◆产业赋能与业态融合◆

基于深度学习算法的萤石氟化钙含量预测研究

李凌云[1]，刘海庆[1]，郝　彬[1]，任建吉[2]

（1. 多氟多新材料股份有限公司，河南 焦作　454000；2. 河南理工大学 软件学院，河南 焦作　454000）

［摘　要］萤石作为一种重要的工业矿物，其品质主要由其中氟化钙的含量决定。为了有效评估和预测萤石中的氟化钙含量，提出一种基于自适应噪声完全集合经验模态分解（CEEMDAN）和卷积神经网络（CNN）以及双向长短期记忆神经网络（BiLSTM）相结合的深度学习算法用于萤石中氟化钙含量的预测。实验结果表明，所提方法具有较高的预测精度和稳定性，为萤石的质量评估和工业应用提供重要参考。

［关键词］萤石；氟化钙；卷积神经网络；自适应噪声完全集合经验模态分解；双向长短期记忆神经网络

［中图分类号］O6-39　　［文献标志码］A　　［文章编号］2097-4566（2024）09-0073-04

Research on predicting the calcium fluoride content in fluorite based on deep learning algorithm

LI Lingyun[1], LIU Haiqing[1], HAO Bin[1], REN Jianji[2]

(1. *Do-Fluoride New Materials Co., Ltd., Jiaozuo* 454000, *China*;
2. *School of Software, Henan Polytechnic University, Jiaozuo* 454000, *China*)

Abstract: Fluorite is an important industrial mineral, its quality is primarily determined by the content of calcium fluoride within it. To effectively assess and predict the calcium fluoride content in fluorite, a deep learning algorithm that combines CEEMDAN, CNN and BiLSTM networks is proposed. Experimental results indicate that the proposed method achieves high prediction accuracy and stability, providing significant references for the quality assessment and industrial application of fluorite.

Key words: fluorite; calcium fluoride; convolutional neural network(CNN); complete ensemble empirical mode decomposition with adaptive noise(CEEMDAN); bidirectional long and short term memory neural network(BiLSTM)

萤石作为一种关键工业矿物，在氟化工产品制备中扮演着至关重要的角色。萤石的主要成分为氟化钙，是制备如六氟磷酸锂等重要氟化合物的主要氟源。六氟磷酸锂是高效能电池，主要是电动汽车和移动通信设备电池所必需的关键材料，其生产过程中对氟元素的需求量极大[1]。因此，萤石中氟化钙的含量对于整个氟化工产业链的效率和产出具有决定性的影响。萤石的质量不仅直接影响下游产品的质量和稳定性，而且对优化生产流程、降低成本以及提升产品竞争力具有重要意义。随着全球对高性能电池需求量的不断增长，对氟化工原料的需求量也逐渐上升，因此，对萤石品质的持续监控和改进变得尤为重要。

传统的萤石氟化钙含量测定方法通常依赖于实验室分析，耗时长且成本高昂[2]。随着人工智能技术的发展，深度学习算法不断更新，能够处理时间序列数据的模型层出不穷，如长短时记忆网络（LSTM）、门控循环单元（GRU）以及时序卷积网络（TCN），已被证明在材料科学的成分分析和质量控制中具有极高的预测精度和实用价值。这些算法能从历史数据中学习和提取复杂的特征和趋势，提供精确的预测模型，并允许在实时环境下进行动态调整和优化，从而更好地应对复杂的工业生产条件[3]。

为了进一步提高预测模型的性能，本研究提出了一种基于完全集合经验模态分解（CEEMDAN）、卷积神经网络（CNN）[4]和双向长短时记忆网络（BiLSTM）[5]的组合深度学习框架，称为CEEMDAN-CNN-BiLSTM。该方法首先利用CEEMDAN处理原始时间序列数据，增强关键信号成分并抑制噪声，然后通过CNN层提取处理后数据的空间特征，最

［收稿日期］2024-08-20

［作者简介］李凌云(1976-)，女，河南焦作人，高级工程师，高级经济师，长期从事氟化工研究管理工作。
E-mail：DFD3153@dfdchem.com

终通过 BiLSTM 层来捕获特征间的时间依赖性。该方法的集成优势在于其能够结合各个算法的独特优点，提供更为精确和稳定的预测结果。通过该模型的训练，不仅可以提高萤石氟化钙含量预测的准确性，还可以为六氟磷酸锂原料的制备流程优化提供重要的决策支持，并显著提高最终产品的质量和市场竞争力。该研究为萤石的质量评估和氟化工产业提供了新的科学方法和实用工具，展示了深度学习技术在原材料质量控制及其工业应用中的巨大潜力。

1 CEEMDAN-CNN-BiLSTM 方法介绍

CEEMDAN-CNN-BiLSTM 框架通过 3 个阶段的协同工作（见图 1），实现了对萤石氟化钙含量预测的高精度和高稳定性。首先，CEEMDAN 方法对原始数据进行预处理，强化关键信号成分。CNN 通过 C（卷积层）和 Conv（卷积与池化）进一步提取这些加工后数据的空间特征，并将这些特征传递给 BiLSTM 层。在 BiLSTM 层利用这些特征通过 x_t（输入值）、h_t（隐藏层）、O_t（输出层）结合其对时间依赖性的深入理解，进行预测输出。

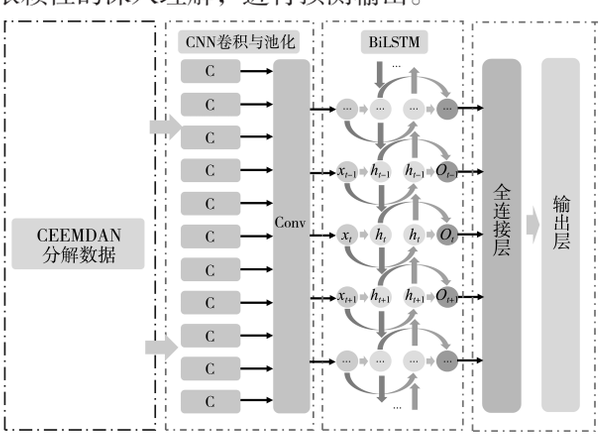

图 1 萤石氟化钙含量 CEEMDAN-CNN-BiLSTM 预测架构

Fig. 1 CEEMDAN-CNN-BiLSTM prediction architecture of calcium fluoride content in fluorite

1.1 CEEMDAN

CEEMDAN 是一种信号处理方法，用于将复杂信号分解为若干个称为固有模态函数（*IMF*）的成分。CEEMDAN 是 EMD（经验模态分解）的改进版，它通过引入白噪声来避免 EMD 在分解过程中可能出现的模式混叠问题。CEEMDAN 的优势在于它能够保持信号的自适应性，同时通过增加噪声的方式增强分解的鲁棒性。CEEMDAN 的分解过程可以通过以下步骤描述：

第一步，初始信号加噪声，给定信号 $x(t)$，引入白噪声 $w_i(t)$，然后对噪声扰动信号进行 EMD 分解，得到第一模态 $IMF_1^{(i)}(t)$：

$$x(t)+\varepsilon_1 w_i(t) \xrightarrow{\text{EMD}} IMF_1^{(i)}(t)。 \tag{1}$$

其中，ε_1 是噪声的幅度，i 表示第 i 次引入噪声。

第二步，将多次噪声扰动分解的 *IMF* 平均，得到第一个 $IMF_1(t)$：

$$IMF_1(t)=\frac{1}{N}\sum_{i=1}^{N} IMF_1^{(i)}(t)。 \tag{2}$$

其中，N 是噪声扰动的次数。

第三步，原始信号减去第一个 $IMF_1(t)$ 计算剩余信号 $r_1(t)$：

$$r_1(t)=x(t)-IMF_1(t)。 \tag{3}$$

第四步，对剩余信号 $r_1(t)$ 加入噪声，重复 EMD 分解，得到第二模态 $IMF_2^{(i)}(t)$：

$$r_1(t)+\varepsilon_2 w_i(t) \xrightarrow{\text{EMD}} IMF_2^{(i)}(t)。 \tag{4}$$

对结果取平均，得到第二个 $IMF_2(t)$：

$$IMF_2(t)=\frac{1}{N}\sum_{i=1}^{N} IMF_2^{(i)}(t)。 \tag{5}$$

第五步，递归分解，重复上述过程，依次提取出所有 *IMF*，直到信号中的所有模态都被提取出来或剩余信号趋于零。

第六步，噪声幅度自适应调整：噪声幅度 ε_k 随分解过程逐步调整，确保不同模态的分离精度。

1.2 CNN-BiLSTM

CNN 是一种深度学习网络结构，主要用于图像处理领域。通过其特有的卷积层来自动提取输入数据的重要特征，无需手动提取特征工程。在卷积层中，卷积核（滤波器）在输入数据上滑动，计算核与数据的点积，形成特征图，CNN 结构如图 2 所示。这一操作可以用式（6）表示：

$$\boldsymbol{F}\otimes w=\sum_{i=1}^{H}\sum_{j=1}^{W} F(i,j)\cdot w(i,j)。 \tag{6}$$

其中，$\boldsymbol{F}$ 表示输入的特征矩阵，w 是卷积核，H 和 W 分别代表卷积核的高度和宽度。通过应用激活函数和池化层，CNN 能够进一步增强模型的非线性特性和抽象能力。

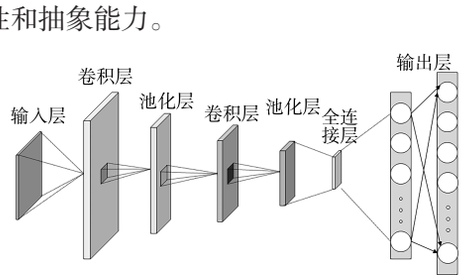

图 2 CNN 结构

Fig. 2 CNN structure

BiLSTM 是 LSTM [6] 的一个变种，能够更好地捕捉时间序列数据中的长期依赖关系。传统的

LSTM 只能从前到后处理信息，而 BiLSTM 通过增加一个反向传播的 LSTM 单元（如图 3 所示）能同时从过去和未来学习信息，极大地提高了模型对时间序列的理解能力。BiLSTM 的核心是其门控机制，包括遗忘门 f_t（式 7）、输入门 i_t（式 8）、输出门 O_t（式 9）、候选状态 g_t（式 10）、最终细胞状态 C_t（式 11）、隐藏状态 h_t（式 12），公式表述如下：

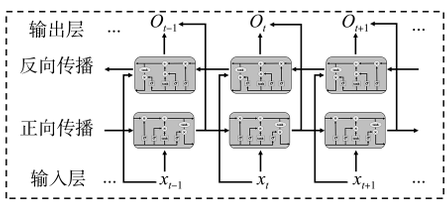

图 3　BiLSTM 结构

Fig. 3　BiLSTM structure

$$f_t=\sigma(\boldsymbol{W}_f\cdot[h_{t-1},x_t]+b_f); \tag{7}$$

$$i_t=\sigma(\boldsymbol{W}_i\cdot[h_{t-1},x_t]+b_i); \tag{8}$$

$$O_t=\sigma(\boldsymbol{W}_o\cdot[h_{t-1},x_t]+b_o); \tag{9}$$

$$g_t=\tanh(\boldsymbol{W}_g\cdot[h_{t-1},x_t]+b_g); \tag{10}$$

$$C_t=g_t\cdot i_t+C_{t-1}\cdot f_t; \tag{11}$$

$$h_t=O_t\cdot\tanh(C_t)。 \tag{12}$$

其中，当前输入 x_t 和前一个时刻的隐藏状态 h_{t-1} 连接为一个向量 $[h_{t-1},x_t]$，$\boldsymbol{W}_f$、$\boldsymbol{W}_i$、$\boldsymbol{W}_g$、$\boldsymbol{W}_o$ 分别表示遗忘门、输入门、候选状态和输出门的权重矩阵，b_f、b_i、b_g、b_o 分别表示这些门的偏置项，tanh 为激活函数。

2　实验数据集

在本研究中，使用的萤石品质数据集来源于某工厂实际采集的氟化钙含量数据。这些数据是从连续的生产过程中获取的，旨在评估和预测萤石的化学品质，主要包含其氟化钙成分的变化。数据集涵盖了多个生产周期，包括氟化钙含量的多个测量值，以及与萤石品质相关的其他化学和物理参数。

2.1　数据预处理

为了确保所使用数据集的质量，最大化预测模型的准确性与稳定性，首先对原始数据进行了预处理操作，关键在于排除可能由设备故障、操作误差或外部环境变化引起的非典型数据点。本研究采用基于 3-sigma 的统计方法进行异常值检测，具体而言，在假设数据呈正态分布的情况下，任何超出平均值±3 倍标准差范围的数据点都可以被视为异常值。该规则的数学表达式为：

$$|x-\mu|>3\sigma。 \tag{13}$$

其中，μ 是样本平均值，σ 是样本标准差，x 是单个数据点。通过此方法，能有效地标识并剔除那些可能影响分析结果的极端异常值，从而提升数据集的整体质量和后续分析的可靠性。

2.2　CEEMDAN 处理

在数据预处理之后，进一步采用 CEEMDAN 来处理萤石品质数据集。在 CEEMDAN 处理中，每次分解过程中都加入一定量的白噪声，重复多次以形成一个集合。然后，对集合中的所有 *IMF* 进行平均，以消除噪声的影响。这种方法不仅提高了固有模态函数的可靠性，也使得分解结果更加稳定和可重复，分解结果如图 4 所示。

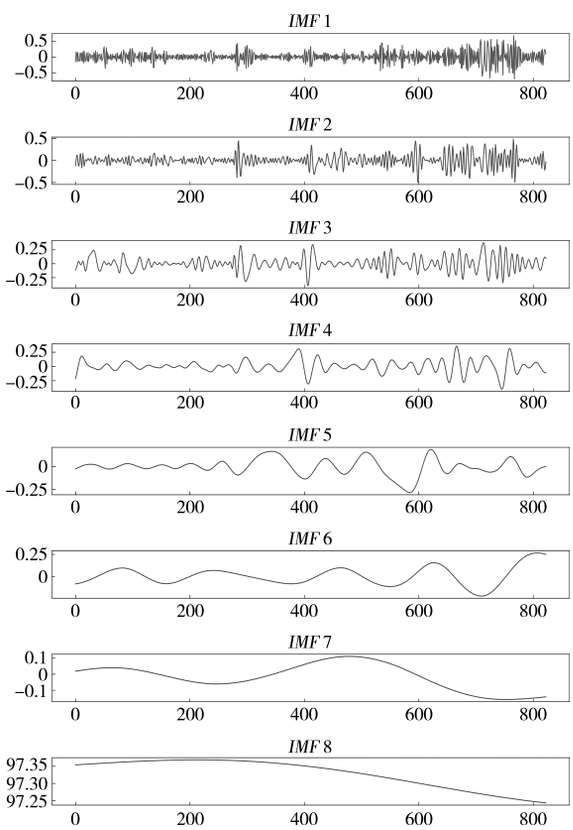

图 4　CEEMDAN 分解结果

Fig. 4　CEEMDAN decomposition results

3　实验与评估

3.1　实验设置及环境

所有实验部分都在同一实验环境下运行，实验环境配置如表 1 所示。神经网络结构参数如表 2 所示。在给定按时间顺序组成的萤石氟化钙含量向量情况进行计算。

表 1　实验环境配置

Table 1　Experimental environment configuration

配置项	详情	配置项	详情
CPU	Intel(R) Xeon(R) Gold 5218 CPU @ 2.30 GHz, 2.29 GHz	操作系统	Win11 x64
GPU	NVIDIA GeForce RTX 2080 Ti	CUDA 版本	12.1
内存	64.0 GB	PyTorch 版本	2.2.2

表2　神经网络结构参数

Table 2　Network structure parameters

编号	网络层	描述
1	Input	Input_size=8
2	CNN Layer 1	Input_channels=IMFs，Output_channels=32，Kernel_size=3，Stride=1，Padding=1，Activation=ReLU
3	CNN Layer 2	Input_channels=32，Output_channels=64，Kernel_size=3，Stride=1，Padding=1，Activation=ReLU
4	Max Pooling	Pool_size=2，Stride=2
5	BiLSTM Layer 1	Input_size=64，Hidden_size=128，Layers=1，Bidirectional=True
6	BiLSTM Layer 2	Input_size=128，Hidden_size=256，Layers=1，Bidirectional=True
7	FC Layer	Input_size=256，Output_size=1，Activation=Sigmoid

3.2　评估函数

本实验采用决定系数（R^2）、平均绝对误差（*MAE*）和均方误差（*MSE*）作为评估函数，它们的定义如下。

$$R^2=1-\frac{\sum_i(\hat{y}_i-y_i)^2}{\sum_i(\bar{y}_i-y_i)^2}\ ; \tag{14}$$

$$MAE=\frac{1}{m}\sum_{i=1}^{m}|y_i-\hat{y}_i|; \tag{15}$$

$$MSE=\frac{1}{m}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2。 \tag{16}$$

其中y_i是第i个实际值，$\hat{y}_i$是第i个预测值，$\bar{y}_i$是所有实际值的平均值，m是实际值的总数。

4　实验结果与分析

为检验所提算法是否具有较高的预测准确性，在同样的实验环境下将所提算法实验结果与其他深度学习算法的实验结果进行对比。表3展示了不同算法R^2、*MSE*、*MAE*的对比，图5展示了不同算法预测值和真实值的拟合情况。

表3　不同算法实验结果

Table 3　Experimental results of different algorithms

算法	R^2	*MSE*	*MAE*
CEEMDAN-CNN-BILSTM	0.965	0.003 0	0.047
GRU	0.895	0.009 1	0.071
CNN	0.941	0.004 2	0.052
BiLSTM	0.922	0.006 8	0.055
TCN	0.911	0.007 7	0.065
LSTM	0.939	0.005 3	0.054

由表3和图5可知，CEEMDAN-CNN-BiLSTM算法优于大多数现有方法。根据表3，相比现有方法最好的结果，所提算法的R^2提升了2.6%，*MSE*降低了28.6%，*MAE*降低了9.6%。这是因为所提方法引入了CEEMDAN和CNN，使得所提算法同时考虑了关键信号成分和数据中的空间特征，最后采用BiLSTM来增强模型捕捉潜在规律的能力。

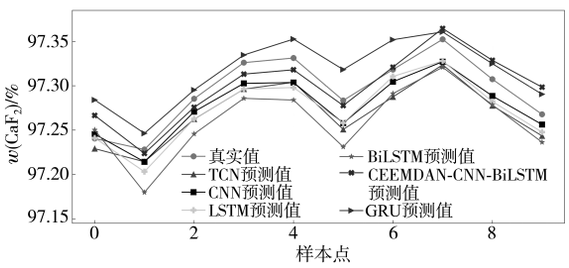

图5　预测值与真实值对比

Fig. 5　Comparison of predicted values and true values

5　结论

CEEMDAN-CNN-BiLSTM算法适用于稳定产出的萤石矿，以及已经检测合格但存在含量偏差的萤石原料，在萤石氟化钙含量预测中具有更高的准确性和稳定性，为萤石的质量评估和工业应用提供重要参考，有利于提高氟化工产业链的效率。由于采集的萤石品质数据量受限，该研究还存在一定的局限性。未来，可以提升萤石化验与抽检频率，从而获取更优的数据集以获取更优的预测结果。

［参考文献］

［1］ 宋丽萍.六氟磷酸锂合成工艺[J].化工设计通讯，2018，44(11)：112-114.
SONG L P. Synthesis Process of Lithium Hexafluorophosphate [J]. Chemical Engineering Design Communications, 2018, 44 (11): 112-114.

［2］ 于辉，曹建虎，高伟伟，等.仪器分析法测定萤石球中氟化钙含量[J].物理测试，2022，40(6)：30-32.
YU H, CAO J H, GAO W W, et al. Determination of Calcium Fluoride Content in Fluorite Spheres by Instrumental Analysis Method [J]. Physics Examination and Testing, 2022, 40(6): 30-32.

［3］ 夏传波，田兴磊，王志明，等.萤石成分分析方法的标准现状与研究进展[J].中国无机分析化学，2023，13(8)：813-824.
XIA C B, TIAN X L, WANG Z M, et al. Standard Status and Research Progress of Fluorite Composition Analysis Method [J]. Chinese Journal of Inorganic Analytical Chemistry, 2023, 13 (8): 813-824.

［4］ LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

［5］ SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks [J]. IEEE Transactions on Signal Processing, 1997, 45 (11): 2673-2681.

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