Pure Mathematics 理论数学, 2024, 14(6), 460-471
Published Online June 2024 in Hans. https://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2024.146264

Hans汉斯

融合CEEMDAN与CNN-GRU-Attentions的复合模型在碳价格预测中的应用研究

陈　昱

上海出版印刷高等专科学校，信息与智能工程系，上海

收稿日期：2024年5月11日；录用日期：2024年6月13日；发布日期：2024年6月30日

摘　要

为了提升碳价格预测的精确性，本研究引入了一种集成模型，结合了自适应噪声完备集合经验模态分解(CEEMDAN)、卷积神经网络(CNN)、改进长短期记忆网络(GRU)和注意力机制(Attentions)。该模型利用GRU的更新门简化了LSTM结构，降低了模型的参数量和复杂度。通过CNN的特征提取能力和GRU对时间序列的捕捉，模型能够模拟碳价格在时间和空间上的依赖性。注意力机制的加入进一步提升了模型对关键历史特征的识别能力。实验结果表明，该CNN-GRU-Attentions模型在拟合度、MAE、MAPE和RMSE等指标上均优于传统模型，具有较高的预测精度和实际应用潜力。

关键词

碳价格预测，CNN-GRU-Attention模型，时间序列分析

Research on the Application of a Composite Model Integrating CEEMDAN with CNN-GRU-Attentions for Carbon Price Forecasting

Yu Chen

Department of Information and Intelligent Engineering, Shanghai Publishing and Printing College, Shanghai

Received: May 11th, 2024; accepted: Jun. 13th, 2024; published: Jun. 30th, 2024

文章引用: 陈昱. 融合 CEEMDAN 与 CNN-GRU-Attentions 的复合模型在碳价格预测中的应用研究[J]. 理论数学, 2024, 14(6): 460-471. DOI: 10.12677/pm.2024.146264

Abstract

To enhance the accuracy of carbon price forecasting, this study introduces an integrated model that combines Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Convolutional Neural Networks (CNN), Gated Recurrent Units (GRU), and Attentions mechanisms. The model utilizes the GRU's update gate to simplify the Long Short-Term Memory (LSTM) structure, reducing the model's parameter count and complexity. By leveraging the feature extraction capabilities of CNNs and GRU's capture of time series, the model can simulate the temporal and spatial dependencies of carbon prices. The incorporation of the Attention mechanism further enhances the model's ability to identify key historical features. Experimental results demonstrate that the CNN-GRU-Attention model outperforms traditional models in terms of fit, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE), indicating its higher predictive accuracy and practical application potential.

Keywords

Carbon Price Forecasting, CNN-GRU-Attention Model, Time Series Analysis

1. 前言

21 世纪以来，随着全球对温室气体排放的控制与减排政策不断加强，碳交易市场作为创新的环境经济政策工具，对推动企业提高资源利用效率、促进节能减排具有显著作用。中国建立碳交易市场，既是响应国际社会减排要求的重要举措，也是实现国内“双碳”目标的关键政策之一。然而，碳交易市场的稳定性和效率性在很大程度上取决于碳价格的合理预测与调控，这直接关系到市场的健康发展和政策的有效实施[1] [2]。

碳价格预测对于完善市场机制、指导政策制定和辅助投资者决策具有重要的理论和实践价值。准确的碳价格预测能够帮助政府制定合理的碳排放配额分配政策，引导企业进行有效的节能减排投资，同时为市场参与者提供决策支持，降低交易成本和风险[3] [4]。

尽管已有研究在碳价格预测方面取得了一定的进展，但仍存在一些亟待解决的问题。现有模型往往忽视了碳价格数据的非线性、高波动性等复杂特征，导致预测精度不足。此外，随着人工智能技术的快速发展，如何有效融合传统统计方法与新兴的机器学习算法，提高预测模型的准确性和鲁棒性，成为了当前研究的热点和难点[5] [6]。

针对现有研究的不足，本研究提出了一种融合自适应噪声完备集合经验模态分解(CEEMDAN)和卷积神经网络–门控循环单元–注意力机制(CNN-GRU-Attention)的复合型碳价格预测模型。本研究的创新之处在于：首先，通过 CEEMDAN 分解技术对碳价格数据进行预处理，有效提取不同时间尺度上的特征信息，增强了模型对数据复杂性的适应能力；其次，利用 CNN-GRU-Attention 模型结合了卷积神经网络强大的特征提取能力、GRU 网络对时间序列的捕捉能力以及注意力机制对关键历史特征的识别能力，提高了模型的预测精度和实际应用潜力。本研究不仅为碳市场建设提供了新的理论支持和实践指导，也为碳价格预测领域的学术探索提供了新的视角，有望弥补现有研究的不足之处[7]-[10]。

2. 理论准备

2.1. CEEMDAN 分解

2.1.1. CEEMDAN

自适应噪声完备集合经验模态分解(CEEMDAN)并不是在 EEMD 与 CEEMD 方法上改进而来的，而是从 EMD 的基础上加以改进，但是借用了 EEMD 方法中加入高斯噪声和通过多次叠加并平均以抵消噪声的思想。他们的区别简单来说，EEMD 方法是将添加白噪声后的 m 个信号直接做 EMD 分解，然后相对应的 IMF 间直接求均值。

CEEMDAN 方法是每求完一阶 IMF 分量，又重新给残值加入白噪声(或白噪声的 IMF 分量)并求此时的 IMF 分量均值，并逐次迭代。具体分解过程可分为三步，第一步对原始信号复制 m 次并分别加入随机白噪声，形成 m 个新的信号；第二步分别进行常规 EMD 分解，将每个信号分解出的第一个 IMF 序列求平均后得到第一个 IMF 序列；第三步将原始序列减第一个 IMF 序列得到新序列；第四步重复上述过程，得到所有的 IMF 序列和残差序列。相对于 EEMD 方法，它有以下几个优势：一是完备性，即把分解后的各个分量相加能够获得原信号，CEEMDAN 在较小的平均次数下就可以有很好的完备性，而对于 EEMD 方法，较小的平均次数时，重构误差较大，完备性不强；二是计算速度更快，正是因为 CEEMDAN 相较于 EEMD 该方法不需要太多的平均次数，可以有效提升程序运算速度；三是模态分解结果更好，EEMD 分解可能会出现多个幅值很小的低频 IMF 分量，这些分量对于信号分析意义不大，CEEMDAN 方法可以减少这些分量数目。

CEEMDAN 算法的过程如下：

(1) 对于原始信号 $x(t)$，向其中添加标准正态分布白噪声。设 $\omega_i(t)$为高斯白噪声，ε 为噪声标准差。假设总共添加了 I 次白噪声，则第 i 个信号序列 $x_i(t)$可以表示为：

$$x_i(t) = x(t) + \varepsilon_0 \omega_i(t), i = 1,2,\cdots,I$$

(2) 然后对每一个重构后的信号 $x_i(t)(i = 1,2,\cdots,I)$ 进 EMD 分解，则可以得到 IMF_{ij}，表示第 i 个重构信号分解后得到的第 j 个 IMF 分量。将这些 IMF_{ij} 中的第一个分量即 IMF_{i1} 进行平均，可以得到 CEEMDAN 分解的第一个模态分量为：

$$\mathrm{IMF}_1(t) = \frac{1}{I}\sum_{i=1}^{I}\mathrm{IMF}_{i1}(1)$$

计算第一阶段的残差 $r_1(t)$：

$$r_1(t) = x(t) - \mathrm{IMF}_1(t)$$

(3) 设 $E_k(\cdot)$为通过 EMD 分解获得的第 k 阶模态分量，对 $r_1(t)$构造新序列 $Lr_1(t) + \varepsilon_1 E_1(\omega_i(t))$，对该组新序列进行 EMD 分解，提取所有分解结果当中的第一个分量即 $E_1(r_1(t) + \varepsilon_1 E_1(\omega_i(t)))$，求取均值即可得到第二个模态分量：

$$\mathrm{IMF}_2 = \frac{1}{I}\sum_{i=1}^{I} E_1\left(r_1(t) + \varepsilon_1 E_1\left(\omega_i(t)\right)\right)$$

计算第二阶段的残差 $r_2(t)$：

$$r_2(t) = r_1(t) - \mathrm{IMF}_2(t)$$

(4) 当得到 k 个 *IMF* 分量，第 k 个余量信号可以计算为：

$$r_k(t) = r_{k-1}(t) - \mathrm{IMF}_k(t)$$

则第 $k+1$ 个模态分量表示为：

$$\mathrm{IMF}_{k+1}(t)=\frac{1}{I}\sum_{i=1}^{I}E_1\left(r_k(t)+\varepsilon_k E_k\left(\omega_i(t)\right)\right)$$

(5) 重复式上上个式子，直到残余成分不再满足分解条件。最后，原始信号 $x(t)$被分解为下式，其中 R 为最终的残差：

$$x(t)=\sum_{i=1}^{K}\mathrm{IMF}_1(t)+R(t)$$

通过计算步骤可以看出，CEEMDAN 算法在计算 IMF_2 及之后的分量时并不通过直接向原始信号中添加白噪声的方式，而是对上一步的残余信号添加正负等幅值自适应噪声对，故而不会对原始信号的准确性造成干扰。另一方面，向残余信号中添加的也并非普通的高斯噪声，而是经过 EMD 分解的分量，因此具有更小的标准差。

综上，CEEMDAN 不仅能把分解后的各个分量相加能够获得原信号，在较小的平均次数下就可以有很好的完备性，而且它在分解过程中添加的噪声对不会干扰分解质量，同时具有较高的分解效率。

2.1.2. 排列熵

熵(Entropy, En)是一种较为流行的复杂度计算指标，常被用来描述时间序列系统中的混乱程度。20 世纪 40 年代，香农在信息论中引入熵概念，创建了信息熵，以概率分布的形式对系统的不确定性进行衡量，提出后被应用到各个领域。随后以信息熵为基础的各种熵以不同形式逐渐涌现，排列熵即为其中之一，其基本思想是输入序列的顺承关系，具有原理简单、计算速度快、抗噪声能力强等优点，因此将排列熵应用于对时间序列复杂度的检测可以获得较好的效果。假设 $\{x(i),i=1,2,\cdots,N\}$ 为一组时间序列信号，则排列熵的具体算法步骤描述如下：

对一维信号序列 $\{x(i),i=1,2,\cdots,N\}$ 进行相空间重构，得到重构矩阵 Y 为：

$$Y=\begin{bmatrix} x(1) & x(1+\tau) & \cdots & x(1+(m-1)\tau) \\ x(2) & x(2+\tau) & \cdots & x(2+(m-1)\tau) \\ \vdots & \vdots & \ddots & \vdots \\ x(j) & x(j+\tau) & \cdots & x(j+(m-1)\tau) \\ \vdots & \vdots & \ddots & \vdots \\ x(K) & x(K+\tau) & \cdots & x(K+(m-1)\tau) \end{bmatrix}$$

式中：

K——代表重构向量的个数($j=1,2,\cdots,K$)；

M——代表嵌入维数；

τ——代表延迟时间。

矩阵 Y 中的每一行 $\{x(j)\ x(j+\tau)\ \cdots\ x(j+(m-1)\tau)\}$ 表示一个重构分量，按照升序方式对每个重构分量中的元素进行排序可得式(3-15)。

$$x(j+(j_1-1)\tau)\le\cdots\le x(j+(j_m-1)\tau)$$

式中， $j_1,j_2,\cdots,j_m$ 分别代表重构分量中各个元素所在的列。

若存在 $x(j+(j_n-1)\tau)=x(j+(j_l-1)\tau)$ 且 $j_n<j_l$ 的情况，则按列值序号大小进行排序，即当 $j_n<j_l$ 时，规定 $x(j+(j_n-1)\tau)\le x(j+(j_l-1)\tau)$ 。

因此，对于 Y 矩阵中的每一个重构分量都可以得到对应的一组符号序列，如下式所示。

$$s(l)=\{j_1,j_2,\cdots,j_m\}$$

式中，$l=1,2,\cdots,k$，且 $k\le m!$，m 维相空间总计可以映射出 $m!$ 中不同符合序列 $\{j_1,j_2,\cdots,j_m\}$，符合序列 $s(l)$仅仅是其中一种排列。

若 k 中不同符号序列出现的概率分别为 $p_1,p_2,\cdots,p_k$，则原始信号序列 $\{x(i),i=1,2,\cdots,N\}$ 的排列熵表达式如式()。

$$h_{PE}(m)=-\sum_{j=1}^{k}p_j\ln p_j$$

一般将其进行归一化处理，得到最终排列熵值如式()所示。

$$H_{PE}=\frac{h_{PE}}{\ln(m!)}$$

式中，H_{PE} 的取值范围为[0,1]，其值越小代表信号序列越简单，随机性越低；反之，说明子序列越规律，随机程度较低。

2.2. CNN-GRU-A 模型

2.2.1. 卷积神经网络(CNN)

卷积神经网络(CNN)是一种经常用于图像、文本和信号输入的深度学习算法，由提取对象特征的堆叠层组成。CNN 由卷积层、池化层和全连接层组成。卷积层是 CNN 中特征高效提取的关键，在本文中卷积层被主要应用于对光伏功率相关数据的特征提取。下图 1 为 CNN 的基本架构。

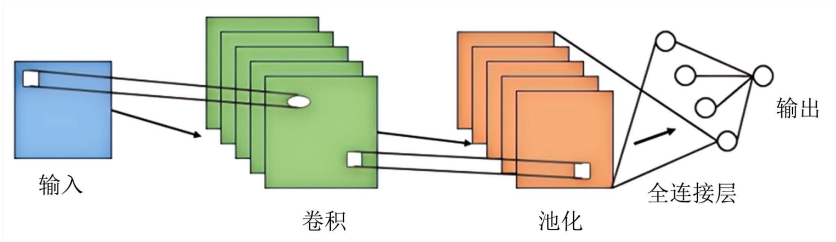

Figure 1. CNN basic architecture
图 1. CNN 基本架构

2.2.2. 门控循环单元(GRU)

GRU 是长短期记忆网络(LSTM)的一种变体，在学术界得到广泛应用。GRU 采用复位门和更新门来克服长期依赖问题，并结合数据单元和隐藏层状态来解决梯度消失的问题。相较于 LSTM，GRU 网络仅含有两个门结构，减少了训练参数的数量。这使得 GRU 网络更易于收敛，减轻了 LSTM 网络的过度拟合问题，同时保持了 LST 网络在预测任务中的卓越性能。下图 2 为 GRU 神经单元的结构。具体计算公式如下：

$$\begin{cases} r_1=\sigma\left(W_r\times\left[h_{t-1},x_t\right]\right), \\ z_1=\sigma\left(W_z\times\left[h_{t-1},x_t\right]\right), \\ \tilde{h}_t=\tanh\left(W\times\left[r_t*h_{t-1},x_t\right]\right), \\ h_t=\left(1-z_t\right)*h_{t-1}+z_t*\tilde{h}_t, \end{cases}$$

式中：z_t 表示更新门，主要用于遗忘和记忆；r_t 为复位门，用于确定是否将当前状态与先前信息合并；h_t 表示中间记忆状态；W_z 和 W_r 分别为更新门和复位门的连接权值；σ 和 tanh 为激活功能。

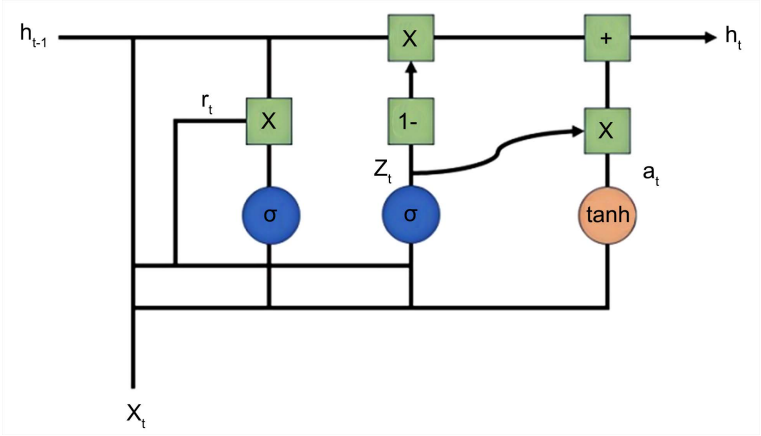

Figure 2. Structure diagram of GRU neural unit
图 2. GRU 神经单元的结构

2.2.3. 注意力机制(Attentions)

若序列数据过长，其长期学习能力会随着训练轮数的增加而逐渐衰减，因而很难保留全部有用信息。Attention 机制是模仿人脑注意力的资源分配机制而提出的，人的大脑信号处理机制在当前状态会重点关注一些目标区域，并且减少或者忽略对其它区域的关注。这样的机制会使网络在信息处理上获得更高的效率和更高的准确性。在深度学习中，Attentions 机制的核心思想是巧妙合理地改变对信息的注意力，忽略无关信息并放大关键信息。

Attentions 机制的结构如下图 3 所示，相比于简单解码器只通过一个统一的特征向量来计算隐含状态，Attention 机制在解码部分的每一个时间输入不同的。根据输入信息来计算概率分布，再进行加权平均。同时，Attentions 机制当前时刻的计算不依赖于前一时刻的计算结果，因此它可以进行并行处理，这是单一的 RNN 结构无法解决的。

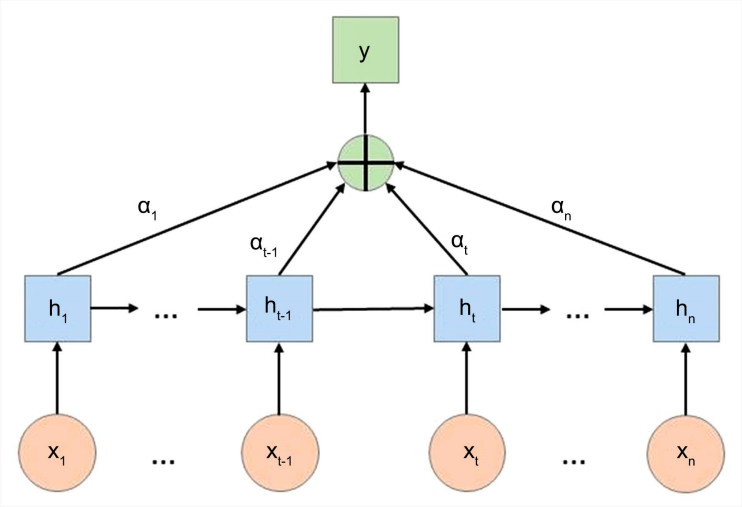

Figure 3. Attention mechanism structure diagram
图 3. Attention 机制结构示意图

3. 基于 CEEMDAN 分解和 CNN-GRU-Attention 加权组合预测建模

3.1. 实验环境

文中采用编程语言 Matlab 对深圳市碳价格数据进行 CEEMDAN 分解以及采用编程语言 Python 编写并开源基于 Keras 人工神经网络库的 TensorFlow 框架。实验平台配置为：AMD Ryzen 7 5800H 处理器，3.20 GHz*，16.0 GB 内存，Windows10 专业版 64 位(DirectX 64)操作系统。

3.2. 实验数据

选取深圳区域碳市场的碳配额价格数据进行检验，原始碳日价格序列来源于国泰安数据库，时间窗为 2013 年 8 月 5 日~2023 年 1 月 12 日，数据以天级别进行更新，剔除非交易日数据，共计 4608 条数据。碳价格数据包括开盘价、最高价、最低价、成交价、收盘价，均以人民币元为单位。其中，收盘价不仅可以展现当日交易行情，还能为后一交易日的价格提供参考依据，是市场投资者，尤其是短期投资者重点关注的价格指标，因此，收盘价具有重要观测意义，本文选取碳价格的日收盘价进行预测。

3.3. 总体研究思路

为了提高碳价格预测的精度，本文提出一种基于 CEEMDAN 分解和 CNN-GRU-Attention 组合模型的碳价格预测方法。该方法能够有效地提升预测的准确性。预测模型的流程如下图 4 所示，

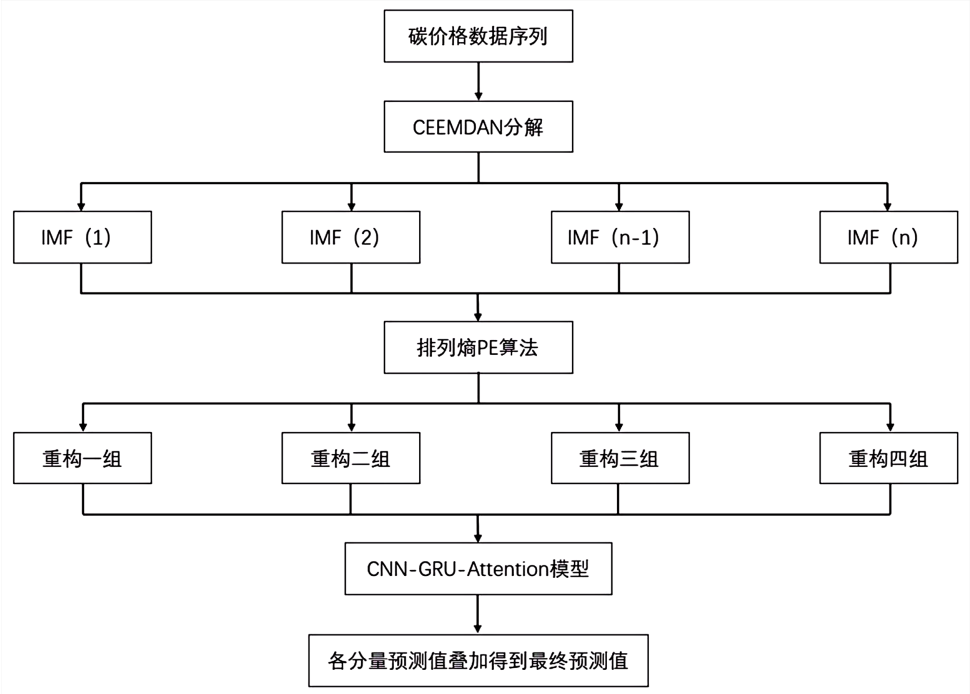

Figure 4. Prediction flow chart based on CEEMDAN decomposition and CNN-GRU-Attention combined model
图 4. 基于 CEEMDAN 分解和 CNN-GRU-Attention 组合模型的预测流程图

3.4. 模型结构、参数设置和训练过程

模型结构包括输入层、两个卷积层、一个池化层、一个全连接层、一个输出层和一个注意层。CNN 层卷积核设置为 64，内核大小为 12，并添加最大池化层，激活函数为 ReLU。GRU 网络每层的隐藏单元

个数为 60，并加入 Dropout 防止过拟合，Dropout 值设置为 0.3。

训练过程中，首先将数据输入模型，通过 CNN 层提取特征，然后送入 GRU 层捕获长期依赖关系。注意层学习隐藏层状态的重要性，最后通过全连接层得到碳价格预测结果。

具体步骤如下：

1) 自适应噪声完备集合经验模态分解数据。

由于碳价格数据受到众多外在客观因素的影响，其往往呈现出高随机性、高波动性的复杂变化特征，增加了模型的预测难度。为了最小化噪声信号对预测结果的干扰，采用 CEEMDAN 分解算法对数据序列进行平滑处理是一种可行的方法，可以有效提升程序运算速度。CEEMDAN 对碳价格序列进行分解，结果如图 5 所示：

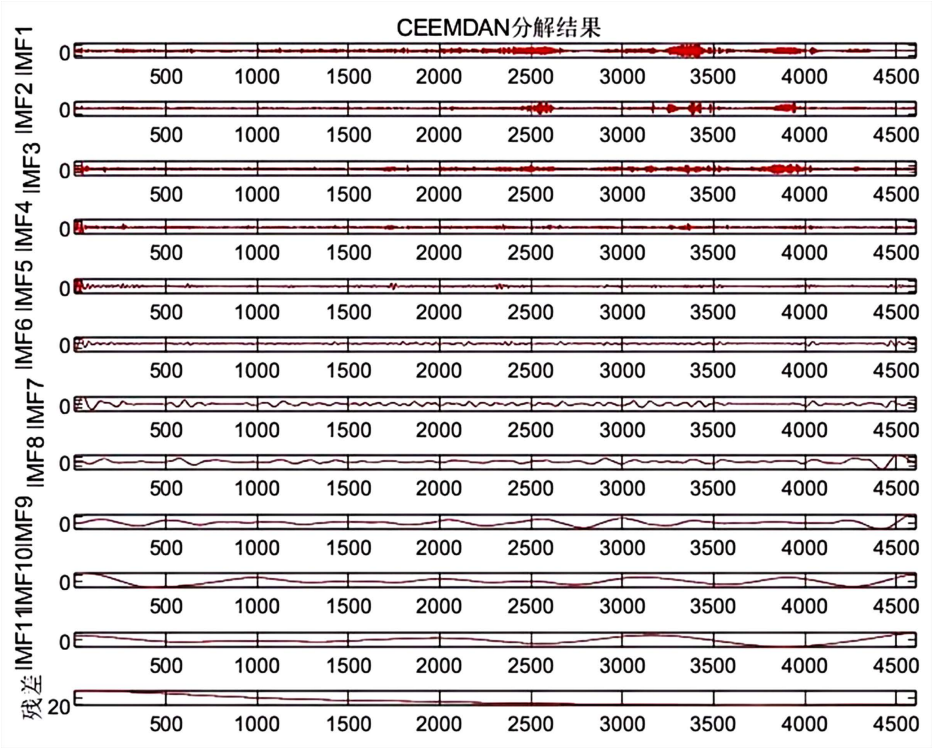

Figure 5. CEEMDAN decomposition results of carbon price
图 5. 碳价格 CEEMDAN 分解结果

从上图可以看出，CEEMDAN 分解得到了 11 组 IMF 分量和一组 RES 分量，各个序列按频率从高到低依次排列，而且不同频率的序列波动具有一定的规律性。每个子序列中存在多时间尺度信号的现象基本消除，单一频率的数据特征将更有利于模型学习，获得更好的预测精度和稳定性。

2) 排列熵重构。

排列熵是一种评价时间序列信号复杂度的方法，可以定量的对一定长度的时间序列进行检测。采用其对 CEEMDAN 分解得到的多个子序列分量进行复杂度分析，计算得到的排列熵值越大，说明序列波动越大；排列熵值越小，说明序列随机性越低，然后根据排列熵值相近的程度对子序列分量进行合并。

在利用排列熵进行序列复杂度计算时，需对相空间重构时的两个参数：嵌入维数 m、延迟时间τ进行选择。若 m 取值较大，不仅会增加计算时间，同时降低了对信号微小变化的敏感性；若 m 取值较小，重构向量的状态数减少，将会导致对信号动力学突变的检测精度下降。Bandt 等[11]建议 m 取 3~7 为宜，本

文选用的嵌入维数 $m = 5$。郑小霞等[12]指出，当延迟时间 τ 在 1~6 之间变化时，排列熵波动较小，因此本文 τ 的取值为 1。

对上一小节 CEEMDAN 分解后的各个子序列进行排列熵值计算，归一化后的结果如表 1 所示。同时依据所得数据绘制不同子序列分量的排列熵值分布图，如表 1 所示。

Table 1. The permutation entropy of each subsequence component
表 1. 各个子序列分量排列熵值

序列分量	PE 值
IMF1	0.95576
IMF2	0.8671
IMF3	0.8813
IMF4	0.70186
IMF5	0.50358
IMF6	0.35222
IMF7	0.26062
IMF8	0.21189
IMF9	0.17216
IMF10	0.15906
IMF11	0.15271
RES	0.10211

结合表 1 和图 4 及图 5 可知，随着序列号的增加，每个模态分量的 PE 值在逐渐减小，说明其复杂程度在逐步降低。通过比较归一化后的各个序列间熵值的接近程度，将归一化排列熵值相差在 0.2 以内的序列进行合并重组，残余分量是一个单调曲线，可以合并在低频分量中。从表 1 可以看出，其中 IMF1、IMF2、IMF3、IMF4 的排列熵值相差最大为 0.1983，可以合并为一组；IMF5 序列的排列熵值与其余分量都相差较大，则单独成为一组；IMF6、IMF7、IMF8 复杂度相近，可合并为一组；最后将剩余的复杂度都相对接近的 IMF9~IMF11 分量和 RES 分量叠加构成新的一组。经过重构后得到 4 组新的子序列，后续预测模型的数量也从 12 次降为了 4 次，整体模型的计算效率得到提高。

3) 模型预测。

将分解后的对每个重构子序列放入 CNN-GRU-Attention 组合模型中分别进行预测；CNN-GRU 结合了 CNN 和 GRU 的优点，并在此基础上加入注意力机制，提出来 CNN-GRU-Attention 组合神经网络模型，该模型包括一个输入层、两个卷积层、一个池化层、一个全连接层、一个输出层和一个注意层。设置一维 CNN 层卷积核为 64，内核大小为 12，并在一维 CNN 中添加最大池化层，激活函数为 ReLU。卷积层和池化层中的 stride = 1。设置 GRU 网络，每层的隐藏单元个数为 60，并在 GRU 网络中加入 Dropout，设置其值为 0.3，应用 Dropout 层可以防止模型过拟合。本文提出的基于注意力机制的 CNN-GRU 模型具体实施步骤如下：首先，将数据输入模型，通过 CNN 的输入层、卷积层和池化层以提取输入数据的特征，将得到的特征展平成一维阵列，作为时间序列输入 GRU 层。其次，CNN 层的输出被送入第一个 GRU 层，该层捕获变量之间的长期依赖关系，并将每个 GRU 层的所有隐藏状态作为注意层的输入，注意层能够学习这些隐藏层状态的重要性。最后，通过全连接层得到碳价格预测结果，如下图 6 所示。

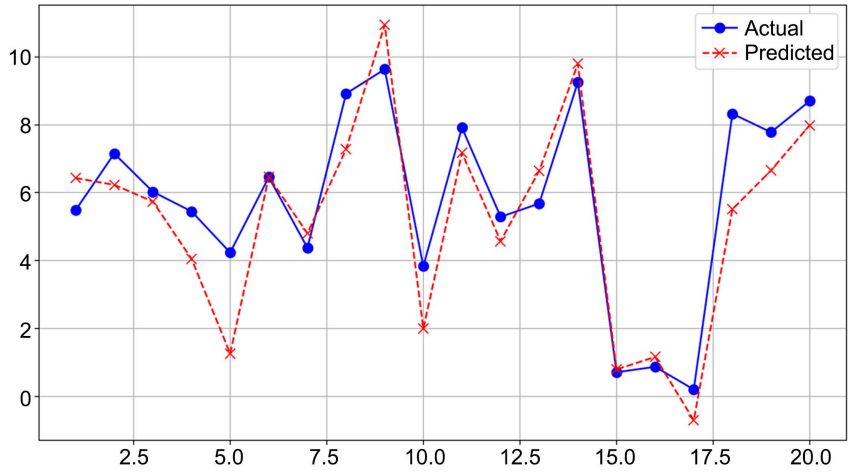

Figure 6. Prediction results of CNN-GRU-Attention model
图 6. CNN-GRU-Attention 模型预测结果

4) 参数选择。

为了评估 CNN-GRU-Attention 模型的预测精度，选用均方根误差(RMSE)、平均绝对百分误差(MAPE)、平均绝对误差(MAE)和拟合度(r^2)对算法模型进行评价。其计算公式如下：

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{t=1}^{m}\left(z_t-x_t\right)^2}{m}}$$

$$\mathrm{MAPE}=\sum_{t=1}^{m}\left|\frac{z_t-x_t}{z_t}\right|\frac{100}{m}$$

$$\mathrm{MAE}=\frac{\sum_{t=1}^{m}\left|z_t-x_t\right|}{m}$$

$$\mathrm{r}^2=1-\frac{\sum_{t=1}^{m}\left(z_t-x_t\right)^2}{\sum_{t=1}^{m}\left(z_t-\overline{x}_t\right)^2}$$

式中，z_t 和 x_t 分别表示第 t 次测试的真实值和算法模型输出值，m 为测试总次数。

3.5. 实验结果及分析

为了清晰地呈现实验结果，本研究采用表格形式展示了 CNN-GRU-Attention 模型与现有模型(GRU 模型和 CNN-GRU 模型)在碳价格预测方面的对比。表 2 展示了模型的拟合度和各项评价指标的具体数值，突出了本模型在预测精度上的优势。

Table 2. Comparison of prediction errors of different models
表 2. 不同模型预测误差对比

评价指标/Evaluation Indicators	GRU Model	CNN-GRU Model	CNN-GRU-Attention Model
MAE (Mean Absolute Error)	1.034	0.309	**0.151**
MAPE (Mean Absolute Percentage Error)	66.179	41.203	**13.014**
RMSE (Root Mean Square Error)	1.333	0.505	**0.192**
R^2 (Coefficient of Determination)	0.913	0.988	**0.998**

*注：MAE 表示平均绝对误差，MAPE 表示平均绝对百分误差，RMSE 表示均方根误差，R^2 表示拟合度。

从表 2 中可以看出，CNN-GRU-Attention 模型在 MAE、MAPE 和 RMSE 等评价指标上均优于 GRU 模型和 CNN-GRU 模型。特别是，CNN-GRU-Attention 模型的 MAE 为 0.151，MAPE 为 13.014，RMSE 为 0.192，显著低于其他两种模型，显示出更高的预测精度。此外，该模型的 R^2 值为 0.998，接近完美拟合，进一步证明了其在拟合度上的优势。

Table 3. Comparison of prediction errors between original data sequence and decomposed reconstructed data sequence
表 3. 原始数据序列与分解重构数据序列的预测误差对比

评价指标/Evaluation Indicators	Original Data Sequence	Decomposed Reconstructed Data Sequence
MAE (Mean Absolute Error)	4.338	**0.151**
MAPE (Mean Absolute Percentage Error)	25.923	**13.014**
RMSE (Root Mean Square Error)	6.256	**0.192**
R^2 (Coefficient of Determination)	0.858	**0.998**

从表 3 中可以看出，经过 CEEMDAN 分解重构后的数据序列在 CNN-GRU-Attention 模型中的预测误差明显降低，这表明 CEEMDAN 分解能够有效提取碳价格数据中的关键特征，提高模型的预测精度。

综上，基于 CEEMDAN 分解和 CNN-GRU-Attention 组合预测模型的各项评价指标均优于其他模型，可以同时显示预测准确性和稳定性。

3.6. 模型探讨

CNN-GRU-Attention 模型的优势在于其综合运用了多种先进的机器学习技术。CEEMDAN 分解技术的应用提高了模型对碳价格数据中不同时间尺度特征的提取能力。CNN 的引入增强了模型对局部特征的捕捉，而 GRU 网络则有效处理了时间序列数据的长期依赖问题。此外，注意力机制的加入使得模型能够更加关注对预测有重要影响的历史信息，进一步提升了预测的准确性。

尽管 CNN-GRU-Attention 模型在实验中表现出色，但仍存在一些局限性。首先，模型的复杂性较高，需要大量的训练数据和计算资源。其次，模型对输入特征的敏感性较高，对噪声和异常值的处理能力有待提高。此外，模型的解释性较差，对于为何做出特定的预测缺乏直观的理解。模型的泛化能力需要在更多的碳交易市场中进行验证。模型对于非市场因素，如政策变动的响应需要进一步研究。模型的实施需要专业的数据科学团队和相应的技术支持。

3.7. 改进和优化建议

为了改进和优化 CNN-GRU-Attention 模型，未来的研究可以考虑以下几个方向：探索更高效的模型结构，减少模型的参数数量，以降低计算复杂度。引入正则化技术，如 Dropout 或 L1/L2 正则化，以提高模型对噪声和异常值的鲁棒性。利用解释性机器学习方法，提高模型的可解释性，帮助用户理解模型的预测过程。

4. 结论与展望

本研究提出了一种创新的复合模型，该模型结合了 CEEMDAN 分解、卷积神经网络(CNN)、门控循环单元(GRU)和注意力机制(Attentions)，用于提高碳价格预测的准确性。实验结果表明，该模型在拟合度、平均绝对误差(MAE)、平均绝对百分比误差(MAPE)和均方根误差(RMSE)等关键评价指标上均优于现有模型，具有显著的预测优势。

本模型的开发对于碳市场的参与者具有重大意义，它不仅为政府和政策制定者提供了一个强有力的工具，以更准确地预测和制定碳排放政策，还帮助企业和投资者做出更明智的决策，优化资源配置和降低交易风险。

尽管本研究的模型已经展现出卓越的性能，但仍有改进空间。未来的研究可以探索以下几个方向：增强模型的泛化能力，通过在不同的碳交易市场和时间段进行测试。提高模型对市场动态和政策变化的适应性和鲁棒性。增强模型的可解释性，以便更好地理解预测结果背后的逻辑。

该模型的潜在应用领域广泛。政策制定：辅助政府机构制定和调整碳排放政策。企业战略：帮助企业评估节能减排措施的成本效益。金融投资：为投资者提供精确的市场趋势分析，优化投资决策。

综上所述，本研究的复合模型在碳价格预测领域具有重要的实用性和广阔的应用前景。未来的研究将继续推动模型的发展，以满足不断变化的市场需求。

参考文献

[1] Chevallier, J. (2009) Nonparametric Tests of Trends for Carbon Prices Forecast. *Energy Economics*, **31**, 535-543.

[2] Koop, G. and Tole, L. (2008) The Dynamic Demand for Carbon Emission Quotas. *Journal of Environmental Economics and Management*, **56**, 117-130.

[3] Byun, S. and Cho, S. (2013) Forecasting CO_2 Emission Allowance Prices Using GARCH Models. *Energy Economics*, **36**, 54-63.

[4] Han, L., *et al*. (2018) Forecasting Carbon Prices Using a Limited Distributional Lag Model and a Genetic Algorithm. *Journal of Cleaner Production*, **172**, 2747-2757.

[5] Atsalakis, G.S., *et al*. (2015) Carbon Trading Prediction Using Computational Intelligence Techniques. *Energy*, **81**, 701-710.

[6] Abdi, M. and Taghipour, S. (2016) A Probabilistic Neural Network Model for CO_2 Price Forecasting. *Energy Conversion and Management*, **111**, 227-236.

[7] Tsai, H.T. and Kuo, C.H. (2015) A Radial Basis Function Neural Network for Carbon Price Forecasting. *Applied Energy*, **137**, 218-228.

[8] 关晓轲. 基于灰色预测模型的碳交易价格预测研究[J]. 系统工程, 2017, 35(2): 95-100.

[9] 汪文隽, 等. 基于多元 GARCH-BEKK 模型的中国碳市场溢出效应研究[J]. 管理科学学报, 2018, 21(2): 1-12.

[10] 李沙沙. 基于多层感知机神经网络的欧盟碳交易市场碳价预测模型[J]. 系统工程理论与实践, 2019, 39(6): 1465-1474.

[11] Daskalakis, G. (2013) On the Efficiency of the European Carbon Market: New Evidence from Phase II. *Energy Policy*, **54**, 369-375. https://doi.org/10.1016/j.enpol.2012.11.055

[12] Tavoni, M., Kriegler, E., Riahi, K., van Vuuren, D.P., Aboumahboub, T., Bowen, A., *et al*. (2014) Post-2020 Climate Agreements in the Major Economies Assessed in the Light of Global Models. *Nature Climate Change*, **5**, 119-126. https://doi.org/10.1038/nclimate2475