Received July 18, 2017, accepted August 7, 2017, date of publication August 24, 2017, date of current version November 14, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2740968

Cleaning Method for Status Monitoring Data of Power Equipment Based on Stacked Denoising Autoencoders

JIEJIE DAI, HUI SONG, GEHAO SHENG, (Member, IEEE), AND XIUCHEN JIANG
Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Corresponding author: Jiejie Dai (secess@163.com)

This work was supported in part by The National Natural Science Foundation of China under Grant 51477100, in part by the National High Technology Research and Development Program (863 Program) under Grant 2015AA050204, and in part by the State Grid Science and Technology Program.

ABSTRACT Currently, the cleaning process for power equipment monitoring data is cumbersome and often leads to loss of information. To address these problems, a data cleaning method based on stacked denoising autoencoder (SDAE) networks is proposed in this paper. SDAE networks have a strong ability to denoise and restore corrupted data and have a strong feature extraction capability. The status monitoring data of equipment under normal conditions are trained by SDAE to obtain the cleaning parameters and the reconstruction errors. An upper threshold of the reconstruction errors obtained from training samples is determined through Kernel density estimation. A tolerance window width is added to achieve rapid anomaly detection. The abnormal data are classified as outliers, missing data, or fault status data according to the relationship between the reconstruction error and the threshold and between the duration of abnormal data and the tolerance window. To verify the effectiveness of the proposed method, the SDAE model is used to process the data for the dissolved gas concentration in transformer oil and the temperature of the transmission line. The results show that the proposed method can effectively identify and repair outliers and missing information. The model can perform rapid anomaly detection when the equipment is running abnormally.

INDEX TERMS Power equipment, status data, data cleaning, stacked denoising autoencoders, anomaly detection.

I. INTRODUCTION

With improvements in information technology of the power industry, the status monitoring data for power equipment have shown a massive growth trend. These data provide information used for equipment status assessment, fault analysis and prediction. However, the operating environments of transformers, transmission lines, gas insulated switchgears (GIS) and other power equipment are complex and diverse [1], [2], [3]. The original monitoring data consist of partial abnormal points because of environmental interference and the limitations of measurement techniques. Abnormal data can also be caused by equipment defects [4]. These data contain important connotative information and cannot be treated the same as ordinary noise or missing data. It is therefore highly desirable to classify and clean abnormal data [5], [6].

Currently, various methods are used to clean and repair power data. Neighborhood-based techniques, such as moving average and smoothing techniques, have been successfully used to detect and clean power load curve data [7], [8], [9]. Usually, status monitoring data for power equipment are treated as a time series. Through statistics or data mining, intervention models have been used to search for anomalous sequences [10]. Sliding a limited window across time series data has also been used to identify outliers [11]. However, in some cases, the length of abnormal data can vary in different periods. Therefore, a proper window size cannot be determined in advance. Intelligent algorithms such as neural networks, support vector machines, fuzzy algorithms and others methods have been proposed to improve data quality by cleaning outliers or missing data [12], [13], [14]. The implementation of these methods is complex, and engineering applications are limited.

The original information often contains different types of anomalous signals [15], [16]. The data cleaning model should be able to extract effective information from abnormal data.

Based on the deep learning idea [17], [18], in 2008, Vincent and Bengio proposed a denoising autoencoder (DAE) algorithm to reduce noise as one of the criteria for learning [19]. A DAE takes a corrupted version of the original input feed into the input layer. Thus, the DAE is trained to reconstruct the input from a corrupted version of it. This approach enhances the stability and robustness of the disturbance system. DAEs are stacked to form a hierarchical structure (stacked denoising autoencoder, SDAE) [20]. An SDAE is used to obtain more advanced features of the deep network. In the presence of "dirty" data, the repair ability of SDAE is particularly prominent.

In this paper, we propose a data cleaning method based on SDAE. The SDAE cleaning model is able to automatically identify and repair outliers and missing data. When the equipment is in the abnormal state, the model is applicable for the state data flow, and the abnormal operating state can be rapidly detected. This facilitates the status assessment and fault diagnosis of the equipment.

II. SDAE ALGORITHM

An autoencoder (AE) is a neural network that includes an input layer, a hidden layer, and an output layer [20]. An AE is structured by its encoder part and decoder part. The encoder maps an input vector into hidden representation. The decoder maps the hidden representation back to a reconstruction vector of the input vector. Training an AE is the process of minimizing the reconstruction error.

A DAE is very similar to an AE except for the input part. A DAE takes a corrupted version of the original input feed into the input layer [20]. The corrupted version $\tilde{x}$ of an input vector x is obtained by stochastic mapping, and $\tilde{x}$ is fed to the input and encoded according to (1).

$$y = f_\theta(\tilde{x}) = s(W\tilde{x} + b) \tag{1}$$

where y is the hidden representation, f_θ is the encode mapping, θ is the encoder parameters containing W and b, W is the weight matrix between the input and hidden layer, and b is the bias vector of the hidden layer.

The hidden representation y is decoded according to (2).

$$z = g_{\theta'}(y) = s(W'y + b') \tag{2}$$

where z is the reconstruction vector, $g_{\theta'}$ is the decode mapping, θ' is the decoder parameters containing W' and b', W' is the weight matrix between the hidden and output layer, and b' is the bias vector of the output layer.

The reconstruction vector z cannot reproduce the input x exactly. The reconstruction error is calculated by the following equation.

$$L(X, Z) = \begin{cases} H(B(x)|B(z)), & x \in \{0, 1\} \\ ||x - z||^2, & x \in \mathbb{R} \end{cases} \tag{3}$$

where X is a set of input vectors x, Z is the corresponding set of reconstructed vector z, H represents the Bernoulli cross entropy [21], and $B(x)$ and $B(z)$ are the mean values of x and z, respectively.

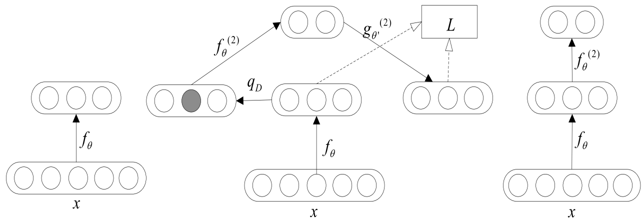


FIGURE 1. Stacking denoising autoencoders. After training a first level DAE, the learned encoding function f_θ is used on the clean input (left). The resulting representation is used to train a second level DAE (middle) to learn a second level encoding function $f_\theta^{(2)}$. From there, the procedure can be repeated (right).

When the reconstruction error is minimized, the optimal parameters are obtained. The common characteristic of the input vector x and the reconstruction vector z is maximized. However, during the training process, f_θ is the feature mapping of the corrupted data. Therefore, the DAE model is trained to achieve the denoising effect.

An SDAE is structured with stacked DAEs, as shown in Fig. 1 [20]. The SDAE can help to extract features from the input layer. The training algorithm is summarized as shown in Algorithm 1.

Algorithm 1 Training Stacked Denoising Autoencoder

Input: data set $X = \{x\}, x \in \mathbb{R}$
Output: reconstruction set $Z = \{z\}$, θ and θ'

Step 1) Set the training network parameters: the number of network layers L, pretraining iterations κ, learning rate α, weight-decay λ, number of visible nodes and hidden nodes and the fine-tuning iterations κ'
Step 2) Initialize the parameters θ and θ', x vectors were normalized to $\bar{x}$
Step 3) Pretraining
    Stochastic mapping: $\bar{x} \rightarrow \tilde{x}$
    for $j = 1$ to κ do
        for $i = 1$ to L do
            Perform forward propagation to compute Z
            y_{i} is fed to the $(i+1)th$ DAE
        end for
    end for
Step 4) Fine-tuning
    for $j = 1$ to κ'
        Compute the reconstruction error of output layer $\delta_{\mathbf{L}}$
        for $i = \mathrm{L} - 1$ to 1 do
            Compute the reconstruction error $\delta_{\mathbf{i}}$
        end for
        for $i = 1$ to L do
            update parameter
            $\nabla_{W_i} = \delta_i (f(\tilde{x}))^T$, $\nabla_{b_i} = \delta_i$,
            $W_i \leftarrow W_i - \alpha \nabla_{W_i} - \lambda W_i$, $b_i \leftarrow b_i - \alpha \nabla_{b_i}$
        end for
    end for

In this paper, we use the status monitoring data for power equipment as the input vector. In the training process, minimizing the reconstruction error is used as the tuning standard. When the model optimization parameters are used to clean data, the reconstruction error is a criterion for measuring the quality of the data. The output of SDAE is the reconstruction data corresponding to the input.

III. DATA CLEANING METHOD BASED ON SDAE

A. PRINCIPLE ANALYSIS

In normal running conditions, the status monitoring data of power equipment concentrate close to a non-linear low-dimensional manifold [20], [22], indicated by $(\times)$ in Fig. 2. Large measurement errors occur due to meter malfunction or system disturbance. The isolated outliers from the expected value are presented in the monitoring data. Similarly, due to communication failure and other unknown factors, some data are missing. These data deviate from the manifold distribution of normal data, indicated by $(\cdot)$ in Fig. 2. Parts of the input status monitoring data are randomly corrupted in the SDAE training process. For undamaged data x_i, the SDAE model learns through its deep structure and extracts its distribution characteristics. For the subsets of randomly corrupted data $\tilde{x}_i$, SDAE predicts the real value based on the uncorrupted value and extracts its implicit distribution characteristics. Therefore, the SDAE cleaning model has the ability to satisfy the probability distribution of the training sample. The SDAE model trained by the normal status data maps the distribution to the desired manifold or near the manifold. The reconstruction error satisfies the following relation: $||\tilde{x}_i - \tilde{z}_i||^2 \gg ||x_i - z_i||^2$, where $\tilde{z}_i$ is the reconstruction representation of $\tilde{x}_i$ and z_i is the reconstruction representation of x_i.

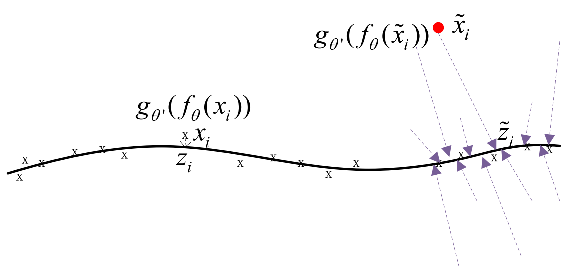


FIGURE 2. **Geometric interpretation of data cleaning.**

The incipient fault of the power equipment is usually reflected by changes in the monitoring status trend. For example, when the transformer iron core is grounding, the grounding current of the core gradually increases. As the fault severity increases, the grounding current increases from several milliamps to several hundreds or thousands of milliamps [23]. The SDAE parameters θ and θ' trained by the normal status data do not contain the characteristic mapping of the anomaly. If these parameters are used to clean the data of the abnormal running state, the reconstruction error will be too large within a long time range.

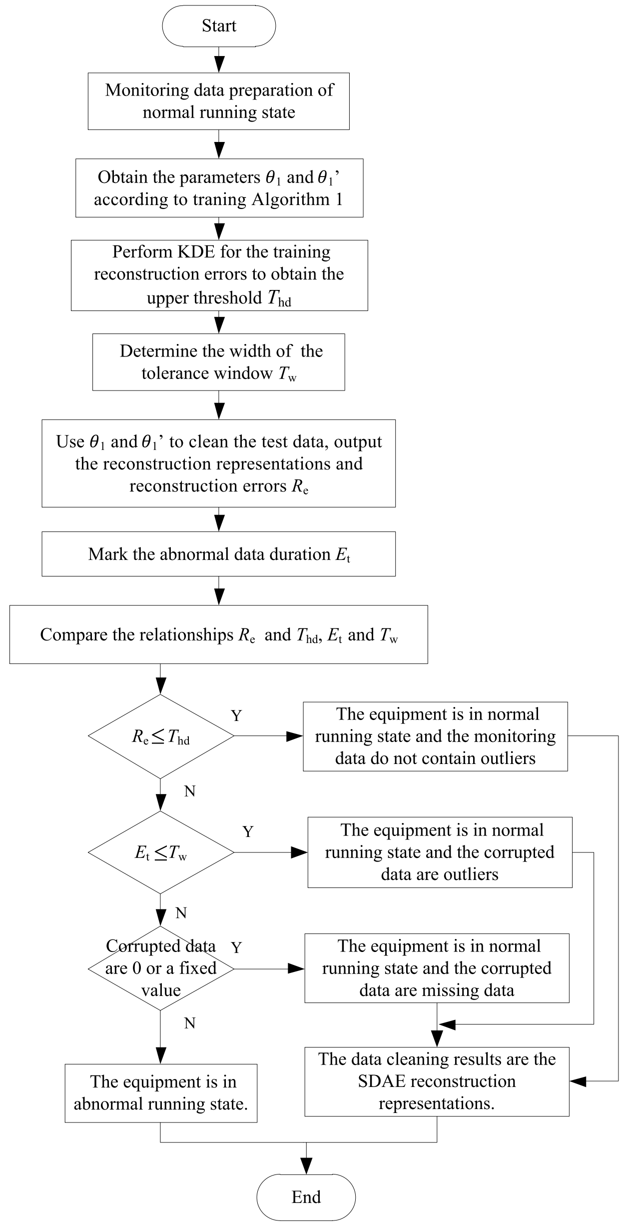


FIGURE 3. **Flowchart of the proposed method.**

According to the above analysis, the SDAE cleaning model is trained on the basis of the normal status data of power equipment. The reconstruction error and the large reconstruction error duration of the training samples are used as the evaluation criterion of the data type. An upper threshold of the reconstruction error and the tolerance window of large reconstruction error duration are set.

Samples whose corresponding reconstruction error exceeds the threshold and the large error duration within the tolerance window are identified as isolated outliers. Samples whose values are 0 or a fixed value and have large reconstruction error duration beyond the tolerance window are identified as missing data. These two types of data are considered the corrupted data in the normal status of the equipment. The SDAE model performs denoising based on the implicit distribution characteristics. Therefore, the SDAE reconstruction representation can be used as the cleaning value of the isolated outliers and the missing point. The trend in the samples is determined to be either increasing or decreasing, and the large reconstruction error duration beyond the tolerance window is identified. These samples are then identified as abnormal status data, and the cleaning model detects the abnormal state. This can be timely feedback to the maintenance staff, acting as a rapid anomaly monitoring effect. The model is not suitable for abnormal state data cleaning. Therefore, we keep the original data for follow-up research.

B. REALIZATION FOR DATA CLEANING

As described above, the SDAE cleaning details are illustrated in Fig. 3.

1) Perform the SDAE training according to Algorithm 1 for the normal status of the power equipment, and obtain the parameters θ and θ' of the cleaning model.

2) Calculate the reconstruction errors of the training samples from the cleaning model. Perform kernel density estimation (KDE) [24], [25] for the reconstruction errors, and determine the upper threshold of the reconstruction errors T_{hd} and the tolerance window T_{w}.

3) Use the SDAE model to clean the monitoring data. The reconstruction error R_{e} and the error duration E_{t} are obtained and compared with T_{hd} and T_{w}. The data type is determined by the following rules.

① $R_{\mathrm{e}} \leq T_{\mathrm{hd}}$: The data are in a normal status and do not contain "dirty" data. Therefore, this part of the data is uncorrupted.

② $R_{\mathrm{e}} > T_{\mathrm{hd}}$ and $E_{\mathrm{t}} \leq T_{\mathrm{w}}$: The data are in a normal status and contain "dirty" data, where the "dirty" data are isolated outliers.

③ $\mathrm{R_e} > T_{\mathrm{hd}}$, $E_{\mathrm{t}} > T_{\mathrm{w}}$ and data are 0 or a fixed value: The data are in a normal status and contain "dirty" data, where the "dirty" data are missing points.

④ $\mathrm{R_e} > T_{\mathrm{hd}}$, $E_{\mathrm{t}} > T_{\mathrm{w}}$ and the data show some trend (increasing or decreasing): The data are in an abnormal status.

4) For the data of cases ① ② ③ in 3), the reconstruction representation of the SDAE model is used as the replacing value.

5) In case ④ in 3), the data are kept raw because the cleaning models are not suitable for addressing this issue.

IV. CASE STUDIES

A. CLEANING DATA FOR THE DISSOLVED GAS CONCENTRATION IN TRANSFORMER OIL

Oil chromatographic data sets from an on-line monitoring device in a 220kV substation are used in our experiments. These data sets are daily observations of the gases dissolved in the transformer insulation oil for the three years from 2013 to 2015. In September 2015, the total hydrocarbon concentration exceeded the attention value [26], and the growth rate was high.

The total hydrocarbon data from 2013 to 2014 are used as the normal status samples X_{train} to train the model. The X_{train} values are learned by SDAE in accordance with Algorithm 1 to construct the cleaning model. The number of input neurons is 90. There are three hidden layers with 60, 40 and 60 neurons each. The learning epochs are 1000. The weights are initialized to small random values chosen from a zero-mean Gaussian with a standard deviation of approximately 0.01, and the biases are initialized to 0. The learning rate is set to 0.1, and the weight-decay is 0.0001. Based on experiments, it is found that when the corruption level is low (less than 5%), the repairing effect of outliers is excellent. However, the repairing effect of missing points is poor. When the corruption level is too high, the overall fitting accuracy of the sequence data is reduced. By adjusting the corruption level, we found that the model with a corruption level of 10% has high fitting accuracy and strong ability of repairing dirty data. So, the model assumes a corruption level of 10%.

To evaluate the training performance of the SDAE model, we use the mean absolute error (MAE) to measure the error of the training data.

$$MAE = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}|x_{ij} - z_{ij}|}{m \times n} \tag{4}$$

where m is the total number of input vectors and n is the dimension of each input vector.

The MAE of the measured value X_{train} and the reconstructed representation Z_{train} is 3.97%; thus, the absolute percentage accuracy is 96.03%. We can think that 3.97% data in the training set cannot be completely restored because of the large reconstruction error. Meanwhile, the background noises affect the accuracy to a certain extent. The confidence level is increased to avoid misinterpreting background noises as anomaly, so the confidence interval is set to 0.97. The cumulative distribution of reconstruction errors using KDE is shown in Fig. 4. The kernel estimator is a Gaussian type. The upper threshold R_{e} is 0.005 609 at the confidence level of 0.97, indicated by (■) in Fig. 4. Based on site visits and consultation with the onsite maintenance personnel, the tolerance window is set to 6.

The trained model is used to clean the monitoring data from 2015. To facilitate the presentation of these data, we select part of the data from 2015, named as X_{test} as an example.

We use the SDAE model to clean the normalized values of X_{test}, recording the reconstructed representations and reconstruction errors, as shown in Fig. 5. According to the cleaning procedure described in section 3.2, the types of abnormal data are determined as shown in Table 1.

At sample times of 12-13, 45 and 127, the abnormal data can easily be identified as outliers. At sample times of 85 to 92, the reconstruction errors are greater than the threshold, and E_{t} is beyond of the tolerance window. However, the data are a fixed value, so the data in this time period are missing. According to section 3.2, X_{test} is determined to be in a normal operation state, and the "dirty" data use

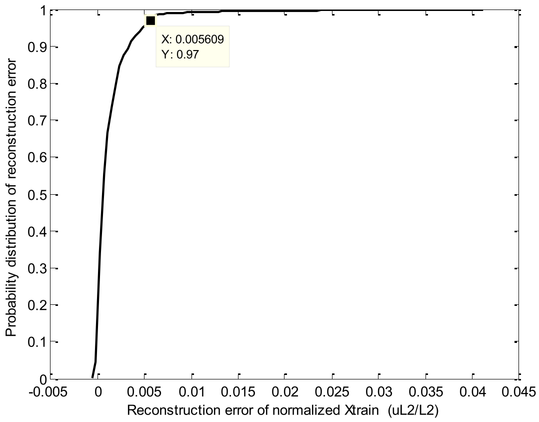


FIGURE 4. Distribution of reconstruction errors calculated by KDE.

TABLE 1. Abnormal data types of X_{test}.

Sample time	Abnormal data type
12-13	outlier
45	outlier
85-92	missing data
127	outlier

SDAE reconstruction as a data cleaning value. After the data are cleaned by the SDAE model, the data are consistent with the overall distribution characteristics. The background noises are dramatically reduced, and the data curve shows a smoothing distribution. The removal of pseudo-information and pseudo-trends rules can provide more realistic data for the follow-up state assessment. Compared with the proposed SDAE model, wavelet and s-transform denoising methods are used to clean X_{test}. In the wavelet transforms method, the db4 mother wavelet is used to decompose the data into three levels, and soft thresholding performs to remove the noise. At sample times of 12-13, 45 and 127, the wavelet method repairs the outliers to the desired manifold. However, at sample times of 85 to 92, wavelet transforms method cannot achieve the goal of data repairing. By s-transform denoising method, the data trend is cleaned smoothly, but the denoising effect at or near abnormal points is poor. S-transform method retains the abnormal information when the data changes dramatically. So it does not apply to clean the status monitoring data of power equipment.

The model is also used to clean the data of the total hydrocarbon concentration of this transformer from September, 2015. The reconstruction error abruptly increases, as shown in Fig. 6. According to section 3.2, the transformer is in an abnormal running state at this time.

In fact, during the overhaul of this transformer, the maintenance personnel found that the A-phase lead connector of the transformer on the high-voltage side had a few broken areas (shown in Fig. 7). When the power equipment is in an abnormal state, the cleaning model can achieve fault rapid detection.

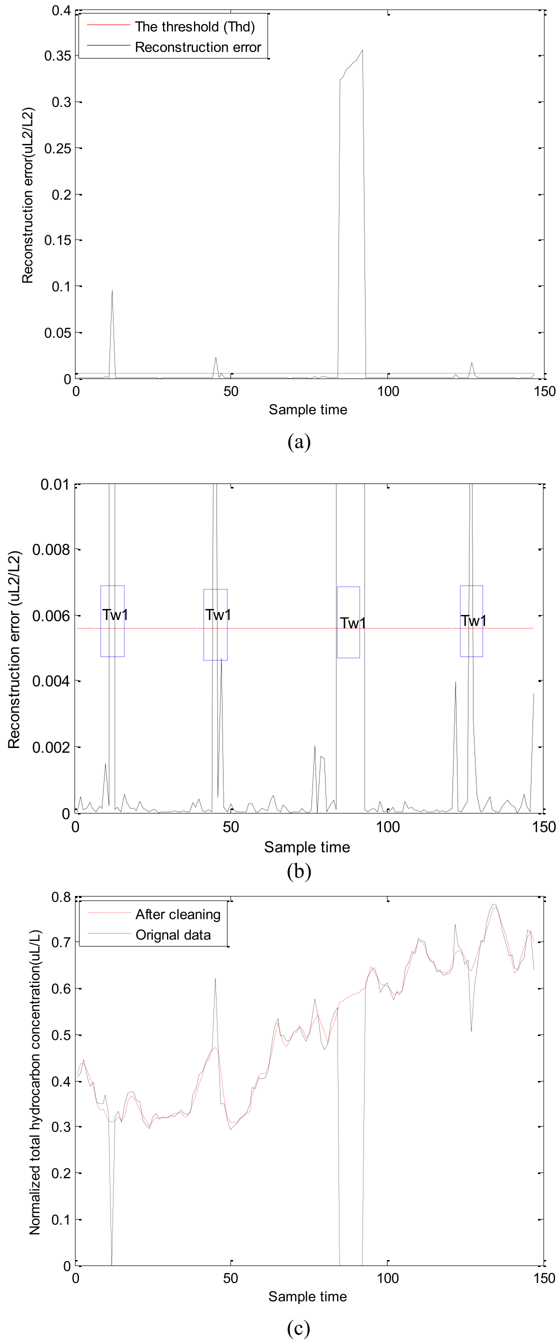


FIGURE 5. Cleaning X_{test} using the proposed method. (a) The relationship between the reconstruction errors R_e of X_{test} and the threshold T_{hd}. (b) The relationship between E_t and T_w through enlarging part of (a). (c) The cleaning result of X_{test}.

B. CLEANING THE TEMPERATURE DATA OF THE TRANSMISSION LINE

Next, we analyze the average temperature data of a 110 kV LGJX-300/40 transmission line. We use the temperature data X_{temp} to build the SDAE model. The number of input neurons is 144, and there are three hidden layers with 100, 72 and 100 neurons each. The learning epochs are 1000. The weights are initialized to small random values chosen from a zero-mean Gaussian with a standard deviation of

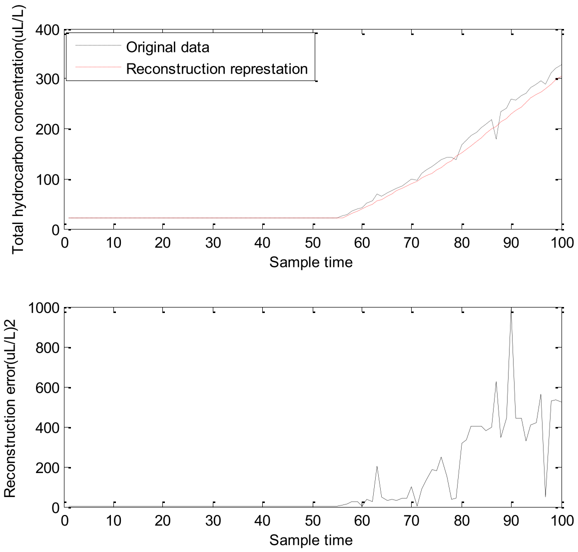


FIGURE 6. Results of using the model to clean the abnormal running status data.

FIGURE 7. A-phase lead connector of the transformer on the high-voltage side.

approximately 0.01, and the biases are initialized to 0. The learning rate is set to 0.1, and the weight-decay is 0.0001. Taking the fitting accuracy and reconstruction ability into account, the model assumes a corruption level of 10% based on experiments.

The MAE of the measured value X_{temp} and the reconstruction value Z_{temp} is 3.81%; thus, the absolute percentage accuracy is 96.19%. We can think that 3.81% data in the training set cannot be completely restored because of the large reconstruction error. Meanwhile, the background noises affect the accuracy to a certain extent. The confidence level is increased to avoid misinterpreting background noises as anomaly, so the confidence interval is set to 0.97. The cumulative distribution of the reconstruction errors using KDE is shown in Fig. 8. The kernel estimator is a Gaussian type. The upper threshold R_e is 0.002 456 at the confidence level of 0.97, indicated by (■) in Fig. 8. Based on site visits and consultation with the onsite maintenance personnel, the tolerance window is set to 6.

The temperature data are cleaned based on the SDAE model according to the steps presented in section 3.2 and Fig. 9. The types of abnormal data are determined as shown

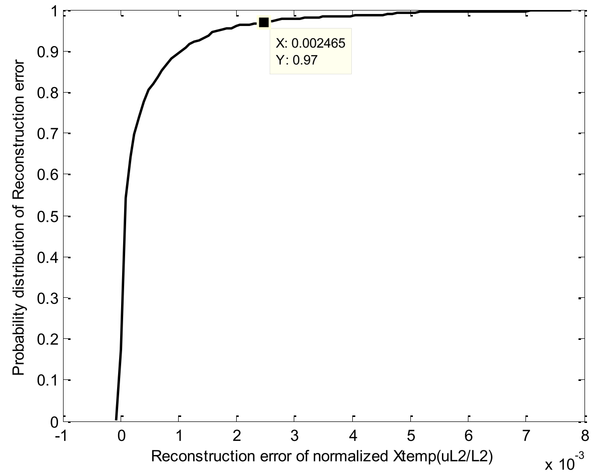


FIGURE 8. Distribution of reconstruction errors calculated by KDE.

TABLE 2. Abnormal data types of X_{temp}.

Sample time	Abnormal data type
8-9	outlier
30-32	missing data
82	outlier

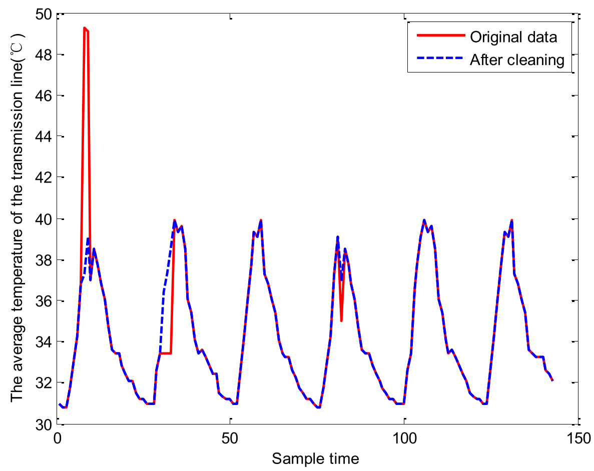


FIGURE 9. Cleaning results of the transmission line temperature data.

in Table 2. The test data are in a normal operation state. At sample times of 30, 31 and 32, there are missing data, and at sample times of 8, 9 and 82, there are outliers. The SDAE reconstructed representations are used as the cleaning results. The three points 33.45 °C, 33.45 °C, 33.45 °C at sample times of 30, 31 and 32 are corrected to 36.43 °C, 37.35 °C and 38.58 °C, respectively. The points 49.28 °C, 49.07 °C and 35.01 °C at sample times of 8, 9 and 82 are corrected to 37.27 °C, 39.08 °C and 37.00 °C, respectively. After cleaning the data using the proposed SDAE model, the data curve matches the actual distribution.

According to [27], the temperature model of the transmission line can be described as follows:

$$T = 3.57 * 10^{(-5)} I^2 - 0.11727I - 0.003705F + 0.13615v + 85.74814 \quad (5)$$

where T is the average temperature of the transmission line, I is the line load, F is the line tension, and v denotes the wind speed.

Based on the tension, load, and wind speed data from the on-line monitoring device, we compute the line temperatures at sample times of 8, 9, 30, 31, 32 and 82 to be 37.42 °C, 39.27 °C, 36.18 °C, 37.16 °C, 38.16 °C, 36.85 °C, respectively, which match the results of our approach.

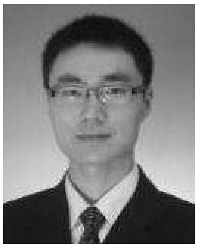
V. CONCLUSION

In this paper, we view the cleaning of power status monitoring data as a corrupted data recovery problem. Based on the ability of SDAE to effectively reconstruct "dirty" data and its significant abnormal status representation learning ability, we propose a deep learning model to clean data. A deep network is first constructed with the normal operating state data using SDAE. KDE is then used to determine the upper threshold of the reconstruction errors. A tolerance window width is set and compared with the abnormal data duration. By comparing the relationship between the reconstruction error and the upper threshold and between the tolerance window and the duration of abnormal data, the abnormal data are classified as outliers, missing data or abnormal running state data. The outliers and missing data are repaired by the SDAE reconstructed representations. When the equipment is defective or fails, the cleaning model rapidly detects the anomaly. The model not only cleans series monitoring data but also provides a new method for detecting anomalies in power equipment.

REFERENCES

[1] R. Hussein, K. BashirShaban, and A. H. El-Hag, "Denoising of acoustic partial discharge signals corrupted with random noise," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 23, no. 3, pp. 1453–1459, Jun. 2016.
[2] X. Chen, Y. Qian, Y. Xu, G. Sheng, and X. Jiang, "Energy estimation of partial discharge pulse signals based on noise parameters," *IEEE Access*, vol. 4, pp. 10270–10279, Jan. 2017.
[3] Q. Wang, D. Kundur, H. Yuan, Y. Liu, J. Lu, and Z. Ma, "Noise suppression of corona current measurement from HVdc transmission lines," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 2, pp. 264–275, Feb. 2016.
[4] Y. Yan, G. Sheng, Y. Liu, X. Jiang, X. Sun, and Y. Sun, "A method for fast mining abnormal state information of power equipment based on time series analysis," in *Proc. IEEE PES Gen. Meet. Conf. Expo.*, Jul. 2014, pp. 1–5.
[5] R. Misra, A. Baral, and A. Lahiri, "Denoising neutral current of a power transformer measured during impulse test by framelet technique," in *Proc. IEEE 10th Int. Conf. Properties Appl. Dielectr. Mater.*, Jul. 2012, pp. 1–6.
[6] Q. Wang, D. Kundur, H. Yuan, Y. Liu, J. Lu, and Z. Ma, "Smart transformer for smart grid—Intelligent framework and techniques for power transformer asset management," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 1026–1034, Mar. 2015.
[7] J. Chen, W. Li, A. Lau, J. Cao, and K. Wang, "Automated load curve data cleansing in power systems," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 213–221, Sep. 2010.
[8] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*. New York, NY, USA: Wiley, 2006.
[9] Z. Guo, W. Li, A. Lau, T. Inga-Rojas, and K. Wang, "Detecting X-outliers in load curve data in power systems," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 875–884, May 2012.
[10] C. Brighenti and M. Á. Sanz-Bobi, "Auto-regressive processes explained by self-organized maps. Application to the detection of abnormal behavior in industrial processes," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2078–2090, Dec. 2011.
[11] E. Keogh, J. Lin, and A. Fu, "HOT SAX: Finding the most unusual time series subsequence: Algorithms and applications," in *Proc. ICDM*, 2005, pp. 440–449.
[12] M. Yang and J. Ma, "Data completing of missing wind power data based on adaptive BP neural network," in *Proc. PMAPS*, Oct. 2016, pp. 1–6.
[13] W. Shi *et al.*, "Improving power grid monitoring data quality: An efficient machine learning framework for missing data prediction," in *Proc. HPCC-CSS-ICESS*, Aug. 2015, pp. 417–422.
[14] X. Yang, G. Zhang, J. Lu, and J. Ma, "A kernel fuzzy c-means clustering-based fuzzy support vector machine algorithm for classification problems with outliers or noises," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 1, pp. 105–115, Feb. 2011.
[15] S. Meng, L.-T. Huang, and W.-Q. Wang, "Tensor decomposition and PCA jointed algorithm for hyperspectral image denoising," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 7, pp. 897–901, Jul. 2016.
[16] C. Varon, C. Alzate, and J. A. K. Suykens, "Noise level estimation for model selection in kernel PCA denoising," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 11, pp. 2650–2663, Nov. 2015.
[17] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[18] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.
[19] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103.
[20] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengioand, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, 2010.
[21] I. Sason, "On the entropy of sums of Bernoulli random variables via the Chen–Stein method," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2012, pp. 542–546.
[22] T. Shnitzer, R. Talmon, and J. J. Slotine, "Manifold learning with contracting observers for data-driven time-series analysis," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 904–918, Feb. 2017.
[23] W. H. Kersting and W. Carr, "Grounded wye-delta transformer bank backfeed short-circuit currents," *IEEE Trans. Ind. Appl.*, vol. 53, no. 1, pp. 65–70, Jan. 2017.
[24] S. Wang, J. Wang, and F. L. Chung, "Kernel density estimation, kernel methods, and fast learning in large data sets," *IEEE Trans. Cybern.*, vol. 44, no. 1, pp. 1–20, Jan. 2014.
[25] D. Lu, Z. Bao, and Z. Li, "Load sampling for SCUC based on principal component analysis and kernel density estimation," in *Proc. IEEE PES Gen. Meet.*, Jul. 2016, pp. 1–5.
[26] *IEEE Guide for the Interpretation of Gases Generated in Oil-Immersed Transformers*, IEEE Standard C57.104-2008, 2008.
[27] J. L. Ren, "Dynamic overhead transmission line rating based on tension measurement," (in Chinese), Ph.D. dissertation, School Electron., Inf. Elect. Eng., Shanghai Jiao Tong Univ., Shanghai, China, 2008.

JIEJIE DAI was born in Shandong, China. She received the B.E. degree from Shandong University, Jinan, China, in 2009, the M.E. degree from Beihang University, Beijing, China, in 2012. She is currently pursuing the Ph.D. with Shanghai Jiao Tong University, Shanghai, China. Her interests include the condition monitoring of power apparatus.



HUI SONG was born in Shandong, China. He received the B.E. degree from Shandong University, Jinan, China, in 2009, and the M.E. degree from Shanghai Jiao Tong University, Shanghai, China, in 2012, where he is currently pursuing the Ph.D. degree. His interests include condition monitoring, partial discharge, and insulation diagnostics.

GEHAO SHENG (M'08) was born in Hunan, China. He received the B.E., M.S., and Ph.D. degrees in electric power system and automation from the Huazhong University of Science and Technology, Wuhan, China, in 1996, 1999, and 2003, respectively. From 2003 to 2005, he was a Post-Doctoral Researcher with the Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China. He is currently a Professor with the Department of Electrical Engineering, Shanghai Jiao Tong University. His research interest is the condition monitoring of power apparatus.

XIUCHEN JIANG was born in Shandong, China. He received the B.E. degree in high voltage and insulation technology from Shanghai Jiao Tong University, Shanghai, China, in 1987, the M.S. degree in high voltage and insulation technology from Tsinghua University, Beijing, China, in 1992, and the Ph.D. degree in electric power system and automation from Shanghai Jiao Tong University in 2001. He is currently a Professor with the Department of Electrical Engineering, Shanghai Jiao Tong University. His research interests include electrical equipment online monitoring and condition-based maintenance and automation.