2024 3rd International Conference on Energy and Power Engineering, Control Engineering (EPECE) | 979-8-3503-6609-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/EPECE63428.2024.00033

Gas Concentration Prediction of Power System Based on EMA-Autoformer

Xiaohu Wang[1,a], Fei Nie[1,b], Xiaoyuan Zhang[2,c], Weibin Wang[1,d], Hengwei Lei[1,e*]
[1]PipeChina West East Gas Pipeline Company, Shanghai, China
[2]Huazhong University of Science and Technology, Wuhan, Hubei, China.
[a]wangxh25@pipechina.com.cn; [b]niefei@pipechina.com.cn; [c]xiaoyuanz@hust.edu.cn; [d]wangwb04@pipechina.com.cn;
[e*]leihw@pipechina.com.cn.

***Abstract*—As a core component in power systems, the monitoring and early warning of electric transformer operation status are of great importance. Oil chromatography analysis is an effective method for detecting internal faults of electric transformers, and accurate prediction of oil chromatography gas concentration is crucial for judging the operation status of electric transformers and detecting potential faults. This study proposes an EMA-Autoformer algorithm based on an improved model with self-attention and exponential moving average. Firstly, the exponential moving average method is used to fill the missing values in oil chromatography gas data. Secondly, the self-attention mechanism based on auto-correlation algorithm in the Autoformer model is employed to extract the time correlation of gas concentration time series, enabling accurate prediction of gas concentration. Experimental results show that the accuracy of the EMA-Autoformer algorithm in predicting oil chromatography gas concentration significantly surpasses that of time series prediction models such as Autoformer, Refomer, and Informer. Moreover, compared to Refomer and Informer, Autoformer performs better on the gas concentration data processed by EMA.**

Keywords-Gas concentration prediction, Exponential moving average, Auto-correlation, Self-attention mechanism

I. Introduction

The stable operation of power system equipment is of significant importance for ensuring production safety. Among them, electric transformers are widely used in power systems, which can achieve voltage and current transformation and energy transmission, thus ensuring the normal maintenance and operation of power systems. However, in actual operation, the safe operation of electric transformer equipment may be affected by external environmental factors or the operating status of internal electrical equipment. Oil chromatography analysis technology is an important means to detect the operating status of electric transformer equipment. By analyzing the gas concentration in the oil, the operating condition of the equipment can be judged, and the potential fault risk can be predicted. Traditional gas concentration detection methods depend on manual inspection and laboratory analysis, which have problems of long detection cycles and inability to effectively predict gas concentration. As an adaptive and efficient prediction method, deep neural networks have achieved significant results in many fields. Applying deep neural networks to the prediction of gas concentration in oil chromatography equipment can effectively achieve real-time detection and early warning of equipment operating status.

Taking the power system of the West-East Gas Transmission station as an example, a large amount of oil chromatography gas concentration data was collected, and the improved model with self-attention mechanism was introduced into the gas concentration prediction field for the first time, proposing the EMA-Autoformer algorithm. Firstly, the missing values in the oil chromatography gas concentration time series were processed using the exponential moving average (EMA) method[1]; subsequently, the local time correlation and global correlation of the sequence were extracted using the self-correlated self-attention mechanism model (Autoformer)[2] to achieve accurate prediction of equipment gas concentration. Finally, the performance of the proposed model was evaluated by comparing the prediction errors of various models, providing an effective method for predicting the gas concentration of oil chromatography in power systems.

II. Related Work

In recent years, gas concentration prediction methods have mainly been based on machine learning algorithms, deep learning algorithms, and combinations of these two types of algorithms.

In order to improve the prediction accuracy of the least squares support vector machine algorithm (LSSVM) for gas concentration, Zhu et al.[3] proposed an enhanced sparrow search algorithm (ISSA) and compared it with the particle swarm optimization algorithm (PSO) and the original sparrow search algorithm (SSA). The newly proposed model significantly enhances the prediction accuracy of gas concentration compared to the original model.

Zhang et al.[4] proposed a fully-connected temporal multilayer graph convolutional network, which first utilizes multi-graph convolutional layers (MGC) to learn the topological structure of the gas sensor network, enhancing the extraction capability of data spatial features. Subsequently, gated recurrent units (GRU) are employed to extract the temporal features of gas time series, and finally, fully connected layers are used to process these features.

Liu et al.[5] proposed a method that combines Pearson's correlation coefficient and long-short term memory (LSTM) networks. This approach utilizes Pearson's correlation coefficient to select features of gas concentration data, followed by feeding the feature sequences into the LSTM network to extract temporal correlations. By doing so, predictions of gas concentration at future time steps can be made, with adaptive moment estimation (Adam) employed for parameter optimization.

Dey et al.[6] proposed a model that combines t-distributed stochastic neighbor embedding (t-SNE), variational autoencoder (VAE), and bidirectional LSTM networks. Firstly,

979-8-3503-6609-9/24/$31.00 ©2024 IEEE
DOI 10.1109/EPECE63428.2024.00033

the t-SNE algorithm is employed to reduce the dimensionality of gas concentration data. Secondly, the VAE algorithm is utilized to reconstruct the internal features of low-dimensional gas concentration. Finally, the obtained gas concentration features are input into the bidirectional LSTM network to make predictions on gas concentration.

However, in existing gas concentration prediction methods, machine learning algorithms have lower complexity than deep learning algorithms, which often lead to issues of underfitting or overfitting. Among the existing deep learning algorithms, recurrent neural networks (RNNs) dominated by LSTM or GRU are always employed due to their ability to effectively extract local correlations in sequences[7, 8]. However, these networks cannot simultaneously consider the global correlations, resulting in the loss of some important information. Moreover, existing methods cannot utilize effective strategies to fill missing values in data, thereby breaking the continuity of gas concentration time series. The proposed EMA-Autoformer method fills missing values in data using the EMA algorithm to enhance temporal continuity, and extracts both local and global correlations in gas concentration time series using the Autoformer algorithm.

III. Model structure and method

In practical scenarios, power system equipment may encounter data missing issues when collecting gas concentration data from oil chromatography due to complex external factors or internal device influences. This study first fills the missing values in the time series of oil chromatography gas concentration data reasonably and then utilizes a network capable of accurately extracting time-related features to predict future gas concentrations. For the first time, an improved model with self-attention mechanisms is applied to the processing of gas concentration time series. The proposed EMA-Autoformer method fills missing values in gas concentration data using the EMA algorithm, forming a sequence with stronger temporal continuity. Furthermore, a self-attention mechanism model based on auto-correlation is employed to extract the strong temporal continuity, enabling accurate predictions of gas concentrations.

A. EMA

Exponential moving average algorithm, as a weighted moving average algorithm with exponential decay, can be used to fill missing values in gas concentration. Its core idea is to fill the missing value at the current time by taking a weighted sum of values from previous moments, assigning larger weights to data that are closer in time and smaller weights to data that are further away in time. The weights exponentially decay over time[1]. This method can effectively fill missing values in gas concentration.

Due to the input data being vector sequences, each dimension of the vector represents the concentration value of a certain gas. When filling missing values, the vector sequences are divided into dimensions, and each gas forms a separate concentration time series for processing. In each concentration time series of a gas, starting from a non-missing value before the missing value, an exponential moving average is performed gradually towards the end of the sequence until the missing value moment is reached. The formula for calculating the filled value at the missing value position is:

$$EMA_t = \beta EMA_{t-1} + (1-\beta)\theta_t \qquad (t \le N) \quad (1)$$

Where EMA_t represents the exponential moving average result at time t starting from the exponential moving average origin ($EMA_0 = 0$), β denotes the weighted values, θ_t represents the true gas concentration at time t, and N is the sequence length used for calculating the moving average. The final result of EMA_t will be used as the filled value for missing value positions, ensuring a better temporal continuity in the sequence.

B. Autoformer

The overall structure of the Autoformer is similar to that of the Transformer[9], consisting of encoders and decoders. The overall structure is shown in Figure 1, where the decompose module is used for sequence decomposition, the Auto-correlation module is the self-attention layers based on auto-correlation, also referred to as auto-correlation layer, the Add & Norm module represents residual and normalization layers, and the Feed Forward module refers to the feed-forward layers. In the original Transformer model, the q, k, and v vectors generated for each vector are calculated directly through matrix multiplication to obtain attention scores and outputs. The distinguishing feature of the Autoformer is its use of a sequence decomposition architecture to divide sequences into seasonal and trend components, and the repeated use of sequence decomposition modules and self-attention modules based on auto-correlation to calculate attention scores and outputs in both encoders and decoders.

1) Sequence decomposition

The sequence decomposition module can decompose time series into trend and seasonal components, achieving better results by deeply processing the seasonal and trend components separately. The seasonal and trend components are achieved through simple moving average[2].

$$X_t = MovingAverage(Padding(X)) \qquad (2)$$

$$X_s = X - X_t \qquad (3)$$

Where X_t represents the trend component of the input sequence X, X_s denotes its seasonal component, $MovingAverage()$ refers to the simple moving average algorithm, and $Padding()$ represents the operation of edge supplementing the sequence. We use $X_t, X_s = decompose(X)$ to represent the above equation.

The input to the model encoder is the gas concentration time series X_{in} of the past L time steps. The initial input to the decoder includes the initialization of the trend component and the initialization of the seasonal component, both of which are formed by the latter half of X_{in}, denoted as $X_{in}[\frac{L}{2}:L]$. Their calculation formulas are as follows:

$$X_{t\_init}, X_{s\_init} = decompose(X_{in}[\tfrac{L}{2}:L]) \qquad (4)$$

In this formula, X_{t_init} and X_{s_init} represent the initialization of the trend and seasonal components, respectively. The mean of the input data is added to the tail of

the trend component, and zeros are added to the tail of the seasonal component to form a sequence of length $\frac{3L}{2}$. This sequence serves as the initial input to the decoder at the initial time.

In the encoder and decoder of the model, the output of the auto-correlation layers or feed-forward layers needs to go through the sequence decomposition module to become the input of the next layer. The calculation method is as follows:

$$Y_t, Y_s = decompose(Y) \tag{5}$$

In this equation, Y represents the output sequence of the model's auto-correlation layer or feed-forward layer, and Y_t and Y_s denote the trend and seasonal components obtained from Y, respectively. The calculation in each layer of the encoder only uses the seasonal components Y_s. The output obtained is used for cross-attention calculation with the intermediate layer output of the decoder, helping the decoder focus on information with greater global relevance. In addition to using the seasonal components Y_s to calculate self-attention and feed-forward output in the decoder, the trend components Y_t is also added to the output result to supplement the changing trend information of the sequence.

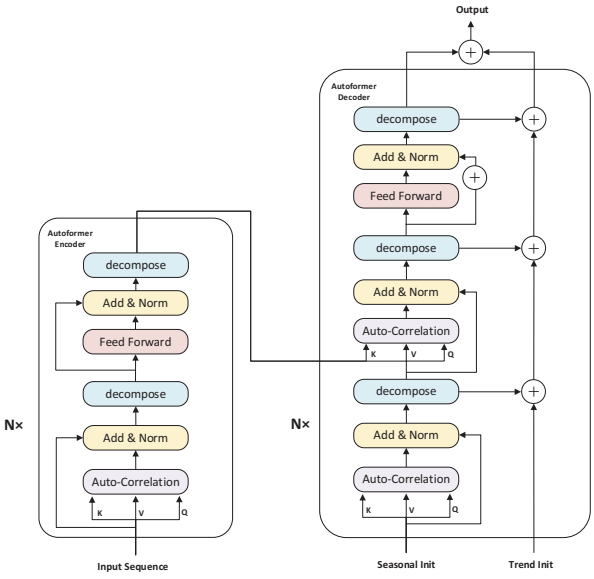

Figure 1. The overall structure of the Autoformer model.

2) Auto-correlation

In the Autoformer, the auto-correlation layer is the self-attention layer based on auto-correlation, which utilizes the auto-correlation values of the query (q) and key (k) vectors to represent the attention scores between vectors. Additionally, it employs an auto-correlation-based method to compute the weighted sum of the value (v) vector, resulting in the output[2]. The auto-correlation layer captures the local time correlation of the gas concentration sequence by using the attention scoring method based on auto-correlation calculations. Furthermore, it obtains the global correlation of the entire sequence through the self-attention mechanism.

Based on the theory of stochastic processes[10], the calculation formula for the auto-correlation $A_{XX}(\tau)$ of a sequence is as follows:

$$A_{XX}(\tau) = \lim_{L\to\infty} \frac{1}{L} \sum_{t=1}^{L} X_t X_{t-\tau} \tag{6}$$

$A_{XX}(\tau)$ represents the time-delay correlation between discrete-time sequences X_t and $X_{t-\tau}$, which largely indicates the intrinsic local temporal correlation within the sequences. In this method, the largest k values in the auto-correlation results are selected as the attention scores to obtain the weighted sum as the output.

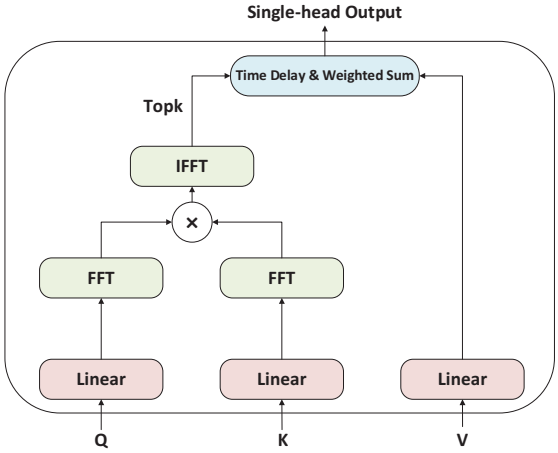

Figure 2. The calculation process of the auto-correlation layer.

The calculation process of the auto-correlation layer is shown in Figure 2, where FFT and IFFT modules represent fast Fourier transform and its inverse transform, respectively. "Topk" represents the top k largest values and their corresponding delays in the auto-correlation. The Time Delay & Weighted Sum module indicates that the input sequence is shifted cyclically according to the delays in Topk, and the weighted sum of the auto-correlation corresponding to the delays is calculated. Taking one head of the multi-head attention as an example, we first generate q, k, v vectors for each vector in the input vector sequence of length L, and concatenate them to form Q, K, V sequences. Subsequently, it is necessary to calculate the attention score matrix based on auto-correlation using Q and K sequences. According to the Wiener–Khinchin theorem, the auto-correlation of a sequence can be calculated using Fourier transform[10]. Considering sequences Q and K as time series Q_t and K_t, the calculation formula for the attention score based on auto-correlation is as follows:

$$S_{Q,K}(f) = \mathcal{F}(Q_t)\mathcal{F}^*(K_t) \tag{7}$$

$$A_{Q,K}(\tau) = \mathcal{F}^{-1}(S_{Q,K}(f)) \tag{8}$$

In this formula, $A_{Q,K}(\tau)$ and $S_{Q,K}(f)$ represent the auto-correlation and its frequency-domain representation of sequences Q and K, respectively. $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier forward and inverse transforms, while $*$ indicates conjugation.

Then, the following formula is used for calculating self-attention based on auto-correlation:

$$A_{Q,K}(\tau_1),\ldots,A_{Q,K}(\tau_k) = Topk(A_{Q,K}(\tau)) \quad (9)$$

$$\hat{A}_{Q,K}(\tau_1),\ldots,\hat{A}_{Q,K}(\tau_k) = softmax(A_{Q,K}(\tau_1),\ldots,A_{Q,K}(\tau_k)) \quad (10)$$

$$Output = \sum_{i=1}^{k} R(V,\tau_i)\,\hat{A}_{Q,K}(\tau_i) \quad (11)$$

Where $Topk(\,)$ represents selecting the top k values from the auto-correlation. $R(V,\tau_i)$ denotes the result of performing a circular shift of sequence V with a length of τ_i. The output is obtained by computing the weighted sum of the auto-correlation corresponding to its time delay.

The above illustration demonstrates the calculation method for the single-head output in the multi-head attention mechanism. By concatenating the output vector sequences of each head in order according to the aforementioned method and feeding them into the fully connected layer for adjustment, the final output sequence can be obtained.

IV. Results and Discussion

In order to provide a clearer understanding of the experimental results, this section first introduces the structure of the gas concentration dataset. Subsequently, the performance of the algorithm is evaluated using commonly used metrics in the field of time series forecasting, namely Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). A comparison is made with existing algorithms to demonstrate the effectiveness of the EMA-Autoformer method in the field of gas concentration prediction.

A. Description of the dataset.

The dataset used in this study originates from the electric transformer oil chromatography data of the West-East Gas Transmission station power system. This dataset consists of time series containing multiple time points, with each time point represented by an 8-dimensional vector. Each dimension represents the concentration of hydrogen, carbon monoxide, carbon dioxide, methane, ethylene, acetylene, ethane, and water vapor at that specific time point, with units of parts per million by volume (ppmv). Furthermore, the dataset will be divided into training, testing, validation sets for model training, testing, and validation, respectively. Due to external factors or internal device factors, some gas concentrations in the data at certain time points are missing. When performing gas concentration prediction, the gas concentration sequence from past moments is input into the network model to predict the gas concentration values at future moments.

B. Experimental Results and Evaluation.

This study evaluates the performance of the EMA-Autoformer algorithm by comparing the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) between predicted sequences and true sequences in the testing set. The EMA-Autoformer algorithm is compared with the Autoformer algorithm to demonstrate the improvement achieved by using the EMA algorithm in conjunction with the model. Additionally, a comparison is made with the Refomer and Informer models, both with and without the EMA algorithm, to highlight the effectiveness of combining the two algorithms in the EMA-Autoformer approach. The MAE and MSE for each algorithm are presented in Table 1.

Table 1. Formatting sections, subsections and subsubsections.

	Autoformer		Reformer		Informer	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
w/o EMA	0.2033	0.6599	0.2040	0.6483	0.1903	0.6511
w/ EMA	0.0236	0.0304	0.0508	0.0664	0.0334	0.0422

Compared to directly using the Autoformer algorithm to process gas concentration data without any preprocessing, the MAE of the EMA-Autoformer algorithm decreased by 88.39%, and the RMSE decreased by 95.39%. This indicates that the EMA algorithm can significantly enhance the time continuity of gas concentration data with missing values, allowing the Autoformer model to extract more accurate time-related patterns. As a result, the accuracy of the model for predicting gas concentration has been significantly improved. Meanwhile, compared to the Reformer and Informer models using EMA for data preprocessing, the MAE of the EMA-Autoformer algorithm decreased by 53.54% and 29.34%, respectively, and the RMSE decreased by 54.21% and 27.96%, respectively. This suggests that the Autoformer model performs better on gas concentration prediction when using gas data processed by EMA. By comparing the MAE and RMSE errors of the algorithms in the table, it can be known that the EMA-Autoformer algorithm has a good effect on gas concentration prediction, enabling effective monitoring and early warning of the safe operation status of equipment.

V. Conclusion

This study addresses the existing problems in gas concentration prediction methods and introduces an improved model using self-attention mechanism in this field for the first time. A novel EMA-Autoformer prediction method is proposed, which fills the missing values in gas chromatography data using the EMA algorithm and predicts gas concentrations using the Autoformer model. This approach achieves high prediction accuracy and provides an effective method for detecting and alerting the operational status of power system equipment.

Experimental results demonstrate that the proposed EMA-Autoformer method has significant advantages over other algorithms in oil chromatography gas concentration prediction tasks. This indicates the effectiveness of the EMA algorithm in handling incomplete gas data and the superiority of the Autoformer model in capturing local and global correlation information in gas concentration time series. It also highlights the better performance of Autoformer on gas data processed by EMA.

References

[1] Gardner Jr, Everette S. "Exponential smoothing: The state of the art." Journal of forecasting 4.1 (1985): 1-28;

[2] Wu, Haixu, et al. "Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting." Advances in Neural Information Processing Systems 34 (2021): 22419-22430;

[3] Zhu Zehao, et al.Concentration Prediction of Multi-component Gases Based on Improved Sparrow Search Algorithm.Journal of Physics: Conference Series 2650.1(2023);

[4] Zhang Yanmei, et al.FTM-GCN: A novel technique for gas concentration predicting in space with sensor nodes.Sensors and Actuators: B. Chemical 399.(2024);

[5] Liu Chao, et al.LSTM-Pearson Gas Concentration Prediction Model Feature Selection and Its Applications.Energies 16.5(2023):2318-2318;

[6] Dey, Prasanjit, et al. "t-SNE and variational auto-encoder with a bi-LSTM neural network-based model for prediction of gas concentration in a sealed-off area of underground coal mines." Soft Computing 25 (2021): 14183-14207;

[7] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780;

[8] Chung, Junyoung, et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling." ar**v preprint ar**v:1412.3555 (2014);

[9] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017);

[10] Brockwell, Peter J., and Richard A. Davis, eds. Introduction to time series and forecasting. New York, NY: Springer New York, 2002.