;$\boldsymbol{k}_j$,$\boldsymbol{v}_j$分别为矩阵$\boldsymbol{K}$和矩阵$V$的第$j$行.

为了量化查询相邻的稀缺性,使用了KL(Kull-back-Leibler)散度计算$\boldsymbol{Q}$向量的概率分布与均匀分布的熵,其中第i个稀疏性度量的定义式为

$$M(\boldsymbol{q}_i,\boldsymbol{K})=\ln\sum_{j=1}^{L_K}\mathrm{e}^{\frac{\boldsymbol{q}_i\boldsymbol{k}_j^{\top}}{\sqrt{d}}}-\frac{1}{L_K}\sum_{j=1}^{L_K}\frac{\boldsymbol{q}_i\boldsymbol{k}_j^{\top}}{\sqrt{d}}. \tag{18}$$

概率稀疏注意力机制可表示为

$$A(\boldsymbol{Q},\boldsymbol{K},V)=\mathrm{Softmax}\left(\frac{\overline{\boldsymbol{Q}}\,\boldsymbol{K}^{\top}}{\sqrt{d}}\right)V, \tag{19}$$

式中:$\overline{\boldsymbol{Q}}$为一个和查询矩阵$\boldsymbol{Q}$形状大小相同的稀疏矩阵,只包含稀疏度量$M(\boldsymbol{q}_i,\boldsymbol{K})$下的前$m$个主要查询向量$\boldsymbol{q}_i$,$m$由一个恒定的采样因子$c$控制,$m=c\ln L_Q$.

针对长序列存在的输入之间的长范围依赖,通过注意力蒸馏机制在内存有限的情况下将单个层级特征时间维度减半,蒸馏机制是采用一维卷积以及最大池化对时间维度进行下采样,从d层到$d+1$层计算公式为

$$X_{d+1}^{t}=\mathrm{MaxPool}\left\{\mathrm{ELU}\left[\mathrm{Convld}\left(\left[X_d^t\right]_{AB}\right)\right]\right\}, \tag{20}$$

式中:$\left[X_d^t\right]_{AB}$为多头概率稀疏自注意力机制的关键步骤;Convld为一维卷积操作;ELU为激活函数;MaxPool为最大池化操作.蒸馏操作对每一层数序列长度都逐层减半,减少了计算的内存使用.

Informer采用生成式的一步解码器(Decoder),如图3所示,由遮挡的多头注意力机制和遮挡的多头稀疏注意力机制组成,通过掩码操作遮挡多头注意力,避免了模型的自回归现象.

该Decoder的输入为

$$X_{\mathrm{DE}}^{t}=\mathrm{Concat}\left(X_{\mathrm{token}}^{t},X_0^t\right)\in R^{(L_{\mathrm{token}}+L_y)\times d_{\mathrm{model}}}, \tag{21}$$

式中:X_{DE}^{t}为学习输入的序列;X_{token}^{t}为已知标签序列;X_0^t为结果的占位符;Concat表示序列拼接操作.

3.4 多模态信息融合网络架构设计

3.4.1 自注意力模块

自注意力网络用于建模各个模态内关系,本文数值数据与文本数据各使用一个Informer单元,其注意力计算公式见式(17).

在多头注意力机制下,第i个注意力头为

$$\mathrm{head}_i=A(\boldsymbol{Q}\boldsymbol{W}_i^{Q},\boldsymbol{K}\boldsymbol{W}_i^{K},V\boldsymbol{W}_i^{V}), \tag{22}$$

式中:$\boldsymbol{W}_i^{Q}$,$\boldsymbol{W}_i^{K}$,$\boldsymbol{W}_i^{V}$分别为第i个头的查询投影矩阵、键投影矩阵、值投影矩阵.

随后将所有单头注意力拼接并线性变换得到多头注意力:

$$MT(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Concat}(\mathrm{head}_1,\cdots,\mathrm{head}_C)\boldsymbol{W}^O, \tag{23}$$

式中：C为注意力头个数；$\boldsymbol{W}^O$为投影权重矩阵.

分别对数值与文本模态取得单模态表示如下：

$$\boldsymbol{f}_n=MT(\boldsymbol{Q}_n,\boldsymbol{K}_n,\boldsymbol{V}_n),\ \boldsymbol{f}_t=MT(\boldsymbol{Q}_t,\boldsymbol{K}_t,\boldsymbol{V}_t), \tag{24}$$

式中：$\boldsymbol{Q}_n=\boldsymbol{X}\boldsymbol{W}_n^Q$；$\boldsymbol{K}_n=\boldsymbol{X}\boldsymbol{W}_n^K$；$\boldsymbol{V}_n=\boldsymbol{X}\boldsymbol{W}_n^V$；$\boldsymbol{Q}_t=\boldsymbol{U}\boldsymbol{W}_t^Q$；$\boldsymbol{K}_t=\boldsymbol{U}\boldsymbol{W}_t^K$；$\boldsymbol{V}_t=\boldsymbol{U}\boldsymbol{W}_t^V$. 2个自注意力模块输出的$\boldsymbol{f}_n$,$\boldsymbol{f}_t$将用于后续的价格预测.

3.4.2　交叉注意力模块

为刻画不同模态之间的相互作用，本文构建了跨模态交叉注意力机制[37]. 考虑到数值序列能够直接反映煤炭价格的变化状态与结构性约束，在时序预测中承担对目标变量演化的主要解释作用，本文以数值模态作为查询($\boldsymbol{Q}$)，以文本模态作为键($\boldsymbol{K}$)与值($\boldsymbol{V}$)，在时间对齐的基础上，动态检索、按需提取与当前数值状态相关的文本语义证据，将外部信息注入统一的状态表征中. 该设计使注意力权重以数值状态为条件进行自适应分配，实现对关键语义的筛选，并围绕预测目标所依赖的主要信息进行跨模态匹配[38-39]. 具体公式为

$$A_{n\leftarrow t}(\boldsymbol{Q}_n,\boldsymbol{K}_t,\boldsymbol{V}_t)=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}_n\boldsymbol{K}_t^{\top}}{\sqrt{d}}\right)\boldsymbol{V}_t. \tag{25}$$

第i个头与多头交叉注意力分别为

$$\mathrm{head}_i^{(n\leftarrow t)}=A_{n-t}(\boldsymbol{Q}_n\boldsymbol{W}_i^{Q_n},\boldsymbol{K}_t\boldsymbol{W}_i^{K_t},\boldsymbol{V}_t\boldsymbol{W}_i^{V_t}). \tag{26}$$

$$MT_{n\leftarrow t}=\mathrm{Concat}(\mathrm{head}_1^{(n\leftarrow t)},\cdots,\mathrm{head}_C^{(n\leftarrow t)})\boldsymbol{W}_{n\leftarrow t}^O. \tag{27}$$

最后，得到跨模态交互特征：

$$\boldsymbol{f}_{n\leftarrow t}=MT_{n\leftarrow t}. \tag{28}$$

3.4.3　模态融合

在融合层中，将自注意力与交叉注意力模块的特征向量进行融合，形成一个融合特征向量. 令3类特征集合，即$\boldsymbol{S}=\{\boldsymbol{f}_n,\boldsymbol{f}_t,\boldsymbol{f}_{n\leftarrow t}\}$，设可学习查询向量$\boldsymbol{g}\in\boldsymbol{R}^d$，融合权重与融合特征向量为

$$\alpha_o=\frac{\exp(\boldsymbol{g}^{\top}\boldsymbol{W}_f\boldsymbol{f}_o)}{\sum\limits_{u\in M}\exp(\boldsymbol{g}^{\top}\boldsymbol{W}_f\boldsymbol{f}_u)}, \tag{29}$$

式中：$\alpha_o\in(0,1)$为融合权重，满足$\sum\limits_{o\in M}\alpha_o=1$；$o,u\in M$为索引标签，$M=\{n,t,n\leftarrow t\}$为索引集合；$\boldsymbol{W}_f\in\boldsymbol{R}^{d\times d}$为线性投影矩阵；$\boldsymbol{f}_o$为$M$中某个特定模态的特征向量；$\boldsymbol{f}_u$为$M$中遍历所有模态的特征向量.

$$\boldsymbol{f}_{\mathrm{merged}}=\sum_{o\in M}\alpha_o\boldsymbol{f}_o, \tag{30}$$

式中　$\boldsymbol{f}_{\mathrm{merged}}\in\boldsymbol{R}^d$为融合特征向量.

3.4.4　结果预测

在融合层得到的融合表征$\boldsymbol{f}_{\mathrm{merged}}$基础上构建回归头，对价格进行预测，回归头一次性输出未来H周的价格向量.

$$\hat{\boldsymbol{y}}_{t+1:t+H}=\boldsymbol{W}_p\boldsymbol{f}_{\mathrm{merged}}+\boldsymbol{b}_p, \tag{31}$$

式中：$\hat{\boldsymbol{y}}_{t+1:t+H}\in\boldsymbol{R}^H$为未来1～$H$周的价格预测结果；$\boldsymbol{W}_p\in\boldsymbol{R}^{H\times d}$为预测权重矩阵；$\boldsymbol{b}_p\in\boldsymbol{R}^H$为预测偏置向量.

$$L_{\mathrm{MSE}}=\frac{1}{BH}\sum_{b=1}^{B}\sum_{h=1}^{H}(\hat{y}_{t+h}^{(b)}-y_{t+h}^{(b)})^2, \tag{32}$$

式中：L_{MSE}为损失函数；B为批量大小；H为需要连续预测的步长(周数).

4　试验结果分析

本文将处理后的数据按照7∶1∶2的比例划分训练集、验证集和测试集，输入不同模型进行试验. 试验采用的操作系统为Windows11，CPU为IntelCorei7-13700H，运行内存为32 GB，GPU为NVIDIA GeForce RTX 4060Laptop 8 GB，试验的环境配置为Python 3.6.10，Tensorflow 2.6.2，Keras 2.6.0，pandas 1.1.5.

4.1　预测评价指标

为了清晰直观地对不同模型的预测效能进行比较，本文选取了平均绝对误差(MAE，简写为E_{MA})和均方根误差($RMSE$，简写为E_{RMS})2种常用评价指标，对各模型的预测结果进行综合评估. 指标计算公式如下：

$$E_{\mathrm{MA}}=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right|, \tag{33}$$

$$E_{\mathrm{RMS}}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}, \tag{34}$$

式中：y_i为第i条样本的真实观测值；$\hat{y}_i$为模型对该样本的预测值；n为测试集中的样本总数.

4.2　消融试验

本文Informer模型的初始参数如表3所示.

表3　模型参数

Table 3　Model parameters

参数	设定值
向量维度嵌入维度(d_model)	256
注意力头数(heads)	4
批次大小(Batch_size)	32
学习率(Learning rate)	0.001
迭代次数(Epochs)	120

在其他参数保持不变的前提下，对多头注意力机制的head数及Batch_size进行调整，以确定煤炭价格预测模型的最优参数.

4.2.1 多头注意力机制的head数对模型预测效果的影响

多头注意力的head数是一个重要超参数，对模型的性能和效率有着显著的影响. 接下来对不同的heads进行了一系列的试验，并比较了在不同heads条件下预测模型的预测能力(见表4). 根据试验需求，设置了2,4,6,8这4种，固定预测步长为4. 由表4可知，预测误差呈“U形”趋势，head数量为4时*MAE*与*RMSE*均最小，head数量超过4时轻微过拟合且训练变慢，故head数量为4时模型预测效果最好.

表4 Head数对模型预测效果的影响

Table 4 Impact of the number of heads on model prediction performance

heads	*MAE*	*RMSE*
2	3.674 2	5.103 9
4	3.561 8	4.872 6
6	3.623 7	4.981 4
8	3.719 5	5.146 8

4.2.2 Batch_size对模型的影响

Batch_size为每轮训练的样本数. 过大会导致显存不足、收敛变慢，过小会导致频繁更新参数致使稳定性下降. 根据试验需求，设置了不同数量的Batch_size，包括8,16,32,64，固定预测步长为4，模型预测效果的影响见表5. 综合对比，Batch_size为32时预测误差最小. Batch_size为32时虽收敛，但泛化略弱、对调参更敏感. 不同Batch_size间差异较小，均处于可接受范围.

表5 Batch_size对模型预测效果的影响

Table 5 Impact of Batch_size on model prediction performance

Batch_size	*MAE*	*RMSE*
8	3.642 1	4.989 7
16	3.598 4	4.913 5
32	3.561 8	4.872 6
64	3.684 9	5.091 2

4.3 模型对比试验

本文构建了基于Informer的多模态交叉注意力预测模型，通过多次对比试验，对所构建的预测模型的性能进行分析和测试，探究本模型能否显著提升煤炭价格预测能力.

4.3.1 数值模态下各模型对比

在数值模态试验中，在本文数值数据集上横向对比了ARIMAX, RNN, LSTM, Transformer与Informer的预测性能(见表6)，以检验Informer在长时间序列建模中的优势. 结果显示，Informer模型在各个预测步长相比其他模型误差最低，表明Informer模型对长依赖与季节性更为敏感、拟合更稳健. 由图4可知，随着预测步长H增大，各模型误差均小幅上升，但Informer的上升斜率最缓，进一步验证了其在长时序场景下的稳定性.

表6 数值模态下不同模型误差对比

Table 6 Comparison of errors of different models in numerical mode

模型	*H*=1		*H*=2		*H*=4		*H*=8	
	MAE	*RMSE*	*MAE*	*RMSE*	*MAE*	*RMSE*	*MAE*	*RMSE*
ARIMAX	5.873 4	6.683 8	7.089 2	8.149 7	9.704 1	11.233 6	16.743 5	19.213 8
RNN	13.395 7	14.660 1	15.562 3	17.256 9	20.027 4	22.258 2	24.444 6	27.062 0
LSTM	3.810 5	4.549 8	7.335 9	8.298 2	14.616 7	16.671 3	24.272 8	27.355 1
Transformer	1.607 6	2.277 4	2.867 3	3.712 8	5.009 5	6.052 1	8.607 2	10.294 9
Informer	1.415 5	2.238 7	2.403 4	3.581 0	4.248 8	5.810 3	7.303 6	9.489 1

4.3.2 交叉注意力与简单融合对比

为验证交叉注意力机制的有效性，以简单融合(数值矩阵与文本表示矩阵简单拼接)为基准，在每个预测步长上，采用DM检验，基于逐期损失差并

用 Newey-West 校正获得各步长的 p 值(见表 7). 若 $p<0.05$,则判定交叉注意力显著优于简单融合.

表 8 为交叉注意力与简单融合的误差对比. 由表 8 可见,交叉注意力在所有预测步长上的 *MAE*,*RMSE* 均更低,其中在预测步长为 4 时提升效果最显著. DM 检验结果中,在预测步长为 2,4,8 时均在 5% 显著性水平下通过,说明交叉注意力相对简单融合的改进具有统计显著性. 故认为交叉注意力整体优于简单融合,且在中长步长 4,8 上的优势更稳健.

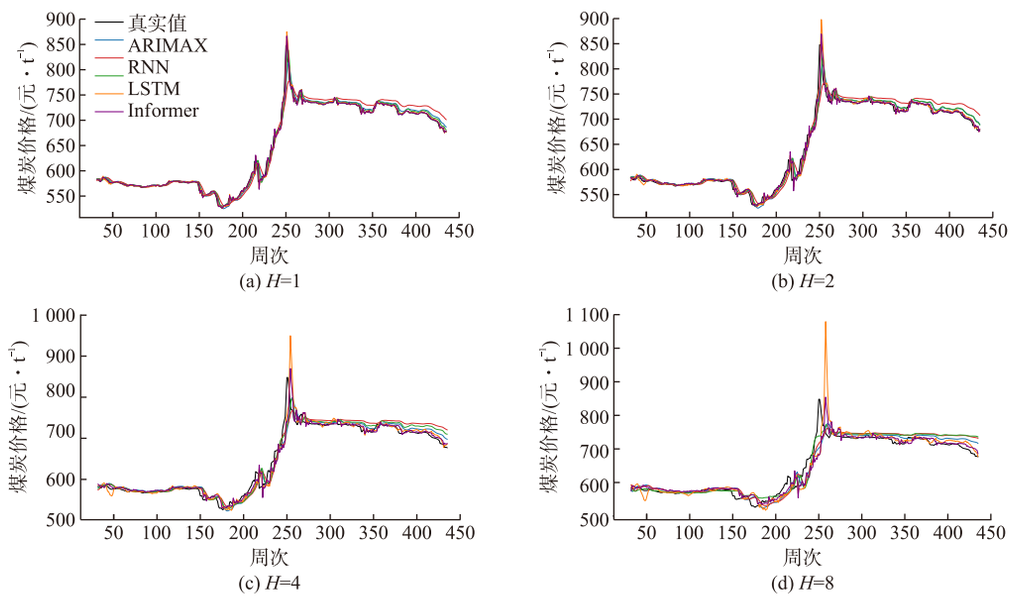


图 4　数值模态下不同模型预测结果

Fig. 4　Prediction results of different models in numerical mode

表 7　DM 检验(交叉注意力与简单融合)

Table 7　DM test (cross-attention and simple fusion)

预测步长	p 值	预测步长	p 值
1	0.061 5	4	0.003 6
2	0.023 1	8	0.011 4

表 8　交叉注意力与简单融合误差对比

Table 8　Comparison of errors between cross-attention and simple fusion

模型	*H*=1		*H*=2		*H*=4		*H*=8	
	MAE	*RMSE*	*MAE*	*RMSE*	*MAE*	*RMSE*	*MAE*	*RMSE*
简单融合	1.303 9	2.061 7	2.214 7	3.298 6	3.914 2	5.352 8	6.724 8	8.737 9
交叉注意力	1.186 7	1.873 9	2.012 4	2.998 7	3.561 8	4.872 6	6.128 4	7.962 3

图 5 为交叉注意力与简单融合预测结果. 相比简单融合,交叉注意力模型在峰值及拐点的预测偏差相对较小,整体误差呈现一定程度的收敛.

4.3.3　文本编码器对比

为评估文本表示对多模态预测的影响,在参数相同的条件下对比 BERT 与 RoBERTa 模型,优化器、学习率、窗口、步长等保持一致(见表 9). 同时构造“将 20% 标题随机替换为无关短句”的噪声数据集,用以检验模型的鲁棒性(见表 10).

由表 10 可知,RoBERTa 在所有预测步长上均优于 BERT,在加入 20% 文本噪声时,RoBERTa 的退化幅度 ΔMAE 为 0.1721、$\Delta RMSE$ 为 0.2784,小于 BERT 的退化幅度. 因此,本文采用 RoBERTa 作为文本编码器,以在多步预测中兼顾准确性与稳健性.

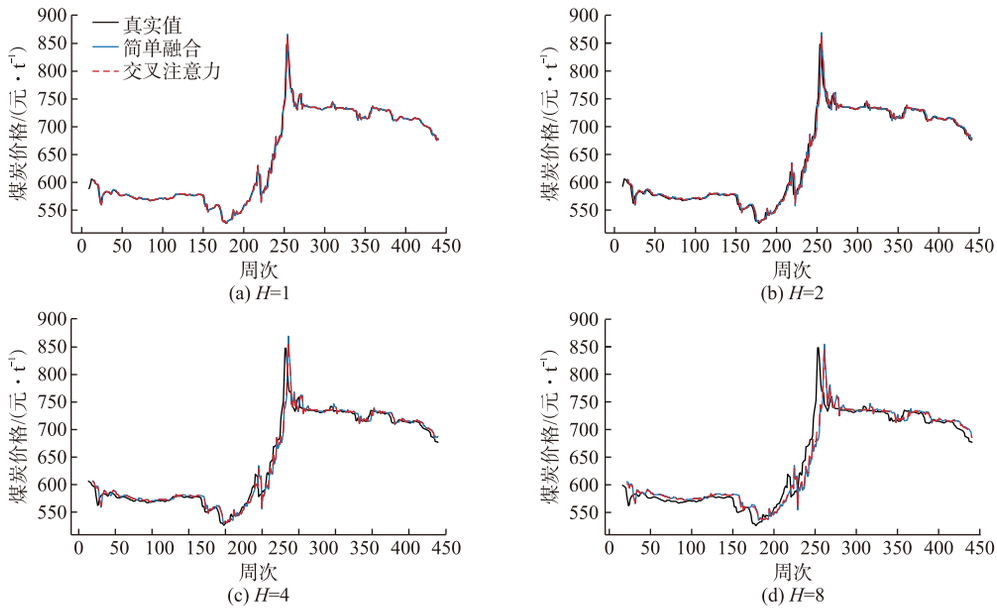


图 5 交叉注意力与简单融合预测结果

Fig. 5 Prediction results of cross-attention and simple fusion

表 9 文本编码器对模型误差影响

Table 9 Impact of text encoder on model error

模型	$H=1$		$H=2$		$H=4$		$H=8$	
	MAE	*RMSE*	*MAE*	*RMSE*	*MAE*	*RMSE*	*MAE*	*RMSE*
BERT	1.269 8	1.942 7	2.086 5	3.128 4	3.674 9	5.111 3	6.236 5	8.128 9
RoBERTa	1.186 7	1.873 9	2.012 4	2.998 7	3.561 8	4.872 6	6.128 4	7.962 3

表 10 模型在噪声条件下的鲁棒性分析

Table 10 Analysis of model robustness under noisy conditions

文本编码器	原始数据集		20% 噪声		ΔMAE	$\Delta RMSE$
	MAE	*RMSE*	*MAE*	*RMSE*		
BERT	3.674 9	5.111 3	3.889 6	5.438 2	0.214 7	0.326 9
RoBERTa	3.561 8	4.872 6	3.733 9	5.151 0	0.172 1	0.278 4

5 结 论

1) 本文从供给、需求与宏观经济 3 个维度分析煤炭价格的影响因素，以 2017—2025 年环渤海 5 500 K 动力煤指数为研究对象，提出一种基于 RoBERTa、Informer 与交叉注意力模块构成的多模态煤炭价格预测模型.

2) 在数值模态下，将 ARIMAX，RNN，LSTM，Transformer 与 Informer 进行对比，结果表明 Informer 在不同预测步长上误差相对较低，证明其更适合捕捉煤炭价格的长期依赖和季节性特征.

3) 在多模态条件下，将交叉注意力与简单融合进行比较并进行 DM 检验，结果显示交叉注意力在中长步长上的预测效果最为突出，能够更充分地利用新闻文本对煤炭价格波动的解释力.

4) 通过对比 BERT 与 RoBERTa，并在加入 20% 噪声的文本数据上测试鲁棒性，发现 RoBERTa 在煤炭市场信息扰动较大的背景下更具稳定性.

5) 本文所构建的煤炭价格预测模型可在供需波动背景下为煤炭企业的采购、库存安排与合同定价提供一定的参考，同时亦可为动力煤现货及衍生品市场参与者的价格研判提供辅助性依据.

参考文献：

[1] 燕志鹏，于泽民，顾新莲.我国碳排放价格与煤炭期货价格的传导机制研究[J].经济问题，2022(6)：67-74.
YAN Zhipeng, YU Zemin, GU Xinlian. Research on conduction mechanism between carbon price and coal future price in China[J]. On Economic Problems, 2022(6): 67-74.

[2] 国家统计局.中华人民共和国2024年国民经济和社会发展统计公报[EB/OL].(2025-02-28)[2025-07-01]. https://www.gov.cn/lianbo/bumen/202502/content_7008605.htm
National Bureau of Statistics of China. Statistical Communiqué of the People's Republic of China on the 2024 National Economic and Social Development [EB/OL]. https://www.gov.cn/lianbo/bumen/202502/content_7008605.htm

[3] 张赟懿，苏　彤，王思泉，等.行稳方能致远：中国煤炭市场运行特征与价格政策评估[J].管理世界，2025, 41(2): 101-119.
ZHANG Yunyi, SU Tong, WANG Siquan, et al. Only by standing up can move forward: Assessing the operation characteristics and pricing policy of China's coal market[J]. Journal of Management World, 2025, 41(2): 101-119.

[4] 杜克锐，李旻旸，王思泉，等.煤炭价格冲击与新能源业务扩张叠加下的企业融资约束[J].经济研究，2024,59(12):4-20.
DU Kerui, LI Minyang, WANG Siquan, et al. The impact of coal price shocks and new energy business expansion on corporate financing constraints[J]. Economic Research Journal, 2024, 59(12): 4-20.

[5] 于　左，史俊荣，罗天羽.竞争企业间部分所有权、去产能政策与中国煤炭价格畸高[J].经济学(季刊)，2024, 24(3): 861-877.
YU Zuo, SHI Junrong, LUO Tianyu. Partial ownership of competitive enterprises, de-capacity policy and distortional overcharge on coal in China[J]. China Economic Quarterly, 2024, 24(3): 861-877.

[6] 刘　畅，王旭冉.中国煤炭价格波动与去产能政策的选择[J].西安交通大学学报(社会科学版)，2020, 40(3): 101-108.
LIU Chang WANG Xuran. Influencing factors of coal price fluctuation in China: Based on SVAR model[J]. Journal of Xi'an Jiaotong University (Social Sciences), 2020, 40(3): 101-108.

[7] GUO X P, SHI J X, REN D F. Coal price forecasting and structural analysis in China[J]. Discrete Dynamics in Nature and Society, 2016, 2016(1): 1256168.

[8] WANG C, XU G, SUN C L, et al. Modeling and forecasting of coal price based on influencing factors and time series[J]. Journal of Cleaner Production, 2024, 467: 143030.

[9] JIN X, SONG A Y, HE M F. Short-term thermal coal price forecasting based on EL-CLSTM model [J]. IEEE Access, 2024, 12: 147364-147371.

[10] DING Z H, FENG C C, LIU Z H, et al. Coal price fluctuation mechanism in China based on system dynamics model[J]. Natural Hazards, 2017, 85(2): 1151-1167.

[11] YANG S L, LIU Q H, DUAN Z P. Study on the price elasticity problem of China′s coal supply and demand[C]//IEEE. 2012 International Conference on Information Management, Innovation Management and Industrial Engineering. Sanya, 2012, 3: 275-277.

[12] ZHU S Q, CHI Y Y, GAO K Y, et al. Analysis of influencing factors of thermal coal price[J]. Energies, 2022, 15(15): 5652.

[13] ZHAO Z Y, ZHU J, XIA B. Multi-fractal fluctuation features of thermal power coal price in China[J]. Energy, 2016, 117: 10-18.

[14] JIANG S M, YANG C, GUO J T, et al. ARIMA forecasting of China's coal consumption, price and investment by 2030[J]. Energy Sources, Part B: Economics, Planning, and Policy, 2018, 13(3): 190-195.

[15] ZHANG B, MA J H. Coal price index forecast by a new partial least-squares regression[J]. Procedia Engineering, 2011, 15: 5025-5029.

[16] SHI R Y, WANG D, ZHAO Y Y. The effect of international energy market shocks on coal price of China based on the fuzzy integrated vector auto regressive and error correction model[J]. Journal of Intelligent & Fuzzy Systems, 2021, 40(4): 8451-8461.

[17] YU L, LAI K K, WANG S Y, et al. Oil price forecasting with an EMD-based multiscale neural network learning paradigm[C]// International Conference on Computational Science. Springer: Berlin, 2007: 925-932.

[18] 刘　硕，何永秀，陶卫君，等.遗传BP神经网络的煤价预测与煤价风险规避策略[J].华北电力大学学报，2009,36(6):75-80.
LIU Shuo, HE Yongxiu, TAO Weijun, et al. Coal price forecasting based on GABP neural network and risk resistance strategy[J]. Journal of North China Electric Power University, 2009, 36(6): 75-80.

[19] KÖLMEK M A, NAVRUZ İ. Forecasting the day-ahead price in electricity balancing and settlement market of Turkey by using artificial neural networks [J]. Turkish Journal of Electrical Engineering and Computer Sciences, 2015, 23(3): 841-852.

[20] 李朋林,梁露露.基于BP神经网络的煤炭价格影响因素及预测研究[J].数学的实践与认识,2015,45(17):113-126.
LI Penglin, LIANG Lulu. Research on the influence factors and prediction of coal price based on BP neural network [J]. Mathematics in Practice and Theory, 2015, 17: 113-126.

[21] GUPTA N, NIGAM S. Crude oil price prediction using artificial neural network [J]. Procedia Computer Science, 2020, 170: 642-647.

[22] 余乐安.基于人工智能的预测与决策优化理论和方法研究[J].管理科学,2022,35(1):60-66.
YU Lean. Forecasting and decision optimization theory and methods based on artificial intelligence [J]. Journal of Management Science, 2022, 35(1): 60-66.

[23] 李星毅,李仲飞,李其谦,等.基于机器学习的资产收益率预测研究综述[J].中国管理科学,2025,33(1):311-322.
LI Xingyi, LI Zhongfei, LI Qiqian, et al. A review of research on asset return prediction based on machine learning [J]. Chinese Journal of Management Science, 2025, 33(1): 311-322.

[24] 毕文杰,扶春娟.基于机器学习的Airbnb房源价格预测及影响因素研究:以北京市为例[J].运筹与管理,2022,31(9):217-224.
BI Wenjie, FU Chunjuan. Price estimation and determinants research of Airbnb with machine learning: Based on data from Beijing[J]. Operations Research and Management Science, 2022, 31(9): 217-224.

[25] MANDAL P, SRIVASTAVA A K, SENJYU T, et al. A new recursive neural network algorithm to forecast electricity price for PJM day-ahead market [J]. International Journal of Energy Research, 2010, 34(6): 507-522.

[26] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[27] ALAMEER Z, FATHALLA A, LI K, et al. Multistep-ahead forecasting of coal prices using a hybrid deep learning model [J]. Resources Policy, 2020, 65: 101588.

[28] QIAO W B, YANG Z. Forecast the electricity price of US using a wavelet transform-based hybrid model [J]. Energy, 2020, 193: 116704.

[29] HUANG W Y, GAO T X, HAO Y, et al. Transformer-based forecasting for intraday trading in the Shanghai crude oil market: Analyzing open-high-low-close prices[J]. Energy Economics, 2023, 127: 107106.

[30] FATALIYEV K, LIU W. MCASP: multi-modal cross attention network for stock market prediction [C]// Proceedings of the 21st Annual Workshop of the Australasian Language Technology Association. 2023: 67-77.

[31] ZHOU H Y, ZHANG S H, PENG J Q, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(12): 11106-11115.

[32] GONG M J, YAN C C, XU W, et al. Short-term wind power forecasting model based on temporal convolutional network and Informer[J]. Energy, 2023, 283: 129171.

[33] 施进炜,张 程,原冬芸.基于数据修正的概率稀疏自注意短期风电功率预测[J].智慧电力,2023,51(10):54-61.
SHI Jinwei, ZHANG Cheng, YUAN Dongyun. Short-term wind power prediction based on data correction with probabilistic sparse self-attention [J]. Smart Power, 2023, 51(10): 54-61.

[34] 秦 进,胡 冉,毛成辉,等.高速铁路日常客运量的EMD-Informer组合预测方法[J].铁道学报,2024,46(2):1-11.
QIN Jin, HU Ran, MAO Chenghui, et al. EMD-Informer method for prediction of daily passenger flow of high-speed railways[J]. Journal of the China Railway Society, 2024, 46(2): 1-11.

[35] 杨新涅,游 超,朱承元.基于Informer模型的航班延误预测[J].科学技术与工程,2025,25(19):8282-8288.
YANG Xinsheng, YOU Chao, ZHU Chengyuan. Flight delay prediction based on Informer modeling [J]. Science Technology and Engineering, 2025, 25(19): 8282-8288.

[36] ZHOU H Y, LI J X, ZHANG S H, et al. Expanding the prediction capacity in long sequence time-series forecasting [J]. Artificial Intelligence, 2023, 318: 103886.

[37] TSAI Y H H, BAI S J, LIANG P P, et al. Multimodal transformer for unaligned multimodal language sequences[C]// Proceedings of the 57th conference of the Association for Computational Linguistics. Florence, 2019: 6558.

[38] TAN H, BANSAL M. Lxmert: Learning cross-modality encoder representations from transformers [C]// 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hongkong, 2019: 5100-5111.

[39] LU J S, BATAR D, PARIKH D, et al. Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks[C]//Advances in Neural Information Processing Systems32 (NIPS 2019. Vancouver, 2019: 32.

(责任编辑 沈 菊 仝 腾)