*Artificial Intelligence Technologies and Applications*
*C. Chen (Ed.)*


*doi:10.3233/FAIA231388*

# A Remote Out-of-Tolerance Classification Method of Gateway Electric Energy Meters Based on Feature Selection and Deep Belief Network

Chong LI[1], Hao WANG, Chuan LI, Hongtao SHEN, Junlong WANG, Juchuan GUO
*Marketing service center of State Grid Hebei Electric Power Co., Ltd, Shijiazhuang 050035, China*

**Abstract.** Accurate identification of the out-of-tolerance of the gateway electric energy meter plays a vital role in ensuring the stability of the new power system and maintaining the fairness of the transaction. As a key metering device in power generation enterprises, network connections between power grids, inter-provincial network connections, and other metering devices, the importance of checkpoint energy measurement is self-evident. Therefore, studying the causes of metering errors and the classification of outlier checkpoint metering devices is significant. A remote out-of-tolerance gateway electric energy meters classification method based on feature selection and deep belief network is proposed. This paper first uses Spearman and Pearson correlation coefficients to select the main influencing factors of checkpoint energy metering error and then trains a deep belief network model to complete the detection of outlier checkpoint energy metering. Finally, experimental results verify the superiority of the proposed method.

**Keywords.** classification;feature selection;deep belief network

## 1. Introduction

Research on key technologies for online monitoring of checkpoint metering device status for new power systems can enable the evaluation of the performance of electric energy metering devices under actual operating conditions and establish a new management and operation mechanism for checkpoint metering devices [1]. This helps managers timely and accurately grasp the actual operating status of checkpoint metering devices and solve problems in existing work, such as high safety risks, intense periodic calibration work, poor timeliness of anomaly detection, and low degree of information [2].

The current research on checkpoint energy meters is relatively limited. Sun et al. [3] propose a method for identifying abnormal measurement data of checkpoint energy meters based on pseudo-anomalous point identification, which can effectively identify abnormal data. During the operation of the checkpoint meter, they will cause certain measurement errors. The article uses a high sampling rate ADC and selects an appropriate voltage range to provide a basis for improving the accuracy and reliability of

[1] Correspondent Author: Chong LI, Marketing service center of State Grid Hebei Electric Power Co., Ltd, Shijiazhuang 050035, China; email: chunglee3181@126.com

the checkpoint meter. Chen et al. [4] propose an estimation method based on dual forgetting factors. This method is based on the characteristics of actual power consumption data and the characteristics of the parameters that need to be estimated. It introduces a dynamic dual forgetting factor recursive least squares (RLS) method and uses real-time information to adjust the estimation results repeatedly. It can adapt to situations where the speed of change in multiple parameter estimation is different, and the method's effectiveness has been verified through numerical simulations.

To solve the problems, a remote out-of-tolerance classification method of gateway electric energy meters based on feature selection and deep belief network is proposed. This article first uses Spearman and Pearson correlation coefficients to select the main influencing factors of checkpoint energy metering error and then trains a deep belief network model to complete the detection of outlier checkpoint energy metering. Finally, experimental results verify the superiority of the proposed method.

## 2. A Remote out-of-tolerance Classification Method of Gateway Electric Energy Meters Based on Feature Selection and Deep Belief Network

### *2.1. Feature Selection Method Based on Spearman and Pearson correlation Coefficients*

The influencing factors of the error in checkpoint energy meters are divided into two categories. The first category includes six intrinsic factors: voltage, current, power factor, commissioning time, economic properties of the monitored checkpoint, and the unified batch of checkpoint metering over-limit rate. The second category includes four extrinsic factors: temperature, humidity, air pressure, and holiday types, which can affect the parameters such as voltage and current and consequently affect the meter's error.

A feature selection process was implemented using Spearman and Pearson correlation coefficients to characterize the correlations between these ten factors and the over-limit state of checkpoint energy meters (no over-limit as 0, with over-limit as 1). The complex correlation relationships among the ten parameters during the operation of checkpoint energy meters needed to be identified to eliminate data redundancy and correlation. A single correlation coefficient cannot objectively characterize the close relationship between data variables. Hence, a comprehensive correlation coefficient was selected in this study. The Spearman correlation coefficient was used with the Pearson correlation coefficient to characterize the correlation relationships among checkpoint energy meter state data variables. The formula for the Spearman correlation coefficient $\rho_1$ is as follows:

$$\rho_1 = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)} \tag{1}$$

$\rho_1$ is the Spearman correlation coefficient between two variables; $N$ is the sample size; $d$ is the difference between the rank of the two corresponding data points.

The formula for calculating the Pearson correlation coefficient $\rho_2$ is shown below:

$$\rho_2 = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu_X)(X-\mu_Y)}{\sigma_X \sigma_Y} \tag{2}$$

In the formula, $X$ and $Y$ represent different influencing factors for the over-limit state of checkpoint energy meters; cov($X$, $Y$) represents the covariance; $\sigma_Y$ and $\sigma_Y$ represent the standard deviation of $X$ and $Y$, respectively. $\mu_X$ and $\mu_Y$ represent the mean value of $X$ and $Y$, respectively.

The comprehensive correlation coefficient between state parameters is represented by $r_{xy}$. It can be described as follows:

$$r_{xy} = \frac{1}{2}(\rho_1 + \rho_2) \tag{3}$$

When the comprehensive correlation coefficient $r_{xy}$ between the over-limit state influencing factors of checkpoint energy meters is greater than or equal to the threshold $\theta$ set by humans, the factor will be retained. Otherwise, it will be removed.

After feature selection, a sequence of over-limit state influencing factors for checkpoint energy meters with high correlation with the over-limit state will be obtained, denoted by $x=\{x_1, x_2, \ldots, x_c\}$, where $c$ is the number of over-limit state influencing factors of checkpoint energy meters after feature selection.

### *2.2. Data Preprocessing*

Considering the influences of different dimensions on the data, the data was normalized before being fed into the Deep Belief Network (DBN), using selected features including voltage, current, power factor, operating time, and temperature. As these parameters have varying numerical scales and units [5], to accelerate the training speed of the network, a normalization function was used to preprocess the input data, scaling them to [-1, 1]. The normalization formula is as follows:

$$v_i = 2 \times \frac{x_i - x_{i,\min}}{x_{i,\max} - x_{i,\min}} - 1 \tag{4}$$

In the formula, $x_i$ represents the sequence of influential factors of the energy meter after feature selection. In contrast, $x_{i,\max}$ and $x_{i,\min}$ represent the maximum and minimum values of $x_i$. The normalized parameters of the energy meter were then arranged into a matrix v, represented by the following expression:

$$v = \begin{bmatrix} v_1 & v_2 & \cdots & v_c \end{bmatrix}^{\mathrm{T}} \tag{5}$$

### *2.3. Construction of an Energy Meter Abnormality Classification Model Based on a Deep Belief Network (DBN)*

An Energy Meter Abnormality Classification Model based on Deep Belief Network

(DBN) was established to determine if energy meters were abnormal. The DBN consisted of Restricted Boltzmann Machines (RBM) and fully-connected layers. RBMs, which consist of hidden and visible layers that are fully connected in an all-to-all manner, are seen as feature extractors that can effectively extract features from input data. The DBN structure is shown in Figure 1.

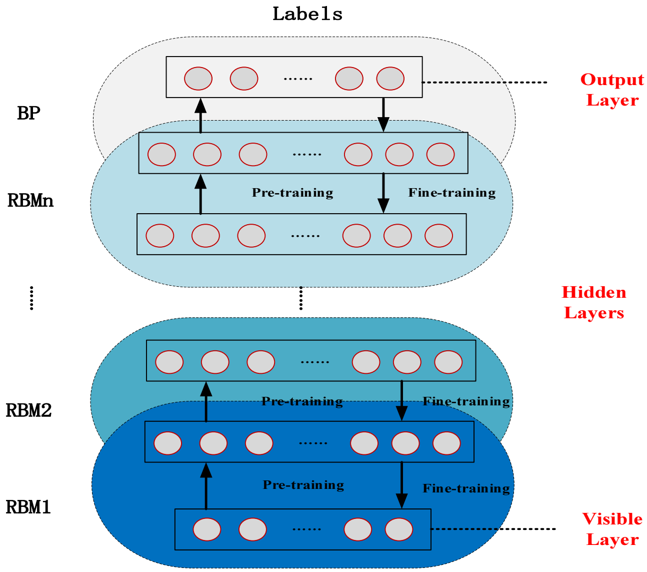


**Figure 1** DBN Structure Diagram

In this study, the RBM had *n* visible units and *m* hidden units, with the normalized and matrix-formatted energy meter abnormality factors *v* as input, and the features extracted by the RBM layer were denoted as *h*. The energy of the RBM layer was defined as follows:

$$\mathrm{E}(v,h\,|\,\theta) = -\sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{m} b_j h_j - \sum_{i=1}^{n}\sum_{j=1}^{m} v_i W_{ij} h_j \tag{6}$$

$$w_{i,j} = \begin{bmatrix} w_{1,1} & w_{2,1} & \cdots & w_{i,1} \\ w_{1,2} & w_{2,2} & \cdots & w_{i,2} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1,j} & w_{2,j} & \cdots & w_{i,j} \end{bmatrix} \tag{7}$$

where $\theta = \{\mathrm{W_{ij}}, a_i, b_j\}$ is the parameter of the RBM, $v_i$ represents the normalized input of the energy meter data, $h_j$ represents the states of the *j-th* hidden layer neuron after visible layer input, $W_{ij}$ indicates the weight of each visible unit neuron *i* and hidden unit neuron *j*, $a_i$ represents the bias term associated with visible unit neuron *i*, and $b_j$ represents the bias term associated with hidden unit neuron *j*. With the parameters determined, based on this energy function, the joint probability density distribution of the input data and the hidden layer $(v,h)$ can be obtained:

$$p(v,h\mid\theta)=\frac{e^{-E(v,h|\theta)}}{Z(\theta)},Z(\theta)=\sum_{v,h}e^{-E(v,h|\theta)} \tag{8}$$

The energy function conforms to the Boltzmann distribution, where $Z(\theta)$ is the normalization factor representing the sum of energy of all possible conditions. With well-trained weights, the distribution of probabilities of correctly identifying the energy meter data matrix in the visible layer can be obtained. That is to say, the marginal distribution can be calculated.

$$p(v\mid\theta)=\frac{1}{Z(\theta)}\sum_{h}e^{-E(v,h|\theta)} \tag{9}$$

The activation probability of the $j^{th}$ hidden unit is given by:

$$p(h_j=1\mid v,\theta)=\sigma(b_j+\sum_i v_iW_{ij}) \tag{10}$$

In the equation, σ denotes the activation function, $p$ is the activation probability, $b_j$ represents the bias term associated with the j-th hidden layer neuron, $v_i$ is the normalized input data after being matrix-formatted, $W_{ij}$ refers to the weight of each visible unit neuron i and hidden unit neuron j.

The activation probability of the *i-th* visible unit is given by:

$$p(v_i=1\mid h,\theta)=\sigma(a_i+\sum_j W_{ij}h_j) \tag{11}$$

In the equation, $a_i$ represents the bias term associated with the *i-th* visible unit neuron, $h_j$ denotes the state of the *j-th* hidden unit neuron representing the matrix data of the electricity meter at the interface after being input into the visible layer, and other symbols have the same meanings as before.

We use the output of the last layer of RBM as the input of the fully connected layer:

$$m_i=\sigma(l_i+\sum_j k_{ij}p_j) \tag{12}$$

$$k_{i,j}=\begin{bmatrix} k_{1,1} & k_{2,1} & \cdots & k_{i,1} \\ k_{1,2} & k_{2,2} & \cdots & k_{i,2} \\ \vdots & \vdots & \ddots & \vdots \\ k_{1,j} & k_{2,j} & \cdots & k_{i,j} \end{bmatrix} \tag{13}$$

$$Z_i=\sigma(l_i+\sum_{i=1}^{n}k_{ij}m_i) \tag{14}$$

In the equation, $k_{ij}$ denotes the weight matrix of the fully connected neural network, $l_i$ represents the bias term, and $Z_i$ represents the network output. The parameters are updated continuously using the gradient descent method. Figure 2 shows model topology diagram.

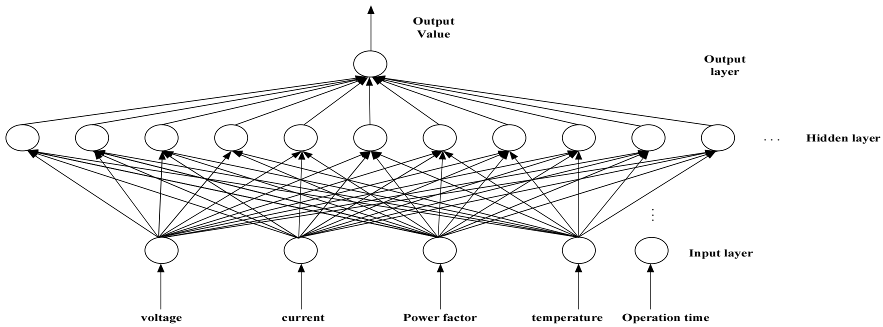


**Figure 2** Model Topology Diagram

After building the network, the processed normal and faulty electricity meter data are separately fed into the model for training. RBM pre-training can generate very good initial parameter values and reduce the difficulty of model learning. After pre-training layer by layer, the error back propagation algorithm adjusts the discriminative performance. The collected factors affecting electricity meter faults are input into the trained DBN model to obtain the final judgment result. The formula is as follows:

$$Z=\sigma(l_i+k_{ij}m_i) \tag{15}$$

where $Z$ represents the judgment result of electricity meter faults, for which 0 denotes no fault and 1 represents a fault.

## 3. Experiment and Analysis

In this paper, some historical monitoring data of meter bases in a metrology center in 2022 were taken as an example to identify the operating status of electricity meters using two methods. The Matthews correlation coefficient (MCC) was used to compare these methods with the BP method. Table 1 shows the results:

**Table 1** Algorithm Comparison (MCC)

| Algorithm type | MCC |
|---|---|
| Proposed method | 0.7501 |
| BP neural network algorithm | 0.5911 |

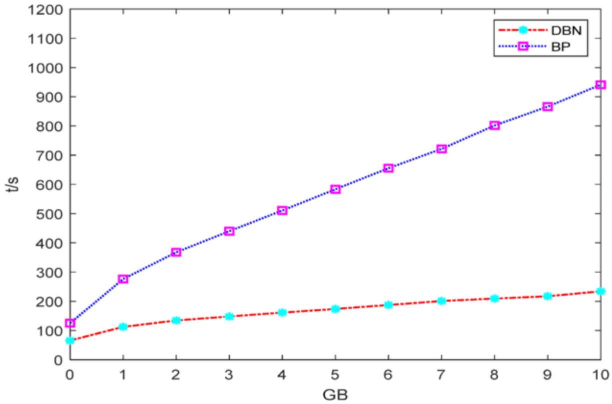


**Figure 3** Processing time comparison chart

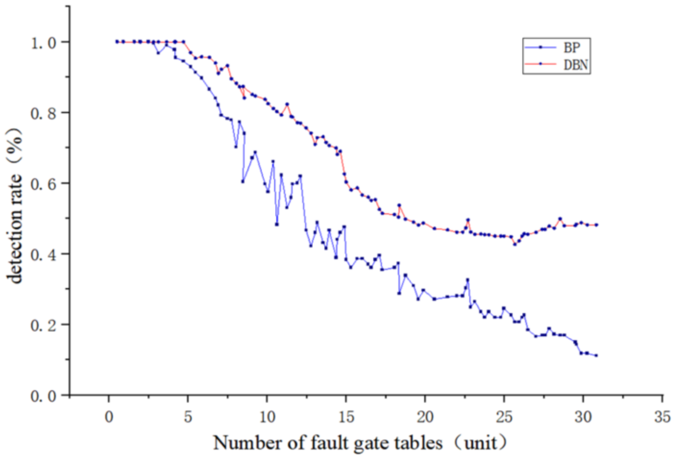


**Figure 4** Algorithm detection rate

Figure 3 and Figure 4 show that the deep belief network algorithm performs better than the BP neural network in terms of processing speed and the fault detection rate of malfunctioning meters. Compared with the BP model, the accuracy of the out-of-tolerance classification of the gateway electric energy meter of the DBN model is increased by 21.20%. Moreover, we use the DBN network results in a more stable detection rate of electricity meters with faults, which depends on the appropriate data preprocessing and the characteristics of the DBN network.

## 4. Conclusion

A remote out-of-tolerance gateway electric energy meters classification method based on feature selection and deep belief network is proposed. This paper first uses Spearman and Pearson correlation coefficients to select the main influencing factors of checkpoint energy metering error and then trains a deep belief network model to complete the detection of outlier checkpoint energy metering. Compared with the BP model, the accuracy of the out-of-tolerance classification of the gateway electric energy meter of the DBN model is increased by 21.20%. Experimental results show that the DBN network is feasible for error diagnosis of boundary meters and has a faster computation speed and better detection rate for metering errors than the BP neural network.

## Acknowledgment

This paper was supported by the State Grid Science Technology Project “Research on key technologies of gateway metering device status on-line monitoring for new power system” (kj2022-034).

## References

[1] Zakaret C, Peladarinos N, Cheimaras V, et al. Blockchain and Secure Element, a Hybrid Approach for Secure Energy Smart Meter Gateways [J]. Sensors, 2022, 22(24): 9664.
[2] Chen K, Deng S, Li W. Measurement error analysis and accuracy improvement method based on gateway meter [C]//Journal of Physics: Conference Series. IOP Publishing, 2022, 2355(1): 012004.
[3] Śmiałkowski Tomasz,Czyżewski Andrzej. Detection of Anomalies in the Operation of a Road Lighting System Based on Data from Smart Electricity Meters[J]. Energies,2022,15(24).
[4] Chen Kaiming, Deng Siyang, Li Wenjing. Measurement error analysis and accuracy improvement method based on gateway meter [J]. Journal of Physics: Conference Series, 2022, 2355: (1).
[5] Zander M,E. H B D V,Gertjan K, et al. Reduction of Static Electricity Meter Errors by Broadband Compensation of Voltage and Current Channel Differences[J]. IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT,2021,70.