Dynamic Monitoring Method Based on Comparative Study of Power and Environmental Protection Indicators

Xiaojiao Liang[✉], Chunling Ma, Chuanguo Ma, and Shaofei Xin

State Grid Shandong Electric Power Company Dongying Power Supply Company, Dongying, China
bendideren@163.com

Abstract. In order to promote my country's comprehensive green development, this paper conducts a research on the correlation mapping between electricity consumption and environmental protection data. Based on the existing short-term electricity load forecasting model, the electricity load forecast data is divided into active electricity and reactive electricity, so as to realize the comprehensive tracking of the operation status of the enterprise's electrical equipment. The comparative learning algorithm is used to obtain the association mapping of the electricity consumption-environmental data distribution, so that the Encoder data of the pollution discharge data can be obtained based on the electricity consumption. Subsequently, the generator generates enterprise short-term emission forecast data based on the coded data, and combines the Wasserstein distance with the mean square error of similar samples to construct a loss function to improve the quality of the generated data. Analysis and comparison of calculation examples through simulation experiments has show the effectiveness and feasibility of the proposed.

Keywords: Electricity load · Environmental protection index · Comparative study · GAN · Monitoring warning

1 Introduction

At present, the environmental protection testing index data of enterprises has the characteristics of many points, wide area, and long line. Relying on the inspection and supervision of the national law enforcement department can often only make up for it. Moreover, the single monitoring method inevitably has the occurrence of fraudulent monitoring data and destruction of monitoring equipment. As an essential energy in the production activities of enterprises [1], which has mirror image and high coverage, which can timely, accurately and comprehensively reflect the production status of enterprises and the use of environmental protection equipment [2]. Accurate short-term pollution forecast can assist environmental protection departments to formulate response strategies in advance to avoid the occurrence of severe events such as smog and acid rain. Due to the inevitable correlation between input energy and production, this paper analyzes

the correlation between electricity consumption data and environmental protection indicators of key pollutant discharge monitoring enterprises. Through comparison mapping and data fitting, the short-term forecast of enterprise pollutant discharge can be realized.

At present, relevant institutions have begun to pay attention to the prediction of enterprise pollution discharge based on enterprise electricity consumption. However, the research content is relatively small. Paper [3] uses regression analysis method to build an environmental impact prediction model based on industry electricity consumption, and realizes the monitoring and early warning of air pollution prevention and control for key enterprises. In addition, Paper [4] studies air pollution prevention and audit methods based on big data visualization technology, and analyzes the relationship between enterprise electricity consumption data and air pollution; Paper [5] analyzes the impact of air pollution on short-term electricity consumption on the demand side. Paper [6] adjusts the power generation control strategy according to the real-time monitoring data of pollution emissions, thereby reducing the amount of air pollution emissions. The above literatures all show that there is a correlation between electricity consumption data and environmental protection data. However, no effective mapping relationship between electricity consumption and environmental protection has been established. At present, relatively accurate short-term forecasts of electricity consumption data can be achieved, and short-term forecasts of pollutant emissions can be completely based on this. Based on the existing research foundation and technology, the article carries out short-term forecast of enterprise pollutant discharge. The main improvements are as follows:

(1) Industrial electricity involves rotating high-power electrical appliances such as motors, so production equipment will generate reactive power. Even with compensating devices, there will be power fluctuations [7]. Therefore, in order to measure the operation status and production volume of production equipment, the electricity load is divided into reactive power and active power.

(2) In order to maximize the correlation analysis of electricity consumption and environmental protection characteristics, the article adopts comparative learning [8] to realize the correlation mapping between electricity consumption and environmental protection features. By zooming in similar instances in the projected space and pushing away dissimilar instances, the distance distribution of mapping codes preserves the differentiated pollutant discharge information of enterprises. Even when enterprises lack sample data, they can make reasonable predictions based on similar samples and their own characteristics.

(3) In order to improve the global similarity and consistency between the predicted data and the actual data, the Wassersein distance is used when the predicted data is generated based on the mapped coded data combined with the mean square error of similar samples, the loss function of the generator is constructed, and the data distribution of the generated curve is further drawn closer to the real data, effectively avoiding the distortion of the predicted data.

2   Short-term Emissions Forecasting Process

This paper obtains electricity consumption and environmental protection monitoring data based on related systems, and uses web crawling technology to obtain data on weather,

holidays and emission reduction policies. After all kinds of data are preprocessed, similar enterprises are classified [9]. The overall data processing flow as follows (Fig. 1):
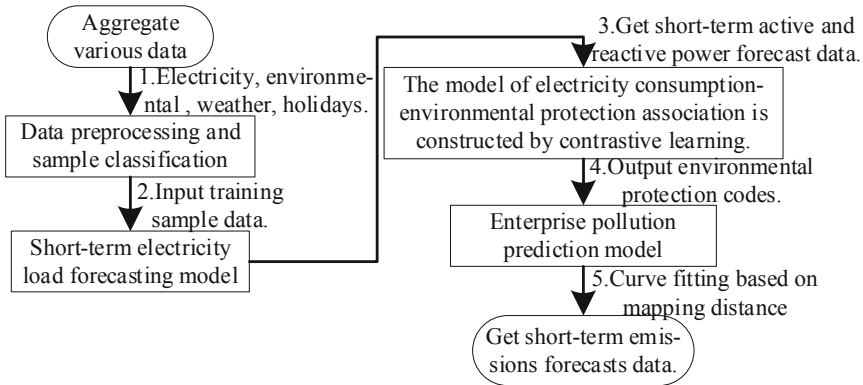

Fig. 1. Short-term emission forecast

First, in order to obtain the short-term active and reactive power forecast data, the short-term electricity load prediction model based on two-layer LSTM [10] processes the input data which includes active power, reactive power, meteorological data, and holiday data. Subsequently, this paper uses the comparative learning algorithm to realize the correlation mapping between electricity load and environmental protection index data, so as to construct a similar mapping model of electricity consumption-environmental protection correlation. Contrastive learning can achieve differentiated feature extraction based on data differences between enterprises. After the above prediction data is input into the model, the encoder of the pollutant discharge prediction data is output. Finally, the encoder and the recent actual sewage monitoring data of the enterprise are input into the enterprise sewage prediction model, which uses the combination of Wasserstein distance and similar historical sample data as the loss function. The model can predict the short-term pollutant discharge data of enterprises based on historical electricity consumption data and pollutant discharge data.

2.1   Building a Short-Term Electricity Load Forecasting Model

In this paper, the LSTM algorithm is used to build a short-term electricity load prediction model. In order to effectively track the operation status of enterprise equipment, the historical electricity load data of the enterprise is divided into active power and reactive power data. The model training data set contains enterprise history active power, reactive power, data, the temperature data, holiday data, The output is the predicted value of short-term active and reactive power. Take the company's active power and reactive power data at time t, combined with temperature data, holiday data, and emission reduction policy data (such as corporate pollution exceeding the standard and being restricted from running) as the data input of the LSTM model, and the input layer is 100 memory neurons.

The predicted output value of the active power and reactive power of the previous neural unit is h_{t-1}, and each neuron input has a corresponding weight. After sigmoid activation, a value of 0–1 is obtained, that is, three gate values. The calculation formula as follows:

$$
\begin{aligned}
q_t &= \sigma(W_q x_t' + U_q h_{t-1}) \\
r_t &= \sigma(W_r x_t' + U_r h_{t-1}) \\
o_t &= \sigma(W_o x_t' + U_o h_{t-1})
\end{aligned}
\tag{1}
$$

where the output value of the input gate is q_t, the output value of the forget gate is r_t, and the output value of the output gate is o_t, W_q, U_q, W_r, U_r, W_o, U_o are the corresponding parameters of each gate.

The input value x_t' and the output h_{t-1} of the previous unit have corresponding weights W_c and U_c, and tanh activation function is used. The output value is equivalent to obtaining the new memory value $\tilde{c}_t$ obtained based on the input value at time t. The calculation formula is as follows:

$$
\tilde{c}_t = \tanh(W_c x_t' + U_c h_{t-1})
\tag{2}
$$

$$
c_t = r_t \circ c_{t-1} + q_t \circ \tilde{c}_t
\tag{3}
$$

where r_t is the output value of the forget gate at time t, q_t is the output value of the input gate at time t, and c_{t-1} is the final memory value of the previous neuron at time t-1. The final output value of h_t is

$$
h_t = o_t \circ \tanh(c_t)
\tag{4}
$$

The hidden layer adopts the same single-layer structure as the input layer, the output layer is a two-dimensional output, and the output data is the predicted value of the short-term enterprise active power and reactive power $\hat{x}_t$.

2.2 Constructing a Similarity Mapping Model for Electricity Consumption and Environmental Protection

The historical active power, reactive power data and historical sewage monitoring data of the enterprise constitute training data set τ, and M sample data of enterprises are randomly selected from the training data set to form a branch training set A, which generates a total of two branches that act on the upper and lower branches. The two training sets are the historical sewage monitoring data set A_1, the historical active power data and the historical reactive power data A_2. Each branch contains the sample data of M enterprises, the sample data of the upper branch is fixed, while the data of the lower branch is dynamically updated. Randomly extract a certain type of historical pollutant discharge monitoring data $P^t = \left[p_1^t, p_2^t, \cdots, p_n^t\right]$ of a certain enterprise and input it into the upper branch, and input the active power and reactive power data at the corresponding time into the lower branch.

$$
X^t = \begin{bmatrix} x_1^t & x_2^t & \cdots & x_n^t \end{bmatrix} = \begin{bmatrix} x_{A1}^t, x_{A2}^t, \cdots, x_{An}^t \\ x_{R1}^t, x_{R2}^t, \cdots, x_{Rn}^t \end{bmatrix}
\tag{5}
$$

where t represents the daily sampling time point p_1^t, x_{A1}^t and x_{R1}^t are the same dimension column vector of the sampling sequence, n represents the daily monitoring sampling number of a certain enterprise, p_i^t and x_i^t are positive examples of each other, $i, j \in (1, 2, \cdots, n)$. During training, any other input sample data p_j and x_j of the same enterprise or different enterprises in the upper branch and the lower branch are negative examples of p_i^t and x_i^t.

Subsequently, a representation learning architecture is constructed by which the training data is projected into a hyperplane representation space. In order to shorten the distance of positive examples and push the distance of negative examples farther, the upper and lower branches adopt a two-tower asymmetric structure, and the internal levels and the number of neurons are different. Historical sewage monitoring data set A_1, and the data enters the upper branch, through the feature Encoder (using the Transform encoder Encoder structure), the Encoder module has a total of 4 layers, thereby mapping the input sample data to a hyperplane Represents the vector z_p in the space. The architecture of the first part of the lower branch is the same as that of the upper branch, but the number of neurons in the middle layers is different, and the parameters are not shared. The output is represented as $h_x = f_{\theta'}(x)$, and then input the nonlinear transformation structure Projector, the output is $z_x = g_{\theta'}(h_x)$. The matrix internal structure of vectors z_x and z_p is consistent with the quantity (Fig. 2).
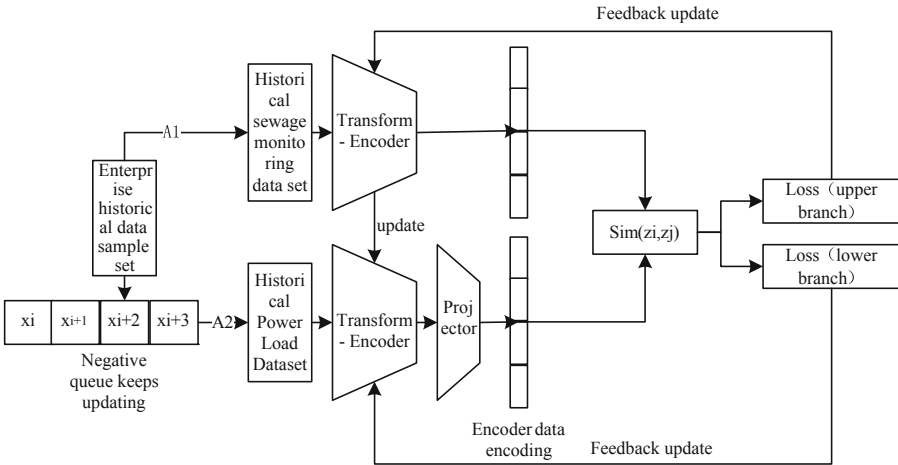

Fig. 2. Training process of electricity-environmental protection correlation mapping.

In the representation space, in order to realize the correlation analysis between the electricity data and the environmental protection indicators, it is necessary to make the representation vector z after the mapping of the positive electricity data and the pollutant discharge monitoring data overlap or be as close as possible, and the distance from the negative sample is farther away. Therefore, L2 is used as the distance metric to represent

the similarity between space vectors $S(z_i, z_j)$:

$$S(z_i, z_j) = \frac{z_i^T z_j}{\|z_i\|_2 \|z_j\|_2} \tag{6}$$

In order to make the distance between positive examples close and the distance between any negative examples far in the representation space, the following loss function is used. Upper branch loss function:

$$loss = -\log \frac{e^{S(z_{xi}, z_{pj})/\tau}}{\sum\limits_{j=1, i\neq j}^{M \times n} e^{S(z_{pi}, z_{pj})/\tau}} \tag{7}$$

Lower branch loss function:

$$loss = -\log \frac{e^{S(z_{xi}, z_{pj})/\tau}}{\sum\limits_{j=1, i\neq j}^{M \times n} e^{S(z_{xi}, z_{xj})/\tau}} \tag{8}$$

where τ is the temperature hyperparameter, which represents the difficulty recognition penalty coefficient, which is used to widen the distance between positive samples and negative samples with high similarity, so as to avoid model collapse due to parameter adjustment based on high-difficulty negative samples. $S\left(z_{x_i}, z_{p_i}\right)$ represents the similarity between positive samples, that is, the similarity between the electricity data samples and the pollutant discharge monitoring data samples of an enterprise in the same day, $S\left(z_{P_i}, z_{p_j}\right)$ represents the similarity between positive and negative samples of sewage monitoring, $S\left(z_{x_i}, z_{x_j}\right)$ represents the similarity between positive and negative samples of electricity data, that is, the positive electricity data samples and other negative electricity data samples of an enterprise in the same time period, or The similarity between the positive and other negative pollution monitoring data samples. The smaller the objective loss function, the better. Therefore, the larger the numerator, the greater the similarity between positive examples. While sample data of different enterprises or different time periods are similar in spatial mapping. so as to maintain the individual differences of the sample data while realizing the association analysis.

2.3   Construction of Enterprise Pollution Prediction Model

In the training stage of the enterprise pollutant discharge prediction model, if the similarity between the Encoder data of the input sample and the Encoder data of the training sample is within a certain threshold range, it means that the similarity between the two is high, that is, the similar sample data exists in the historical pollutant discharge data sample. Therefore, in the process of data fitting, the influence of the weight of similar samples increases, and the pollutant discharge monitoring data is more inclined to fit the pollutant discharge forecast data with similar samples; if the similarity exceeds the

threshold, it means that the existing enterprise pollutant discharge samples are all different from the input samples, so the loss is The function is more inclined to generate pollution prediction data through the generator based on the encoder output encoding of electricity. The training process for this part is as follows (Fig. 3):
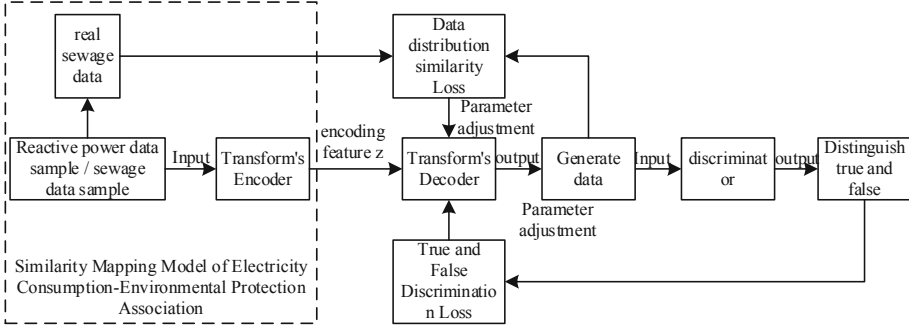

Fig. 3. Training process for generating pollutant emission prediction data.

In order to measure the distance between the prediction and the real data, this paper uses the Wasserstein distance as a component of the loss function. Wasserstein distance is used to measure the distance between two probability distributions. That is, the shortest distance to move date from distribution 1 to distribution 2. The advantage over KL divergence and JS divergence is that the distance can still be measured when the two data distributions have no or negligible overlap. At this time, the JS divergence is constant, and the KL divergence is meaningless. The Wasserstein distance is calculated the loss function as follows:

$$loss = E\left(\underset{Encoder}{var}(p_1, p_2, \cdots, p_h, \tilde{p})\right)$$
$$+ \ell \underset{\tilde{p} \sim P_g, p \sim P_r}{E}\left[D(p, \tilde{p})\right]$$
$$+ \lambda \underset{\hat{p} \sim P_{\hat{p}}}{E}\left[\left(\left\|\nabla_{\hat{p}} D(p, \hat{p})\right\|_2 - 1\right)^2\right] \tag{9}$$

where P_g is the set of generated data, P_r is the set of actual data, $\tilde{p}$ is the generated data sample, p is the actual data sample, and $p_1, p_2, \cdots, p_h$ is the similar sample encoded by the Encoder of the actual data sample p, and represents the sample cluster set with high similarity of historical pollutant discharge data. It can be determined by setting the threshold. $\underset{Encoder}{var}(p_1, p_2, \cdots, p_h, \tilde{p})$ represents the mean square error between the generated data and the historical pollutant discharge data of similar enterprises. By minimizing this value, the approximate data fitting generation of the pollutant discharge data of similar enterprises is realized. ℓ is the weight data generated based on the data volume b of similar enterprises. $\ell = a/e^b$, a is a custom parameter. When there is a lot of similar enterprise data, the value of ℓ decreases. At this time, the generated data is

more inclined to rely on the pollutant discharge data of similar enterprises. When there is little or no similar enterprise data, the value of ℓ increases, and the generated data depends on the generator. $\lambda \mathop{E}\limits_{\hat{p}\sim P_{\hat{p}}} \left[\left(\left\| \nabla_{\hat{p}} D(p, \hat{p}) \right\|_2 - 1 \right)^2 \right]$ is the weight gradient penalty, which is used to ensure that the generated data is close to the real sample data, but does not exceed the real sample data, $\left\| \nabla_{\hat{p}} D(p, \hat{p}) \right\|_2$ is the gradient penalty coefficient, and $\hat{p}$ is the sampling composition data between the generated data and the real data. Gradient penalty to stabilize the gradient change, $\hat{p} = \varepsilon p + (1 - \varepsilon)\tilde{p}$, λ is the weight; $\mathop{E}\limits_{\tilde{p}\sim P_g, p \sim P_\gamma} \left[D(p, \tilde{p}) \right]$ is the average Wasserstein distance between all generated data in the set and its corresponding actual data samples, $D(p, \tilde{p})$ is the Wasserstein distance between the generated samples and the actual samples, the calculation formula is as follows:

$$D(p, \tilde{p}) = \min_{\sigma \in \sum_N} \left(\sum_i \left\| p_i - \tilde{p}_{\sigma(i)} \right\|^2 \right) \tag{10}$$

where p_i represents the actual sewage monitoring data at time t, $\tilde{p}$ is the generated data, $\tilde{p}_{\sigma(t)}$ is the real data at time t mapped to the generated sewage monitoring data at time $\sigma(t)$ when the data was generated, since the actual data and the generated data have the same number of elements, and the data measures are the same, so the Wasserstein distance is measured by the quadratic difference, and $\sum_N$ represents the set formed by all the sorted arrangements of N elements.

The training process is as follows:

(1) Initialize generator parameters θ_g and discriminators θ_d;
(2) In the iterative training process, the generator is trained once, and the discriminator is trained several times. The process is as follows:

Train the discriminator, fix the generator parameters, and update the maximizing discriminator $loss_d$:

$$\begin{aligned} loss_d = &E\left(\mathop{\text{var}}\limits_{Encoder} (p_1, p_2, \cdots, p_h, \tilde{p}) \right) \\ &+ \ell \mathop{E}\limits_{\tilde{p}\sim P_g, p \sim P_r} \left[D(p, \tilde{p}) \right] \\ &+ \lambda \mathop{E}\limits_{\hat{p}\sim P_{\hat{p}}} \left[\left(\left\| \nabla_{\hat{p}} D(p, \hat{p}) \right\|_2 - 1 \right)^2 \right] \end{aligned} \tag{11}$$

$$\theta_d \leftarrow \theta_d + \eta \nabla loss_d \tag{12}$$

where η is the learning rate.I In order to realize the curve fitting process based on similar sample data, the mean square error discrimination condition of historical similar samples is added. Train the generator, fix the discriminator parameters, and update the minimized generator $loss_g$:

$$loss_g = \mathop{E}\limits_{\tilde{p}\sim p_g, p \sim p_r} \left[D(p, \tilde{p}) \right] \tag{13}$$

$$\theta_g \leftarrow \theta_g - \eta \nabla loss_g \qquad (14)$$

After the training is completed, the electricity consumption-environmental protection correlation similarity mapping model inputs the electricity data and some pollutant discharge sampling point data, outputs the Encoder code. These codes into the enterprise pollutant discharge prediction model, and outputs the enterprise pollutant discharge forecast data.

3   Case Analysis

3.1   Sample Data Preprocessing

The data collection frequency of enterprise power consumption data and environmental protection data is 24 points/day. The power collection data is the historical active power data and reactive power data obtained based on the multi-function energy meter. Reactive power is the power consumption generated by converting electrical energy into capacitance or inductance during the generation of enterprise operating equipment, including forward reactive + reverse reactive power. The environmental monitoring indicator is the nitrogen oxide concentration in the air pollution indicator. Data preprocessing is performed on the acquired data. The 2385 key sewage monitoring enterprises in the province are classified and modeled based on the industry, and the historical data from January 2019 to December 2020 is selected as the model training sample set, and the historical data from January 2021 to May 2021 is used as the model. To verify the sample set, in order to avoid the imbalance of sample categories, only 2/5 of the historical samples of the normal production of the enterprise are selected for training and verification, and the historical data in mid-June 2021 is used as the model test data sample set to realize the enterprise within a short period of one week. Evaluating the effect of pollutant discharge forecasting.

3.2   Data Spatial Distribution Visualization

In the training phase, after the active, reactive power and NOx data are input into the electricity consumption-environmental protection correlation similarity mapping model, the Encoder code of the enterprise pollution prediction model generator is obtained from the upper and lower branches. In order to visualize the spatial distribution law of NOx data extracted by the comparative learning algorithm, that is, on the premise of retaining the detailed information of enterprise sewage, the clustering and classification of sample data are realized, and the representation space is relatively uniformly distributed, and electricity data is used as an enhanced sample data, through comparative learning to achieve the fitting of the distribution law of nitrogen oxide emission monitoring data in the same period.

The article uses a nonlinear dimensionality reduction algorithm, that is, t-distributed random neighbor embedding to reduce the 6-dimensional Encoder data to three-dimensional space, and uses matlab to correlate the Encoder codes generated by the NOx data and electricity consumption data in the three-dimensional space, as shown in the following figure:
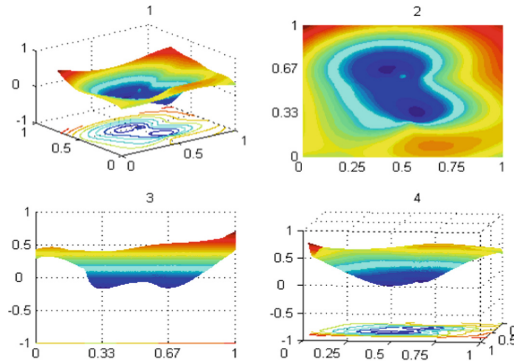
Fig. 4. 3D view of encoder data.

In Fig. 4, view 1 is the azimuth –37.5°, the elevation 30°. View 2 is the azimuth 0°, the elevation 90°. View 3 is the azimuth 90°, the elevation 0°. View 4 is the azimuth –7°, elevation –10°. The upper branch obtains the Encoder of the enterprise pollutant discharge prediction model generator, which is located in the upper area of view 2 and the left area of view 3. The lower branch obtains the Encoder of the enterprise pollutant discharge prediction model generator based on the active and reactive power data. The encoding is located in the lower area of view 2 and the right area of view 3. Figure 4 shows the symmetry and correlation of the Encoder code generated by the "twin-tower" structure from different perspectives. Since only 829 key pollutant discharge enterprises can realize daily monitoring of nitrogen oxides, the remaining 1,556 key pollutant discharge monitoring enterprises have upper branches. The NOx data samples are missing, limited by the limited amount of NOx data, which leads to deviations in the code of the lower branch fitting the sewage data based on the electricity consumption data, as shown in the yellow and red data areas in Fig. 4, this part of the area of the average nitrogen oxide emission index per unit of electricity of enterprises is lower than 5.89, and most of them are manufacturing enterprises. Therefore, the daily monitoring data is insufficient, so without corresponding pollution data mapping. Figure 4 shows that through comparative learning, the Encoder data distribution of the pollutant discharge index data can be extracted based on the power consumption data of the enterprise, so as to provide the recurring code of the pollutant discharge data for the enterprise pollutant discharge prediction model generator.

3.3   Comparative Analysis of Results

In the process of model data processing, the article mainly improves from the following three points:

(1)   Divide the power load data into active and reactive power data.
(2)   The input code of the enterprise pollutant discharge prediction model generator is obtained by comparing the electricity consumption data with NOx data.
(3)   The loss function uses the Wasserstein distance and combines the dynamic weights to adjust the influence of similar samples to achieve curve fitting.
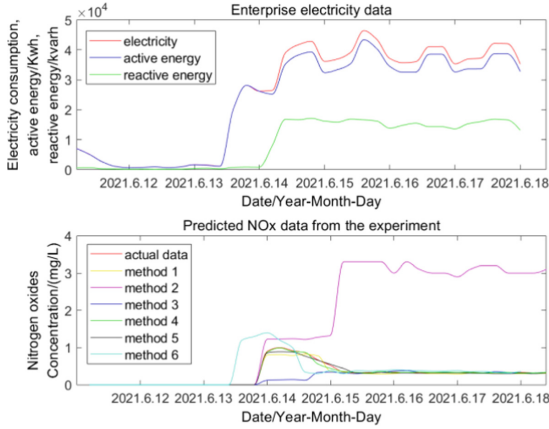
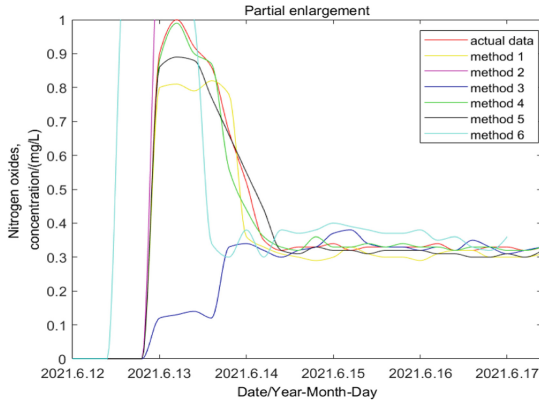**Fig. 5.** Comparison chart of experimental results.

Fig. 6. Partial enlarged view of forecast data.

As shown in Fig. 5 and Fig. 6, after the electricity consumption-environmental protection correlation similarity mapping model and enterprise pollution prediction model are trained, the electricity consumption data is input into the electricity consumption-environmental protection correlation similarity mapping model, and the Encoder data is output, and then the Encoder data is input. Enterprise pollutant discharge forecast model, output forecast data of enterprise nitrogen oxide discharge. The example enterprise is a steel product processing enterprise, so it is less affected by external weather and holidays. The enterprise is based on the production plan. Therefore, the short-term electricity consumption forecast data is basically consistent with the actual electricity consumption data, so the actual electricity consumption is not shown in the experiment. Minor deviation of data from predicted electricity consumption data. The comparison experiment method is as follows:

Method 1: without comparative learning. Under the condition of sufficient NOx training data, the active and reactive power data as model input, and the experimental model uses the VAE-GAN method;

Method 2: Unlike method 1, NOx sample is insufficient;

Method 3: The NOx sample is sufficient, electricity load as the input data;

Method 4: The NOx sample is sufficient, the method in the text is adopted;

Method 5: The NOx sample is insufficient, the method in the text is adopted;

Method 6: The NOx sample is sufficient, Wassersein distance replaced by kl.

As shown in Fig. 5 and Fig. 6, the method 4 and method 5 using the method in this paper have a high similarity to the curve fitting of the actual monitoring data, especially when the method 4 has sufficient training data for nitrogen oxide monitoring, the fitting best effect; Method 1 does not use comparative learning, but adopts the VAE-GAN method. The active and reactive electricity data is used as the model input, and the NOx data is used as the learning and reproduction data for training. Under the condition of sufficient training samples, the effect is also relatively Ideal, but in the case of insufficient NOx training sample data, as shown in Method 2, fitting prediction will be made based on the emission data of enterprises with similar power consumption scales, rather than the comparison of electricity consumption and environmental protection indicators in the article Learning to obtain the similarity code, generating a large deviation in the prediction results based on the similarity of electricity consumption only, and failing to capture the unique information of the enterprise's own sewage; Method 3 is to use the electricity load instead of the active and reactive power data in the text as the input of the model, as shown in the Fig. 5, at 8:00 on June 13, 2021, the enterprise enters the production and operation state, and the simulation demonstrates the decontamination The equipment is not turned on, causing the actual nitrogen oxide index to exceed the standard. The reactive power data can better reflect this imagination, while the overall power load data cannot achieve effective prediction based on the operation of the equipment; Method 6 uses the kl divergence method for the loss function. Although the kl divergence can reflect the distribution of the data well, it is difficult to track the lag characteristics of the sewage data, and the model training process takes a long time and the parameter adjustment is difficult, so it is easy to generate The phenomenon that the gradient is zero, as shown in Fig. 6, the fitting effect of the model after the training of method 6 has a certain deviation, and because the KL divergence distance is difficult to measure the lag of the data, the KL divergence data in Table 1 not counted.

This paper selects 295 key polluting enterprises for 7-day nitrogen oxide emission concentration prediction, uses RMSE (Root Mean Squared Error) root mean square error (normalized mean), and evaluates the effect of emission concentration prediction on the results of 12 experiments. The KL divergence is used to measure the information loss of a predicted data distribution compared to the actual data distribution. The smaller the data, the higher the curve fitting degree, and the fitting curve is integrated to obtain the forecast of the cumulative pollutant discharge of the enterprise. The prediction accuracy of the predicted value and the actual pollutant discharge is evaluated by the mean absolute percentage error (MAPE, Mean Absolute Percentage Error), as shown in the following Table 1:

Table 1. Curve similarity evaluation table

Method name	RMSE	KL divergence	MAPE
Method 1	3.08%	$2.34*10{-2}$	10.65%
Method 2	68.30%	1.58	462.65%
Method 3	10.68%	1.43	56.33%
Method 4	1.03%	$7.8*10{-3}$	4.56%
Method 5	1.74%	$9.88*10{-3}$	6.73%
Method 6	15.17%	/	78.82%

According to the RMSE and KL divergence, the curve fitting of the predicted value obtained by the method 4 and method 5 described in this paper is the best. A MAPE of 0% indicates a perfect model, and a MAPE greater than 100% indicates an inferior model. Table 1 shows that Method 4 and Method 5 can predict relatively accurate corporate pollutant emissions. In summary, the method in this paper can effectively realize the prediction of pollutant discharge monitoring data based on electricity consumption data.

In order to further verify the impact of the similar enterprise sample data in the comparative learning and loss function on the prediction model. In this paper, under the assumption that an enterprise lacks NOx samples, the effect of the number of similar enterprises on generating forecast data is tested. The L1 distance and mean error are obtained by comparing with the real data, as shown in the following figure (Fig. 7):
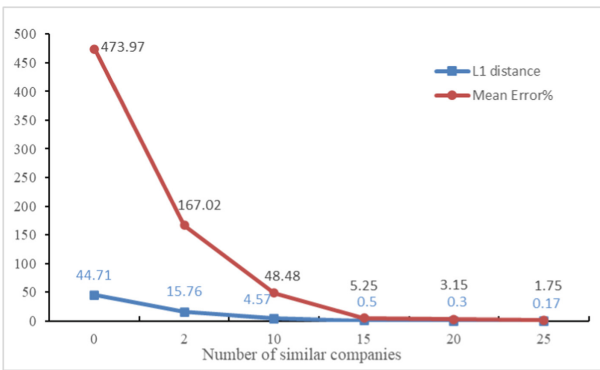
Fig. 7. Similar enterprise data affects the generation of forecast data.

The above figure shows that in the absence of a company's NOx samples and no similar companies, that is, only based on electricity consumption data as input and based on GAN to generate pollution forecast data, the L1 distance and mean error are both bad. However, when the number of similar enterprises is 2, both the L1 distance and the mean error decrease rapidly. With the increase of similar enterprises, the mapping between the sample data in the comparative learning process is more dense, so that the

encoded data can be generated based on the nearest similar samples. The sample weights of similar enterprises in the loss function can assist the model to generate data on the basis of similar enterprise data, and further ensure the stability of the data generated by GAN.

4   Conclusion

Through the comparative learning algorithm, the correlation mapping between electricity consumption data and environmental protection indicators is realized., the lower branch electric power data of the positive example enhanced sample of the upper branch environmental protection data fits the distribution of the corresponding positive example mapping data, and realizes that the positive examples attract and the negative examples repel each other, and there is no corresponding upper branch positive example or the electricity data samples with few positive examples of sewage discharge in the upper branch are relatively evenly distributed based on the similarity and distance of environmental protection sampling data and electricity data between enterprises, so as to ensure that when the training samples of pollution discharge data monitoring are missing, similar enterprise data can be obtained based on the electricity consumption-environmental protection correlation mapping, and the pollutant discharge data can be fitted by the generated data and the similar enterprise data to generate the short-term pollutant discharge forecast data of the enterprise.

Since it is still in the initial stage of exploration to predict the pollutant discharge of enterprises based on electricity consumption data. And some companies have the problem of missing environmental protection data, which affects the stability of the generated data by GAN. In the future, with the increase of environmental monitoring data of various enterprises, the distribution of similar sample data will be more dense, which can further improve the quality of generated data. In addition, this paper only predicts the nitrogen oxide indicator data. With the increase of point source emission monitoring data, the prediction of multiple indicators can be explored, so as to provide technical research reference for environmental monitoring.

References

1. Du, Y., Yang, X., Guo, L., et al.: Coupling evaluation of green development and safety and efficiency of distribution network incontext of double carbon. Sci. Technol. Eng. **21**(30), 12973–12981 (2021)
2. Matindife, L., Sun, Y., Wang, Z.: Disaggregated power system signal recognition using capsule network. In: Zhang, H., Zhang, Z., Wu, Z., Hao, T. (eds.) NCAA 2020. CCIS, vol. 1265, pp. 345–356. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-7670-6_29
3. Sun, K., Chen, Z., Fan, M., et al.: Design of special action plan for pollution prevention and control of key enterprises based on electric power big data mining. Distrib. Utilization **38**(04), 28–36 (2021)
4. Chen, W., Gao, J.: Research on air pollution prevention audit method based on big data visualization analysis technology. Finan. Acc. **04**(07), 65–68 (2019)
5. He, Z., Zhang, T., Hu, Y., et al.: Short-term electric load forecasting model considering the influence of air pollution prevention and control policy. Smart Power **47**(05), 1–9 (2019)

6. An, J., Chen, Q., Dai, F., et al.: Power green dispatch strategies for air pollution prevention and control. Power Syst. Technol. **45**(02), 605–612 (2021)
7. Babuta, A., Gupta, B., Kumar, A., et al.: Power and energy measurement devices: a review, comparison, discussion, and the future of research. Measurement **172**, 108961 (2021)
8. Ting, C., Simon, K., Mohammad, N., et al: A simple framework for contrastive learning of visual representations. In: Proceeding of the 37th International Conference on Machine Learning. Cambridge MA: JMLR, pp. 1597–1607 (2020)
9. Huang, S., Kang, Z., Xu, Z.: Deep K-Means: a simple and effective method for data clustering. In: Neural Computing for Advanced Applications. NCAA 2020. Communications in Computer and Information Science, vol.1265, pp. 272–283 (2020) https://doi.org/10.1007/978-981-15-7670-6_23
10. Xu, Y., Chhim, L., Zheng, B., Nojima, Y.: Stacked deep learning structure with bidirectional long-short term memory for stock market prediction. In: Zhang, H., Zhang, Z., Wu, Z., Hao, T. (eds.) NCAA 2020. CCIS, vol. 1265, pp. 447–460. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-7670-6_37