Research and Application of Oil Chromatography Data Prediction Method based on WATL

Yuanhui Zhang

Department of Computer, North China Electric Power University, Baoding, 071000, Hebei China.

383965312@qq.com

Abstract

As the key core equipment of power system, the safe and stable operation of transformer is of great significance. Oil chromatographic data is an important index to measure the running state of power transformers. It is widely used in engineering practice to predict the running state of power transformers by oil chromatographic curve changes. Based on Wavelet decomposition, Temporal Convolutional Network and Long short-term memory Network are presented in this paper Memory technology uses the principle of Encoder-Decoder for fusion, and introduces the attention mechanism, and puts forward the oil chromatography data prediction method based on WATL. Secondly, in order to fully learn the distribution law of oil chromatography data and solve the time-dependent problem of oil chromatography data, In this paper, using the principle of encoder-decoder, the time convolutional network is used as the encoder part to encode the low frequency component and the high frequency component respectively, and the output of the last hidden layer of the time convolutional network is used as the input of the decoder's long and short term memory neural network to make predictions on different components respectively. Then, in order to reflect the different correlation between different feature quantities of oil chromatographic data, attention mechanism was introduced in the decoding process, so that the weight of features that greatly affected the current output was added to the model, and the final predicted results were obtained by wavelet reconstruction. Compared with the single long and short memory network and autoregressive algorithm model, this method has a great improvement in index performance under the premise of taking calibration coefficient, mean square error and absolute error as evaluation indexes. Finally, according to the good results obtained by the WATL model in the experiment, a Web-based oil chromatography prediction system was designed and developed. The model was applied to the system, and the short-term curve trend of the transformer was predicted by using the historical oil chromatography data to generate the short-term operating state trend report of the transformer.

Keywords

Transformer; Oil Chromatogram; Time Convolutional Network; Long and Short Memory Network; Wavelet Decomposition; Attention Mechanism.

1. Introduction

In the past, the traditional power grid has formed a relatively mature protection system in the operation, maintenance and maintenance of transformers and other equipment. In the development process of the maintenance system of power transmission and transformation equipment, there have appeared three maintenance systems for power transformers, namely, fault maintenance (after-the-fact maintenance), planned maintenance and in-condition maintenance [15]. With the rapid development of sensor technology, computer technology and micro-electronic technology, it is possible to monitor and measure the power equipment in operation online, and then obtain the information of the equipment in real time. Through the analysis and comprehensive processing of the information obtained, the operating state (or power supply reliability) of power transformers and other equipment can be predicted and evaluated in real time according to its numerical characteristics, so as to detect potential faults in equipment as early as possible [16]. Thus, the third state maintenance is born, which can monitor the installation of power equipment in online operation in real time. However, because the online monitoring algorithm is not very mature, most of them use statistical methods such as the ratio of three, and the cost of the monitoring device is very high. In most cases, the transformer running state can not be accurately obtained.

Oil chromatographic data as an important indicator of measuring transformer running state, the use of historical monitoring of transformer oil chromatographic parameters related to data, modeling projections for the future quantity of oil chromatographic condition, can effectively grasp the development trend of transformer running, the preventive measures drawn up under the adverse trend of development of transformer faults has important reference meaning [17]. Therefore, using oil chromatographic data prediction, through the oil chromatographic variation curve to understand the transformer operation state in advance, as early as possible to find the potential transformer operation fault, improve the reliability of the power network, is an important topic that the power sector has been concerned about.

2. Related Work

Oil chromatographic data is an important index to measure the running state of power transformers. Predicting the running state of transformers through the change of oil chromatographic curves can find the potential faults of transformers early. The traditional oil chromatography prediction algorithm is not high in prediction accuracy and does not consider the different degree of interaction between characteristic gases. The Watl algorithm proposed in this paper solves the problem of long-term dependent information loss by encoding oil chromatographic data by using time convolutional network (TCN). It is very suitable for predicting oil chromatographic data by using long and short term memory network (LSTM) in theory, because the output content of LSTM is not only related to the current input content. It also has a lot to do with the state of history. In order to reflect the different degree of influence of the correlation between various gases on the output, the attention mechanism was introduced in this paper to allocate the weight of each characteristic gas in the oil chromatography, so as to improve the prediction accuracy and reliability.

2.1 Basic principle of WATL prediction model

A time series prediction is defined in a given input sequence $x_0, x_1, \ldots, x_t$ under the condition, the corresponding output is expected to be predicted each time $y_{t+1}, y_{t+2}, \ldots, y_T$ the key constraint is that in order to predict the output y_t over time t, we are restricted to using only previously observed input $x_0, \ldots x_t$, Formally, the time series modeling network is an arbitrary function $f : x^T \rightarrow y^{T+1}$ generate mappings:

$$\hat{y}_0, \ldots, \hat{y}_T = f(x_0, \ldots, x_T)$$

If only dependent $x_0, \ldots, x_{t\text{-}1}$ constraints are satisfied y_t, and not in any "future" input $x_{t+1}, \ldots, x_T$. The learning objective in sequence modeling is to find a network $f : L\,(y_0, \ldots, y_t, f(x_0, \ldots, x_T))$, minimize the expected loss between the actual output and the forecast, where the input sequence and the output sequence are plotted according to some distribution.

2.2 The basic structure of the WATL algorithm network.

The network structure diagram based on the WATL algorithm is shown in Figure 1. It can be clearly seen that the network is divided into three main parts: encoding, decoding and attention. Firstly, the input time series is decomposed by wavelet to obtain the high frequency and low frequency components of the time series. Then, the time convolutional network TCN is used to encode each component, and the encoded information will be used as the input of the next layer of the long and short-term memory network LSTM, which will begin to decode the encoded information according to the label of time series. In the process of decoding, the attention mechanism is introduced to make the current network pay more attention to the features that affect the output of the current time and increase the weight of these features. Finally, wavelet reconstruction is used to predict time series.

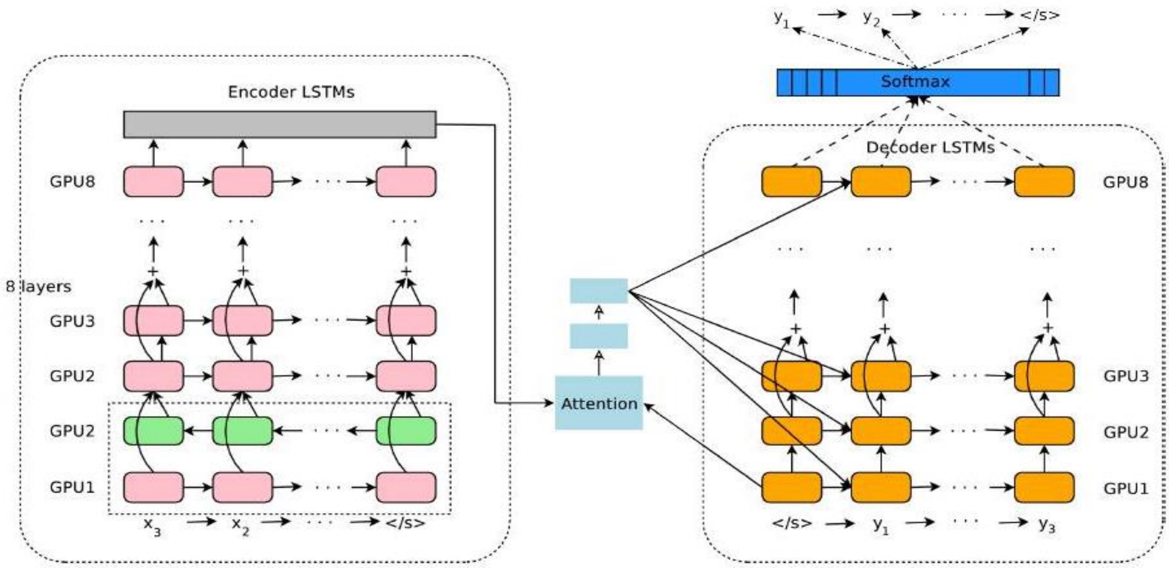

Figure 1. Network structure of walt algorithm

2.3 Watl model fusion based on Encoder-Decoder technology

The structure of Encoder-Decoder is shown in the figure. It is an end-to-end model application framework. It can be clearly seen from the figure that this model framework is divided into Encoder part and Decoder part. The Encoder part is to encode the input time series, and then obtain the hidden state at each moment, and then collect the information and send it to the decoder for decoding.

Let's say I put in a time series $x_1, x_2, \dots, x_T$ Where n represents the length of time $y_1, y_2, \dots, y_T$ is the target value of the corresponding time dimension T. In the time series, the state of the hidden hidden layer at the current moment is jointly determined by the state of the hidden layer at the previous moment and the input at the current moment. There is a mapping relationship between them, and the mapping relationship is assumed to be h, the formula is as follows:

$$h_t = f(h_{t-1}, x_t)$$

h_t is the current hidden layer state, h_{t-1} and x_t represents the input of the state of the hidden layer at the last moment and the current moment respectively. Suppose you have some nonlinear function q in your encoder, the generated semantic encoding after encoding is c, the formula is as follows:

$$c = q(\{f_{h1}, \dots f_{Tx}\})$$

When the calculation at the last moment is completed, the hidden layer state of the previous moment cannot be seen, so the hidden layer state of the last moment is coded as the semantic code, that is:

$$c = h_{Tx}$$

The process of decoding is to predict the output y_t at the next moment based on the given semantic encoding c and the output sequence $y_1, y_2 \dots y_{t-1}$ that has been generated. In fact, it is to decompose the joint probability of the generated time series $y = \{y_1, y_2, \dots y_{t-1}\}$ into the conditional probability in order:

$$p(y) = \prod_{t=1}^{T} p(y_t | \{y_1, y_2, \dots, y_{t-1}\}, c)$$

And each conditional probability can be written as:

$$p(y_t | \{y_1, y_2, \dots, y_{t-1}\}, c) = g(y_{t-1}, s_t, c)$$

Where s_t represents the hidden state of the decoder at the t moment. Formula is as follows:

$$s_t = f(s_{t-1}, y_{t-1}, c_t)$$

Encoder and Decoder are connected end-to-end through intermediate semantic coding. In this mode, we need the encoder to improve the semantic vector of a fixed length when we input the decoder. A fixed length does not fully represent the entire sequence of information, so that the information carried by the first input will be diluted by the information entered later. This becomes more severe when the input sequence is longer. In the final decoding process, sufficient information of the input sequence is not obtained completely, and the decoding accuracy will naturally decrease.

2.4 Attentional mechanism

Attention mechanism plays an important role in today's deep learning, which can be summarized into the following three parts.

1) This kind of attention structure is in line with human's basic cognition of things, just like comparing neural network computing to the brain's information processing. However, unlike neural network, the attention mechanism is not a black box test. The network automatically selects the features to be satisfied and cannot be artificially intervention, so it is highly interpretable.

2) Attention machine allows our model to focus on relevant contents when extracting information, automatically filter invalid information for the model, and establish a dependency relationship between input and output without going through a loop, which enhances the degree of parallelization of the model and greatly improves the running speed of the model.

3) It breaks some limitations in the traditional neural network, such as the system performance decline when a model increases with the length of the input sequence, or the computational efficiency of the model is low when the model disrupts the input sequence, etc.

4) It overcomes the limitation that Encoder-Decoder needs fixed length semantic vectors to establish dependencies. The attention mechanism can model sequence data with variable length well, which further enhances its ability to capture remote dependent information, reduces the depth of hierarchy and improves the accuracy effectively. Its basic frame diagram is shown in Figure 2:

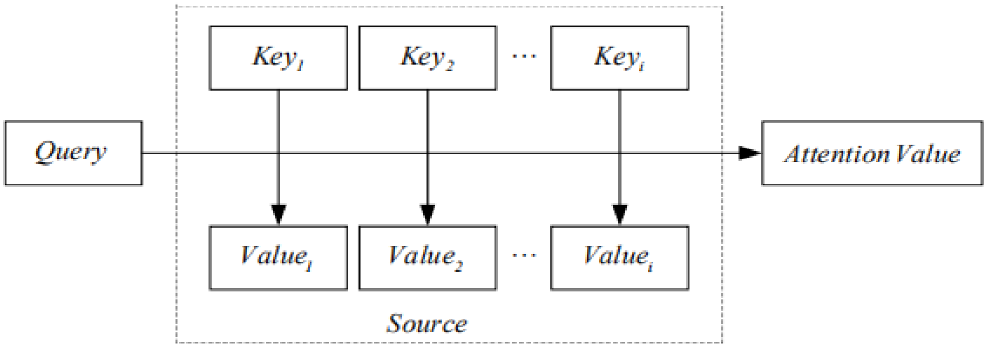

Figure 2. Basic framework

3. Implementation of Watl Algorithm (Oil Chromatography)

3.1 Data set introduction and preprocessing

Oil chromatographic data are generally collected by monitoring devices, but there are abnormal values in the collected oil chromatographic data due to the wide variety of sensors and the external environmental influence in the data transmission process. These outliers need to be preprocessed before they can be converted into the input format required by the model and input into the network. The cleaned data can improve the prediction accuracy of the model because there is not too much noise. Therefore, data preprocessing can ensure the quality of data. In order to lay a foundation for better modeling and analysis in the future, the following data cleaning work is carried out in this paper. Cleaning process as shown:

Invalid value processing: use box diagram to deal with excessive dimension values and outliers generated by sensor faults and abnormal transformer operation. A box chart is a statistical chart used to show the dispersion of a set of data. It is mainly used to reflect the distribution characteristics of original data, and can also compare the distribution characteristics of multiple groups of data. The boxplot is made by finding the upper edge, lower edge, median and two quartiles of a set of data. Then, connect the two quartiles to draw the box; The upper edge and the lower edge are connected with the box body, with the median in the middle of the box body. The box diagram structure is shown in.

1. First of all, we need to get the maximum, minimum, average and quartile of this set of data.

2. Next, draw 5 line segments on the coordinate axis respectively according to the values of the maximum, upper quartile, median, lower quartile and minimum. Connect the two ends of the upper quartile and the lower quartile to form a rectangle, and then connect the upper edge (maximum) and the lower edge (minimum) with vertical lines from the upper and lower ends of the rectangle respectively.

3. Next, calculate the range of moderate and extreme outliers by using the quartile range mentioned above (the upper quartile is Q3=3*(n+1)/4, the lower quartile is Q1=(n+1)/4, and the quartile range is IQR=Q3-Q1).Median Q2=(n+1)/2, the range of mild outliers: the upper limit is 1.5*IQR, the lower limit is Q1-1.5*IQR, and the value within this range is the mild outliers. Range of extreme outliers: the upper limit is Q3+3*IQR, the lower limit is Q1-3*IQR, and the values within this range are extreme outliers.

4. Remove extreme outliers.

Missing value processing: Due to the randomness of missing value, this paper chose to use the mean value of connected 10 days to fill in the missing value caused by the communication device and the above-mentioned artificial deletion.

Data normalization processing: the oil chromatographic data is the characteristic data of the running state of the transformer. When the transformer has a corresponding fault, the corresponding characteristic gas content will be increased. Therefore, there is no unified dimension for the nine characteristic gases. In order to eliminate the influence of convergence brought by the dimension of characteristic gases on model training, and not to change the original distribution of data, Z-score normalization is used in this paper to process the experimental data set. The standardization is as follows:

$$x' = \frac{(x-u)}{\delta}$$

Here is the characteristic gas x, $u = \frac{1}{n}\sum_i^n x_i$, $\delta = \sqrt{\frac{1}{n}\sum_i^n (x_i - u)^2}$, these are the characteristic mean and the characteristic standard deviation.

we can clearly see that there is noise in the historical data of C2H4, and there is an obvious sudden increase near the data of the 500th and the 1000th, which belongs to extreme outliers. In addition, we can see the negative value of the sea patrol in a few parts on this horizontal axis. For this reason, the appellate data cleaning process is adopted to realize data cleaning:

3.2 WATL modeling

Aiming at the problem that the high noise in oil chromatography prediction leads to the general prediction accuracy of the model, the method of wavelet decomposition and reconstruction is added before training the WATL model in this chapter. The low frequency and high frequency components of the original data are extracted by wavelet decomposition. The high frequency component is the detail component, which is also the place where the noise is mainly concentrated. Each component is predicted separately and then accumulated to get the final result, which can effectively solve the problem of noise in the signal. The concrete operation structure diagram of the WATL model is shown in.

The modeling process is as follows:

(1) First, the original oil chromatographic data were preprocessed and divided into training set and test set;

(2) Then the original data of the training set is decomposed into three components by wavelet transform;

(3) The step (2) is decomposed into three sub-sequences, which are one low-frequency sequence and two high-frequency time sequences respectively; The temporal convolutional network is used to extract the characteristic information of time series and realize the encoding.

(4) The encoded information of the three wavelet components output in step (3) is respectively used as the input of LSTM network to realize decoding;

(5) The attention mechanism was added to the output of Step (4) to get the predicted results of the three components, and the predicted values were reverse-normalized, and the three predicted values were added and averaged to get the final predicted value of oil chromatography.

3.3 Evaluation index of WATL model

Different from classification, regression is to predict continuous real values, that is, the output value is continuous real values, so the accuracy rate cannot be used as an evaluation index in the index selection of the model. The commonly used continuous value prediction and evaluation indexes include mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), calibration coefficient (R2), etc. This paper mainly uses MSE, MAE and R2 to measure the performance of the model.

Mean square error refers to the expected value of the square of the difference between the estimated value of the parameter and the true value of the parameter. MSE is a commonly used continuous value prediction evaluation index. The formula of MSE is as follows:

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

Where m is the number of samples, yi is the true label of the sample, and yhat is the prediction label of the sample. Because the mean absolute error has one less square term than the mean square error, it reflects the extent to which differences between individual samples affect the overall error. The smaller the mean square error, the stronger the fitting ability of the model. However, this will lead to a large value of the loss function, leading to the problem of slow convergence of the model. Moreover, the mean absolute error is only the average of the absolute values of the differences between two samples. Do not reflect differences in the true distribution of the data. Therefore, this paper continues to add the calibration coefficient to the above two evaluation indexes, and the formula of the calibration coefficient is as follows.

$$R^2 = 1 - \frac{\sum_{i=0}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=0}^{m}(y_i - \bar{y}_i)^2}$$

y_i is the sample true label, $\hat{y}$ is the sample prediction label, $\bar{y}$ is the mean of the real sample. As a whole, the numerator is the class mean square error, and the denominator is the class variance. The numerator reflects the mean of the population error, and the denominator reflects the fluctuations between real samples. Therefore, the calibration coefficient well reflects the fluctuation of the prediction error in the real sample. It can be seen from the formula that the value of the calibration coefficient is between 0 and 1. The larger the calibration coefficient is, the stronger the fitting generalization ability of the model is.

4. Experience of Result and Comparison

The experimental data set is from the ABC phase DGA monitoring data of a 110kV transformer in Henan from 2011 to 2018, and a total of more than 2,758 cases of data were collected. Some samples are shown in Table 1:

Table 1. Some Example of oil chromatogram

samples	CH4	C2H4	C2H2	CO	O2	N2	C2H6	CO2
1	1.037	1.824	0	542.835	0	0	1.364	0
2	1.185	2.313	0	476.909	0	0	1.742	0
3	1.467	2.781	0	335.126	0	0	2.404	0
4	1.664	2.313	0	288.661	0	0	3.496	0
5	1.425	1.687	0	380.931	0	0	1.383	0

It can be seen from the table that C2H2, O2, N2 and CO2 are all zero, so this paper uses five normal characteristic gases to verify the feasibility of the WATL algorithm.

In order to prove the reliability of the attentional mechanism in WATL, this paper extracted the weight of the attentional machine at the end of the WATL model and used a bar chart to display it, as shown in the figure:

Combined with the correlation between gases in the table, it can be clearly seen from the figure that gas C2H4 itself and CO have the largest weight on the results, satisfying the relationship between the strength of the correlation of gases. It shows that the final output of the model is based on the trend of CO and C2H4, which is in line with the original intention and expected effect of introducing attention mechanism in this paper.

In order to further prove that the attention mechanism can improve the prediction accuracy of the WATL model, this paper sets the output of all characteristic gases to the model under the attention mechanism module of the WATL model as equal weights. When the others remain unchanged, the C2H4 test set is used to verify the results under the parameter of the optimal value, and the effect is shown in the figure 5:

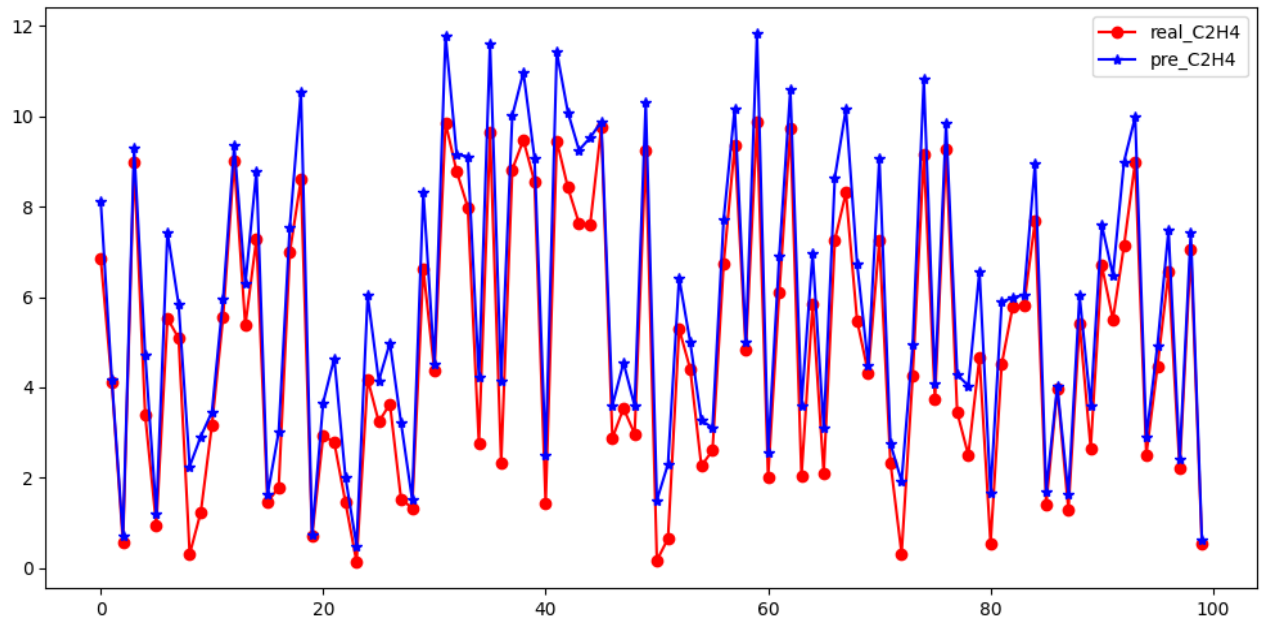

Figure 5. Result Of prediction

It can be seen from the figure that, in the absence of attention mechanism, although the predicted curve roughly fitted the change trend of the real curve, compared with the WATL model, there was a large deviation from the real value in some time steps when the attention mechanism was not added. For this reason, this paper outputs the weight of each attention mechanism in the model, as shown in the figure:

As can be seen from the figure, each input feature has an equal weight of attention to the output.

Table 2. Comparsion results

Evaluation index (%)	model			
	GM(1,1)	ARIMA(2,1,3)	WTL	WATL
mse	2.47	6.47	1.26	0.323
MAE	26.5	25.9	18.7	4.85
R2	69	45	85	96

It can be obtained from Table 2 that the mean square error and absolute error of GM (1,1) model are 2.24% and 26% when the model is predicting C2H4.5. The calibration coefficient is 69%, the mean square error of ARIMA (2,1,3) is 6.47%, the absolute error is 25.9%, and the calibration coefficient is 45%. The mean square error of WTL model without attention mechanism is 1.26%, the absolute error is 18.7%, the calibration coefficient is 85%, and the mean square error of WATL model is 0.323%. The absolute error was 4.85% and the calibration coefficient was 96%. It can be seen that the performance of the three evaluation indexes of WATL is better than that of the other three models, among which ARIMA has the worst performance. It shows that the WATL model proposed in this paper is feasible in predicting C2H4 characteristic gas. In order to further prove that the Watl model is superior to the other three models in the prediction accuracy of C2H4, this paper uses gas C2H4 in the last 10 days as the test set to verify the average difference of the prediction accuracy of the four models. The results are shown in the table

Table 3. Forecast results

time	real(ppl)	GM(1,1)	ARIMA(2,1,3)	WTL	WATL
2018/10/8	6.701	5.631	6.214	6.320	6.657
2018/10/9	5.506	3.570	5.247	6.286	6.035
2018/10/10	7.142	6.521	5.324	6.358	6.987
2018/10/11	8.987	6.832	6.254	6.987	7.687
2018/10/12	5.513	7.586	6.328	7.358	6.934
2018/10/13	4.453	6.186	7.215	6.354	5.681
2018/10/14	6.556	6.358	5.397	5.398	5.987
2018/10/15	6.202	5.673	6.924	6.035	6.346
2018/10/16	7.040	7.873	6.358	6.593	6.834
2018/10/17	5.346	6.257	5.986	6.476	6.548
Mean_diff	D_Value_mean	1.206	1.208	1.059	0.679

As can be seen from Table 3, the average difference of predicted C2H4 values of the four models in the next 10 days can be seen. The average differences of GM (1,1), LSTM and ARIMA (2,1,3) were all above 1, while that of WALTL model was below 1. The validity of the WATL model in the prediction accuracy of C2H4 is further proved.

In order to further prove the effectiveness of the WATL model proposed in this paper on other gases and other gases, four models are respectively used to predict CH4, H2, CO and C2H6. The prediction of four gases is based on the correlation between characteristic gases as shown in the table. For example, H2, CH4, C2H4 and C2H6 have high correlation with them. CH4, C2H6, C2H4 and H2 were used as the characteristic gases for predicting H2, and H2 was used as the prediction gas label. The data of the first 2768 cases were also used as the training set and the last 100 cases as the test. Figure 6 of H2 prediction is shown below, as well as the performance of the four models on the four gas evaluation indexes.

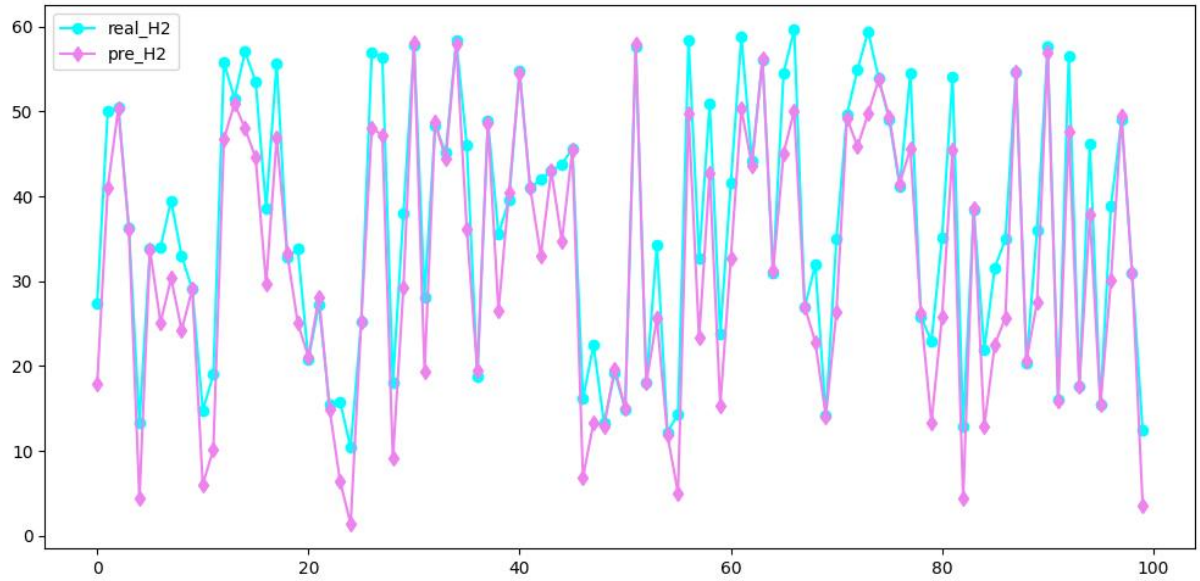

Figure 6. H2 prediction results

Indicators of the four models in H2, CH4, C2H6 and CO respectively, The validity and reliability of the Watl model in oil chromatographic prediction are proved.

5. Conclusion

In this chapter, the design principle of WATL is introduced in detail, and the WATL algorithm is applied to the prediction of oil chromatography, and the prediction model of WATL oil chromatography is built. Then introduces the modeling process of the WATL algorithm. Before the experiment, the section of this chapter also introduces the experimental environment, the configuration of model parameters, and the necessary oil chromatographic data cleaning work. Finally, the prediction of the five characteristic gases of oil chromatography H2, CH4, C2H6, C2H4 and CO was realized. The prediction effects of GM (1,1), ARIMA (2,1,3) and WTL were compared on the basis of mean square error, calibration coefficient and absolute error as evaluation indexes. The effectiveness and reliability of the proposed WATL in oil chromatographic prediction are proved.

References

[1] LIU Xiufeng. Application of GM (1,1) Optimization Model in Prediction of Dissolved Gas Concentration in Transformer Oil [D]. Xihua University,2013.

[2] Liao R J, Zheng H B, Grzybowski S, et al. Fuzzy information granulated particle swarm optimization-support vector machine regression for the trend forecasting of dissolved gases in oil-filled transformers[J]. IET Electric Power Applications, 2011, 5(2):230-237.

[3] Wang Huifang, Zhao Wanfang, Du Zhendong. Economic Life Prediction of Power Transformer Based on Life Data [J]. Power System Technology, 2015,39 (3): 810-816.

[4] ZHAO Wenqing, ZHU Lingyu, GAO Shuguo, et al.Research on Fault Diagnosis Method of Power Transformer Based on Multi-source Information Fusion [J]. Power Information and Communication Technology, 2018, 16 (10): 25-30

[5] CHANG Fangyuan, LI Yuling, LI Erxia, et al. Research and Design of Distribution Transformer Monitoring Terminal Based on SoC Chip [J]. Power Information and Communication Technology, 2018, 16 (10): 60-64

[6] Liao R J, Zheng H B, Grzybowski S, et al. Fuzzy information granulated particle swarm optimization-support vector machine regression for the trend forecasting of dissolved gases in oil-filled transformers[J]. IET Electric Power Applications, 2011,5(2):230-237.

[7] Wang Huifang, Zhao Wanfang, Du Zhendong. Economic life of power transformer based on life data Prediction [J]. Power System Technology, 2015,39 (3): 810-816.

[8] Sun Jianfeng, Ge Rui, Zheng Li, Hu Chaofan. Safety Operation Analysis of State Grid in 2010 [J]. China Electric Power, 2011,44(05):1-4

[9] MA Ke, WANG Yiyu, DONG Yu. Analysis of State Grid Safety Operation in 2009 [J]. China Electric Power, 2010,43(11):1-4.

[10] Luo Zhiqiang, Dong Yu, Hu Chaofan. Safety Operation Analysis of State Grid in 2008 [J]. China Electric Power, 2009, 42(05):8-12.

[11] Matsuura Qiushi et al.Insulation diagnosis technology in operation of power equipment [J]. Technical report of the Japan Electrical Institute, 1992.

[12] Xu Jing, Gao Feng, Wang Jing et al.Power System Technology, 2000,24(8):48-52. (in Chinese)

[13] 'forum. Research on On-line Status Analysis and Intelligent Diagnosis System of Power Transmission and Transformation Equipment [D]. North China Electric Power University,2013.

[14] Sun Caixin, Chen Weigen, Li Jian, et al.Online Monitoring and Fault Diagnosis of Gas in Oil of Electrical Equipment [M]. Beijing: Science Press,2003.)

[15] Jiang Xiuchen, Sheng Ge Hao. Research and Application of Power Equipment State Big Data Analysis [J]. High Voltage Technology, 2018,44 (4) : 1041-1050.